Embracing randomness: within- and between-host evolution of two RNA viruses

by

Katarina M. Braun

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Cellular and Molecular Biology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2021

Date of final oral exam: 25 May 2021

The dissertation is approved by the following members of the Final Oral Committee:

Thomas C. Friedrich, Professor, Pathobiological Sciences

Caitlin Pepperell, Associate Professor, Medicine and Medical Microbiology and Immunology

Shelby O'Connor, Associate Professor, Pathology and Laboratory Medicine

Bret Payseur, Professor, Genetics

Andrew Mehle, Associate Professor, Medical Microbiology and Immunology

Table of Contents

Acknowledgments

The path to a PhD is a windy one. It requires persistence, curiosity, resilience, determination, and, probably most importantly, the mentorship and support of quite a few people. It feels appropriate to begin this dissertation document with a statement of recognition and gratitude to all of those people.

To my thesis committee – Tom, Andy, Shelby, Caitlin, and Bret: thank you all for your unwavering support and guidance along the journey that is graduate school. Andy, thank you for your positivity, your infectious curiosity, and your patient guidance on all of the barcoded flu projects. Shelby, thank you for your sequencing advice, your mentorship, and for giving me the push I needed to wrap this thing up. Caitlin, thank you for helping me work through my crazy F30 ideas and for modeling the type of physician-scientist I hope to become. Bret, thank you for your mentorship and for helping me discover my passion for evolutionary biology. Tom, you have been a phenomenal mentor, a kind and thoughtful advisor, a model scientist, a consistent cheerleader, and a friend. I am so grateful you agreed to take me on as a graduate student, even though I was a little reserved and unsure at the beginning. You have given me some pretty incredible opportunities to explore interesting science, to pursue projects longer than I probably should have, to travel to present my science, and to pivot when it made sense. Thank you for pushing me when I needed to be pushed and encouraging me to pull back when things felt overwhelming and, most importantly, for always having my back.

To the Friedrich lab – you have all been incredible mentors, colleagues, and friends and I feel lucky to have been a part of a lab that is so supportive and filled with talented and caring individuals. Gabrielle and Andrea, I knew absolutely nothing when I started in lab and you both taught me so much and provided endless support along the way. Macy and Emma, thank you for helping me fumble through learning to be a mentor. It was wonderful to watch you both grow as scientists. Louise, thank you for your mentorship, support, and your words of encouragement even after you moved on to your post-doctoral training. You are an inspirational scientist and I hope to continue to learn from you moving forward. Luis, Katie, and Kasen, thank you for your camaraderie, commiseration, and friendship! Gage, you are an incredible scientist and a dear friend and I am so grateful to have had the opportunity to work with you and learn from you over the past year. I can't wait to see what you do next. Also, thanks for keeping me young, throwing fishies at me, and teaching me that I truly love Taylor Swift. Chelsea, I could not have done it without you. Thank you for a*lways* being there for me. You patiently and skillfully supported me through anxiousness, insecurity, and fear when I needed it and excitement and joy when I needed that. You are an exceptional scientist, leader, and advocate and I look forward to seeing what you do next.

To AVRL and collaborators – I am grateful to have been a part of the AVRL community, which has always been a welcoming and energizing environment to work in. Thank you to everyone at AVRL who has provided support and guidance along the way (including the AVRL dog family). Thank you to all of the collaborators who made so many of my projects possible. Katia, thank you for helping me focus my research questions, for the

beta-binomial model (which I used in almost all of my papers), and for helping me to elevate my science.

To my MSTP family – thank you Anna, Michael, Alex, Jon, and Zach for your camaraderie and support along this crazy MD-PhD journey. I could not have asked for a class of people more inspiring, brilliant, or kind than you all are. Anna, thank you for encouraging me to pursue the gender project with you. I don't think I would have seen that through to completion without your friendship and encouragement.

To MSTP leadership – Anna H, JP, Mark, Caitlin, Scott, Jeniel, and Elizabeth - you have all been incredible and inspirational mentors and leaders and constant sources of encouragement and advice. Chelsea and Nichole (and Paul), your wealth of knowledge and your organizational skills are unparalleled and the MSTP is lucky to have you.

Thank you to the animals and people who have contributed samples and data to the advancement of science generally and to the projects I was involved in throughout my graduate training. Science does not progress without these contributions so I express my humble gratitude to each of you.

To my friends – thank you, April, Josh, Dave, Liza, Alissa, Nicki, and Lauryn. Thank you for being incredible friends, I couldn't have done it without you guys!

To my family – thank you, Mom, Dad, Tor, Xander, Mimi, and the pack of dogs. Thank you for your unwavering support and encouragement. Thank you for reminding me to take breaks, to keep the “big picture” in mind, and to maintain balance because the journey is as important as the destination. Thank you to Sally, Steve, Katie, Eli, Owen, Scarlet, TJ, Ali, and Neva for your support!

To my partner – Dan. More than anyone, you know how hard, and exciting, and exhausting, and daunting, and amazing this experience has been. You have supported me through every failure and setback and have celebrated every success, including the tiny ones. Thank you for being an incredible listener, for having too much confidence in me, for giving me the time and space that a PhD requires, for always being there when I needed to unwind, for helping me stay grounded, for providing comic relief, and for pursuing your own dreams at the same time. I am so lucky to have a partner like you and I could not have done it without you.

Abstract

Pathogenic RNA viruses emerging from zoonotic reservoirs are among the highest threats for global infectious disease control. Every single major epidemic or pandemic in the 21st century has resulted from an emerging or re-emerging zoonotic RNA virus. Severe Acute Respiratory Syndrome virus 1 (SARS-CoV) emerged in 2003, a novel pandemic H1N1 influenza virus in 2009, Middle East Respiratory Syndrome virus (MERS-CoV) in 2012 and 2015, Ebola in 2014, Zika virus in 2015, Yellow fever virus in 2016, and SARS-CoV-2 in 2019.

It is clear the primary drivers of the emergence of these zoonotic RNA viruses are increasing globalization, habitat fragmentation, and encroachment of a continuously growing human population into wildlife habitats [1]. It is notable that this increased interaction between humans and animals likely increases the risk of interspecies transmission among a large number of potential pathogens, yet RNA viruses are the dominant source of emerging human pathogens [2]. The capacity for RNA viruses to rapidly adapt to new host environments and to respond to shifting selective pressures is not completely understood. Current dogma suggests this trait is tied to short generation times and high mutation rates resulting from error-prone viral replication. RNA virus mutability creates diverse viral populations which are more capable than homogenous populations of adapting to new hosts and host environments [3]. However, the generation of viral variation is only the first step. Individual mutations that confer fitness benefits in particular environments must then increase in frequency and/or make their way out of individual hosts and into populations. This stage presents several obstacles that the virus must

overcome and is therefore likely to be rate-limiting for the overall pace of viral evolution and host-switching.

The first three chapters (**chapters 2-4**) of this dissertation focus on investigating the *evolutionary* processes by which zoonotic RNA viruses adapt to mammalian hosts. The results of this work call attention to several significant obstacles that zoonotic RNA viruses must overcome in order to successfully and efficiently emerge in and adapt to human hosts. I suggest these obstacles all derive from the effects of randomness on viral systems. The cumulative impact of these obstacles has critical implications in assessing the pandemic potential of viruses that have already caused human epidemics, like avian influenza viruses, and the adaptive potential of the current pandemic virus, SARS-CoV-2.

The final two chapters (**chapters 5-6**) of this dissertation discuss our work combining principles of viral evolution with epidemiology and population health to investigate the early patterns of SARS-CoV-2 spread in the state of Wisconsin.

Taken together, this work suggests the effects of randomness on viral populations within and between individual hosts are a previously underappreciated brake to the pace of viral adaptation and host-switching for influenza A virus (IAV) and SARS-CoV-2. Additionally, this work underscores the value of genomic epidemiology early in a pandemic to understand patterns of viral transmission in different populations and to assess the impact of public health guidelines and interventions on a rolling basis

Chapter 1:

Introduction

The emergence of zoonotic viruses is one of the greatest threats to global health security. More than half of all known human pathogens can be traced to a zoonotic source [2,4,5]. The list of known zoonotic viral pathogens is likely a very small fraction of the total possible viral pathogens currently lurking in animal reservoirs. Even as the global population continues to contend with the staggering loss of human life that has resulted from explosive emergence and spread of SARS-CoV-2, we must also anticipate the yet-to-emerge pathogens.

It is clear that a successful and sustainable global pathogen surveillance network will require a One Health approach focused on maintaining healthy human populations, healthy animal populations, and balanced ecosystems [6,7]. Although surveillance systems to detect novel pathogens in high-risk populations and other One Health initiatives continue to improve and expand [8,9], we still do not understand the processes underpinning the emergence and reemergence of zoonotic RNA viruses [10]. The work in this dissertation focuses on the evolutionary processes contributing to and impeding zoonotic RNA virus emergence and adaptation to human populations. Specifically, I focus on Influenza A virus, which has emerged multiple times from its avian reservoir to cause both sporadic epidemics and multiple pandemics, as well as SARS-CoV-2, which emerged from its zoonotic reservoir in 2019 and is responsible for the current global pandemic.

Overview of Avian Influenza Viruses, focusing on H7N9

The influenza virus is composed of eight negative-sense, RNA gene segments coated with nucleoprotein and associated with a heterotrimeric polymerase complex packaged with a protein capsid and a host-derived envelope. Influenza A virus's (IAV) named subtypes are defined by its two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA), which are responsible for host cell entry and exit, respectively. HA binds host receptors, sialic acids, to trigger receptor-mediated endocytosis, virion uncoating, and release into the cytoplasm following acidification of the endosome. NA is a sialidase and cleaves the HA-sialic-acid bond during viral exit, allowing progeny virions to be released from the host cell following replication. In total, there are 18 known IAV HA subtypes (H1-H18) and 10 known NA subtypes (N1-N10).

Aquatic birds (the *Anseriformes* and *Charadriiformes* orders) are the natural reservoir for avian influenza viruses, harboring a vast viral population and substantial viral diversity [11]. Interestingly, H17, H18, and N10 have been exclusively identified in bats [12]. The avian IAV reservoir is particularly diverse because different virus subtypes co-infecting a single host can swap gene segments and be co-packaged together into a new progeny virus through a process called reassortment [13]. Reassortment has been shown to occur at a high rate between viruses belonging to the same subtype [14–16], viruses belonging to different subtypes [17], and even among viruses from different host species [18]. Indeed, most HA-NA constellations have been detected while sampling the avian reservoir and in environmental samples (watersheds) [19], suggesting reassortment is commonplace in the

avian reservoir and is acting to continuously shuffle existing diversity to generate new virus combinations.

The capacity of influenza to undergo reassortment has important implications for host switching because this process can create novel viruses to which the human population has no prior immunity, called antigen shift. Notably, three of the influenza virus pandemics (1957: H2N2, 1968: H3N2, 2009: H1N1) over the past century are hypothesized to have arisen, at least in part, from reassortment events between avian viruses and swine or human viruses [20,21].

Only three IAV subtypes have achieved sustained human-to-human spread (H3N2, H1N1, H2N2) and two of these now circulate seasonally as “seasonal influenza” (H1N1 and H3N2). Seasonal influenza epidemics cause 3 million to 5 million severe cases and 300,000 to 500,000 deaths globally most years [22]. In a typical year, the United States alone sees 140,000 to 710,000 influenza-related hospitalizations and 12,000 to 56,000 deaths [23], with the highest burden of disease affecting the very young, the very old, and people with coexisting medical conditions. Interestingly, influenza infections and deaths recorded during the 2020-2021 season were historically low and this was attributed to the public health interventions in place to reduce the spread of SARS-CoV-2 (masking, distancing) [24].

A subset of fully avian influenza viruses has occasionally “spilled over” into human populations in the absence of reassortment and without achieving sustained human-to-

human spread, including H5N1, H5Nx, H9N2, H7N7, and H7N9 viruses. Highly pathogenic avian H5N1 virus was detected among domestic geese in China in 1996 and in the first human in Hong Kong one year later [25]. Sporadic H5N1 infections throughout Southeast Asia have since followed, totaling 862 infections (from January 2003 to 15 April 2021) and 455 deaths (53% case fatality rate) [26]. H7N9 AIVs have been endemic in chickens since the virus's emergence in China in February 2013 [27]. Since then, H7N9 viruses have caused 1,568 confirmed human infections with a case fatality rate approaching 40% across six epidemic waves [28]. Although the avian reservoir harbors extensive viral diversity, relatively few subtypes have ever been identified in human hosts, and an even smaller subset has caused epidemics or pandemics. The key factors underlying these phenotypic differences remain unknown. Relatedly, there are no reliable methods to assess the "pandemic potential" of avian IAVs sampled from birds, environmental sources, and human spillover infections.

Overview of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

SARS-CoV-2 is one of seven coronaviruses known to infect humans [29]. It is a member of the family *Coronaviridae* and from the genera *betacornavirus*. *Betacoronaviruses* exclusively infect mammals, including humans, and cause a range of respiratory and gastrointestinal diseases [29,30].

Four *betacoronaviruses* cause mild upper respiratory illnesses and are common etiologic agents of the "common cold". These are HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 [31,32]. Three more concerning *betacoronaviruses* have emerged from their

animal reservoirs since the start of the 21st century. Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) emerged in China in 2002-2003 from a yet unknown animal reservoir. Middle East respiratory syndrome (MERS) emerged from dromedary camels in Saudi Arabia in 2012. SARS-CoV-2 emerged in 2019, likely from bats [33], and as of 6 May 2021, there have been more than 155 million confirmed cases of coronavirus disease 19 (COVID-19), the disease caused by SARS-CoV-2, including 3,250,648 deaths [34].

SARS-CoV-2 is primarily transmitted via droplet and airborne transmission [35,36]. The clinical presentation associated with SARS-CoV-2 varies widely from asymptomatic to multi-system inflammatory syndrome and death. The most common clinical signs are fever, fatigue, and cough [37–40] and the most specific signs are ageusia–loss of taste, and anosmia–loss of smell, which only occur in around 5% of COVID-19 infections [41,42]. The highest proportion of cases in the United States has generally been in the 18-to-24 year age group, although death rates are dramatically higher in individuals >60 and especially in individuals >80 years [43]. The incubation period ranges from 1-14 days and the average time from infection to symptom onset is 5 days [44,45].

SARS-CoV-2 is a positive-sense, single-stranded, enveloped virus. Spike (S) is the surface protein that binds the host receptor, angiotensin-converting enzyme 2, which is expressed on respiratory epithelium [46–49]. Transmembrane serine protease 2 (TMPRSS2) and/or furin [50–53] are required to proteolytically activate Spike to facilitate host-cell entry. Once intracellular, viral RNA is replicated and translated into proteins and enzymes which will be packaged into newly produced viral particles. The first two-thirds of the SARS-

CoV-2 genome is occupied by two open reading frames (ORF1a, ORF1b), which encode 15-16 non-structural proteins (nsp), including nsp14, which provides crucial 3'-5' exonuclease activity to proofread the copied RNA, maintain the integrity of the 30kb genome, while driving down the mutation rate [54–59]. Coronaviruses have the largest genome of any RNA virus and Nsp14 is highly conserved within the Coronaviridae family [60]. This suggests that the RNA proofreading function is important to the maintenance of the large-for-an-RNA-virus genome size and to ensure ongoing replication competence [61,62].

Viruses evolve in individuals and populations

In order to evolve, RNA viruses must first generate genetic variation, or mutations. Like all RNA viruses, IAV and SARS-CoV-2 have high mutation rates: 1.8×10^{-3} - 2.28×10^{-3} [63–65] and 2.28×10^{-3} [66,67] substitutions/site/year, respectively. Other sources of viral variation include cytidine deaminases [68] and recombination, although the impact of homologous recombination in IAV is negligible [69]. Coronaviruses, on the other hand, do appear to shuffle genetic diversity through homologous recombination as their replication strategy includes the generation of subgenomic RNAs, which predispose to recombination events [70]. Virus generation times, the time between virus entry to virus production, are typically very short, so the generation of viral diversity via RNA polymerases and homologous recombination is compounded with each generation [71,72]. The fate of individual mutations, broadly speaking, then depends on a combination of deterministic factors, like selection, as well as stochastic (random) processes, like genetic drift [73]. Selection acts to increase the frequency of beneficial mutations (positive selection)

or decrease the frequency of deleterious mutations (negative or purifying selection). Genetic drift is stochastic changes in allele frequencies due to the effects of random sampling. The relative contributions of selection and genetic drift depend largely on the effective population size, which is a useful construct that corresponds to the number of individuals in a population that contribute mutations or genetic variation to the next generation [73,74]. Specifically, selection is strongest in large populations and genetic drift is strongest in small populations. The cumulative effect of these processes can be partially reflected by the substitution rate, or the rate at which viral mutations achieve fixation in a population. IAV's substitution rate is approximately twice that of SARS-CoV-2 (1 mutation per 7 days vs 15 days), which is likely tied to SARS-CoV-2's proofreading exonuclease activity.

At first glance, RNA viruses have everything you need for rapid and deterministic evolution: a high mutation rate, short generation times, and large population sizes. Despite this, adaptive evolution of viruses within individual hosts has rarely been observed and very few zoonotic RNA viruses have successfully emerged in human populations, suggesting there are barriers to efficient within-host evolution that are not yet understood.

While deterministic evolution does not appear to play a significant role at the level of individual hosts, positive selection has been shown to be a major driver of influenza evolution on a global scale [75–77]. The evolutionary dynamics that transform viral variation within individual hosts into global genetic diversity are poorly understood.

Virological barriers to viral adaptation: avian IAV and SARS-CoV-2

One of the primary determinants of a virus's ability to infect a new host or host species is the accessibility and distribution of the receptor required for host cell entry. Influenza virus's surface protein, hemagglutinin, binds sialic acid receptors to mediate virus entry. Avian IAV binds α-2,3-linked sialic acid receptors, which are distributed along the avian gastrointestinal tract [78]. To improve infection and transmission in mammalian hosts, avian IAV can switch or broaden its receptor repertoire to include α-2,6 linked sialic acids, which predominate in the mammalian upper respiratory tract. A mixture of α-2,3 and α-2,6-linked sialic acids can be found in the mammalian lower respiratory tract, although this compartment is less accessible to airborne viruses. Although we typically dichotomize sialic acid receptors into avian-type (α-2,3) and mammalian-type (α-2,6), sialic acid chains vary widely in length and orientation along most internal epithelial surfaces [79,80]. It is generally believed that α-2,6 binding is required for efficient respiratory droplet transmission and pandemic spread, [81,82] although the 1918-H1N1 virus, H5N1, and H7N9 exhibit the capability to dually bind α-2,3 and α-2,6-linked sialic acid receptors [83,84].

SARS-CoV-2 uses the human angiotensin I converting enzyme (ACE2) as a primary receptor for host cell entry. Spike is cleaved into two subunits (S1 and S2) at a polybasic furin cleavage site (RRAR), which triggers a conformational change and facilitates ACE2 recognition and binding as well as S2 fusion with the host membrane. Unfortunately, ACE2 is relatively ubiquitous and conserved across a large number of mammals, making it a useful receptor for SARS-CoV-2 to expand its host range [85,86]. Interestingly, though,

the furin cleavage site in spike is not present in SARS-CoVs infecting bats or pangolins, which likely harbored the most recent ancestor to SARS-COV-2 [87]. This suggests that changes to the spike glycoprotein may be one of the major determinants of host range and cell- and tissue tropism for SARS-CoV-2. Similarly, the addition of a furin cleavage site to influenza hemagglutinin has been shown to convert a low-pathogenic avian influenza virus to a more highly pathogenic phenotype [88].

In addition to receptor preference, host restriction factors, co-factors, and conditions in the host environment (i.e. humidity and temperature) can all pose significant barriers to host adaptation. One well-known example of this for influenza is in the stability of the polymerase heterotrimer (PB1, PB2, and PA) and its functional association with viral RNA and the viral nucleoprotein to form the ribonucleoprotein (RNP) complexes that mediate RNA-dependent RNA polymerase activity. IAV viral replication in humans requires a lysine at position 627 in PB2 while glutamic acid predominates in most avian viruses. The lysine at this position appears to play a role in mediating virus temperature sensitivity [89] as well as interactions with human-specific restriction factors [90–92]. Accordingly, PB2 E627K enhances viral replication and pathogenicity in mammalian systems [93–95] and has been identified in the majority of viruses successfully isolated and sequenced from H7N9 human spillover infections [96].

Host-specific restriction factors limiting the replication and transmission of SARS-CoV-2 remain unknown but may include viral inhibition of a robust interferon (IFN)-1 and IFN-III response [49].

Immunological barriers to viral adaptation: avian IAV and SARS-CoV-2

To achieve robust infection, a virus must also contend with preexisting and a host's cross-protective humoral and cell-mediated immunity.

Antibodies against influenza HA have long been known to be protective, although often variably protective, against influenza virus infection and severe disease outcomes [97–99]. Antibodies against influenza HA do not typically offer lasting protection because HA, specifically the globular head, is highly plastic and tolerates new point mutations, which can interfere with antibody recognition [100] and/or enhance receptor binding [101] through a process called antigenic drift. Though less well-studied, T-cell responses significantly contribute to protection against influenza virus infection as well as the induction of robust antibody responses [102].

Seasonal influenza viruses are ubiquitous; most individuals are exposed to their first flu virus in childhood [103]. This first exposure is typically followed by subsequent infections and, often, yearly vaccination so the depth and breadth of pre-existing influenza immune profiles are varied and unique to each individual. Interestingly, the subtype and lineage of an individual's first infection may play an outsized role in shaping their lifelong immune profile by inhibiting de-novo antibody responses against subsequent divergent viruses, a phenomenon called "original antigenic sin" [104] and, more recently, "imprinting" [105]. This phenomenon has been shown to play a role in avian influenza virus spillover infections as well. A study by Gostic et al [106] showed age-dependent differences in disease severity

of H5N1 and H7N9 spillover infections and hypothesized these differences could be explained by group match/mismatch between the first virus infecting an individual and the subsequent spillover infection. Specifically, individuals infected with an H5 virus (group 1) were afforded some immune protection if their first influenza infection was likely an H1N1 seasonal virus (group 1), using birth year as a proxy. A similar protective effect was seen between the H3N2 birth-year group and H7N9 spillover infections (group 2), suggesting an individual's pre-existing immune repertoire likely impacts the clinical outcomes and selective pressures involved in subsequent infections [106].

Less is known about the immune response to SARS-CoV-2, but infection does appear to elicit innate and adaptive responses. Pathogen recognition receptors (PRRs) present on immune cells non-specifically recognize the virus and lead to an IFN cascade [107]. Humoral responses initially involve responses to the N protein, followed by the S protein [108]. Seasonal coronavirus IgM and IgG antibodies do not appear to recognize SARS-CoV-2 epitopes, but SARS-CoV epitopes are cross-reactive with SARS-CoV-2. Antibodies generated following SARS-CoV-2 infection [109] and following vaccination [110] afford significant protection from reinfection, although the duration of this protection remains unclear. In addition, multiple "variants of concern" (VOC) have begun to emerge (in fall 2020) and a subset of these reduce the ability of antibodies to neutralize virus (though not to the level that would render them inefficacious) [111] and another subset increases transmissibility [112,113]. The impact that these VOCs will have on the natural history and selective pressures on SARS-CoV-2 remains unknown.

Evolutionary barriers to viral adaptation: IAV and SARS-CoV-2

RNA viruses, including IAV, are often described as “rapidly evolving”, due to their fast generation times, frequently large and genetically diverse populations, and high mutation rates. Virologists and infectious disease scientists therefore expect beneficial mutants, such as antigenic escape variants, to enjoy a marked fitness advantage and rapidly become dominant. While little is known about SARS-CoV-2, IAV mutants can rapidly adapt to selective pressures in cell culture [114–116]. Likewise, on the global scale, seasonal IAV evolution is characterized by antigenic drift, in which escape variants rise to high frequency in the human population and initiate subsequent epidemics [76,117,118]. Perplexingly, however, there is little evidence of adaptive evolution of IAV within individual humans [119–123].

Rare instances of within-host IAV adaptive evolution have been described in prolonged infections in immunocompromised individuals and in atypical cases of drug resistance [124–126]. Together these observations suggest that positive selection (natural selection favoring new mutations) is weak in individual IAV-infected hosts. Adaptive variants and escape variants must arise within individuals and make their way out of individuals before spreading through a population. This raises the question: why is IAV antigenic evolution detectable in cell culture and in human populations, but not at the level of individual hosts? Perhaps a clue lies with a recent influenza household cohort study that determined that genetic drift, i.e. stochastic processes, could explain the observed patterns of within-host diversity [122]. In support of this idea, another study found the emergence of drug-resistant IAV in cell culture was inhibited by pervasive genetic drift when virus titers were low [116].

These studies suggest the role of genetic drift in slowing IAV evolution on the scale of individual hosts underappreciated. Identifying the principal barriers impeding the preservation and transmission of adaptive avian IAV and SARS-CoV-2 will improve predictive models for vaccine design as well as for ongoing epidemic and pandemic surveillance efforts.

A number of factors likely contribute to inefficient adaptive evolution within individual viral infections. The vast majority of the IAV and SARS-CoV-2 genomes are coding, so the vast majority of new mutations will range in phenotypic impact from mildly deleterious to lethal, which limits viral evolutionary flexibility. Further, mutations do not occur in isolation so the phenotypic impact of any mutation also depends on any co-occurring (linked) mutations and their combined impact [127].

Even when a beneficial mutation or haplotype, a group of mutations in linkage, arises, its growth within the viral population is far from certain. The fate of a new mutation depends on two factors: (1) the strength of its advantage (selection coefficient, s) and (2) the effective population size (Ne). Ne is the number of genetically distinct variants reproducing in a population. Note that Ne is distinct from census size, the overall number of individuals: an infection initiated by 10 genetically identical viruses has a census size of 10 and a Ne of 1. Ne correlates inversely with the strength of natural selection and is directly proportional to genetic drift – small populations are more susceptible to the effects of genetic drift than large populations. In a haploid population, the deterministic effect of selection will remain negligible until $2 \times Ne \times s$ exceeds 1 [128]. In the case of spillover

viruses not fully adapted to mammalian hosts, viral replication may be inefficient, and even beneficial mutations could be lost by chance if the viral population size is not sufficiently large; this is genetic drift.

Finally, even if a beneficial variant arises and achieves consensus frequency (majority) in a viral population, onward transmission can be very challenging. In most RNA virus systems, between host transmission corresponds with a dramatic reduction in population size, called a transmission bottleneck [129–132]. Narrow transmission bottlenecks cause a founder effect and purge low-frequency intrahost single nucleotide variants (iSNVs), regardless of their fitness. Conversely, wide transmission bottlenecks allow more viruses to initiate infection, reducing the chance that beneficial or rare variants are lost at the time of transmission. For example, in humans, airborne transmission of seasonal influenza viruses has been shown to involve a narrow transmission bottleneck, with new infections founded by as few as 1-2 genetically distinct viruses [122,133–136]. 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) [137]. When transmission involves the transfer of very few variants, 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, the transmission of these variants has rarely been observed in nature [120,138]. In this way, narrow transmission bottlenecks are generally expected to slow the pace of seasonal IAV adaptation [139,140] and may have similar effects on avian IAV and SARS-CoV-2. Overall, understanding the

size of the transmission bottleneck is important for evaluating the probability that novel variants arising within an individual host infection will be transmitted onward.

An overview of genomic epidemiology

Viral deep sequencing is the workhorse method involved in assessing viral diversity, which is needed to analyze viral evolution within hosts. Sequencing viruses collected from groups of infected individuals can additionally be used to reconstruct virus family trees (phylogenies), which can be used to infer paths of transmission. The practice of combining viral genomic data with epidemiological data is referred to as “genomic epidemiology”. According to the CDC, genomic epidemiology is “the use of pathogen genomic data to determine the distribution and spread of an infectious disease in a specific population and the application of this information to control health problems" [141]. Genomic epidemiology first emerged as a critical tool while reconstructing patterns of viral spread and evaluating the efficacy of public health strategies during the Ebola and Zika outbreaks [142–152], and has more recently played a critical role in the fight against SARS-CoV-2. Viral sequencing can augment traditional methods of contact tracing. Specifically, because the SARS-CoV-2 genome acquires new mutations at a relatively constant rate, mutational patterns can be used to make inferences about likely paths of transmission within populations [54–59,153]. Additionally, the mutational differences contributing to a set of viral sequences (reflected in a phylogeny) can be used to infer key parameters like the basic reproductive number R_0, an important measure of the transmission potential of a virus [154]. Such an analysis would allow you to evaluate the rate of viral spread and, with

longitudinal data, the effectiveness of a public health intervention in a given population [155,156].

A brief outline of the chapters to follow

Chapter 2 discusses our characterization of within- and between-host diversity of H7N9 avian influenza viruses replicating in a mammalian model – the ferret. We find that H7N9 within-host diversity is under purifying selection in ferrets, variants are rarely transmitted onward, and the transmission bottleneck is even narrower for H7N9 viruses than H1N1 viruses. We find no evidence of natural selection favoring new or mammalian-adaptive mutations within ferrets or arising during transmission. These findings suggest that purifying selection, randomness, and very narrow bottlenecks combine to severely constrain the ability of H7N9 viruses to effectively adapt to mammalian hosts in typical spillover infections, even with onward airborne transmission.

In **chapter 3**, we evaluate the forces shaping SARS-CoV-2 viral evolution within and between hosts a different mammalian model – the domestic cat. Similar to our findings in chapter 2, we show that SARS-CoV-2 genetic variation is predominantly influenced by genetic drift and purifying selection within individual hosts and by narrow transmission bottlenecks between hosts. In addition, we identify a notable variant at amino acid position 655 in Spike (H655Y) that arises rapidly in transmitting cats, persists at intermediate frequencies in these cats, and becomes fixed following transmission in two of three transmission pairs. This same variant has been shown to confer escape from human

monoclonal antibodies and circulates in humans on multiple SARS-CoV-2 genetic lineages.

In **chapter 4**, we investigate whether novel SARS-CoV-2 variants arise and transmit efficiently among acutely infected humans. Employing a comprehensive approach to exclude spurious, pipeline-induced variants, we find that very limited variation is generated and transmitted during acute SARS-CoV-2 infection. Most infections in our dataset are characterized by fewer than 5 total iSNVs, the majority of which are low-frequency. Most iSNVs are not detected in global consensus genomes and are rarely detected in downstream branches on local or global phylogenetic trees. Even among putative household transmission pairs, iSNVs are shared infrequently, and we estimate that a small number of viruses found infection after most transmission events (a narrow transmission bottleneck). The combination of low within-host diversity, tight transmission bottlenecks, and infrequent propagation along transmission chains may slow the rate of novel variant emergence among acutely infected individuals.

While the emergence of novel variants of concern should be monitored closely, our data suggest that rapid accumulation of novel mutations within-host is not the norm during acute infection. Like influenza viruses, a significant portion of variation generated within a host is likely lost during transmission. This implies that even if novel, beneficial variants are generated de novo, these variants are unlikely to spread beyond that individual. When this process is expanded across the global population, the combination of limited diversity

within-host and tight transmission bottlenecks should slow the pace at which novel, beneficial variants could emerge.

Chapter 5 discusses our use of genomic epidemiology to characterize the initial SARS-CoV-2 outbreaks in the two most populous counties in Wisconsin, Dane and Milwaukee Counties. These two counties provided a “natural experiment” to understand the impact of the “Safer at Home” Executive Order on within- and between-county SARS-CoV-2 transmission in two nearby US counties with distinguishing demographic features. We show that despite their proximity, distinct viral lineages drove each county’s outbreak. The number of viral introductions into each county differed as well. In addition, we show that the “Safer at Home” order decreased viral spread in both counties by at least 40%. These findings have implications for the application of targeted public health guidance.

Finally, **chapter 6** discusses our investigation into the most common source of infection in healthcare personnel at a major academic medical institution in the Upper Midwest of the United States during spring, 2020. We use viral sequencing to show that healthcare personnel were most likely to become infected with SARS-CoV-2 through community exposure rather than patient contact. These findings support the success of the CDC’s infection control guidelines to protect healthcare personnel. These findings also underscore the importance of ongoing measures to reduce community spread through mask-wearing, physical distancing, robust testing programs, and rapid vaccine distribution.

Chapter 2:

Stochastic processes constrain adaptation of wildtype H7N9 avian influenza viruses to mammalian hosts

Katarina M. Braun[1], Chelsea Crooks[1], Luis Antonio Haddock III[1], Gabrielle L. Barry[1], Joe Lalli[1], Gabriele Neumann[1,2], Tokiko Watanabe[2], Yoshihiro Kawaoka[1,2], Thomas C. Friedrich[1]

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

[2] Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, Japan

Abstract

H7N9 influenza viruses have caused over 1,500 human spillover infections and can be transmitted by respiratory droplet in ferrets. Although these viruses seem poised to adapt to humans and cause widespread outbreaks, no such event has occurred. Critical insights have been offered regarding the molecular determinants restricting successful host-switching of avian H7N9 viruses to human hosts, but little is known about the evolutionary constraints. To address this, we deep sequence time series swabs collected from 23 ferrets infected with H7N9, including seven transmission events. We compare these findings against nine ferrets infected with seasonal H1N1, including four transmission

events. We find that H7N9 within-host diversity is under purifying selection in ferrets, variants are rarely transmitted onward, and the transmission bottleneck is even narrower for H7N9 viruses than H1N1 viruses. We find no evidence of natural selection favoring new or mammalian-adaptive mutations within ferrets or arising during transmission. Together, these findings suggest that purifying selection, randomness, and very narrow bottlenecks combine to severely constrain the ability of H7N9 viruses to effectively adapt to mammalian hosts in typical spillover infections, even with onward airborne transmission.

Introduction

The potential emergence of a novel avian influenza virus in humans poses a significant public health and economic threat [20,157–159]. Despite significant advances in influenza surveillance and forecasting [160–162], we still do not understand the evolutionary processes underlying the emergence of pandemic influenza viruses [157,159]. H7N9 avian influenza viruses (AIVs) naturally circulate in aquatic birds and have been endemic in chickens since the virus's emergence in China in February, 2013 [27]. Since then, H7N9 viruses have caused 1,568 confirmed human infections with a case fatality rate approaching 40% across six epidemic waves [28]. During the fifth and largest epidemic wave, some low-pathogenicity avian influenza (LPAI) H7N9 viruses acquired a novel motif in hemagglutinin (HA) which both facilitates systemic virus replication in chickens and enhances pathogenicity in mammals [163–167]; these viruses are designated highly pathogenic avian influenza (HPAI) viruses. Many posit and are concerned ongoing human spillover infections may facilitate mammalian adaptation of H7N9 AIVs, eventually resulting in a mammalian-transmissible, and therefore pandemic, virus.

High pandemic potential is currently assigned to both H7N9 and H5Nx AIVs [138,158,168–173]. H7N9 viruses appear particularly threatening because, unlike H5N1 viruses, H7N9 viruses can be transmitted between ferrets via respiratory droplet without first needing to acquire mammalian-adapting mutations [170,174,175]. In addition, H7N9 viruses bind human-type receptors, in which sialic acids are linked to galactose in an **α**(2,6) pattern [135,170]. It is therefore unclear why, despite causing over 1,500 human spillover infections, there are no documented cases of human-to-human transmission of H7N9 viruses [176]. Some have speculated that the lack of human-to-human transmission can be explained by a number of factors including H7N9 residual binding to avian-type (**α**(2,3)) sialic acid receptors, viral fusion occurring at a higher-than-optimal pH for human-transmissible viruses, reduced polymerase activity at the human upper respiratory tract temperature (33 °C), and variability in HA glycosylation patterns [135,177–180]. Another study cites limited within-host diversity of LPAI H7N9 viruses in ferrets, compared to chickens, as a potential barrier to rapid mammalian adaptation [135]. Although these studies offer critical insights for the molecular determinants restricting host-switching of H7N9 viruses to mammalian hosts, it remains unclear whether additional evolutionary barriers further restrict mammalian adaptation and transmissibility of LPAI and HPAI H7N9 viruses.

In 2017, Dr. Kawaoka's group characterized the replication and pathogenicity of H7N9 viruses in ferrets [170]. Using time series samples originally collected in this study, we sought to investigate the evolutionary dynamics of LPAI and HPAI H7N9 avian influenza viruses replicating and transmitting in a mammalian system. We performed whole-

genome deep sequencing in duplicate and evaluated H7N9 population dynamics in seven ferret transmission events and in an additional nine infections not resulting in transmission. Importantly, we compare the viral genetic diversity of these AIVs in a mammalian system to seasonal H1N1 in four ferret transmission events and an additional non-transmitting infection [170,181]. In contrast to our initial predictions, we find no evidence for mammalian adaptation in ferrets and postulate that stochastic forces play a significant role in limiting avian influenza virus host-switching. We conclude the evolutionary barrier to emergence of an H7N9 AIV capable of sustained spread in humans is quite high. We speculate pandemic preparedness resources might be best directed toward ongoing poultry vaccination [182,183], safety regulations for wet markets [184], and ecological restoration [185] as opposed to broad surveillance for particular AIV lineages or variants.

Materials and methods

Ferrets transmission experiments & sample collection and availability

No new transmission experiments were performed as part of this study. We took advantage of nasal wash samples collected from ferrets participating in 2017 and 2020 studies conducted by Imai and colleagues to assess the transmissibility of H7N9 viruses in mice, ferrets, and non-human primates [170,181]. In these previously-described studies, four groups of four ferrets were directly inoculated with various H7N9 viruses (1×10^6 pfu) and one group of two ferrets was infected with an H1N1pdm seasonal virus for comparison (inoculated or index ferrets). The H7N9 viruses included a high-pathogenic human isolate – A/Guangdong/17SF003/2016 (“GD/3”), two recombinant viruses which possess arginine or lysine at position 289 (H7 numbering) to confer neuraminidase-

inhibitor sensitivity or resistance, respectively, on the background of the GD/3 consensus sequence – rGD/3-NA289R and rGD/3-NA289K ("rGD/3"), and a low-pathogenic H7N9 virus – A/Anhui/1/2013 ("Anhui/1"). The H1N1 comparator group was infected with a representative 2009 pandemic virus – A/California/04/2009 ("CA04").

Four (GD/3, rGD/3-NA289R, rGD/3-NA289K, Anhui/1) or six (CA04) serologically-confirmed naive ferrets (exposed/contact ferrets) were placed in a cage adjacent to an infected ferret (separated by ~5cm) on day 2 post infection. Pairs of ferrets were individually co-housed in adjacent wireframe cages which allow for spread of virus by respiratory droplet, but not by direct or indirect (via fomite) contact. Nasal washes were collected from infected ferrets on day 1 after inoculation and from contact ferrets on day 1 after co-housing, and then every other day (for up to 15 days) for virus titration. Virus titers in nasal washes were determined by plaque assay on MDCK cells. Viral RNA was available for isolation from nasal wash samples collected from index ferrets on days 1, 3, 5 and 7 post-infection and from contact ferrets on days 3, 5, 7, 9, 11, 13, and 15 post-infection.

Viruses

A/Guangdong/17SF003/2016 was propagated in embryonated chicken eggs to prepare a virus stock after being isolated from a fatal human case treated with oseltamivir [186]. We sequenced the GD/3 stock virus to verify consensus and sub-consensus variants (see details in section below). The GD/3 stock consensus sequence differs from the human isolate consensus sequence (GISAID isolate ID: EPI_ISL_249309) at nine sites (eight out

of nine were non-synonymous changes). A/Anhui/1/2013 was also propagated in embryonated chicken eggs after being isolated from an early human infection [174]. A/California/04/2009 was propagated in MDCK cells and was originally obtained from the Centers for Disease Control (CDC) [187]. Recombinant viruses, rGD3-NA289K and rGD3-NA289R, were generated by plasmid-based reverse genetics as previously described [188].

Template preparation

Total nucleic acids including viral RNA (vRNA) were extracted from nasal washes and were reverse transcribed using SSIV VILO (Invitrogen, USA) and the Uni12 primer (AGCAAAAGCAGG) in a total reaction volume of 20 µl . The complete reverse transcription protocol can be found here: https://github.com/tcflab/protocols/blob/master/VILO_Reverse_Transcription_h7n9_GLB_2019-02-15.md.

Single-stranded cDNA was used as a template for PCR amplification to amplify all eight genes using segment specific primers using high-fidelity Phusion 2X DNA polymerase (New England BioLabs, Inc., USA). PCR was performed by incubating the reaction mixtures at 98°C for 30 s, followed by 35 cycles of 98°C for 10 s, 51 - 72°C depending on gene segment for 30 s, 72°C for 120 s, followed by a final extension step at 72°C for 5 min. The complete PCR protocol, including segment-specific annealing temperatures and primer sequences, can be found here: https://github.com/tcflab/protocols/blob/master/Phusion_PCR_h7n9_GLB_2019-02-21.md.

PCR products were separated by electrophoresis on a 1% agarose gel (Qiagen, USA). The bands corresponding to full-length gene segments were excised and the DNA was recovered using QIAquick gel extraction kit (Qiagen, USA). To control for RT-PCR and sequencing errors, especially in low-titer samples, all samples were prepared in complete technical replicate starting from vRNA [189,190]. After completing replicate whole-genome sequencing for all 90 samples, we sequenced samples with low or no coverage, typically from low-titer samples, a third time and merged sequencing reads with the first two replicates with the goal of minimizing holes in our dataset. Samples for which we prepared a third complete deep-sequencing library included GD3 ferret 1 day 5 (all genes other than NA), GD3 ferret 1 day 7 (NA), GD3 ferret 5 day 7 (PA and NA), GD3 ferret 7 day 1 (PB1 and NA), rGD3 ferret 9 day 3 (NA), rGD3 ferret 15 day 7 (PA), rGD3 ferret 18 day 11 (PB1), rGD3 ferret 22 day 3 (MP), rGD3 ferret 23 day 7 (PA), Anhui ferret 25 day 7 (HA), Anhui ferret 31 day 3 (PB1), CA04 ferret 33 day 5 (NS) and day 7 (PB1).

Deep sequencing

Gel-purified PCR products were quantified using Qubit dsDNA high-sensitivity kit (Invitrogen, USA) and were diluted in elution buffer to a concentration of 1 ng/µl. All segments originating from the same samples with a non-zero concentration as determined by hsDNA Qubit (Invitrogen, USA) were pooled equimolarly and these genome pools were again quantified by Qubit. Each equimolar genome pool was diluted to a final concentration of 0.2 ng/µl (1 ng in 5 µl volume). Each sample (90 complete genomes in technical or duplicate or triplicate in addition to the stock virus) were made compatible for deep sequencing using the Nextera XT DNA sample preparation kit

(Illumina, USA). Specifically, each sample or genome was enzymatically fragmented and tagged with short oligonucleotide adapters, followed by 15 cycles of PCR for template indexing. Individual segments with undetectable concentrations by Qubit were tagmented and indexed separately to maximize recovery of complete genomes. Samples were purified using two consecutive AMPure bead cleanups (0.5x and 0.7x) and were quantified once more using Qubit dsDNA high-sensitivity kit (Invitrogen, USA). If quantifiable at this stage, independent gene segments were pooled into their corresponding genome pools. The average sample fragment length and purity was determined using Agilent High Sensitivity DNA kit and the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). After passing quality control measures, genomes were pooled into six groups of ~30, which were sequenced on independent sequencing runs. Libraries of 30 genomes 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. Denatured pooled libraries were diluted to a final concentration of 16 pM, apart from the first library which was diluted to 12pM, with a PhiX-derived control library accounting for 1% of total DNA loaded onto the flowcell. Then 600 µl of diluted-denatured library was loaded onto a 600-cycle v3 reagent cartridge. Average quality metrics were recorded, reads were demultiplexed, and FASTQ files were generated on Illumina's BaseSpace platform [191].

Sequence data analysis – quality filtering and variant calling

FASTQ files were processed using custom bioinformatic pipelines, available at this GitHub address https://github.com/tcflab/Sniffles2. Briefly, read ends were trimmed to achieve an average read quality score of Q30 and a minimum read length of 100 bases

using Trimmomatic [192]. Paired-end reads were merged and mapped to a reference sequence using Bowtie2 [193]. GD/3 and rGD/3 samples were mapped to the consensus sequence of the A/Guangdong/17SF006/2016 human isolate (GISAID isolate ID: EPI_ISL_249309) [174]. Anhui/1 samples were mapped to the consensus sequence of the A/Anhui/1/2013 human isolate (GISAID isolate ID: EPI_ISL_138739) [174]. CA04 samples were mapped to A/California/04/2009 reference sequence (GISAID isolate ID: EPI_ISL_29618). Alignment files were randomly subsampled to 200,000 reads per genome using seqtk to ensure even coverage and reduce resequencing bias [194]. The sequence depth per gene in each sample is shown in Supplementary Figure 1. Single nucleotide variants (iSNVs) were called with Varscan [195] using a frequency threshold of 1%, a minimum coverage of 100 reads, and a base quality threshold of Q30 or higher. Variants were called independently for technical replicate pairs and only iSNVs (intrahost single nucleotide variants) called in both replicates, “intersection iSNVs”, were retained [196]. If an iSNV was only found in one replicate, it was discarded. iSNV frequency is reported as the average frequency found across both replicates. iSNVs are annotated to determine the impact of each variant on the amino acid sequence. iSNVs were annotated in ten open reading frames: PB2 (polymerase basic protein 2), PB1 (polymerase basic protein 1), PA (polymerase acidic), HA (hemagglutinin), NP (nucleoprotein), NA (neuraminidase), M1 (matrix protein 1), M2 (matrix protein 2), NS1 (non-structural protein 1), and NEP (nuclear export protein), though for some analyses M1 and M2 are jointly represented as MP (matrix proteins) an NS1 and NEP are jointly represented as NS (non-structural proteins).

Sequence data analysis – diversity statistics

Nucleotide diversity was calculated using π summary statistics. π quantifies the average number of pairwise differences per nucleotide site among a set of sequences and was calculated using SNPGenie [197,198]. SNPGenie adapts the Nei and Gojobori method of estimating nucleotide diversity (π), and its synonymous (π_S) and nonsynonymous (π_N) partitions from next-generation sequencing data [199]. As most random nonsynonymous mutations are likely to be disadvantageous, we expect $\pi_N = \pi_S$ points toward neutrality suggesting that allele frequencies are determined primarily by genetic drift. $\pi_N < \pi_S$ indicates purifying selection is acting to remove new deleterious mutations, and $\pi_N > \pi_S$ indicates diversifying selection is favoring new mutations and may indicate positive selection is acting to preserve multiple amino acid changes [200]. We used paired t-tests to evaluate the hypothesis that $\pi_N = \pi_S$ within gene segments.

Sequence data analysis – estimating transmission bottleneck size

The beta-binomial model, explained in detail in Sobel-Leonard's 2017 paper [137], was used to infer effective transmission bottleneck size (Nb), the number of virions comprising the founding viral population at the onset of infection in the recipient host that successfully establish lineages persisting to the first sampling time point. In this model, the probability of iSNV transmission is determined by iSNV frequency in the index at the time of sampling. The probability of transmission is the probability that each iSNV is included at least once in a sample size equal to the bottleneck. The model incorporates sampling noise arising from a finite number of reads and therefore accounts for the possibility of false-negative variants that are not called in contact animals due to conservative variant

calling thresholds (≥1% in both technical replicates). The frequencies of transmitted variants are permitted to change from immediately following transmission to the first sampling time point according to a beta distribution. The beta-binomial model makes several limiting assumptions. The model assumes viral genetic diversity is neutral and variant frequencies are not impacted by selection and that variant sites are independent, which is unlikely to be true within a given gene segment because homologous recombination is not known to occur in influenza viruses [69]. In addition, the beta-binomial model ignores variants that arise *de novo* within contact animals to avoid artificially inflating bottleneck sizes. Overall, bottleneck size estimates from the beta-binomial model are conservative estimates. Computer code for estimating transmission bottleneck sizes using the beta-binomial approach has been adapted from the original scripts, available here: https://github.com/koellelab/betabinomial_bottleneck.

Figures

All figures were generated using Python Matplotlib and packages including plotly, seaborn, numpy, and scipy and were edited using Adobe Illustrator for clarity and readability. All derived data and computer code used to generate figures is available in the GitHub repository accompanying this manuscript [201].

Results

H1N1 viruses transmit more frequently than H7N9 viruses among the ferrets evaluated here

We isolated and sequenced viral RNA (vRNA) from nasal washes collected from two previously published studies [170,181]. Among 5 donor ferrets infected with H1N1, 4 successfully transmitted to a naive recipient ferret (80%). By comparison, 7 out of 12 ferrets infected with H7N9 AIV transmitted to a recipient ferret (58.3%) (**Figure 1, Supplemental Figure 2**). These group sizes are small and the H1N1 transmission rate is not significantly different from the H7N9 transmission rate (p=0.12; Mann-Whitney U). This is notable, however, given that the H7N9 viruses evaluated are wildtype AIV sequences capable of infecting and transmitting in ferrets, a mammalian model with similar sialic-acid receptor distribution and respiratory system anatomy to humans [202].

Rates of transmission varied substantially between H7N9 virus subgroups. One of four ferrets resulted in transmission when infected with either the LPAI human isolate (A/Anhui/1/2013; "Anhui/1") or the HPAI human isolate (A/Guangdong/17SF003/2016; "GD/3"). We also evaluated two recombinant viruses reverse engineered to contain a neuraminidase-inhibitor escape mutation (NA-289K) or wildtype (NA-289R) at position 289 in neuraminidase (NA) on the background consensus sequence of GD/3. Two of four ferrets infected with the neuraminidase-escape variant, rGD/3-NA289K, transmitted to the donor ferret and three of four ferrets infected with the wildtype variant, rGD3/NA289R, transmitted to the donor ferret (**Supplemental Figure 2**).

H7N9 within-host diversity is dominated by low-frequency iSNVs

Patterns of viral genetic variation provide rich information about how variants emerge within, and transmit beyond, individual hosts. We mapped sequencing reads against the

inoculating virus sequence and called within-host variants present in ≥1% of sequencing reads in both technical replicates, called intersection iSNVs (see methods for additional details). All coding region changes are reported using H7 numbering for the H7N9 viruses and H1 numbering for the H1N1 viruses, consistent with the numbering schemes used in the Nextstrain. We identified 867 unique iSNVs across all donor and recipient ferrets and all timepoints (482 synonymous, 382 nonsynonymous, and 3 stop mutations).

The average number of unique iSNVs per ferret across all available time points varied significantly across virus groups (Anhui/1, CA04, GD/3, and rGD/3) ($p=6.83x10^{-10}$; one-way ANOVA) (**Figure 2a**). The number of unique iSNVs was lowest in the CA04 group, ranging from 3-83 iSNVs per ferret (n=9 ferrets). Unsurprisingly, the number of unique iSNVs in the clonal recombinant H7N9 virus group was also low, ranging from 1-43 per ferret (n=13 ferrets). Viral diversity was highest in the H7N9 isolate groups with 85-195 and 27-142 unique iSNVs per ferret in the Anhui/1 group (n=5 ferrets) and GD/3 group (n=5 ferrets), respectively. The total count of unique iSNVs within a single ferret is relatively stable over time (**Supplementary Figure 3**).

Most iSNVs were detected at <10% frequency (**Figure 2b**). Compared to expectations under a neutral model, iSNVs detected within each virus group (Anhui/1, CA04, GD/3, and rGD/3) were present in excess at low frequencies. This pattern is consistent with purifying selection and population expansion acting on intrahost viral intrahost populations.

The frequency, genome location, and annotation (synonymous vs nonsynonymous) for each iSNV detected in hemagglutinin (HA), the receptor binding protein, is shown in **Figure 2c**. iSNVS in all other gene segments are plotted in **Supplementary Figure 4**.

H7N9 viral populations are subject to purifying selection in donor and recipient ferret hosts

We used a common measure of nucleotide diversity, π, within individual ferrets to assess signals of H7N9 viruses adapting or diversifying within mammalian hosts. This summary statistic quantifies the average number of pairwise differences per nucleotide site among a set of viral sequences. In particular, we compared the nucleotide diversity at synonymous sites (πS) to nucleotide diversity at nonsynonymous sites (πN) to assess the evolutionary forces acting on viral populations within individual hosts. In general, $\pi N/\pi S < 1$ indicates that, on average, purifying selection is acting to remove deleterious mutations from the viral population, and $\pi N/\pi S > 1$ indicates that diversifying selection is favoring new mutations, which might be expected in the case of an avian influenza virus adapting to a mammalian host. When πN approximates πS, this suggests that allele frequencies are determined primarily by genetic drift, stochastic shifts in allele frequencies primarily determined by population size [203].

As is expected for a fit viral population, πS exceeded or was equal to πN in the ferrets infected with H1N1 viruses (**Figure 3a**, orange). πS was significantly greater than πN in PB2, PA, and NA and πN and πN never significantly exceeded πS, suggesting these viruses are under a combination of purifying selection and genetic drift in ferrets.

Somewhat surprisingly, the H7N9 viruses were also under a combination of purifying selection and genetic drift in ferrets. πS significantly exceeded πN in all genes apart from NA in the GD/3 group and all genes apart from NA and HA in Anhui/1 (**Figure 3a**, blue and turquoise). It is notable that HPAI and LPAI H7N9 are not subject to diversifying selection as it suggests these avian influenza viruses are relatively fit in mammalian hosts.

We compared nucleotide diversity in donor-recipient pairs to evaluate population forces before and after transmission. Genetic diversity is lost during transmission. We found genome-wide genetic diversity (π) is lower in recipient ferrets compared to donor ferrets in the H1N1 group ($p=0.125$, paired t-test) and significantly lower in recipient ferrets compared to donor ferrets in the H7N9 group ($p=0.005$; paired t-test). As we have done previously [204,205], we looked for selective sweeps by comparing the change in πN and πS for each gene in paired donor and recipient ferrets. Within each gene segment, πN did not differ significantly between donor and recipient and πS similarly did not different between donor and recipient. This was true across all H1N1 transmission pairs (**Figure 3b**) and all H7N9 transmission pairs (**Figure 3c**). Taken together, this suggests that while genetic diversity is purged during the transmission event, this diversity is purged equally across the genome with no evidence for a selective reduction in any particular gene segment.

Airborne transmission results in a dramatic shift of iSNV frequencies

We took advantage of time series data to track iSNV frequency trends within each donor ferret and following airborne transmission into the associated recipient ferret. Strikingly, frequencies of specific H7N9 mutations in donor ferrets do not predict their likelihood of transmission nor do they predict iSNV frequency post-transmission. For example, one polymorphic site at position 137 in HA involving a glycine to glutamic acid mutation (G137E) in the GD/3 transmission pair was present at 81% one day after inoculation in the donor ferret and decreased to a sub-consensus frequency (39.3%) by 7 DPI. Despite this marked downward trend in the donor animal, G137E was transmitted to the recipient ferret and was found at ≥99% from first time point post-infection onward (**Figure 4**). Another polymorphic site in the matrix protein (M1) with an arginine-to-lysine mutation at position 210 (R210K) was conversely never detected in the donor ferret above 1%, yet was nearly fixed (97.5%) at the first time point post infection in the recipient ferret. Interestingly, M1 R210K then decreased in frequency in the recipient ferret and was found at 54.5% at 9 DPI, suggesting this mutation might have conferred a mild fitness cost. We observed similar patterns in synonymous variants. For example, a synonymous A-to-G change at nucleotide 2,037 in the polymerase basic protein 1 (PB1) was found at 5.57% at 1 DPI, 5.1% at 3 DPI, 4.76% at 5 DPI, 1.87% at 7 DPI and was nearly fixed immediately following transmission, but again decreased in frequency to 57.57% at 9 DPI in the recipient ferret.

Transmission dynamics of individual H7N9 GD/3 variants remained stochastic even in the case of amino acids under positive selective pressure in humans. A valine to isoleucine change at amino acid position 219 in M1 is thought to play a role in avian influenza virus

adaptation to mammals [206] and accordingly increased in frequency from 34.7% to 84.3% in the donor ferret, but nonetheless failed to transmit to the recipient and then amazingly arose *de novo* once again in the recipient ferret. Similar iSNV dynamics were observed in the Anhui/1 transmission pair as well in the recombinant GD/3 pairs (**Figure 4**, **Figure 5a**).

Unlike iSNV dynamics in the H7N9 transmission events, multiple iSNVs in the H1N1 CA04 donor ferrets remained polymorphic immediately following transmission (e.g. HA D127E and S183P (H1 numbering)) (**Figure 4**, **Figure 5a**). It is clear that airborne transmission of H7N9 viruses in ferrets dramatically alters the viral population, stochastically allowing minor variants to become dominant in subsequent generations despite lacking a putative fitness benefit, and conversely preventing known mammalian-adaptive consensus-level variants from transmitting.

Airborne transmission of H7N9 viruses in ferrets is characterized by a very narrow transmission bottleneck

The number of viruses that found infection is a crucial determinant of the pace at which novel, beneficial variants can emerge at the level of the population. Narrow transmission bottlenecks cause a founder effect and purge low-frequency iSNVs, regardless of their fitness. Conversely, wide transmission bottlenecks allow more viruses to initiate infection, reducing the chance that beneficial or rare variants are lost. Understanding the size of the transmission bottleneck is therefore important for evaluating the probability that novel variants arising within an individual host infection will be transmitted onward. To infer

transmission bottleneck sizes, we applied the beta-binomial inference method [137]. To do this, we used the first time point available in the recipient host and the time point immediately preceding this in the associated donor host (see methods for details).

The vast majority of iSNVs detected in all donor ferrets were lost during transmission and were not found in the recipient ferret. A very small number of iSNVs in the Anhui/1 and GD/3 donor ferrets transmit and are found fixed (at 100% frequency) in the recipient ferret (**Figure 5a**). Most notably, two synonymous iSNVs at 3.5% (A-to-G at nt 270 in PB1) and 3.6% (C-to-T at nt 1735 in PB1) in the GD/3 donor ferret transmit and were fixed immediately following transmission. This pattern where iSNVs are dichotomously either lost or fixed following transmission is consistent with a very narrow transmission bottleneck [137]. The majority of H1N1 iSNVs were similarly lost during transmission, although we found five iSNVs that were shared at sub-consensus frequencies (<50%) among donor-recipient pairs (**Figure 5a**).

While bottleneck size estimates varied modestly between ferret pairs, we found consistent support for fewer than 11 viruses initiating infection in all recipient ferrets. We found the combined maximum likelihood estimate for the mean transmission bottleneck size for the CA04 (H1N1) pairs (n=4 pairs) was 6 (95% CI: 3-11) (**Figure 5b**). We evaluated seven transmission events in the H7N9 group; one Anhui/1 pair, one GD/3 pair, and five rGD/3 pairs. However, two of the rGD/3 transmission events (pairs 9 and 11) were uninformative because the donor ferret had no polymorphic sites. The combined maximum likelihood

estimate for the mean transmission bottleneck size for the H7N9 group (n=5 pairs) was 1 (95% CI: 1-3) (**Figure 5b**).

Although group sizes are quite small, we compared the mean transmission bottleneck sizes for the H1N1 group to the H7N9 group and found modest evidence that the H7N9 transmission bottlenecks in a mammalian model system are even narrower than the H1N1 group (p=0.054; unpaired t-test). Overall, our data suggest the vast majority of H7N9 iSNVs arising in ferret hosts are lost during transmission and because so few viruses found infection following transmission, any iSNV that happens to be present in a transmitting virus' genome will likely become fixed in the post-transmission viral population.

Discussion

The evolutionary processes by which avian influenza viruses adapt to mammalian hosts are poorly understood despite the critical importance of these mechanisms in assessing the pandemic potential of avian influenza viruses. Our study examined the viral dynamics of wildtype LPAI and HPAI H7N9 viruses in a ferret model, a well-studied mammalian system which closely resembles human respiratory physiology [207]. Relatively few studies have evaluated evolutionary dynamics of avian viruses in *in vivo* mammalian models, particularly because such studies could not be conducted during the gain-of-function research pause. In this study we hypothesized that an avian virus replicating and transmitting in a mammalian system would be under strong selective pressure to become more mammalian, but to our surprise we do not detect any evidence to support this

hypothesis. We instead find evidence that HPAI and LPAI viruses are subject to mild purifying selection in ferret hosts, which is a signature classically associated with a virus that is already well-adapted to its host.

In some ways, H7N9 avian influenza isolates do appear well-adapted to mammalian hosts; they replicate to high-titre in the upper and lower respiratory compartment of ferrets [170], achieve infection via airborne transmission between ferrets [135,175,177], and are responsible for > 1,500 human spillover infections [174]. Yet there is no evidence that H7N9 viruses are capable of sustained human-to-human transmission [176] and there are very few documented human spillover infections since the fifth epidemic wave in 2017 [28]. It follows that there are significant barriers to more efficient H7N9 mammalian infection and transmission. Molecular barriers have been previously identified and include mixed avian- and human-receptor preferences, fusion instability at high endosomal pHs in humans, and reduced polymerase activity at lower temperatures in the upper respiratory compartment [135,177–180]. In this study, we identified and characterized evolutionary barriers which we posit combine with molecular barriers to severely constrain the ability of wildtype H7N9 viruses to effectively adapt to mammalian hosts in typical spillover infections. These constraints are most apparent during airborne transmission of influenza where variants under possible positive selection in index ferrets and even putative mammalian-adapting variants, like PB2 D701N and M1 V219I, are not preferentially transmitted to contact ferrets.

We are not the first to document that transmission of H7N9 viruses in ferrets involves a stringent and attenuating transmission bottleneck. Zaraket et al also showed evidence of this in ferrets using a LPAI H7N9 isolate [135]. Additionally, other studies, including two from our lab, have used similar methods to characterize the evolution of avian-like viruses in ferret models. One of these studies evaluated a H5N1 laboratory reassortant virus with an H5 HA segment and the other seven segments originating from a human H1N1 virus [205]. The other study evaluated a genetically-modified avian virus resembling the 1918-pandemic virus [204,208]. Upon initial comparison, the results of these prior studies appear to be in part at odds with the results of this study. These studies involving reassortant H5N1 and 1918-like viruses both detect evidence of selective sweeps on HA during airborne transmission, as evidenced in part by a greater reduction in HA genetic diversity than any other gene segment following transmission, suggesting that selection acted to favor transmission and/or replication of only a subset of HA sequences from index animals in contacts infected by respiratory droplets. We detect no such signal in this study and more broadly detect no strong evidence for the role of positive or directional selection acting on H7N9 viruses within or between ferret hosts. However, Wilker and Dinis et al note in their discussion that the replication and transmission of wiltype avian viruses, in contrast to engineered reassortant avian viruses, in mammals may result in patterns of selection that differ from those observed in their study [205]. We believe this rationale precisely explains the differences in population dynamics observed across these studies.

It is helpful to imagine multiple “fitness peaks” and “fitness valleys” across a landscape which captures the interaction of the virus and host genotypes [123]. When placing wildtype

H7N9 viruses on this landscape it is likely that the tallest fitness peak can be found when wildtype avian H7N9 viruses infect avian hosts, though our results suggest wildtype H7N9 viruses are at least moderately fit in the context of a mammalian host as well. In contrast, the reassortant and genetically modified H5N1 and 1918-like viruses are not viruses found in nature and may be located in relative “fitness valleys" on this same fitness landscape. Each of these viruses is likely subject to the same biological constraints, including short-lived infections, the vast majority of new mutations conferring mildly deleterious to lethal phenotypes [209–211], and narrow transmission bottlenecks which permit very few viruses from making their way into subsequent hosts [122,133–136]. However, the overall impact of this molecular biology may differ significantly depending on where the virus is located in the virus-host genotype landscape. If wildtype H7N9 viruses replicating in mammalian hosts are already located on a relative fitness peak, any new mutations are exceptionally unlikely to confer a sufficient benefit to be positively selected in the setting of an acute infection. Additionally, wildtype H7N9 viruses in ferrets randomly establishing successful infection following airborne transmission are very likely to carry neutral and/or deleterious variants which then achieve de facto fixation because so few viruses successfully passed through the transmission bottleneck. These predictions are consistent with the results of this study. In contrast, diversifying selection and selective sweeps are much more likely to be detected in the context of viruses located in a relative “fitness valley”, which is consistent with results observed in the studies evaluating reassortant H5N1 and 1918-like avian viruses in mammalian models. It has been previously hypothesized that epistasis is crucial to the evolution of influenza viruses and mutations that promote human adaptation in one viral and host genetic background may

not be well-tolerated in others [74,212]. Additionally, given that we and similar studies identify (even *de novo*) mammalian-adaptive mutations in the context of relatively few ferrets suggest that generation of mammalian-adaptive mutations is not the rate-limiting step in adaptation of avian viruses to mammalian hosts and instead is consistent with predictions made by Russell et al who hypothesized H5N1 viruses would generate human-adapting mutations during infection, but these mutations would remain at low frequencies and fail to be transmitted [158].

The results of this study may have implications for assessing the adaptive potential of avian influenza viruses in the setting of human spillover infections. Though it is not unimaginable for a wildtype avian virus to quickly adapt to humans and achieve sustained human-to-human transmission, the results of this study suggest there are significant evolutionary barriers for wildtype avian viruses to do so. Each of the known influenza pandemics have resulted from a major reassortant event typically by way of a "mixing vessel", rather than adaptation of avian influenza viruses in the setting of a human spillover infection [20,21]. It is true that avian virus spillover infections are responsible for significant individual morbidity and mortality in southeast Asia, however these spillover infections may be less concerning for ongoing and further human adaptation of avian influenza viruses than previously recognized. Our results additionally emphasize the importance of population and One Health interventions to reduce the opportunity for avian and mammalian viruses to co-infect a single host – these interventions include, but are not limited to, continued poultry vaccination, culling, poultry movement restrictions, best practices at live animal markets and others [213]. Like most ferret studies, the results of

these experiments are limited by relatively small sample sizes and biological uncertainties regarding the possible differences between viral infection and transmission between ferrets and humans. Results described here cannot be mapped directly onto avian influenza virus infections in human hosts and should continue to be explored and corroborated by additional investigations, including targeted virological and epidemiological research [214].

Figures, tables, and supplemental material

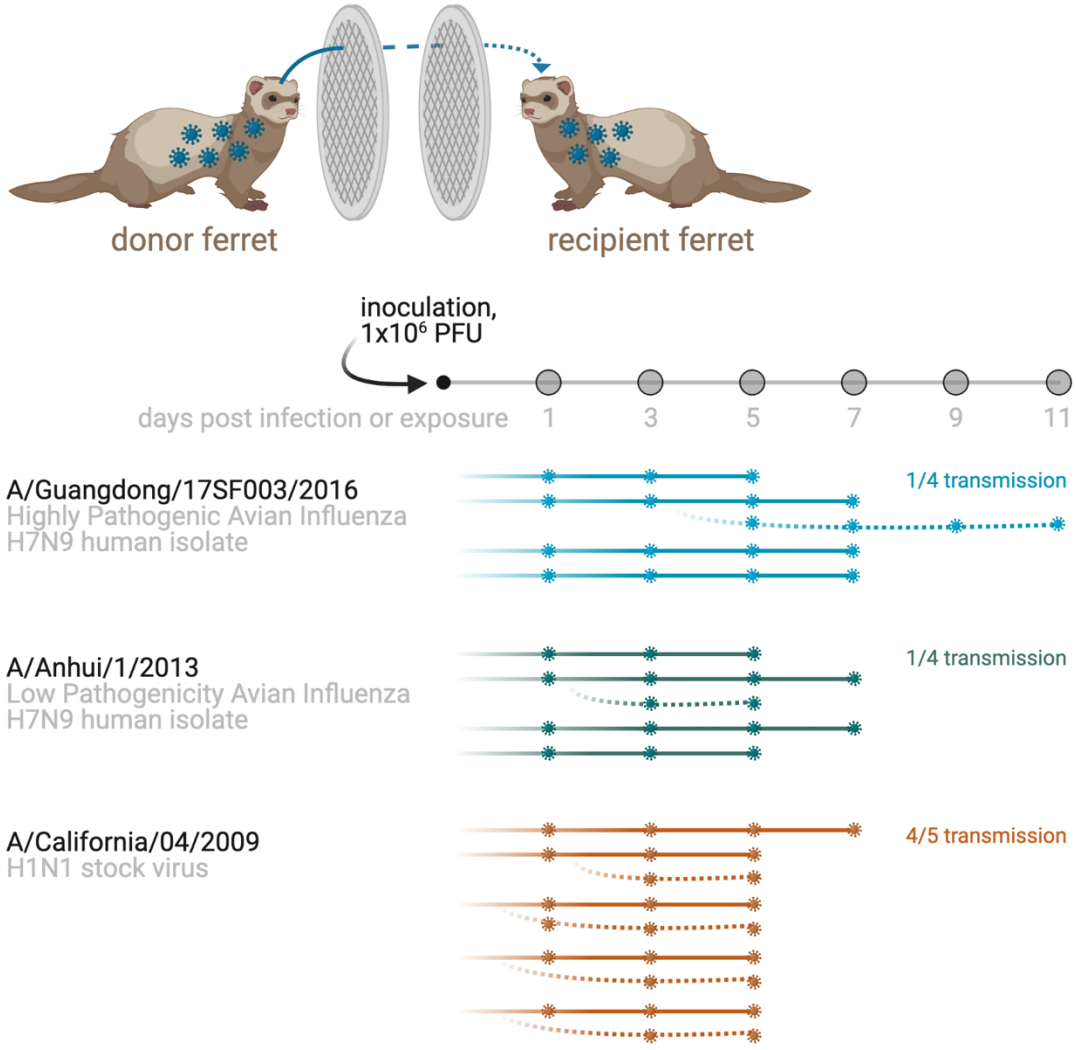


Figure 1. Overview of the experimental system and sampling timeline

Schematic depicting sampling timeline for donor and recipient ferrets. Ferrets were inoculated intranasally with 106 PFU of a HPAI H7N9 isolate (A/Guangdong/17SF003/2016; blue), a LPAI H7N9 isolate (A/Anhui/1/2013; turquoise), or a H1N1pdm virus (A/California/04/2009; orange) on day 1 post infection (DPI). Recombinant HPAI H7N9 viruses (rGD3-NA289R, rGD3-NA289K) are not depicted here and can be seen in Supplemental Figure 1. One day after infection, one naive recipient

ferret was paired with each donor ferret. Nasal washes were collected from donor (solid line) and recipient (dotted line) ferrets up to 15 DPI. Small virions denote days on which live virus was detected by plaque assay. Viral RNA was extracted from these same days and was prepared in duplicate for whole-genome sequencing.

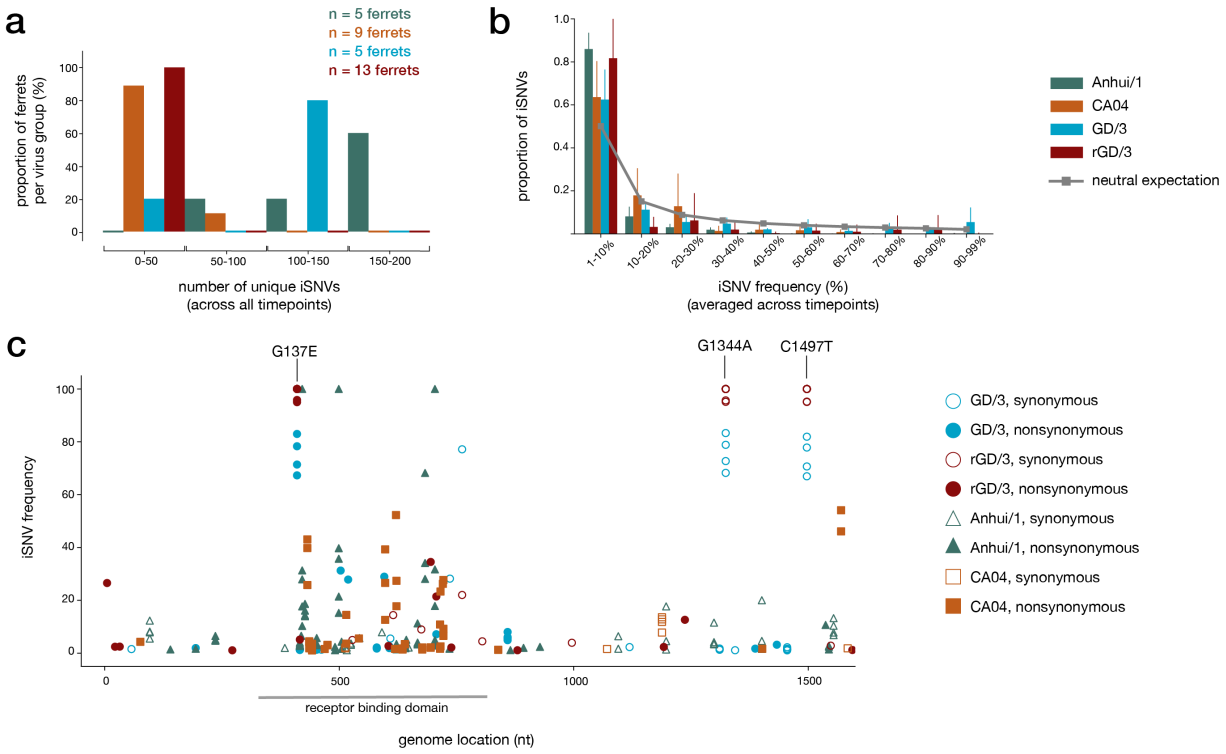


Figure 2. Frequency and location of intrahost single nucleotide variants

a. A histogram displaying the average number of unique iSNVs detected across all available timepoints. The y-axis displays the proportion of ferrets with various numbers of unique iSNV (x-axis bins) compared to the total group size across four virus groups. Virus groups are the LPAI Anhui/1 group (turquoise; n=5), the H1N1 CA04 group (orange; n=9),

the HPAI GD/3 group (blue; n=5) and the recombinant HPAI rGD/3 group (red; n=13). b. The proportion of iSNVs that were detected at various within-host frequency bins is shown for each virus group. Error bars represent the variance in the proportion of total within-host iSNVs across individual ferrets within each group. The solid grey line indicates the expected proportion of variants in each frequency bin under a neutral model. c. All iSNVs detected in hemagglutinin (HA) across all virus groups. GD/3 and rGD/3 iSNVs are plotted using circles, Anhui/1 iSNVs are plotted using triangles, and CA04 iSNVs are plotted with squares. Synonymous iSNVs are denoted with open symbols and nonsynonymous iSNVs are denoted with closed symbols. Three iSNVs found in multiple HPAI samples at high frequencies are labeled; G137E and two synonymous mutations at nucleotides 1,344 and 1,497. iSNVs in all other gene segments can be found in Supplementary Figure 3.

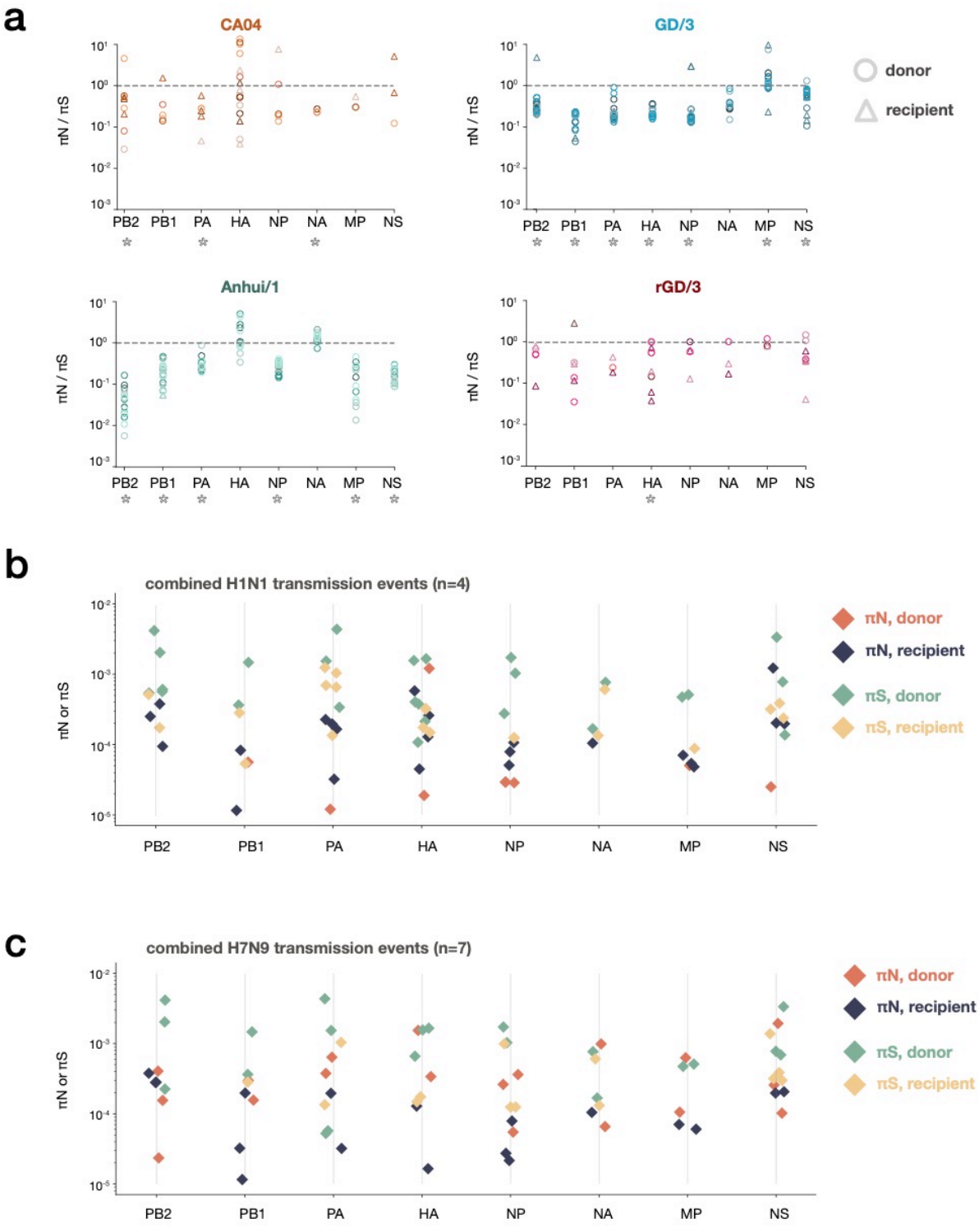

a
CA04
GD/3
Anhui/1
rGD/3
πN / πS
PB2 PB1 PA HA NP NA MP NS
donor
recipient
b
combined H1N1 transmission events (n=4)
πN or πS
πN, donor
πN, recipient
πS, donor
πS, recipient
c
combined H7N9 transmission events (n=7)

Figure 3. Patterns of viral genetic diversity within ferret hosts

a. πN / πS nucleotide diversity is plotted for each gene segment. Each datapoint represents a single ferret. Circles denote donor ferrets and triangles denote recipient ferrets. The dotted grey line represents where πN is equal to πS (y=1). A grey star is plotted below each gene on the x-axis when πS is significantly greater than πN, suggesting that gene segment is under purifying selection. b. πN and πS in the H1N1 donors and recipients are plotted for each gene segment. c. πN and πS in the H7N9 donors and recipients are plotted for each gene segment. πN and πS in the donor ferrets are denoted by the salmon and green diamonds, respectively. πN and πS in the recipient ferrets are denoted by the dark blue and yellow diamonds, respectively.

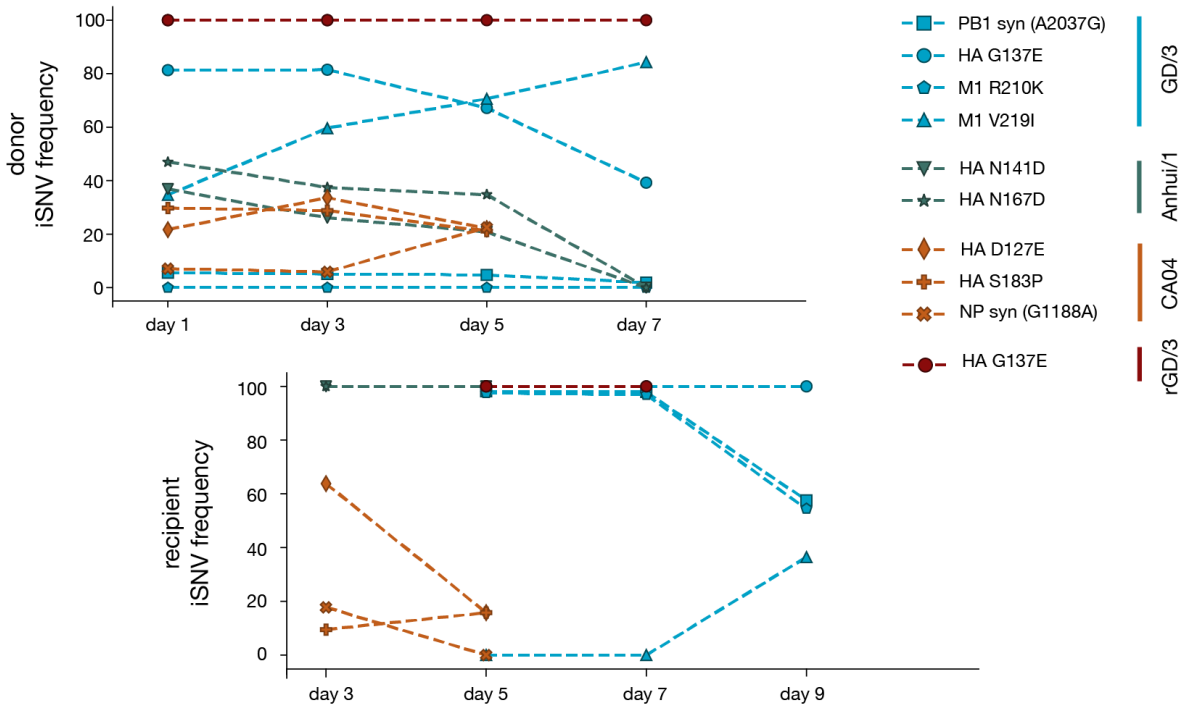


Figure 4. iSNV frequency dynamics across the transmission event.

The frequencies of individual iSNVs are plotted over time in donor ferrets (top plot) and following transmission into the associated recipient ferret (bottom plot). Colors denote virus groups and markers denote particular iSNVs. iSNV are plotted as y=0 at time points when an iSNV was not detected ≥1% frequency and are absent at time points when no viral RNA was recovered for deep sequencing.

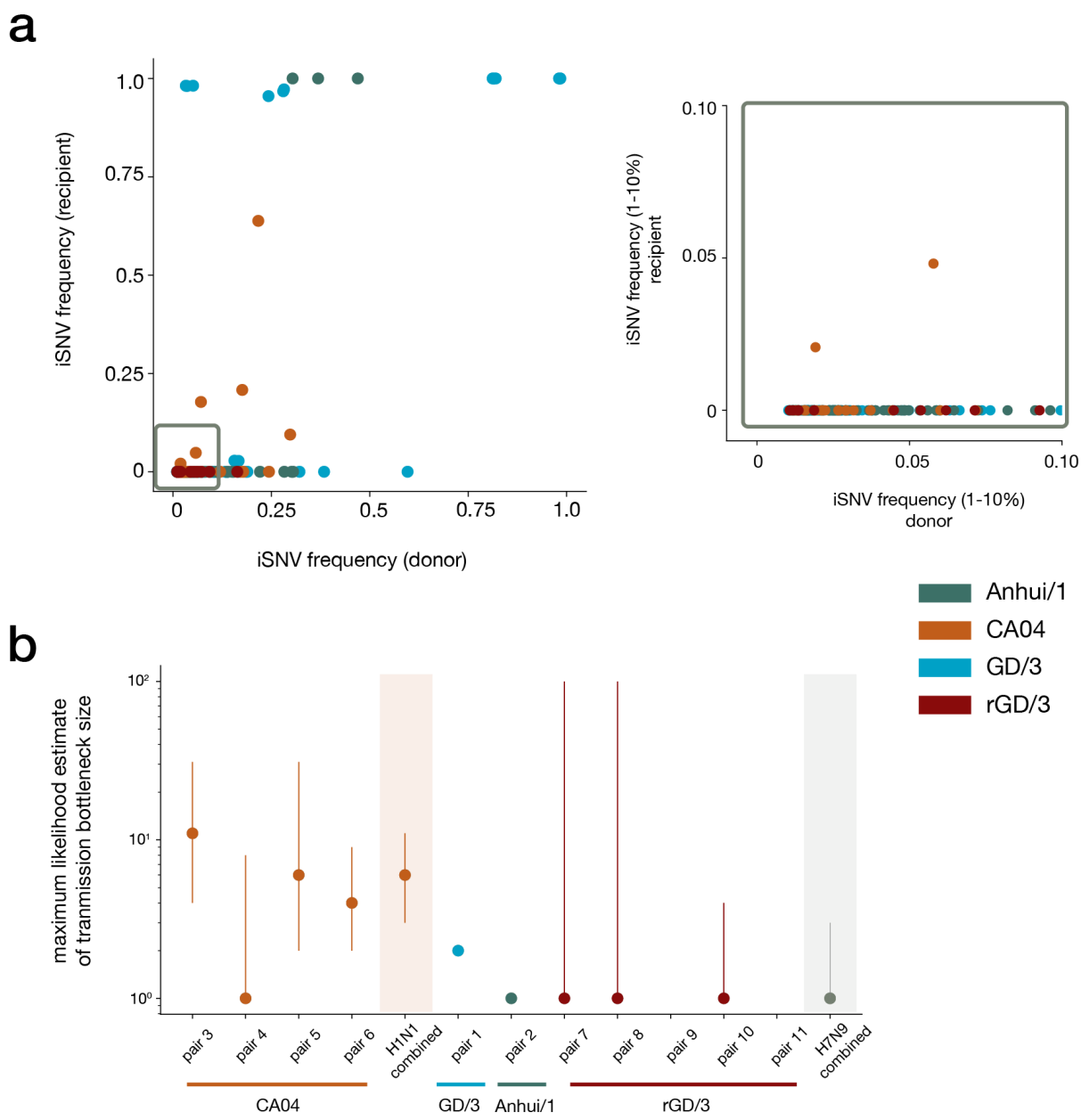
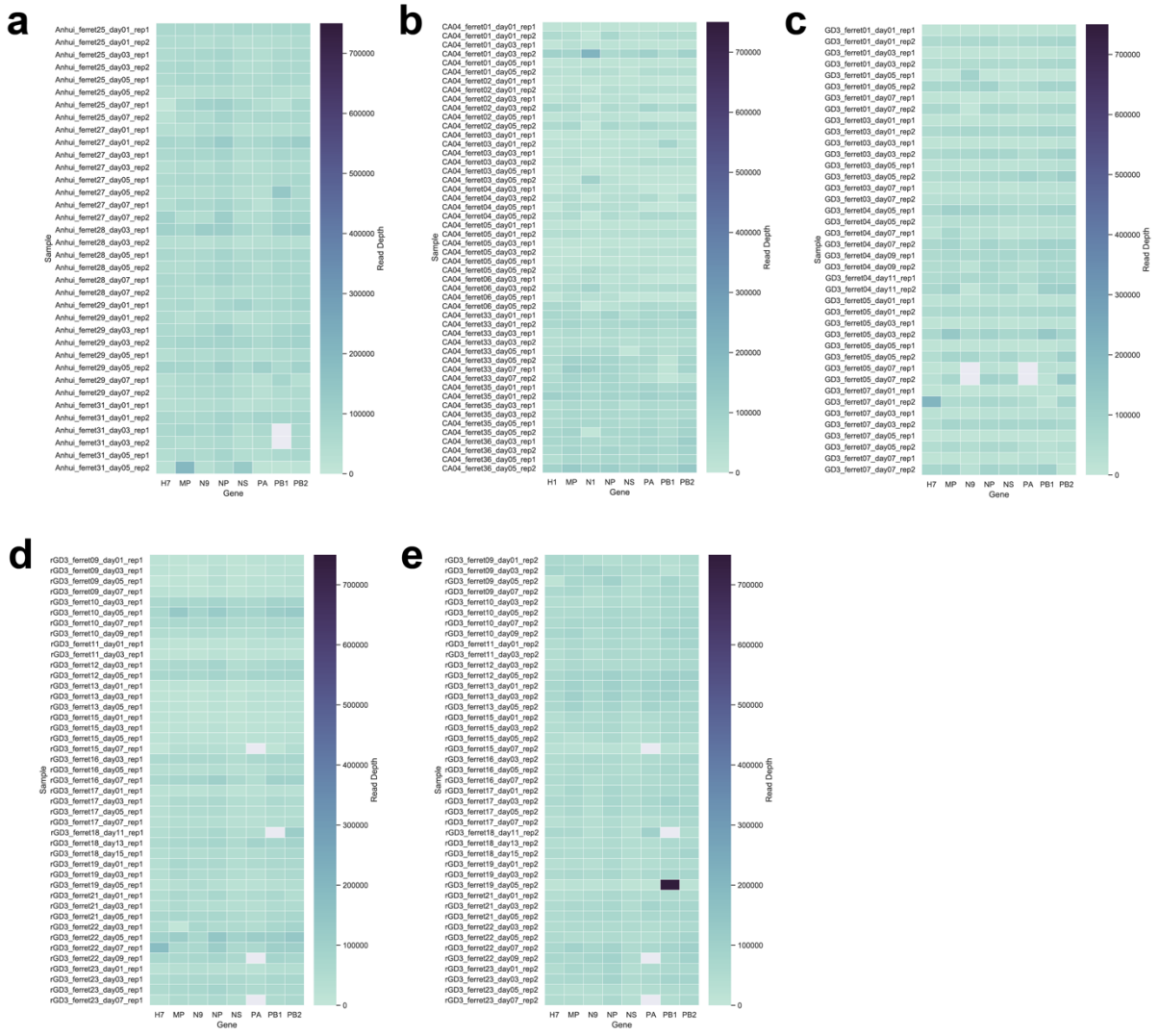


Figure 5. H1N1 and H7N9 transmission bottlenecks in ferret donor-recipient pairs

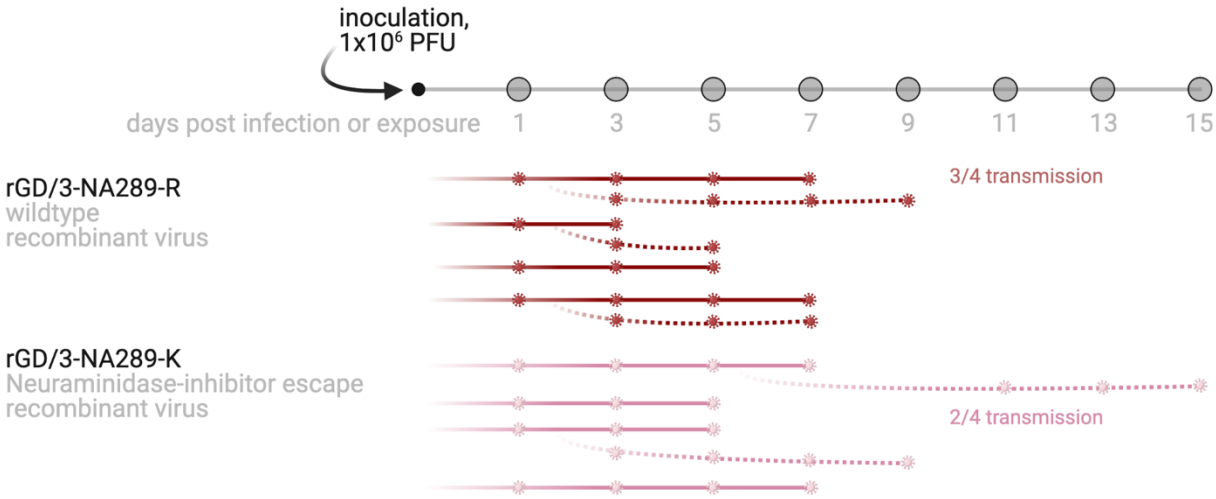
a. "TV plots" showing intersection iSNV frequencies in all 11 donor-recipient pairs. The grey box highlights low-frequency iSNVs (1-10%). b. Maximum likelihood estimates for mean transmission bottleneck size in individual donor-recipient pairs. Colors denote virus groups. Bottleneck sizes could not be estimated for a few pairs (rGD/3 pair 9 and pair 11) because there were no polymorphic sites detected in the donor. The combined H1N1

estimate was calculated using pairs 3, 4, 5 and 6. The combined H7N9 estimate was calculated using pairs 1, 2, 7, 8 and 10.



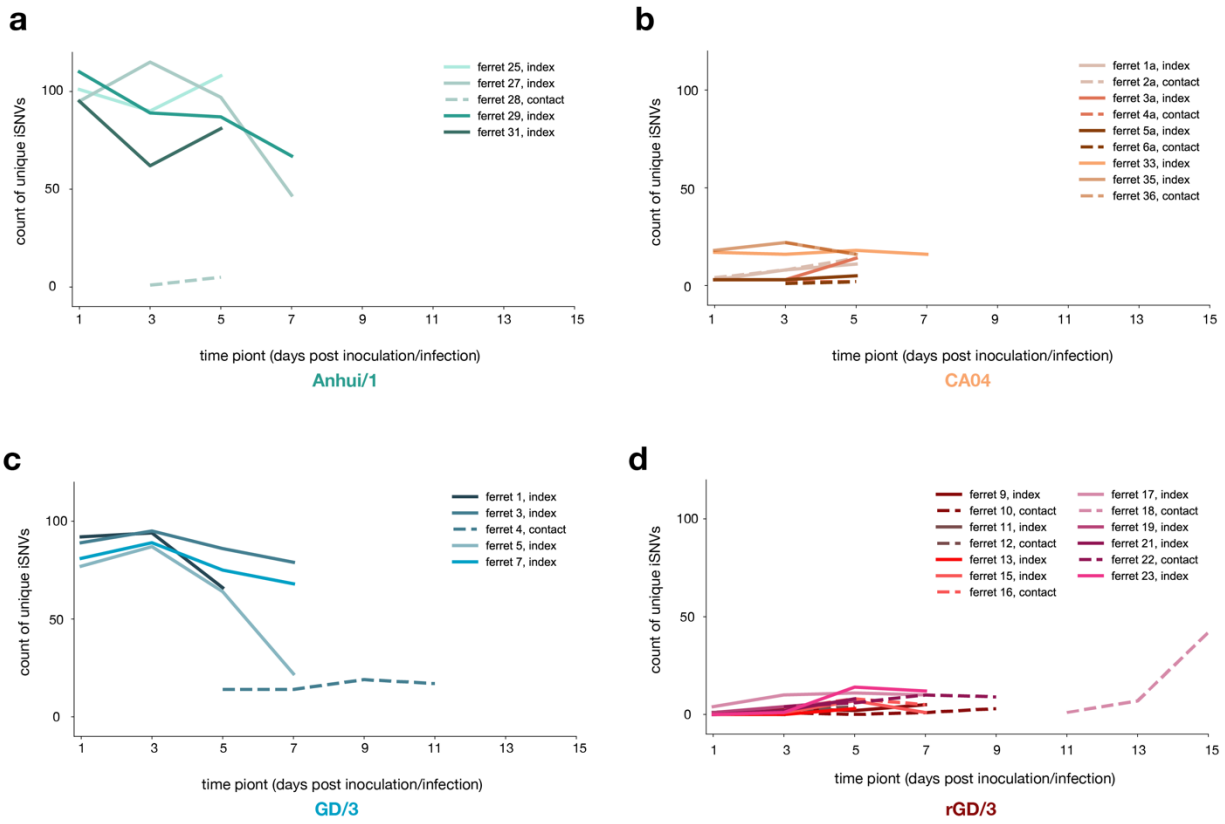
Supplemental Figure 1. Read number coverage heat maps.

Coverage is represented by color heatmap. Gene segments are shown along the x-axis. Individual samples are shown along the y-axis. Read coverage for **a.** Anhui/1, **b.** CA04, **c.** GD/3, **d.** rGD/3 replicate 1, and **e**. rGD/3 replicate 2.



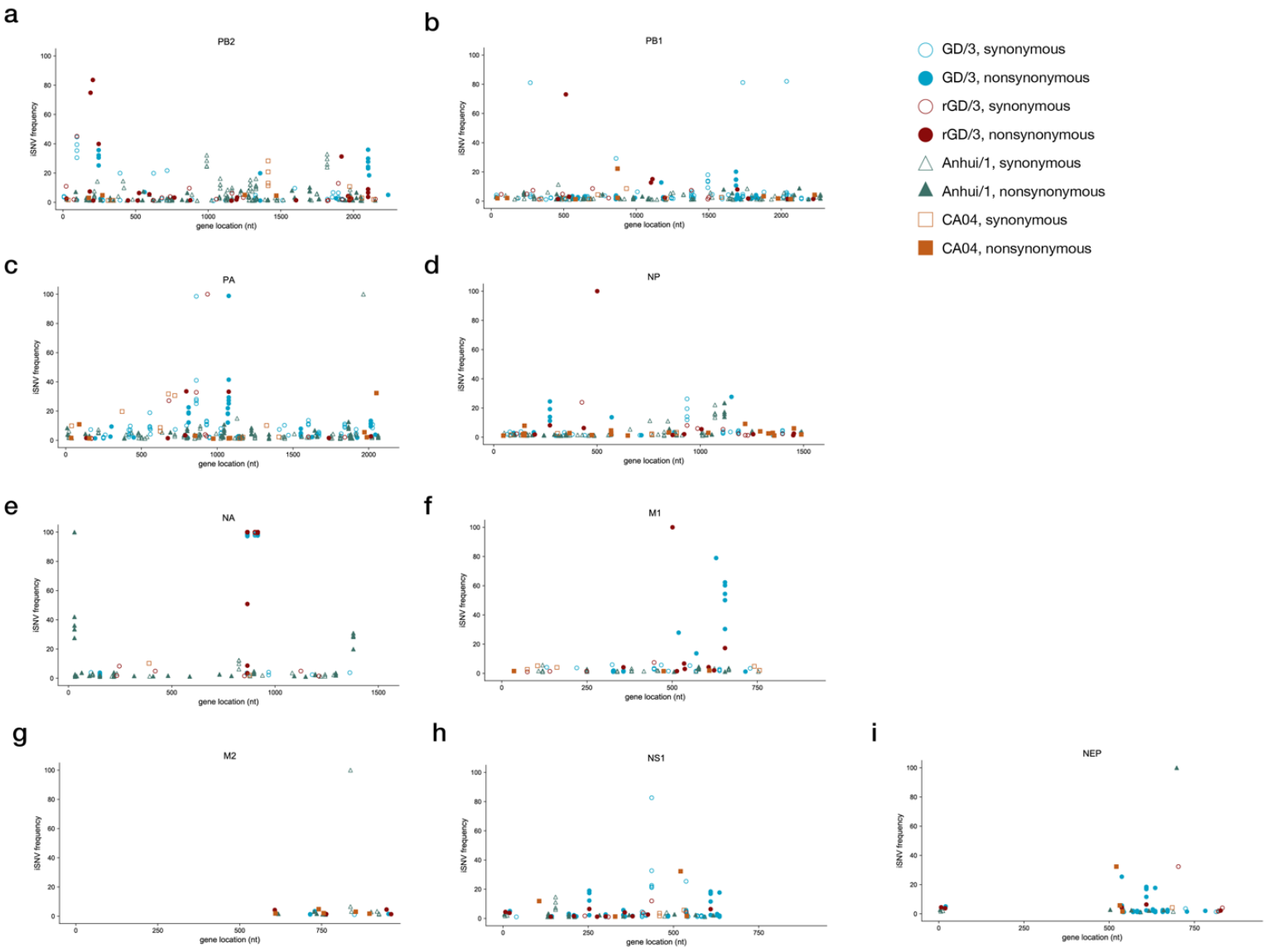
Supplemental Figure 2. Overview of the experimental system and sampling timeline for recombinant viruses

Schematic depicting sampling timeline for donor and recipient ferrets for recombinant H7N9 viruses. Ferrets were inoculated intranasally with 10^6 PFU of rGD3-NA289R (red) or rGD3-NA289K (pink). One day after infection, one naive recipient ferret was paired with each donor ferret. Nasal washes were collected from donor (solid line) and recipient (dotted line) ferrets up to 15 DPI. Small virions denote days on which live virus was detected by plaque assay. Viral RNA was extracted from these same days and was prepared in duplicate for whole-genome sequencing.



Supplemental Figure 3. Count of unique iSNVs within individual ferrets and over time

Count of unique iSNVs within individual ferrets and over time in **a**. the LPAI H7N9 Anhui/1 group, **b.** the H1N1 CA04 group, **c.** the HPAI H7N9 GD/3 group, and **d.** the recombinant HPAI rGD/3 group. Donor ferrets are denoted with solid lines and recipient ferrets are denoted with dashed lines.



Supplemental Figure 4. Patterns of viral diversity within ferret hosts

All iSNVs detected in **a**. PB2 (polymerase basic protein 2) **b**. PB1(polymerase basic protein 2) **c**. PA (polymerase acidic protein) **d**. NP (nucleoprotein) **e**. NA (neuraminidase) **f**. M1 (matrix protein 1) **g**. M2 (matrix protein 2) **h**. NS1 (non-structural protein 1) **i**. NEP (nuclear export protein or non-structural protein 2). GD/3 and rGD/3 iSNVs are plotted using circles, Anhui/1 iSNVs are plotted using triangles, and CA04 iSNVs are plotted with squares. Synonymous iSNVs are denoted with open symbols and nonsynonymous iSNVs are denoted with closed symbols.

Chapter 3:

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[6], David H. O'Connor[2], Thomas C. Friedrich[1,5]

# These authors contributed equally

* Corresponding author; tfriedri@wisc.edu

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

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

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

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

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

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

Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. PLoS Pathog. 2021 Feb 26;17(2):e1009373. doi: 10.1371/journal.ppat.1009373. PMID: 33635912; PMCID: PMC7946358.

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 [215–219]. Among these species, natural transmission has only been observed in mink, cats, and ferrets [215,220,221]. Transmission from humans to mink and back to humans has also recently been documented [222]. Infectious virus has been recovered from various upper- and mid-respiratory tissues in cats and ferrets, including nasal turbinates, soft palate, tonsils, and trachea [215,220]. However, only in cats has infectious virus been recovered from

lung parenchyma, where infection is most commonly linked to severe disease in humans [215,220,223,224].

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 [122,133–136,139,204,205,225,226]. 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 [122,133–136]. 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) [137]. 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 [120,138]. In this way, narrow transmission bottlenecks are generally expected to slow the pace of seasonal influenza virus adaptation [139,140] 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 [54,227–229]. 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 [54,227–233]. 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.

Methods

Ethics statement

No animal experiments were specifically performed for this study. We used residual nasal swabs collected from domestic cats as part of a previously published study [234].

Animal studies were approved prior to the start of the study by the Institutional Animal Care and Use Committee and performed in accordance with the Animal Care and Use Committee guidelines at the University of Wisconsin-Madison.

Domestic cat experiments

No animal experiments were specifically performed for this study. We used residual nasal swabs collected from domestic cats as part of a previously published study [234]. Animals used in this study were specific-pathogen-free animals from a research colony maintained at the University of Wisconsin-Madison and were negative for feline coronavirus. As previously described by Halfmann et al, domestic cats were housed in 0.56 m x 0.81 m x 1.07 m cages in a laboratory with 65% humidity at 23oC, and with at least 15.2 air exchanges per hour. Weight and body temperature (through implanted transponders) were measured daily (days 1–14). Under ketamine and dexdomitor anesthesia, three cats were inoculated with 5.2 x 105 plaque-forming units (PFU of SARS-CoV-2 given by a combination of inoculation routes for every animal (nasal [100 µl per nare], tracheal [500 µl], oral [500 µl], and ocular [50 µl per eye]). To reverse the effects of the anesthesia, antisedan was administered to the animals after completion of the inoculation. Nasal swabs were collected daily during the study (days 1–10).

Nucleic acid extraction

For each sample, approximately 140 µL of viral transport medium was passed through a 0.22 µm 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 H2O.

Complementary DNA (cDNA) generation

Complementary DNA (cDNA) was synthesized using a modified ARTIC Network approach [235,236]. Briefly, RNA was reverse transcribed with SuperScript IV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) using random hexamers and dNTPs. Reaction conditions were as follows: 1 µL of random hexamers and 1 µL 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 [235,236]. 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 S3 Table. 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 [235,236]. Following amplification, samples were pooled together before TruSeq 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 30 µL 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 25 µL of resuspension buffer. Lastly, samples were amplified using 8 PCR cycles, cleaned with a 1:1 bead clean-up, and eluted in 30 µL 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. Read coverage for index cat samples is plotted in S6 Fig and for contact samples in S7 Fig. 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) [237]. 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. Variant concordance across technical replicates is plotted in S8 Fig for index cats and S9 Fig for contact cats. 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 [234]. 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 (**S2 Table**). 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) [238]. 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 [197]. 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 [199]. We tested the null hypothesis that πN = πS within each gene using an unpaired t-test (S1 Table). The code to replicate these results can be found in the `diversity_estimates.ipynb` Jupyter Notebook in the `code` directory of the GitHub repository [200].

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 [239]. 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.

Focal Nextstrain build of S H655Y sequences

The focal H655Y build (S5 Fig) was prepared as described in Hodcroft et al. (2020), with different mutations targeted for the S:655 mutation [240]. 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.

Code and data availability

Code to replicate analyses and re-create most figures is available at https://github.com/katarinabraun/SARSCoV2_transmission_in_domestic_cats. Fig 1 was created by hand in Adobe Illustrator and S6 and S7 Figs 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. 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).

Financial disclosure

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 Author TCF. Author GKM is supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Results

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

Recently, members of our team inoculated three domestic specific-pathogen free cats with a second-passage SARS-CoV-2 human isolate from Tokyo (hCoV-19/Japan/UT-NCGM02/2020) [234]. Each index cat was co-housed with a 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 **S1A Fig** and infectious viral titers are shown in S1B Fig.

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) [235,236]. Variants were required to be present in technical replicates at ≥3% and ≤97% of sequencing reads [190] (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 [231]. 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 (**S2 Fig**). 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 [239]. Especially in the setting of an acute viral infection, exponential population growth is also expected to result in an excess of low-frequency variants [241]. 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 [239]. 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 [203]. Broadly speaking, excess nonsynonymous polymorphism (πN/πS > 1) points toward diversifying or positive selection while excess synonymous polymorphism (πN/πS < 1) indicates purifying selection. When πN / πS 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 (**S1 Table, S10 Fig, S11 Fig**). 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 (**S1 Table**). 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 [242].

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 **S2 Fig** and all indels over time are shown in **S3 Fig.** 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 (**S2 Fig**).

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 [135]. 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 [137]. 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 [242]. 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 [134].

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 [122]. 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 [54,227–229], but it is in line with other estimates suggesting that few SARS-CoV-2 viruses are transmitted between humans [227,228]. 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. 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 [243,244]. 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 [137]. 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 individual infections in this way as well [245–249]. The relationship between SARS-CoV-2 viral genetic diversity and COVID-19 clinical severity remains unclear. Some have proposed a direct relationship between particular viral lineages and COVID-19 severity [250], while others postulate that host factors, like age and comorbidities such as hypertension, diabetes, and preexisting respiratory system disease, are more likely to explain variable clinical outcomes [251].

Although within-host diversity was limited in the cats evaluated here, we identify two notable variants. S H655Y and E S67S were found at 0.85% and 0.34% in the stock, but were preferentially amplified in all three index cats and were detectable at intermediate frequencies at the first-day post-inoculation. Interestingly, S H655Y is not found in any of the 18 full-genome domestic cat, tiger, and lion SARS-CoV-2 sequences available on GISAID (**S4 Fig**). S H655Y has, however, been reported in a variety of other settings, including transmission studies in a hamster model, SARS-CoV-2 tissue culture experiments [252–255], and in a stock virus passaged on Vero E6 cells [BioProject PRJNA645906, experiment numbers SRX9287152 (p1), SRX9287151 (p2),

SRX9287154 (p3a); BioProject PRJNA627977]. S H655Y additionally persisted in vivo in rhesus macaques challenged with one of these stock viruses [BioProject PRJNA645906, experiment number SRX9287155]. As of 28 December, 2020, S H655Y has been detected in 1,070 human SARS-CoV-2 viruses across 18 different countries in sequences deposited in GISAID. The majority of these sequences come from the United Kingdom (n=886) (**S5B Fig, S5C Fig**). It is important to note, however, that sampling of SARS-CoV-2 sequences is heavily biased and sequences from the COVID-19 Genomics UK consortium (COG-UK) are currently overrepresented in GISAID. Additionally, S H655Y is the 16th most common variant detected in Spike among publicly-available SARS-CoV-2 sequences (Tze Chuen Lee R. Spike glycoprotein mutation surveillance. GISAID. https://www.epicov.org/epi3/cfrontend#2ea2a6). Sequences containing S H655Y variant are found in two distinct European clusters, EU1 and EU2, suggesting it has arisen more than once (**S5A Fig**).

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 (**S12 Fig**) [252,253,256]. 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 [252]. 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 [234]. 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 outgrowth 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 for 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 [122]. 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.

SARS-CoV-2 viruses can replicate and be shed via the respiratory tract. Differences in cell types, receptor distribution, temperature and humidity along the length of the respiratory tract may favor the emergence of different viral variants. If viral populations vary genetically across anatomic location, virus collected from different parts of the respiratory tract could result in different bottleneck size estimates. In this study, we had access to nasal swabs and were therefore were only able to evaluate genetic diversity arising in the upper respiratory tract. Others have previously documented foci of influenza virus in the lower respiratory tract appear to be independent from upper respiratory tract infections [257,258]. Current insights into potential differences in the genetic composition, structure, and evolution in the upper vs. lower respiratory tract remain incomplete for both influenza viruses and SARS-CoV-2.

Large SARS-CoV-2 outbreaks in mink have been reported recently, some with “concerning” mutations that may evade human humoral immunity [259]. These mink outbreaks have resulted in the Danish authorities’ decision to cull 17 million mink as a safeguard against spill-back transmission into humans. 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 [260]. 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 [237]. 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 [252,261]. 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.

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). We thank Chelsea Crooks for her careful reading of and comments on this manuscript.

Figures, tables, and supplemental material

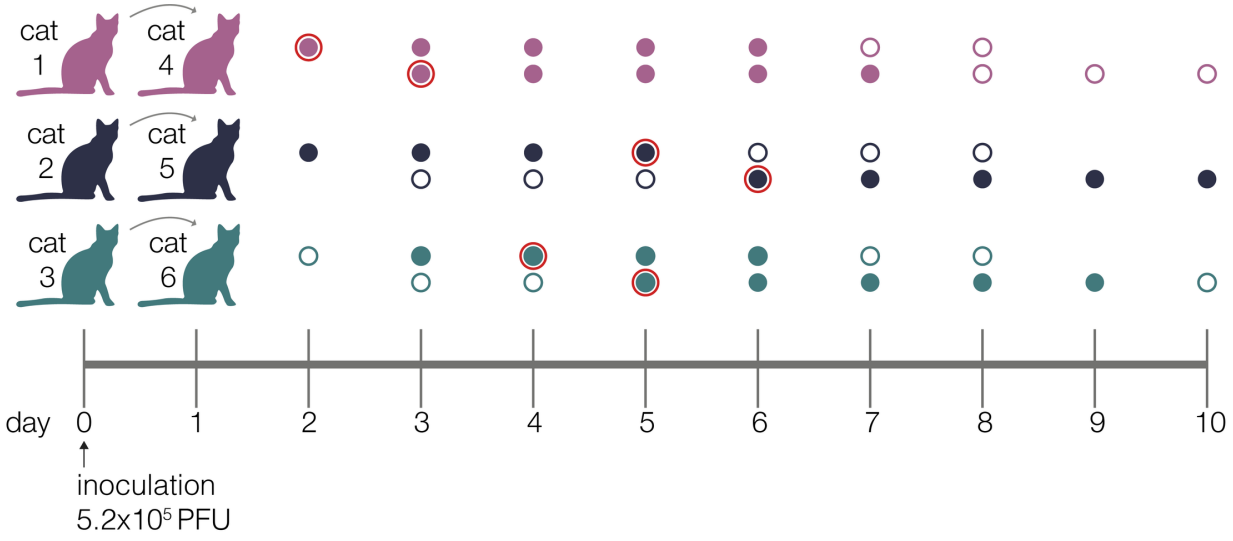


Figure 1. Experimental timeline.

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 represents 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-binomial estimate to calculate transmission bottleneck sizes.

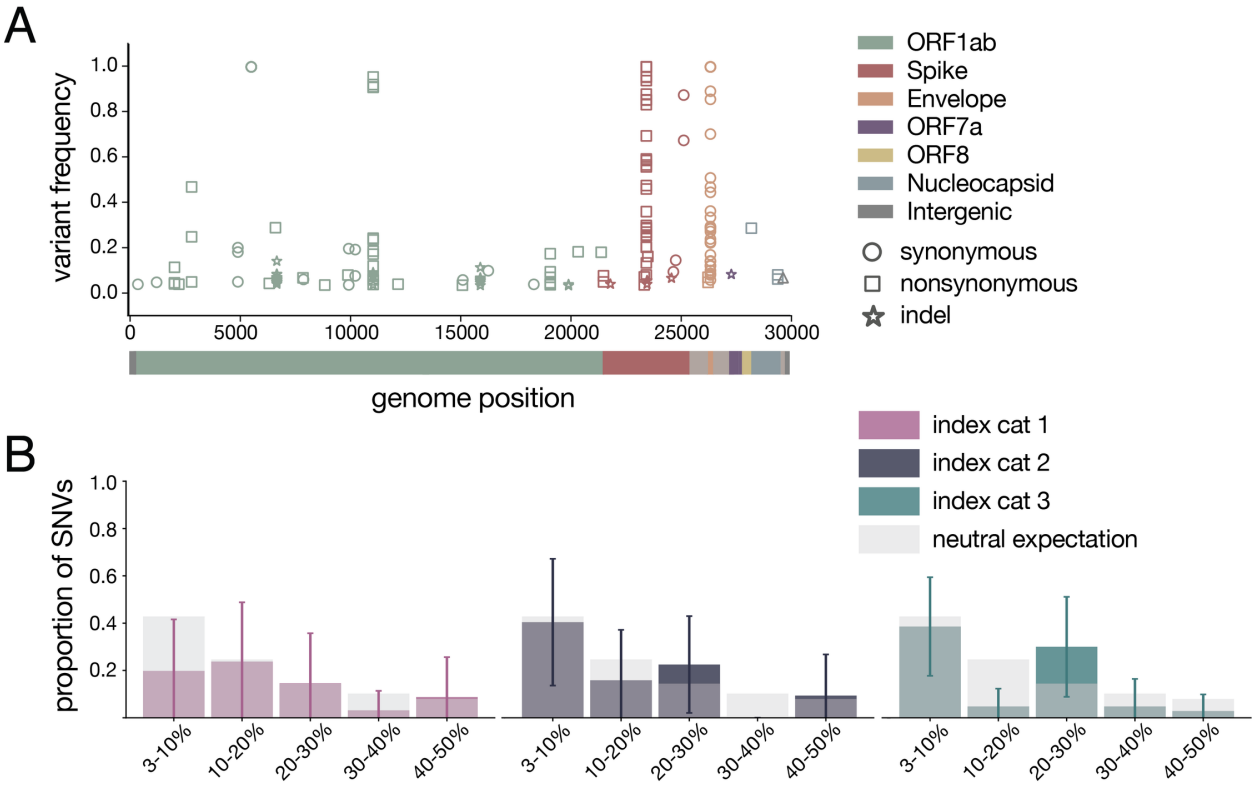


Figure 2. Within-host diversity of SARS-CoV-2 viruses in domestic cats.

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.

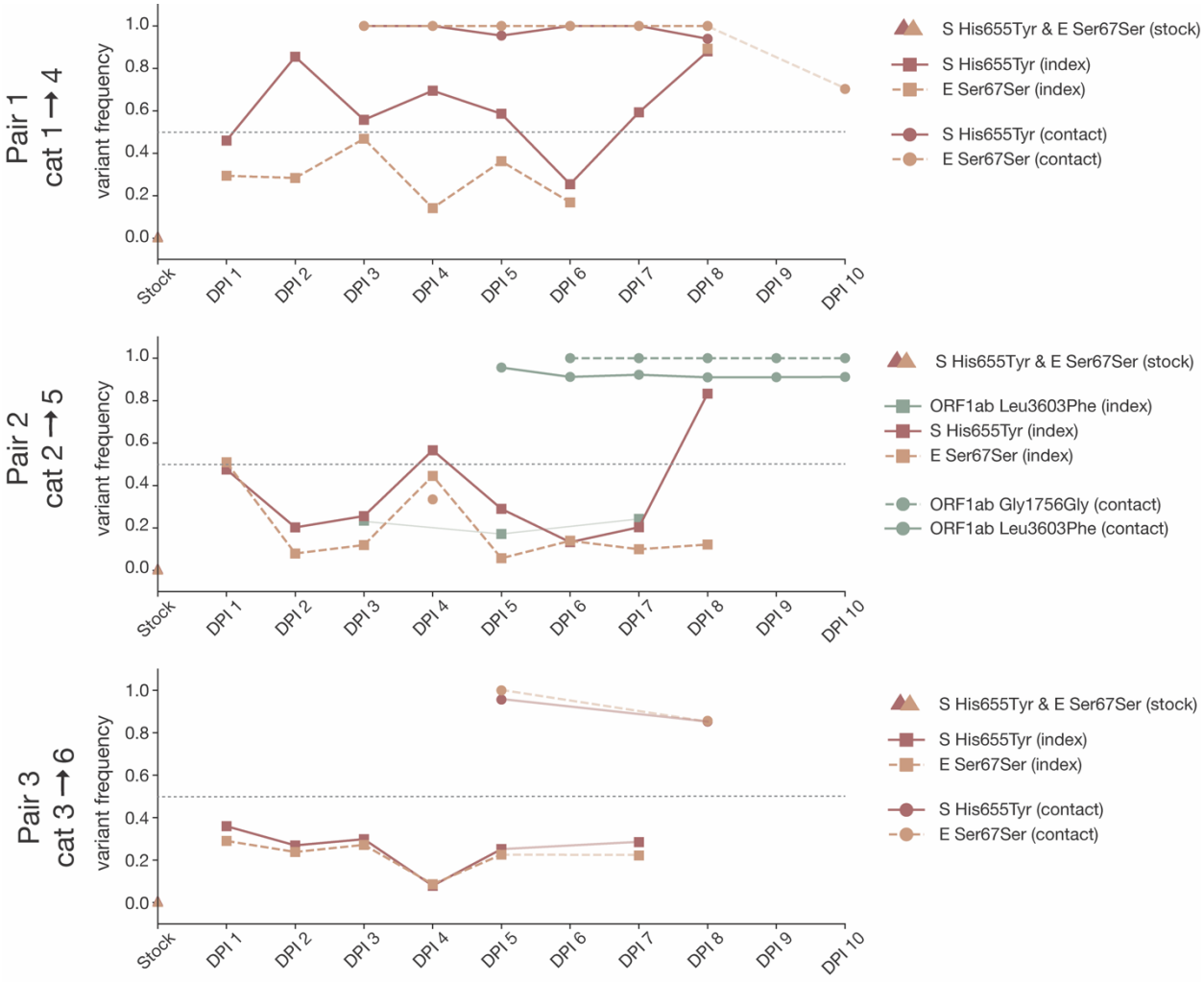


Figure 3. Frequency of iSNVs over time in each index and contact cat.

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** and all indels over time are shown in **Supplementary Figure 3**. 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.

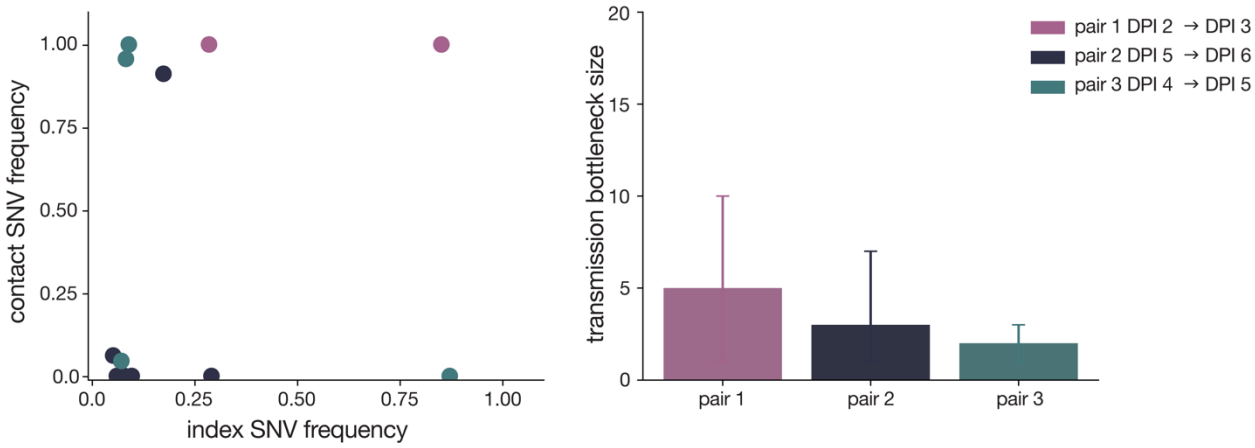
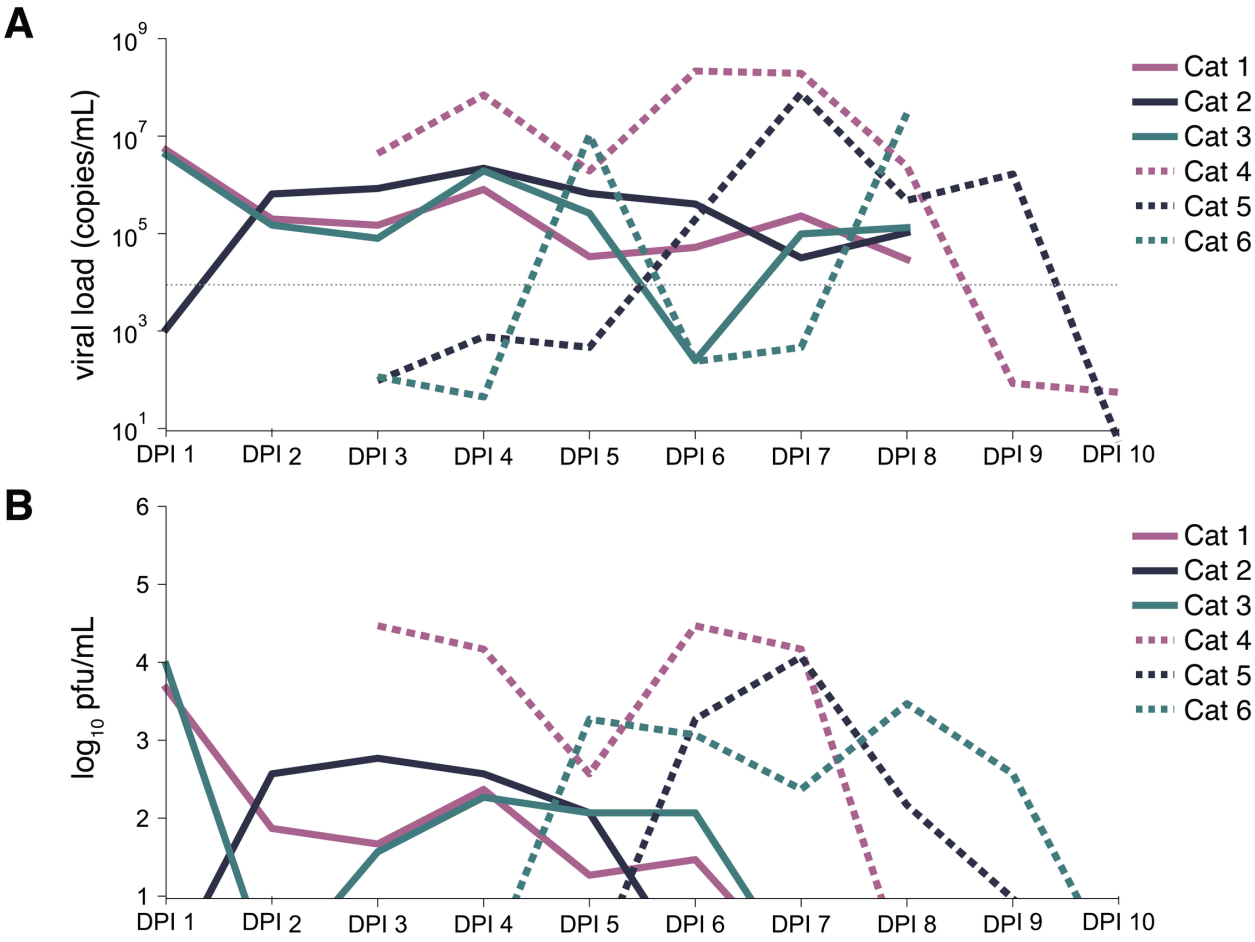
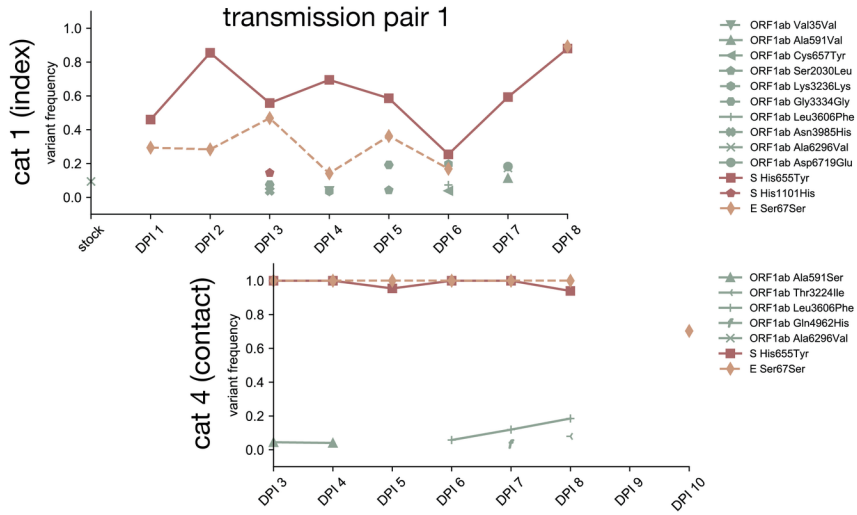


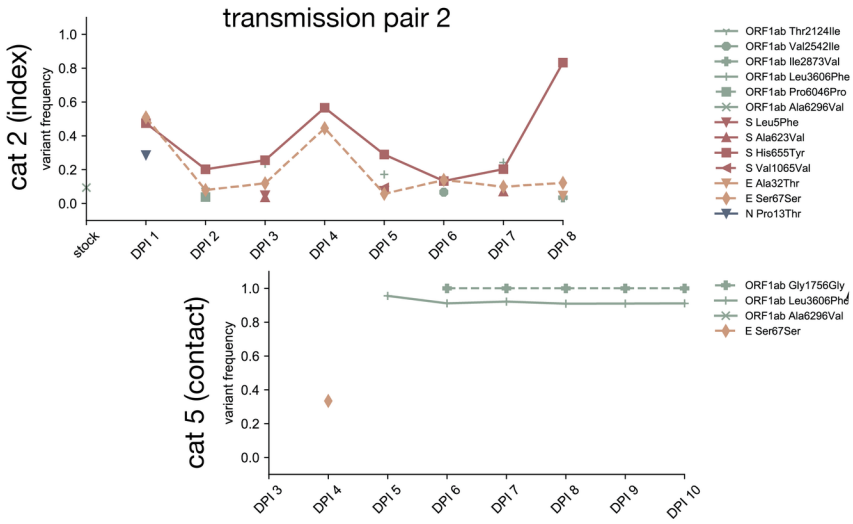
Figure 4. SARS-CoV-2 transmission is defined by a narrow bottleneck.

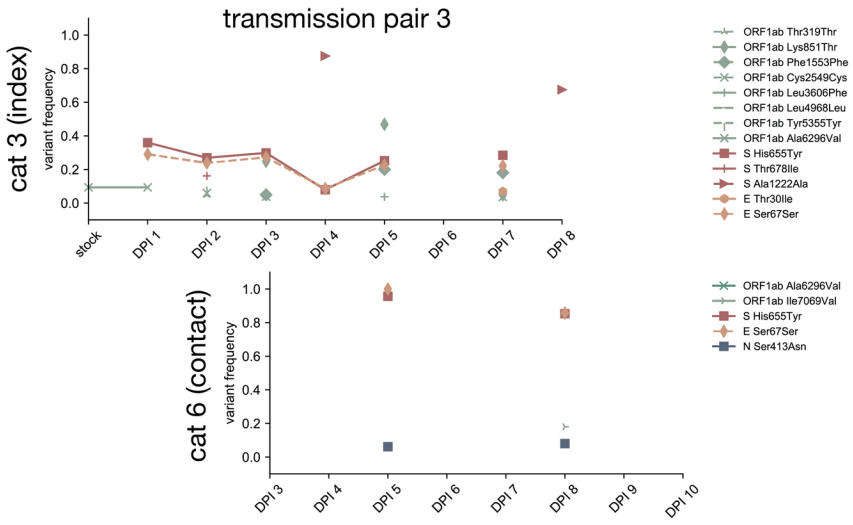
Variant frequencies in the index cats (x-axis) compared with frequencies of the same variants in the corresponding contact cats (y-axis) that were used in the beta-binomial estimate are shown on the left. Estimates of SARS-CoV-2 transmission bottleneck with 99% confidence intervals shown on the right.



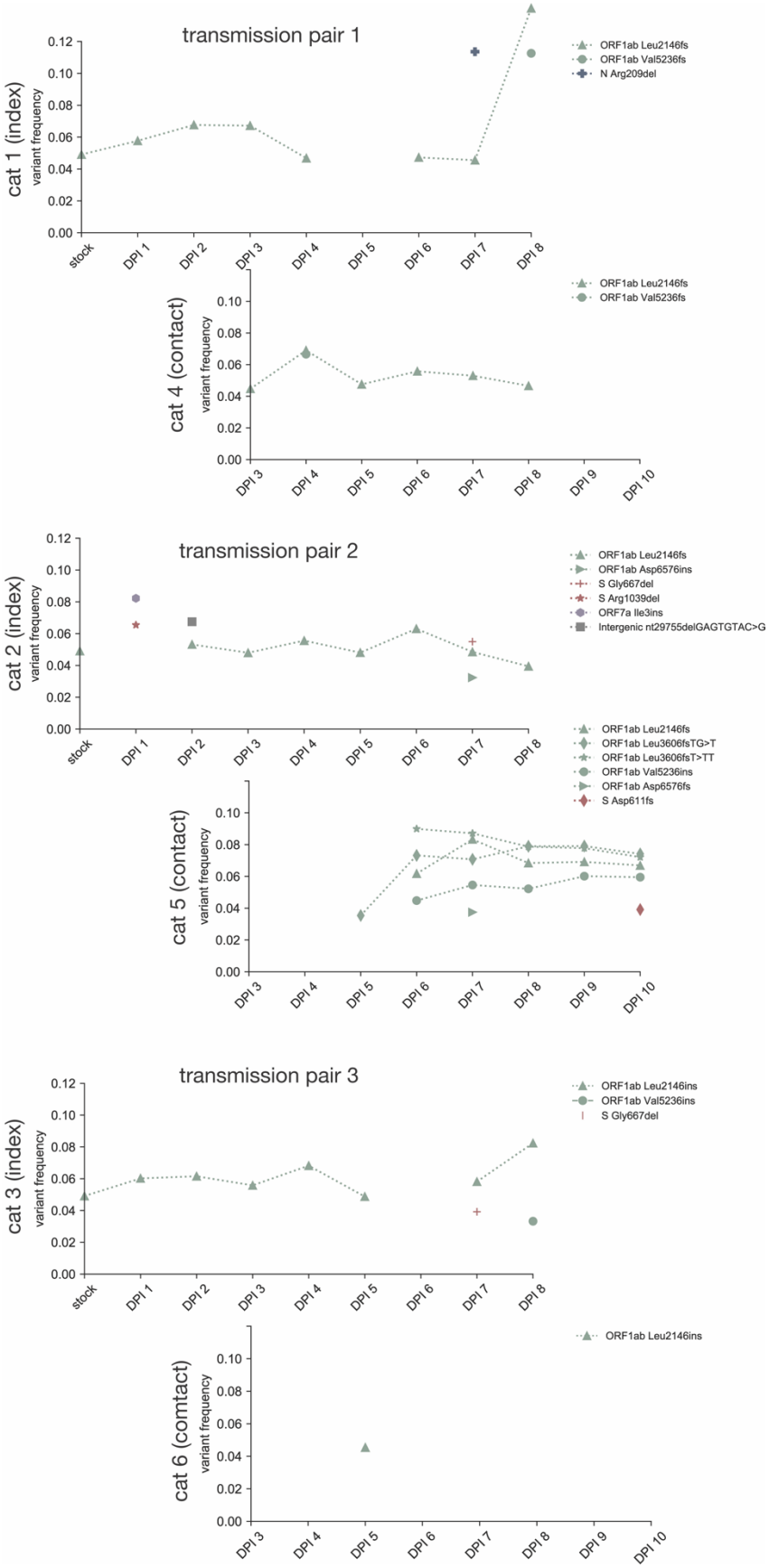
S1 Fig. Viral loads and viral titers over time. A) Viral RNA burden over time for each cat. Index cats are represented by a solid line and contact cats are represented by a dashed line. Transmission pairs are denoted by color. The grey, horizontal dotted line represents when less than ~100 copies/μL are input into the reverse transcription reaction. B) Infectious viral titer over time. Index cats are represented by a solid line and contact cats are represented by a dashed line. Transmission pairs are denoted by color.


transmission pair 1
cat 1 (index)
variant frequency
1.0
0.8
0.6
0.4
0.2
0.0
stock
DPI 1
DPI 2
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
ORF1ab Val35Val
ORF1ab Ala591Val
ORF1ab Cys657Tyr
ORF1ab Ser2030Leu
ORF1ab Lys3236Lys
ORF1ab Gly3334Gly
ORF1ab Leu3606Phe
ORF1ab Asn3985His
ORF1ab Ala6296Val
ORF1ab Asp6719Glu
S His655Tyr
S His1101His
E Ser67Ser
cat 4 (contact)
variant frequency
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
DPI 9
DPI 10
ORF1ab Ala591Ser
ORF1ab Thr3224Ile
ORF1ab Leu3606Phe
ORF1ab Gln4962His
ORF1ab Ala6296Val
S His655Tyr
E Ser67Ser

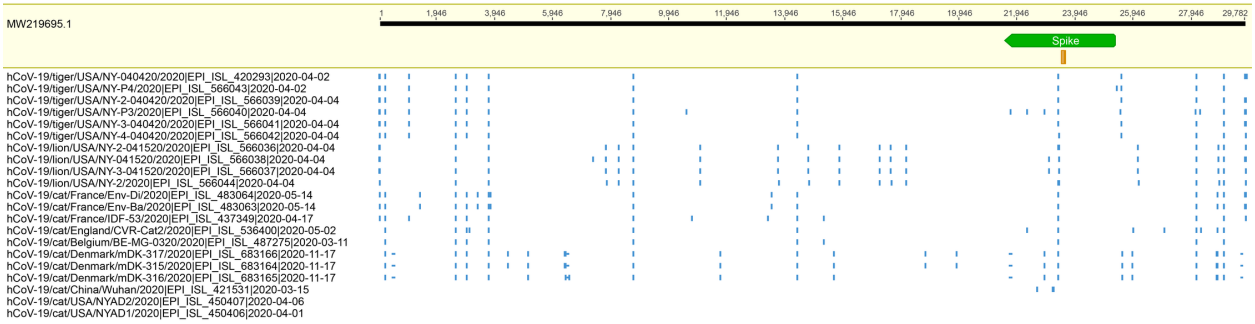

transmission pair 2
cat 2 (index)
variant frequency
1.0
0.8
0.6
0.4
0.2
0.0
stock
DPI 1
DPI 2
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
ORF1ab Thr2124Ile
ORF1ab Val2542Ile
ORF1ab Ile2873Val
ORF1ab Leu3606Phe
ORF1ab Pro6046Pro
ORF1ab Ala6296Val
S Leu5Phe
S Ala623Val
S His655Tyr
S Val1065Val
E Ala32Thr
E Ser67Ser
N Pro13Thr
cat 5 (contact)
variant frequency
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
DPI 9
DPI 10
ORF1ab Gly1756Gly
ORF1ab Leu3606Phe
ORF1ab Ala6296Val
E Ser67Ser


transmission pair 3
cat 3 (index)
variant frequency
1.0
0.8
0.6
0.4
0.2
0.0
stock
DPI 1
DPI 2
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
ORF1ab Thr319Thr
ORF1ab Lys851Thr
ORF1ab Phe1553Phe
ORF1ab Cys2549Cys
ORF1ab Leu3606Phe
ORF1ab Leu4968Leu
ORF1ab Tyr5355Tyr
ORF1ab Ala6296Val
S His655Tyr
S Thr678Ile
S Ala1222Ala
E Thr30Ile
E Ser67Ser
cat 6 (contact)
variant frequency
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
DPI 9
DPI 10
ORF1ab Ala6296Val
ORF1ab Ile7069Val
S His655Tyr
E Ser67Ser
N Ser413Asn

S2 Fig. Longitudinal frequency of iSNVs detected in all cats and at all timepoints. Each variant is colored based on gene location. Nonsynonymous variants are plotted with solid lines and synonymous variants are plotted with dashed lines. Days with viral loads too low to yield high quality sequences are shown by the gaps in data (i.e. cat 3 day 6 and cat 4 day 9).

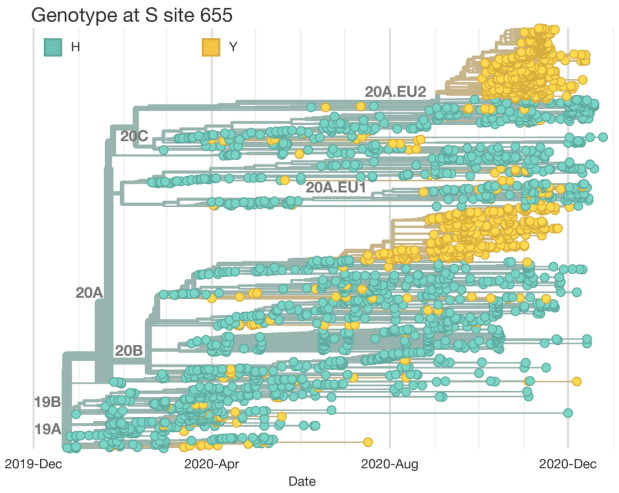

transmission pair 1
cat 1 (index)
variant frequency
0.12
0.10
0.08
0.06
0.04
0.02
0.00
stock
DPI 1
DPI 2
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
ORF1ab Leu2146fs
ORF1ab Val5236fs
N Arg209del
cat 4 (contact)
variant frequency
DPI 3
DPI 4
DPI 5
DPI 6
DPI 7
DPI 8
DPI 9
DPI 10
ORF1ab Leu2146fs
ORF1ab Val5236fs
transmission pair 2
cat 2 (index)
variant frequency
ORF1ab Leu2146fs
ORF1ab Asp6576ins
S Gly667del
S Arg1039del
ORF7a Ile3ins
Intergenic nt29755delGAGTGTAC>G
cat 5 (contact)
variant frequency
ORF1ab Leu2146fs
ORF1ab Leu3606fsTG>T
ORF1ab Leu3606fsT>TT
ORF1ab Val5236ins
ORF1ab Asp6576fs
S Asp611fs
transmission pair 3
cat 3 (index)
variant frequency
ORF1ab Leu2146ins
ORF1ab Val5236ins
S Gly667del
cat 6 (comtact)
variant frequency
ORF1ab Leu2146ins

S3 Fig. Longitudinal frequency of indels detected in all cats and at all timepoints. Each indel is colored based on gene location. Days with viral loads too low to yield high quality sequences are shown by the gaps in data (i.e. cat 3 day 6 and cat 4 day 9). Note the y-axis range is 0-12%, not 0-100%, to facilitate readability.



S4 Fig. Sequence alignment of all tiger, lion, and domestic cat sequences available in GISAID as of December 2020. Sequences were aligned against MW219695.1, the inoculum virus used in these experiments. Consensus-level differences are highlighted with a blue vertical marker. Indels are noted with a horizontal vertical marker. The spike open reading frame is annotated with a green marker and site amino acid 655 in Spike is highlighted with the orange box. None of these sequences contain a consensus mutation at residue 655 in Spike.

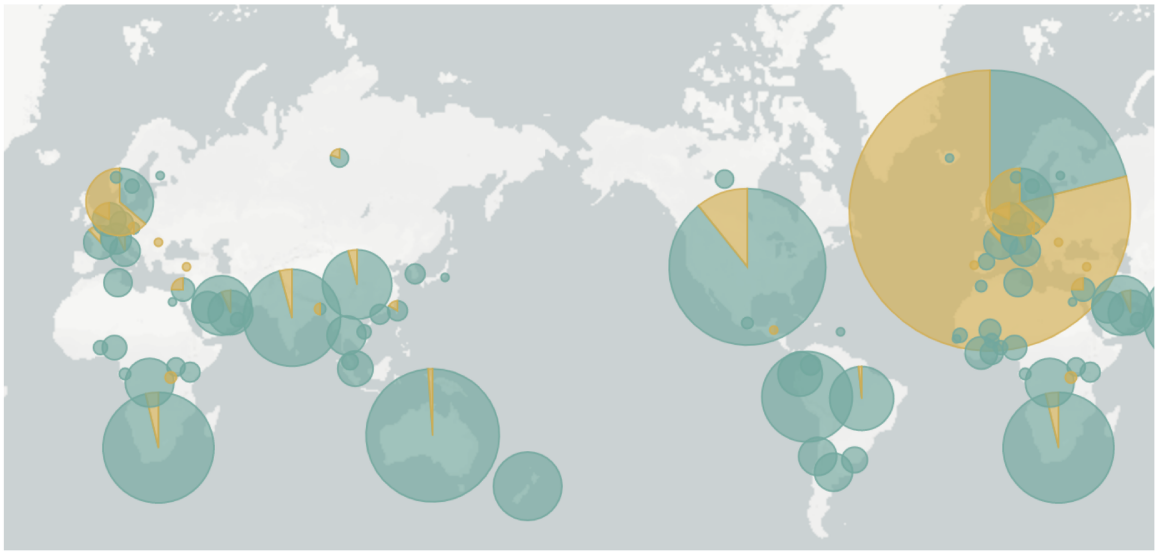
A



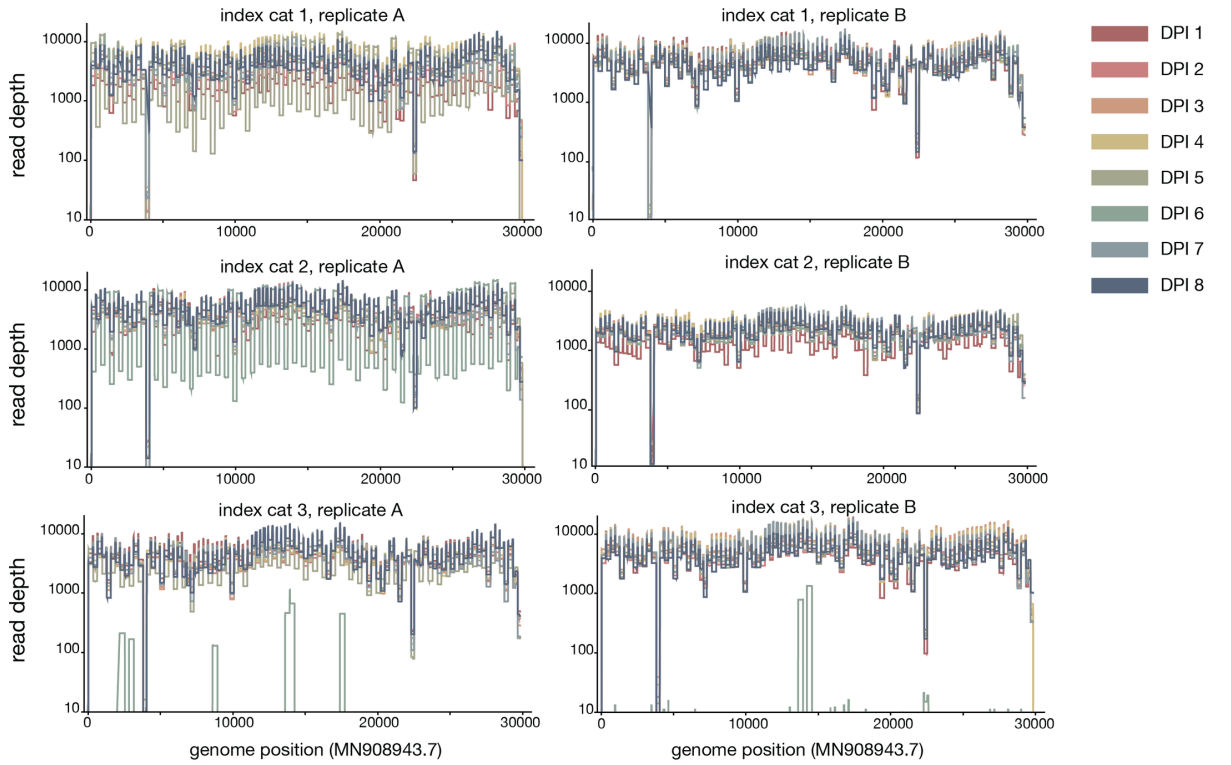
B

Country	Date of first detected sequence with H655Y	Number of sequences	Date of most recent detected sequence with H655Y
United Kingdom	2020-03-13	943	2020-12-13
Denmark	2020-04-13	44	2020-11-30
USA	2020-03-01	42	2020-11-12
South Africa	2020-06-11	8	2020-10-9
India	2020-05-11	6	2020-9-2
China	2020-01-26	3	2020-3-13
Netherlands	2020-05-07	3	2020-12-7
Australia	2020-07-31	3	2020-8-12
Israel	2020-03-24	2	2020-4-5
United Arab Emirates	2020-05-15	2	2020-5-18
France	2020-05-26	2	2020-8-14
Rwanda	2020-10-17	2	2020-10-22
Taiwan	2020-03-19	1	2020-3-19
Portugal	2020-04-02	1	2020-4-2
Belize	2020-04-09	1	2020-4-9
Romania	2020-04-09	1	2020-4-9
Turkey	2020-04-26	1	2020-4-26
Brazil	2020-06-30	1	2020-6-30
Bangladesh	2020-07-12	1	2020-7-12
Czech Republic	2020-11-04	1	2020-11-4
Russia	2020-11-16	1	2020-11-16
Italy	2020-12-02	1	2020-12-2

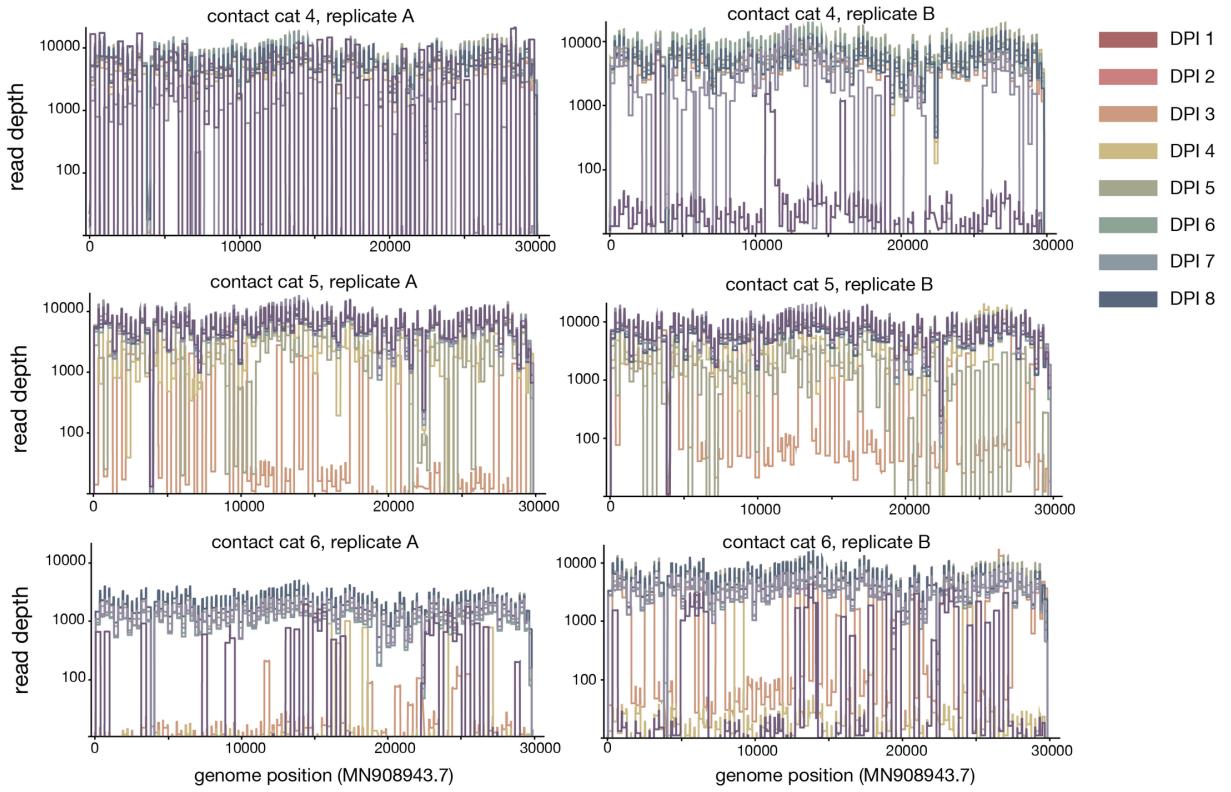
C



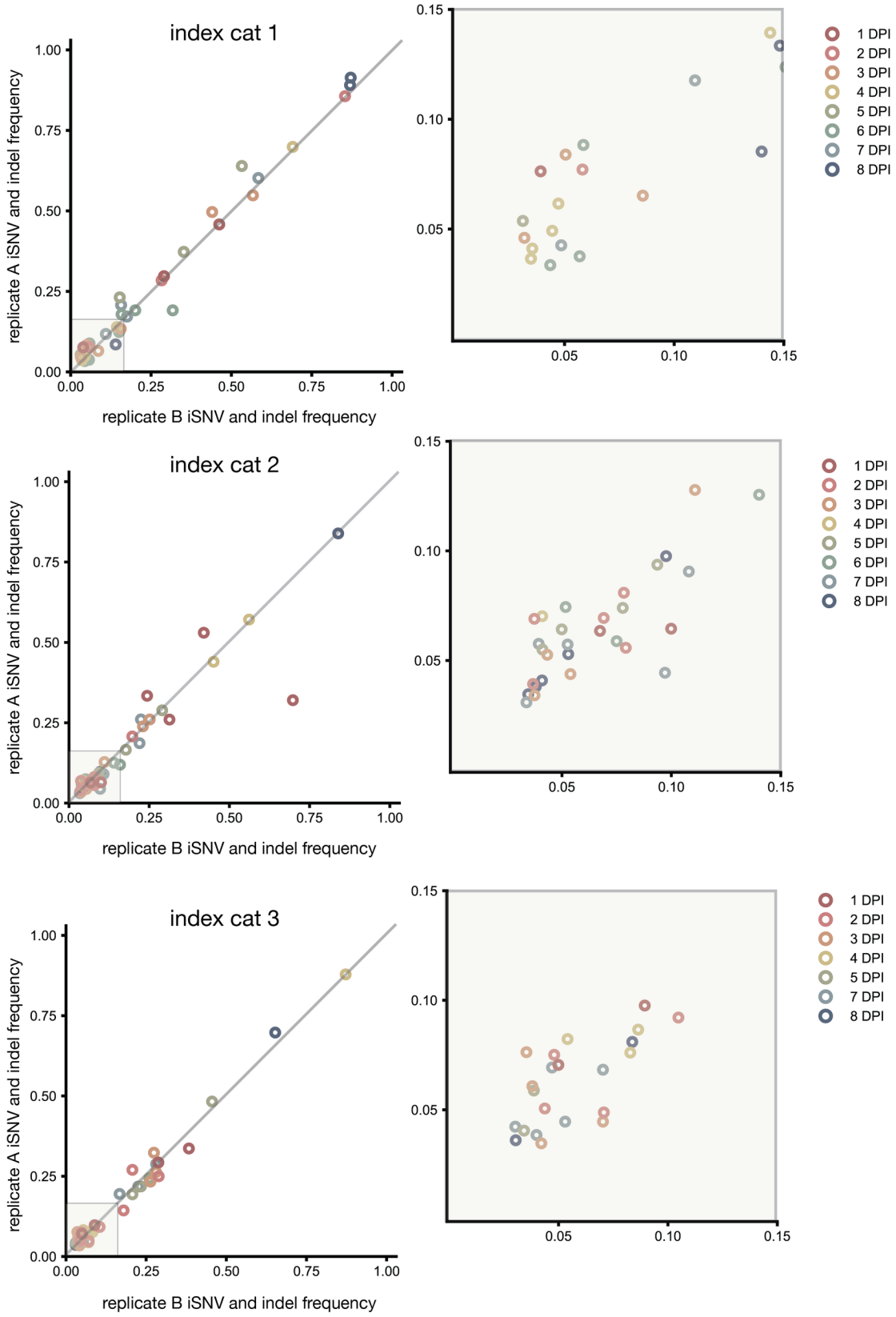
S5 Fig. Geographic dispersion of Spike H655Y variant. A) A time-resolved phylogeny focused on viruses that contain Spike H655Y. Viruses that contain histidine (H) at Spike 655 are colored in teal. Viruses with tyrosine (Y) at Spike 655 are colored in yellow. B) Counts of SARS-CoV-2 viruses that contain Spike H655Y, broken down by country. C) Map highlighting the number viruses from each country. The size of the circle represents the number of sequences from the appropriate country contained in the phylogeny.



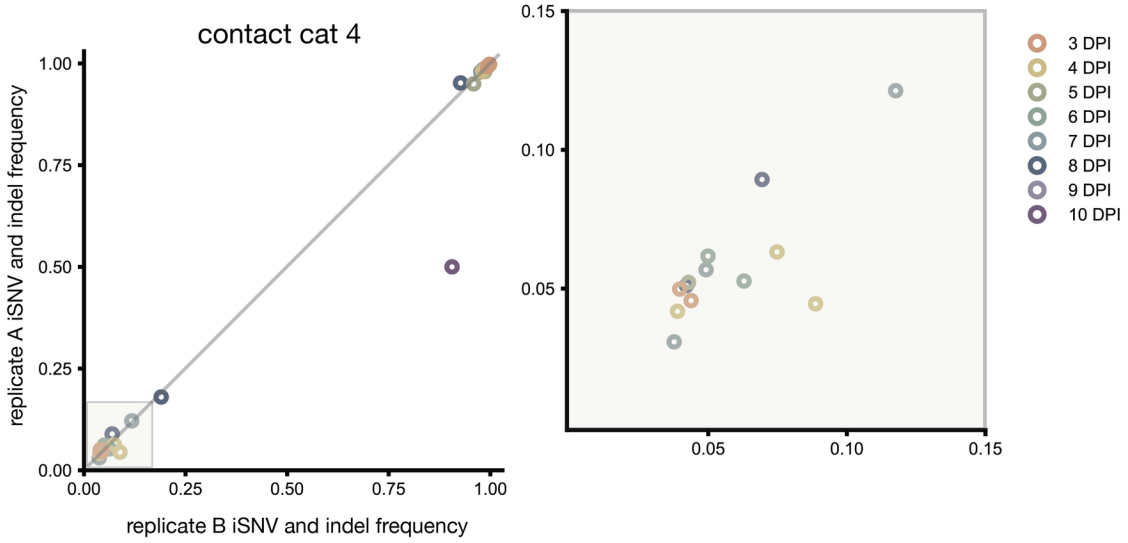
S6 Fig. Read depth across the SARS-CoV-2 genome in index cats. Each day is represented by a different color. Replicate A is shown in the left column and replicate B is shown in the right column.

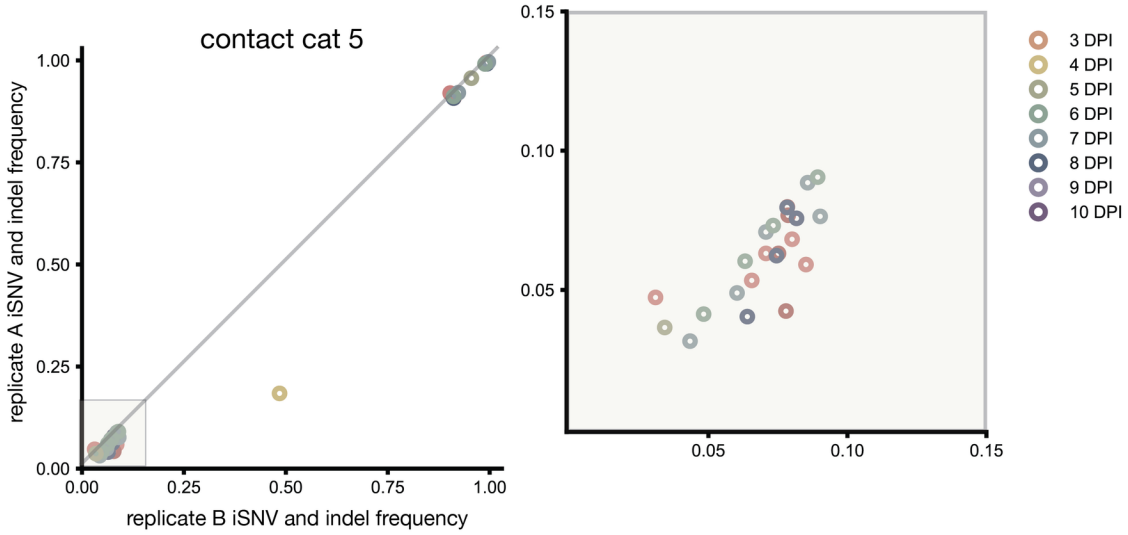


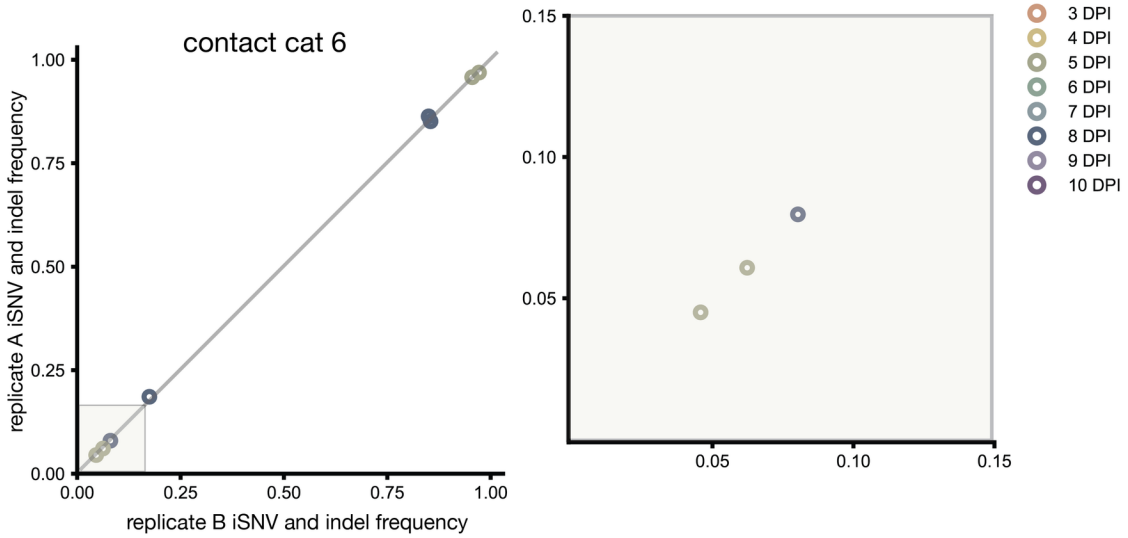
S7 Fig. Read depth across the SARS-CoV-2 genome in contact cats. Each day is represented by a different color. Replicate A is shown in the left column and replicate B is shown in the right column.


index cat 1
index cat 2
index cat 3
replicate A iSNV and indel frequency
replicate B iSNV and indel frequency
0.00
0.25
0.50
0.75
1.00
0.05
0.10
0.15
1 DPI
2 DPI
3 DPI
4 DPI
5 DPI
6 DPI
7 DPI
8 DPI

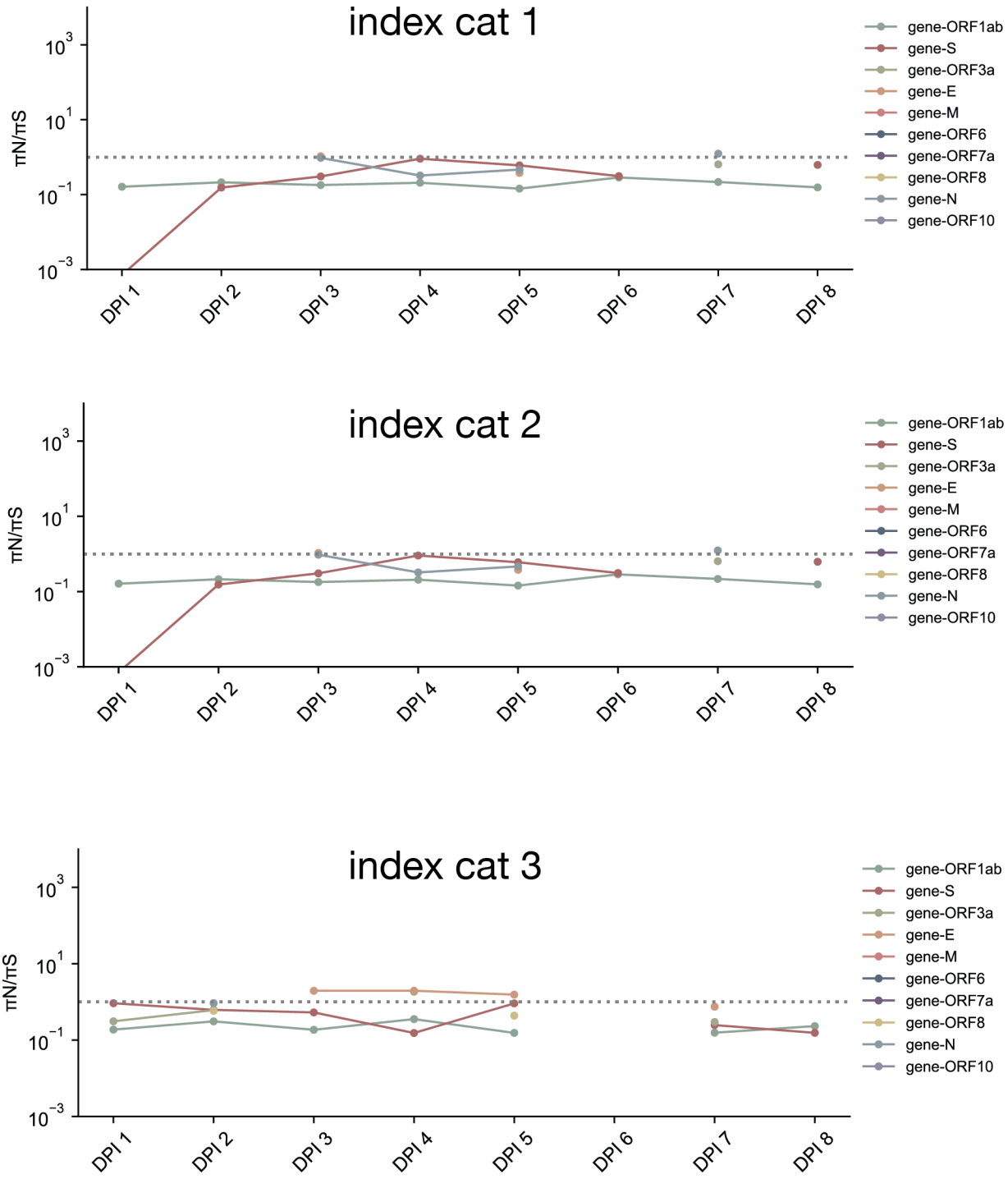
S8 Fig. Intersection variants found across technical replicates in index cats. The frequency of each variant per replicate is shown here. The diagonal line represents the 1:1 intersection of replicate variants. The subplot to the right of each primary plot is a zoomed-in view of the low-frequency variants (3-15%). Each timepoint is denoted by a different color.


contact cat 4
replicate A iSNV and indel frequency
replicate B iSNV and indel frequency
0.00
0.25
0.50
0.75
1.00
0.05
0.10
0.15
3 DPI
4 DPI
5 DPI
6 DPI
7 DPI
8 DPI
9 DPI
10 DPI

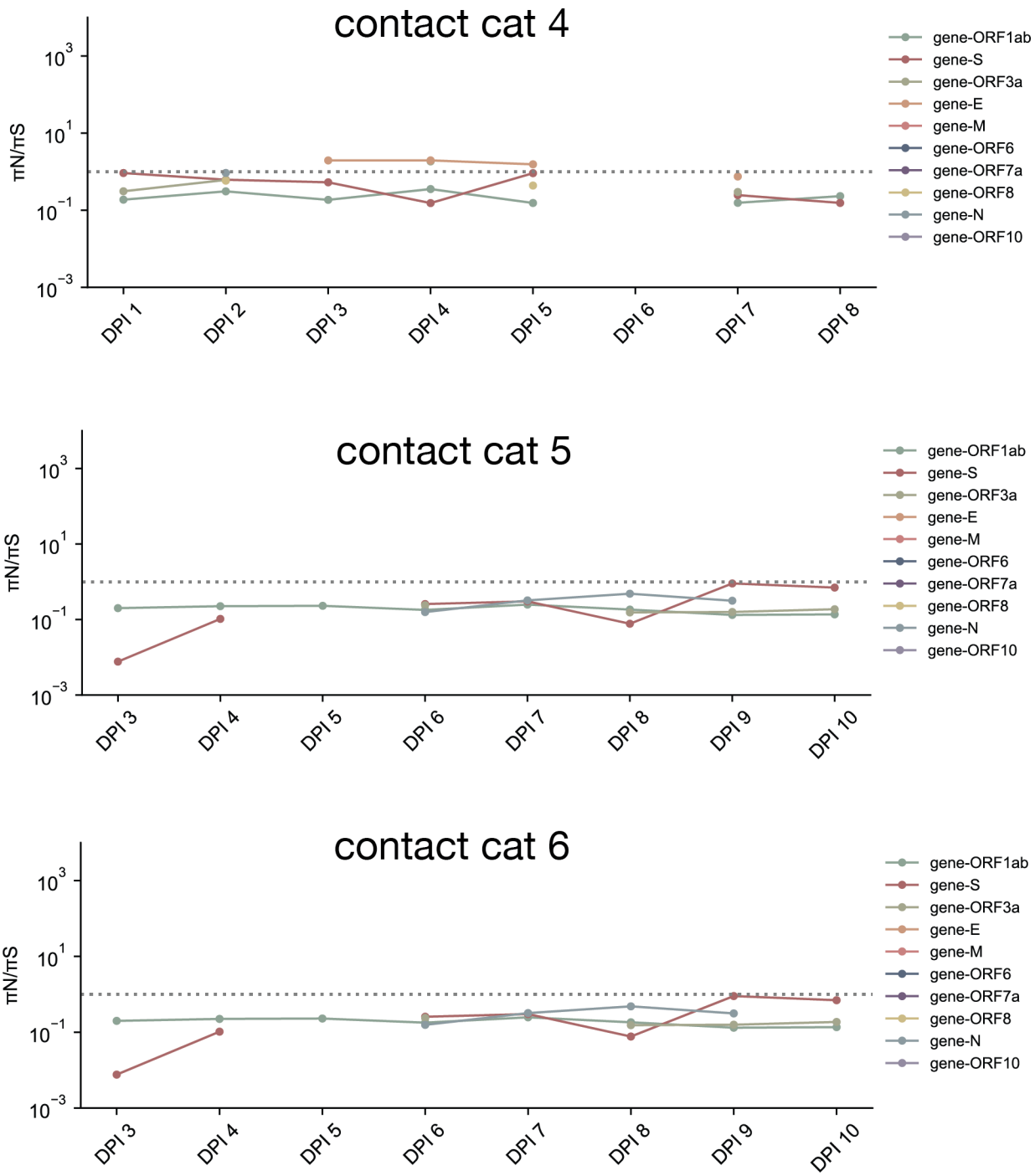

contact cat 5
replicate A iSNV and indel frequency
replicate B iSNV and indel frequency
0.00
0.25
0.50
0.75
1.00
0.05
0.10
0.15
3 DPI
4 DPI
5 DPI
6 DPI
7 DPI
8 DPI
9 DPI
10 DPI


contact cat 6
replicate A iSNV and indel frequency
replicate B iSNV and indel frequency
0.00
0.25
0.50
0.75
1.00
0.05
0.10
0.15
3 DPI
4 DPI
5 DPI
6 DPI
7 DPI
8 DPI
9 DPI
10 DPI

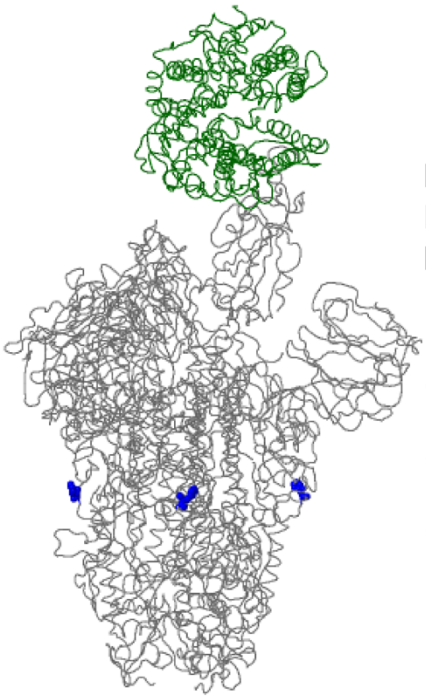
S9 Fig. Intersection variants found across technical replicates in contact cats. The frequency of each variant per replicate is shown here. The diagonal line represents the 1:1 intersection of replicate variants. The subplot to the right of each primary plot is a zoomed-in view of the low-frequency variants (3-15%). Each timepoint is denoted by a different color.



S10 Fig. Longitudinal pairwise nonsynonymous nucleotide diversity divided by pairwise synonymous nucleotide diversity in index cats. Line color denotes gene. The horizontal dotted gray line is plotted at y = 1 or when πN ~ πS.



S11 Fig. Longitudinal pairwise nonsynonymous nucleotide diversity divided by pairwise synonymous nucleotide diversity in contact cats. Line color denotes gene. The horizontal dotted gray line is plotted at y = 1 or when πN ~ πS.



ACE2 human rececptor
CoV spike glycoprotein trimer
Spike H655Y

Hanada* grouping scheme	Radical replacement?	H residue group	Y residue group
Polarity and Volume	Radical	Polar and relatively large	Nonpolar and relatively large
Charge and Aromaticity	Radical	Basic	Neutral and aromaticity
Charge and Polarity	Radical	Basic and polarity	Neutral and polarity

S12 Fig. SARS-CoV-2 spike glycoprotein crystal structure. Spike H655Y is highlighted in blue. The table to the right of the crystal structure includes summary information regarding the impact of a histidine to tyrosine change on amino acid charge, volume, and aromaticity. * Qualitative definitions of radical amino acid replacements, based on three alternative residue groupings, see Hanada et al., 2006 [262]. The crystal structure and summary information were generated using GISAID's CoVserver mutation analysis tool.

	mean πS	std πS	mean πN	std πN	πN/πS	statistic, p-value
index cat 1						
ORF1ab	0.015670	0.005609	0.003019	0.001124	0.195265	(statistic=6.23, pvalue=2.11e-05)
S	0.001564	0.000600	0.000644	0.000467	0.413995	(statistic=3.34, pvalue=0.005)
ORF3a	0.005367	0.000001	0.005878	0.002267	0.641899	(statistic=-0.303, pvalue=0.772)
E	0.011707	0.010139	0.011930	0.005719	0.719601	(statistic=-0.0456, pvalue=0.965)
M			0.002138	0.000059		(statistic=nan, pvalue=nan)
ORF6			0.007395			(statistic=nan, pvalue=nan)
ORF7a	0.011992	0.000487				(statistic=nan, pvalue=nan)
ORF8	0.031186	0.012756				(statistic=nan, pvalue=nan)
N	0.005292	0.002941	0.003036	0.001403	0.744626	(statistic=1.410, pvalue=0.196)
ORF10						(statistic=nan, pvalue=nan)
index cat 2						
ORF1ab	0.025570	0.007229	0.005423	0.001061	0.219071	(statistic=7.799, pvalue=1.837e-06)
S	0.004651	0.002567	0.001476	0.000556	0.457442	(statistic=3.419, pvalue=0.0042)
ORF3a	0.008660	0.003052	0.003867	0.001630	0.535879	(statistic=3.549, pvalue=0.0053)
E	0.008410	0.009696	0.015842	0.010396	436.372255	(statistic=-1.191, pvalue=0.261)
M			0.002479	0.000841		(statistic=nan, pvalue=nan)
ORF6			0.007468			(statistic=nan, pvalue=nan)
ORF7a	0.011872					(statistic=nan, pvalue=nan)
ORF8	0.030343	0.012115	0.005673	0.002620	0.224566	(statistic=2.733, pvalue=0.0292)
N	0.006988	0.004962	0.001398	0.000535	0.346156	(statistic=2.240, pvalue=0.066)
ORF10						(statistic=nan, pvalue=nan)
index cat 3						
ORF1ab	0.022195	0.005336	0.004752	0.000970	0.225357	(statistic=8.509, pvalue=1.988e-06)
S	0.003619	0.001697	0.001584	0.001032	0.505292	(statistic=2.711, pvalue=0.0189)
ORF3a	0.005516	0.000037	0.004351	0.003509	0.764893	(statistic=0.649, pvalue=0.533)
E	0.016174	0.000532	0.023985	0.007123	1.553581	(statistic=-2.145, pvalue=0.0642)
M	0.006492		0.002078	0.000007		(statistic=nan, pvalue=nan)
ORF6						(statistic=nan, pvalue=nan)
ORF7a	0.012294	0.000305				(statistic=nan, pvalue=nan)
ORF8	0.020154	0.006891	0.008740	0.002221	0.509847	(statistic=2.708, pvalue=0.0352)
N	0.005320	0.001974	0.003336		0.925883	(statistic=nan, pvalue=nan)
ORF10						(statistic=nan, pvalue=nan)

S1 Table. Nonsynonymous and synonymous nucleotide diversity estimates in index cats.

	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	Cat 6
π DPI 1	0.000246	0.000290	0.000433			
π DPI 2	0.000314	0.000705	0.000546			
π DPI 3	0.000458	0.000557	0.000781	0.000712	0.000153	
π DPI 4	0.000577	0.000650	0.000568	0.000796	0.000206	0.000037
π DPI 5	0.000489	0.000513	0.000540	0.001007	0.000149	0.000576
π DPI 6	0.000430	0.000720		0.000917	0.000854	0.000156
π DPI 7	0.000365	0.000541	0.000683	0.000721	0.000876	0.000025
π DPI 8	0.000214	0.000591	0.000458	0.000879	0.000872	0.000720
π DPI 9				0.000125	0.000965	0.000070
π DPI 10				0.000371	0.000932	0.000000
mean π	0.000387	0.000571	0.000573	0.000691	0.000626	0.000226
std π	0.000117	0.000128	0.000114	0.000279	0.000355	0.000273

S2 Table. Genome-wide pairwise nucleotide diversity estimates in index and contact cats

name	pool	sequence	length	%gc	tm (use 65)
nCoV-2019_1_LEFT	nCoV-2019_1	ACCAACCAACTTTCGATCTCTTGT	24	41.67	60.69
nCoV-2019_1_RIGHT	nCoV-2019_1	CATCTTTAAGATGTTGACGTGCCTC	25	44	60.45
nCoV-2019_2_LEFT	nCoV-2019_2	CTGTTTTACAGGTTCGCGACGT	22	50	61.67
nCoV-2019_2_RIGHT	nCoV-2019_2	TAAGGATCAGTGCCAAGCTCGT	22	50	61.74
nCoV-2019_3_LEFT	nCoV-2019_1	CGGTAATAAAGGAGCTGGTGGC	22	54.55	61.32
nCoV-2019_3_RIGHT	nCoV-2019_1	AAGGTGTCTGCAATTCATAGCTCT	24	41.67	60.32
nCoV-2019_4_LEFT	nCoV-2019_2	GGTGTATACTGCTGCCGTGAAC	22	54.55	61.56
nCoV-2019_4_RIGHT	nCoV-2019_2	CACAAGTAGTGGCACCTTCTTTAGT	25	44	60.97
nCoV-2019_5_LEFT	nCoV-2019_1	TGGTGAAACTTCATGGCAGACG	22	50	61.39
nCoV-2019_5_RIGHT	nCoV-2019_1	ATTGATGTTGACTTTCTCTTTTTGGAGT	28	32.14	60.17
nCoV-2019_6_LEFT	nCoV-2019_2	GGTGTTGTTGGAGAAGGTTCCG	22	54.55	61.64

nCoV-2019_6_RIGHT	nCoV-2019_2	TAGCGGCCTTCTGTAAAACACG	22	50	61.18
nCoV-2019_7_LEFT	nCoV-2019_1	ATCAGAGGCTGCTCGTGTTGTA	22	50	61.73
nCoV-2019_7_LEFT_alt0	nCoV-2019_1	CATTTGCATCAGAGGCTGCTCG	22	54.55	62.44
nCoV-2019_7_RIGHT	nCoV-2019_1	TGCACAGGTGACAATTTGTCCA	22	45.45	60.95
nCoV-2019_7_RIGHT_alt5	nCoV-2019_1	AGGTGACAATTTGTCCACCGAC	22	50	61.07
nCoV-2019_8_LEFT	nCoV-2019_2	AGAGTTTCTTAGAGACGGTTGGGA	24	45.83	61
nCoV-2019_8_RIGHT	nCoV-2019_2	GCTTCAACAGCTTCACTAGTAGGT	24	45.83	60.56
nCoV-2019_9_LEFT	nCoV-2019_1	TCCCACAGAAGTGTTAACAGAGGA	24	45.83	61.18
nCoV-2019_9_LEFT_alt4	nCoV-2019_1	TTCCCACAGAAGTGTTAACAGAGG	24	45.83	60.44
nCoV-2019_9_RIGHT	nCoV-2019_1	ATGACAGCATCTGCCACAACAC	22	50	61.71
nCoV-2019_9_RIGHT_alt2	nCoV-2019_1	GACAGCATCTGCCACAACACAG	22	54.55	62.26
nCoV-2019_10_LEFT	nCoV-2019_2	TGAGAAGTGCTCTGCCTATACAGT	24	45.83	61.12
nCoV-2019_10_RIGHT	nCoV-2019_2	TCATCTAACCAATCTTCTTCTTGCTCT	27	37.04	60.31
nCoV-2019_11_LEFT	nCoV-2019_1	GGAATTTGGTGCCACTTCTGCT	22	50	61.66
nCoV-2019_11_RIGHT	nCoV-2019_1	TCATCAGATTCAACTTGCATGGCA	24	41.67	61.35
nCoV-2019_12_LEFT	nCoV-2019_2	AAACATGGAGGAGGTGTTGCAG	22	50	61.08
nCoV-2019_12_RIGHT	nCoV-2019_2	TTCACTCTTCATTTCCAAAAAGCTTGA	27	33.33	60.36
nCoV-2019_13_LEFT	nCoV-2019_1	TCGCACAAATGTCTACTTAGCTGT	24	41.67	60.56
nCoV-2019_13_RIGHT	nCoV-2019_1	ACCACAGCAGTTAAAACACCCT	22	45.45	60.36
nCoV-2019_14_LEFT	nCoV-2019_2	CATCCAGATTCTGCCACTCTTGT	23	47.83	60.62
nCoV-2019_14_LEFT_alt4	nCoV-2019_2	TGGCAATCTTCATCCAGATTCTGC	24	45.83	61.47
nCoV-2019_14_RIGHT	nCoV-2019_2	AGTTTCCACACAGACAGGCATT	22	45.45	60.42
nCoV-2019_14_RIGHT_alt2	nCoV-2019_2	TGCGTGTTTCTTCTGCATGTGC	22	50	62.76
nCoV-2019_15_LEFT	nCoV-2019_1	ACAGTGCTTAAAAAGTGTAAAAGTGCC	27	37.04	61.32
nCoV-2019_15_LEFT_alt1	nCoV-2019_1	AGTGCTTAAAAAGTGTAAAAGTGCCT	26	34.62	60.13
nCoV-2019_15_RIGHT	nCoV-2019_1	AACAGAAACTGTAGCTGGCACT	22	45.45	60.16
nCoV-2019_15_RIGHT_alt3	nCoV-2019_1	ACTGTAGCTGGCACTTTGAGAGA	23	47.83	61.57
nCoV-2019_16_LEFT	nCoV-2019_2	AATTTGGAAGAAGCTGCTCGGT	22	45.45	60.82
nCoV-2019_16_RIGHT	nCoV-2019_2	CACAACTTGCGTGTGGAGGTTA	22	50	61.32
nCoV-2019_17_LEFT	nCoV-2019_1	CTTCTTTCTTTGAGAGAAGTGAGGACT	27	40.74	60.69
nCoV-2019_17_RIGHT	nCoV-2019_1	TTTGTTGGAGTGTTAACAATGCAGT	25	36	60.11
nCoV-2019_18_LEFT	nCoV-2019_2	TGGAAATACCCACAAGTTAATGGTTTAAC	29	34.48	60.69
nCoV-2019_18_LEFT_alt2	nCoV-2019_2	ACTTCTATTAAATGGGCAGATAACAACTGT	30	33.33	61.38
nCoV-2019_18_RIGHT	nCoV-2019_2	AGCTTGTTTACCACACGTACAAGG	24	45.83	61.51
nCoV-2019_18_RIGHT_alt1	nCoV-2019_2	GCTTGTTTACCACACGTACAAGG	23	47.83	60.3
nCoV-2019_19_LEFT	nCoV-2019_1	GCTGTTATGTACATGGGCACACT	23	47.83	61.18
nCoV-2019_19_RIGHT	nCoV-2019_1	TGTCCAACTTAGGGTCAATTTCTGT	25	40	60.4
nCoV-2019_20_LEFT	nCoV-2019_2	ACAAAGAAAACAGTTACACAACAACCA	27	33.33	60.68
nCoV-2019_20_RIGHT	nCoV-2019_2	ACGTGGCTTTATTAGTTGCATTGTT	25	36	60.28

nCoV-2019_21_LEFT	nCoV-2019_1	TGGCTATTGATTATAAACACTACACACCC	29	37.93	61.49
nCoV-2019_21_LEFT_alt2	nCoV-2019_1	GGCTATTGATTATAAACACTACACACCCT	29	37.93	61.29
nCoV-2019_21_RIGHT	nCoV-2019_1	TAGATCTGTGTGGCCAACCTCT	22	50	60.83
nCoV-2019_21_RIGHT_alt0	nCoV-2019_1	GATCTGTGTGGCCAACCTCTTC	22	54.55	61.2
nCoV-2019_22_LEFT	nCoV-2019_2	ACTACCGAAGTTGTAGGAGACATTATACT	29	37.93	61.25
nCoV-2019_22_RIGHT	nCoV-2019_2	ACAGTATTCTTTGCTATAGTAGTCGGC	27	40.74	60.73
nCoV-2019_23_LEFT	nCoV-2019_1	ACAACTACTAACATAGTTACACGGTGT	27	37.04	60.26
nCoV-2019_23_RIGHT	nCoV-2019_1	ACCAGTACAGTAGGTTGCAATAGTG	25	44	60.57
nCoV-2019_24_LEFT	nCoV-2019_2	AGGCATGCCTTCTTACTGTACTG	23	47.83	60.37
nCoV-2019_24_RIGHT	nCoV-2019_2	ACATTCTAACCATAGCTGAAATCGGG	26	42.31	61.19
nCoV-2019_25_LEFT	nCoV-2019_1	GCAATTGTTTTTCAGCTATTTTGCAGT	27	33.33	60.73
nCoV-2019_25_RIGHT	nCoV-2019_1	ACTGTAGTGACAAGTCTCTCGCA	23	47.83	61.3
nCoV-2019_26_LEFT	nCoV-2019_2	TTGTGATACATTCTGTGCTGGTAGT	25	40	60.28
nCoV-2019_26_RIGHT	nCoV-2019_2	TCCGCACTATCACCAACATCAG	22	50	60.42
nCoV-2019_27_LEFT	nCoV-2019_1	ACTACAGTCAGCTTATGTGTCAACC	25	44	60.8
nCoV-2019_27_RIGHT	nCoV-2019_1	AATACAAGCACCAAGGTCACGG	22	50	61.13
nCoV-2019_28_LEFT	nCoV-2019_2	ACATAGAAGTTACTGGCGATAGTTGT	26	38.46	60.13
nCoV-2019_28_RIGHT	nCoV-2019_2	TGTTTAGACATGACATGAACAGGTGT	26	38.46	60.91
nCoV-2019_29_LEFT	nCoV-2019_1	ACTTGTGTTCCTTTTTGTTGCTGC	24	41.67	61.39
nCoV-2019_29_RIGHT	nCoV-2019_1	AGTGTACTCTATAAGTTTTGATGGTGTGT	29	34.48	60.69
nCoV-2019_30_LEFT	nCoV-2019_2	GCACAACTAATGGTGACTTTTTGCA	25	40	61.19
nCoV-2019_30_RIGHT	nCoV-2019_2	ACCACTAGTAGATACACAAACACCAG	26	42.31	60.3
nCoV-2019_31_LEFT	nCoV-2019_1	TTCTGAGTACTGTAGGCACGGC	22	54.55	62.03
nCoV-2019_31_RIGHT	nCoV-2019_1	ACAGAATAAACACCAGGTAAGAATGAGT	28	35.71	60.69
nCoV-2019_32_LEFT	nCoV-2019_2	TGGTGAATACAGTCATGTAGTTGCC	25	44	61.09
nCoV-2019_32_RIGHT	nCoV-2019_2	AGCACATCACTACGCAACTTTAGA	24	41.67	60.56
nCoV-2019_33_LEFT	nCoV-2019_1	ACTTTTGAAGAAGCTGCGCTGT	22	45.45	61.58
nCoV-2019_33_RIGHT	nCoV-2019_1	TGGACAGTAAACTACGTCATCAAGC	25	44	61.08
nCoV-2019_34_LEFT	nCoV-2019_2	TCCCATCTGGTAAAGTTGAGGGT	23	47.83	61.02
nCoV-2019_34_RIGHT	nCoV-2019_2	AGTGAAATTGGGCCTCATAGCA	22	45.45	60.03
nCoV-2019_35_LEFT	nCoV-2019_1	TGTTCGCATTCAACCAGGACAG	22	50	61.39
nCoV-2019_35_RIGHT	nCoV-2019_1	ACTTCATAGCCACAAGGTTAAAGTCA	26	38.46	60.69
nCoV-2019_36_LEFT	nCoV-2019_2	TTAGCTTGGTTGTACGCTGCTG	22	50	61.44
nCoV-2019_36_RIGHT	nCoV-2019_2	GAACAAAGACCATTGAGTACTCTGGA	26	42.31	60.74
nCoV-2019_37_LEFT	nCoV-2019_1	ACACACCACTGGTTGTTACTCAC	23	47.83	60.93
nCoV-2019_37_RIGHT	nCoV-2019_1	GTCCACACTCTCCTAGCACCAT	22	54.55	61.48
nCoV-2019_38_LEFT	nCoV-2019_2	ACTGTGTTATGTATGCATCAGCTGT	25	40	60.86
nCoV-2019_38_RIGHT	nCoV-2019_2	CACCAAGAGTCAGTCTAAAGTAGCG	25	48	61.13
nCoV-2019_39_LEFT	nCoV-2019_1	AGTATTGCCCTATTTTCTTCATAACTGGT	29	34.48	61

nCoV-2019_39_RIGHT	nCoV-2019_1	TGTAACTGGACACATTGAGCCC	22	50	60.55
nCoV-2019_40_LEFT	nCoV-2019_2	TGCACATCAGTAGTCTTACTCTCAGT	26	42.31	61.25
nCoV-2019_40_RIGHT	nCoV-2019_2	CATGGCTGCATCACGGTCAAAT	22	50	62.09
nCoV-2019_41_LEFT	nCoV-2019_1	GTTCCCTTCCATCATATGCAGCT	23	47.83	60.75
nCoV-2019_41_RIGHT	nCoV-2019_1	TGGTATGACAACCATTAGTTTGGCT	25	40	60.75
nCoV-2019_42_LEFT	nCoV-2019_2	TGCAAGAGATGGTTGTGTTCCC	22	50	61.08
nCoV-2019_42_RIGHT	nCoV-2019_2	CCTACCTCCCTTTGTTGTGTTGT	23	47.83	60.69
nCoV-2019_43_LEFT	nCoV-2019_1	TACGACAGATGTCTTGTGCTGC	22	50	60.93
nCoV-2019_43_RIGHT	nCoV-2019_1	AGCAGCATCTACAGCAAAAGCA	22	45.45	61.14
nCoV-2019_44_LEFT	nCoV-2019_2	TGCCACAGTACGTCTACAAGCT	22	50	61.66
nCoV-2019_44_LEFT_alt3	nCoV-2019_2	CCACAGTACGTCTACAAGCTGG	22	54.55	60.67
nCoV-2019_44_RIGHT	nCoV-2019_2	AACCTTTCCACATACCGCAGAC	22	50	60.87
nCoV-2019_44_RIGHT_alt0	nCoV-2019_2	CGCAGACGGTACAGACTGTGTT	22	54.55	62.77
nCoV-2019_45_LEFT	nCoV-2019_1	TACCTACAACTTGTGCTAATGACCC	25	44	60.57
nCoV-2019_45_LEFT_alt2	nCoV-2019_1	AGTATGTACAAATACCTACAACTTGTGCT	29	34.48	60.94
nCoV-2019_45_RIGHT	nCoV-2019_1	AAATTGTTTCTTCATGTTGGTAGTTAGAGA	30	30	60.01
nCoV-2019_45_RIGHT_alt7	nCoV-2019_1	TTCATGTTGGTAGTTAGAGAAAGTGTGTC	29	37.93	61.53
nCoV-2019_46_LEFT	nCoV-2019_2	TGTCGCTTCCAAGAAAAGGACG	22	50	61.38
nCoV-2019_46_LEFT_alt1	nCoV-2019_2	CGCTTCCAAGAAAAGGACGAAGA	23	47.83	61.35
nCoV-2019_46_RIGHT	nCoV-2019_2	CACGTTCACCTAAGTTGGCGTA	22	50	60.86
nCoV-2019_46_RIGHT_alt2	nCoV-2019_2	CACGTTCACCTAAGTTGGCGTAT	23	47.83	61.17
nCoV-2019_47_LEFT	nCoV-2019_1	AGGACTGGTATGATTTTGTAGAAAACCC	28	39.29	61.42
nCoV-2019_47_RIGHT	nCoV-2019_1	AATAACGGTCAAAGAGTTTTAACCTCTC	28	35.71	60.06
nCoV-2019_48_LEFT	nCoV-2019_2	TGTTGACACTGACTTAACAAAGCCT	25	40	61.09
nCoV-2019_48_RIGHT	nCoV-2019_2	TAGATTACCAGAAGCAGCGTGC	22	50	60.74
nCoV-2019_49_LEFT	nCoV-2019_1	AGGAATTACTTGTGTATGCTGCTGA	25	40	60.57
nCoV-2019_49_RIGHT	nCoV-2019_1	TGACGATGACTTGGTTAGCATTAATACA	28	35.71	61.05
nCoV-2019_50_LEFT	nCoV-2019_2	GTTGATAAGTACTTTGATTGTTACGATGGT	30	33.33	60.59
nCoV-2019_50_RIGHT	nCoV-2019_2	TAACATGTTGTGCCAACCACCA	22	45.45	60.95
nCoV-2019_51_LEFT	nCoV-2019_1	TCAATAGCCGCCACTAGAGGAG	22	54.55	61.34
nCoV-2019_51_RIGHT	nCoV-2019_1	AGTGCATTAACATTGGCCGTGA	22	45.45	61.14
nCoV-2019_52_LEFT	nCoV-2019_2	CATCAGGAGATGCCACAACTGC	22	54.55	61.83
nCoV-2019_52_RIGHT	nCoV-2019_2	GTTGAGAGCAAAATTCATGAGGTCC	25	44	60.62
nCoV-2019_53_LEFT	nCoV-2019_1	AGCAAAATGTTGGACTGAGACTGA	24	41.67	60.69
nCoV-2019_53_RIGHT	nCoV-2019_1	AGCCTCATAAAACTCAGGTTCCC	23	47.83	60.31
nCoV-2019_54_LEFT	nCoV-2019_2	TGAGTTAACAGGACACATGTTAGACA	26	38.46	60.18
nCoV-2019_54_RIGHT	nCoV-2019_2	AACCAAAAACTTGTCCATTAGCACA	25	36	60.11
nCoV-2019_55_LEFT	nCoV-2019_1	ACTCAACTTTACTTAGGAGGTATGAGCT	28	39.29	61.43
nCoV-2019_55_RIGHT	nCoV-2019_1	GGTGTACTCTCCTATTTGTACTTTACTGT	29	37.93	60.54

nCoV-2019_56_LEFT	nCoV-2019_2	ACCTAGACCACCACTTAACCGA	22	50	60.49
nCoV-2019_56_RIGHT	nCoV-2019_2	ACACTATGCGAGCAGAAGGGTA	22	50	61.21
nCoV-2019_57_LEFT	nCoV-2019_1	ATTCTACACTCCAGGGACCACC	22	54.55	61.16
nCoV-2019_57_RIGHT	nCoV-2019_1	GTAATTGAGCAGGGTCGCCAAT	22	50	61.26
nCoV-2019_58_LEFT	nCoV-2019_2	TGATTTGAGTGTTGTCAATGCCAGA	25	40	61.44
nCoV-2019_58_RIGHT	nCoV-2019_2	CTTTTCTCCAAGCAGGGTTACGT	23	47.83	61.06
nCoV-2019_59_LEFT	nCoV-2019_1	TCACGCATGATGTTTCATCTGCA	23	43.48	61.42
nCoV-2019_59_RIGHT	nCoV-2019_1	AAGAGTCCTGTTACATTTTCAGCTTG	26	38.46	60.02
nCoV-2019_60_LEFT	nCoV-2019_2	TGATAGAGACCTTTATGACAAGTTGCA	27	37.04	60.53
nCoV-2019_60_RIGHT	nCoV-2019_2	GGTACCAACAGCTTCTCTAGTAGC	24	50	60.44
nCoV-2019_61_LEFT	nCoV-2019_1	TGTTTATCACCCGCGAAGAAGC	22	50	61.5
nCoV-2019_61_RIGHT	nCoV-2019_1	ATCACATAGACAACAGGTGCGC	22	50	61.25
nCoV-2019_62_LEFT	nCoV-2019_2	GGCACATGGCTTTGAGTTGACA	22	50	61.91
nCoV-2019_62_RIGHT	nCoV-2019_2	GTTGAACCTTTCTACAAGCCGC	22	50	60.35
nCoV-2019_63_LEFT	nCoV-2019_1	TGTTAAGCGTGTTGACTGGACT	22	45.45	60.16
nCoV-2019_63_RIGHT	nCoV-2019_1	ACAAACTGCCACCATCACAACC	22	50	61.85
nCoV-2019_64_LEFT	nCoV-2019_2	TCGATAGATATCCTGCTAATTCCATTGT	28	35.71	60.11
nCoV-2019_64_RIGHT	nCoV-2019_2	AGTCTTGTAAAAGTGTTCCAGAGGT	25	40	60.1
nCoV-2019_65_LEFT	nCoV-2019_1	GCTGGCTTTAGCTTGTGGGTTT	22	50	61.92
nCoV-2019_65_RIGHT	nCoV-2019_1	TGTCAGTCATAGAACAAACACCAATAGT	28	35.71	60.9
nCoV-2019_66_LEFT	nCoV-2019_2	GGGTGTGGACATTGCTGCTAAT	22	50	61.21
nCoV-2019_66_RIGHT	nCoV-2019_2	TCAATTTCCATTTGACTCCTGGGT	24	41.67	60.45
nCoV-2019_67_LEFT	nCoV-2019_1	GTTGTCCAACAATTACCTGAAACTTACT	28	35.71	60.43
nCoV-2019_67_RIGHT	nCoV-2019_1	CAACCTTAGAAACTACAGATAAATCTTGGG	30	36.67	60.4
nCoV-2019_68_LEFT	nCoV-2019_2	ACAGGTTCATCTAAGTGTGTGTGT	24	41.67	60.14
nCoV-2019_68_RIGHT	nCoV-2019_2	CTCCTTTATCAGAACCAGCACCA	23	47.83	60.31
nCoV-2019_69_LEFT	nCoV-2019_1	TGTCGCAAAATATACTCAACTGTGTCA	27	37.04	61.43
nCoV-2019_69_RIGHT	nCoV-2019_1	TCTTTATAGCCACGGAACCTCCA	23	47.83	61.14
nCoV-2019_70_LEFT	nCoV-2019_2	ACAAAAGAAAATGACTCTAAAGAGGGTTT	29	31.03	60.13
nCoV-2019_70_RIGHT	nCoV-2019_2	TGACCTTCTTTTAAAGACATAACAGCAG	28	35.71	60.27
nCoV-2019_71_LEFT	nCoV-2019_1	ACAAATCCAATTCAGTTGTCTTCCTATTC	29	34.48	60.54
nCoV-2019_71_RIGHT	nCoV-2019_1	TGGAAAAGAAAGGTAAGAACAAGTCCT	27	37.04	60.8
nCoV-2019_72_LEFT	nCoV-2019_2	ACACGTGGTGTTTATTACCCTGAC	24	45.83	61.04
nCoV-2019_72_RIGHT	nCoV-2019_2	ACTCTGAACTCACTTTCCATCCAAC	25	44	60.97
nCoV-2019_73_LEFT	nCoV-2019_1	CAATTTTGTAATGATCCATTTTTGGGTGT	29	31.03	60.29
nCoV-2019_73_RIGHT	nCoV-2019_1	CACCAGCTGTCCAACCTGAAGA	22	54.55	62.45
nCoV-2019_74_LEFT	nCoV-2019_2	ACATCACTAGGTTTCAAACTTTACTTGC	28	35.71	60.68
nCoV-2019_74_RIGHT	nCoV-2019_2	GCAACACAGTTGCTGATTCTCTTC	24	45.83	60.85
nCoV-2019_75_LEFT	nCoV-2019_1	AGAGTCCAACCAACAGAATCTATTGT	26	38.46	60.24

nCoV-2019_75_RIGHT	nCoV-2019_1	ACCACCAACCTTAGAATCAAGATTGT	26	38.46	60.69
nCoV-2019_76_LEFT	nCoV-2019_2	AGGGCAAACTGGAAAGATTGCT	22	45.45	60.76
nCoV-2019_76_LEFT_alt3	nCoV-2019_2	GGGCAAACTGGAAAGATTGCTGA	23	47.83	61.87
nCoV-2019_76_RIGHT	nCoV-2019_2	ACACCTGTGCCTGTTAAACCAT	22	45.45	60.42
nCoV-2019_76_RIGHT_alt0	nCoV-2019_2	ACCTGTGCCTGTTAAACCATTGA	23	43.48	60.69
nCoV-2019_77_LEFT	nCoV-2019_1	CCAGCAACTGTTTGTGGACCTA	22	50	60.75
nCoV-2019_77_RIGHT	nCoV-2019_1	CAGCCCCTATTAAACAGCCTGC	22	54.55	61.59
nCoV-2019_78_LEFT	nCoV-2019_2	CAACTTACTCCTACTTGGCGTGT	23	47.83	60.55
nCoV-2019_78_RIGHT	nCoV-2019_2	TGTGTACAAAAACTGCCATATTGCA	25	36	60.22
nCoV-2019_79_LEFT	nCoV-2019_1	GTGGTGATTCAACTGAATGCAGC	23	47.83	60.92
nCoV-2019_79_RIGHT	nCoV-2019_1	CATTTCATCTGTGAGCAAAGGTGG	24	45.83	60.62
nCoV-2019_80_LEFT	nCoV-2019_2	TTGCCTTGGTGATATTGCTGCT	22	45.45	60.89
nCoV-2019_80_RIGHT	nCoV-2019_2	TGGAGCTAAGTTGTTTAACAAGCG	24	41.67	60.02
nCoV-2019_81_LEFT	nCoV-2019_1	GCACTTGGAAAACTTCAAGATGTGG	25	44	61.24
nCoV-2019_81_RIGHT	nCoV-2019_1	GTGAAGTTCTTTTCTTGTGCAGGG	24	45.83	60.73
nCoV-2019_82_LEFT	nCoV-2019_2	GGGCTATCATCTTATGTCCTTCCCT	25	48	61.52
nCoV-2019_82_RIGHT	nCoV-2019_2	TGCCAGAGATGTCACCTAAATCAA	24	41.67	60.02
nCoV-2019_83_LEFT	nCoV-2019_1	TCCTTTGCAACCTGAATTAGACTCA	25	40	60.46
nCoV-2019_83_RIGHT	nCoV-2019_1	TTTGACTCCTTTGAGCACTGGC	22	50	61.33
nCoV-2019_84_LEFT	nCoV-2019_2	TGCTGTAGTTGTCTCAAGGGCT	22	50	61.61
nCoV-2019_84_RIGHT	nCoV-2019_2	AGGTGTGAGTAAACTGTTACAAACAAC	27	37.04	60.36
nCoV-2019_85_LEFT	nCoV-2019_1	ACTAGCACTCTCCAAGGGTGTT	22	50	61.03
nCoV-2019_85_RIGHT	nCoV-2019_1	ACACAGTCTTTTACTCCAGATTCCC	25	44	60.51
nCoV-2019_86_LEFT	nCoV-2019_2	TCAGGTGATGGCACAACAAGTC	22	50	61.07
nCoV-2019_86_RIGHT	nCoV-2019_2	ACGAAAGCAAGAAAAAGAAGTACGC	25	40	61.01
nCoV-2019_87_LEFT	nCoV-2019_1	CGACTACTAGCGTGCCTTTGTA	22	50	60.16
nCoV-2019_87_RIGHT	nCoV-2019_1	ACTAGGTTCCATTGTTCAAGGAGC	24	45.83	60.81
nCoV-2019_88_LEFT	nCoV-2019_2	CCATGGCAGATTCCAACGGTAC	22	54.55	61.58
nCoV-2019_88_RIGHT	nCoV-2019_2	TGGTCAGAATAGTGCCATGGAGT	23	47.83	61.4
nCoV-2019_89_LEFT	nCoV-2019_1	GTACGCGTTCCATGTGGTCATT	22	50	61.5
nCoV-2019_89_LEFT_alt2	nCoV-2019_1	CGCGTTCCATGTGGTCATTCAA	22	50	62.01
nCoV-2019_89_RIGHT	nCoV-2019_1	ACCTGAAAGTCAACGAGATGAAACA	25	40	60.91
nCoV-2019_89_RIGHT_alt4	nCoV-2019_1	ACGAGATGAAACATCTGTTGTCACT	25	40	60.74
nCoV-2019_90_LEFT	nCoV-2019_2	ACACAGACCATTCCAGTAGCAGT	23	47.83	61.58
nCoV-2019_90_RIGHT	nCoV-2019_2	TGAAATGGTGAATTGCCCTCGT	22	45.45	60.82
nCoV-2019_91_LEFT	nCoV-2019_1	TCACTACCAAGAGTGTGTTAGAGGT	25	44	60.93
nCoV-2019_91_RIGHT	nCoV-2019_1	TTCAAGTGAGAACCAAAAGATAATAAGCA	29	31.03	60.03
nCoV-2019_92_LEFT	nCoV-2019_2	TTTGTGCTTTTTAGCCTTTCTGCT	24	37.5	60.14
nCoV-2019_92_RIGHT	nCoV-2019_2	AGGTTCCTGGCAATTAATTGTAAAAGG	27	37.04	60.53

nCoV-2019_93_LEFT	nCoV-2019_1	TGAGGCTGGTTCTAAATCACCCA	23	47.83	61.59
nCoV-2019_93_RIGHT	nCoV-2019_1	AGGTCTTCCTTGCCATGTTGAG	22	50	60.55
nCoV-2019_94_LEFT	nCoV-2019_2	GGCCCCAAGGTTTACCCAATAA	22	50	60.56
nCoV-2019_94_RIGHT	nCoV-2019_2	TTTGGCAATGTTGTTCCTTGAGG	23	43.48	60.18
nCoV-2019_95_LEFT	nCoV-2019_1	TGAGGGAGCCTTGAATACACCA	22	50	61.1
nCoV-2019_95_RIGHT	nCoV-2019_1	CAGTACGTTTTTGCCGAGGCTT	22	50	61.95
nCoV-2019_96_LEFT	nCoV-2019_2	GCCAACAACAACAAGGCCAAAC	22	50	61.82
nCoV-2019_96_RIGHT	nCoV-2019_2	TAGGCTCTGTTGGTGGGAATGT	22	50	61.36
nCoV-2019_97_LEFT	nCoV-2019_1	TGGATGACAAAGATCCAAATTTCAAAGA	28	32.14	60.22
nCoV-2019_97_RIGHT	nCoV-2019_1	ACACACTGATTAAAGATTGCTATGTGAG	28	35.71	60.17
nCoV-2019_98_LEFT	nCoV-2019_2	AACAATTGCAACAATCCATGAGCA	24	37.5	60.5
nCoV-2019_98_RIGHT	nCoV-2019_2	TTCTCCTAAGAAGCTATTAAAATCACATGG	30	33.33	60.01

S3 Table. ARTIC v3 primer sequences.

Chapter 4:

Limited within-host diversity and tight transmission bottlenecks limit SARS-CoV-2 evolution in acutely infected individuals

Katarina Braun[1]*, Gage Moreno[3]*, Cassia Wagner[2], Molly A. Accola[5], William M. Rehrauer[5], David Baker[3,4], Katia Koelle[6], David H. O'Connor[3,4], Trevor Bedford[2], Thomas C. Friedrich[1#], Louise H. Moncla[2#]

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

[2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America

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

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

[5]University of Wisconsin School of Medicine and Public Health, Madison, WI, United States of America and the William S. Middleton Memorial Veterans Hospital

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

*These authors contributed equally

[#]Co-corresponding

Submitted to Nature Communications 2021-05-06.

Abstract

The emergence of divergent SARS-CoV-2 lineages has raised concern that novel variants eliciting immune escape or enhanced transmissibility could emerge within individual hosts. Though growing evidence suggests that novel variants arise during prolonged infections, most infections are acute. Understanding how efficiently variants emerge and transmit among acutely-infected hosts is therefore critical for predicting the pace of long-term SARS-CoV-2 evolution. To characterize how within-host diversity is generated and propagated, we combine extensive laboratory and bioinformatic controls with metrics of within- and between-host diversity to 133 SARS-CoV-2 genomes from acutely-infected individuals. We find that within-host diversity is low and transmission bottlenecks are narrow, with very few viruses founding most infections. Within-host variants are rarely transmitted, even among individuals within the same household, and are rarely detected along phylogenetically linked infections in the broader community. These findings suggest that efficient selection and transmission of novel SARS-CoV-2 variants is unlikely during typical, acute infection.

Introduction

The recent emergence of variants of concern has spurred uncertainty about how severe acute respiratory coronavirus 2 (SARS-CoV-2) will evolve in the longer term. SARS-CoV-2 acquires a fixed consensus mutation approximately every 11 days as it replicates in a

population [153]. Recently, however, lineages of SARS-CoV-2 have arisen harboring more variants than expected based on this clock rate, with some variants conferring enhanced transmissibility and/or antibody escape [263,264]. The emergence of these lineages has raised concern that SARS-CoV-2 may rapidly evolve to evade vaccine-induced immunity, and that vaccines may need to be frequently updated. A current leading hypothesis posits that these lineages may have emerged during prolonged infections. Under this hypothesis, longer infection times, coupled with antibody selection [265], may allow more time for novel mutations to be generated and selected before transmission. Studies of SARS-CoV-2 [265–269] and other viruses [270,271] support this hypothesis. Longitudinal sequencing of SARS-CoV-2 from immunocompromised or persistently infected individuals accordingly reveals an accumulation of single-nucleotide variants (iSNVs) and short insertions and deletions (indels) during infection [265–267,272]. In influenza virus and norovirus infections, variants that arose in immunocompromised patients were later detected globally, suggesting that long-term infections may mirror global evolutionary dynamics [270,273]. Mutations defining novel variant lineages resulting in enhanced transmissibility and/or immune escape in SARS-CoV-2 Spike, like Δ69/70, N501Y and E484K, have already been documented arising in persistently infected and immunocompromised individuals [265,266].

While prolonged infections occur, the vast majority of SARS-CoV-2 infections are acute [274]. Viral evolutionary capacity is limited by the duration of infection [275], and it is not yet clear whether the evolutionary patterns observed during prolonged SARS-CoV-2 infections also occur in acutely infected individuals. Replication-competent virus has

rarely been recovered from individuals with mild to moderate coronavirus disease 2019 (COVID-19) beyond ~10 days following symptom onset [47,276]. Multiple studies of influenza viruses show that immune escape variants are rarely detected during acute infection, even within vaccinated individuals [120–122]. Detailed modeling of influenza dynamics suggests that the likelihood of within-host mutation emergence depends on the interplay of immune response timing, the de-novo mutation rate, and the number of virus particles transmitted between hosts [275]. Understanding the speed with which SARS-CoV-2 viruses acquire novel mutations that may escape population immunity will be critical for formulating future vaccine updates. If novel immune-escape variants emerge primarily within long-term infections, then managing long-term infections in an effort to reduce any onward transmission may be critically important. Conversely, if novel variants are efficiently selected and transmitted during acute infections, then vaccine updates may need to occur frequently.

While understanding the process of variant generation and transmission is critically important, a clear consensus on how frequently variants are shared and transmitted between individuals has been elusive. Estimates of SARS-CoV-2 diversity within hosts have been highly variable, and comparing results among labs has been complicated by sensitivity to variant-calling thresholds and inconsistent laboratory controls [56,136,189,228]. Some data suggest that SARS-CoV-2 genetic diversity within individual hosts during acute infections is limited [56,242] and shaped by genetic drift and purifying selection [57,228,229,277]. Estimates of the size of SARS-CoV-2 transmission bottlenecks [54,228,278] have ranged considerably, and recent validation work has shown that estimates of within-host

diversity and transmission bottleneck sizes are highly sensitive to sequencing protocols and data analysis parameters, like the frequency cutoff used to define/identify within-host variants [56,279]. Clarifying the extent to which within-host variants arise and transmit among acutely infected individuals, while controlling for potential error, will be critical for assessing the speed at which SARS-CoV-2 evolves and adapts.

To characterize how within-host variants are generated and propagated, we employ extensive laboratory and bioinformatic controls to characterize 133 SARS-CoV-2 samples collected from acutely-infected individuals in Wisconsin, United States. By comparing patterns of intrahost single nucleotide variants (iSNVs) to densely-sampled consensus genomes from the same geographic area, we paint a clear picture of how variants emerge and transmit within communities and households. We find that overall within-host diversity is low during acute infection, and that iSNVs detected within hosts almost never become dominant in later-sampled sequences. We find that iSNVs are infrequently transmitted, even between members of the same household, and we estimate that transmission bottlenecks between putative household pairs are narrow. This suggests that most iSNVs are transient and very rarely transmit beyond the individual in which they have originated. Our results imply that during typical, acute SARS-CoV-2 infections, the combination of limited intrahost genetic diversity and narrow transmission bottlenecks may slow the pace by which novel variants arise, are selected, and transmit onward. Finally, most individual infections likely play a minor role in SARS-CoV-2 evolution, consistent with the hypothesis that novel variants are more likely to arise in rare instances of prolonged infection.

Materials and Methods

Sample approvals and sample selection criteria

Samples selected for iSNV characterization were derived from 150 nasopharyngeal (NP) swab samples collected from March 2020 through July 2020, originating from the University of Wisconsin Hospital and Clinics and the Milwaukee Health Department Laboratories. Submitting institutions provided a cycle threshold (Ct) or relative light unit (RLU) for all samples. Sample metadata, including GISAID and SRA accession identifiers, are available in Supplemental Table 2.

We obtained a waiver of HIPAA Authorization and were approved to obtain the clinical samples along with a Limited Data Set by the Western Institutional Review Board (WIRB #1-1290953-1) and the FUE IRB 2016-0605. This limited dataset contains sample collection data and county of collection. Additional sample metadata, e.g. race/ethnicity, were not shared.

Diagnostic assays for the samples included in this study were performed at the University of Wisconsin Hospital and Clinical diagnostic laboratory using CDC's diagnostic RT-PCR [280], the Hologic Panther SARS-CoV-2 assay [281], or the Aptima SARS-CoV-2 assay [282].

Nucleic acid extraction

Viral RNA (vRNA) was extracted from 100 µl of VTM using the Viral Total Nucleic Acid Purification kit (Promega, Madison, WI, USA) on a Maxwell RSC 48 instrument and eluted in 50 µL of nuclease-free H2O.

Complementary DNA (cDNA) generation and PCR

Complementary DNA (cDNA) was synthesized according to a modified ARTIC Network approach [235,283]. RNA was reverse transcribed with SuperScript IV VILO (Invitrogen, Carlsbad, CA, USA) according to manufacturer guidelines [235,283]. A SARS-CoV-2-specific multiplex PCR for Nanopore sequencing was performed using the ARTIC v3 primers (Supplemental 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).

TruSeq Illumina library prep and sequencing for minor variants

All Wisconsin surveillance samples were prepped and sequenced by Oxford Nanopore Technologies (details below) and a subset described in this paper were additionally prepped for sequencing on an Illumina MiSeq. These SARS-CoV-2 samples (n=150) consisted of household pairs as well as a random sampling of the surveillance cohort selective for enhanced iSNV characterization. Amplified cDNA was purified and made compatible for sequencing on an Illumina MiSeq according to the TruSeq Nano DNA manufacturer instructions (Illumina, USA). The average DNA fragment length and purity was determined using the Agilent High Sensitivity DNA kit and the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). Samples were pooled at equimolar concentrations to a final concentration of 4 nM. All libraries were run on a 500-cycle v2 flow cell. The samples included in this study were sequenced across seven distinct MiSeq

runs. Each sample was library prepped and sequenced in technical replicate. Replicates were true replicates in that we started from two aliquots taken from the original samples.

Oxford nanopore library preparation and sequencing for consensus sequences

All consensus-level surveillance sequencing of SARS-CoV-2 was performed using Oxford Nanopore sequencing (n=3,351) as described previously [55].

Processing raw ONT data

Sequencing data was processed using the ARTIC bioinformatics pipeline scaled up using on campus computing cores (https://github.com/artic-network/artic-ncov2019). The entire ONT analysis pipeline is available at https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin.

Processing raw Illumina data

Raw FASTQ files were analyzed using the workflow available in the following GitHub repository – https://github.com/gagekmoreno/SARS_CoV-2_Zequencer. Reads were paired and merged using BBMerge (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmerge-guide/) and mapped to the Wuhan-Hu-1/2019 reference (Genbank accession MN908947.3) 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/). Variants were called at ≥0.01% in high-quality reads

(phred score >30) that were ≥100 base pairs in length and supported by a minimum of 10 reads. The total minimum read support was set to 10 to generate initial VCF files with complete consensus genomes for the few samples where coverage fell below 100 reads in a few areas. Substantial downstream variant cleaning was performed as outlined below.

iSNV quality control

BBMap's output VCF files were cleaned using custom Python scripts, which can be found in the GitHub accompanying this manuscript (https://github.com/lmoncla/ncov-WI-within-host). First, any samples without technical replicates were excluded. Next, we discarded all iSNVs which occurred at primer-binding sites (**Supplemental Table 3**). These "recoded" VCFs can be found in the GitHub repository in "data/vcfs-recode". We then filtered these recoded VCF files and for variants with (1) 100x coverage; (2) found at ≥3% frequency; (3) and found between nucleotides 54 and 29,837 (based on the first and last ARTIC v3 amplicon). We excluded all indels from our analysis, including those that occur in intergenic regions.

We inspected our filtered iSNV datasets across replicate pairs. We visually inspected each replicate pair VCF and plotted replicate frequencies against each other (available in the GitHub repository). This identified a few samples which were outliers for having very limited overlap in their iSNV populations. This could be traced to low coverage or amplicon drop-out in each sample. FASTQs for these samples are available in GenBank, but we have excluded them from downstream analyses presented here (n=11; tube/filename

identifier 65, 124, 125, 303, 316, 1061, 1388, 1103, 1104, 1147, and 1282) (iSNVs in technical replicates are shown for sample 1104 in **Supplemental Figure 4b**).

We generated one cleaned VCF file by averaging the frequencies found for overlapping iSNVs and discarding all iSNVs which were only found in one replicate. In addition to the SARS-CoV-2 diagnostic swabs, we sequenced a SARS-CoV-2 synthetic RNA control (Twist Bioscience, San Francisco, CA) representing the Wuhan-Hu-1 sequence (Genbank: MN908947.3) in technical replicate at 1x106 template copies per reaction in order to identify spurious variants introducing during library prep and sequencing. We then excluded variants detected in the synthetic RNA control (**Supplemental Table 4**) from all downstream analyses. Notably, this filter removed a single variant at nucleotide position 6,669 from our analysis [56]. Finally, within-host variants called at ≥50% and <97% frequency comprise consensus-level mutations relative to the Wuhan-Hu-1/2019 reference sequence. To ensure that the corresponding minor variant was reported we report the opposite minor allele at a frequency of 1 - the consensus variant frequency. For example, a C to T variant detected at 75% frequency relative to the Wuhan-1 reference was converted to a T to C variant at 25% frequency.

Processing of the raw sequence data, mapping, and variant calling with the Washington pipeline

To assess the sensitivity of our iSNV calls to bioinformatic pipelines, we generated VCF files using an independent bioinformatic pipeline. Raw reads were assembled against the SARS-CoV-2 reference genome Wuhan-Hu-1/2019 (Genbank accession MN908947.3;

the same reference used for the alternative basecalling method) to generate pileup files using the bioinformatics pipeline available at https://github.com/seattleflu/assembly. Briefly, reads were trimmed with Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic) [192] in paired end mode, in sliding window of 5 base pairs, discarding all reads that were trimmed to <50 base pairs. Trimmed reads were mapped using Bowtie 2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) [193], and pileups were generated using samtools mpileup (http://www.htslib.org/doc/samtools-mpileup.html). Variants were then called from pileups using varscan mpileup2cns v2.4.4 (http://varscan.sourceforge.net/using-varscan.html#v2.3_mpileup2cns). Variants were called at ≥1% frequency, with a minimum coverage of 100, and were supported by a minimum of 2 reads.

Phylogenetic analysis

All available full-length sequences from Wisconsin through February 16, 2021 were used for phylogenetic analysis using the tools implemented in Nextstrain custom builds (https://github.com/nextstrain/ncov) [145,284]. Phylogenetic trees were built using the standard Nextstrain tools and scripts [145,284]. We used custom python scripts to filter and clean metadata. A custom “Wisconsin” profile was made to create a Wisconsin-centric subsampled build to include representative sequences. The scripts and output are available at https://github.com/gagekmoreno/Wisconsin-SARS-CoV-2.

Household pairs permutation test

For household groups, we performed all pairwise comparisons between members of the household, excluding pairs for which the consensus genomes differed by >2 nucleotide changes. We determined this cutoff by modeling the probability that 2 consensus genomes separated by one serial interval differ by n mutations. We model this process as Poisson-distributed with lambda equal to the expected number of substitutions per serial interval, as described previously 34. We chose to model this expectation using the serial interval rather than the generation interval for the following reason.

The sequence data we have represent cases that were sampled via passive surveillance, usually from individuals seeking testing after developing symptoms. Differences in the genome sequences from two individuals therefore represent the evolution that occurred between the sampling times of those two cases. Although neither the serial interval nor the generation interval perfectly matches this sampling process, we reasoned that the serial interval, or the time between the symptom onsets of successive cases, may more accurately capture how the data were sampled. We evaluated probabilities across a range of serial interval and clock rates. For serial interval, we use the values inferred by He et al, of a mean of 5.8 days with a 95% confidence interval of 4.8-6.8 days [285]. For substitution rate, we employ estimates from Duchene et al, who estimate a mean substitution rate of 1.10 x 10-3 substitutions per site per year, with a 95% credible interval of 7.03 x 10-4 and 1.15 x 10-3 1. To model the expectation across this range of values, we evaluate the probabilities for serial intervals at the mean (5.8), as well as for 4, 5, 6, 7, and 8 days, and substitution rates at the mean (1.10 x 10-3) and at the bounds of the 95% credible interval. For each combination of serial interval and substitution rate, we

calculate the expected substitutions in one serial interval as: (substitution rate per site per year * genome length/365 days) *serial interval. The results using the mean serial interval (5.8 days) and substitution rate (1.10 x 10-3) are shown in the main text, while the full set of combinations is shown in the supplement. Under this model, the vast majority of consensus genomes derived from cases separated by a single serial interval are expected to differ by ≤2 mutations. The probability that two genomes that are separated by one serial interval differ by 3 mutations ranges from 0.0016-0.059. Only in the case of an 8-day serial interval with the highest bound of the substitution rate do we infer a probability of 3 mutations that is greater than 0.05. We therefore classified all pairs of individuals from each household that differed by ≤2 consensus mutations and who were tested within 14 days of each other as putative transmission pairs.

To determine whether putative household transmission pairs shared more variants than individuals without an epidemiologic link, we performed a permutation test. At each iteration, we randomly selected a pair of samples (with replacement) and computed the proportion of variants they share as: (2 x total number of shared variants) / (the total number of variants detected among the two samples). For example, if sample A contained 5 iSNVs relative to the reference (Wuhan-1, Genbank accession MN908947.3), sample B harbored 4 iSNVs, and 1 iSNV was shared, then the proportion of sample A and B's variants that are shared would be 2/9 = 0.22. We performed 10,000 iterations in which pairs were sampled randomly to generate a null distribution. We then compared the proportion of variants shared by each putative household transmission pair to this null

distribution. The proportion of variants shared by a household pair was determined to be statistically significant if it was greater than 95% of random pairs.

Transmission bottleneck calculation

The beta-binomial method [137], was used to infer the transmission bottleneck size Nb. Nb quantifies the number of virions donated from the index individual to the contact (recipient) individual that successfully establish lineages in the recipient that are present at the sampling time point. The beta-binomial method assumes variant sites are independent, which may not be true given that SARS-CoV-2 contains a continuous genome thought to undergo limited recombination [286]. In addition, the beta-binomial method assumes that identical variants found in the index and contact are shared as a result of transmission, though it is possible that identical variants occurring in a donor and a recipient individual occurred independently of one another and are not linked through transmission. We consider this possibility at one site in particular which commonly appears at low frequencies in donor-recipient pairs. Code for estimating transmission bottleneck sizes using the beta-binomial approach has been adapted from the original scripts (https://github.com/koellelab/betabinomial_bottleneck) and is included in the GitHub accompanying this manuscript (https://github.com/lmoncla/ncov-WI-within-host).

We calculated individual transmission bottleneck size estimates for each household transmission pair as were identified in the household permutation test (n=28). We used the date of symptom onset and/or date of sample collection to assign donor and recipient within each pair. Within each pair, if the date of symptom onset differed by ≥3 days, we

assigned the individual with the earlier date as the donor. If this information was unavailable or uninformative (<3 days) for both individuals in a pair, we looked at the date of sample collection and if these dates differed by ≥3 days, we assigned the individual with the earlier date as the donor. If this information was also not available or was not informative (<3 days), we calculated the bottleneck size with each individual as a donor. These bidirectional comparisons are denoted with an “a” or “b” appended to the filename (n=16 pairs were analyzed bidirectionally). In total, we analyzed 44 pairs (including bidirectional comparisons). Metadata and GISAID accession numbers for each pair are described in **Supplemental Table 4**.

Combined transmission bottleneck size estimates (as seen in Figure 6c) were estimated as described in the supplemental methods in Martin & Koelle [279]. Briefly, overall transmission bottleneck sizes were estimated based on the assumption that transmission bottleneck sizes are distributed according to a zero-truncated Poisson-distribution and bidirectional bottleneck estimates were each assigned 50% of the weight in this calculation compared to the unidirectional pairs. Matlab code to replicate the combined bottleneck estimates can be found in the GitHub accompanying this paper (https://github.com/lmoncla/ncov-WI-within-host).

Enumerating mutations along the phylogeny

We used the global Nextstrain [145] phylogenetic tree (nextstrain.org/ncov/global) accessed on February 24, 2021 to query whether mutations detected within-host are detected on the global tree. We accessed the tree in JSON format and traverse the tree using baltic

[287]. To determine the fraction of within-host variants detected on the phylogenetic tree, we traversed the tree from root to tip, gathering each mutation that arose on the tree in the process. For each mutation, we counted the number of times it arose on internal and terminal nodes. We then compared the fraction of times each iSNV identified within-host was detected on an internal node vs. a terminal node. To determine whether particular iSNVs were enriched at internal nodes, we compared the frequency of that iSNV's detection against the overall ratio of mutations arising on internal vs. terminal nodes in the phylogeny with a Fisher's exact test.

To query whether iSNVs ever became dominant in tips sampled downstream, we used a transmission metric developed previously [288]. Using the tree JSON output from the Nextstrain pipeline [145], we traversed the tree from root to tip. We collapsed very small branches (those with branch lengths less than 1 x 10-16) to obtain polytomies. For each tip for which we had within-host data that lay on an internal node, i.e., had a branch length of nearly 0 (< 1 x 10-16), we then determined whether any subsequent tips occurred in the downstream portion of the tree, i.e., tips that fall along the same lineage but to the right of the parent tip. We then traversed the tree and enumerated every mutation that arose from the parent tip to each downstream tip. If any mutations along the path from the parent to downstream tip matched a mutation found within-host in the parent, this was classified as a potential instance of variant transmission. A diagram of how "downstream tips" and mutations were classified is shown in **Figure 4a.**

Linear regression model

To determine the relative contributions of phylogenetic divergence, geographic distance, clade membership, and household membership to the probability of sharing within-host variants, we fit linear regression models to the data in R. As our outcome variable, we performed pairwise comparisons for each pair of samples in the dataset (including household and non-household pairs) and compute the proportion of variants shared for each pair. We then model the proportion of shared variants as the combined function of 4 predictor variables as follows: Proportion of variants shared ~ $\beta 0 + \beta 1x1 + \beta 2x2 + \beta 3x3 + \beta 4x4$, where x1 represents a 0 or 1 value for household, where a 1 indicates the same household and a 0 indicates no household relationship. X2 denotes the divergence, i.e., the branch length in mutations between tip A and tip B as a continuous variable, x3 indicates the great circle distance in kilometers between the location of sample collection as a continuous variable, and x4 denotes a 0 or 1 for whether the two tips belong to the same clade (same clade coded as a 1, different clade coded as a 0). We fit a univariate model for each variable independently, a model with an intercept alone, and a combined model using the Rethinking package in R (https://www.rdocumentation.org/packages/rethinking/versions/1.59). We perform model comparison with the WAIC metric and select the combined model as the one with the best fit. We compute mean coefficient estimates and 95% highest posterior density intervals (HPDI) by sampling and summarizing 10,000 values from the posterior distribution.

Data and code availability

Consensus genomes have been deposited in GISAID with accession numbers available in Supplemental Table 1. Raw Illumina reads are available in the Short Read Archive

under bioproject PRJNA718341. All raw Nanopore reads are available in the Short Read Archive under bioproject PRJNA614504. All code used to analyze the data and generate the figures shown in this manuscript are available at https://github.com/lmoncla/ncov-WI-within-host.

Results

Within-host variation is limited and sensitive to iSNV-calling parameters

Viral sequence data provide rich information about how variants emerge within, and transmit beyond, individual hosts. Viral nucleotide variation generated during infection provides the raw material upon which selection can act. However, viral sequence data are sensitive to multiple sources of error [56,136,189], which has obscured easy comparison among existing studies of SARS-CoV-2 within-host evolution. Here, we take several steps to minimize sources of error and to assess the robustness of our results against variable within-host single nucleotide variant (iSNV)-calling parameters.

First, we identified spurious iSNVs introduced by our library preparation pipeline by sequencing in duplicate a clonal, synthetic RNA transcript identical to our reference genome (MN90847.3). We considered only variants found in both technical replicates, which we refer to as “intersection iSNVs”. We detected 7 intersection iSNVs at ≥1% frequency (**Supplemental Table 1**); 2 of these were previously identified by a similar experiment in Valesano et al. [56]. We excluded all 7 of these iSNVs from downstream analyses. To exclude laboratory contamination, we sequenced a no-template control (water) with each large sequencing batch and confirmed that these negative controls

contained <10x coverage across the SARS-CoV-2 genome (**Supplemental Figure 1**, **Supplemental Figure 2**). To ensure that spurious variants were not introduced by our bioinformatic pipelines, we validated our iSNV calls using a second pipeline which employs distinct trimming, mapping, and variant calling softwares. We found near-equivalence between the two pipelines' iSNV calls (R2=0.998; **Supplemental Figure 3a**), providing additional independent support for our bioinformatic pipeline to accurately call iSNVs.

Viral iSNV calls are also sensitive to the variant-calling threshold (i.e., a minimum frequency at which iSNVs must occur to be considered non-artefactual) applied [189] and the number of viral input copies. Work by Grubaugh et al. [190] showed highly accurate iSNV calls with tiled amplicon sequencing using technical replicates and a 3% frequency threshold. Consistent with this observation, we observed a near-linear correlation between iSNVs called in each replicate at a 3% frequency threshold (R2=0.992) (**Figure 1a**). Unsurprisingly, we find the proportion of intersection iSNVs compared to all iSNVs within a given sample increases as the frequency threshold increases (**Supplemental Figure 3b**). Additionally, the majority of iSNVs detected in our clonal RNA controls occur <3% frequency (**Supplemental Figure 3c**).

Consistent with previous studies, we observed a negative correlation between Ct and the overlap in variants between replicates such that high-Ct (i.e., low vRNA copy number) samples had fewer intersection iSNVs called in each replicate (**Figure 1b**) [189,190]. Although we do not have access to absolute quantification for viral input copies for our

sampleset, we can use results of semi-quantitative clinical assays on the sequenced specimens as a proxy for viral RNA (vRNA) concentration. Using input data from two different clinical assay platforms, we find no correlation between viral input copies and the number of intersection iSNVs detected (**Supplemental Figure 3d** and **Supplemental Figure 3e**).

Based on these observations, we chose to use a 3% iSNV frequency cutoff for all downstream analyses, and report only iSNVs that were detected in both technical replicates, at a frequency ≥3%. Using these criteria, we found limited SARS-CoV-2 genetic diversity in most infected individuals: 22 out of 133 samples did not harbor even a single intersection iSNV at ≥3% frequency. Among the 111 samples that did harbor within-host variation, the average number of iSNVs per sample was 3.5 (median=3, range=1-11) (**Figure 1c).** Most iSNVs were detected at <10% frequency (**Figure 1d**). Compared to expectations under a neutral model, every type of mutation we evaluated (synonymous, nonsynonymous, intergenic region, and stop) was present in excess at low frequencies, consistent with purifying selection or population expansion within the host (**Figure 1d**). Taken together, our results confirm that the number of iSNVs detected within-host are dependent on variant-calling criteria. Once rigorous laboratory and bioinformatic controls are applied, we find that most infections are characterized by very few iSNVs, and primarily low-frequency variants.

Recurrent iSNVs consist of Wuhan-1 reversions and common polymorphic sites

Previous studies of SARS-CoV-2 evolution have noted the unusual observation that iSNVs are sometimes shared across multiple samples. Understanding the source and frequency of shared iSNVs is important for measuring the size of transmission bottlenecks and for identifying potential sites of selection. In our dataset, most iSNVs were unique to a single sample (**Figure 2a**). However, 41 iSNVs were detected in at least 2 samples. These “shared iSNVs” were detected across multiple sequencing runs (**Supplemental Figure 5**), and were absent in our negative controls, suggesting they are unlikely to be artefacts of method error. Most of the shared iSNVs we detect fall into two categories: iSNVs that occur within or adjacent to a homopolymer region (8/41 iSNVs, **Figure 2b**, yellow and purple bars), or iSNVs that represent “Wuhan-1 reversions” (31/41 iSNVs, **Figure 2b**, blue and purple bars). iSNVs in or near homopolymer regions were defined as those that fall within or one nucleotide outside of a span of at least 3 identical nucleotide bases. Shared iSNVs were more commonly detected in A/T homopolymer regions than in G/C homopolymer regions. We classified iSNVs as “Wuhan-1 reversions” when a sample’s consensus sequence had a near-fixed variant (50-97% frequency) relative to the Wuhan-1 reference, with the original Wuhan-1 nucleotide present as an iSNV. Overall, this suggests that shared variants in our dataset may be at least partially explained by viral polymerase incorporation errors, potentially in A/T-rich regions, and at sites that are frequently polymorphic.

The most commonly detected iSNVs in our dataset represent Wuhan-1 reversion at nucleotide sites 241 (detected 18 times; within/adjacent to a homopolymer region) and 3037 (detected 21 times; not in a homopolymer region). Both of these sites are

polymorphic deep in the SARS-CoV-2 phylogeny near the branch point for clade 20A (Nextstrain clade nomenclature). Within-host polymorphisms at sites 241 and 3037 were also detected in recent studies in the United Kingdom and Austria [54,228]. T241C and T3037C are both synonymous variants, and have emerged frequently on the global SARS-CoV-2 phylogenetic tree, suggesting that these sites may be frequently polymorphic within and between hosts across multiple geographic areas (**Figure 2c**).

Most within-host variation does not contribute to consensus diversity

The emergence of divergent SARS-CoV-2 lineages has raised concerns that new variants may be selected during infection and efficiently transmitted onward. We next sought to characterize whether iSNVs arising within hosts contribute to consensus diversity sampled later in time. Using the Wisconsin-specific phylogenetic tree (**Supplemental Figure 6**), we queried whether iSNVs detected within hosts are ever found at consensus in tips sampled downstream. For each Wisconsin tip that lay on an internal node and for which we had within-host data, we traversed the tree from that tip to each subtending tip. We then enumerated each mutation that occurred along that path, and compared whether any mutations that arose on downstream branches matched iSNVs detected within-host (see **Figure 3a** for a schematic). Of the 110 Wisconsin tips harboring within-host variation, 93 occurred on internal nodes. Of those, we detect only a single instance in which an iSNV detected within a host was later detected at consensus. C1912T (a synonymous variant) was present in USA/WI-UW-214/2020 at ~4% frequency, and arose on the branch leading to USA/WI-WSLH-200068/2020 (**Figure 3b**). USA/WI-UW-214/2020 is part of a large polytomy, so this does not necessarily suggest that USA/WI-UW-214/2020

and USA/WI-WSLH-200068/2020 fall along the same transmission chain. These results indicate that despite relatively densely sampling consensus genomes from related viruses from Wisconsin, we do not find evidence that iSNVs frequently rise to consensus along phylogenetically linked infections.

If iSNVs arising during infection are adaptive and efficiently transmitted, then they should be found frequently in consensus genomes, and may be enriched on internal nodes of the phylogenetic tree. For each within-host variant detected in our dataset, we queried the number of times it occurred on the global SARS-CoV-2 phylogeny on tips and internal nodes. We then compared the ratio of detections on tips vs. internal nodes to the overall ratio of mutations on tips vs. internal nodes on the phylogeny. 42% (77/185) of iSNVs are present at least once at consensus level on the global phylogeny (**Supplemental Figure 7**). When present, iSNVs from our dataset that also occur in consensus genomes on the global tree tend to be rare, and predominantly occur on terminal nodes (**Figure 3c**, **Supplemental Figure 7**). Overall, iSNVs that are also found at consensus are present on internal nodes and tips at a ratio similar to that of consensus mutations overall (ratio of mutations on phylogeny nodes:tips = 4,637:17,200; ratio of iSNVs on nodes:tips = 128:411, p=0.16, Fisher's exact test). Although this is the predominant pattern, we detect one exception. C28887T is present in one sample in our dataset at a frequency of ~6%, but is found on 10 internal nodes and 15 tips (p = 0.028, Fisher's exact test) (Figure 3c). C28887T encodes a threonine-to-isoleucine change at position 205 in the N protein, and is a clade-defining mutation for the B.1.351 lineage. Although the functional impact of this mutation is not completely understood, N T205I may increase stability of the N protein

32,33. Despite the detection within-host and subsequent emergence of N205I globally, this iSNV was only detected in our dataset in one sample at low frequency. In general, across our dataset, the frequency with which iSNVs were detected within-host vs. on the phylogenetic tree is not correlated (**Figure 2c**). This suggests that although putative functional mutations may arise within a host, these events are rare. iSNV detection within a host, at least in typical acute infections, may therefore have limited utility for predicting future variant emergence. Together, these data suggest that with rare exception, most within-host variants are purged over time, and typically do not contribute to consensus-level diversity sampled later in time. As such, these findings suggest that most iSNVs are not selectively beneficial and are not efficiently transmitted.

Variation is shared among some household samples, but is likely insufficient for transmission resolution

Household studies provide the opportunity to investigate transmission dynamics in a setting of known epidemiologic linkage. We analyzed 44 samples collected from 19 households from which multiple individuals were infected with SARS-CoV-2. To define putative transmission pairs from our household dataset, we modeled the expected number of mutations that should differ between consensus genomes given one serial interval as previously described[289] (see Methods for details and rationale). We estimate that members of a transmission pair should generally differ by 0 to 2 consensus mutations (**Figure 4a**), and classify all such pairs within a household as putative transmission pairs. While most samples derived from a single household had near-identical consensus genomes, we observed a few instances in which consensus genomes differed

substantially. In particular, USA/WI-UW-476/2020 differed from both other genomes from the same household by 11 mutations, strongly suggesting that this individual was independently infected.

To determine whether putative household transmission pairs shared more within-host variation than randomly sampled pairs of individuals, we performed a permutation test. We randomly sampled individuals with replacement and computed the proportion of iSNVs shared among random pairs to generate a null distribution (**Figure 4b**, grey bars). We then computed the proportion of variants shared among each putative household transmission pair. Finally, we compared the distribution of shared variants among household pairs and random pairs (**Figure 4b**). 90% of random pairs do not share any iSNVs. Although household pairs share more iSNVs than random pairs on average, half (14/28) of all household pairs share no iSNVs at all. Only 7 out of 28 of household pairs share more iSNVs than expected by chance ($p < 0.05$).

While we hypothesized that putative transmission linkage would be the best predictor of sharing iSNVs, other processes could also result in shared iSNVs. For example, if transmission bottlenecks are wide and iSNVs are efficiently transmitted along transmission chains, then iSNVs may be propagated during community transmission. If so, then iSNVs should be shared among samples that are phylogenetically close together. If transmission chains circulate within local geographic areas, then iSNVs may be commonly shared by samples from the same geographic location. Finally, if iSNVs are

strongly constrained by genetic backbone, then variants may be more likely to be shared across samples from the same clade.

To measure the contribution of these factors, we computed the proportion of iSNVs shared by each pair of samples in our dataset (including household and non-household samples), and model the proportion of shared iSNVs as the combined effect of phylogenetic divergence between the tips (i.e., the branch length in mutations between tips), clade membership, geographic distance between sampling locations, and household membership. Phylogenetic divergence and geographic distance between sampling locations have minimal predicted impact on iSNV sharing (**Figure 4c** and **Supplemental Figure 9**). The strongest predictor of sharing iSNVs is being sampled from the same household, which increased the predicted proportion of shared iSNVs by 0.22 (0.16 - 0.27, 95% HPDI). Belonging to the same clade increases the predicted proportion of shared iSNVs by 0.043 (0.033 - 0.053, 95% HPDI), likely because sharing a within-host variant is contingent on sharing the same consensus base. Taken together, being sampled from the same household is the strongest predictor of sharing iSNVs, and some household pairs share more variation than expected by chance. However, these effects are modest. Given the low overall diversity within hosts and presence of shared iSNVs, the degree of sharing we observe is unlikely sufficient for inferring transmission linkage independent of epidemiologic investigation.

Transmission bottlenecks are likely narrow, and sensitive to variant calling threshold

The number of viral particles that found infection is a crucial determinant of the pace at which novel, beneficial variants can emerge. Narrow transmission bottlenecks can induce a founder effect that purges low-frequency iSNVs, regardless of their fitness. Conversely, wide transmission bottlenecks result in many viral particles founding infection, reducing the chance that beneficial variants are lost. Understanding the size of the transmission bottleneck is therefore important for evaluating the probability that novel SARS-CoV-2 variants arising during acute infection will be transmitted onward. To infer transmission bottleneck sizes, we applied the beta-binomial inference method [137]. We inferred transmission directionality using the date of symptom onset or date of sample collection (see methods for details). If this information was not informative, we calculated a bottleneck size bi-directionally evaluating each individual as the possible donor. In total, we performed 40 transmission bottleneck size estimates in 28 putative household pairs.

iSNV frequencies in donor and recipient pairs are plotted in **Figure 5a**. Most iSNVs detected in the donor are either lost or fixed following transmission in the recipient. However, there are a few low-frequency and near-fixed iSNVs which are shared in donor-recipient pairs. The combined maximum likelihood estimate for mean transmission bottleneck size at our defined 3% frequency threshold is 15 (95% CI: 11-21), although results vary across pairs (**Figure 5b**). Prior transmission bottleneck estimates have changed based on the variant-calling threshold employed [54,279]. To determine whether our estimates were sensitive to our choice of a 3% variant threshold, we evaluated bottleneck sizes using variant thresholds ranging from 1% to 20%. We estimate the highest mean transmission bottleneck size when we employ a 1% frequency threshold

(38, 95% CI: 33-43), and lowest when we use a ≥7% frequency threshold (2, 95% CI: 1-4) (**Figure 5c; Supplemental Figure 10**). The finding of larger bottleneck sizes at a 1% threshold may be due to increased false-positive iSNVs at lower thresholds, in agreement with our findings that a majority of iSNVs detected in the clonal RNA control occurred at frequencies <3%. Importantly though, while variant threshold clearly impacts estimated bottleneck size, our estimates are quite consistent. Even across a wide range of thresholds, our transmission bottleneck size estimates range from 2-43, and never exceed 50.

The beta-binomial inference method assumes that shared variation in donor-recipient pairs is due to transmission. However, it is possible that shared low-frequency iSNVs are recurring mutations (i.e. homoplasies) that should be excluded from the beta-binomial analysis. One site in particular, a synonymous change at nucleotide 15,168 in ORF1ab, was commonly found at low frequencies in donor-recipient pairs. To account for the possibility that this variant is a homoplasy rather than shared via transmission, we dropped this site from our dataset and re-calculated bottleneck sizes. While bottleneck size estimates decrease in individual pairs where this variant is found (**Supplemental Figure 10c**), the average bottleneck size across all transmission pairs remains low (mean = 9, 95% CI: 6-14).

It is possible that some of the pairs evaluated were not direct transmission pairs. Instead individuals may be part of the same transmission chain or share a common source of infection. We reasoned if two individuals were infected from a common source, then they

may have developed symptoms around the same time. In contrast, if one individual infected the other, then their symptom onset dates should be staggered. To assess this, we compared bottleneck sizes to the time between symptom onset in donor-recipient pairs for which symptom onset dates were available (n=17) (**Supplemental Figure 11**). We observed no clear trend between bottleneck size and symptom onset intervals. Finally, all bottleneck estimates are inherently limited by access to a single time point from each donor and recipient. Because it is impossible to know the exact date of infection and transmission, the donor iSNV frequencies may not reflect the true diversity present at the time of transmission. Taken together, we find that even among household pairs, the number of transmitted viruses is likely small. Although bottleneck size estimates vary by variant calling threshold, we find consistent support for fewer than 50 viruses founding infection and suspect that the majority of transmission events are founded by very few viruses (<10). Our data suggest that iSNVs generated within-host are generally lost during the transmission event, and are not efficiently propagated among epidemiologically linked individuals.

Discussion

The emergence of divergent SARS-CoV-2 lineages has called into question the role of within-host selection in propagating novel variants. Our results suggest that very limited variation is generated and transmitted during acute SARS-CoV-2 infection. Most infections in our dataset are characterized by fewer than 5 total intersection iSNVs, the majority of which are low-frequency. Most iSNVs are not detected in global consensus genomes, and are rarely detected in downstream branches on the phylogenetic tree. We

show that even among putative household transmission pairs, iSNVs are shared infrequently, and we estimate that a small number of viruses found infection after most transmission events. The combination of low overall within-host diversity, tight transmission bottlenecks, and infrequent propagation along transmission chains may slow the rate of novel variant emergence among acutely infected individuals. Importantly, our results imply that the accumulation of multiple iSNVs is unlikely during typical, acute infection. Together, our findings are consistent with a regime in which typical acute infections play a limited role in the generation and spread of new SARS-CoV-2 variants, and argue for the need to better understand the role of prolonged infections as a source of consequential new variants. Targeted interventions to prevent the number of long-term infections and to prevent transmission from persistently infected individuals may be particularly fruitful for slowing the rate of emergence of novel variants of concern.

Relatively few studies have reported on SARS-CoV-2 within-host diversity, and their results have varied. SARS-CoV-2 within-host sequence data appear to be particularly vulnerable to method error, including sensitivity to cycle threshold [56,228], putative false positive iSNV calls in control runs [56], an uncertain degree of recurrent mutations shared across unrelated samples [54,228,278,290], and variation between technical replicates. Complicating matters, each lab employs its own sample preparation and variant calling pipelines, making comparison across datasets challenging, and concern has been raised regarding recurrent errors that are platform- and lab-specific [291]. iSNVs that recur in nature pose a challenge because they result in the same data pattern that would be expected

from recurrent pipeline errors. We have attempted to employ multiple, overlapping controls to mitigate errors that could arise from sample preparation, bioinformatic processing, and improper variant thresholds. In particular, our results emphasize the importance of duplicate sequencing for any studies relying on low-frequency iSNVs to infer biological processes. Like Valesano et al. [56] we observe that SARS-CoV-2 variant calls are sensitive to Ct and variant-calling criteria. We echo their expressed caution in interpreting SARS-CoV-2 within-host data in the absence of pipeline-specific controls.

Similar to work reported by others [56,228,290], we find that most samples harbor very few iSNVs, and that most variants are low-frequency. Although we employ distinct methods, we corroborate findings by Lythgoe & Hall et al.[228] that iSNVs do not cluster geographically or phylogenetically, suggesting that they are not transmitted efficiently within communities. One difference is that we detect a higher number of shared/recurrent iSNVs in our dataset than reported by Lythgoe & Hall et al. [228], Valesano et al. [56], and Shen et al. [290], but fewer than Popa & Genger et al. [54] and James et al. [278]. While some degree of shared iSNVs is reported across most SARS-CoV-2 datasets [54,56,228,278,290] the exact frequency of shared sites is highly variable. The higher number of shared iSNVs in our results may be partially accounted for by our method of variant reporting. While most studies mapped reads to the Wuhan-1 reference and report variants present at <50% frequency [54,56,228,290], we converted consensus-level variants to their low-frequency counterparts, and counted the minor allele for near-fixed variants. The higher level of shared iSNVs we observe could also be explained by sampling many closely related, cohabiting individuals. Though relatively few, some household transmission pairs do

share iSNVs, likely accounting for some of the shared variation we observe. Future work will be necessary to determine the precise degree to which iSNVs recur across unrelated individuals and the extent to which factors like viral copy number, time of infection, host factors including pre-existing immunity, and sequencing pipeline influence these estimates.

Four other groups have previously estimated the size of the SARS-CoV-2 transmission bottleneck, although the total number of transmission events evaluated to date across studies remains small (~66). Lythgoe & Hall et al. (n=14 pairs) [292], James & Ngcapu et al. (n=11 pairs) [278], and Wang et al. (n=2 pairs) [227] report narrow bottlenecks, in which infection is founded by fewer than 10 viruses. Popa & Genger et al. (n=39 pairs) [54] report bottleneck sizes ranging from 10 to 5000, and an average size of 1000. Reanalysis of the Popa & Genger data using a more conservative variant dataset resulted in an average bottleneck size of 1-3 [279]. Similarly, we find a combined average bottleneck size of 15 using a 3% frequency threshold, and 2 using a 7% frequency threshold. Thus, current evidence is converging to support narrow transmission bottlenecks for SARS-CoV-2, similar to influenza virus [122,133,135]. Still, these estimates rely on a small number of putative transmission events, including the pairs analyzed here. Genuine differences in the SARS-CoV-2 transmission bottleneck size, depending on route of transmission [134] and host factors may exist.

When transmission bottlenecks are narrow, even beneficial variants present at low frequencies in the transmitting host are likely to be lost. However, the recent emergence

of multiple divergent lineages, some of which increase infectiousness, underscore that transmission of such variants clearly can occur [293]. This raises the question: how did these variants make their way out of individual hosts? Narrow transmission bottlenecks generally purge within-host diversity through a founder effect. Although rare, a low-frequency variant that successfully passes through a transmission bottleneck could quickly become the dominant variant in the next host. Such events would become increasingly common as the total number of infected individuals and transmission events occurring in the population climbs, making it possible to observe these rare events.

The model outlined above aligns with the hypothesis that prolonged SARS-CoV-2 infection leads to accumulation of intrahost mutations [265–269]. Prolonged infections may permit additional cycles of viral replication, allowing for more variants to be generated and more time for selection to increase the frequency of beneficial variants. Even a modest increase in frequency within a donor enhances the likelihood of a beneficial variant becoming fixed following transmission in the setting of a narrow transmission bottleneck. Alternatively, it is possible for selection to act during transmission such that some viruses harboring a particular mutation or group of mutations are preferentially transmitted [204]. In a previous study evaluating SARS-CoV-2 genetic diversity within and between domestic cats, we documented modest evidence supporting preferential transmission of a particular nonsynonymous variant in Spike [57]. However, we saw no evidence for selective bottlenecks in this study. Additional studies evaluating the SARS-CoV-2 transmission bottleneck are needed, in particular in the setting of long-term infections and immunocompromised hosts.

Our findings that within-host variation is limited and infrequently transmitted are important. Our data, combined with findings from others, suggest that rapid accumulation of novel mutations within-host is not the norm during acute infection. Like influenza viruses, a significant portion of variation generated within one infected host is likely lost during transmission. The combination of within-host limited diversity and tight transmission bottlenecks should slow the pace at which novel, beneficial variants could emerge during transmission among acutely infected individuals. Future studies that compare within-host diversity in individuals with and without SARS-CoV-2 antibodies will be necessary to evaluate whether immunity imposes signatures of within-host selection. Finally, given the increasing appreciation for the potential role of long infections to promote variant emergence, within-host data may provide its maximum benefit for dissecting the process of variant evolution during prolonged infections.

Acknowledgments

LHM is supported by NIAID grant number K99 AI147029-01. GKM is supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359).

Figures, tables, and supplemental material

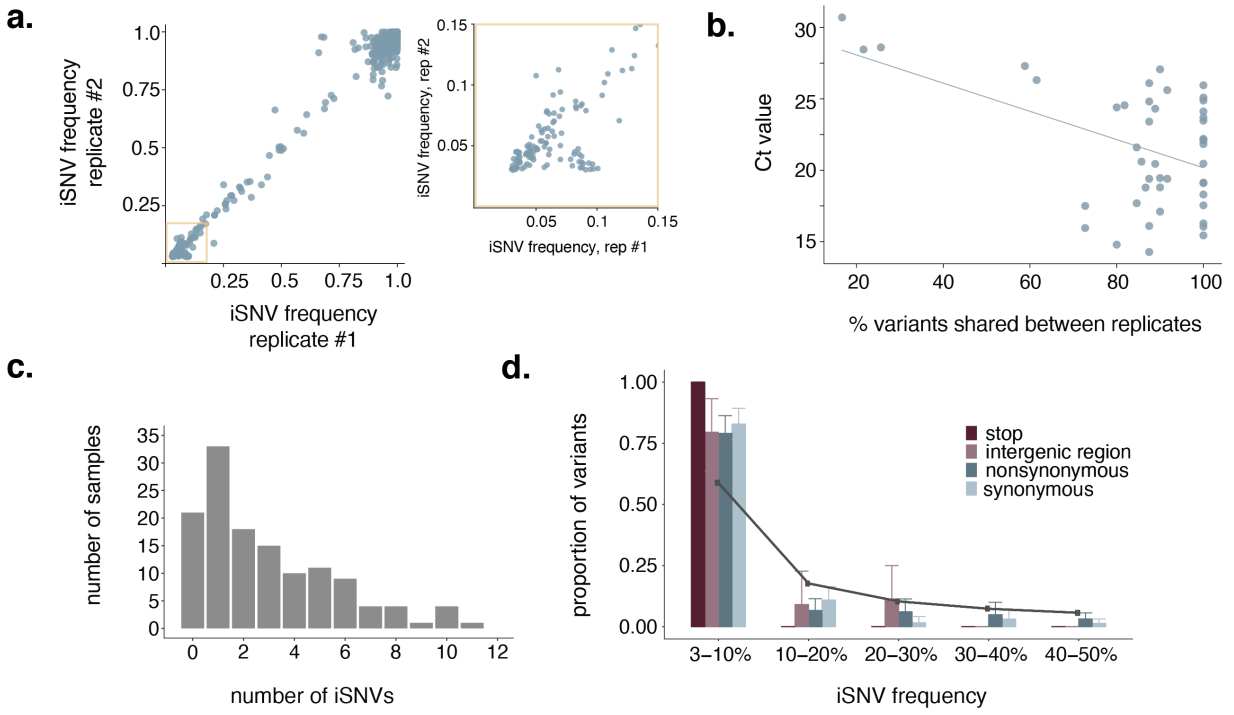


Figure 1: Within host variation is limited after data quality control

a. iSNV frequencies in replicate 1 are shown on the x-axis and frequencies in replicate 2 are shown on y-axis. The yellow box highlights low-frequency iSNVs (3-15%), which is expanded out to the right. **b.** The Ct value is compared to the percent of iSNVs shared between technical replicates. The blue line is a line of best fit to highlight the observed negative trend. **c.** Distribution of the number of total iSNVs detected per sample. Many samples harbor no iSNVs at all, and the maximum number of iSNVs in a single sample was 11. **d.** The proportion of iSNVs that were detected at various within-host frequency bins is shown. Error bars represent the variance in the proportion of total within-host iSNVs within that frequency bin across samples in the dataset as calculated by bootstrapping. There was a single stop variant in the entire dataset, so no error bar is

shown for the stop category. The solid grey line indicates the expected proportion of variants in each frequency bin under a neutral model.

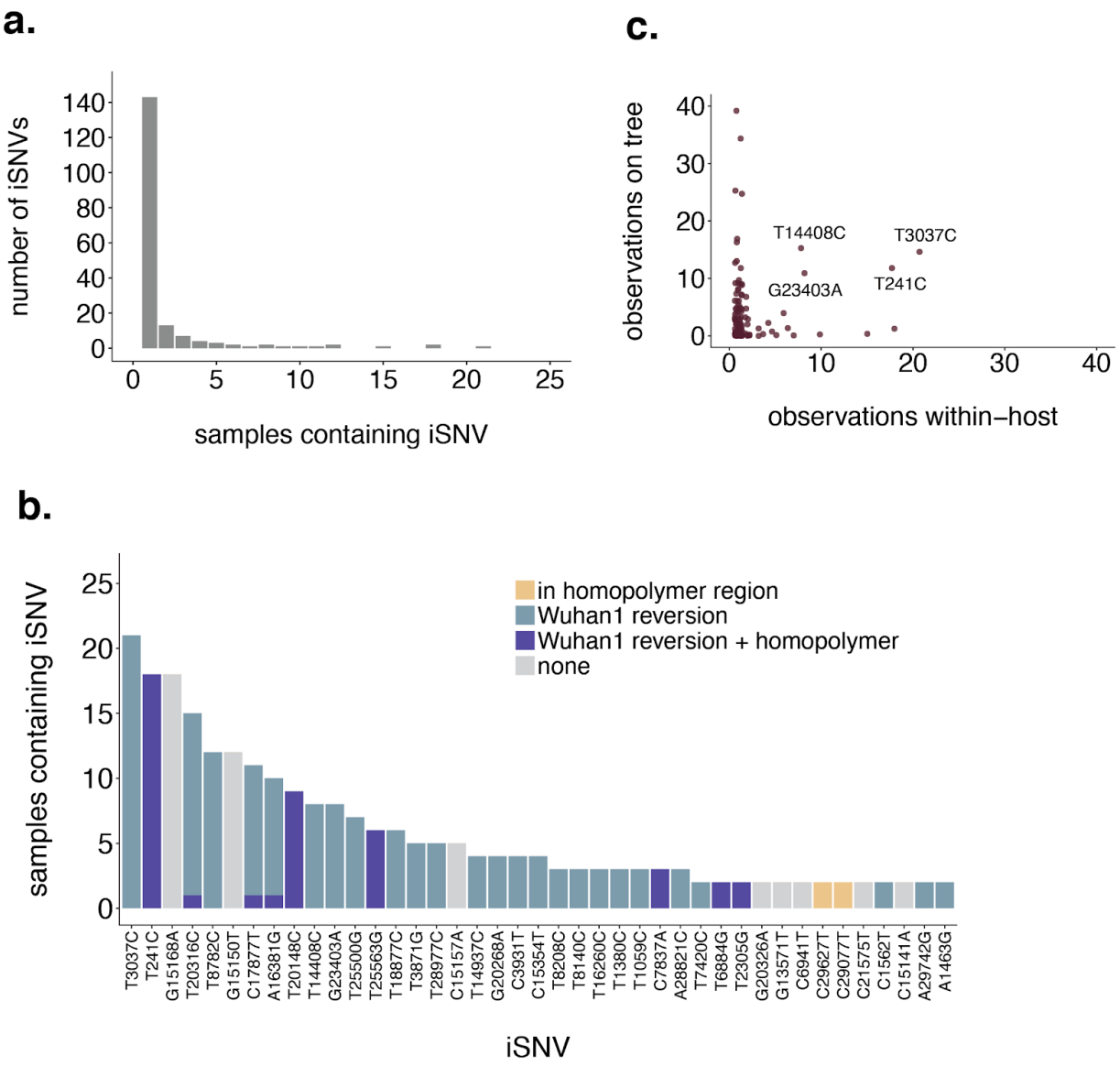


Figure 2: Shared iSNVs represent homopolymers and common polymorphic sites

a. The number of iSNVs (y-axis) present within n individuals (x-axis) is shown. The vast majority of iSNVs are found in only a single sample. 6 iSNVs are shared by at least 10 samples. **b.** Each iSNV detected in at least 2 samples is shown. Variants that occur within, or 1 nucleotide outside of, a homopolymer region (classified as a span of the same base

that is at least 3 nucleotides long) are colored in yellow. Variants that represent the minor allele for variants that were nearly fixed at consensus (annotated here as “Wuhan1 reversions”) are shown in blue, and variants that were both Wuhan1 reversions and occurred in homopolymer regions are colored in purple. **c.** For each unique iSNV detected within a host, the x-axis represents the number of samples in which that iSNV was detected, and the y-axis represents the number of times it is present on the global SARS-CoV-2 phylogenetic tree. The counts on the phylogenetic tree represent the number of times the mutation arose along internal and external branches. The variants labeled with text are those that are detected at least 5 times within-host and at least 5 times on the phylogeny. Two of the most commonly detected iSNVs, T3037C and T241C (shown as the furthest to the left in panel b), are also frequently detected on the phylogenetic tree.

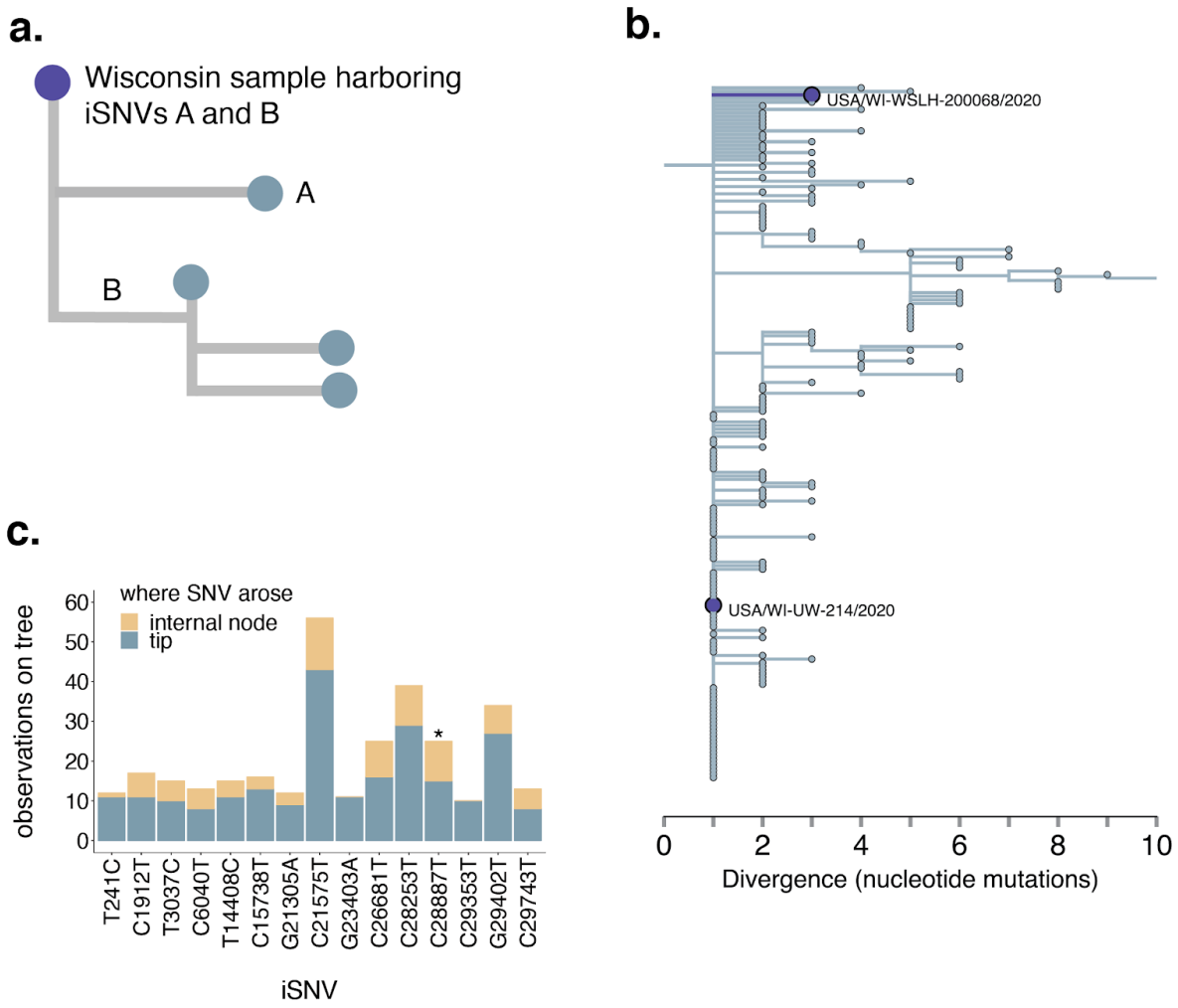


Figure 3: Variants are not common in consensus sequences or in downstream branches

a. We traversed the Wisconsin-focused full-genome SARS-CoV-2 phylogeny from root to tip. For each Wisconsin tip for which we had within-host data, we queried whether any of the iSNVs detected in that sample were ever detected in downstream branches at consensus. In this example, the purple tip represents a Wisconsin sample for which we have within-host data. This sample harbors 2 iSNVs, A and B. iSNV A arises on a tip that falls downstream from the starting, purple tip. iSNV B is present on a downstream branch

leading to an internal node. Both A and B would be counted as instances in which an iSNV was detected at consensus in a downstream branch. **b.** In the Wisconsin-specific phylogenetic tree, we applied the metric described in **a.** Among 110 Wisconsin samples that harbored within-host variation, 93 occurred on internal nodes. Of those, we detect one instance in which a mutation detected as an iSNV in one sequence was detected in a downstream consensus sequence. (C1912T, an iSNV in USA/WI-UW-214/2020, was detected downstream in USA/WI-WSLH-200068/2020.) **c.** For each iSNV identified in the study (in at least 1 sample), we enumerated the number of times that variant occurred on the global SARS-CoV-2 phylogeny on an internal node (yellow) or on a tip (blue). The results for every variant are shown in **Supplemental Figure 6**. Here, we show only the variants that were detected at least 10 times on the global phylogeny. Each such iSNV is found at internal nodes and tips at a ratio comparable to overall mutations on the tree, except for C28887T, which is enriched on internal nodes ($p=0.028$, Fishers' exact test). * indicates p-value < 0.05.

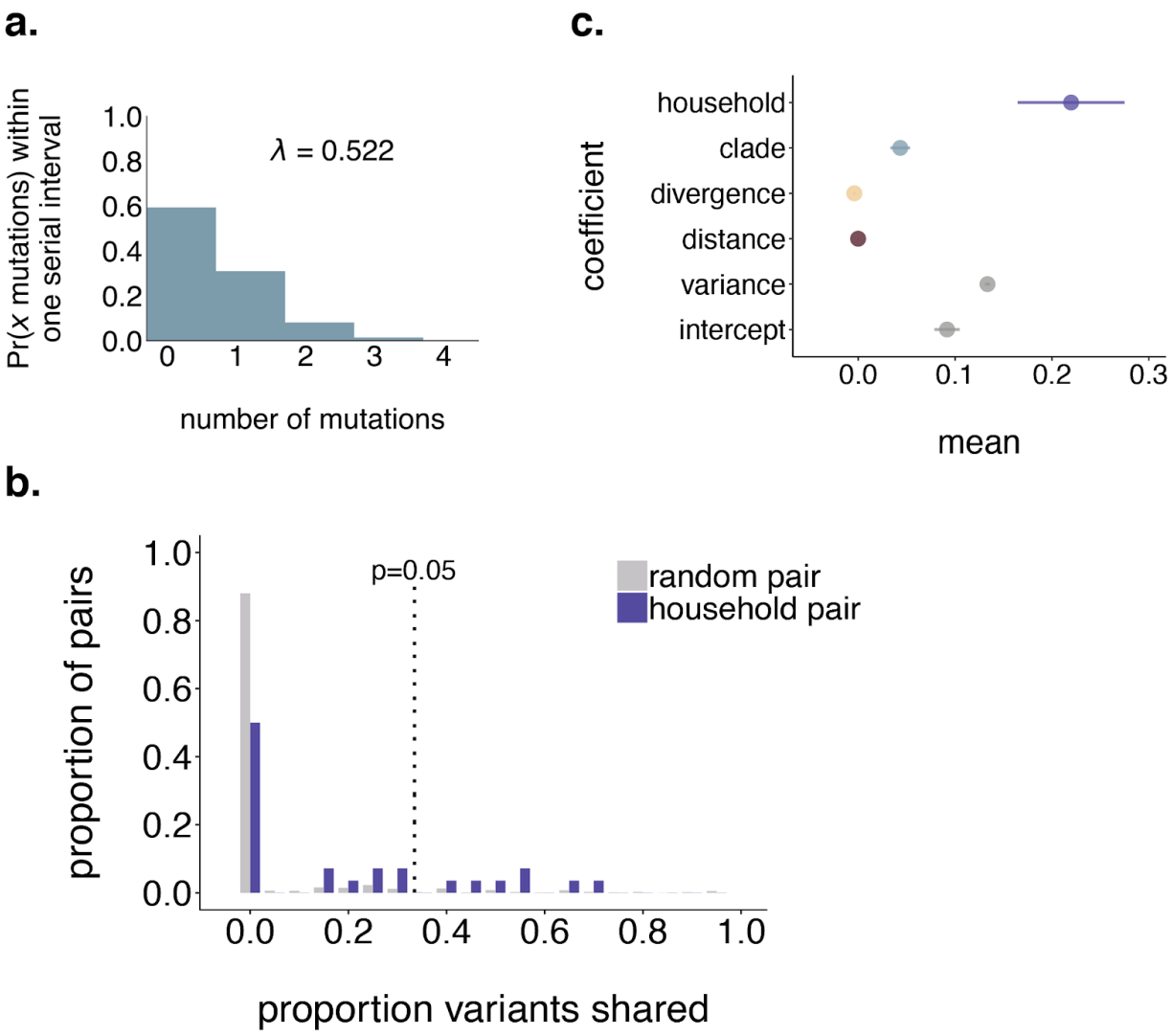


Figure 4: Household pairs share a modest degree of within-host variation

a. We modeled the probability that 2 consensus genomes will share x mutations as Poisson-distributed with lambda equal to the number of mutations expected to accumulate in the SARS-CoV-2 genome over 5.8 days [285] given a substitution rate of 1.10 x 10^{-3} substitutions per site per year [153]. Exploration of how these probabilities change using a range of plausible serial intervals and substitution rates is shown in **Supplemental Figure 8**. The vast majority of genomes that are separated by one serial

interval are expected to differ by ≤2 consensus mutations. **b.** The proportion of random pairs (grey) and putative household transmission pairs (purple) is shown on the y-axis vs. the proportion of iSNVs shared. The dotted line indicates the 95th percentile among the random pairs. Household pairs that share a greater proportion of iSNVs than 95% of random pairs (i.e., are plotted to the right of the dotted line) are considered statistically significant at p=0.05. iSNVs had to be present at a frequency of ≥3% to be considered in this analysis. **c.** We assessed the impact of household membership, clade membership, phylogenetic divergence, and geographic distance on the proportion of iSNVs shared between each pair of samples in our dataset. The mean of each estimated coefficient in the combined linear regression model including all predictors is shown on the x-axis, with lines of spread indicating the range of the estimated 95% highest posterior density interval (HPDI).

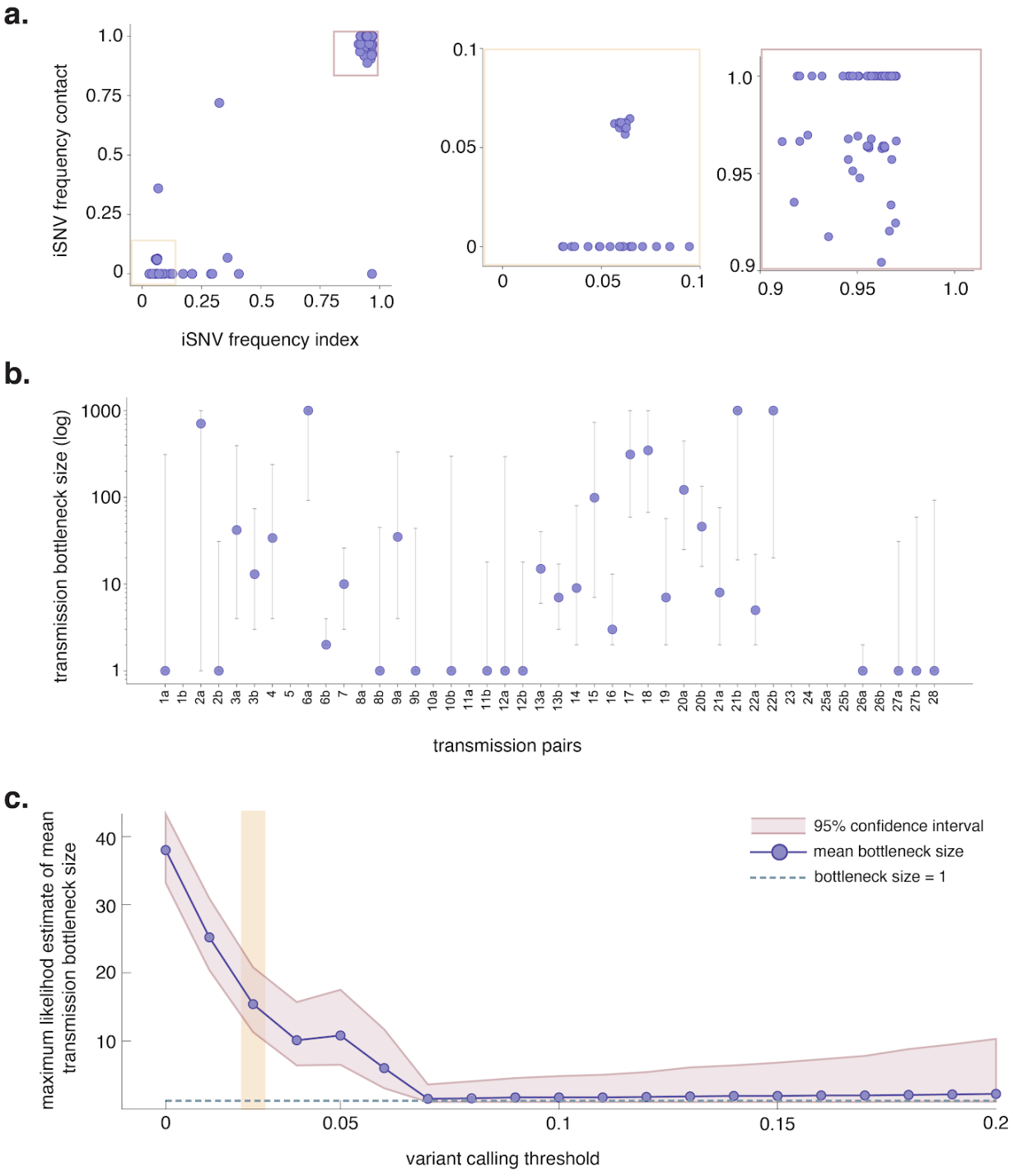
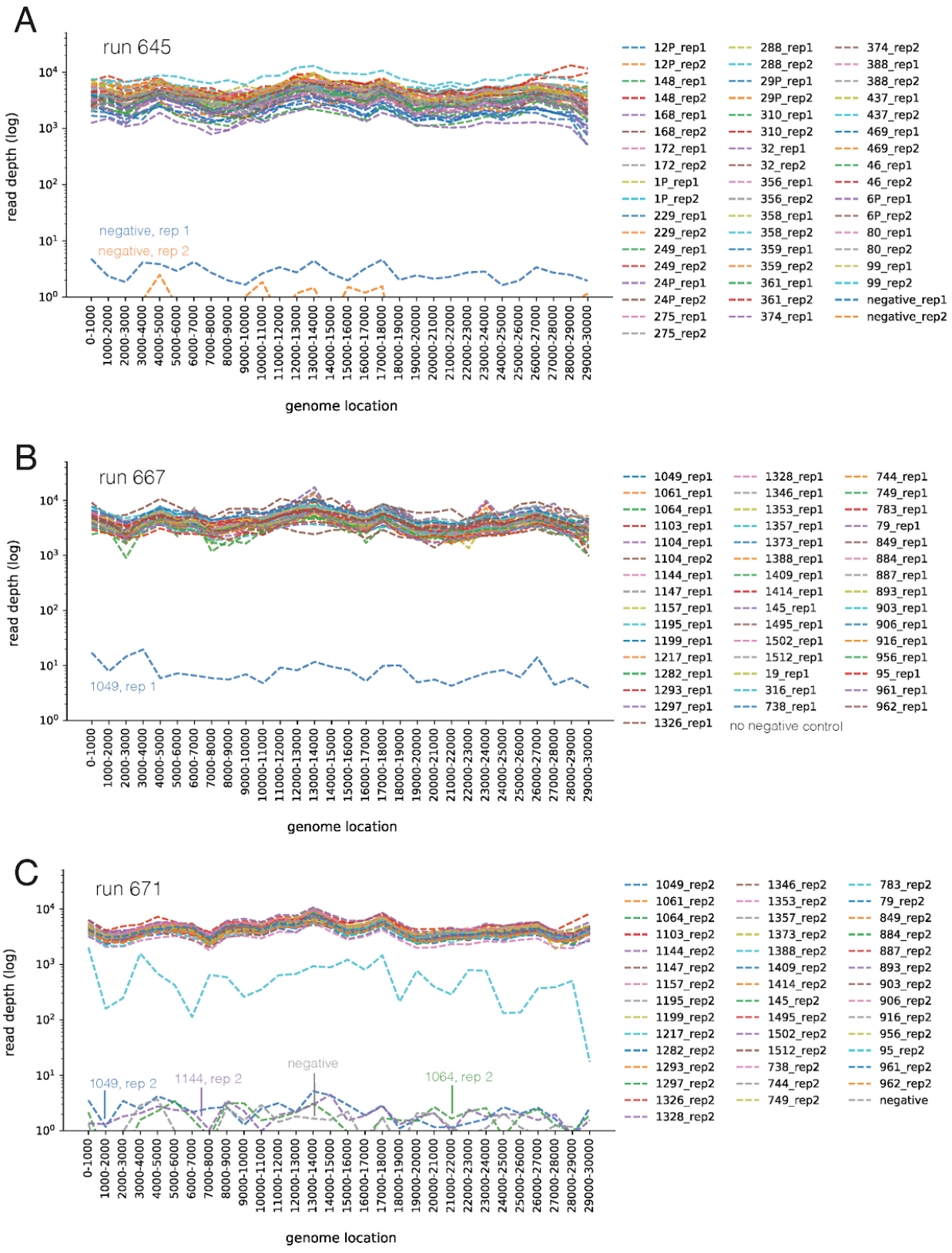


Figure 5: SARS-CoV-2 transmission bottlenecks in household transmission pairs

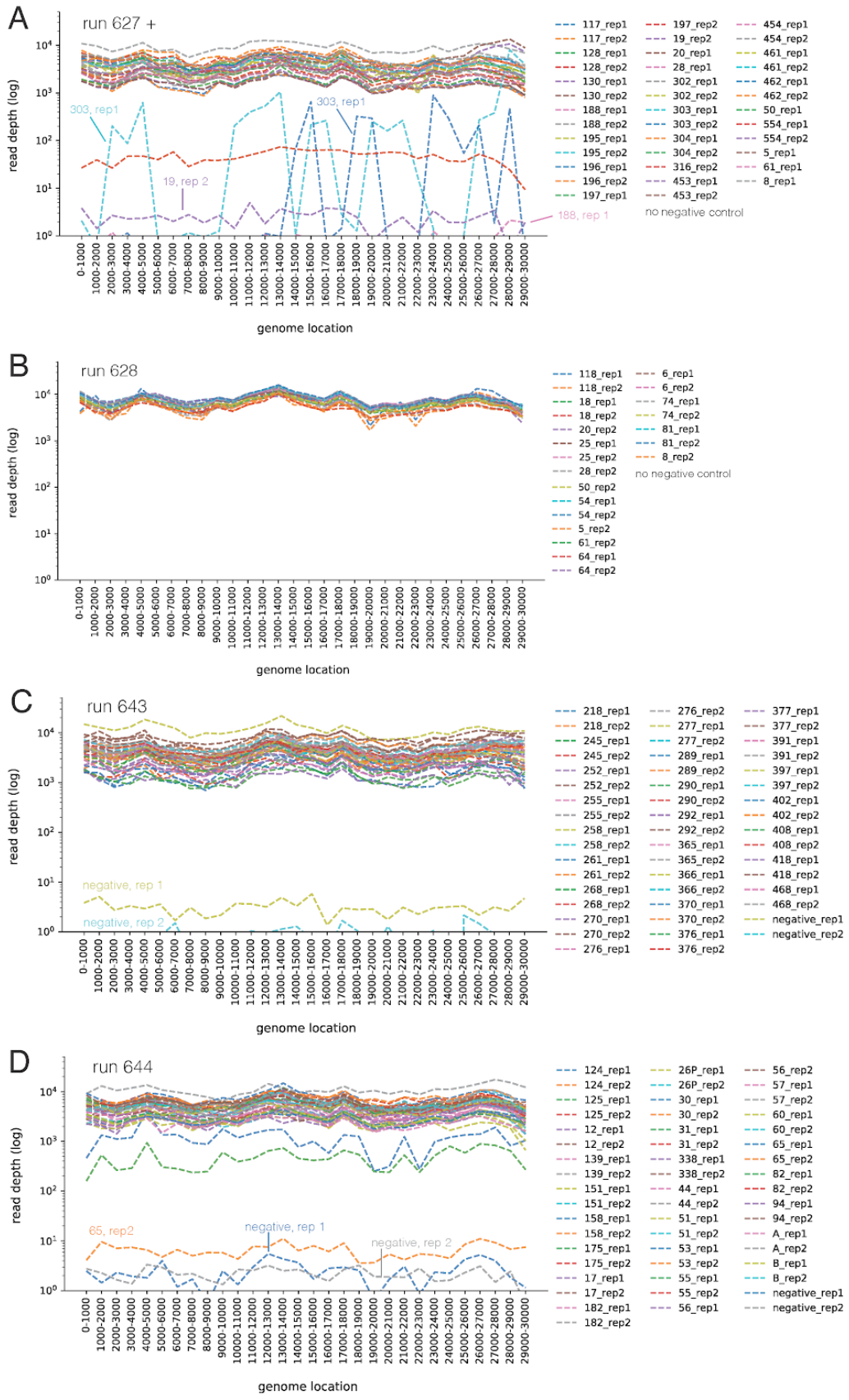
a. "TV plots" showing intersection iSNV frequencies in all 44 donor-recipient pairs using a 3% frequency threshold. The yellow box highlights low-frequency iSNVs (3-10%) and

the mauve box highlights high-frequency iSNVs (90-100%). **b.** Maximum likelihood estimates for mean transmission bottleneck size in individual donor-recipient pairs. Bottleneck sizes could not be estimated for a few pairs (e.g. pairs 5, 10a, 11a, etc) because there were no polymorphic sites detected in the donor. **c.** Bidirectional comparisons are denoted with an “a” and “b” following the pair number. Combined maximum likelihood estimates across all 44 donor-recipient pairs plotted against variant calling thresholds ranging from 1-20%. The purple line shows combined estimates at each variant calling threshold shown and the mauve band displays the 95% confidence interval for this estimate. The dashed grey line indicates a bottleneck size equal to 1. The vertical yellow band highlights the combined transmission bottleneck size using a 3% variant calling threshold.


A
run 645
read depth (log)
negative, rep 1
negative, rep 2
genome location
12P_rep1
12P_rep2
148_rep1
148_rep2
168_rep1
168_rep2
172_rep1
172_rep2
1P_rep1
1P_rep2
229_rep1
229_rep2
249_rep1
249_rep2
24P_rep1
24P_rep2
275_rep1
275_rep2
288_rep1
288_rep2
29P_rep1
29P_rep2
310_rep1
310_rep2
32_rep1
32_rep2
356_rep1
356_rep2
358_rep1
358_rep2
359_rep1
359_rep2
361_rep1
361_rep2
374_rep1
374_rep2
388_rep1
388_rep2
437_rep1
437_rep2
469_rep1
469_rep2
46_rep1
46_rep2
6P_rep1
6P_rep2
80_rep1
80_rep2
99_rep1
99_rep2
negative_rep1
negative_rep2
B
run 667
read depth (log)
1049, rep 1
genome location
1049_rep1
1061_rep1
1064_rep1
1103_rep1
1104_rep1
1104_rep2
1144_rep1
1147_rep1
1157_rep1
1195_rep1
1199_rep1
1217_rep1
1282_rep1
1293_rep1
1297_rep1
1326_rep1
1328_rep1
1346_rep1
1353_rep1
1357_rep1
1373_rep1
1388_rep1
1409_rep1
1414_rep1
145_rep1
1495_rep1
1502_rep1
1512_rep1
19_rep1
316_rep1
738_rep1
744_rep1
749_rep1
783_rep1
79_rep1
849_rep1
884_rep1
887_rep1
893_rep1
903_rep1
906_rep1
916_rep1
956_rep1
95_rep1
961_rep1
962_rep1
no negative control
C
run 671
read depth (log)
1049, rep 2
1144, rep 2
negative
1064, rep 2
genome location
1049_rep2
1061_rep2
1064_rep2
1103_rep2
1144_rep2
1147_rep2
1157_rep2
1195_rep2
1199_rep2
1217_rep2
1282_rep2
1293_rep2
1297_rep2
1326_rep2
1328_rep2
1346_rep2
1353_rep2
1357_rep2
1373_rep2
1388_rep2
1409_rep2
1414_rep2
145_rep2
1495_rep2
1502_rep2
1512_rep2
738_rep2
744_rep2
749_rep2
783_rep2
79_rep2
849_rep2
884_rep2
887_rep2
893_rep2
903_rep2
906_rep2
916_rep2
956_rep2
95_rep2
961_rep2
962_rep2
negative

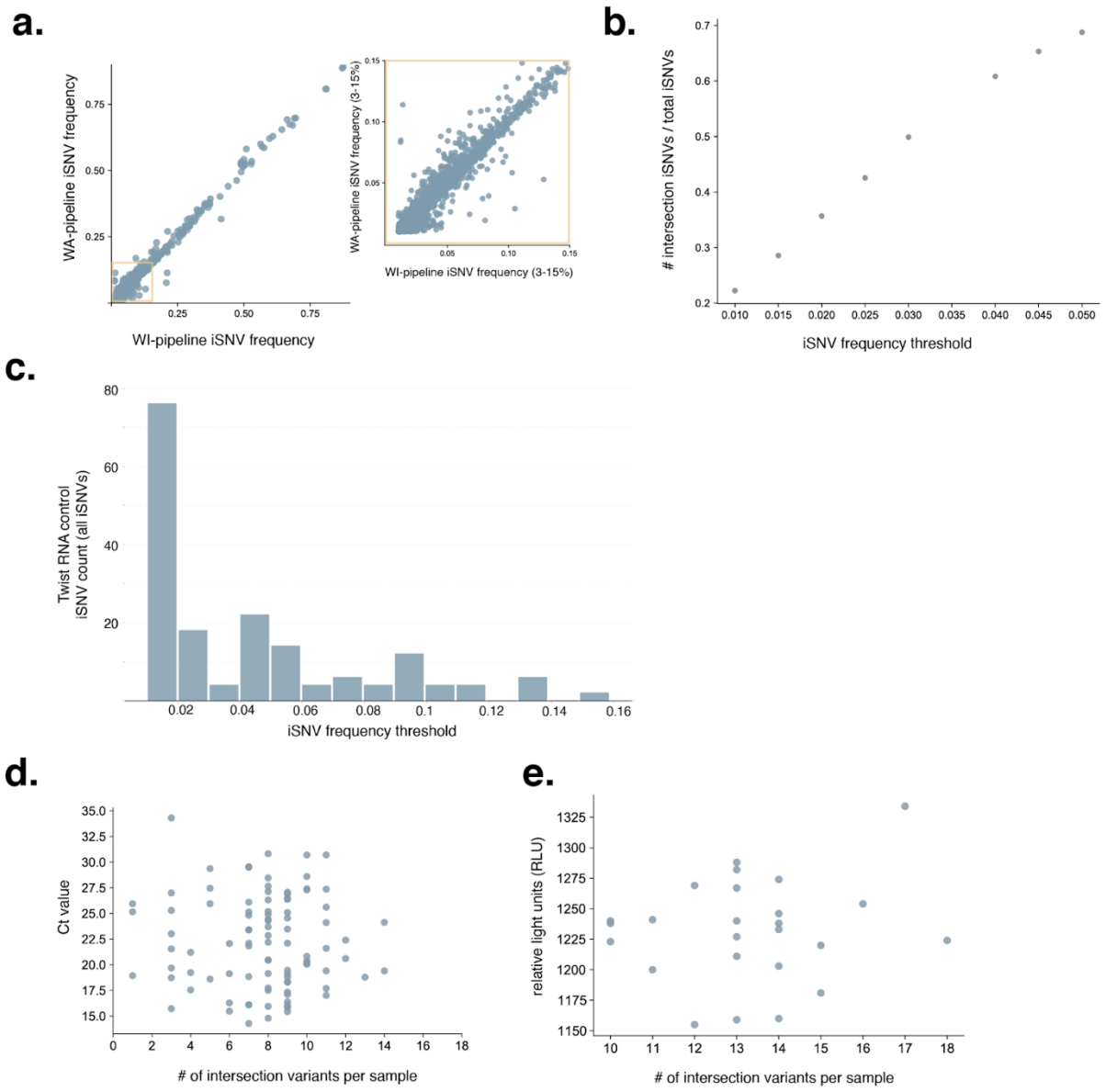
Supplemental Figure 1: Read depth

Read depth by genome location in 1,000-bp bins for MiSeq runs **a.** 627, **b**. 628, **c.** 643, and **d.** 644. Water controls and low-coverage samples are labeled. Samples included in each run are labeled according to the color to the right of each plot.


A
run 627 +
read depth (log)
303, rep1
303, rep1
19, rep 2
188, rep 1
no negative control
genome location
B
run 628
read depth (log)
no negative control
genome location
C
run 643
read depth (log)
negative, rep 1
negative, rep 2
genome location
D
run 644
read depth (log)
65, rep2
negative, rep 1
negative, rep 2
genome location

Supplemental Figure 2: Read depth

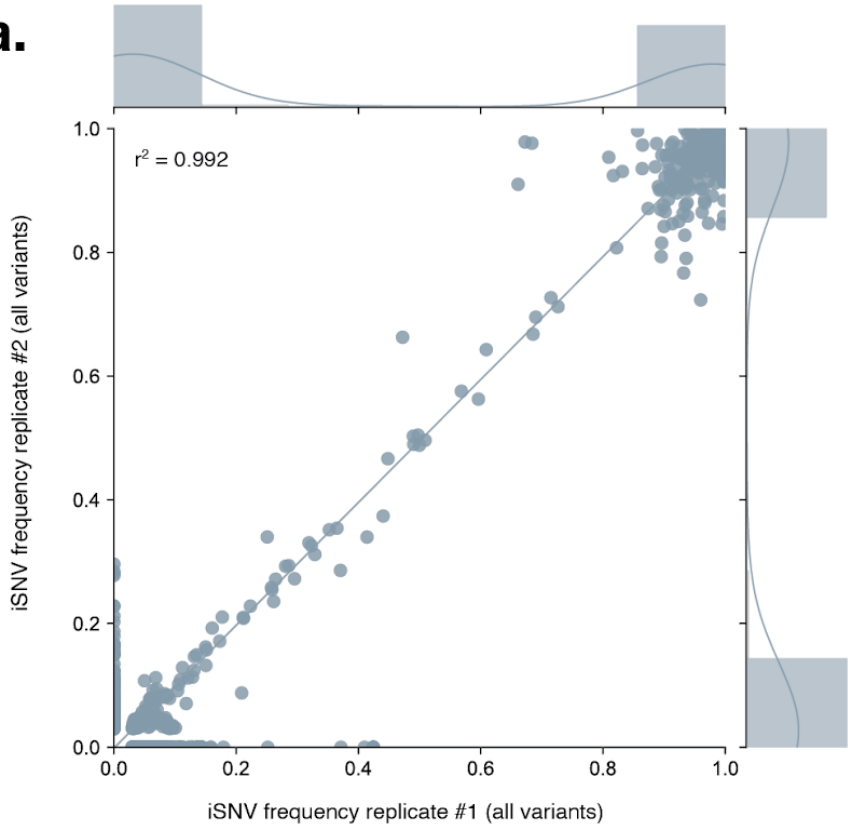
Read depth by genome location in 1,000-bp bins for MiSeq runs **a.** 645, **b**. 667, and **c.** 671. Water controls and low-coverage samples are labeled within each plot. Samples included in each run are labeled according to the color to the right of each plot.



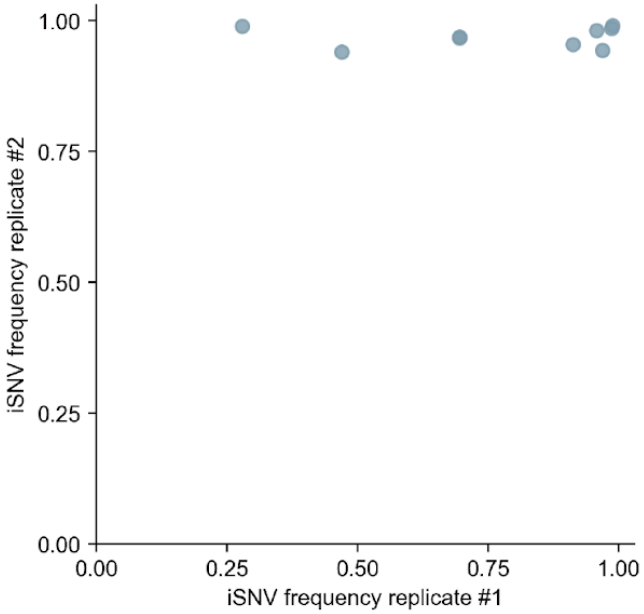
Supplemental Figure 3: Additional iSNV quality control information

a. Variant frequencies generated using the Wisconsin bioinformatic pipelines are shown on the x-axis and frequencies generated using the Washington bioinformatic pipeline are shown on the y-axis. The yellow box highlights low-frequency variants (3-15%), which is expanded out to the right. **b**. Proportion of intersection iSNVs relative to the total number of iSNVs increases as variant frequency threshold increases. **c**. The total number of iSNVs detected across both Twist RNA control replicates compared to the iSNV frequency threshold. The majority of iSNVs detected in these clonal samples occur <3% frequency. Note that the iSNVs reported in **Supplemental Table 1** are intersection iSNVs only. The identities of all iSNVs detected ≥1% frequency in the Twist RNA control can be found in the GitHub accompanying this manuscript. **d.** The number of intersection variants is compared to the Ct value for all samples where a Ct value was available. Out of 133 total samples, Ct values were available for 94. **e.** The number of intersection variants is compared to the RLU (relative light unit) value for all samples where a RLU value was available.

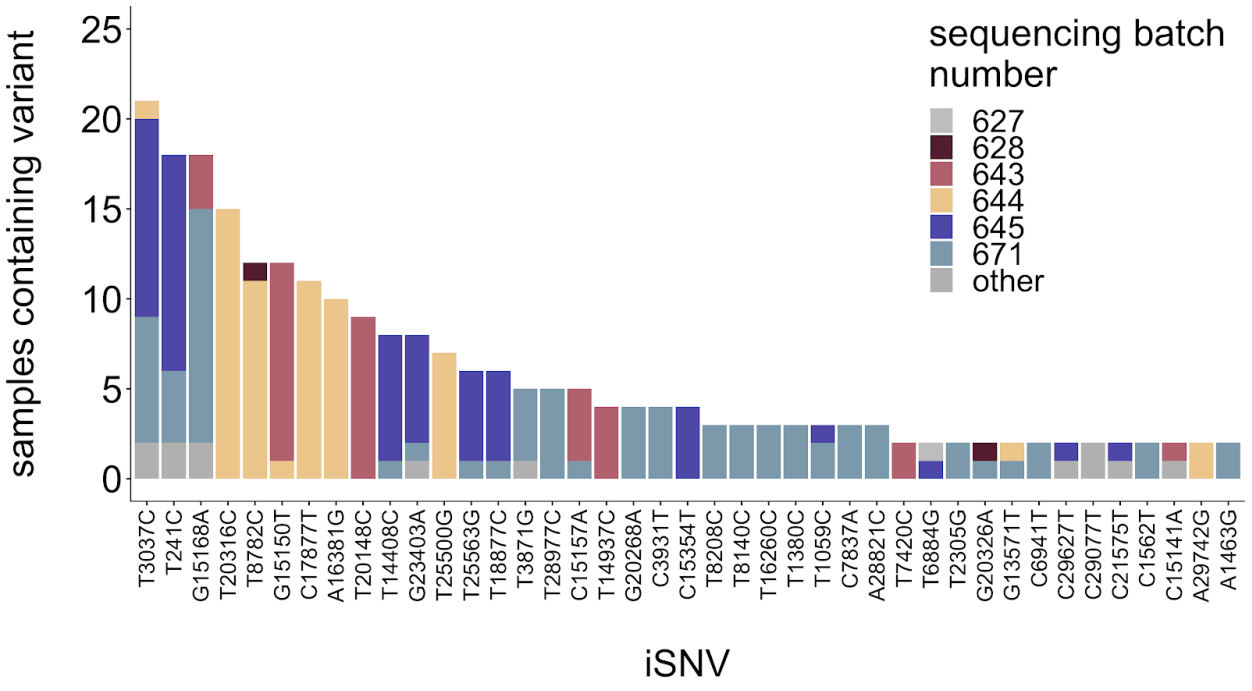
a.


r² = 0.992
iSNV frequency replicate #2 (all variants)
iSNV frequency replicate #1 (all variants)
0.0
0.2
0.4
0.6
0.8
1.0

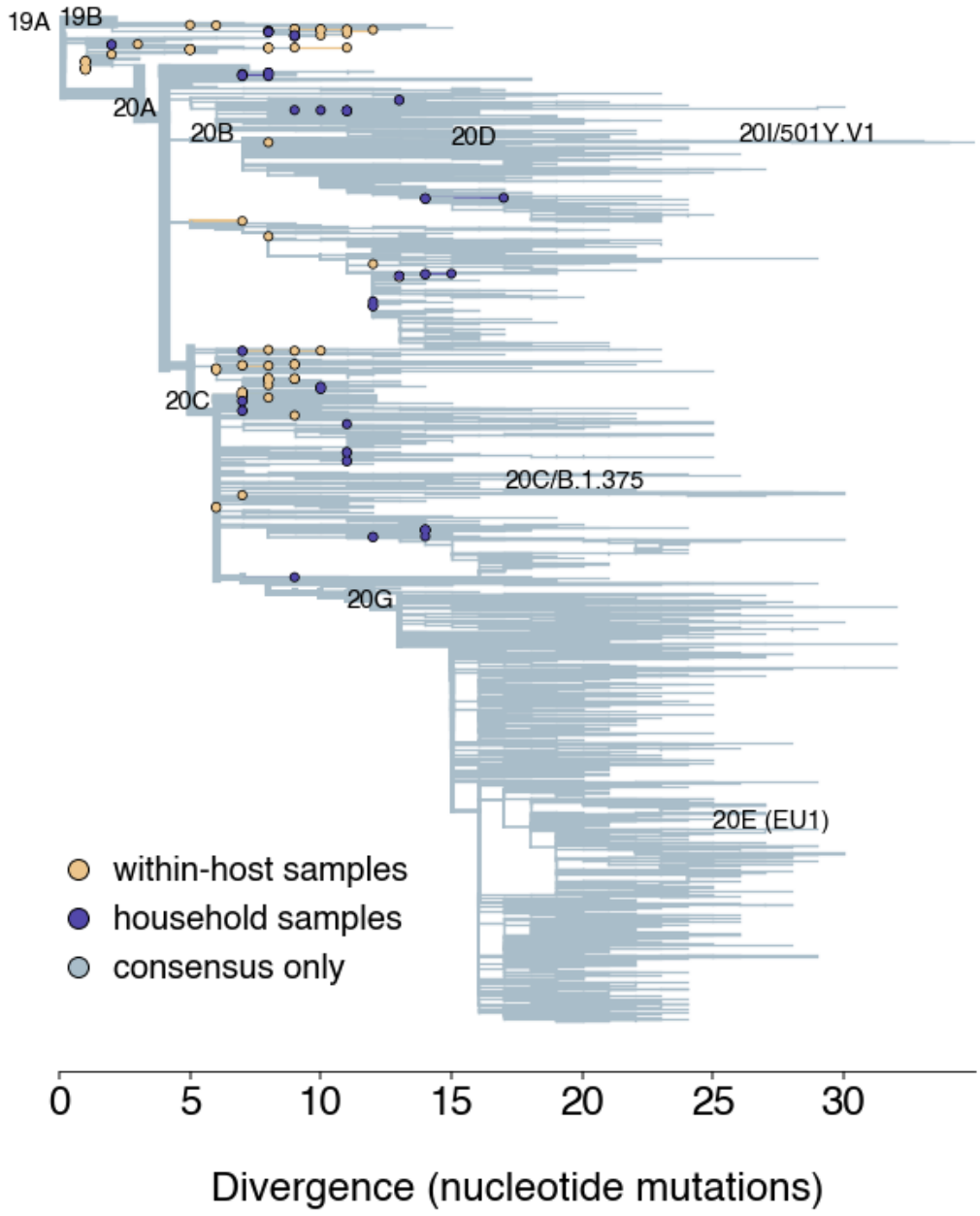
b.


iSNV frequency replicate #2
iSNV frequency replicate #1
0.00
0.25
0.50
0.75
1.00

Supplemental Figure 4: iSNVs in technical replicates across all samples. a. Variant frequencies in replicate 1 are shown on the x-axis and frequencies in replicate 2 are shown on y-axis. This plot includes all variants found in both replicates and not just the intersection variants as shown **Figure 1a**. **b.** Example of one sample with very poor overlap between technical replicates; this sample (sample 1104) was excluded from the experimental dataset.

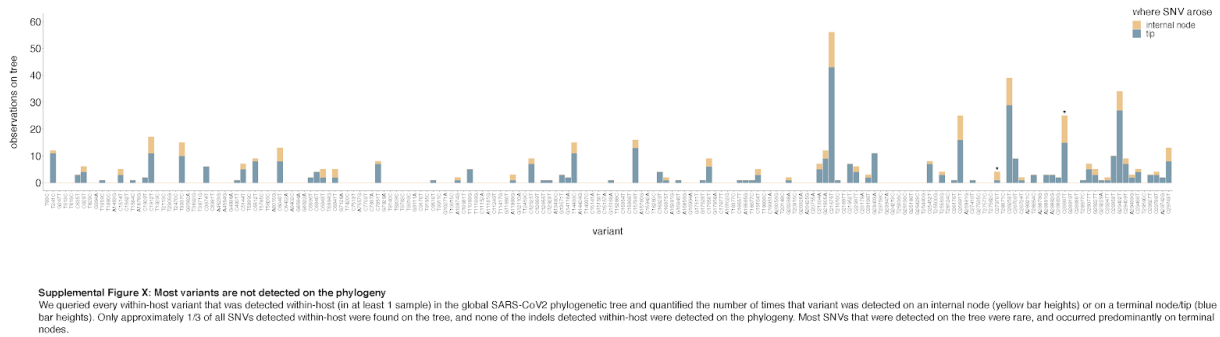


Supplemental Figure 5: iSNVs do not cluster by sequencing run. iSNVs detected in at least 2 samples are shown on the x-axis and are plotted against the number of times they are detected in our dataset. Each iSNV bar is colored according to the number of times it was detected within each sequencing batch.


19A
19B
20A
20B
20D
20I/501Y.V1
20C
20C/B.1.375
20G
20E (EU1)
within-host samples
household samples
consensus only
0
5
10
15
20
25
30
Divergence (nucleotide mutations)

Supplemental Figure 6: Wisconsin divergence phylogeny

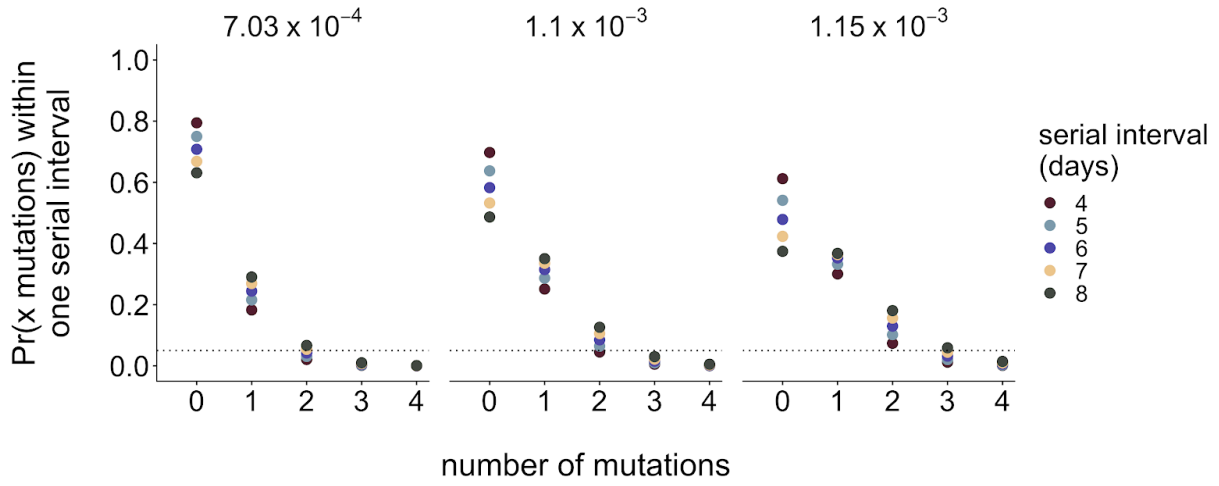
A full-genome phylogenetic tree built showing X Wisconsin consensus sequences with the Nextstrain pipeline is shown. The x-axis represents divergence expressed as the number of nucleotide mutations. Nextstrain clade labels are shown on the corresponding branch. Yellow tips represent Wisconsin samples that were Illumina sequenced in duplicate and analyzed in this manuscript. Purple tips represent samples from households.



Supplemental Figure X: Most variants are not detected on the phylogeny
We queried every within-host variant that was detected within-host (in at least 1 sample) in the global SARS-CoV2 phylogenetic tree and quantified the number of times that variant was detected on an internal node (yellow bar heights) or on a terminal node/tip (blue bar heights). Only approximately 1/3 of all SNVs detected within-host were found on the tree, and none of the indels detected within-host were detected on the phylogeny. Most SNVs that were detected on the tree were rare, and occurred predominantly on terminal nodes.

Supplemental Figure 7: Most iSNVs are not detected on the phylogeny

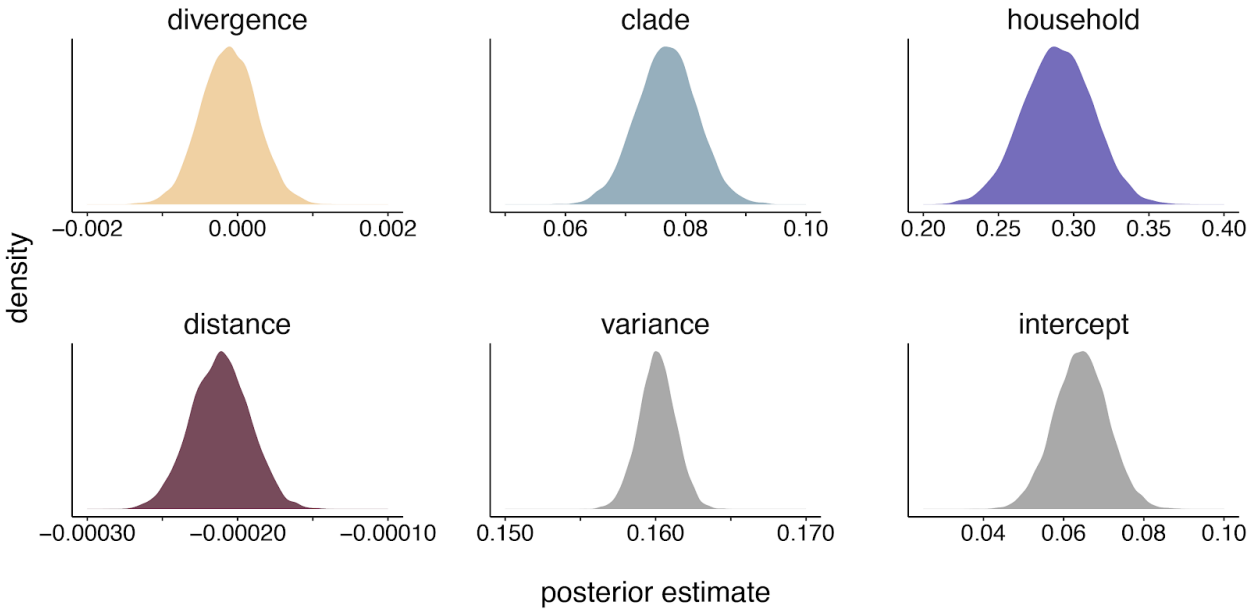
We queried every iSNV that was detected within-host (in at least 1 sample) in the global SARS-CoV-2 phylogenetic tree and quantified the number of times that iSNV was detected on an internal node (yellow bar heights) or on a terminal node/tip (blue bar heights). Only approximately 1/3 of all SNVs detected within-host were found on the tree, and none of the indels detected within-host were detected on the phylogeny. Most SNVs that were detected on the tree were rare, and occurred predominantly on terminal nodes.



Supplemental Figure 8: Modeling the expected number of mutations distinguishing genomes separated by one serial interval

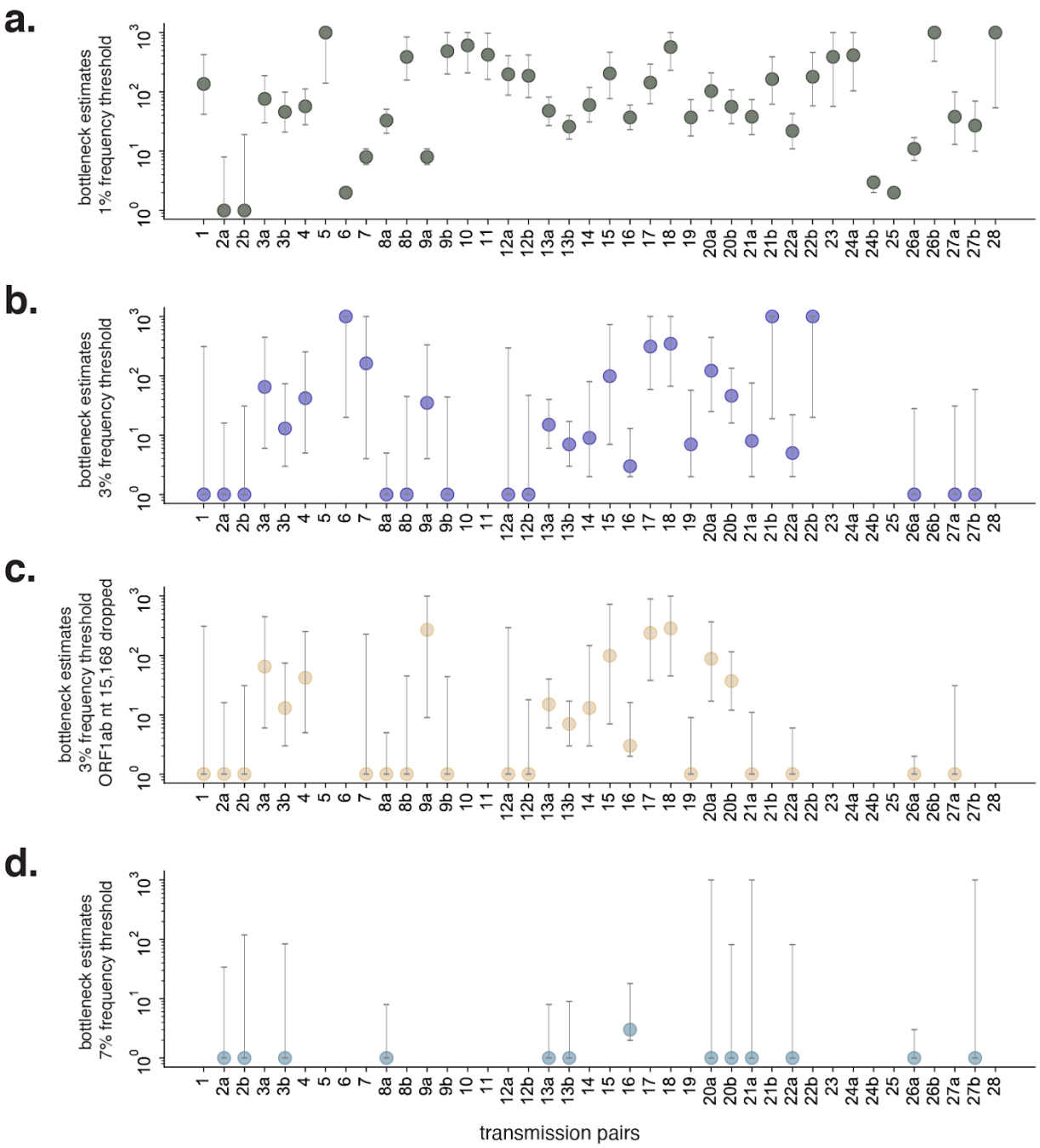
To define whether infections sampled from the same household might be true transmission pairs, we explored the expected number of consensus mutations that should differ between genomes separated by one serial interval. We modeled the probability that 2 consensus genomes will share x mutations as Poisson distributed with lambda equal to the number of mutations expected to accumulate in the SARS-CoV-2 genome over a single serial interval, given a known substitution rate. He et al. estimate a serial interval for SARS-CoV-2 of 5.8 days, with a 95% confidence interval between 4.8-6.8 days [285]. We therefore evaluated serial intervals of 4, 5, 6, 7, and 8 days. For the substitution rate, we use estimates from Duchene et al [153], who estimate a mean substitution rate of 1.10 x 10^{-3} substitutions per site per year, with a 95% credible interval of 7.03 x 10^{-4} and 1.15 x 10^{-3}. We evaluated the probabilities that two consensus genomes differ by 0, 1, 2, 3, and 4 mutations given serial intervals ranging from 4-8, and clock rates at the mean, and upper and lower bounds of the 95% credible interval. For each calculated probability, the

serial interval is represented by color and the substitution rate is shown above each plot. The dotted line represents a probability of 0.05. Given these combinations of values, the vast majority of consensus genomes are expected to differ by 0-2 mutations.



Supplemental Figure 9: Posterior density estimates for regression coefficients

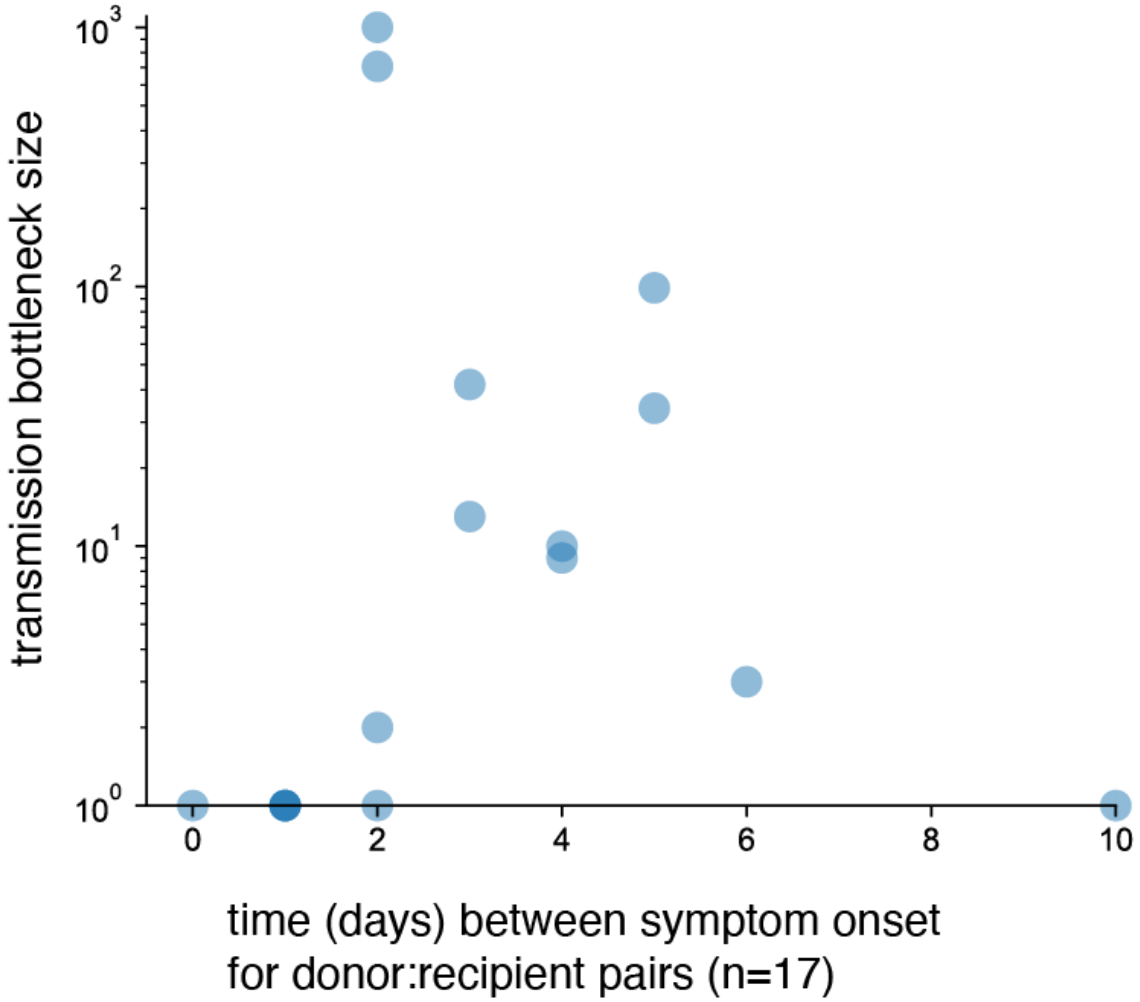
For each regression coefficient evaluated in the combined regression model, the full posterior distribution is shown as a density plot. The posterior distribution of the estimated variance and intercept are also shown.



Supplemental Figure 10: Sensitivity testing of transmission bottleneck estimates

Maximum likelihood estimates for mean transmission bottleneck size in individual donor-recipient pairs using **a.** 1% frequency threshold, **b.** 3% frequency threshold, **c.** excluding site 15,168 as a possible homoplasy with a 3% frequency threshold, and **d.** 7% frequency

threshold. Data are not shown for donor-recipient pairs where no bottleneck estimate could be generated due to lack of variant data. Bidirectional comparisons are indicated with an “a” and “b” following the pair number.



Supplemental Figure 11: Variance in transmission bottleneck size cannot be explained by time between symptom onset in donor:recipient pairs. We plotted transmission bottleneck size on the y-axis against time (days) between symptom onset in 17 donor-recipient pairs on the x-axis for which we had symptom metadata.

Mutation									
Gene	**Reference amino acid**	**Amino acid position**	**Variant amino acid**	**Reference nucleotide**	**Nucleotide position**	**Variant nucleotide**	**rep1 percent**	**rep2 percent**	**Average percent**
orf1ab	Ser	1029	Cys	A	3350	T	0.0406	0.0441	0.04235
orf1ab	Trp	2135	*Stop	G	6669	A	0.0304	0.0347	0.03255
orf1ab	Gly	2863	Val	G	8853	T	0.0103	0.011	0.01065
orf1ab	Thr	2967	Ser	A	9164	T	0.0125	0.0109	0.0117
M	Leu	90	*Stop	T	26791	A	0.1329	0.1368	0.13485
M	Met	90	Val	A	26793	G	0.1313	0.1354	0.13335
M	Trp	92	Arg	T	26796	A	0.131	0.1352	0.1331

Supplemental Table 1. iSNVs detected in replicate sequencing of the synthetic RNA control (Twist-Biosciences).

							Nanopore data	Illumina data			
Strain	**Tube**	**State**	**County**	**Collection Date**	**GISAID Accession**	**Genbank Accession**	**BioProject**	**BioProject**	**N1 Ct value**	**N2 Ct value**	**RLU**
USA/WI-UW-06/2020	A	Wisconsin	Dane County	2020-3-21	EPI_ISL_417200	MT772088	PRJNA614504	PRJNA718341	26.53	27.29	-
USA/WI-UW-07/2020	B	Wisconsin	Dane County	2020-3-21	EPI_ISL_417201	MT772089	PRJNA614504	PRJNA718341	16.28	16.49	-
USA/WI-UW-11/2020	1P	Wisconsin	Dane County	2020-3-15	EPI_ISL_417505	MT706133	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-29/2020	5	Wisconsin	Dane County	2020-3-24	EPI_ISL_421287	MT706150	PRJNA614504	PRJNA718341	16.14	16.05	-
USA/WI-UW-30/2020	6	Wisconsin	Columbia County	2020-3-26	EPI_ISL_421288	MT706151	PRJNA614504	PRJNA718341	24.76	25.41	-
USA/WI-UW-14/2020	6P	Wisconsin	Dane County	2020-3-16	EPI_ISL_417513	MT706136	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-32/2020	8	Wisconsin	Dane County	2020-3-24	EPI_ISL_421290	MT706153	PRJNA614504	PRJNA718341	24.27	24.83	-
USA/WI-UW-34/2020	12	Wisconsin	Dane County	2020-3-26	EPI_ISL_421292	MT706155	PRJNA614504	PRJNA718341	27.81	29.4	-
USA/WI-UW-17/2020	12P	Wisconsin	Dane County	2020-3-13	EPI_ISL_417517	MT706139	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-38/2020	17	Wisconsin	Dane County	2020-3-25	EPI_ISL_421296	MT706159	PRJNA614504	PRJNA718341	27.98	29.23	-
USA/WI-UW-40/2020	19	Wisconsin	Dane County	2020-3-24	EPI_ISL_421298	MT706161	PRJNA614504	PRJNA718341	23.52	24.46	-
USA/WI-UW-39/2020	18	Wisconsin	Dane County	2020-3-22	EPI_ISL_421297	MT706160	PRJNA614504	PRJNA718341	18.2	20.06	-
USA/WI-UW-41/2020	20	Wisconsin	Dane County	2020-3-25	EPI_ISL_421299	MT706162	PRJNA614504	PRJNA718341	24.32	25.31	-
USA/WI-UW-21/2020	24P	Wisconsin	Dane County	2020-3-16	EPI_ISL_417508	MT706142	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-45/2020	25	Wisconsin	Dane County	2020-3-22	EPI_ISL_421303	MT706166	PRJNA614504	PRJNA718341	20.22	20.7	-
USA/WI-UW-22/2020	26P	Wisconsin	Dane County	2020-3-13	EPI_ISL_417514	MT706143	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-48/2020	28	Wisconsin	Dane County	2020-3-25	EPI_ISL_421306	MT706169	PRJNA614504	PRJNA718341	23.02	23.79	-
USA/WI-UW-24/2020	29P	Wisconsin	Dane County	2020-3-15	EPI_ISL_417512	MT706145	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-50/2020	30	Wisconsin	Green County	2020-3-25	EPI_ISL_421308	MT706171	PRJNA614504	PRJNA718341	19.13	19.78	-
USA/WI-UW-51/2020	31	Wisconsin	Dane County	2020-3-20	EPI_ISL_421309	MT706172	PRJNA614504	PRJNA718341	17.11	17.3	-

USA/WI-UW-52/2020	32	Wisconsin	Dane County	2020-3-18	EPI_ISL_421310	MT706173	PRJNA614504	PRJNA718341	15.98	16.57	-
USA/WI-UW-61/2020	44	Wisconsin	Dane County	2020-3-23	EPI_ISL_421319	MT706182	PRJNA614504	PRJNA718341	23.28	24.11	-
USA/WI-UW-63/2020	46	Wisconsin	Dane County	2020-3-24	EPI_ISL_421321	MT706184	PRJNA614504	PRJNA718341	24.54	25.21	-
USA/WI-UW-65/2020	50	Wisconsin	Dane County	2020-3-22	EPI_ISL_421323	MT706186	PRJNA614504	PRJNA718341	15.54	15.33	-
USA/WI-UW-66/2020	51	Wisconsin	Dane County	2020-3-24	EPI_ISL_421324	MT706187	PRJNA614504	PRJNA718341	26.52	27.61	-
USA/WI-UW-67/2020	53	Wisconsin	Dane County	2020-3-25	EPI_ISL_421325	MT706188	PRJNA614504	PRJNA718341	25.62	27.01	-
USA/WI-UW-68/2020	54	Wisconsin	Dane County	2020-3-24	EPI_ISL_421326	MT706189	PRJNA614504	PRJNA718341	15.96	16.13	-
USA/WI-UW-69/2020	55	Wisconsin	Dane County	2020-3-19	EPI_ISL_421327	MT706190	PRJNA614504	PRJNA718341	15.83	16.07	-
USA/WI-UW-70/2020	56	Wisconsin	Dane County	2020-3-19	EPI_ISL_421328	MT706191	PRJNA614504	PRJNA718341	20.12	20.77	-
USA/WI-UW-71/2020	57	Wisconsin	Dane County	2020-3-24	EPI_ISL_421329	MT706192	PRJNA614504	PRJNA718341	18.93	18.64	-
USA/WI-UW-73/2020	60	Wisconsin	Dane County	2020-3-24	EPI_ISL_421331	MT706194	PRJNA614504	PRJNA718341	23.69	24.93	-
USA/WI-UW-74/2020	61	Wisconsin	Dane County	2020-3-20	EPI_ISL_421332	MT706195	PRJNA614504	PRJNA718341	14.19	14.36	-
USA/WI-UW-76/2020	64	Wisconsin	Dane County	2020-3-22	EPI_ISL_421334	MT706197	PRJNA614504	PRJNA718341	17.49	17.59	-
USA/WI-UW-77/2020	65	Wisconsin	Dane County	2020-3-19	EPI_ISL_421335	MT706198	PRJNA614504	PRJNA718341	20.19	20.65	-
USA/WI-UW-84/2020	74	Wisconsin	Dane County	2020-3-24	EPI_ISL_421343	MT706205	PRJNA614504	PRJNA718341	23.12	23.82	-
USA/WI-UW-85/2020	79	Wisconsin	Dane County	2020-4-2	EPI_ISL_425142	MT706206	PRJNA614504	PRJNA718341	24.4	-	-
USA/WI-UW-86/2020	80	Wisconsin	Dane County	2020-4-2	EPI_ISL_425143	MT706207	PRJNA614504	PRJNA718341	22.1	-	-
USA/WI-UW-87/2020	81	Wisconsin	Dane County	2020-4-2	EPI_ISL_425144	MT706208	PRJNA614504	PRJNA718341	22.1	-	-
USA/WI-UW-88/2020	82	Wisconsin	Dane County	2020-4-5	EPI_ISL_425145	MT706209	PRJNA614504	PRJNA718341	25.29	25.93	-
USA/WI-UW-96/2020	94	Wisconsin	Dane County	2020-4-1	EPI_ISL_425153	MT706216	PRJNA614504	PRJNA718341	17.33	18.05	-
USA/WI-UW-97/2020	95	Wisconsin	Dane County	2020-3-30	EPI_ISL_425154	MT706217	PRJNA614504	PRJNA718341	27.3	-	-
USA/WI-UW-99/2020	99	Wisconsin	Dane County	2020-4-2	EPI_ISL_425156	MT706219	PRJNA614504	PRJNA718341	18.8	-	-

USA/WI-UW-110/2020	117	Wisconsin	Dane County	2020-3-31	EPI_ISL_425167	MT706230	PRJNA614504	PRJNA718341	22.2	-	-
USA/WI-UW-111/2020	118	Wisconsin	Dane County	2020-3-31	EPI_ISL_425168	MT706231	PRJNA614504	PRJNA718341	25.61	26.29	-
USA/WI-UW-116/2020	124	Wisconsin	Dane County	2020-3-30	EPI_ISL_425173	MT706236	PRJNA614504	PRJNA718341	31.81	33.31	-
USA/WI-UW-117/2020	125	Wisconsin	Dane County	2020-3-30	EPI_ISL_425174	MT706237	PRJNA614504	PRJNA718341	28.2	-	-
USA/WI-UW-119/2020	128	Wisconsin	Dane County	2020-4-10	EPI_ISL_425176	MT706239	PRJNA614504	PRJNA718341	14.76	14.82	-
USA/WI-UW-120/2020	130	Wisconsin	Dane County	2020-4-13	EPI_ISL_427427	MT706240	PRJNA614504	PRJNA718341	18.3	-	-
USA/WI-UW-255/2020	139	Wisconsin	Dane County	2020-4-2	EPI_ISL_428729	MT706248	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-124/2020	145	Wisconsin	Dane County	2020-4-7	EPI_ISL_427431	MT706252	PRJNA614504	PRJNA718341	17.5	-	-
USA/WI-UW-127/2020	148	Wisconsin	Dane County	2020-4-7	EPI_ISL_427434	MT706255	PRJNA614504	PRJNA718341	19.1	-	-
USA/WI-UW-129/2020	151	Wisconsin	Dane County	2020-4-6	EPI_ISL_427436	MT706257	PRJNA614504	PRJNA718341	30.7	-	-
USA/WI-UW-132/2020	158	Wisconsin	Rock County	2020-4-10	EPI_ISL_427439	MT706260	PRJNA614504	PRJNA718341	17.1	-	-
USA/WI-UW-140/2020	168	Wisconsin	Dane County	2020-4-9	EPI_ISL_427447	MT706268	PRJNA614504	PRJNA718341	21.43	-	-
USA/WI-UW-144/2020	172	Wisconsin	Dane County	2020-4-6	EPI_ISL_427451	MT706272	PRJNA614504	PRJNA718341	23.92	24.31	-
USA/WI-UW-146/2020	175	Wisconsin	Dane County	2020-4-6	EPI_ISL_427453	MT706274	PRJNA614504	PRJNA718341	26.1	-	-
USA/IL-UW-149/2020	182	Illinois	Winnebago County	2020-4-7	EPI_ISL_427456	MT706277	PRJNA614504	PRJNA718341	12.6	-	-
USA/WI-UW-154/2020	188	Wisconsin	Monroe County	2020-4-12	EPI_ISL_427461	MT706282	PRJNA614504	PRJNA718341	21.5	-	-
USA/WI-UW-158/2020	195	Wisconsin	Milwaukee County	2020-3-15	EPI_ISL_428254	MT706286	PRJNA614504	PRJNA718341	23.93	24.33	-
USA/WI-UW-159/2020	196	Wisconsin	Milwaukee County	2020-3-15	EPI_ISL_428255	MT706287	PRJNA614504	PRJNA718341	20.57	21.06	-
USA/WI-UW-160/2020	197	Wisconsin	Milwaukee County	2020-3-15	EPI_ISL_428256	MT706288	PRJNA614504	PRJNA718341	18.6	19.08	-
USA/WI-UW-179/2020	218	Wisconsin	Milwaukee County	2020-3-21	EPI_ISL_428275	MT706307	PRJNA614504	PRJNA718341	19.89	18.58	-
USA/WI-UW-188/2020	229	Wisconsin	Milwaukee County	2020-3-23	EPI_ISL_428284	MT706316	PRJNA614504	PRJNA718341	30.05	29.06	-
USA/WI-UW-201/2020	245	Wisconsin	Ozaukee County	2020-3-25	EPI_ISL_428297	MT706329	PRJNA614504	PRJNA718341	21.19	21.23	-

USA/WI-UW-205/2020	249	Wisconsin	Milwaukee County	2020-3-25	EPI_ISL_428301	MT706333	PRJNA614504	PRJNA718341	27.03	27.23	-
USA/WI-UW-208/2020	252	Wisconsin	Milwaukee County	2020-3-25	EPI_ISL_428304	MT706336	PRJNA614504	PRJNA718341	25.72	24.87	-
USA/WI-UW-211/2020	255	Wisconsin	Milwaukee County	2020-3-25	EPI_ISL_428307	MT706339	PRJNA614504	PRJNA718341	26.72	27.21	-
USA/WI-UW-214/2020	258	Wisconsin	Ozaukee County	2020-3-25	EPI_ISL_428310	MT706342	PRJNA614504	PRJNA718341	21.56	21.52	-
USA/WI-UW-217/2020	261	Wisconsin	Milwaukee County	2020-3-25	EPI_ISL_428313	MT706345	PRJNA614504	PRJNA718341	29.14	29.6	-
USA/WI-UW-223/2020	268	Wisconsin	Milwaukee County	2020-3-26	EPI_ISL_428319	MT706351	PRJNA614504	PRJNA718341	26.67	28.25	-
USA/WI-UW-225/2020	270	Wisconsin	Milwaukee County	2020-3-26	EPI_ISL_428321	MT706353	PRJNA614504	PRJNA718341	18.55	18.91	-
USA/WI-UW-230/2020	275	Wisconsin	Milwaukee County	2020-3-26	EPI_ISL_428326	MT706358	PRJNA614504	PRJNA718341	24.4	25.99	-
USA/WI-UW-231/2020	276	Wisconsin	Milwaukee County	2020-3-26	EPI_ISL_428327	MT706359	PRJNA614504	PRJNA718341	19.42	19.94	-
USA/WI-UW-232/2020	277	Wisconsin	Milwaukee County	2020-3-27	EPI_ISL_428328	MT706360	PRJNA614504	PRJNA718341	19.54	18.33	-
USA/WI-UW-238/2020	283	Wisconsin	Milwaukee County	2020-3-27	EPI_ISL_428334	MT706366	PRJNA614504	PRJNA718341	26.38	25.52	-
USA/WI-UW-242/2020	288	Wisconsin	Milwaukee County	2020-3-28	EPI_ISL_428338	MT706370	PRJNA614504	PRJNA718341	17.93	17.52	-
USA/WI-UW-243/2020	289	Wisconsin	Ozaukee County	2020-3-28	EPI_ISL_428339	MT706371	PRJNA614504	PRJNA718341	25.29	25.04	-
USA/WI-UW-244/2020	290	Wisconsin	Ozaukee County	2020-3-28	EPI_ISL_428340	MT706372	PRJNA614504	PRJNA718341	19.65	19.25	-
USA/WI-UW-246/2020	292	Wisconsin	Milwaukee County	2020-3-28	EPI_ISL_428342	-	PRJNA614504	PRJNA718341	22.16	21.97	-
Tube-302	302	Wisconsin	Dane County	2020-4-16	-	-	-	PRJNA718341	29.5	-	-
Tube-303	303	Wisconsin	Monroe County	2020-4-16	-	-	-	PRJNA718341	-	-	-
Tube-304	304	Wisconsin	Dane County	2020-4-19	-	-	-	PRJNA718341	30.7	-	-
USA/WI-UW-348/2020	310	Wisconsin	Dane County	2020-4-15	EPI_ISL_450702	MT506887	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-351/2020	316	Wisconsin	Dane County	2020-4-14	EPI_ISL_450705	MT506890	PRJNA614504	PRJNA718341	25.51	26.64	-
USA/WI-UW-367/2020	338	Wisconsin	Dane County	2020-4-1	EPI_ISL_450721	MT506906	PRJNA614504	PRJNA718341	-	-	-
USA/WI-UW-273/2020	356	Wisconsin	Milwaukee County	2020-3-31	EPI_ISL_436567	MT706381	PRJNA614504	PRJNA718341	20.29	20.47	-

USA/WI-UW-275/2020	358	Wisconsin	Milwaukee County	2020-4-1	EPI_ISL_436569	MT706383	PRJNA614504	PRJNA718341	18.54	18.14	-
USA/WI-UW-276/2020	359	Wisconsin	Milwaukee County	2020-4-1	EPI_ISL_436570	MT706384	PRJNA614504	PRJNA718341	19.27	19.02	-
USA/WI-UW-277/2020	361	Wisconsin	Milwaukee County	2020-4-2	EPI_ISL_436571	MT706385	PRJNA614504	PRJNA718341	20.28	20.06	-
USA/WI-UW-278/2020	365	Wisconsin	Milwaukee County	2020-4-3	EPI_ISL_436572	MT706386	PRJNA614504	PRJNA718341	16.08	16.12	-
USA/WI-UW-279/2020	366	Wisconsin	Milwaukee County	2020-4-3	EPI_ISL_436573	MT706387	PRJNA614504	PRJNA718341	15.6	15.35	-
USA/WI-UW-282/2020	370	Wisconsin	Milwaukee County	2020-4-6	EPI_ISL_436576	MT706390	PRJNA614504	PRJNA718341	15.39	14.91	-
USA/WI-UW-285/2020	374	Wisconsin	Milwaukee County	2020-4-6	EPI_ISL_436579	MT706393	PRJNA614504	PRJNA718341	27.55	27.17	-
USA/WI-UW-286/2020	376	Wisconsin	Milwaukee County	2020-4-6	EPI_ISL_436580	MT706394	PRJNA614504	PRJNA718341	25.2	25.09	-
USA/WI-UW-287/2020	377	Wisconsin	Milwaukee County	2020-4-6	EPI_ISL_436581	MT706395	PRJNA614504	PRJNA718341	23.43	23.35	-
USA/WI-UW-296/2020	388	Wisconsin	Milwaukee County	2020-4-9	EPI_ISL_436590	MT706404	PRJNA614504	PRJNA718341	16.13	15.52	-
USA/WI-UW-299/2020	391	Wisconsin	Milwaukee County	2020-4-13	EPI_ISL_436593	MT706407	PRJNA614504	PRJNA718341	18.84	18.37	-
USA/WI-UW-302/2020	397	Wisconsin	Milwaukee County	2020-4-13	EPI_ISL_436596	MT706410	PRJNA614504	PRJNA718341	17.6	17.08	-
USA/WI-UW-306/2020	402	Wisconsin	Milwaukee County	2020-4-14	EPI_ISL_436600	MT706414	PRJNA614504	PRJNA718341	30.42	31.22	-
USA/WI-UW-310/2020	408	Wisconsin	Milwaukee County	2020-4-15	EPI_ISL_436604	MT706418	PRJNA614504	PRJNA718341	20.64	19.45	-
USA/WI-UW-315/2020	418	Wisconsin	Milwaukee County	2020-4-17	EPI_ISL_436609	MT706423	PRJNA614504	PRJNA718341	26.33	26.42	-
USA/WI-UW-323/2020	437	Wisconsin	Milwaukee County	2020-4-23	EPI_ISL_436617	MT706430	PRJNA614504	PRJNA718341	26.2	28.6	-
USA/WI-UW-333/2020	453	Wisconsin	Milwaukee County	2020-3-24	EPI_ISL_436627	MT706439	PRJNA614504	PRJNA718341	27.45	26.54	-
USA/WI-UW-334/2020	454	Wisconsin	Milwaukee County	2020-3-24	EPI_ISL_436628	MT706440	PRJNA614504	PRJNA718341	23.24	22.79	-
USA/WI-UW-337/2020	461	Wisconsin	Milwaukee County	2020-3-26	EPI_ISL_436631	MT706443	PRJNA614504	PRJNA718341	26.46	26.5	-
USA/WI-UW-338/2020	462	Wisconsin	Milwaukee County	2020-3-26	EPI_ISL_436632	MT706444	PRJNA614504	PRJNA718341	27.72	27.6	-
USA/WI-UW-340/2020	468	Wisconsin	Milwaukee County	2020-4-1	EPI_ISL_436634	MT706446	PRJNA614504	PRJNA718341	35.15	33.47	-
USA/WI-UW-341/2020	469	Wisconsin	Milwaukee County	2020-4-2	EPI_ISL_436635	MT706447	PRJNA614504	PRJNA718341	23.18	22.53	-

USA/WI-UW-749/2020	1147	Wisconsin	Dane County	2020-6-30	EPI_ISL_491422	-	PRJNA614504	PRJNA718341	-	-	1287
USA/WI-UW-756/2020	1157	Wisconsin	Dane County	2020-7-5	EPI_ISL_495461	MT795871	PRJNA614504	PRJNA718341	-	-	1233
USA/WI-UW-780/2020	1195	Wisconsin	Dane County	2020-7-3	EPI_ISL_495484	MT795891	PRJNA614504	PRJNA718341	-	-	1269
USA/WI-UW-784/2020	1199	Wisconsin	Dane County	2020-7-6	EPI_ISL_495488	-	PRJNA614504	PRJNA718341	-	-	1238
USA/WI-UW-798/2020	1217	Wisconsin	Dane County	2020-7-6	EPI_ISL_495502	-	PRJNA614504	PRJNA718341	-	-	1223
USA/WI-UW-855/2020	1282	Wisconsin	Dane County	2020-7-9	EPI_ISL_509861	MT846545	PRJNA614504	PRJNA718341	-	-	1247
USA/WI-UW-861/2020	1293	Wisconsin	Dane County	2020-7-14	EPI_ISL_509864	MT846550	PRJNA614504	PRJNA718341	-	-	1210
USA/WI-UW-863/2020	1297	Wisconsin	Dane County	2020-7-13	EPI_ISL_509866	MT846552	PRJNA614504	PRJNA718341	-	-	1159
USA/WI-UW-874/2020	1326	Wisconsin	Dane County	2020-7-13	EPI_ISL_509876	MT846562	PRJNA614504	PRJNA718341	-	-	1227
USA/WI-UW-876/2020	1328	Wisconsin	Dane County	2020-7-13	EPI_ISL_509878	MT846564	PRJNA614504	PRJNA718341	-	-	1241
USA/WI-UW-893/2020	1346	Wisconsin	Dane County	2020-7-12	EPI_ISL_509895	MT846581	PRJNA614504	PRJNA718341	-	-	1181
USA/WI-UW-895/2020	1353	Wisconsin	Dane County	2020-7-13	EPI_ISL_509897	MT846583	PRJNA614504	PRJNA718341	-	-	1200
USA/WI-UW-897/2020	1357	Wisconsin	Dane County	2020-7-12	EPI_ISL_509899	MT846585	PRJNA614504	PRJNA718341	-	-	1224
USA/WI-UW-906/2020	1373	Wisconsin	Dane County	2020-7-15	EPI_ISL_509907	MT846593	PRJNA614504	PRJNA718341	-	-	1246
USA/WI-UW-916/2020	1388	Wisconsin	Dane County	2020-7-12	EPI_ISL_509917	MT846603	PRJNA614504	PRJNA718341	-	-	1209
USA/WI-UW-927/2020	1409	Wisconsin	Dane County	2020-7-13	EPI_ISL_509927	MT846614	PRJNA614504	PRJNA718341	-	-	1160
USA/WI-UW-931/2020	1414	Wisconsin	Dane County	2020-7-13	EPI_ISL_509931	MT846618	PRJNA614504	PRJNA718341	-	-	1221
USA/WI-UW-986/2020	1495	Wisconsin	Dane County	2020-7-16	EPI_ISL_509982	MT846672	PRJNA614504	PRJNA718341	-	-	1267
USA/WI-UW-991/2020	1502	Wisconsin	Dane County	2020-7-16	EPI_ISL_509986	MT846677	PRJNA614504	PRJNA718341	-	-	1265
USA/WI-UW-997/2020	1512	Wisconsin	Dane County	2020-7-16	EPI_ISL_509991	MT846683	PRJNA614504	PRJNA718341	-	-	1220
USA/WI-UW-721/2020	1103	Wisconsin	Dane County	2020-7-1	EPI_ISL_491396	-	PRJNA614504	PRJNA718341	-	-	1177
USA/WI-UW-722/2020	1104	Wisconsin	Dane County	2020-6-30	EPI_ISL_491397	-	PRJNA614504	PRJNA718341	-	-	1276
USA/WI-UW-747/2020	1144	Wisconsin	Dane County	2020-7-2	EPI_ISL_491420	MT772518	PRJNA614504	PRJNA718341	-	-	1222

Supplemental Table 2. Sample identifiers and accession numbers. This table includes strain name, tube/filename, state of collection, county of collection, collection date, GISAID accession number, Genbank accession number, as well as Ct values and RLU values where available for each sample included in this study.

name	pool	sequence	length	%gc	tm (use 65)
nCoV-2019_1_LEFT	nCoV-2019_1	ACCAACCAACTTTCGATCTCTTGT	24	41.67	60.69
nCoV-2019_1_RIGHT	nCoV-2019_1	CATCTTTAAGATGTTGACGTGCCTC	25	44	60.45
nCoV-2019_2_LEFT	nCoV-2019_2	CTGTTTTACAGGTTCGCGACGT	22	50	61.67
nCoV-2019_2_RIGHT	nCoV-2019_2	TAAGGATCAGTGCCAAGCTCGT	22	50	61.74
nCoV-2019_3_LEFT	nCoV-2019_1	CGGTAATAAAGGAGCTGGTGGC	22	54.55	61.32
nCoV-2019_3_RIGHT	nCoV-2019_1	AAGGTGTCTGCAATTCATAGCTCT	24	41.67	60.32
nCoV-2019_4_LEFT	nCoV-2019_2	GGTGTATACTGCTGCCGTGAAC	22	54.55	61.56
nCoV-2019_4_RIGHT	nCoV-2019_2	CACAAGTAGTGGCACCTTCTTTAGT	25	44	60.97
nCoV-2019_5_LEFT	nCoV-2019_1	TGGTGAAACTTCATGGCAGACG	22	50	61.39
nCoV-2019_5_RIGHT	nCoV-2019_1	ATTGATGTTGACTTTCTCTTTTTGGAGT	28	32.14	60.17
nCoV-2019_6_LEFT	nCoV-2019_2	GGTGTTGTTGGAGAAGGTTCCG	22	54.55	61.64
nCoV-2019_6_RIGHT	nCoV-2019_2	TAGCGGCCTTCTGTAAAACACG	22	50	61.18
nCoV-2019_7_LEFT	nCoV-2019_1	ATCAGAGGCTGCTCGTGTTGTA	22	50	61.73
nCoV-2019_7_LEFT_alt0	nCoV-2019_1	CATTTGCATCAGAGGCTGCTCG	22	54.55	62.44
nCoV-2019_7_RIGHT	nCoV-2019_1	TGCACAGGTGACAATTTGTCCA	22	45.45	60.95
nCoV-2019_7_RIGHT_alt5	nCoV-2019_1	AGGTGACAATTTGTCCACCGAC	22	50	61.07
nCoV-2019_8_LEFT	nCoV-2019_2	AGAGTTTCTTAGAGACGGTTGGGA	24	45.83	61
nCoV-2019_8_RIGHT	nCoV-2019_2	GCTTCAACAGCTTCACTAGTAGGT	24	45.83	60.56
nCoV-2019_9_LEFT	nCoV-2019_1	TCCCACAGAAGTGTTAACAGAGGA	24	45.83	61.18
nCoV-2019_9_LEFT_alt4	nCoV-2019_1	TTCCCACAGAAGTGTTAACAGAGG	24	45.83	60.44
nCoV-2019_9_RIGHT	nCoV-2019_1	ATGACAGCATCTGCCACAACAC	22	50	61.71
nCoV-2019_9_RIGHT_alt2	nCoV-2019_1	GACAGCATCTGCCACAACACAG	22	54.55	62.26
nCoV-2019_10_LEFT	nCoV-2019_2	TGAGAAGTGCTCTGCCTATACAGT	24	45.83	61.12
nCoV-2019_10_RIGHT	nCoV-2019_2	TCATCTAACCAATCTTCTTCTTGCTCT	27	37.04	60.31
nCoV-2019_11_LEFT	nCoV-2019_1	GGAATTTGGTGCCACTTCTGCT	22	50	61.66
nCoV-2019_11_RIGHT	nCoV-2019_1	TCATCAGATTCAACTTGCATGGCA	24	41.67	61.35
nCoV-2019_12_LEFT	nCoV-2019_2	AAACATGGAGGAGGTGTTGCAG	22	50	61.08
nCoV-2019_12_RIGHT	nCoV-2019_2	TTCACTCTTCATTTCCAAAAAGCTTGA	27	33.33	60.36
nCoV-2019_13_LEFT	nCoV-2019_1	TCGCACAAATGTCTACTTAGCTGT	24	41.67	60.56
nCoV-2019_13_RIGHT	nCoV-2019_1	ACCACAGCAGTTAAAACACCCT	22	45.45	60.36
nCoV-2019_14_LEFT	nCoV-2019_2	CATCCAGATTCTGCCACTCTTGT	23	47.83	60.62
nCoV-2019_14_LEFT_alt4	nCoV-2019_2	TGGCAATCTTCATCCAGATTCTGC	24	45.83	61.47
nCoV-2019_14_RIGHT	nCoV-2019_2	AGTTTCCACACAGACAGGCATT	22	45.45	60.42
nCoV-2019_14_RIGHT_alt2	nCoV-2019_2	TGCGTGTTTCTTCTGCATGTGC	22	50	62.76
nCoV-2019_15_LEFT	nCoV-2019_1	ACAGTGCTTAAAAAGTGTAAAAGTGCC	27	37.04	61.32
nCoV-2019_15_LEFT_alt1	nCoV-2019_1	AGTGCTTAAAAAGTGTAAAAGTGCCT	26	34.62	60.13
nCoV-2019_15_RIGHT	nCoV-2019_1	AACAGAAACTGTAGCTGGCACT	22	45.45	60.16
nCoV-2019_15_RIGHT_alt3	nCoV-2019_1	ACTGTAGCTGGCACTTTGAGAGA	23	47.83	61.57

nCoV-2019_16_LEFT	nCoV-2019_2	AATTTGGAAGAAGCTGCTCGGT	22	45.45	60.82
nCoV-2019_16_RIGHT	nCoV-2019_2	CACAACTTGCGTGTGGAGGTTA	22	50	61.32
nCoV-2019_17_LEFT	nCoV-2019_1	CTTCTTTCTTTGAGAGAAGTGAGGACT	27	40.74	60.69
nCoV-2019_17_RIGHT	nCoV-2019_1	TTTGTTGGAGTGTTAACAATGCAGT	25	36	60.11
nCoV-2019_18_LEFT	nCoV-2019_2	TGGAAATACCCACAAGTTAATGGTTTAAC	29	34.48	60.69
nCoV-2019_18_LEFT_alt2	nCoV-2019_2	ACTTCTATTAAATGGGCAGATAACAACTGT	30	33.33	61.38
nCoV-2019_18_RIGHT	nCoV-2019_2	AGCTTGTTTACCACACGTACAAGG	24	45.83	61.51
nCoV-2019_18_RIGHT_alt1	nCoV-2019_2	GCTTGTTTACCACACGTACAAGG	23	47.83	60.3
nCoV-2019_19_LEFT	nCoV-2019_1	GCTGTTATGTACATGGGCACACT	23	47.83	61.18
nCoV-2019_19_RIGHT	nCoV-2019_1	TGTCCAACTTAGGGTCAATTTCTGT	25	40	60.4
nCoV-2019_20_LEFT	nCoV-2019_2	ACAAAGAAAACAGTTACACAACAACCA	27	33.33	60.68
nCoV-2019_20_RIGHT	nCoV-2019_2	ACGTGGCTTTATTAGTTGCATTGTT	25	36	60.28
nCoV-2019_21_LEFT	nCoV-2019_1	TGGCTATTGATTATAAACACTACACACCC	29	37.93	61.49
nCoV-2019_21_LEFT_alt2	nCoV-2019_1	GGCTATTGATTATAAACACTACACACCCT	29	37.93	61.29
nCoV-2019_21_RIGHT	nCoV-2019_1	TAGATCTGTGTGGCCAACCTCT	22	50	60.83
nCoV-2019_21_RIGHT_alt0	nCoV-2019_1	GATCTGTGTGGCCAACCTCTTC	22	54.55	61.2
nCoV-2019_22_LEFT	nCoV-2019_2	ACTACCGAAGTTGTAGGAGACATTATACT	29	37.93	61.25
nCoV-2019_22_RIGHT	nCoV-2019_2	ACAGTATTCTTTGCTATAGTAGTCGGC	27	40.74	60.73
nCoV-2019_23_LEFT	nCoV-2019_1	ACAACTACTAACATAGTTACACGGTGT	27	37.04	60.26
nCoV-2019_23_RIGHT	nCoV-2019_1	ACCAGTACAGTAGGTTGCAATAGTG	25	44	60.57
nCoV-2019_24_LEFT	nCoV-2019_2	AGGCATGCCTTCTTACTGTACTG	23	47.83	60.37
nCoV-2019_24_RIGHT	nCoV-2019_2	ACATTCTAACCATAGCTGAAATCGGG	26	42.31	61.19
nCoV-2019_25_LEFT	nCoV-2019_1	GCAATTGTTTTTCAGCTATTTTGCAGT	27	33.33	60.73
nCoV-2019_25_RIGHT	nCoV-2019_1	ACTGTAGTGACAAGTCTCTCGCA	23	47.83	61.3
nCoV-2019_26_LEFT	nCoV-2019_2	TTGTGATACATTCTGTGCTGGTAGT	25	40	60.28
nCoV-2019_26_RIGHT	nCoV-2019_2	TCCGCACTATCACCAACATCAG	22	50	60.42
nCoV-2019_27_LEFT	nCoV-2019_1	ACTACAGTCAGCTTATGTGTCAACC	25	44	60.8
nCoV-2019_27_RIGHT	nCoV-2019_1	AATACAAGCACCAAGGTCACGG	22	50	61.13
nCoV-2019_28_LEFT	nCoV-2019_2	ACATAGAAGTTACTGGCGATAGTTGT	26	38.46	60.13
nCoV-2019_28_RIGHT	nCoV-2019_2	TGTTTAGACATGACATGAACAGGTGT	26	38.46	60.91
nCoV-2019_29_LEFT	nCoV-2019_1	ACTTGTGTTCCTTTTTGTTGCTGC	24	41.67	61.39
nCoV-2019_29_RIGHT	nCoV-2019_1	AGTGTACTCTATAAGTTTTGATGGTGTGT	29	34.48	60.69
nCoV-2019_30_LEFT	nCoV-2019_2	GCACAACTAATGGTGACTTTTTGCA	25	40	61.19
nCoV-2019_30_RIGHT	nCoV-2019_2	ACCACTAGTAGATACACAAACACCAG	26	42.31	60.3
nCoV-2019_31_LEFT	nCoV-2019_1	TTCTGAGTACTGTAGGCACGGC	22	54.55	62.03
nCoV-2019_31_RIGHT	nCoV-2019_1	ACAGAATAAACACCAGGTAAGAATGAGT	28	35.71	60.69
nCoV-2019_32_LEFT	nCoV-2019_2	TGGTGAATACAGTCATGTAGTTGCC	25	44	61.09
nCoV-2019_32_RIGHT	nCoV-2019_2	AGCACATCACTACGCAACTTTAGA	24	41.67	60.56
nCoV-2019_33_LEFT	nCoV-2019_1	ACTTTTGAAGAAGCTGCGCTGT	22	45.45	61.58

nCoV-2019_33_RIGHT	nCoV-2019_1	TGGACAGTAAACTACGTCATCAAGC	25	44	61.08
nCoV-2019_34_LEFT	nCoV-2019_2	TCCCATCTGGTAAAGTTGAGGGT	23	47.83	61.02
nCoV-2019_34_RIGHT	nCoV-2019_2	AGTGAAATTGGGCCTCATAGCA	22	45.45	60.03
nCoV-2019_35_LEFT	nCoV-2019_1	TGTTCGCATTCAACCAGGACAG	22	50	61.39
nCoV-2019_35_RIGHT	nCoV-2019_1	ACTTCATAGCCACAAGGTTAAAGTCA	26	38.46	60.69
nCoV-2019_36_LEFT	nCoV-2019_2	TTAGCTTGGTTGTACGCTGCTG	22	50	61.44
nCoV-2019_36_RIGHT	nCoV-2019_2	GAACAAAGACCATTGAGTACTCTGGA	26	42.31	60.74
nCoV-2019_37_LEFT	nCoV-2019_1	ACACACCACTGGTTGTTACTCAC	23	47.83	60.93
nCoV-2019_37_RIGHT	nCoV-2019_1	GTCCACACTCTCCTAGCACCAT	22	54.55	61.48
nCoV-2019_38_LEFT	nCoV-2019_2	ACTGTGTTATGTATGCATCAGCTGT	25	40	60.86
nCoV-2019_38_RIGHT	nCoV-2019_2	CACCAAGAGTCAGTCTAAAGTAGCG	25	48	61.13
nCoV-2019_39_LEFT	nCoV-2019_1	AGTATTGCCCTATTTTCTTCATAACTGGT	29	34.48	61
nCoV-2019_39_RIGHT	nCoV-2019_1	TGTAACTGGACACATTGAGCCC	22	50	60.55
nCoV-2019_40_LEFT	nCoV-2019_2	TGCACATCAGTAGTCTTACTCTCAGT	26	42.31	61.25
nCoV-2019_40_RIGHT	nCoV-2019_2	CATGGCTGCATCACGGTCAAAT	22	50	62.09
nCoV-2019_41_LEFT	nCoV-2019_1	GTTCCCTTCCATCATATGCAGCT	23	47.83	60.75
nCoV-2019_41_RIGHT	nCoV-2019_1	TGGTATGACAACCATTAGTTTGGCT	25	40	60.75
nCoV-2019_42_LEFT	nCoV-2019_2	TGCAAGAGATGGTTGTGTTCCC	22	50	61.08
nCoV-2019_42_RIGHT	nCoV-2019_2	CCTACCTCCCTTTGTTGTGTTGT	23	47.83	60.69
nCoV-2019_43_LEFT	nCoV-2019_1	TACGACAGATGTCTTGTGCTGC	22	50	60.93
nCoV-2019_43_RIGHT	nCoV-2019_1	AGCAGCATCTACAGCAAAAGCA	22	45.45	61.14
nCoV-2019_44_LEFT	nCoV-2019_2	TGCCACAGTACGTCTACAAGCT	22	50	61.66
nCoV-2019_44_LEFT_alt3	nCoV-2019_2	CCACAGTACGTCTACAAGCTGG	22	54.55	60.67
nCoV-2019_44_RIGHT	nCoV-2019_2	AACCTTTCCACATACCGCAGAC	22	50	60.87
nCoV-2019_44_RIGHT_alt0	nCoV-2019_2	CGCAGACGGTACAGACTGTGTT	22	54.55	62.77
nCoV-2019_45_LEFT	nCoV-2019_1	TACCTACAACTTGTGCTAATGACCC	25	44	60.57
nCoV-2019_45_LEFT_alt2	nCoV-2019_1	AGTATGTACAAATACCTACAACTTGTGCT	29	34.48	60.94
nCoV-2019_45_RIGHT	nCoV-2019_1	AAATTGTTTCTTCATGTTGGTAGTTAGAGA	30	30	60.01
nCoV-2019_45_RIGHT_alt7	nCoV-2019_1	TTCATGTTGGTAGTTAGAGAAAGTGTGTC	29	37.93	61.53
nCoV-2019_46_LEFT	nCoV-2019_2	TGTCGCTTCCAAGAAAAGGACG	22	50	61.38
nCoV-2019_46_LEFT_alt1	nCoV-2019_2	CGCTTCCAAGAAAAGGACGAAGA	23	47.83	61.35
nCoV-2019_46_RIGHT	nCoV-2019_2	CACGTTCACCTAAGTTGGCGTA	22	50	60.86
nCoV-2019_46_RIGHT_alt2	nCoV-2019_2	CACGTTCACCTAAGTTGGCGTAT	23	47.83	61.17
nCoV-2019_47_LEFT	nCoV-2019_1	AGGACTGGTATGATTTTGTAGAAAACCC	28	39.29	61.42
nCoV-2019_47_RIGHT	nCoV-2019_1	AATAACGGTCAAAGAGTTTTAACCTCTC	28	35.71	60.06
nCoV-2019_48_LEFT	nCoV-2019_2	TGTTGACACTGACTTAACAAAGCCT	25	40	61.09
nCoV-2019_48_RIGHT	nCoV-2019_2	TAGATTACCAGAAGCAGCGTGC	22	50	60.74
nCoV-2019_49_LEFT	nCoV-2019_1	AGGAATTACTTGTGTATGCTGCTGA	25	40	60.57
nCoV-2019_49_RIGHT	nCoV-2019_1	TGACGATGACTTGGTTAGCATTAATACA	28	35.71	61.05

nCoV-2019_50_LEFT	nCoV-2019_2	GTTGATAAGTACTTTGATTGTTACGATGGT	30	33.33	60.59
nCoV-2019_50_RIGHT	nCoV-2019_2	TAACATGTTGTGCCAACCACCA	22	45.45	60.95
nCoV-2019_51_LEFT	nCoV-2019_1	TCAATAGCCGCCACTAGAGGAG	22	54.55	61.34
nCoV-2019_51_RIGHT	nCoV-2019_1	AGTGCATTAACATTGGCCGTGA	22	45.45	61.14
nCoV-2019_52_LEFT	nCoV-2019_2	CATCAGGAGATGCCACAACTGC	22	54.55	61.83
nCoV-2019_52_RIGHT	nCoV-2019_2	GTTGAGAGCAAAATTCATGAGGTCC	25	44	60.62
nCoV-2019_53_LEFT	nCoV-2019_1	AGCAAAATGTTGGACTGAGACTGA	24	41.67	60.69
nCoV-2019_53_RIGHT	nCoV-2019_1	AGCCTCATAAAACTCAGGTTCCC	23	47.83	60.31
nCoV-2019_54_LEFT	nCoV-2019_2	TGAGTTAACAGGACACATGTTAGACA	26	38.46	60.18
nCoV-2019_54_RIGHT	nCoV-2019_2	AACCAAAAACTTGTCCATTAGCACA	25	36	60.11
nCoV-2019_55_LEFT	nCoV-2019_1	ACTCAACTTTACTTAGGAGGTATGAGCT	28	39.29	61.43
nCoV-2019_55_RIGHT	nCoV-2019_1	GGTGTACTCTCCTATTTGTACTTTACTGT	29	37.93	60.54
nCoV-2019_56_LEFT	nCoV-2019_2	ACCTAGACCACCACTTAACCGA	22	50	60.49
nCoV-2019_56_RIGHT	nCoV-2019_2	ACACTATGCGAGCAGAAGGGTA	22	50	61.21
nCoV-2019_57_LEFT	nCoV-2019_1	ATTCTACACTCCAGGGACCACC	22	54.55	61.16
nCoV-2019_57_RIGHT	nCoV-2019_1	GTAATTGAGCAGGGTCGCCAAT	22	50	61.26
nCoV-2019_58_LEFT	nCoV-2019_2	TGATTTGAGTGTTGTCAATGCCAGA	25	40	61.44
nCoV-2019_58_RIGHT	nCoV-2019_2	CTTTTCTCCAAGCAGGGTTACGT	23	47.83	61.06
nCoV-2019_59_LEFT	nCoV-2019_1	TCACGCATGATGTTTCATCTGCA	23	43.48	61.42
nCoV-2019_59_RIGHT	nCoV-2019_1	AAGAGTCCTGTTACATTTTCAGCTTG	26	38.46	60.02
nCoV-2019_60_LEFT	nCoV-2019_2	TGATAGAGACCTTTATGACAAGTTGCA	27	37.04	60.53
nCoV-2019_60_RIGHT	nCoV-2019_2	GGTACCAACAGCTTCTCTAGTAGC	24	50	60.44
nCoV-2019_61_LEFT	nCoV-2019_1	TGTTTATCACCCGCGAAGAAGC	22	50	61.5
nCoV-2019_61_RIGHT	nCoV-2019_1	ATCACATAGACAACAGGTGCGC	22	50	61.25
nCoV-2019_62_LEFT	nCoV-2019_2	GGCACATGGCTTTGAGTTGACA	22	50	61.91
nCoV-2019_62_RIGHT	nCoV-2019_2	GTTGAACCTTTCTACAAGCCGC	22	50	60.35
nCoV-2019_63_LEFT	nCoV-2019_1	TGTTAAGCGTGTTGACTGGACT	22	45.45	60.16
nCoV-2019_63_RIGHT	nCoV-2019_1	ACAAACTGCCACCATCACAACC	22	50	61.85
nCoV-2019_64_LEFT	nCoV-2019_2	TCGATAGATATCCTGCTAATTCCATTGT	28	35.71	60.11
nCoV-2019_64_RIGHT	nCoV-2019_2	AGTCTTGTAAAAGTGTTCCAGAGGT	25	40	60.1
nCoV-2019_65_LEFT	nCoV-2019_1	GCTGGCTTTAGCTTGTGGGTTT	22	50	61.92
nCoV-2019_65_RIGHT	nCoV-2019_1	TGTCAGTCATAGAACAAACACCAATAGT	28	35.71	60.9
nCoV-2019_66_LEFT	nCoV-2019_2	GGGTGTGGACATTGCTGCTAAT	22	50	61.21
nCoV-2019_66_RIGHT	nCoV-2019_2	TCAATTTCCATTTGACTCCTGGGT	24	41.67	60.45
nCoV-2019_67_LEFT	nCoV-2019_1	GTTGTCCAACAATTACCTGAAACTTACT	28	35.71	60.43
nCoV-2019_67_RIGHT	nCoV-2019_1	CAACCTTAGAAACTACAGATAAATCTTGGG	30	36.67	60.4
nCoV-2019_68_LEFT	nCoV-2019_2	ACAGGTTCATCTAAGTGTGTGTGT	24	41.67	60.14
nCoV-2019_68_RIGHT	nCoV-2019_2	CTCCTTTATCAGAACCAGCACCA	23	47.83	60.31
nCoV-2019_69_LEFT	nCoV-2019_1	TGTCGCAAAATATACTCAACTGTGTCA	27	37.04	61.43

nCoV-2019_69_RIGHT	nCoV-2019_1	TCTTTATAGCCACGGAACCTCCA	23	47.83	61.14
nCoV-2019_70_LEFT	nCoV-2019_2	ACAAAAGAAAATGACTCTAAAGAGGGTTT	29	31.03	60.13
nCoV-2019_70_RIGHT	nCoV-2019_2	TGACCTTCTTTTAAAGACATAACAGCAG	28	35.71	60.27
nCoV-2019_71_LEFT	nCoV-2019_1	ACAAATCCAATTCAGTTGTCTTCCTATTC	29	34.48	60.54
nCoV-2019_71_RIGHT	nCoV-2019_1	TGGAAAAGAAAGGTAAGAACAAGTCCT	27	37.04	60.8
nCoV-2019_72_LEFT	nCoV-2019_2	ACACGTGGTGTTTATTACCCTGAC	24	45.83	61.04
nCoV-2019_72_RIGHT	nCoV-2019_2	ACTCTGAACTCACTTTCCATCCAAC	25	44	60.97
nCoV-2019_73_LEFT	nCoV-2019_1	CAATTTTGTAATGATCCATTTTTGGGTGT	29	31.03	60.29
nCoV-2019_73_RIGHT	nCoV-2019_1	CACCAGCTGTCCAACCTGAAGA	22	54.55	62.45
nCoV-2019_74_LEFT	nCoV-2019_2	ACATCACTAGGTTTCAAACTTTACTTGC	28	35.71	60.68
nCoV-2019_74_RIGHT	nCoV-2019_2	GCAACACAGTTGCTGATTCTCTTC	24	45.83	60.85
nCoV-2019_75_LEFT	nCoV-2019_1	AGAGTCCAACCAACAGAATCTATTGT	26	38.46	60.24
nCoV-2019_75_RIGHT	nCoV-2019_1	ACCACCAACCTTAGAATCAAGATTGT	26	38.46	60.69
nCoV-2019_76_LEFT	nCoV-2019_2	AGGGCAAACTGGAAAGATTGCT	22	45.45	60.76
nCoV-2019_76_LEFT_alt3	nCoV-2019_2	GGGCAAACTGGAAAGATTGCTGA	23	47.83	61.87
nCoV-2019_76_RIGHT	nCoV-2019_2	ACACCTGTGCCTGTTAAACCAT	22	45.45	60.42
nCoV-2019_76_RIGHT_alt0	nCoV-2019_2	ACCTGTGCCTGTTAAACCATTGA	23	43.48	60.69
nCoV-2019_77_LEFT	nCoV-2019_1	CCAGCAACTGTTTGTGGACCTA	22	50	60.75
nCoV-2019_77_RIGHT	nCoV-2019_1	CAGCCCCTATTAAACAGCCTGC	22	54.55	61.59
nCoV-2019_78_LEFT	nCoV-2019_2	CAACTTACTCCTACTTGGCGTGT	23	47.83	60.55
nCoV-2019_78_RIGHT	nCoV-2019_2	TGTGTACAAAAACTGCCATATTGCA	25	36	60.22
nCoV-2019_79_LEFT	nCoV-2019_1	GTGGTGATTCAACTGAATGCAGC	23	47.83	60.92
nCoV-2019_79_RIGHT	nCoV-2019_1	CATTTCATCTGTGAGCAAAGGTGG	24	45.83	60.62
nCoV-2019_80_LEFT	nCoV-2019_2	TTGCCTTGGTGATATTGCTGCT	22	45.45	60.89
nCoV-2019_80_RIGHT	nCoV-2019_2	TGGAGCTAAGTTGTTTAACAAGCG	24	41.67	60.02
nCoV-2019_81_LEFT	nCoV-2019_1	GCACTTGGAAAACTTCAAGATGTGG	25	44	61.24
nCoV-2019_81_RIGHT	nCoV-2019_1	GTGAAGTTCTTTTCTTGTGCAGGG	24	45.83	60.73
nCoV-2019_82_LEFT	nCoV-2019_2	GGGCTATCATCTTATGTCCTTCCCT	25	48	61.52
nCoV-2019_82_RIGHT	nCoV-2019_2	TGCCAGAGATGTCACCTAAATCAA	24	41.67	60.02
nCoV-2019_83_LEFT	nCoV-2019_1	TCCTTTGCAACCTGAATTAGACTCA	25	40	60.46
nCoV-2019_83_RIGHT	nCoV-2019_1	TTTGACTCCTTTGAGCACTGGC	22	50	61.33
nCoV-2019_84_LEFT	nCoV-2019_2	TGCTGTAGTTGTCTCAAGGGCT	22	50	61.61
nCoV-2019_84_RIGHT	nCoV-2019_2	AGGTGTGAGTAAACTGTTACAAACAAC	27	37.04	60.36
nCoV-2019_85_LEFT	nCoV-2019_1	ACTAGCACTCTCCAAGGGTGTT	22	50	61.03
nCoV-2019_85_RIGHT	nCoV-2019_1	ACACAGTCTTTTACTCCAGATTCCC	25	44	60.51
nCoV-2019_86_LEFT	nCoV-2019_2	TCAGGTGATGGCACAACAAGTC	22	50	61.07
nCoV-2019_86_RIGHT	nCoV-2019_2	ACGAAAGCAAGAAAAAGAAGTACGC	25	40	61.01
nCoV-2019_87_LEFT	nCoV-2019_1	CGACTACTAGCGTGCCTTTGTA	22	50	60.16
nCoV-2019_87_RIGHT	nCoV-2019_1	ACTAGGTTCCATTGTTCAAGGAGC	24	45.83	60.81

nCoV-2019_88_LEFT	nCoV-2019_2	CCATGGCAGATTCCAACGGTAC	22	54.55	61.58
nCoV-2019_88_RIGHT	nCoV-2019_2	TGGTCAGAATAGTGCCATGGAGT	23	47.83	61.4
nCoV-2019_89_LEFT	nCoV-2019_1	GTACGCGTTCCATGTGGTCATT	22	50	61.5
nCoV-2019_89_LEFT_alt2	nCoV-2019_1	CGCGTTCCATGTGGTCATTCAA	22	50	62.01
nCoV-2019_89_RIGHT	nCoV-2019_1	ACCTGAAAGTCAACGAGATGAAACA	25	40	60.91
nCoV-2019_89_RIGHT_alt4	nCoV-2019_1	ACGAGATGAAACATCTGTTGTCACT	25	40	60.74
nCoV-2019_90_LEFT	nCoV-2019_2	ACACAGACCATTCCAGTAGCAGT	23	47.83	61.58
nCoV-2019_90_RIGHT	nCoV-2019_2	TGAAATGGTGAATTGCCCTCGT	22	45.45	60.82
nCoV-2019_91_LEFT	nCoV-2019_1	TCACTACCAAGAGTGTGTTAGAGGT	25	44	60.93
nCoV-2019_91_RIGHT	nCoV-2019_1	TTCAAGTGAGAACCAAAAGATAATAAGCA	29	31.03	60.03
nCoV-2019_92_LEFT	nCoV-2019_2	TTTGTGCTTTTTAGCCTTTCTGCT	24	37.5	60.14
nCoV-2019_92_RIGHT	nCoV-2019_2	AGGTTCCTGGCAATTAATTGTAAAAGG	27	37.04	60.53
nCoV-2019_93_LEFT	nCoV-2019_1	TGAGGCTGGTTCTAAATCACCCA	23	47.83	61.59
nCoV-2019_93_RIGHT	nCoV-2019_1	AGGTCTTCCTTGCCATGTTGAG	22	50	60.55
nCoV-2019_94_LEFT	nCoV-2019_2	GGCCCCAAGGTTTACCCAATAA	22	50	60.56
nCoV-2019_94_RIGHT	nCoV-2019_2	TTTGGCAATGTTGTTCCTTGAGG	23	43.48	60.18
nCoV-2019_95_LEFT	nCoV-2019_1	TGAGGGAGCCTTGAATACACCA	22	50	61.1
nCoV-2019_95_RIGHT	nCoV-2019_1	CAGTACGTTTTTGCCGAGGCTT	22	50	61.95
nCoV-2019_96_LEFT	nCoV-2019_2	GCCAACAACAACAAGGCCAAAC	22	50	61.82
nCoV-2019_96_RIGHT	nCoV-2019_2	TAGGCTCTGTTGGTGGGAATGT	22	50	61.36
nCoV-2019_97_LEFT	nCoV-2019_1	TGGATGACAAAGATCCAAATTTCAAAGA	28	32.14	60.22
nCoV-2019_97_RIGHT	nCoV-2019_1	ACACACTGATTAAAGATTGCTATGTGAG	28	35.71	60.17
nCoV-2019_98_LEFT	nCoV-2019_2	AACAATTGCAACAATCCATGAGCA	24	37.5	60.5
nCoV-2019_98_RIGHT	nCoV-2019_2	TTCTCCTAAGAAGCTATTAAAATCACATGG	30	33.33	60.01

Supplemental Table 3. ARTIC v3 primer sequences used to amplify cDNA for library preparation.

	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11	Column 12	Column 13	Column 14
0	tip1	tip2	muts_between_tips	muts	probability_1_serial_interval	tube_IDs	time_between_test	direction_based_on_test_date	time_between_symptoms	direction_based_on_symptoms	#comparisons	pair_number
0	USA/WI-UW-41/2020	USA/WI-UW-48/2020	0	[]	0.749652805157693	20,28	0	28 <-> 20	1	28 <-> 20	2	pair1, pair1a (28 -> 20), pair 1b (20 -> 28)
0	USA/WI-UW-65/2020	USA/WI-UW-32/2020	0	[]	0.749652805157693	50,8	2	8 <-> 50	2	8 <-> 50	2	pair2, pair2a (8 -> 50), pair2b (50 -> 8)
0	USA/WI-UW-69/2020	USA/WI-UW-61/2020	0	[]	0.749652805157693	55,44	4	55 -> 44	3	55 <-> 44	2	pair3, pair3a (55 -> 44), pair3b (44 -> 55)
0	USA/WI-UW-70/2020	USA/WI-UW-67/2020	0	[]	0.749652805157693	56,53	6	56 -> 53	5	56 -> 53	1	pair4
0	USA/WI-UW-74/2020	USA/WI-UW-29/2020	0	[]	0.749652805157693	61,5	4	61 --> 5	4	61 --> 5	1	pair5
0	USA/WI-UW-438/202	USA/WI-UW-432/202	0	[]	0.749652805157693	744,738	2	738 -> 744	2	738 -> 744	2	pair6, pair 6a (738 --> 744), pair 6b (744 --> 738)
0	USA/WI-UW-544/202	USA/WI-UW-551/202	1	['T4917C']	0.21600878707411836	893,884	1	893 <-> 884	4	894 -> 893	1	pair7
0	USA/WI-UW-544/202	USA/WI-UW-575/202	0	[]	0.749652805157693	884,903	1	884 <-> 903	asx	884 <-> 903	2	pair8, pair8a (884 --> 903), pair8b (903 --> 884)
0	USA/WI-UW-551/202	USA/WI-UW-575/202	1	['T4917C']	0.21600878707411836	893,903	0	893 <-> 903	asx	893 <-> 903	2	pair9, pair9a (893 --> 903), pair9b (903 --> 893)
0	USA/WI-UW-546/202	USA/WI-UW-586/202	0	[]	0.749652805157693	887,916	0	887 <-> 916	1	887 <-> 916	2	pair10, pair10a (887 --> 916), pair 10b (916 --> 887)
0	USA/WI-UW-546/202	USA/WI-UW-443/202	0	[]	0.749652805157693	887,749	0	887 <-> 749	0	887 <-> 749	2	pair11, pair 11a (887 --> 749), pair 11b (749 --> 887)
0	USA/WI-UW-586/202	USA/WI-UW-443/202	0	[]	0.749652805157693	916,749	0	916 <-> 749	1	916 <-> 749	2	pair12, pair12a (916 --> 749), pair12b (749 --> 916)
0	USA/WI-UW-577/202	USA/WI-UW-536/202	0	[]	0.749652805157693	906,849	0	906 <-> 849	asx	906 <-> 849	2	pair13, pair13a (906 --> 849), pair13b (849 --> 906)
0	USA/WI-UW-598/202	USA/WI-UW-602/202	0	[]	0.749652805157693	956,962	3	956 --> 962	4	962 --> 956	2	pair14
0	USA/WI-UW-601/202	USA/WI-UW-780/202	0	[]	0.749652805157693	961,1195	8	961 --> 1195	5	961 --> 1195	1	pair15
7	USA/WI-UW-756/202	USA/WI-UW-893/202	0	[]	0.749652805157693	1157,1346	7	1157 --> 1346	6	1157 --> 1346	1	pair16
0	USA/WI-UW-874/202	USA/WI-UW-986/202	0	[]	0.749652805157693	1326,1495	3	1326 --> 1495	asx	1326 <--> 1495	1	pair17
0	USA/WI-UW-874/202	USA/WI-UW-997/202	0	[]	0.749652805157693	1326,1512	3	1326 --> 1512	asx	1326 <--> 1512	1	pair18
0	USA/WI-UW-874/202	USA/WI-UW-991/202	0	[]	0.749652805157693	1326,1502	3	1326 --> 1502	asx	1326 <--> 1502	1	pair19
0	USA/WI-UW-986/202	USA/WI-UW-997/202	0	[]	0.749652805157693	1495,1512	0	1495 <--> 1512	asx	1495 <--> 1512	2	pair20, pair20a (1495 --> 1512), pair20b (1512 --> 1495)
0	USA/WI-UW-986/202	USA/WI-UW-991/202	0	[]	0.749652805157693	1495,1502	0	1495 <--> 1502	asx	1495 <--> 1502	2	pair21, pair21a (1495 --> 1502), pair21b (1502 --> 1495)
0	USA/WI-UW-997/202	USA/WI-UW-991/202	0	[]	0.749652805157693	1512,1502	0	1512 <--> 1502	asx	1512 <--> 1502	2	pair22, pair22a (1512 --> 1502), pair22b (1502 --> 1512)
0	USA/WI-UW-895/202	USA/WI-UW-876/202	0	[]	0.749652805157693	1353,1328	0	1353 <--> 1328	3	1353 --> 1328	1	pair23
0	USA/WI-UW-895/202	USA/WI-UW-863/202	2	['A15942C', 'C25006T']	0.03112093743410756	1353,1297	0	1353 <--> 1297	3	1353 --> 1297	1	pair24
0	USA/WI-UW-876/202	USA/WI-UW-863/202	2	['A15942C', 'C25006T']	0.03112093743410756	1328,1297	0	1297 <--> 1328	0	1297 <--> 1328	1	pair25, pair25a (1297 --> 1328), pair25b (1328 --> 1297)
0	USA/WI-UW-158/202	USA/WI-UW-160/202	0	[]	0.749652805157693	195,197	0	195 <--> 197	NA	195 <--> 197	2	pair26, pair26a (195 --> 197), pair26b (197 --> 195)
0	USA/WI-UW-333/202	USA/WI-UW-334/202	0	[]	0.749652805157693	453,454	0	453 <--> 454	NA	453 <--> 454	2	pair27, pair27a (453 --> 454), pair27b (454 --> 453)
0	USA/WI-UW-119/202	USA/WI-UW-120/202	0	[]	0.749652805157693	128,130	3	128 -> 130	10	130 -> 128	1	pair28

Supplemental Table 4. Household transmission pair metadata including accession numbers, difference in days between symptom onset, difference in days between collection dates, and pair identifier.

Chapter 5:

Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread

Gage K. Moreno[1]*, Katarina M. Braun[2]*, Kasen K. Riemersma[2]*, Michael A. Martin[3,4], Peter J. Halfmann[2,5], Chelsea M Crooks[2], Trent Prall[1], David Baker[1], John J. Baczenas[1,6], Anna S. Heffron[1], Mitchell Ramuta[1], Manjeet Khubbar[7], Andrea M. Weiler[2,6], Molly A. Accola[8], William M Rehrauer[8], Shelby L. O'Connor[1,6], Nasia Safdar[9], Caitlin S. Pepperell[9,10], Trivikram Dasu[7], Sanjib Bhattacharyya[7], Yoshihiro Kawaoka[2,5], Katia Koelle[3], David H. O'Connor[1], Thomas C. Friedrich[2,6]

*These authors contributed equally

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

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

[3]Population Biology, Ecology, and Evolution Graduate Program, Laney Graduate School, Emory University, Atlanta, GA, United States of America

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

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

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

[7]City of Milwaukee Health Department Laboratory, Milwaukee, WI, United States of America

[8]University of Wisconsin School of Medicine and Public Health, Madison, WI, United States of America and the William S. Middleton Memorial Veterans Hospital

[9]Department of Medicine, Division of Infectious Diseases, University of Wisconsin School of Medicine and Public Health, Madison, WI

[10]Department of Medical Microbiology and Immunology, University of Wisconsin-Madison, Madison, WI, United States of America

Nature Communications. 2020 Nov 3;11, 5558 (2020). PMID: 33144575

Abstract

Evidence-based public health approaches that minimize the introduction and spread of new SARS-CoV-2 transmission clusters are urgently needed in the United States and other countries struggling with expanding epidemics. Here we analyze 247 full-genome SARS-CoV-2 sequences from two nearby communities in Wisconsin, USA, and find surprisingly distinct patterns of viral spread. Dane County had the 12th known introduction of SARS-CoV-2 in the United States, but this did not lead to descendant community spread. Instead, the Dane County outbreak was seeded by multiple later introductions, followed by limited community spread. In contrast, relatively few introductions in Milwaukee County led to extensive community spread. We present evidence for reduced

viral spread in both counties following the statewide “Safer-at-Home” order, which went into effect 25 March 2020. Our results suggest patterns of SARS-CoV-2 transmission may vary substantially even in nearby communities. Understanding these local patterns will enable better targeting of public health interventions.

Introduction

The earliest outbreaks of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the United States were seeded by travelers who became infected abroad and initiated chains of community transmission. Several months later, SARS-CoV-2 is now ubiquitous. More than 96% of the 3,144 United States administrative subdivisions (i.e., counties, boroughs, and parishes) have reported at least one SARS-CoV-2 case by June 23, 2020[294]. Movement between administrative subdivisions and states, rather than introduction from abroad, now poses the greatest risk for seeding new clusters of community transmission. However, trends in SARS-CoV-2 caseload and spread are often reported on large geographic scales, such as US states, which obscures the degree to which trends may differ on smaller geographic scales. Finescale spatiotemporal patterns of SARS-CoV-2 spread, particularly below the level of a state or territory, remain poorly defined.

Case counts from diagnostic SARS-CoV-2 testing are used to understand community transmission, but community-level testing may not be widely available and passive surveillance is unlikely to detect asymptomatic or presymptomatic infections. Viral genome sequencing has emerged as a critical tool to overcome these limitations and provides a complementary means of understanding viral transmission dynamics. The

value of this approach was demonstrated during the West African Ebolavirus outbreak in 2014-2016 and again during the emergence of Zika virus in the Americas in 2015-2016[142,143].

The collective global effort to sequence SARS-CoV-2 dwarfs these earlier efforts. As of 28 June 2020, more than 55,000 SARS-CoV-2 sequences collected from over 82 countries have been sequenced and shared publicly on repositories like the Global Initiative on Sharing All Influenza Data (GISAID), enabling real-time phylogenetic analyses encompassing global SARS-CoV-2 diversity [145,295,296]. Patterns of viral sequence variation can also be used to estimate epidemiological parameters, including the total number of infections in a given population and epidemic doubling time, independent of case counts [145–152,297–299]. Here we apply these methods to gain a nuanced view of SARS-CoV-2 transmission within and between regions of the American Upper Midwest.

Dane and Milwaukee Counties are the two most populous counties in the US state of Wisconsin. They are separated by approximately 100 kilometers of rural and suburban communities in Jefferson and Waukesha Counties. An interstate highway that typically carries ~40,000 vehicles a day connects all four of these counties[300]. Madison and Milwaukee are the largest cities in Wisconsin as well as in Dane and Milwaukee Counties, respectively, and are demographically dissimilar [301,302]. On 25 March 2020, the Wisconsin Department of Health Services ordered most individuals to stay at home, closed non-essential businesses, and prohibited most gatherings, an order termed “Safer at Home”

[303–305]. While there were some policies enacted to reduce the viral spread prior to this order [306], the “Safer at Home” order represented the most significant restriction on individuals and businesses. This Executive Order remained in effect until 13 May 2020, when it was struck down by the Wisconsin Supreme Court. From the start of the Executive Order through 21 April 2020, Dane and Milwaukee Counties had the highest documented number of SARS-CoV-2 cases in Wisconsin. Therefore, these two counties provide a “natural experiment” to understand the impact of the “Safer at Home” Executive Order on within- and between-county SARS-CoV-2 transmission in two US counties with distinguishing demographic features.

Here we use our deeply sampled SARS-CoV-2 sequence data to characterize spread in southeastern Wisconsin and, more importantly, illustrate distinct patterns of spatiotemporal SARS-CoV-2 spread in two very nearby communities. We note that this study was not designed prospectively. Moreover, we find that the virus’s basic reproductive number decreased in both counties evaluated during the time in which the “Safer at Home” order was in place, consistent with adoption of physical distancing, use of face coverings, and other related practices [307].

Materials and methods

Sample approvals and sample selection criteria

Sequences for this study were derived from 247 nasopharyngeal (NP) swab samples collected from Dane County between 14 March 2020 through 18 April 2020, and Milwaukee County from 12 March 2020 though 26 April 2020, Wisconsin. Most samples

originated from the University of Wisconsin Hospital and Clinics and the Milwaukee Health Department Laboratories. Available sample metadata, including GISAID accession identifiers, are available in **Supplemental Information 1**.

We worked with residual diagnostic specimens in a biosafety level-3 containment laboratory at the AIDS Vaccine Research Laboratory at the University of Wisconsin – Madison. We obtained a waiver of HIPAA Authorization and were approved to obtain the clinical samples along with a Limited Data Set by the Western Institutional Review Board (WIRB #1-1290953-1). This limited dataset comprised sample collection data and county of collection. Additional sample metadata, e.g. race/ethnicity and income were not shared.

Sample inclusion criteria were retrospectively applied and were threefold: (1) sample had a high-quality consensus sequence (passing GISAID quality control filters), (2) county of origin was Dane county or Milwaukee county, and (3) collection date was on or before our defined endpoint, 18 April 2020.

County-level case data and demographics

We obtained a county-level map of Wisconsin from the State Cartographer's Office (https://www.sco.wisc.edu/maps/wisconsin-outline/). We obtained Wisconsin county-level COVID-19 cumulative case data from the Wisconsin Department of Health Services COVID-19 dashboard (https://data.dhsgis.wi.gov/datasets/covid-19-historical-data-table/, https://cityofmadison.maps.arcgis.com/apps/opsdashboard/index.html#/e22f5ba4f1f94e

0bb0b9529dc82db6a3, and https://county.milwaukee.gov/EN/COVID-19). All Dane and Milwaukee county demographic data came from the Wisconsin Department of Health Services Data & Statistics (https://www.dhs.wisconsin.gov/stats) or the U.S. Census Bureau QuickFacts table (https://www.census.gov/quickfacts/fact/table/).

vRNA isolation for the first confirmed SARS-CoV-2 case in Dane County

The first confirmed case of SARS-CoV-2 in Dane County occurred on 30 January, 2020. This early sample was processed using an early iteration of our SARS-CoV-2 sequencing protocol, as outlined here. All other samples included in this study were processed using the a modified-version of the ARTIC-sequencing protocol, as outlined below. Approximately 140 µL of VTM 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. The sample was treated with TURBO DNase (Thermo Fisher Scientific, Waltham, MA, USA) at 37˚C for 30 min and concentrated to 8µL using the RNA Clean & Concentrator-5 kit (Zymo Research, Irvine, CA, USA). The full protocol for nucleic acid extraction and subsequent cDNA generation is available at https://www.protocols.io/view/sequence-independent-single-primer-amplification-o-bckxiuxn.

Complementary DNA (cDNA) generation for first confirmed SARS-CoV-2 case in Dane County

Complementary DNA (cDNA) was synthesized using a modified Sequence Independent Single Primer Amplification (SISPA) approach described by Kafetzopoulou et al. [308,309]. RNA was reverse-transcribed with SuperScript IV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) using Primer A (5'-GTT TCC CAC TGG AGG ATA-(N_9)-3'). Reaction conditions were as follows: 1µL of primer A was added to 4 µL of sample RNA, heated to 65˚C for 5 minutes, then cooled to 4˚C for 5 minutes. Then 5 µL of a master mix (2 µL 5x RT buffer, 1 µL 10 mM dNTP, 1 µL nuclease free H_2O, 0.5 µL 0.1M DTT, and 0.5 µL SSIV RT) was added and incubated at 42˚C for 10 minutes. For generation of second strand cDNA, 5 µL of Sequenase reaction mix (1 µL 5x Sequenase reaction buffer, 3.85 µL nuclease free H_2O, 0.15 µL Sequenase enzyme) was added to the reaction mix and incubated at 37˚C for 8 minutes. This was followed by the addition of a secondary Sequenase reaction mix (0.45 µl Sequenase Dilution Buffer, 0.15 µl Sequenase Enzyme), and another incubation at 37˚C for 8 minutes. Following incubation, 1µL of RNase H (New England BioLabs, Ipswich, MA, USA) was added to the reaction and incubated at 37˚C for 20 min. Conditions for amplifying Primer-A labeled cDNA were as follows: 5 µL of primer-A labeled cDNA was added to 45 µL of AccuTaq master mix per sample (5 µL AccuTaq LA 10x Buffer, 2.5 µL dNTP mix, 1µL DMSO, 0.5 µL AccuTaq LA DNA Polymerase, 35 µL nuclease free water, and 1 µL Primer B (5′-GTT TCC CAC TGG AGG ATA-3′). Reaction conditions for the PCR were: 98˚C for 30s, 30 cycles of 94˚C for 15 s, 50˚C for 20 s, and 68˚C for 2 min, followed by 68˚C for 10 min.

vRNA isolation

As SARS-CoV-2 cases began to increase in Dane and Milwaukee Counties, we adjusted our sequencing protocol. All samples from 10 March onward were isolated using a Maxwell isolation instrument and subsequently processed using a modified ARTIC tiled amplicon approach [235,310]. Nasopharyngeal swabs received in transport medium (VTM) were briefly centrifuged at 21,130 xg for 30 seconds at room temperature to ensure all residual sample sediments at the bottom of the tube. Viral RNA (vRNA) was extracted from 100 µl of VTM using the Viral Total Nucleic Acid Purification kit (Promega, Madison, WI, USA) on a Maxwell RSC 48 instrument and was eluted in 50 µL of nuclease free H2O.

Complementary DNA (cDNA) generation

Complementary DNA (cDNA) was synthesized using a modified ARTIC Network approach [235,310]. Briefly, vRNA 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 to generate SARS-CoV-2 genomes

A SARS-CoV-2-specific multiplex PCR for Nanopore sequencing was performed, similar to amplicon-based approaches as previously described [235,310]. 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 [235,310]. Following amplification, samples were pooled together before ONT library prep.

Library preparation and sequencing

Amplified PCR product 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 1 ng/µl. A total of 5ng for each sample was then made compatible for deep sequencing using the one-pot native ligation protocol with Oxford Nanopore kit SQK-LSK109 and its Native Barcodes (EXP-NBD104 and EXP-NBD114) [283]. Specifically, samples were end-repaired using the NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs, Ipswich, MA, USA). Samples were then barcoded using 2.5µL of ONT Native Barcodes and the Ultra II End Repair Module. After barcoding, samples were pooled directly into a 1:1 concentration of AMPure XP beads (Beckman Coulter, Brea,

CA, USA) and eluted in 30µL of water. Samples were then tagged with ONT sequencing adaptors according to the modified one-pot ligation protocol [283]. Up to 24 samples were pooled prior to being run on the appropriate flow cell (FLO-MIN106) using the 72hr run script.

Processing raw ONT data

Sequencing data was processed using the ARTIC bioinformatics pipeline (https://github.com/artic-network/artic-ncov2019), with a few modifications. Briefly, we have modified the ARTIC pipeline so that it demultiplexes raw fastq files using qcat as each fastq file is generated by the GridION (https://github.com/nanoporetech/qcat). Once a barcode reaches 100k reads, it will trigger the rest of the ARTIC bioinformatics workflow which will map to the Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 reference (Genbank: MN908947.3) using minimap2. This alignment will then be used to generate consensus sequences and variant calls using medaka (https://github.com/nanoporetech/medaka). The entire ONT analysis pipeline is available at https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin.

Phylogenetic analysis

All 247 available full-length sequences from Dane and Milwaukee County through 26 April 2020 were used for phylogenetic analysis using the tools implemented in Nextstrain custom builds (https://github.com/nextstrain/ncov) [145,284]. Time-resolved and divergence phylogenetic trees were built using the standard Nextstrain tools and scripts [145,284]. We used custom python scripts to filter and clean metadata.

An additional subsampled global phylogeny using all available sequences in GISAID as of 21 June 2020 were input into the Nextstrain pipeline. A custom "Wisconsin" profile was made to create a Wisconsin-centric subsampled build to include representative sequences. To reduce combat bias, we defined representative sequences as 20 sequences from each US state, and 30 sequences from each country, per month per year. This subsampled global build includes 5,377 sequences or roughly 11% of the total sequences in GISAID as of 21 June 2020. All available Wisconsin sequences available on GISAID by 21 June 2020 were incorporated into the subsampled global tree. All of the Wisconsin sequences included in this study are listed in the include.txt to ensure they were represented in the global phylogeny. The scripts and output are available at https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin.

Estimating the number of introductions

To estimate the number of unique introductions into Dane and Milwaukee County we first identified the closest phylogenetic neighbor of each Dane and Milwaukee County sequence in the global (as of 14 June 2020) SARS-CoV-2 phylogenetic tree generated by Dr. Rob Lanfear at the Australian National University. These trees are generated using MAFFT [311] and FastTree [312] and are available at https://github.com/roblanf/sarscov2phylo/. To identify the closest phylogenetic neighbors we first pruned all tips from this tree with ambiguous collection dates (e.g. those given only by month and year as opposed to day, month, and year) and all tips which were excluded from our global alignment using the Nextstrain exclusion criteria (minimum

length of 27000 nucleotides, sequences listed in the “exclude” configuration file, sequences with admin division listed as “USA”) using BioPython. Next, we identified the parent node of each Dane and Milwaukee County tip and then identified the closest phylogenetic neighbor as the other descendant from this node. Aligned neighbor sequences, if not already present, were added to the down-sampled alignment described above, resulting in an alignment of 5,417 sequences. We inferred a maximum likelihood phylogeny of this alignment using IQ-TREE [313] with 1000 Ultrafast bootstrap replicates [314] using the flags -nt 4 -ninit 10 -me 0.05 -bb 1000 -wbtl -czb. The tree was rooted at Wuhan/WH01/2019 and TreeTime {29340210} was used to prune tips from the maximum likelihood tree which did not follow a molecular clock (n_iqd = 4), create a time aligned tree (infer_gtr=True max_iter=2 branch_length_mode=’auto’ resolve_polytomies=False time_marginal=’assign’ vary_rate=0.0004 fixed_clock_rate=0.0008 {2020356}), and infer the geographical locations (Dane County, Milwaukee County, U.S. States, County) of internal nodes (sampling_bias_correction=2.5 to account for undersampling).

To estimate the number of introductions into Dane County and Milwaukee County, this procedure was repeated on 100 of the bootstrap replicate trees. Using each of the 100 bootstrap replicate trees, we identified the earliest node in the path between the root of the tree and each Wisconsin (Dane County, Milwaukee County, and other Wisconsin) tip which was assigned to Wisconsin using the ancestral state reconstruction. Introduction into Wisconsin was assumed to occur mid-way between the earliest Wisconsin node and its parent. The time of introduction was evaluated using the mean estimate as well as the lower and upper limits of the timing for each node. Thus, each bootstrap replicate

contributes three lines to the plots shown in **Fig 3B** and **Fig 3C.** As we do not know whether Wisconsin samples included in the tree from other studies are from Dane or Milwaukee County (or elsewhere in Wisconsin), our estimates for the timing of introduction into each county represent the timing of introduction of that lineage into Wisconsin generally. We conservatively attribute any Dane or Milwaukee County tips or lineages directly descending from a polytomic internal node to a single importation event.

To account for biased sampling within Dane and Milwaukee County we conducted a rarefaction analysis. This was done using the time aligned maximum likelihood tree described above. N (20 to 240, in increments of 20) sequences were randomly sampled from the set of Dane and Milwaukee County sequences and all non-sampled Dane and Milwaukee County sequences were pruned from the tree prior to ancestral state reconstruction and estimation of the number of introductions as described above. Ten replicates for each N were conducted.

Code to replicate this analysis is available at https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin. Results were visualized using Matplotlib [315], Seaborn (https://github.com/mwaskom/seaborn), and Baltic (https://github.com/evogytis/baltic).

Phylodynamic analysis

Bayesian phylogenetic inference and dynamic modelling were performed with BEAST2 software (v2.6.2) [316] and the PhyDyn package (v1.3.6) [150]. The phylodynamic analysis infers SARS-CoV-2 phylogenies of sequences within a region of interest and exogenous

sequences representing the global phylogeny, and uses tree topology to inform a SEIJR compartmental model. For the Bayesian phylogenetic analysis, an HKY substitution model (gamma count=4; K lognormal prior (μ=1, S=1.25)) and a strict molecular clock (uniform prior 0.0005 to 0.005 substitution/site/year) were used. To select the exogenous sequences, a maximum-likelihood global phylogeny was generated with IQTree and randomly downsampled in a time-stratified manner by collection week. Closest cophenetic neighbors for each of the Wisconsin sequences were additionally included, if not present already. Only sequences with coverage of the entire coding region and less than 1% of N base calls were used. For the Dane County analyses, 107 local and 107 exogenous SARS-CoV-2 sequences were used. For the Milwaukee County analyses, 117 local and 129 exogenous SARS-CoV-2 sequences were used.

The SEIJR model dynamics are defined by the following ordinary differential equations:

$$dS/dt = -(\beta I(t) + \beta\tau J(t))\frac{S(t)}{S(t)+E(t)+I(t)+J(t)+R(t)}$$

$$dE/dt = (\beta I(t) + \beta\tau J(t))\frac{S(t)}{S(t)+E(t)+I(t)+J(t)+R(t)} - \gamma_0 E(t)$$

$$dI/dt = \gamma_0(1-p_h)E(t) - \gamma_1 J(t)$$

$$dJ/dt = \gamma_0 p_h E(t) - \gamma_1 J(t)$$

$$dR/dt = \gamma_1(E(t)+J(t))$$

The dynamics of the exogenous compartment is defined by:

$$dY/dt = (\beta_{exog} - \gamma_{exog})Y(t)$$

During phylodynamic model fitting, β, β_{exog}, and α are estimated. Estimated R_0 was derived from β as follows.

$$R_0 = (\beta(1 - p_h) + \beta(\tau p_h))/\gamma_1$$

The epidemic growth rate of the phylodynamic model is governed by the system of differential equations, and can thus be informed by SARS-CoV-2-specific transmission parameters. The SEIJR model includes a "high transmission" compartment (J) that accounts for heterogeneous transmission due to superspreading, an important component of SARS-CoV-2 epidemiology [149,317–319]. Published empirical estimates informed parameterization of superspreading and other epidemiological parameters. The mean duration of latent $(1/\gamma_0)$ and infectious periods $(1/\gamma_1)$ was 3 and 5.5 days, respectively [285]. Likewise, the mean duration of infection for the exogenous compartment $(1/\gamma_{exog})$ was fixed at 8.5 days. To model low, medium, and high transmission heterogeneity, the proportion of infectious individuals in the J compartment (p_h) and their transmission rate multiplier (τ) were set to 0.2 and 16, 0.1 and 36, or 0.05 and 76, respectively. These p_h and τ settings result in 20, 10, or 5% of individuals contributing 80% of total infections. The initial size of the S compartment was fixed at 5×10^5 for Dane County and 9.5×10^5 for Milwaukee County. To account for changes in epidemic dynamics after the Executive Orders, a 25% reduction in importation/exportation of sequences was applied at a 25 March breakpoint, per observed reductions in Google

mobility indices for individuals in Wisconsin [320]. Additionally, the estimated R_0 after 25 March was allowed to vary from the pre-intervention R_0 proportionally by a modifier variable, α.

Each analysis was run in duplicate for at least 3 million states in BEAST2. Parameter traces were visually inspected for adequate mixing and convergence in Tracer (v1.7.1). Log files from duplicate runs were merged with LogCombiner and 10% burn-in applied. Similarly, trajectory files from duplicate runs were merged with an in-house R script and 10% burn-in applied. BEAST2 XML files and scripts for exogenous sequence selection and phylodynamic data analysis/visualization are provided in the GitHub repository listed below.

Data availability

Source data have been deposited in the Sequence Read Archive (SRA) under bioproject PRJNA614504. The consensus genome sequences for national and international genomes are available from GISAID (www.gisaid.org; see Supplementary Table 3). Source data, derived data, analysis pipelines, and figures have been made available for replication of these results at a publicly accessible GitHub repository[321]. For the county-level case data and demographic data presented in **Fig. 1,** we obtained a county-level map of Wisconsin from the State Cartographer's Office (https://www.sco.wisc.edu/maps/wisconsin-outline/). We obtained Wisconsin county-level COVID-19 cumulative case data from the Wisconsin Department of Health Services COVID-19 dashboard (https://data.dhsgis.wi.gov/datasets/covid-19-historical-data-

table/, https://cityofmadison.maps.arcgis.com/apps/opsdashboard/index.html#/e22f5ba4f1f94e0bb0b9529dc82db6a3, and https://county.milwaukee.gov/EN/COVID-19). All Dane and Milwaukee county demographic data came from the Wisconsin Department of Health Services Data & Statistics (https://www.dhs.wisconsin.gov/stats) or the U.S. Census Bureau QuickFacts table (https://www.census.gov/quickfacts/fact/table/). Source data are provided with this paper.

Code availability

Code to replicate this analysis is available at https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin. Code to process sequencing data was made available by ARTIC bioinformatics pipeline (https://github.com/artic-network/artic-ncov2019) and uses Minimap2 v2.17 (https://github.com/lh3/minimap2) and medaka v1.03 (https://github.com/nanoporetech/medaka). Phylogenetic trees were built using Nextstrain tools and clade nomenclature (https://github.com/nextstrain/ncov). The global sub-sampled trees were generated using MAFFT v7.464 (https://mafft.cbrc.jp/alignment/software/) , FastTree v2.1.10 {20224823}, and IQ-TREE v1.5.5 (http://www.iqtree.org), and are available at http://github.com/roblanf/sarscov2phylo/. 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). Bayesian phylogenetic inference and dynamic modelling were performed with BEAST2 software v2.6.2 {30958812} and

the PhyDyn package v1.3.6 {30422979}. Parameter traces were visualized in Tracer v1.7.1 (http://tree.bio.ed.ac.uk/software/tracer/).

Results

SARS-CoV-2 epidemics and community demographics in Dane and Milwaukee Counties

Dane County is home to the 12th reported SARS-CoV-2 case in the United States, detected on 30 January 2020. Subsequent cases were not reported until 9 March 2020. By 26 April, Dane County had 405 confirmed SARS-CoV-2 cases and 19 deaths [322]. Milwaukee County reported its first case on 11 March 2020. By 26 April, Milwaukee County had reported 2,629 confirmed SARS-CoV-2 infections and 126 deaths [323] (**Fig 1B**).

Dane County and Milwaukee County are both located in Southern Wisconsin. Milwaukee County is 127 km east of Dane County, measured from center to center. As of 2015, Dane County had a population of 516,850 at a density of 166 people per km^2 compared to 952,150 at 1,522 per km^2 for Milwaukee County (**Fig 1A**) [301,302].

The majority of individuals living in Dane County are White (81.5%). The next largest group identifies as Hispanic or Latinx (6.3%), followed by Asian (6.0%), Black (5.9%), and American Indian (0.3%) [302]. Milwaukee County is less predominantly White (53.3%) with much larger Black (27.2%) and Hispanic or Latinx (14.5%) populations, followed by Asian (4.3%) and American Indian (0.7%) [301]. The percent of individuals ≥65 years old is similar

in Dane County (13.7%) and Milwaukee County (13.6%), while the percent of individuals under 18 years is slightly lower in Dane County (20.4%) than Milwaukee County (24%). In addition, median income and access to healthcare resources is lower in Milwaukee County than in Dane County [324]. The median individual in Milwaukee County is also more likely to experience poverty and to live with comorbidities such as type II diabetes, hypertension, and obesity (**Table 1**) [324]**.**

Dane and Milwaukee County viruses are genetically distinct

If an outbreak is fueled by community spread following a single introduction, one would expect viral genomes to be close phylogenetic relatives, in which case genetic distances measured in any pairwise comparisons of sequences would be low. To examine this, we generated SARS-CoV-2 consensus sequences using the ARTIC Network protocol [235,310] and defined the population of consensus single-nucleotide variants (SNVs) relative to the initial SARS-CoV-2 Wuhan reference (Genbank: MN908947.3).

In Dane County, we identified 155 distinct SNVs across 122 samples evaluated. These SNVs are evenly distributed throughout the genome, and 92.9% (144/155) are located in open reading frames (ORFs). In Dane County, 52.9% (82/155) of consensus SNVs result in an amino acid change (nonsynonymous) and 40% (62/155) do not (synonymous) (**Fig 2A**).

In Milwaukee County, we identified 148 distinct SNVs across 125 samples evaluated. Among the observed consensus SNVs in Milwaukee County, 63.5% (94/148) are nonsynonymous and 31.8% (47/148) are synonymous (**Fig 2B**).

Mean inter-sequence pairwise SNV distance was 7.65 (std 1.83) and 5.02 (std 3.63) among Dane County and Milwaukee County sequences, respectively (**Fig 2C**). Likewise, we detected an average of 4.4 new SNVs per day (sampling period of 35 days) in Dane County and 3.6 new SNVs per day (sampling period of 41 days) in Milwaukee County. Previous reports suggested SARS-CoV-2 is expected to acquire approximately one fixed SNV every fifteen days following a single introduction [325]. Compared to this benchmark, both Dane County and Milwaukee County have "excess" diversity which can be most parsimoniously explained by multiple introductions of divergent viruses. These patterns are consistent with a greater number of introductions of distinct viruses into Dane County compared to Milwaukee County.

To further analyze genetic differences among viruses in the two locations, we assigned clades using the Nextstrain nomenclature. For example, clade 19B is defined by two mutations at nucleotides 8,782 (ORF1ab S2839S) and 28,144 (Spike L84S) relative to a reference SARS-CoV-2 isolate from Wuhan, China (Genbank: MN908947.3). The majority of Dane County sequences (n = 63 sequences; 51.6%) cluster in the 20A clade (**Fig 3A**). This clade is defined by four variants, at nucleotide positions 241 (upstream of the first open reading frame), 3,037 (ORF1a F924F), 14,408 (ORF1b P314L), and 23,403 (S D614G). A minority (n = 31 sequences; 24.8%) of Milwaukee County sequences also cluster in this clade. In contrast, the 19A clade designation is most common (n = 75 sequences; 60.0%) in sequences from Milwaukee County. This clade is distinguished by a U-to-C variant at nucleotide position 29,711 (downstream of ORF10) (**Fig 3B**).

No onward spread from Dane County index case

The first known SARS-CoV-2 case in Wisconsin was a person who was likely infected during travel to Wuhan, Hubei province, China, where they were exposed to family members with confirmed SARS-CoV-2 infections. The patient reported a sore throat shortly before departing China and returning to the US on 30 January 2020. The person wore a mask during the return flight. Upon arrival in the US, the person immediately presented to an emergency department while still wearing a mask. The person was afebrile and had no respiratory or gastrointestinal signs or symptoms, but began to develop mild respiratory symptoms shortly thereafter. The person's condition remained stable and never required hospitalization or advanced care, with symptoms resolving five days later. The person self-quarantined in an isolated room in a home with a dedicated bathroom for 30 days following symptom onset. During this time, nasopharynx samples repeatedly tested positive for SARS-CoV-2 viral RNA.

Documentation of asymptomatic infections of SARS-CoV-2 has led to concerns about the role of cryptic community transmission in the United States [147,326,327]. However, sequencing in other locations in the United States has revealed early introduction events did not always go on to seed downstream community spread [328]. To determine whether SARS-CoV-2 cases detected in Dane County in March might have been due to undetected spread from the first Wisconsin introduction, we compared the sequence of this early case with local and global SARS-CoV-2 sequence diversity. The first Dane County patient's virus contains an in-frame deletion at nucleotide positions 20,298 -

20,300, in a region that codes for the poly(U)-specific endoribonuclease; the impact of this mutation on viral fitness is unknown [329] (**Supplemental Fig 1**). Notably, this deletion was not detected in any other Dane County sequence, nor in any other sample(s) submitted to GISAID as of 18 April 2020. Moreover, there are no branches originating from the index Dane County case on either the global (Wisconsin sequences plus a subsampled set of global sequences) or local phylogenies (Wisconsin sequences only, maximum likelihood) (**Fig 2C, Fig 3A**). Thus, this early case appears to be an example of successful infection control practices.

SARS-CoV-2 outbreak dynamics differ between Milwaukee and Dane Counties

The independent local phylogenies in Dane and Milwaukee County suggested that these two nearby locations had largely distinct SARS-CoV-2 epidemics through April 2020. To better understand the number of introductions and continued transmission dynamics, we generated a time-resolved sub-sampled global phylogeny incorporating Dane County (red tips) and Milwaukee County (blue tips) sequences alongside representative global SARS-CoV-2 sequences, including all other available Wisconsin sequences (purple tips) (**Fig 4A**). Dane County viruses are distributed throughout the tree, consistent with multiple unique introductions. In contrast, Milwaukee County viruses cluster more closely together, consistent with fewer introductions leading to subsequent community transmission.

To estimate the number of introductions into the state and subsequently each county, we used an ancestral state reconstruction of internal nodes. We performed 100 bootstrap replicates to account for uncertainty in the phylogenetic inference. This yielded an

estimate of 59 [59, 63] (median [95% highest density interval (HDI)]) independent introductions of SARS-CoV-2 into the state of Wisconsin. Of these, 29 [28, 31] led to introductions into Dane county whereas only 21 [19, 21] led to introductions into Milwaukee county (**Fig 4B)**. Surprisingly, only 9 [6, 10] of the introductions into Wisconsin were associated with sequences from both counties. Furthermore, these shared introductions accounted for only 20-30% of the samples from Dane and Milwaukee County present in our dataset. Together, our analyses suggest that transmission between Dane and Milwaukee counties has not been a principal component of viral spread within either region. We find that local transmission in Milwaukee County began earlier, with an introduction event in late January/early February leading to a large number of the Milwaukee County sequences (**Fig 4C**). In comparison, most samples collected from Dane County are associated with multiple introductions in late February/early March (**Fig 4C**)**.** Despite the fact that there were more introductions into Dane County, the reported number of cases was considerably less than in Milwaukee County. This indicates that each introduction into Dane County contributed less to onward viral transmission than in Milwaukee County.

To account for sampling bias on our estimates, we randomly sampled sequences from our set of Dane and Milwaukee County samples (N = 20-240, increments of 20) and pruned all other Dane and Milwaukee samples from the maximum likelihood tree. This was repeated 10 times for each N, creating a set of 120 trees. We repeated the ancestral state reconstruction on each of these trees and re-estimated the number of introductions (**Supplemental Fig 2**). The number of estimated introductions into Dane County

continued to increase with the number of sampled sequences, indicating that these data may be undersampling the true number of circulating viral lineages. In contrast, the number of estimated introductions into Milwaukee County decreases more slowly than Dane County, consistent with a small number of introductions. However, we cannot entirely rule out the possibility that the small number of introductions in Milwaukee County may be an artifact of biased sampling, where the available sequences may only represent a portion of the transmission chains and not a true estimation of the total circulating viral population. Because of this, the true number of introductions is likely to change as more sequences become available in each county. Taken together, these results suggest that patterns of SARS-CoV-2 introduction and spread can differ dramatically in two small administrative regions (here, Dane and Milwaukee Counties), despite their close geographic, economic, and political connections.

Spread of SARS-CoV-2 was reduced following Wisconsin's "Safer at Home" Order

We next used viral sequence data to assess the impact of Wisconsin's "Safer at Home" order on SARS-CoV-2 transmission by estimating the basic reproduction number (R_0). Transmission heterogeneity, or superspreading, is thought to play an important role in SARS-CoV-2 epidemics [149,330,331]. We therefore modeled R_0 before and after the "Safer at Home" order in scenarios in which the level of transmission heterogeneity was low, medium, or high. In both counties, under all three scenarios, R_0 fell by at least 40% after 25 March, indicating that the sequencing data support the observed decline in reported cases. In Dane County, estimated median R_0 was reduced by 40% [4, 74], 49% [13, 79], and 60% [30, 83] under low, medium, and high transmission heterogeneity, respectively.

Similarly, in Milwaukee County, estimated median R_0 was reduced by 68% [50, 83], 71% [56, 86], and 72% [60, 84] under low, medium, and high transmission heterogeneity, respectively.

In Dane County, estimated cumulative incidence was best predicted with the medium transmission heterogeneity model, based on alignment with reported incidence (**Fig 5A**), whereas Milwaukee County's cumulative incidence was best predicted with the model using high transmission heterogeneity (**Fig 5B**). A greater role for superspreading events in Milwaukee versus Dane County could be explained by higher population density, higher poverty rates, and/or worse healthcare access (**Table 1**), all of which may increase contact rates and impede physical distancing efforts [331–335]. Assuming moderate transmission heterogeneity in Dane County, estimated R_0 prior to 25 March was 2.24 [1.86, 2.65] and the median estimated cumulative incidence at the end of the study period (26 April) was 4,546 infections [1,187, 23,709] compared to 405 positive tests. In contrast, assuming high transmission heterogeneity in Milwaukee County, estimated R_0 prior to 25 March was 2.82 [2.48, 3.20] and the median cumulative incidence on 26 April was only 3,008 infections [1,483, 7,508] compared to 2,629 positive tests.

With passive SARS-CoV-2 surveillance efforts in both counties likely missing subclinical and asymptomatic SARS-CoV-2 infections, we expect the true cumulative incidence to be considerably greater than the reported incidence, as has been suggested by others [336]. Indeed, estimated cases were ~10x higher than reported cases in Dane County. Given that there were no substantial differences in the surveillance efforts between

counties, we expected more than the 1.1-fold difference in estimated and reported cases in Milwaukee County. Nearly equivalent estimated and reported cumulative incidence in Milwaukee County could be explained by better detection rates, inaccurate model parameters, and/or biased sampling. However, we likely have representative sampling across Milwaukee County, just on a smaller scale in comparison to Dane County. In an effort generate representative sequence data from Milwaukee County, samples were collected from over 35 zip codes and included samples from known outbreaks, community centers, healthcare facilities, congregate settings (long-term care facilities, jails, correction facilities), meat processing/packing plants as well as households in hotspots where SARS-CoV-2 transmission was detected within Milwaukee County (Supplemental Fig 4). With better detection rates, a greater proportion of actual infections would be reported, but given the similar surveillance efforts between counties we expect detection rates to be comparable. Another possible explanation we cannot rule out is that different model parameters are required to more accurately model Milwaukee County's epidemic. Our testing of three superspreading scenarios demonstrated that the superspreading parameters, at least, may be county-specific. In the case of biased sampling, where the available sequences only represent a portion of transmission chains in the county, our model would only estimate the caseload resulting from a subset of transmission chains in Milwaukee County and would underestimate the county-wide caseload. In support of representative county-wide sampling in Dane, but not Milwaukee County, sequences from 26.4% (107/405) of test-positive cases in Dane County, but only 3.9% (117/3008) of test-positive cases in Milwaukee County were available for phylodynamic modelling [322,323].

Discussion

A clear understanding of SARS-CoV-2 transmission patterns in a given location may permit and promote more effective targeting of public health messaging and infection mitigation efforts. Several studies have described how SARS-CoV-2 entered and began circulating within broad geographic regions, like entire countries (England, Brazil, Austria, Australia) or large and populous US states (Bay Area, NYC) [147–149,298,337–340]. But few studies to date have explored how such patterns may differ on finer geographic and temporal scales, even though many interventions will necessarily be highly localized in scope. Here, we examined differences in SARS-CoV-2 introduction and spread in two nearby counties – Dane County and Milwaukee County – as an example of how such patterns may differ even on small geographic scales. Dane County, Wisconsin had one of the earliest detected cases of SARS-CoV-2 infection in the United States, but this infection did not spark community spread. This is probably due to a combination of good infection control practices by healthcare providers, the patient, and sheer luck. Since March 2020 we find evidence for extensive introductions of SARS-CoV-2 into Dane County, none of which led to large-scale transmission clusters by the end of April 2020. As of 18 August 2020, Dane County has had a cumulative prevalence of 124.9 cases per 100,000 residents. In contrast, Milwaukee County, a larger and more densely populated region ~100km away, has had 2,627 cases per 100,000 residents as of 18 August 2020 [341]. Our findings suggest that Milwaukee County's higher caseload stems from greater levels of community spread descendant from fewer introduction points than in Dane County. Strikingly, we see little evidence for mixing of virus populations between these two closely-linked communities in the same US state.

We used patterns of SARS-CoV-2 diversification in a phylodynamic model to estimate the initial reproductive rate of infections in each county before official physical distancing policies were enacted. In this initial phase of the outbreak, the median estimated R_0 trended lower in Dane County than in Milwaukee County (2.24 vs 2.82). Higher overall population density and a higher average number of individuals residing in one dwelling in Milwaukee County could have contributed to a higher reproductive rate and greater community spread. A potential additional explanation for greater community spread is that the average individual in Milwaukee County, compared to Dane County, has access to fewer financial and healthcare resources and is more likely to experience poverty and to live with comorbid conditions, many of which are also risk factors for testing positive for SARS-CoV-2, the latter of which are also risk factors for severe COVID-19 [301,302,342,343]. Additionally, Milwaukee County is home to a higher proportion of Black and Hispanic or Latinx individuals compared to Dane County. Because of race-based discrimination, people belonging to these groups experience worse health outcomes than White individuals, on average, despite being treated in the same healthcare systems [301,302,344,345]. The social vulnerability index (SVI) is a metric ranging designed to determine how resilient a community is when confronted with external stressors like natural disasters or a pandemic [346]. A higher SVI indicates a community is vulnerable to experiencing worsened outcomes secondary to an external stressor (range of zero to one). All of the factors mentioned above contribute to a higher SVI in Milwaukee County (0.8268) compared to Dane County (0.1974) [346]. While the association between SVI and SARS-CoV-2 incidence is not significant, according to a recent study, the SVI components of

socioeconomic and minority status are both predictors of higher SARS-CoV-2 incidence and case fatality rates [347]. These sub-components are likely to be among the main drivers in the outbreak dynamics between Dane and Milwaukee County.

Like most US states, in late March 2020 Wisconsin enacted a set of physical distancing policies aimed at reducing the spread of SARS-CoV-2. Wisconsin's order, termed "Safer at Home," was enacted 25 March 2020. After this timepoint, the estimated R_0 was reduced by 40% or more in both counties. The sequencing data is consistent with the observed reduction in positive tests, as clusters expanded more slowly and new clusters arose more slowly. Throughout this time, we find that the Dane County and Milwaukee County outbreaks were largely independent of one another. Our data reveal only limited mixing of SARS-CoV-2 genotypes between these geographically-linked communities, supporting the notion that public health policies emphasizing physical distancing effectively reduce transmission between communities. Notably, "Safer at Home" ended abruptly 13 May 2020, when it was overturned by the Wisconsin Supreme Court. Additional sequencing and epidemiological data will be necessary to understand whether virus intermingling between these counties increased after the cessation of the Executive Order.

Viral determinants could also affect differential transmission patterns within and between Dane and Milwaukee Counties. If variants with greater transmission potential exist, then early introductions of such a variant into a community could contribute to greater spread there. Recent reports have suggested that a point mutation in the SARS-CoV-2 spike protein-encoding an aspartate-to-glycine substitution at amino acid residue 614 (D164G)

may enhance transmissibility [348–350]. This mutation confers increased infectivity of pseudotyped murine retroviruses in ACE2-expressing HEK293T cells [350] and has been proposed to be increasing in global prevalence, perhaps under natural selection [351]. Importantly, however, the rise in D614G frequency could also be due to founder effects, as viruses bearing the glycine allele may have been the first to establish local transmission in Europe. D614G is one of the mutations defining the 20A clade; these viruses remain dominant in Europe [328], so introductions from Europe into the United States, including into Dane County, predominantly carry D614G. In comparison, in Milwaukee County, the vast majority of viruses have an aspartic acid residue at this site despite much higher levels of community transmission early in the pandemic. This observation does not necessarily indicate that D614G does not impact viral transmissibility; its role may be muted by other determinants of transmission, including demographic and socioeconomic factors. Viruses encoding D614G may displace 614D variants over time in regions like Milwaukee County, where 614D viruses have sustained community spread.

There are some important caveats to this study. Of the total reported positives in each county during the study period, high-quality sequences were available for 30% of test-positive cases in Dane County, but only 5% of test-positive cases in Milwaukee County [322,323]. Despite the deep sampling of SARS-CoV-2 sequences in Wisconsin relative to other regions in the US, even greater targeted sequencing efforts may be required to fully capture the sequence heterogeneity conferred by multiple introduction events and variable superspreading dynamics. It is possible additional sequencing in Milwaukee

County would uncover additional viral lineages, or that the 5% of cases we sequenced do not fully represent the diversity of viruses found throughout the county, skewing our observations. However, in analyzing sample metadata we find no evidence that particular locations within Milwaukee County were dramatically over- or under-sampled relative to their known SARS-CoV-2 prevalence (**Supplemental Fig. 4**). Another potential explanation is that Milwaukee County was under-testing relative to Dane County. Throughout the period analyzed here, the percentage of SARS-CoV-2 tests returning positive in Milwaukee County was ~20%, compared to only ~5% in Dane County [322,323], indicating that a higher proportion of infections might have been missed by testing in Milwaukee County relative to Dane County. As we are only able to sequence test-positive samples, it is possible that under-testing in Milwaukee County limited our ability to capture a complete representation of their epidemic. However, we have no reason to suspect Milwaukee testing regimes were biased toward or against subsets of the overall population. During this time, there were three free community testing sites (supported by the Wisconsin National Guard) and several additional community testing and shelter sites located throughout the city. COVID-19 testing criteria for Milwaukee public health laboratories targeted all sectors of the population per Wisconsin Department of Health Services guidelines [352]. In sum, we have taken steps to minimize systematic sampling bias in Milwaukee County in this study, but we cannot entirely exclude the possibility that the samples available to us for sequencing did not fully capture the diversity of SARS-CoV-2 circulating in Milwaukee County during the study period.

It is also possible that other sequences from these counties relevant to our analyses were collected by other groups. As of 21 June 2020, there were 477 Wisconsin sequences available, but only 351 of these had geolocation information resolved to the county level. Some of the remaining 126 sequences likely originated from Dane County or Milwaukee County, but we cannot include these sequences in our analysis given their geolocation data resolved only to the state level. Currently there is no clearly stated national-level guidance for metadata to be associated with pathogen sequences. Dates and geographic locations with greater than state-level resolution are required to track the emergence and spread of novel pathogens like SARS-CoV-2. Explicit regulatory guidance from the United States enabling the disclosure of sequencing data with county-level geolocation data and sampling dates would enable other institutions to harmonize reporting of viral sequences and improve subsequent studies comparing viral sequences from different locations, as described previously [353]. Such reporting may be especially important for identifying disparities in viral transmission due to socioeconomic vulnerabilities in specific counties that would otherwise be masked using state-level data reporting.

Few previous studies have carefully evaluated patterns of SARS-CoV-2 introduction and spread below the level of US regions or states. Yet, with little US federal guidance, the majority burden of organizing and implementing anti-SARS-CoV-2 public health campaigns has fallen to US cities and counties. Tailoring public health messaging and intervention strategies to specific communities and locations can enhance their efficacy and durability. Our study exemplifies how viral sequence dynamics can enhance our understanding of the finescale patterns of virus introduction and spread, revealing

differences in transmission patterns between even nearby communities that could inform the design of targeted interventions. For example, our data suggest Dane County, which had a large number of introductions but relatively little sustained community spread during the study period, might have benefited most from travel restrictions and/or quarantine for people entering the community. In contrast, our data suggest that community spread was established early in the study period in Milwaukee County, so interventions targeted at interrupting transmission clusters might have had the most impact. These could include limiting indoor community gatherings, targeting messaging or social marketing campaigns promoting masking and other physical distancing measures, and improving access to economic and healthcare resources – not only direct access to care, but also paid leave and other support systems for workers who are ill. To this end, continued efforts to sequence SARS-CoV-2 viruses across multiple spatio-temporal scales remain critical for tracking viral transmission dynamics within and between communities and for guiding "precision medicine" public health interventions to suppress future SARS-CoV-2 outbreaks.

Acknowledgements

We gratefully acknowledge Dr. Trevor Bedford and the entire Nextstrain team for making Nextstrain phylogenetic tools publicly available and for their commitment to tracking the global spread of SARS-CoV-2. We also acknowledge the GISAID team and all its data contributors for maintaining the largest public repository of SARS-CoV-2 sequence- and metadata. We are indebted to Dr. Louise Moncla for her careful reading of and insightful comments on this manuscript. We thank Rachel Mukai, MPH, CHES, Interim Director of

Data and Evaluation and Epidemiology—Clinical Services, City of Milwaukee Health Department, for creating Supplementary Fig. 4. 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 T.C.F. and D.H.O. K.M.B. is supported by F30 AI145182-01A1 from the National Institutes of Allergy and Infectious Disease. G.K.M. is supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359).

Figures, tables, and supplemental material

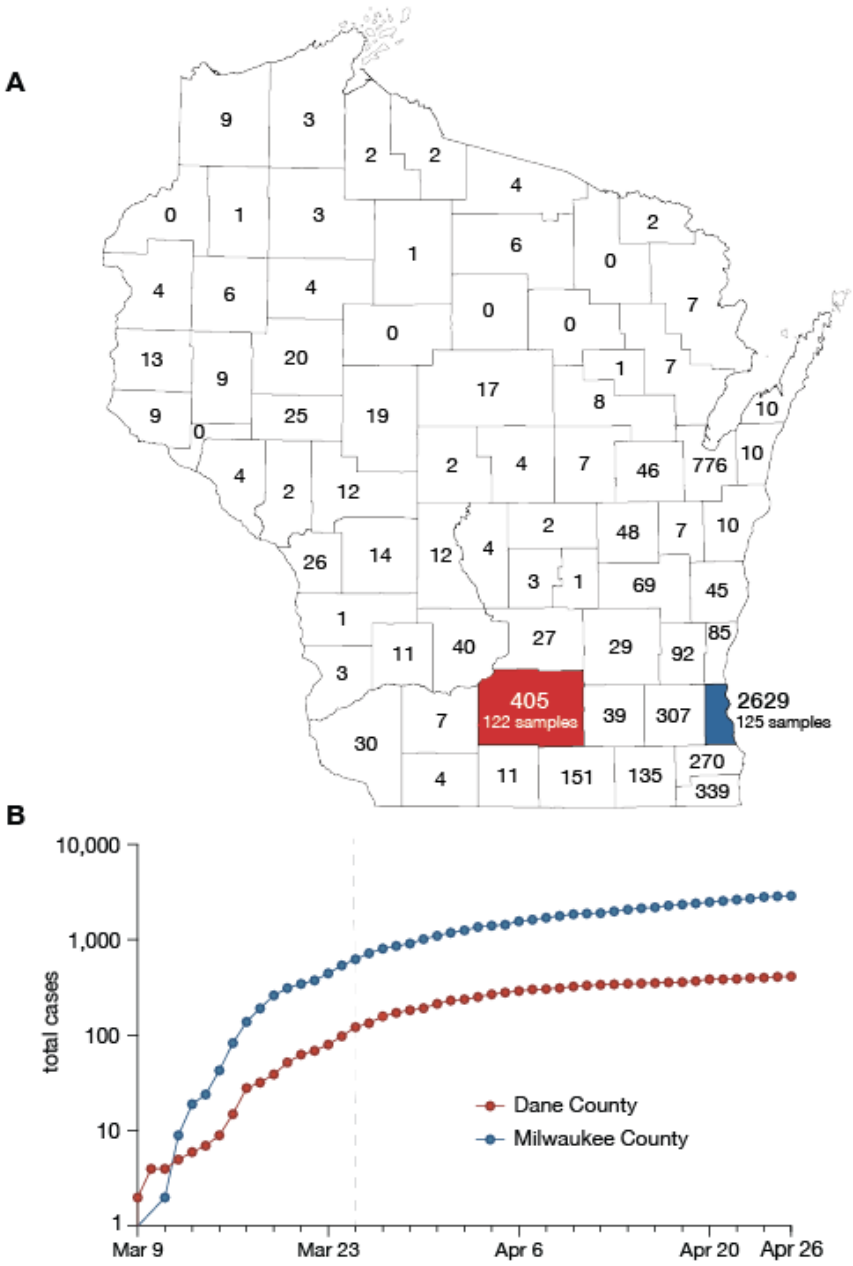


Figure 1.

Demography and epidemiology of SARS-CoV-2 in southern Wisconsin. A) A map of Wisconsin highlighting Dane County (red) and Milwaukee County (blue). Cumulative case counts through 26 April 2020 are reported within each county border. B) Cumulative

SARS-CoV-2 cases in Dane County (red) and Milwaukee County (blue) from 9 March through 26 April. The vertical dashed line indicates the start date of Wisconsin's "Safer at Home" order, which went into effect 25 March 2020.

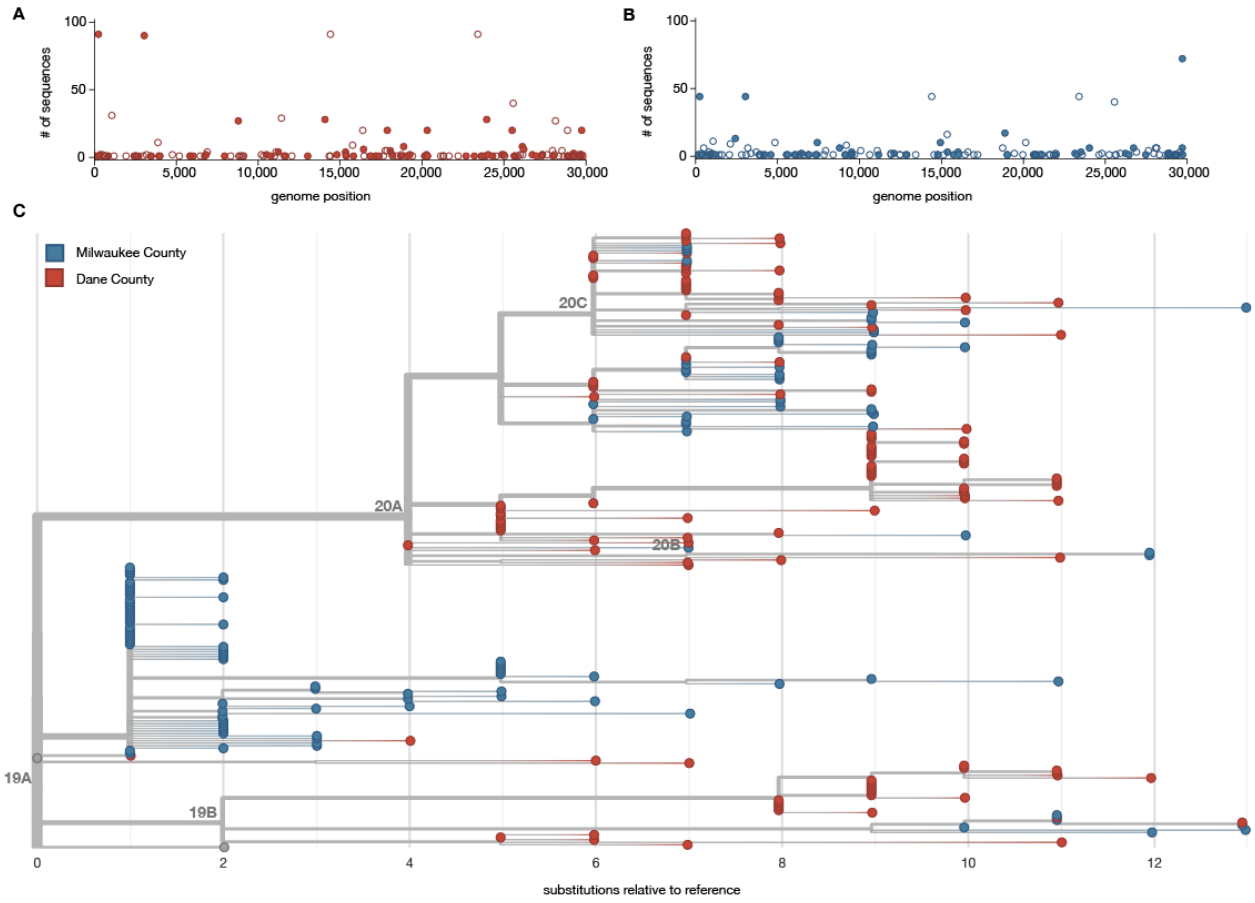


Figure 2.

Characterizing consensus-level variants and sequence divergence among Dane and Milwaukee County sequences. SNVs are annotated relative to the initial Wuhan SARS-CoV-2 reference (Genbank: MN908947.3). A) Frequency of consensus SNVs among the Dane County sequences, represented in red. B) Frequency of consensus SNVs among the Milwaukee County sequences, represented in blue. Open symbols denote

synonymous or intergenic SNVs and closed symbols denote nonsynonymous SNVs. C) A divergence-based phylogenetic tree built using Nextstrain tools for the 122 Dane County (red) and 125 Milwaukee County (blue) sequences. Wisconsin samples are rooted against Wuhan-Hu-1/2019 and Wuhan/WH01/2019.

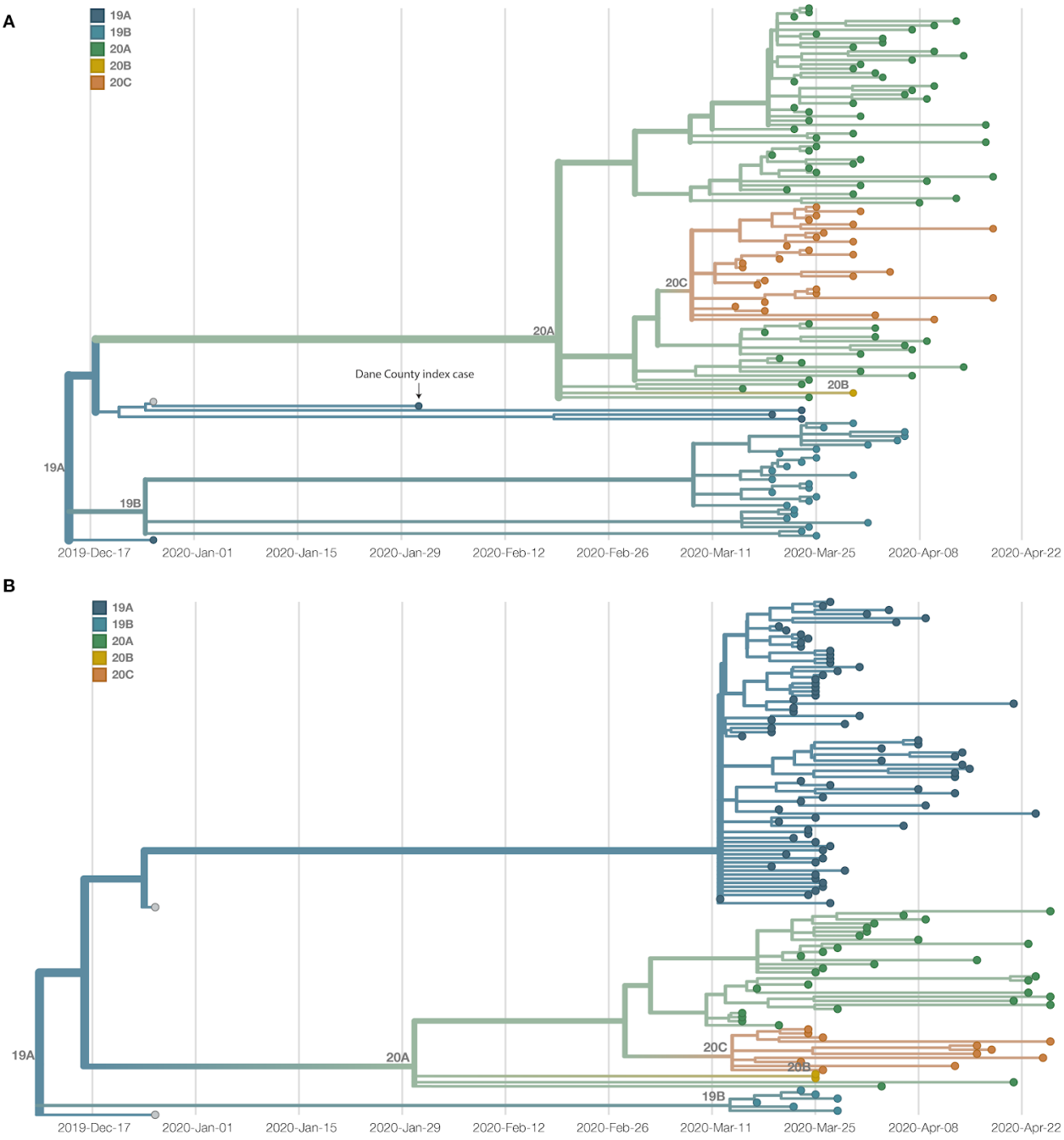


Figure 3.

Dane and Milwaukee County outbreaks are defined by genetically distinct viruses. A) A time-resolved phylogenetic tree built using Nextstrain tools for 122 samples collected in Dane County. B) A time-resolved phylogenetic tree for 125 samples collected in Milwaukee County. Clade is denoted by color. Both phylogenies include Wuhan

sequences (Wuhan-Hu-1/2019 and Wuhan/WH01/2019, denoted in grey) to more effectively time-align each tree.

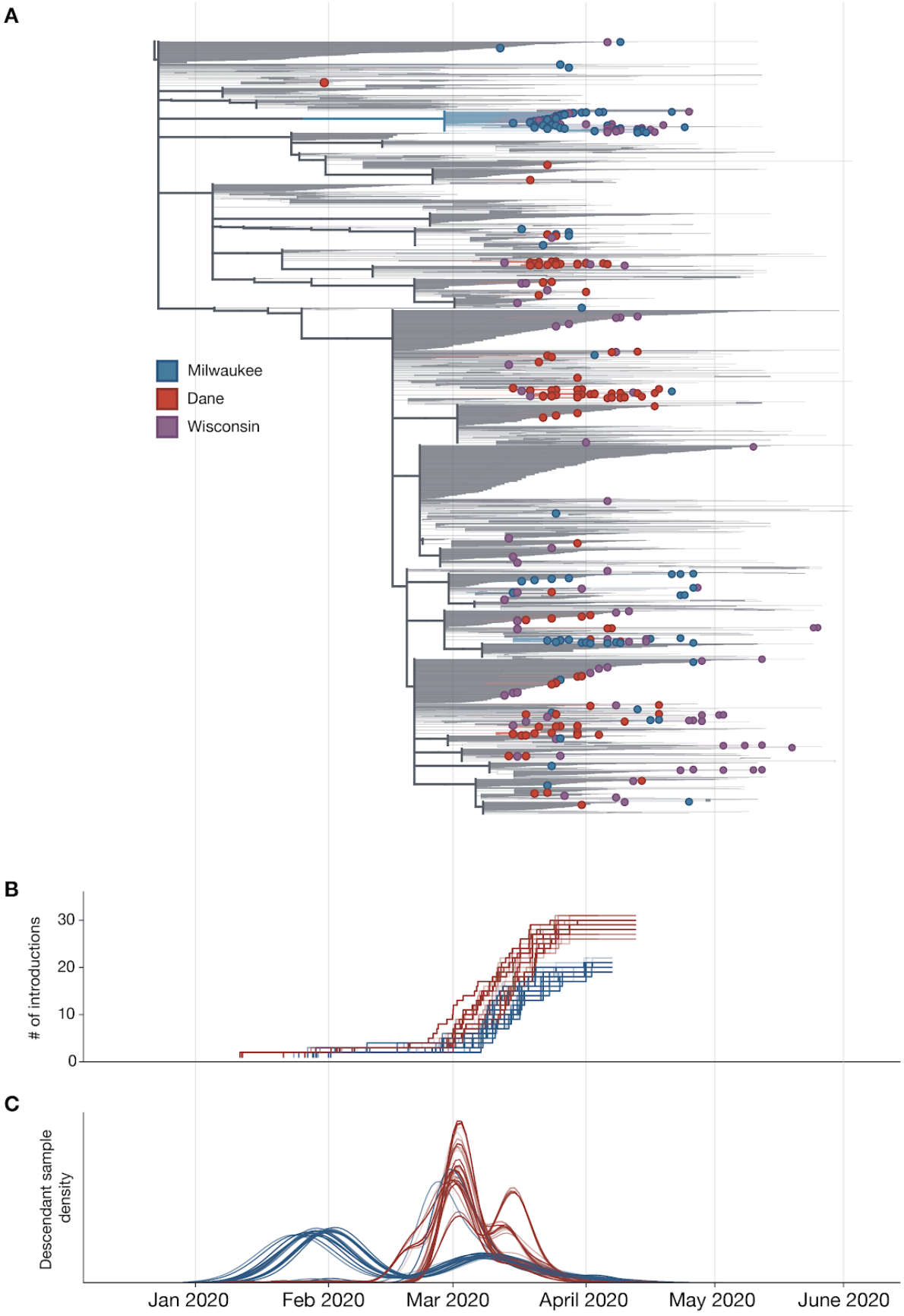

A
Milwaukee
Dane
Wisconsin
B
# of introductions
30
20
10
0
C
Descendant sample density
Jan 2020
Feb 2020
Mar 2020
April 2020
May 2020
June 2020

Figure 4.

Estimate of the number of introduction events into Milwaukee and Dane County and their relative contribution to downstream epidemic dynamics. A) Maximum likelihood (ML) time-resolved tree with subsampled global sequences and closest phylogenetic neighbors relatives included (grey branches). Sequences from Dane and Milwaukee Counties are highlighted in red and blue, respectively. Sequences with geolocation information available to the state level, or that are located outside of Dane and Milwaukee Counties (i.e. La Crosse) are shown in purple. B) Estimated cumulative number of introduction events into each county. C) Gaussian Kernel Density Estimate plots showing the estimated timing of each introduction event (3 curves per replicate: mean and 90% confidence intervals) into Dane County (red) or Milwaukee County (blue). The relative number of samples from each region attributable to an introduction event is represented on the y-axis. Curves are normalized to a cumulative density of one; therefore, y-axis scale is not shown.

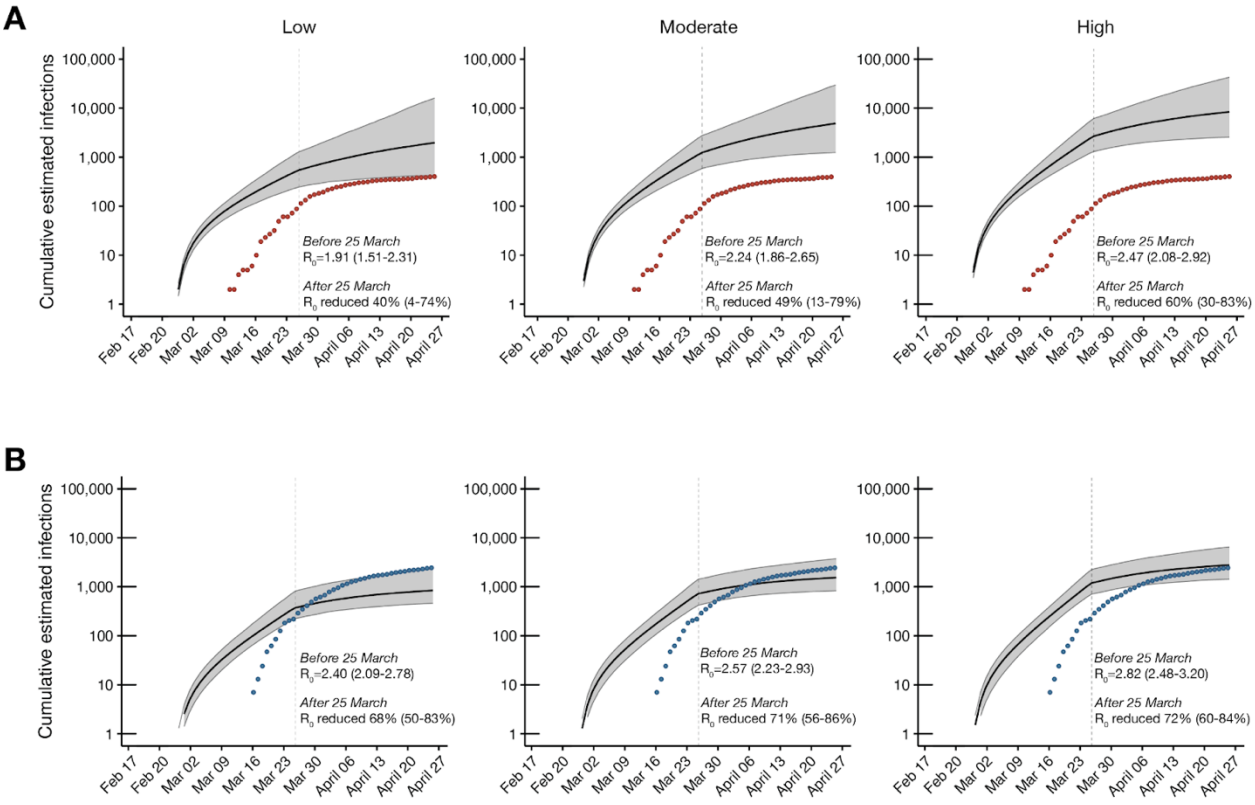


Figure 5.

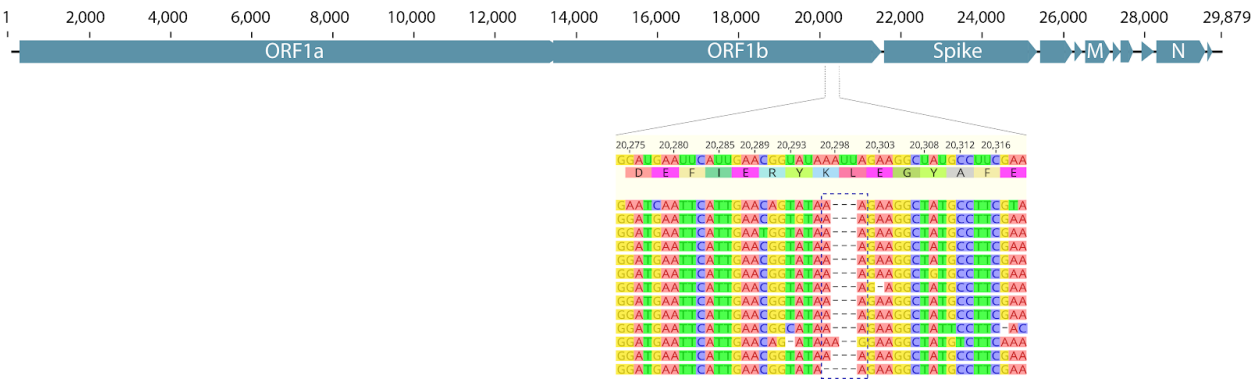
Phylodynamic modelling of regional outbreaks informs regional outbreak dynamics before and after government interventions. Bayesian phylodynamic modelling of cumulative incidence up to 26 April for outbreaks in A) Dane County and B) Milwaukee County under low (left), medium (center), and high (right) transmission heterogeneity conditions. Model parameters for low, medium, and high transmission heterogeneity were fixed such that 20, 10, and 5% of superspreading events contribute 80% of cumulative infections, respectively. Median cumulative incidence (solid black line) is bound by the 95% confidence intervals (CI; gray ribbon). Dots represent reported cumulative positive tests in Dane County (red) and Milwaukee County (blue). Estimated median reproductive numbers (R_0) with 95% HDI are listed for the period before the Wisconsin “Safer at Home”

order was issued on 25 March 2020. Percent reduction in R_0 with 95% HDI is provided for the period after 25 March 2020.

County-level demographic data	Dane	Milwaukee
Population size (2015)	516,850	952,150
Population per square mile (2015)	430	3942
Average number of persons per dwelling (2014-2018)	2.35	2.44
Age (2014-2018):		
% of population under 5	5.6	6.9
% of population under 18	20.4	24
% of population over 65	13.7	13.6
Race/ethnicity (2015):		
White	81.5%	53.3%
African American	5.9%	27.2%
American Indian	0.3%	0.7%
Hispanic	6.3%	14.5%
Asian	6.0%	4.3%
Median income (2015)	$65,416	$45,905
% of population that is uninsured, under 65 (2014-2018)	4.9%	8.2%
Poverty estimate, all ages (2015)	11.2%	20.3%
% of population reported overweight or obese (2012-2016)	54.3% - 58.5%	64.7% - 69%
% of adults reporting diagnosed diabetes (2012-2016)	4.2% - 6.8%	8.6% - 9.8%

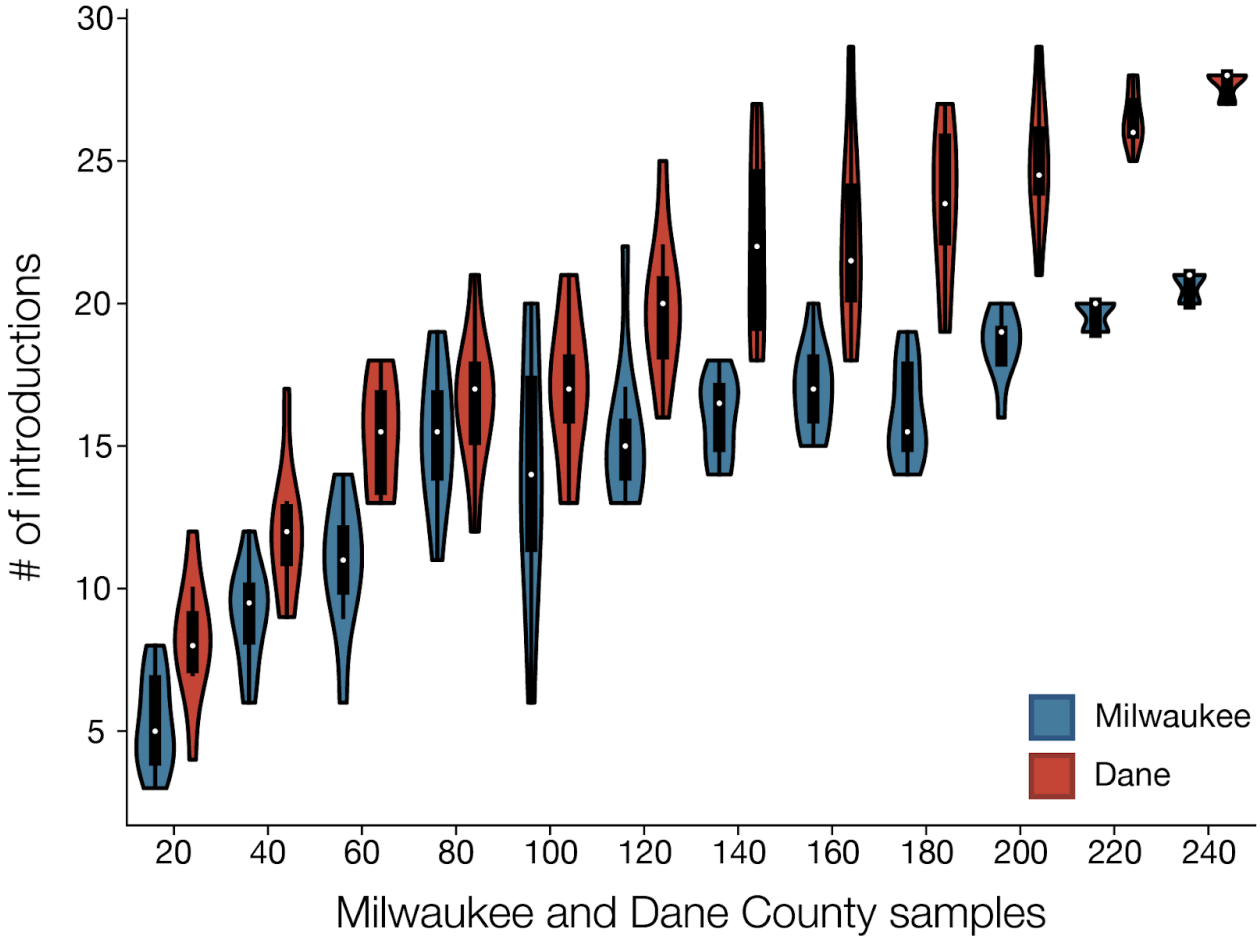
Table 1.

County level demographics for Dane and Milwaukee County.



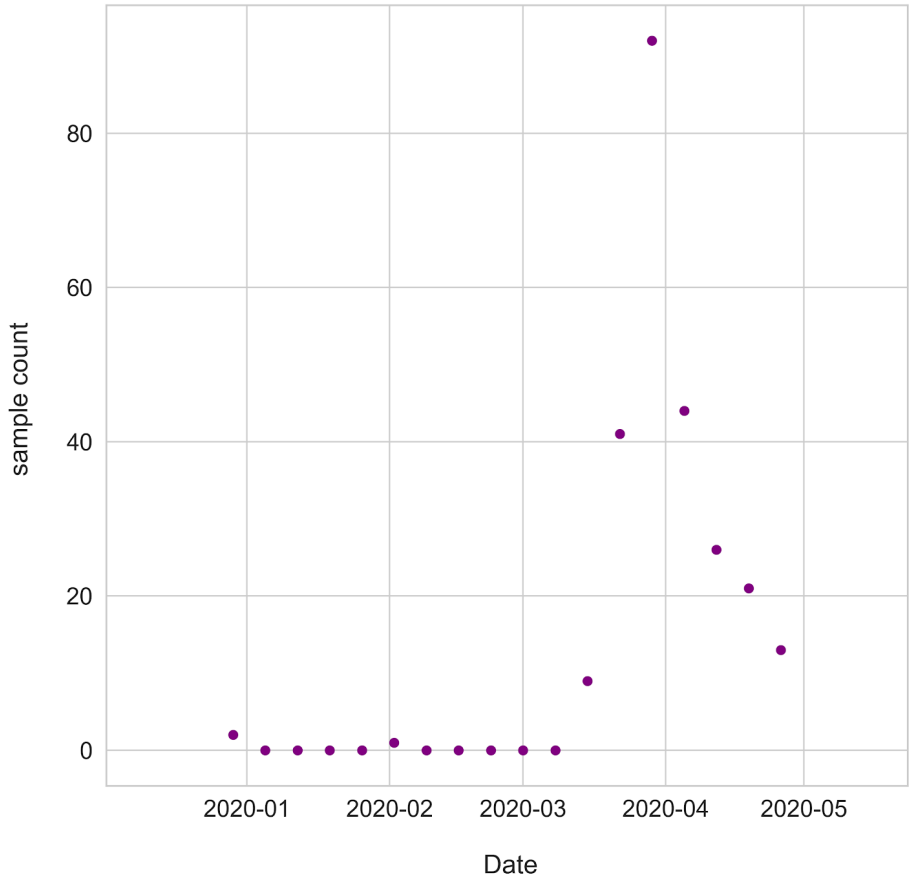
Supplemental Figure 1.

Diagnostic deletion in the index Dane County sample. Consensus-level deletion identified in the Dane County index sample. Zoomed in panel shows nucleotide and amino acid identities of the in-frame deletion.



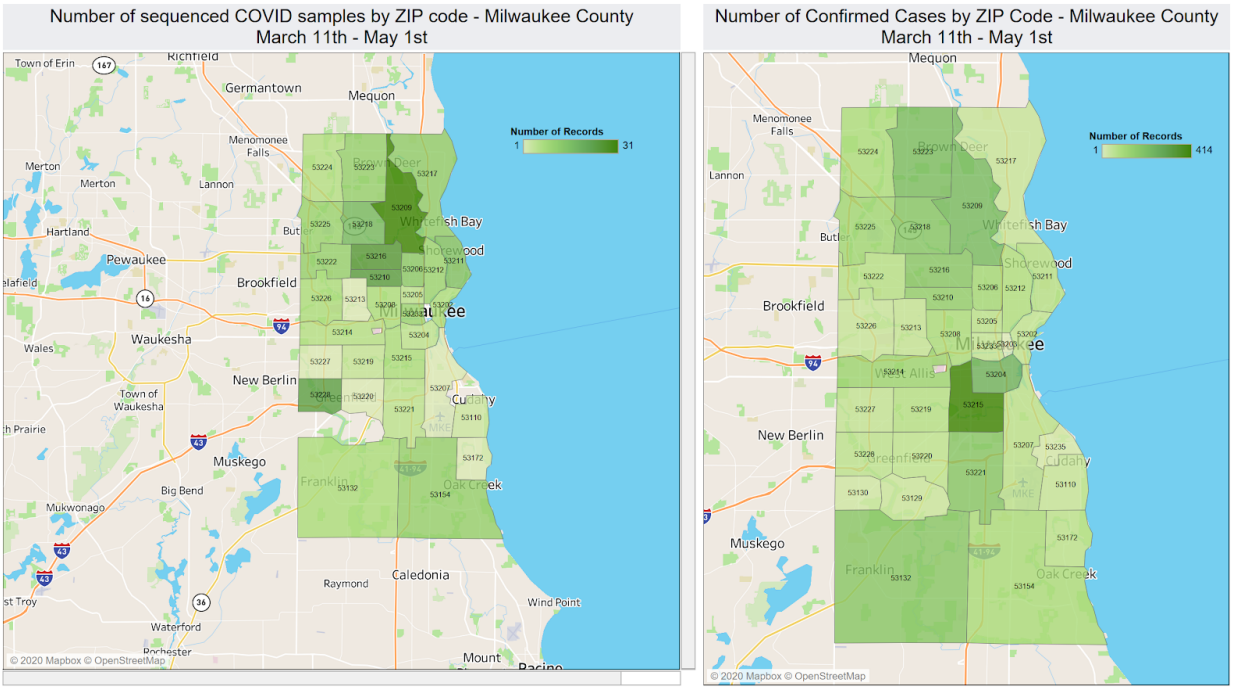
Supplemental Figure 2.

Sampling sensitivity of estimates for the number of introductions into Dane and Milwaukee Counties. Estimates of the number of introductions into Dane and Milwaukee Counties using a time aligned maximum likelihood phylogeny. N sequences (x-axis) were randomly sampled from the available Dane and Milwaukee County samples and the remaining were pruned from the tree. Ten replicates of each N were conducted and the number of introductions (y-axis) was estimated for each.



Supplemental Figure 3.

Temporal distribution of SARS-CoV-2 samples. The total count of samples collected (y-axis) during 1-week intervals (x-axis) from the first documented Wisconsin case through 18 April, 2020. Weeks with no samples are still shown here as indicated by data points at y = 0. Code to recreate this figure can be found in the GitHub repository – `data_raw/supplementary_figure_3.ipynb`.



Supplemental Figure 4.

Geographic distribution of Milwaukee County sequences (left) compared to test positive cases (right) by zip code. Case count data were collated using the Wisconsin Electronic Disease Surveillance System (WEDSS) and data were sorted and plotted using Tableau (https://www.tableau.com/) to create counts and color gradients by ZIP code.

Division	Number of sequences	Number of positive SARS-CoV-2 cases	Percent sequenced (%)
England	23212	265,849	8.73
Wales	5425	17389	31.20
Scotland	5121	18847	27.17
Washington	3957	55600	7.12
California	2189	500600	0.44
Victoria	1522	13469	11.30
New York	1505	411600	0.37
Michigan	1069	86200	1.24
Wisconsin	963	52900	1.82
Louisiana	786	116300	0.68
Singapore	770	54555	1.41
Denmark	735	14306	5.14
Northern Ireland	717	6049	11.85
Netherlands	692	56381	1.23
Utah	688	32400	2.12
New South Wales	656	3832	17.12
Portugal	642	52061	1.23
British Columbia	604	3834	15.75
Reykjavik	601	1932	31.11
Florida	560	469800	0.12
Basque Country	559	15634	3.58
Minnesota	545	54300	1.00
Virginia	522	89900	0.58
Gujarat	447	66777	0.67
Massachusetts	418	117200	0.36
Dane County - April 26th	122	405	30.12
Milwaukee County - April 26th	125	2629	4.75

Supplemental Table 1.

SARS-CoV-2 sequencing depth per admin division. Admin division is defined as the state or country. Percent sequenced (%) is calculated as the number of consensus sequences over the number of documented positive SARS-CoV-2 cases across a variety of geographic locations as of July 31, 2020. The number of sequences per geographic location were obtained on GISAID by downloading the NextMeta file and filtering on the

'Admin Division' search field. We compared the sequencing depth to the sequencing depth obtained by our study (indicated by '- April 26th') to highlight that we are one of the top 25 deepest sequenced divisions.

Location	Number of sequences	Number of positive SARS-CoV-2 cases	Percent sequenced (%)
Yakima County	1704	9971	17.09
San Diego	702	29577	2.37
Sydney	543	3809	14.26
East Baton Rouge Parish	424	11263	3.76
Manhattan	419	222522	0.19
Brooklyn	373	61948	0.60
Valencia	343	12999	2.64
Dane County	334	4145	8.06
King County	322	15946	2.02
Greater Houston Area	320	80914	0.40
Santa Clara County	313	11128	2.81
Donostia-San Sebastian & Vitoria-Gasteiz	535	15634	3.42
Wuhan	257	68138	0.38
Nassau County	202	43482	0.46
Snohomish County	187	6033	3.10
Orange County	184	38131	0.48
Munich	181	51068	0.35
Queens	175	67598	0.26
South Yorkshire	171	9610	1.78
San Francisco	163	7231	2.25
Pierce County	159	5851	2.72
Milwaukee County	153	19332	0.79
Los Angeles County	146	198355	0.07
Hyderabad	144	73050	0.20
Dane County - April 26th	122	405	30.12
Milwaukee County - April 26th	125	2629	4.75

Supplemental Table 2.

SARS-CoV-2 sequencing depth per location. Percent sequenced (%) is calculated as the number of consensus sequences over the number of documented positive SARS-CoV-2

cases across a variety of geographic locations. The number of sequences per geographic location were obtained on GISAID by downloading the NextMeta file and filtering on the 'Location' search field. We compared the sequencing depth to the sequencing depth obtained by our study (indicated by '- April 26th') to highlight that we are one of the top 25 deepest sequenced locations.

strain	gisaid_epi_isl	Genbank accession	sra_accession	srr_run_number	biosample_id	bioproject	date	region	country	division	location	Region exposure	Country exposure	Division exposure
USA/WI1/2020	EPI_ISL_408670	MT039887	SRX7777165	SRR11140745	SAMN14154204	PRJNA607948	2020-1-31	North America	USA	Wisconsin	Dane county	Asia	China	Hubei
USA/WI-UW-02/2020	EPI_ISL_416489	-	SRX7972381	SRR11393278	SAMN14428237	PRJNA614504	2020-3-15	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-04/2020	EPI_ISL_416491	-	SRX7972383	SRR11393276	SAMN14428239	PRJNA614504	2020-3-15	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-05/2020	EPI_ISL_416492	-	SRX7972384	SRR11393275	SAMN14428240	PRJNA614504	2020-3-15	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-03/2020	EPI_ISL_416523	-	SRX7972382	SRR11393277	SAMN14428238	PRJNA614504	2020-3-14	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-06/2020	EPI_ISL_417200	-	SRX7988797	SRR11410125	SAMN14443954	PRJNA614504	2020-3-21	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-07/2020	EPI_ISL_417201	-	SRX7988798	SRR11410124	SAMN14443955	PRJNA614504	2020-3-21	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-08/2020	EPI_ISL_417202	-	SRX7988799	SRR11410123	SAMN14443956	PRJNA614504	2020-3-21	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-09/2020	EPI_ISL_417203	-	SRX7988800	SRR11410122	SAMN14443957	PRJNA614504	2020-3-21	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-10/2020	EPI_ISL_417204	-	SRX7988801	SRR11410121	SAMN14443958	PRJNA614504	2020-3-21	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-25/2020	EPI_ISL_421283	-	SRX8079441	SRR11507378	SAMN14555931	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-26/2020	EPI_ISL_421284	-	SRX8079442	SRR11507377	SAMN14555932	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-27/2020	EPI_ISL_421285	-	SRX8079453	SRR11507366	SAMN14555933	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-28/2020	EPI_ISL_421286	-	SRX8079464	SRR11507355	SAMN14555934	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-29/2020	EPI_ISL_421287	-	SRX8079475	SRR11507344	SAMN14555935	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-31/2020	EPI_ISL_421289	-	SRX8079497	SRR11507322	SAMN14555937	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-32/2020	EPI_ISL_421290	-	SRX8079498	SRR11507321	SAMN14555938	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-33/2020	EPI_ISL_421291	-	SRX8079499	SRR11507320	SAMN14555939	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-34/2020	EPI_ISL_421292	-	SRX8079500	SRR11507319	SAMN14555940	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin

USA/WI-UW-35/2020	EPI_ISL_421293	-	SRX8079443	SRR11507376	SAMN14555941	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-37/2020	EPI_ISL_421295	-	SRX8079445	SRR11507374	SAMN14555943	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-38/2020	EPI_ISL_421296	-	SRX8079446	SRR11507373	SAMN14555944	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-39/2020	EPI_ISL_421297	-	SRX8079447	SRR11507372	SAMN14555945	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-40/2020	EPI_ISL_421298	-	SRX8079448	SRR11507371	SAMN14555946	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-41/2020	EPI_ISL_421299	-	SRX8079449	SRR11507370	SAMN14555947	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-42/2020	EPI_ISL_421300	-	SRX8079450	SRR11507369	SAMN14555948	PRJNA614504	2020-3-18	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-43/2020	EPI_ISL_421301	-	SRX8079451	SRR11507368	SAMN14555949	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-44/2020	EPI_ISL_421302	-	SRX8079452	SRR11507367	SAMN14555950	PRJNA614504	2020-3-17	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-45/2020	EPI_ISL_421303	-	SRX8079454	SRR11507365	SAMN14555951	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-46/2020	EPI_ISL_421304	-	SRX8079455	SRR11507364	SAMN14555952	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-48/2020	EPI_ISL_421306	-	SRX8079457	SRR11507362	SAMN14555954	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-49/2020	EPI_ISL_421307	-	SRX8079458	SRR11507361	SAMN14555955	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-50/2020	EPI_ISL_421308	-	SRX8079459	SRR11507360	SAMN14555956	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Green County	North America	USA	Wisconsin
USA/WI-UW-51/2020	EPI_ISL_421309	-	SRX8079460	SRR11507359	SAMN14555957	PRJNA614504	2020-3-20	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-52/2020	EPI_ISL_421310	-	SRX8079461	SRR11507358	SAMN14555958	PRJNA614504	2020-3-18	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-53/2020	EPI_ISL_421311	-	SRX8079462	SRR11507357	SAMN14555959	PRJNA614504	2020-3-18	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-54/2020	EPI_ISL_421312	-	SRX8079463	SRR11507356	SAMN14555960	PRJNA614504	2020-3-20	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-55/2020	EPI_ISL_421313	-	SRX8079465	SRR11507354	SAMN14555961	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-56/2020	EPI_ISL_421314	-	SRX8079466	SRR11507353	SAMN14555962	PRJNA614504	2020-3-18	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-58/2020	EPI_ISL_421316	-	SRX8079468	SRR11507351	SAMN14555964	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-59/2020	EPI_ISL_421317	-	SRX8079469	SRR11507350	SAMN14555965	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-60/2020	EPI_ISL_421318	-	SRX8079470	SRR11507349	SAMN14555966	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-61/2020	EPI_ISL_421319	-	SRX8079471	SRR11507348	SAMN14555967	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-62/2020	EPI_ISL_421320	-	SRX8079472	SRR11507347	SAMN14555968	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-63/2020	EPI_ISL_421321	-	SRX8079473	SRR11507346	SAMN14555969	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-64/2020	EPI_ISL_421322	-	SRX8079474	SRR11507345	SAMN14555970	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-65/2020	EPI_ISL_421323	-	SRX8079476	SRR11507343	SAMN14555971	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin

USA/WI-UW-66/2020	EPI_ISL_421324	-	SRX8079477	SRR11507342	SAMN14555972	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-67/2020	EPI_ISL_421325	-	SRX8079478	SRR11507341	SAMN14555973	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-68/2020	EPI_ISL_421326	-	SRX8079479	SRR11507340	SAMN14555974	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-69/2020	EPI_ISL_421327	-	SRX8079480	SRR11507339	SAMN14555975	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-70/2020	EPI_ISL_421328	-	SRX8079481	SRR11507338	SAMN14555976	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-71/2020	EPI_ISL_421329	-	SRX8079482	SRR11507337	SAMN14555977	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-72/2020	EPI_ISL_421330	-	SRX8079483	SRR11507336	SAMN14555978	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-73/2020	EPI_ISL_421331	-	SRX8079484	SRR11507335	SAMN14555979	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-74/2020	EPI_ISL_421332	-	SRX8079485	SRR11507334	SAMN14555980	PRJNA614504	2020-3-20	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-75/2020	EPI_ISL_421333	-	SRX8079487	SRR11507332	SAMN14555981	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-76/2020	EPI_ISL_421334	-	SRX8079488	SRR11507331	SAMN14555982	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-77/2020	EPI_ISL_421335	-	SRX8079489	SRR11507330	SAMN14555983	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-78/2020	EPI_ISL_421336	-	SRX8079490	SRR11507329	SAMN14555984	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-79/2020	EPI_ISL_421338	-	SRX8079491	SRR11507328	SAMN14555985	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-80/2020	EPI_ISL_421339	-	SRX8079492	SRR11507327	SAMN14555986	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-81/2020	EPI_ISL_421340	-	SRX8079493	SRR11507326	SAMN14555987	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-82/2020	EPI_ISL_421341	-	SRX8079494	SRR11507325	SAMN14555988	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-84/2020	EPI_ISL_421343	-	SRX8079496	SRR11507323	SAMN14555990	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-85/2020	EPI_ISL_425142	-	SRX8114902	SRR11544850	SAMN14596861	PRJNA614504	2020-4-2	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-86/2020	EPI_ISL_425143	-	SRX8114903	SRR11544849	SAMN14596862	PRJNA614504	2020-4-2	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-87/2020	EPI_ISL_425144	-	SRX8114914	SRR11544838	SAMN14596863	PRJNA614504	2020-4-2	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-88/2020	EPI_ISL_425145	-	SRX8114925	SRR11544827	SAMN14596864	PRJNA614504	2020-4-5	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-89/2020	EPI_ISL_425146	-	SRX8114931	SRR11544821	SAMN14596865	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-90/2020	EPI_ISL_425147	-	SRX8114932	SRR11544820	SAMN14596866	PRJNA614504	2020-4-3	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-91/2020	EPI_ISL_425148	-	SRX8114933	SRR11544819	SAMN14596867	PRJNA614504	2020-4-1	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-92/2020	EPI_ISL_425149	-	SRX8114934	SRR11544818	SAMN14596868	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-93/2020	EPI_ISL_425150	-	SRX8114935	SRR11544817	SAMN14596869	PRJNA614504	2020-4-4	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-94/2020	EPI_ISL_425151	-	SRX8114936	SRR11544816	SAMN14596870	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin

USA/WI-UW-95/2020	EPI_ISL_425152	-	SRX8114904	SRR11544848	SAMN14596871	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-96/2020	EPI_ISL_425153	-	SRX8114905	SRR11544847	SAMN14596872	PRJNA614504	2020-4-1	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-97/2020	EPI_ISL_425154	-	SRX8114906	SRR11544846	SAMN14596873	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-98/2020	EPI_ISL_425155	-	SRX8114907	SRR11544845	SAMN14596874	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-99/2020	EPI_ISL_425156	-	SRX8114908	SRR11544844	SAMN14596875	PRJNA614504	2020-4-2	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-100/2020	EPI_ISL_425157	-	SRX8114909	SRR11544843	SAMN14596876	PRJNA614504	2020-4-3	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-101/2020	EPI_ISL_425158	-	SRX8114910	SRR11544842	SAMN14596877	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-102/2020	EPI_ISL_425159	-	SRX8114911	SRR11544841	SAMN14596878	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-103/2020	EPI_ISL_425160	-	SRX8114912	SRR11544840	SAMN14596879	PRJNA614504	2020-4-3	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-105/2020	EPI_ISL_425162	-	SRX8114915	SRR11544837	SAMN14596881	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-106/2020	EPI_ISL_425163	-	SRX8114916	SRR11544836	SAMN14596882	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-107/2020	EPI_ISL_425164	-	SRX8114917	SRR11544835	SAMN14596883	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-108/2020	EPI_ISL_425165	-	SRX8114918	SRR11544834	SAMN14596884	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-109/2020	EPI_ISL_425166	-	SRX8114919	SRR11544833	SAMN14596885	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-110/2020	EPI_ISL_425167	-	SRX8114920	SRR11544832	SAMN14596886	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-111/2020	EPI_ISL_425168	-	SRX8114921	SRR11544831	SAMN14596887	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-112/2020	EPI_ISL_425169	-	SRX8114922	SRR11544830	SAMN14596888	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-114/2020	EPI_ISL_425171	-	SRX8114924	SRR11544828	SAMN14596890	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-115/2020	EPI_ISL_425172	-	SRX8114926	SRR11544826	SAMN14596891	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-116/2020	EPI_ISL_425173	-	SRX8114927	SRR11544825	SAMN14596892	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-117/2020	EPI_ISL_425174	-	SRX8114928	SRR11544824	SAMN14596893	PRJNA614504	2020-3-30	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-119/2020	EPI_ISL_425176	-	SRX8114930	SRR11544822	SAMN14596895	PRJNA614504	2020-4-10	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-120/2020	EPI_ISL_427427	-	SRX8149929	SRR11582218	SAMN14654585	PRJNA614504	2020-4-13	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-122/2020	EPI_ISL_427429	-	SRX8149905	SRR11582242	SAMN14654587	PRJNA614504	2020-4-7	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-124/2020	EPI_ISL_427431	-	SRX8149923	SRR11582224	SAMN14654589	PRJNA614504	2020-4-7	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-127/2020	EPI_ISL_427434	-	SRX8149926	SRR11582221	SAMN14654592	PRJNA614504	2020-4-7	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-128/2020	EPI_ISL_427435	-	SRX8149927	SRR11582220	SAMN14654593	PRJNA614504	2020-4-7	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-129/2020	EPI_ISL_427436	-	SRX8149928	SRR11582219	SAMN14654594	PRJNA614504	2020-4-6	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin

USA/WI-UW-135/2020	EPI_ISL_427442	-	SRX8149900	SRR11582247	SAMN14654600	PRJNA614504	2020-4-9	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-140/2020	EPI_ISL_427447	-	SRX8149906	SRR11582241	SAMN14654605	PRJNA614504	2020-4-9	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-144/2020	EPI_ISL_427451	-	SRX8149910	SRR11582237	SAMN14654609	PRJNA614504	2020-4-6	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-145/2020	EPI_ISL_427452	-	SRX8149911	SRR11582236	SAMN14654610	PRJNA614504	2020-4-10	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-146/2020	EPI_ISL_427453	-	SRX8149912	SRR11582235	SAMN14654611	PRJNA614504	2020-4-6	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-147/2020	EPI_ISL_427454	-	SRX8149913	SRR11582234	SAMN14654612	PRJNA614504	2020-4-10	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-148/2020	EPI_ISL_427455	-	SRX8149914	SRR11582233	SAMN14654613	PRJNA614504	2020-4-7	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-150/2020	EPI_ISL_427457	-	SRX8149917	SRR11582230	SAMN14654615	PRJNA614504	2020-4-13	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-151/2020	EPI_ISL_427458	-	SRX8149918	SRR11582229	SAMN14654616	PRJNA614504	2020-4-8	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-152/2020	EPI_ISL_427459	-	SRX8149919	SRR11582228	SAMN14654617	PRJNA614504	2020-4-6	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-153/2020	EPI_ISL_427460	-	SRX8149920	SRR11582227	SAMN14654618	PRJNA614504	2020-4-9	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-155/2020	EPI_ISL_427462	-	SRX8149922	SRR11582225	SAMN14654620	PRJNA614504	2020-4-10	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-156/2020	EPI_ISL_428252	-	SRX8151738	SRR11584248	SAMN14656612	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-157/2020	EPI_ISL_428253	-	SRX8151739	SRR11584247	SAMN14656613	PRJNA614504	2020-3-15	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-158/2020	EPI_ISL_428254	-	SRX8151750	SRR11584236	SAMN14656614	PRJNA614504	2020-3-15	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-159/2020	EPI_ISL_428255	-	SRX8151761	SRR11584225	SAMN14656615	PRJNA614504	2020-3-15	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-160/2020	EPI_ISL_428256	-	SRX8151772	SRR11584214	SAMN14656616	PRJNA614504	2020-3-15	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-161/2020	EPI_ISL_428257	-	SRX8151783	SRR11584203	SAMN14656617	PRJNA614504	2020-3-29	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-162/2020	EPI_ISL_428258	-	SRX8151794	SRR11584192	SAMN14656618	PRJNA614504	2020-3-29	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-165/2020	EPI_ISL_428261	-	SRX8151827	SRR11584159	SAMN14656621	PRJNA614504	2020-3-17	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-166/2020	EPI_ISL_428262	-	SRX8151740	SRR11584246	SAMN14656622	PRJNA614504	2020-3-17	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-168/2020	EPI_ISL_428264	-	SRX8151742	SRR11584244	SAMN14656624	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-170/2020	EPI_ISL_428266	-	SRX8151744	SRR11584242	SAMN14656626	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-171/2020	EPI_ISL_428267	-	SRX8151745	SRR11584241	SAMN14656627	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-172/2020	EPI_ISL_428268	-	SRX8151746	SRR11584240	SAMN14656628	PRJNA614504	2020-3-19	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-173/2020	EPI_ISL_428269	-	SRX8151747	SRR11584239	SAMN14656629	PRJNA614504	2020-3-20	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-176/2020	EPI_ISL_428272	-	SRX8151751	SRR11584235	SAMN14656632	PRJNA614504	2020-3-20	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-177/2020	EPI_ISL_428273	-	SRX8151752	SRR11584234	SAMN14656633	PRJNA614504	2020-3-20	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin

USA/WI-UW-178/2020	EPI_ISL_428274	-	SRX8151753	SRR11584233	SAMN14656634	PRJNA614504	2020-3-20	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-179/2020	EPI_ISL_428275	-	SRX8151754	SRR11584232	SAMN14656635	PRJNA614504	2020-3-21	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-181/2020	EPI_ISL_428277	-	SRX8151756	SRR11584230	SAMN14656637	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-182/2020	EPI_ISL_428278	-	SRX8151757	SRR11584229	SAMN14656638	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-183/2020	EPI_ISL_428279	-	SRX8151758	SRR11584228	SAMN14656639	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-184/2020	EPI_ISL_428280	-	SRX8151759	SRR11584227	SAMN14656640	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-185/2020	EPI_ISL_428281	-	SRX8151760	SRR11584226	SAMN14656641	PRJNA614504	2020-3-22	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-186/2020	EPI_ISL_428282	-	SRX8151762	SRR11584224	SAMN14656642	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-187/2020	EPI_ISL_428283	-	SRX8151763	SRR11584223	SAMN14656643	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-188/2020	EPI_ISL_428284	-	SRX8151764	SRR11584222	SAMN14656644	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-189/2020	EPI_ISL_428285	-	SRX8151765	SRR11584221	SAMN14656645	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-190/2020	EPI_ISL_428286	-	SRX8151766	SRR11584220	SAMN14656646	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-191/2020	EPI_ISL_428287	-	SRX8151767	SRR11584219	SAMN14656647	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-192/2020	EPI_ISL_428288	-	SRX8151768	SRR11584218	SAMN14656648	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-195/2020	EPI_ISL_428291	-	SRX8151771	SRR11584215	SAMN14656651	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-196/2020	EPI_ISL_428292	-	SRX8151773	SRR11584213	SAMN14656652	PRJNA614504	2020-3-21	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-197/2020	EPI_ISL_428293	-	SRX8151774	SRR11584212	SAMN14656653	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-200/2020	EPI_ISL_428296	-	SRX8151777	SRR11584209	SAMN14656656	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-205/2020	EPI_ISL_428301	-	SRX8151782	SRR11584204	SAMN14656661	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-207/2020	EPI_ISL_428303	-	SRX8151785	SRR11584201	SAMN14656663	PRJNA614504	2020-3-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-208/2020	EPI_ISL_428304	-	SRX8151786	SRR11584200	SAMN14656664	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-209/2020	EPI_ISL_428305	-	SRX8151787	SRR11584199	SAMN14656665	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-210/2020	EPI_ISL_428306	-	SRX8151788	SRR11584198	SAMN14656666	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-211/2020	EPI_ISL_428307	-	SRX8151789	SRR11584197	SAMN14656667	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-212/2020	EPI_ISL_428308	-	SRX8151790	SRR11584196	SAMN14656668	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-213/2020	EPI_ISL_428309	-	SRX8151791	SRR11584195	SAMN14656669	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-215/2020	EPI_ISL_428311	-	SRX8151793	SRR11584193	SAMN14656671	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-216/2020	EPI_ISL_428312	-	SRX8151795	SRR11584191	SAMN14656672	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin

USA/WI-UW-217/2020	EPI_ISL_428313	-	SRX8151796	SRR11584190	SAMN14656673	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-218/2020	EPI_ISL_428314	-	SRX8151797	SRR11584189	SAMN14656674	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-219/2020	EPI_ISL_428315	-	SRX8151798	SRR11584188	SAMN14656675	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-221/2020	EPI_ISL_428317	-	SRX8151800	SRR11584186	SAMN14656677	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-222/2020	EPI_ISL_428318	-	SRX8151801	SRR11584185	SAMN14656678	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-223/2020	EPI_ISL_428319	-	SRX8151802	SRR11584184	SAMN14656679	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-225/2020	EPI_ISL_428321	-	SRX8151804	SRR11584182	SAMN14656681	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-226/2020	EPI_ISL_428322	-	SRX8151806	SRR11584180	SAMN14656682	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-227/2020	EPI_ISL_428323	-	SRX8151807	SRR11584179	SAMN14656683	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-228/2020	EPI_ISL_428324	-	SRX8151808	SRR11584178	SAMN14656684	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-229/2020	EPI_ISL_428325	-	SRX8151809	SRR11584177	SAMN14656685	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-230/2020	EPI_ISL_428326	-	SRX8151810	SRR11584176	SAMN14656686	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-231/2020	EPI_ISL_428327	-	SRX8151811	SRR11584175	SAMN14656687	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-232/2020	EPI_ISL_428328	-	SRX8151812	SRR11584174	SAMN14656688	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-233/2020	EPI_ISL_428329	-	SRX8151813	SRR11584173	SAMN14656689	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-234/2020	EPI_ISL_428330	-	SRX8151814	SRR11584172	SAMN14656690	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-235/2020	EPI_ISL_428331	-	SRX8151815	SRR11584171	SAMN14656691	PRJNA614504	2020-3-28	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-236/2020	EPI_ISL_428332	-	SRX8151817	SRR11584169	SAMN14656692	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-237/2020	EPI_ISL_428333	-	SRX8151818	SRR11584168	SAMN14656693	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-238/2020	EPI_ISL_428334	-	SRX8151819	SRR11584167	SAMN14656694	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-240/2020	EPI_ISL_428336	-	SRX8151821	SRR11584165	SAMN14656696	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-241/2020	EPI_ISL_428337	-	SRX8151822	SRR11584164	SAMN14656697	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-242/2020	EPI_ISL_428338	-	SRX8151823	SRR11584163	SAMN14656698	PRJNA614504	2020-3-28	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-245/2020	EPI_ISL_428341	-	SRX8151826	SRR11584160	SAMN14656701	PRJNA614504	2020-3-28	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-246/2020	EPI_ISL_428342	-	SRX8151828	SRR11584158	SAMN14656702	PRJNA614504	2020-3-28	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-248/2020	EPI_ISL_428344	-	SRX8151830	SRR11584156	SAMN14656704	PRJNA614504	2020-3-28	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-259/2020	EPI_ISL_428935	-	SRX8155700	SRR11588239	SAMN14669377	PRJNA614504	2020-4-17	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-260/2020	EPI_ISL_428936	-	SRX8155701	SRR11588238	SAMN14669378	PRJNA614504	2020-4-18	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin

USA/WI-UW-270/2020	EPI_ISL_436564	-	SRX8281160	SRR11721857	SAMN14844834	PRJNA614504	2020-3-12	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-272/2020	EPI_ISL_436566	-	SRX8281172	SRR11721845	SAMN14844836	PRJNA614504	2020-3-27	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-273/2020	EPI_ISL_436567	-	SRX8281183	SRR11721834	SAMN14844837	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-274/2020	EPI_ISL_436568	-	SRX8281194	SRR11721823	SAMN14844838	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-275/2020	EPI_ISL_436569	-	SRX8281128	SRR11721889	SAMN14844839	PRJNA614504	2020-4-1	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-276/2020	EPI_ISL_436570	-	SRX8281139	SRR11721878	SAMN14844840	PRJNA614504	2020-4-1	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-277/2020	EPI_ISL_436571	-	SRX8281150	SRR11721867	SAMN14844841	PRJNA614504	2020-4-2	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-278/2020	EPI_ISL_436572	-	SRX8281158	SRR11721859	SAMN14844842	PRJNA614504	2020-4-3	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-279/2020	EPI_ISL_436573	-	SRX8281159	SRR11721858	SAMN14844843	PRJNA614504	2020-4-3	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-280/2020	EPI_ISL_436574	-	SRX8281162	SRR11721855	SAMN14844844	PRJNA614504	2020-4-3	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-285/2020	EPI_ISL_436579	-	SRX8281167	SRR11721850	SAMN14844849	PRJNA614504	2020-4-6	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-288/2020	EPI_ISL_436582	-	SRX8281170	SRR11721847	SAMN14844852	PRJNA614504	2020-4-8	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-289/2020	EPI_ISL_436583	-	SRX8281171	SRR11721846	SAMN14844853	PRJNA614504	2020-4-8	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-290/2020	EPI_ISL_436584	-	SRX8281173	SRR11721844	SAMN14844854	PRJNA614504	2020-4-8	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-291/2020	EPI_ISL_436585	-	SRX8281174	SRR11721843	SAMN14844855	PRJNA614504	2020-4-8	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-293/2020	EPI_ISL_436587	-	SRX8281176	SRR11721841	SAMN14844857	PRJNA614504	2020-4-9	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-296/2020	EPI_ISL_436590	-	SRX8281179	SRR11721838	SAMN14844860	PRJNA614504	2020-4-9	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-299/2020	EPI_ISL_436593	-	SRX8281182	SRR11721835	SAMN14844863	PRJNA614504	2020-4-13	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-301/2020	EPI_ISL_436595	-	SRX8281185	SRR11721832	SAMN14844865	PRJNA614504	2020-4-13	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-302/2020	EPI_ISL_436596	-	SRX8281186	SRR11721831	SAMN14844866	PRJNA614504	2020-4-13	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-303/2020	EPI_ISL_436597	-	SRX8281187	SRR11721830	SAMN14844867	PRJNA614504	2020-4-13	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-304/2020	EPI_ISL_436598	-	SRX8281188	SRR11721829	SAMN14844868	PRJNA614504	2020-4-13	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-305/2020	EPI_ISL_436599	-	SRX8281189	SRR11721828	SAMN14844869	PRJNA614504	2020-4-14	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-306/2020	EPI_ISL_436600	-	SRX8281190	SRR11721827	SAMN14844870	PRJNA614504	2020-4-14	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-308/2020	EPI_ISL_436602	-	SRX8281192	SRR11721825	SAMN14844872	PRJNA614504	2020-4-16	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-309/2020	EPI_ISL_436603	-	SRX8281193	SRR11721824	SAMN14844873	PRJNA614504	2020-4-16	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-310/2020	EPI_ISL_436604	-	SRX8281195	SRR11721822	SAMN14844874	PRJNA614504	2020-4-15	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-313/2020	EPI_ISL_436607	-	SRX8281121	SRR11721896	SAMN14844877	PRJNA614504	2020-4-16	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin

USA/WI-UW-316/2020	EPI_ISL_436610	-	SRX8281124	SRR11721893	SAMN14844880	PRJNA614504	2020-4-18	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-319/2020	EPI_ISL_436613	-	SRX8281127	SRR11721890	SAMN14844883	PRJNA614504	2020-4-21	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-320/2020	EPI_ISL_436614	-	SRX8281129	SRR11721888	SAMN14844884	PRJNA614504	2020-4-21	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-321/2020	EPI_ISL_436615	-	SRX8281130	SRR11721887	SAMN14844885	PRJNA614504	2020-4-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-322/2020	EPI_ISL_436616	-	SRX8281131	SRR11721886	SAMN14844886	PRJNA614504	2020-4-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-323/2020	EPI_ISL_436617	-	SRX8281132	SRR11721885	SAMN14844887	PRJNA614504	2020-4-23	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-324/2020	EPI_ISL_436618	-	SRX8281133	SRR11721884	SAMN14844888	PRJNA614504	2020-4-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-325/2020	EPI_ISL_436619	-	SRX8281134	SRR11721883	SAMN14844889	PRJNA614504	2020-4-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-328/2020	EPI_ISL_436622	-	SRX8281137	SRR11721880	SAMN14844892	PRJNA614504	2020-4-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-329/2020	EPI_ISL_436623	-	SRX8281138	SRR11721879	SAMN14844893	PRJNA614504	2020-4-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-330/2020	EPI_ISL_436624	-	SRX8281140	SRR11721877	SAMN14844894	PRJNA614504	2020-4-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-331/2020	EPI_ISL_436625	-	SRX8281141	SRR11721876	SAMN14844895	PRJNA614504	2020-4-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-332/2020	EPI_ISL_436626	-	SRX8281142	SRR11721875	SAMN14844896	PRJNA614504	2020-4-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-333/2020	EPI_ISL_436627	-	SRX8281143	SRR11721874	SAMN14844897	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-334/2020	EPI_ISL_436628	-	SRX8281144	SRR11721873	SAMN14844898	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-335/2020	EPI_ISL_436629	-	SRX8281145	SRR11721872	SAMN14844899	PRJNA614504	2020-3-24	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-336/2020	EPI_ISL_436630	-	SRX8281146	SRR11721871	SAMN14844900	PRJNA614504	2020-3-25	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-337/2020	EPI_ISL_436631	-	SRX8281147	SRR11721870	SAMN14844901	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-338/2020	EPI_ISL_436632	-	SRX8281148	SRR11721869	SAMN14844902	PRJNA614504	2020-3-26	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-339/2020	EPI_ISL_436633	-	SRX8281149	SRR11721868	SAMN14844903	PRJNA614504	2020-3-31	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-340/2020	EPI_ISL_436634	-	SRX8281151	SRR11721866	SAMN14844904	PRJNA614504	2020-4-1	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-341/2020	EPI_ISL_436635	-	SRX8281152	SRR11721865	SAMN14844905	PRJNA614504	2020-4-2	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-342/2020	EPI_ISL_436636	-	SRX8281153	SRR11721864	SAMN14844906	PRJNA614504	2020-4-4	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-343/2020	EPI_ISL_436637	-	SRX8281154	SRR11721863	SAMN14844907	PRJNA614504	2020-4-5	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-344/2020	EPI_ISL_436638	-	SRX8281155	SRR11721862	SAMN14844908	PRJNA614504	2020-4-6	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-345/2020	EPI_ISL_436639	-	SRX8281156	SRR11721861	SAMN14844909	PRJNA614504	2020-4-9	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-346/2020	EPI_ISL_436640	-	SRX8281157	SRR11721860	SAMN14844910	PRJNA614504	2020-4-21	North America	USA	Wisconsin	Milwaukee county	North America	USA	Wisconsin
USA/WI-UW-347/2020	EPI_ISL_450701	MT506886	SRX8379446	SRR11828929	SAMN14995484	PRJNA614504	2020-4-18	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin

USA/WI-UW-351/2020	EPI_ISL_450705	MT506890	SRX8379463	SRR11828912	SAMN14995488	PRJNA614504	2020-4-14	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-356/2020	EPI_ISL_450710	MT506895	SRX8379447	SRR11828928	SAMN14995493	PRJNA614504	2020-4-14	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-359/2020	EPI_ISL_450713	MT506898	SRX8379450	SRR11828925	SAMN14995496	PRJNA614504	2020-4-18	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin
USA/WI-UW-362/2020	EPI_ISL_450716	MT506901	SRX8379453	SRR11828922	SAMN14995499	PRJNA614504	2020-4-17	North America	USA	Wisconsin	Dane county	North America	USA	Wisconsin

Supplemental Table 3.
GISAID accession numbers and associated information for all consensus sequences included in this manuscript.

name	pool	sequence	length	%gc	tm (use 65)
nCoV-2019_1_LEFT	nCoV-2019_1	ACCAACCAACTTTCGATCTCTTGT	24	41.67	60.69
nCoV-2019_1_RIGHT	nCoV-2019_1	CATCTTTAAGATGTTGACGTGCCTC	25	44	60.45
nCoV-2019_2_LEFT	nCoV-2019_2	CTGTTTTACAGGTTCGCGACGT	22	50	61.67
nCoV-2019_2_RIGHT	nCoV-2019_2	TAAGGATCAGTGCCAAGCTCGT	22	50	61.74
nCoV-2019_3_LEFT	nCoV-2019_1	CGGTAATAAAGGAGCTGGTGGC	22	54.55	61.32
nCoV-2019_3_RIGHT	nCoV-2019_1	AAGGTGTCTGCAATTCATAGCTCT	24	41.67	60.32
nCoV-2019_4_LEFT	nCoV-2019_2	GGTGTATACTGCTGCCGTGAAC	22	54.55	61.56
nCoV-2019_4_RIGHT	nCoV-2019_2	CACAAGTAGTGGCACCTTCTTTAGT	25	44	60.97
nCoV-2019_5_LEFT	nCoV-2019_1	TGGTGAAACTTCATGGCAGACG	22	50	61.39
nCoV-2019_5_RIGHT	nCoV-2019_1	ATTGATGTTGACTTTCTCTTTTTGGAGT	28	32.14	60.17
nCoV-2019_6_LEFT	nCoV-2019_2	GGTGTTGTTGGAGAAGGTTCCG	22	54.55	61.64
nCoV-2019_6_RIGHT	nCoV-2019_2	TAGCGGCCTTCTGTAAAACACG	22	50	61.18
nCoV-2019_7_LEFT	nCoV-2019_1	ATCAGAGGCTGCTCGTGTTGTA	22	50	61.73
nCoV-2019_7_LEFT_alt0	nCoV-2019_1	CATTTGCATCAGAGGCTGCTCG	22	54.55	62.44
nCoV-2019_7_RIGHT	nCoV-2019_1	TGCACAGGTGACAATTTGTCCA	22	45.45	60.95
nCoV-2019_7_RIGHT_alt5	nCoV-2019_1	AGGTGACAATTTGTCCACCGAC	22	50	61.07
nCoV-2019_8_LEFT	nCoV-2019_2	AGAGTTTCTTAGAGACGGTTGGGA	24	45.83	61
nCoV-2019_8_RIGHT	nCoV-2019_2	GCTTCAACAGCTTCACTAGTAGGT	24	45.83	60.56
nCoV-2019_9_LEFT	nCoV-2019_1	TCCCACAGAAGTGTTAACAGAGGA	24	45.83	61.18
nCoV-2019_9_LEFT_alt4	nCoV-2019_1	TTCCCACAGAAGTGTTAACAGAGG	24	45.83	60.44
nCoV-2019_9_RIGHT	nCoV-2019_1	ATGACAGCATCTGCCACAACAC	22	50	61.71
nCoV-2019_9_RIGHT_alt2	nCoV-2019_1	GACAGCATCTGCCACAACACAG	22	54.55	62.26
nCoV-2019_10_LEFT	nCoV-2019_2	TGAGAAGTGCTCTGCCTATACAGT	24	45.83	61.12
nCoV-2019_10_RIGHT	nCoV-2019_2	TCATCTAACCAATCTTCTTCTTGCTCT	27	37.04	60.31
nCoV-2019_11_LEFT	nCoV-2019_1	GGAATTTGGTGCCACTTCTGCT	22	50	61.66
nCoV-2019_11_RIGHT	nCoV-2019_1	TCATCAGATTCAACTTGCATGGCA	24	41.67	61.35
nCoV-2019_12_LEFT	nCoV-2019_2	AAACATGGAGGAGGTGTTGCAG	22	50	61.08

nCoV-2019_12_RIGHT	nCoV-2019_2	TTCACTCTTCATTTCCAAAAAGCTTGA	27	33.33	60.36
nCoV-2019_13_LEFT	nCoV-2019_1	TCGCACAAATGTCTACTTAGCTGT	24	41.67	60.56
nCoV-2019_13_RIGHT	nCoV-2019_1	ACCACAGCAGTTAAAACACCCT	22	45.45	60.36
nCoV-2019_14_LEFT	nCoV-2019_2	CATCCAGATTCTGCCACTCTTGT	23	47.83	60.62
nCoV-2019_14_LEFT_alt4	nCoV-2019_2	TGGCAATCTTCATCCAGATTCTGC	24	45.83	61.47
nCoV-2019_14_RIGHT	nCoV-2019_2	AGTTTCCACACAGACAGGCATT	22	45.45	60.42
nCoV-2019_14_RIGHT_alt2	nCoV-2019_2	TGCGTGTTTCTTCTGCATGTGC	22	50	62.76
nCoV-2019_15_LEFT	nCoV-2019_1	ACAGTGCTTAAAAAGTGTAAAAGTGCC	27	37.04	61.32
nCoV-2019_15_LEFT_alt1	nCoV-2019_1	AGTGCTTAAAAAGTGTAAAAGTGCCT	26	34.62	60.13
nCoV-2019_15_RIGHT	nCoV-2019_1	AACAGAAACTGTAGCTGGCACT	22	45.45	60.16
nCoV-2019_15_RIGHT_alt3	nCoV-2019_1	ACTGTAGCTGGCACTTTGAGAGA	23	47.83	61.57
nCoV-2019_16_LEFT	nCoV-2019_2	AATTTGGAAGAAGCTGCTCGGT	22	45.45	60.82
nCoV-2019_16_RIGHT	nCoV-2019_2	CACAACTTGCGTGTGGAGGTTA	22	50	61.32
nCoV-2019_17_LEFT	nCoV-2019_1	CTTCTTTCTTTGAGAGAAGTGAGGACT	27	40.74	60.69
nCoV-2019_17_RIGHT	nCoV-2019_1	TTTGTTGGAGTGTTAACAATGCAGT	25	36	60.11
nCoV-2019_18_LEFT	nCoV-2019_2	TGGAAATACCCACAAGTTAATGGTTTAAC	29	34.48	60.69
nCoV-2019_18_LEFT_alt2	nCoV-2019_2	ACTTCTATTAAATGGGCAGATAACAACTGT	30	33.33	61.38
nCoV-2019_18_RIGHT	nCoV-2019_2	AGCTTGTTTACCACACGTACAAGG	24	45.83	61.51
nCoV-2019_18_RIGHT_alt1	nCoV-2019_2	GCTTGTTTACCACACGTACAAGG	23	47.83	60.3
nCoV-2019_19_LEFT	nCoV-2019_1	GCTGTTATGTACATGGGCACACT	23	47.83	61.18
nCoV-2019_19_RIGHT	nCoV-2019_1	TGTCCAACTTAGGGTCAATTTCTGT	25	40	60.4
nCoV-2019_20_LEFT	nCoV-2019_2	ACAAAGAAAACAGTTACACAACAACCA	27	33.33	60.68
nCoV-2019_20_RIGHT	nCoV-2019_2	ACGTGGCTTTATTAGTTGCATTGTT	25	36	60.28
nCoV-2019_21_LEFT	nCoV-2019_1	TGGCTATTGATTATAAACACTACACACCC	29	37.93	61.49
nCoV-2019_21_LEFT_alt2	nCoV-2019_1	GGCTATTGATTATAAACACTACACACCCT	29	37.93	61.29
nCoV-2019_21_RIGHT	nCoV-2019_1	TAGATCTGTGTGGCCAACCTCT	22	50	60.83
nCoV-2019_21_RIGHT_alt0	nCoV-2019_1	GATCTGTGTGGCCAACCTCTTC	22	54.55	61.2
nCoV-2019_22_LEFT	nCoV-2019_2	ACTACCGAAGTTGTAGGAGACATTATACT	29	37.93	61.25
nCoV-2019_22_RIGHT	nCoV-2019_2	ACAGTATTCTTTGCTATAGTAGTCGGC	27	40.74	60.73
nCoV-2019_23_LEFT	nCoV-2019_1	ACAACTACTAACATAGTTACACGGTGT	27	37.04	60.26
nCoV-2019_23_RIGHT	nCoV-2019_1	ACCAGTACAGTAGGTTGCAATAGTG	25	44	60.57
nCoV-2019_24_LEFT	nCoV-2019_2	AGGCATGCCTTCTTACTGTACTG	23	47.83	60.37
nCoV-2019_24_RIGHT	nCoV-2019_2	ACATTCTAACCATAGCTGAAATCGGG	26	42.31	61.19
nCoV-2019_25_LEFT	nCoV-2019_1	GCAATTGTTTTTCAGCTATTTTGCAGT	27	33.33	60.73
nCoV-2019_25_RIGHT	nCoV-2019_1	ACTGTAGTGACAAGTCTCTCGCA	23	47.83	61.3
nCoV-2019_26_LEFT	nCoV-2019_2	TTGTGATACATTCTGTGCTGGTAGT	25	40	60.28
nCoV-2019_26_RIGHT	nCoV-2019_2	TCCGCACTATCACCAACATCAG	22	50	60.42
nCoV-2019_27_LEFT	nCoV-2019_1	ACTACAGTCAGCTTATGTGTCAACC	25	44	60.8
nCoV-2019_27_RIGHT	nCoV-2019_1	AATACAAGCACCAAGGTCACGG	22	50	61.13

nCoV-2019_28_LEFT	nCoV-2019_2	ACATAGAAGTTACTGGCGATAGTTGT	26	38.46	60.13
nCoV-2019_28_RIGHT	nCoV-2019_2	TGTTTAGACATGACATGAACAGGTGT	26	38.46	60.91
nCoV-2019_29_LEFT	nCoV-2019_1	ACTTGTGTTCCTTTTTGTTGCTGC	24	41.67	61.39
nCoV-2019_29_RIGHT	nCoV-2019_1	AGTGTACTCTATAAGTTTTGATGGTGTGT	29	34.48	60.69
nCoV-2019_30_LEFT	nCoV-2019_2	GCACAACTAATGGTGACTTTTTGCA	25	40	61.19
nCoV-2019_30_RIGHT	nCoV-2019_2	ACCACTAGTAGATACACAAACACCAG	26	42.31	60.3
nCoV-2019_31_LEFT	nCoV-2019_1	TTCTGAGTACTGTAGGCACGGC	22	54.55	62.03
nCoV-2019_31_RIGHT	nCoV-2019_1	ACAGAATAAACACCAGGTAAGAATGAGT	28	35.71	60.69
nCoV-2019_32_LEFT	nCoV-2019_2	TGGTGAATACAGTCATGTAGTTGCC	25	44	61.09
nCoV-2019_32_RIGHT	nCoV-2019_2	AGCACATCACTACGCAACTTTAGA	24	41.67	60.56
nCoV-2019_33_LEFT	nCoV-2019_1	ACTTTTGAAGAAGCTGCGCTGT	22	45.45	61.58
nCoV-2019_33_RIGHT	nCoV-2019_1	TGGACAGTAAACTACGTCATCAAGC	25	44	61.08
nCoV-2019_34_LEFT	nCoV-2019_2	TCCCATCTGGTAAAGTTGAGGGT	23	47.83	61.02
nCoV-2019_34_RIGHT	nCoV-2019_2	AGTGAAATTGGGCCTCATAGCA	22	45.45	60.03
nCoV-2019_35_LEFT	nCoV-2019_1	TGTTCGCATTCAACCAGGACAG	22	50	61.39
nCoV-2019_35_RIGHT	nCoV-2019_1	ACTTCATAGCCACAAGGTTAAAGTCA	26	38.46	60.69
nCoV-2019_36_LEFT	nCoV-2019_2	TTAGCTTGGTTGTACGCTGCTG	22	50	61.44
nCoV-2019_36_RIGHT	nCoV-2019_2	GAACAAAGACCATTGAGTACTCTGGA	26	42.31	60.74
nCoV-2019_37_LEFT	nCoV-2019_1	ACACACCACTGGTTGTTACTCAC	23	47.83	60.93
nCoV-2019_37_RIGHT	nCoV-2019_1	GTCCACACTCTCCTAGCACCAT	22	54.55	61.48
nCoV-2019_38_LEFT	nCoV-2019_2	ACTGTGTTATGTATGCATCAGCTGT	25	40	60.86
nCoV-2019_38_RIGHT	nCoV-2019_2	CACCAAGAGTCAGTCTAAAGTAGCG	25	48	61.13
nCoV-2019_39_LEFT	nCoV-2019_1	AGTATTGCCCTATTTTCTTCATAACTGGT	29	34.48	61
nCoV-2019_39_RIGHT	nCoV-2019_1	TGTAACTGGACACATTGAGCCC	22	50	60.55
nCoV-2019_40_LEFT	nCoV-2019_2	TGCACATCAGTAGTCTTACTCTCAGT	26	42.31	61.25
nCoV-2019_40_RIGHT	nCoV-2019_2	CATGGCTGCATCACGGTCAAAT	22	50	62.09
nCoV-2019_41_LEFT	nCoV-2019_1	GTTCCCTTCCATCATATGCAGCT	23	47.83	60.75
nCoV-2019_41_RIGHT	nCoV-2019_1	TGGTATGACAACCATTAGTTTGGCT	25	40	60.75
nCoV-2019_42_LEFT	nCoV-2019_2	TGCAAGAGATGGTTGTGTTCCC	22	50	61.08
nCoV-2019_42_RIGHT	nCoV-2019_2	CCTACCTCCCTTTGTTGTGTTGT	23	47.83	60.69
nCoV-2019_43_LEFT	nCoV-2019_1	TACGACAGATGTCTTGTGCTGC	22	50	60.93
nCoV-2019_43_RIGHT	nCoV-2019_1	AGCAGCATCTACAGCAAAAGCA	22	45.45	61.14
nCoV-2019_44_LEFT	nCoV-2019_2	TGCCACAGTACGTCTACAAGCT	22	50	61.66
nCoV-2019_44_LEFT_alt3	nCoV-2019_2	CCACAGTACGTCTACAAGCTGG	22	54.55	60.67
nCoV-2019_44_RIGHT	nCoV-2019_2	AACCTTTCCACATACCGCAGAC	22	50	60.87
nCoV-2019_44_RIGHT_alt0	nCoV-2019_2	CGCAGACGGTACAGACTGTGTT	22	54.55	62.77
nCoV-2019_45_LEFT	nCoV-2019_1	TACCTACAACTTGTGCTAATGACCC	25	44	60.57
nCoV-2019_45_LEFT_alt2	nCoV-2019_1	AGTATGTACAAATACCTACAACTTGTGCT	29	34.48	60.94
nCoV-2019_45_RIGHT	nCoV-2019_1	AAATTGTTTCTTCATGTTGGTAGTTAGAGA	30	30	60.01

nCoV-2019_45_RIGHT_alt7	nCoV-2019_1	TTCATGTTGGTAGTTAGAGAAAGTGTGTC	29	37.93	61.53
nCoV-2019_46_LEFT	nCoV-2019_2	TGTCGCTTCCAAGAAAAGGACG	22	50	61.38
nCoV-2019_46_LEFT_alt1	nCoV-2019_2	CGCTTCCAAGAAAAGGACGAAGA	23	47.83	61.35
nCoV-2019_46_RIGHT	nCoV-2019_2	CACGTTCACCTAAGTTGGCGTA	22	50	60.86
nCoV-2019_46_RIGHT_alt2	nCoV-2019_2	CACGTTCACCTAAGTTGGCGTAT	23	47.83	61.17
nCoV-2019_47_LEFT	nCoV-2019_1	AGGACTGGTATGATTTTGTAGAAAACCC	28	39.29	61.42
nCoV-2019_47_RIGHT	nCoV-2019_1	AATAACGGTCAAAGAGTTTTAACCTCTC	28	35.71	60.06
nCoV-2019_48_LEFT	nCoV-2019_2	TGTTGACACTGACTTAACAAAGCCT	25	40	61.09
nCoV-2019_48_RIGHT	nCoV-2019_2	TAGATTACCAGAAGCAGCGTGC	22	50	60.74
nCoV-2019_49_LEFT	nCoV-2019_1	AGGAATTACTTGTGTATGCTGCTGA	25	40	60.57
nCoV-2019_49_RIGHT	nCoV-2019_1	TGACGATGACTTGGTTAGCATTAATACA	28	35.71	61.05
nCoV-2019_50_LEFT	nCoV-2019_2	GTTGATAAGTACTTTGATTGTTACGATGGT	30	33.33	60.59
nCoV-2019_50_RIGHT	nCoV-2019_2	TAACATGTTGTGCCAACCACCA	22	45.45	60.95
nCoV-2019_51_LEFT	nCoV-2019_1	TCAATAGCCGCCACTAGAGGAG	22	54.55	61.34
nCoV-2019_51_RIGHT	nCoV-2019_1	AGTGCATTAACATTGGCCGTGA	22	45.45	61.14
nCoV-2019_52_LEFT	nCoV-2019_2	CATCAGGAGATGCCACAACTGC	22	54.55	61.83
nCoV-2019_52_RIGHT	nCoV-2019_2	GTTGAGAGCAAAATTCATGAGGTCC	25	44	60.62
nCoV-2019_53_LEFT	nCoV-2019_1	AGCAAAATGTTGGACTGAGACTGA	24	41.67	60.69
nCoV-2019_53_RIGHT	nCoV-2019_1	AGCCTCATAAAACTCAGGTTCCC	23	47.83	60.31
nCoV-2019_54_LEFT	nCoV-2019_2	TGAGTTAACAGGACACATGTTAGACA	26	38.46	60.18
nCoV-2019_54_RIGHT	nCoV-2019_2	AACCAAAAACTTGTCCATTAGCACA	25	36	60.11
nCoV-2019_55_LEFT	nCoV-2019_1	ACTCAACTTTACTTAGGAGGTATGAGCT	28	39.29	61.43
nCoV-2019_55_RIGHT	nCoV-2019_1	GGTGTACTCTCCTATTTGTACTTTACTGT	29	37.93	60.54
nCoV-2019_56_LEFT	nCoV-2019_2	ACCTAGACCACCACTTAACCGA	22	50	60.49
nCoV-2019_56_RIGHT	nCoV-2019_2	ACACTATGCGAGCAGAAGGGTA	22	50	61.21
nCoV-2019_57_LEFT	nCoV-2019_1	ATTCTACACTCCAGGGACCACC	22	54.55	61.16
nCoV-2019_57_RIGHT	nCoV-2019_1	GTAATTGAGCAGGGTCGCCAAT	22	50	61.26
nCoV-2019_58_LEFT	nCoV-2019_2	TGATTTGAGTGTTGTCAATGCCAGA	25	40	61.44
nCoV-2019_58_RIGHT	nCoV-2019_2	CTTTTCTCCAAGCAGGGTTACGT	23	47.83	61.06
nCoV-2019_59_LEFT	nCoV-2019_1	TCACGCATGATGTTTCATCTGCA	23	43.48	61.42
nCoV-2019_59_RIGHT	nCoV-2019_1	AAGAGTCCTGTTACATTTTCAGCTTG	26	38.46	60.02
nCoV-2019_60_LEFT	nCoV-2019_2	TGATAGAGACCTTTATGACAAGTTGCA	27	37.04	60.53
nCoV-2019_60_RIGHT	nCoV-2019_2	GGTACCAACAGCTTCTCTAGTAGC	24	50	60.44
nCoV-2019_61_LEFT	nCoV-2019_1	TGTTTATCACCCGCGAAGAAGC	22	50	61.5
nCoV-2019_61_RIGHT	nCoV-2019_1	ATCACATAGACAACAGGTGCGC	22	50	61.25
nCoV-2019_62_LEFT	nCoV-2019_2	GGCACATGGCTTTGAGTTGACA	22	50	61.91
nCoV-2019_62_RIGHT	nCoV-2019_2	GTTGAACCTTTCTACAAGCCGC	22	50	60.35
nCoV-2019_63_LEFT	nCoV-2019_1	TGTTAAGCGTGTTGACTGGACT	22	45.45	60.16
nCoV-2019_63_RIGHT	nCoV-2019_1	ACAAACTGCCACCATCACAACC	22	50	61.85

nCoV-2019_64_LEFT	nCoV-2019_2	TCGATAGATATCCTGCTAATTCCATTGT	28	35.71	60.11
nCoV-2019_64_RIGHT	nCoV-2019_2	AGTCTTGTAAAAGTGTTCCAGAGGT	25	40	60.1
nCoV-2019_65_LEFT	nCoV-2019_1	GCTGGCTTTAGCTTGTGGGTTT	22	50	61.92
nCoV-2019_65_RIGHT	nCoV-2019_1	TGTCAGTCATAGAACAAACACCAATAGT	28	35.71	60.9
nCoV-2019_66_LEFT	nCoV-2019_2	GGGTGTGGACATTGCTGCTAAT	22	50	61.21
nCoV-2019_66_RIGHT	nCoV-2019_2	TCAATTTCCATTTGACTCCTGGGT	24	41.67	60.45
nCoV-2019_67_LEFT	nCoV-2019_1	GTTGTCCAACAATTACCTGAAACTTACT	28	35.71	60.43
nCoV-2019_67_RIGHT	nCoV-2019_1	CAACCTTAGAAACTACAGATAAATCTTGGG	30	36.67	60.4
nCoV-2019_68_LEFT	nCoV-2019_2	ACAGGTTCATCTAAGTGTGTGTGT	24	41.67	60.14
nCoV-2019_68_RIGHT	nCoV-2019_2	CTCCTTTATCAGAACCAGCACCA	23	47.83	60.31
nCoV-2019_69_LEFT	nCoV-2019_1	TGTCGCAAAATATACTCAACTGTGTCA	27	37.04	61.43
nCoV-2019_69_RIGHT	nCoV-2019_1	TCTTTATAGCCACGGAACCTCCA	23	47.83	61.14
nCoV-2019_70_LEFT	nCoV-2019_2	ACAAAAGAAAATGACTCTAAAGAGGGTTT	29	31.03	60.13
nCoV-2019_70_RIGHT	nCoV-2019_2	TGACCTTCTTTTAAAGACATAACAGCAG	28	35.71	60.27
nCoV-2019_71_LEFT	nCoV-2019_1	ACAAATCCAATTCAGTTGTCTTCCTATTC	29	34.48	60.54
nCoV-2019_71_RIGHT	nCoV-2019_1	TGGAAAAGAAAGGTAAGAACAAGTCCT	27	37.04	60.8
nCoV-2019_72_LEFT	nCoV-2019_2	ACACGTGGTGTTTATTACCCTGAC	24	45.83	61.04
nCoV-2019_72_RIGHT	nCoV-2019_2	ACTCTGAACTCACTTTCCATCCAAC	25	44	60.97
nCoV-2019_73_LEFT	nCoV-2019_1	CAATTTTGTAATGATCCATTTTTGGGTGT	29	31.03	60.29
nCoV-2019_73_RIGHT	nCoV-2019_1	CACCAGCTGTCCAACCTGAAGA	22	54.55	62.45
nCoV-2019_74_LEFT	nCoV-2019_2	ACATCACTAGGTTTCAAACTTTACTTGC	28	35.71	60.68
nCoV-2019_74_RIGHT	nCoV-2019_2	GCAACACAGTTGCTGATTCTCTTC	24	45.83	60.85
nCoV-2019_75_LEFT	nCoV-2019_1	AGAGTCCAACCAACAGAATCTATTGT	26	38.46	60.24
nCoV-2019_75_RIGHT	nCoV-2019_1	ACCACCAACCTTAGAATCAAGATTGT	26	38.46	60.69
nCoV-2019_76_LEFT	nCoV-2019_2	AGGGCAAACTGGAAAGATTGCT	22	45.45	60.76
nCoV-2019_76_LEFT_alt3	nCoV-2019_2	GGGCAAACTGGAAAGATTGCTGA	23	47.83	61.87
nCoV-2019_76_RIGHT	nCoV-2019_2	ACACCTGTGCCTGTTAAACCAT	22	45.45	60.42
nCoV-2019_76_RIGHT_alt0	nCoV-2019_2	ACCTGTGCCTGTTAAACCATTGA	23	43.48	60.69
nCoV-2019_77_LEFT	nCoV-2019_1	CCAGCAACTGTTTGTGGACCTA	22	50	60.75
nCoV-2019_77_RIGHT	nCoV-2019_1	CAGCCCCTATTAAACAGCCTGC	22	54.55	61.59
nCoV-2019_78_LEFT	nCoV-2019_2	CAACTTACTCCTACTTGGCGTGT	23	47.83	60.55
nCoV-2019_78_RIGHT	nCoV-2019_2	TGTGTACAAAAACTGCCATATTGCA	25	36	60.22
nCoV-2019_79_LEFT	nCoV-2019_1	GTGGTGATTCAACTGAATGCAGC	23	47.83	60.92
nCoV-2019_79_RIGHT	nCoV-2019_1	CATTTCATCTGTGAGCAAAGGTGG	24	45.83	60.62
nCoV-2019_80_LEFT	nCoV-2019_2	TTGCCTTGGTGATATTGCTGCT	22	45.45	60.89
nCoV-2019_80_RIGHT	nCoV-2019_2	TGGAGCTAAGTTGTTTAACAAGCG	24	41.67	60.02
nCoV-2019_81_LEFT	nCoV-2019_1	GCACTTGGAAAACTTCAAGATGTGG	25	44	61.24
nCoV-2019_81_RIGHT	nCoV-2019_1	GTGAAGTTCTTTTCTTGTGCAGGG	24	45.83	60.73
nCoV-2019_82_LEFT	nCoV-2019_2	GGGCTATCATCTTATGTCCTTCCCT	25	48	61.52

nCoV-2019_82_RIGHT	nCoV-2019_2	TGCCAGAGATGTCACCTAAATCAA	24	41.67	60.02
nCoV-2019_83_LEFT	nCoV-2019_1	TCCTTTGCAACCTGAATTAGACTCA	25	40	60.46
nCoV-2019_83_RIGHT	nCoV-2019_1	TTTGACTCCTTTGAGCACTGGC	22	50	61.33
nCoV-2019_84_LEFT	nCoV-2019_2	TGCTGTAGTTGTCTCAAGGGCT	22	50	61.61
nCoV-2019_84_RIGHT	nCoV-2019_2	AGGTGTGAGTAAACTGTTACAAACAAC	27	37.04	60.36
nCoV-2019_85_LEFT	nCoV-2019_1	ACTAGCACTCTCCAAGGGTGTT	22	50	61.03
nCoV-2019_85_RIGHT	nCoV-2019_1	ACACAGTCTTTTACTCCAGATTCCC	25	44	60.51
nCoV-2019_86_LEFT	nCoV-2019_2	TCAGGTGATGGCACAACAAGTC	22	50	61.07
nCoV-2019_86_RIGHT	nCoV-2019_2	ACGAAAGCAAGAAAAAGAAGTACGC	25	40	61.01
nCoV-2019_87_LEFT	nCoV-2019_1	CGACTACTAGCGTGCCTTTGTA	22	50	60.16
nCoV-2019_87_RIGHT	nCoV-2019_1	ACTAGGTTCCATTGTTCAAGGAGC	24	45.83	60.81
nCoV-2019_88_LEFT	nCoV-2019_2	CCATGGCAGATTCCAACGGTAC	22	54.55	61.58
nCoV-2019_88_RIGHT	nCoV-2019_2	TGGTCAGAATAGTGCCATGGAGT	23	47.83	61.4
nCoV-2019_89_LEFT	nCoV-2019_1	GTACGCGTTCCATGTGGTCATT	22	50	61.5
nCoV-2019_89_LEFT_alt2	nCoV-2019_1	CGCGTTCCATGTGGTCATTCAA	22	50	62.01
nCoV-2019_89_RIGHT	nCoV-2019_1	ACCTGAAAGTCAACGAGATGAAACA	25	40	60.91
nCoV-2019_89_RIGHT_alt4	nCoV-2019_1	ACGAGATGAAACATCTGTTGTCACT	25	40	60.74
nCoV-2019_90_LEFT	nCoV-2019_2	ACACAGACCATTCCAGTAGCAGT	23	47.83	61.58
nCoV-2019_90_RIGHT	nCoV-2019_2	TGAAATGGTGAATTGCCCTCGT	22	45.45	60.82
nCoV-2019_91_LEFT	nCoV-2019_1	TCACTACCAAGAGTGTGTTAGAGGT	25	44	60.93
nCoV-2019_91_RIGHT	nCoV-2019_1	TTCAAGTGAGAACCAAAAGATAATAAGCA	29	31.03	60.03
nCoV-2019_92_LEFT	nCoV-2019_2	TTTGTGCTTTTTAGCCTTTCTGCT	24	37.5	60.14
nCoV-2019_92_RIGHT	nCoV-2019_2	AGGTTCCTGGCAATTAATTGTAAAAGG	27	37.04	60.53
nCoV-2019_93_LEFT	nCoV-2019_1	TGAGGCTGGTTCTAAATCACCCA	23	47.83	61.59
nCoV-2019_93_RIGHT	nCoV-2019_1	AGGTCTTCCTTGCCATGTTGAG	22	50	60.55
nCoV-2019_94_LEFT	nCoV-2019_2	GGCCCCAAGGTTTACCCAATAA	22	50	60.56
nCoV-2019_94_RIGHT	nCoV-2019_2	TTTGGCAATGTTGTTCCTTGAGG	23	43.48	60.18
nCoV-2019_95_LEFT	nCoV-2019_1	TGAGGGAGCCTTGAATACACCA	22	50	61.1
nCoV-2019_95_RIGHT	nCoV-2019_1	CAGTACGTTTTTGCCGAGGCTT	22	50	61.95
nCoV-2019_96_LEFT	nCoV-2019_2	GCCAACAACAACAAGGCCAAAC	22	50	61.82
nCoV-2019_96_RIGHT	nCoV-2019_2	TAGGCTCTGTTGGTGGGAATGT	22	50	61.36
nCoV-2019_97_LEFT	nCoV-2019_1	TGGATGACAAAGATCCAAATTTCAAAGA	28	32.14	60.22
nCoV-2019_97_RIGHT	nCoV-2019_1	ACACACTGATTAAAGATTGCTATGTGAG	28	35.71	60.17
nCoV-2019_98_LEFT	nCoV-2019_2	AACAATTGCAACAATCCATGAGCA	24	37.5	60.5
nCoV-2019_98_RIGHT	nCoV-2019_2	TTCTCCTAAGAAGCTATTAAAATCACATGG	30	33.33	60.01

Supplemental Table 4.

ARTIC primers.

Chapter 6:

Viral sequencing reveals US healthcare personnel rarely become infected with SARS-CoV-2 through patient contact

Katarina M. Braun[1]*, Gage K. Moreno[2]*, Ashley Buys[3], Elizabeth D. Somsen[2], Max Bobholz[2], Molly A. Accola[3], Laura Anderson[3], William M. Rehrauer[3], David A. Baker[2], Nasia Safdar[4], Alexander J. Lepak[4], David H. O'Connor[2,5#], Thomas C. Friedrich[1,5#]

*These authors contributed equally

#These authors contributed equally; corresponding authors; tfriedri@wisc.edu (T.C.F), dhoconno@wisc.edu (D.H.O)

[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]University of Wisconsin School of Medicine and Public Health, Madison, WI, United States of America and the William S. Middleton Memorial Veterans Hospital

[4]Department of Medicine, Division of Infectious Diseases, University of Wisconsin School of Medicine and Public Health, Madison, WI

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

Clin Infect Dis. 2021 Apr 15:ciab281. doi: 10.1093/cid/ciab281. Epub ahead of print. PMID: 33857303; PMCID: PMC8083259.

Author summary

We investigated SARS-CoV-2 infections in 95 healthcare personnel (HCP). The majority of HCP infections could not be linked to a patient or co-worker. Infection control procedures, consistently followed, offer significant protection to HCP caring for COVID-19 patients.

Abstract

Background

Healthcare personnel (HCP) are at increased risk of infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). We posit current infection control guidelines generally protect HCP from SARS-CoV-2 infection in a healthcare setting.

Methods

In this retrospective case series, we use viral genomics to investigate the likely source of SARS-CoV-2 infection in HCP at a major academic medical institution in the Upper Midwest of the United States between 25 March - 27 December, 2020. We obtain limited epidemiological data through informal interviews and review of the electronic health record. We combine epidemiological information with healthcare-associated viral sequences and with viral sequences collected in the broader community to infer the most likely source of infection in HCP.

Results

We investigated SARS-CoV-2 infection clusters involving 95 HCP and 137 possible patient contact sequences. The majority of HCP infections could not be linked to a patient or co-worker (55/95; 57.9%) and were genetically similar to viruses circulating concurrently in the community. We found 10.5% of infections could be traced to a coworker (10/95). Strikingly, only 4.2% of HCP infections could be traced to a patient source (4/95).

Conclusions

Infections among HCP add further strain to the healthcare system and put patients, HCP, and communities at risk. We found no evidence for healthcare-associated transmission in the majority of HCP infections evaluated here. Though we cannot rule out the possibility of cryptic healthcare-associated transmission, it appears that HCP most commonly becomes infected with SARS-CoV-2 via community exposure. This emphasizes the ongoing importance of mask-wearing, physical distancing, robust testing programs, and rapid distribution of vaccines.

Introduction

Despite the use of personal protective equipment (PPE) and other strategies to mitigate risk, front-line healthcare workers are at increased risk for infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) compared to the general population [354–356]. Healthcare-associated SARS-CoV-2 infections negatively affect healthcare personnel (HCP) through direct health impacts, lost wages, and secondary consequences for their close contacts [357]. Additional repercussions include staffing shortages,

environmental contamination, low morale and other mental health impacts on HCP. Each of these can impact overall quality of care [358,359]. Here we use rapid viral sequencing and forensic genomics to investigate the likely sources of infection in 95 confirmed cases of coronavirus-disease 2019 (COVID-19) in HCP. We further describe how the results of these investigations informed infection control recommendations within a large academic medical system in the midwestern United States.

The US Centers for Disease Control and Prevention (CDC) have released guidelines for infection prevention for HCP interacting directly with patients with SARS-CoV-2 [360]. These guidelines include recommendations for the proper use of PPE, hand hygiene, precautions to be taken during aerosol-generating procedures, environmental infection control practices and many others. These guidelines, and additional institution-specific infection control measures [361], were in place at the institution evaluated here. We posit that these guidelines are generally successful in protecting HCP from SARS-CoV-2 infection in a healthcare setting. Here we test this hypothesis using viral sequences collected from infected HCP, as well as concurrent viral sequences collected from the broader community, to investigate possible sources of infection in a series of HCP.

With a few exceptions [362–364], viral sequencing is not currently standard practice for investigating healthcare-associated SARS-CoV-2 infections, although we and others have highlighted the potential utility of this approach [365–368]. It is currently estimated that SARS-CoV-2 acquires ~2-2.5 consensus mutations per month [369,370]. Viral sequences can therefore be used to infer likely epidemiological relationships. Viruses collected from

transmission pairs or from individuals with a shared source of infection are expected to share higher levels of genetic diversity than individuals who become infected at similar times, but from distinct sources. This was especially true during March - December 2020 in the United States, when transmission rates were high and multiple viruses of distinct genetic lineages co-circulated in many areas [371]. By increasing the resolution of inference, rapid viral sequencing can facilitate a targeted approach to examine SARS-CoV-2 nosocomial outbreaks at the level of the individual and the institution, which others have referred to collectively as “precision epidemiology” [372].

Materials and methods

Sample approvals and sample selection criteria

From 12 March 2020 to 10 January 2021, ~1,172 HCP tested positive for SARS-CoV-2 at a major academic medical institution in the Upper Midwest. Whenever possible, informal interviews and contact tracing information was collected for each HCP infection. HCP viruses and viruses from other individuals involved in each outbreak (patients, co-workers) were sequenced if epidemiological data did not reveal a likely exposure source and if residual swab was available. Individuals who had high-risk exposures to family or community members with confirmed COVID-19 were not sequenced. Individuals who reported high-risk community activities, such as attending a wedding, funeral, indoor bar, or plane travel, were also not sequenced. Relevant patient contacts of individuals with no likely exposure source were identified in the Epic electronic medical record using a comprehensive caregiver trace. This function identifies all patient records accessed by a

HCP being traced. Diagnostic assays for the samples included in this study were performed in a clinical lab using CDC's diagnostic RT-PCR [280], the Hologic Panther SARS-CoV-2 assay [281], or the Aptima SARS-CoV-2 assay [282].

Summary of infection control measures to prevent transmission of SARS-CoV-2 at our institution

Detailed descriptions of all infection control measures implemented to prevent transmission of SARS-CoV-2 at the medical institution evaluated here can be found in a recent report by Lepak et al [361]. Briefly, these guidelines include a universal testing policy for all patients, negative air pressure in all locations where SARS-CoV-2 patients are treated, a limit of one visitor or primary support person per patient per day (required to undergo screening prior to entry), establishment of an employee testing site with required employee self-monitoring for symptoms, maintenance of a log of persons entering the room of a confirmed or suspected COVID-19 patient for contact tracing purposes, detailed PPE guidelines, among others.

Sample preparation and sequencing

Detailed methods descriptions can be found in Moreno et al. [55]. Briefly, viral RNA was extracted using the Viral Total Nucleic Acid Purification kit (Promega, Madison, WI, USA) on a Maxwell RSC 48 instrument. Complementary DNA (cDNA) was synthesized using SuperScript IV Reverse Transcriptase [235,283]. A SARS-CoV-2-specific multiplex PCR was performed using the ARTIC v3 primers [235,283]. DNA was made compatible for sequencing using the one-pot native ligation protocol with Oxford Nanopore kit SQK-LSK109 and its

Native Barcodes (EXP-NBD104 and EXP-NBD114) [283]. Up to 23 samples, with one no-template control (water), were pooled prior to being run on the appropriate Nanopore flow cell (FLO-MIN106) using the 72hr run script.

Processing raw ONT data

Sequencing data was processed using the ARTIC bioinformatics pipeline (https://github.com/artic-network/artic-ncov2019), with a few modifications. Briefly, we have modified the ARTIC pipeline so that it demultiplexes raw fastq files using qcat as each fastq file is generated by the GridION (https://github.com/nanoporetech/qcat). Once a barcode reaches 100k reads, it maps to the Wuhan-Hu-1 reference (Genbank: MN908947.3) using minimap2. This alignment will then be used to generate consensus sequences and variant calls using medaka (https://github.com/nanoporetech/medaka). The analysis pipeline is available at https://github.com/gagekmoreno/SARS-CoV-2-in-Southern-Wisconsin.

Consensus sequence analysis – clade and lineage generation

Samples were excluded from downstream analysis if gaps in the consensus sequence totaled ≥20% of the genome. Each sample's consensus sequence was visually inspected in Geneious Prime (https://www.geneious.com) and/or in Nextstrain's Nextclade online tool (https://clades.nextstrain.org/). We used Pangolin's command-line tool to assign sequences to Pangolin lineages (https://github.com/cov-lineages/pangolin).

Consensus sequence analysis – Southeast Wisconsin Phylogenetic tree

Wisconsin-centric time-resolved and divergence phylogenetic trees (seen in **Supplementary File 1**) were built using the standard Nextstrain tools and scripts [145]. Laboratories responsible for obtaining and genetic sequence data included here, if not our own, are documented in **Supplementary File 2**. An interactive view of this Nextstrain phylogenetic tree can be found here.

Genetic distance comparisons

Full length SARS-CoV-2 sequences available on GISAID as of 10 March, 2020 were obtained and filtered on "Wisconsin" and parsed by date of collection into month bins. We used this dataset as a community comparator set. Consensus mutations were called against Wuhan-Hu-1 reference (Genbank: MN908947.3) using Varscan v2.4.3. HCP and patient samples were similarly binned by month. We performed a permutation test comparing the percent overlap in mutation identities in 100,000 randomly selected pairs from the community comparator set and plotted these values as a distribution in **Figure 5**. We plotted the genetic diversity of n-choose-2 random pairs for healthcare-associated sample, where n is the number of HCP and patient samples available for comparison each month.

Data availability

Accession numbers for all healthcare-associated samples can be found in Supplemental File 1**.** Code to replicate the genetic distance analyses can be found in the GitHub

accompanying this manuscript [373]. **Figures 1A, 2A, and 3A** were created with BioRender (http://biorender.com).

Study approvals

The University of Wisconsin-Madison Institutional Review Board deemed this study quality improvement, rather than research, and considered it exempt from review. Data and metadata were collected as part of routine infection control policy in nosocomial outbreaks and all data were deidentified prior to analysis.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by a COVID-19 Response grants from the Wisconsin Partnership Program at the University of Wisconsin School of Medicine and Public Health awarded to T.C.F. and D.H.O. Author N.S. is supported by the National Institute of Allergy and Infectious Diseases Institute (NIAID) Grant 1DP2AI144244-01.

Results

HCP began testing positive for SARS-CoV-2 at a major academic biomedical institution in the American Upper Midwest in early March 2020. We began sequencing viral genomes from residual nasopharyngeal specimens from the individuals involved in these infection clusters. We focused our analyses on HCP infections and infection clusters that

were highest risk for nosocomial transmission, as when healthcare-associated transmission could not be ruled out using epidemiological data alone (see methods for details). Each investigation included at least one HCP, all known direct and indirect SARS-CoV-2-positive patient contacts where residual swab was available, and occasionally extended to epidemiologically-linked household contacts.

We consider three potential sources of HCP infection: “patient source” (via HCP-patient interactions), “employee source” (via HCP-HCP interactions), and “no evidence of healthcare-associated transmission”. Some HCP infections did not fit neatly into these categories so we have included three additional categories which are defined in full in the **Supplemental File 1**. These additional categories are “combined patient and employee cluster”, “outside community”, and “inconclusive”. In each category, for us to conclude person A was a likely source of infection for person B, persons A and B must have had known contact with each other, must have been tested within 14 days of each other, and must have been infected with viruses differing by no more than a single mutation [285].

From 12 March, 2020 to 10 January, 2021 ~1,172 HCP tested positive for SARS-CoV-2 at the institution we evaluate in this study. In total, we investigated 95 HCP (8.1%) and 137 possible patient contacts collected between 25 March and 27 December, 2020 (n=232). Of these, we were able to generate 87 complete HCP sequences and 87 complete patient contact sequences which were used in downstream analyses (n=174). Of the 87 patient sequences, 4 were included in 2 or more outbreak investigations.

We did not find a closely related virus among co-worker and patient contacts in 55 HCP infections. We identified a specific household or community source of infection in an additional 3 cases (58/95; 61.1%). We find a smaller percentage could be traced to a coworker (10/95; 10.5%) or were part of a patient-employee cluster (12/95; 12.6%). Strikingly, the smallest proportion of HCP infections could be clearly traced to a patient source (4/95; 4.2%). The remaining HCP infections could not be definitively traced to a single source and were therefore inconclusive (11/95; 11.6%) (**Table 1**). Below, we describe one representative example of three distinct transmission scenarios – no evidence of healthcare-associated transmission, HCP-to-HCP, and patient-to-HCP.

In case #20, we compared the viral sequence of a HCP (HCP 20-1), who tested positive on 5 October, to a patient contact who tested positive eight days prior. A comprehensive caregiver trace of HCP 20-1 revealed a single patient contact with diagnosed COVID-19 (patient 20-A) within the 14 days prior HCP 20-1's symptom onset. HCP 20-1 provided direct care to patient 20-A while wearing appropriate PPE and with no reported lapses in PPE. HCP 20-1 was infected with a virus clustering with the 20G clade whereas patient 20-A was infected with a 20A-clade virus. The sequences of these viruses differed at >20 sites, so we concluded these individuals were unlikely to represent a transmission pair (**Figure 1**).

In case #16, we investigated infections in three HCP who worked in the same department and tested positive on 8 September (HCP 16-2), 18 September (HCP 16-1), and 29 September (HCP 16-3). Contact tracing revealed HCP 16-2 worked for two days prior to symptom onset and may have had unmasked contact with HCP 16-1 during overlapping

meal breaks. Contact tracing additionally revealed HCP 16-3 had an exposure event lasting >15 minutes in the outside community prior to testing positive. Viral sequencing in this cluster showed HCP 16-1 and 16-2 were infected with 20G-clade viruses identical at the consensus level, while HCP 16-3 was infected with a genetically dissimilar 20A-clade virus. We therefore concluded HCP 16-2 was a likely source of infection for HCP 16-1, while HCP 16-3 was likely infected elsewhere (**Figure 2**).

Case #10 involved a HCP (HCP 10-1) who provided care for 15 patients diagnosed with COVID-19 in the 14 days prior to symptom onset. HCP 10-1 provided direct care to each of these patients while wearing appropriate PPE with no reported lapses in PPE. We generated consensus sequences from HCP 10-1 and nine patient contacts. There was insufficient viral RNA (vRNA) in the remaining six patient contacts to generate high-quality consensus sequences for comparison. The virus isolated from patient 10-G was identical to the virus from HCP 10-1. Given the known epidemiological association between these two individuals, the time separating sample collections (28 July & 5 August), and identical viral sequences, we concluded patient 10-G is a likely source of infection for HCP 10-1 (**Figure 3**). However, we cannot rule out the possibility that another patient whose sample could not be sequenced also shared an identical virus.

HCP and patient viruses are broadly distributed throughout a phylogenetic tree showing the diversity of circulating viruses collected from the areas surrounding the academic medical center (**Figure 4**). To investigate the possibility that we missed cryptic healthcare-associated transmission, we compared genetic distances between random pairs of healthcare-

associated samples against the genetic distances between randomly paired sequences from the community dataset (grey tips in **Figure 4**) within each month in our study period (**Figure 5**). Overall, healthcare-associated pairs do not share substantially greater sequence identity than randomly paired sequences from the community. This is consistent with a relatively limited role for nosocomial spread of SARS-CoV-2. We additionally plot 14 pairs which are very likely to be true transmission pairs based on epidemiological data (e.g. HCP 2-1 and their household contact) and show these pairs are uniformly genetically identical (see dashed magenta lines **in Figure 5**).

The center where we conducted this case series implemented a number of changes to their institutional infection control guidelines based on these sequencing results [361]. The recommendations for extended reuse of medical grade face masks were clarified and now instruct HCP to consider barrier mask replacement after three days of wear and to inspect the barrier mask prior to each use and to replace if soiled or damaged. N-95s or powered air-purifying respirators (PAPR) are now universally required on inpatient units housing COVID-19-confirmed and suspected patients. In addition, medical-grade face masks, instead of cloth masks, are now required for HCP in all clinical areas, and not just direct patient care areas. This final recommendation was based on likely HCP-to-HCP transmission involving a HCP who was not directly involved in patient care of COVID-19 patients (case #14 in **Supplementary File 1**).

Discussion

HCP across the hospital are involved in caring for people with COVID-19, whether or not they work on an actual COVID-19 ward. With shifting guidelines and PPE shortages that persist today, it is critical to assess the risk that HCP treating people with known SARS-CoV-2 infection will become infected themselves. We used viral genome sequencing to assess the risk that HCP in a large academic medical system treating COVID-19 patients would acquire nosocomial infections. Our results suggest that caring for COVID-19 patients accounted for a minority of HCP infections (n=4). In contrast, HCP at this institution were much more likely to acquire SARS-CoV-2 from infected coworkers (n=10) or outside of the healthcare system (n=58). This result suggests that infection control procedures, consistently followed, offer significant protection to HCP caring for COVID-19 patients in the United States. A similar conclusion was drawn by recent studies evaluating healthcare-associated infections in the Netherlands and in the UK, suggesting this conclusion may hold across healthcare systems [358,374]. These results are further supported by another recent study which found the most important risk factor for HCP SARS-CoV-2 seropositivity was cumulative COVID-19 incidence in surrounding communities, not workplace factors [375].

This study has important limitations. We were able to generate high-quality sequence information for a minority of documented COVID-19 cases in HCP (87/1,172; 7.4%) during our study period (25 March - 27 December, 2020). Our dataset is therefore incomplete and may not be entirely representative of viruses circulating in this healthcare setting, particularly for asymptomatic cases. Similarly, we did not sequence viruses from

all SARS-CoV-2-positive patients who were treated at the medical center where we conducted this study. Given this limitation, we were often able to exclude patient contacts and co-workers as likely sources of infection in HCP, but we were rarely able to pinpoint the exact source of infection. It is therefore possible we have underestimated the true rate of SARS-CoV-2 transmission in this healthcare setting. However, the finding that randomly paired HCP and patient sequences do not have greater sequence identity than randomly paired sequences from across the surrounding community suggests to us that we have not severely underestimated nosocomial transmission. Our ability to determine the source of infections in these outbreaks was also often limited by incomplete contact tracing data; undocumented exposures between HCP may have occurred inside and outside of the workplace.

This study examined SARS-CoV-2 infections in HCP from a single academic medical center so our conclusions may not be broadly generalizable. However, another recent study evaluated healthcare-associated infections in the Netherlands and similarly found no evidence for widespread nosocomial transmission of SARS-CoV-2, suggesting our conclusions may hold across institutions and healthcare systems [374]. Further, we were not able to differentiate between routes of infection (airborne, droplet, contact) with the limited epidemiological data available to us in this study.

Sampling and contact tracing of nosocomial outbreaks is often coordinated by local hospitals and/or departments of health while expertise in viral sequencing, bioinformatics, and phylogenetics can more often be found in academic laboratories. Successful

application of precision epidemiology requires the integration of these areas. This is possible now at academic medical institutions like ours, but presents more of a challenge at smaller, rural, and private patient care centers. Federal support should be provided to help establish and maintain these collaborations in the current pandemic and in anticipation of future outbreaks.

Here we demonstrated how rapid whole-genome sequencing of current SARS-CoV-2 outbreaks in hospitals can be used retrospectively to reconstruct the likely source of HCP infection and prospectively to adjust and improve infection control practices and guidelines. The approach we describe here need not be limited to investigation of pandemic virus outbreaks. Key concepts from genome sequencing and routine pathogen surveillance can be applied to any nosocomial pathogen and inform changes to infection control practices. Overall, while we do find examples of patient-to-HCP and HCP-to-HCP spread, we found no evidence of healthcare-associated transmission in a majority of HCP infections, emphasizing the importance of ongoing measures to reduce community spread through mask-wearing, physical distancing, robust testing programs, and rapid distribution of vaccines.

Acknowledgements

We gratefully acknowledge Anna Heffron for assisting with sample transport. We also thank all healthcare workers and infection control teams for their ongoing dedication to patient and community health and wellness. We gratefully acknowledge the originating laboratories responsible for obtaining the specimens and the submitting laboratories

where genetic sequence data were generated and shared via the GISAID initiative (**Supplementary File 2**).

Figures, tables, and supplemental material

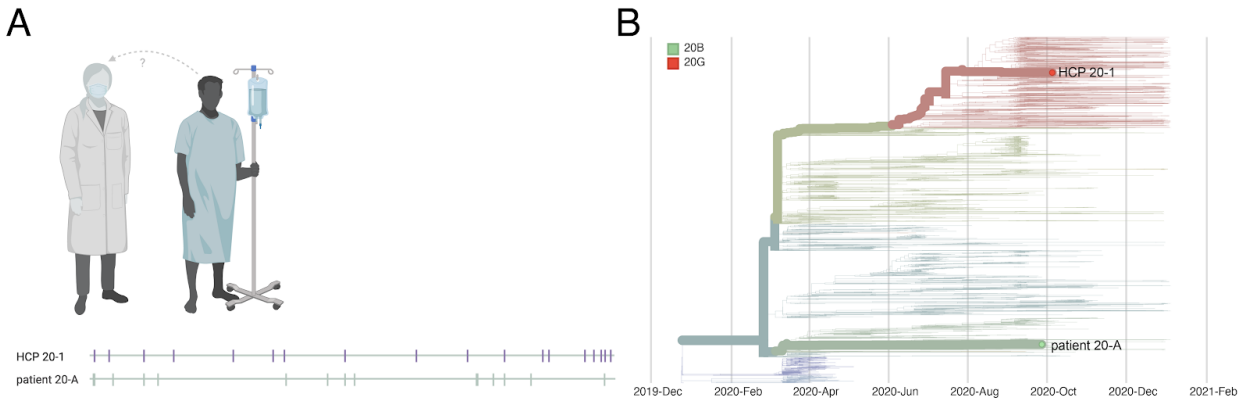


Figure 1. Graphical representation of case #20.

A. Virus sequences are aligned against SARS-CoV-2 reference sequence Wuhan-Hu-1 (MN908947.3). Vertical markers denote the location of consensus nucleotide differences between patient viruses and the reference. B. A time-resolved phylogenetic tree built using Nextstrain tools with all Wisconsin sequences available as of 2021-01-15. Viruses involved in this case are denoted with thick branches and labeled tips. Color denotes clade.

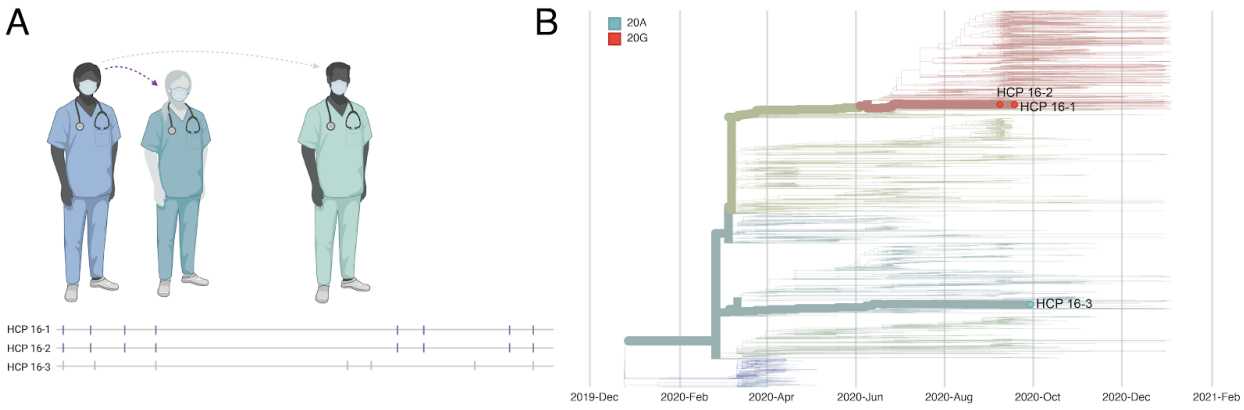


Figure 2. Graphical representation of case #16.

A. Virus sequences are aligned against SARS-CoV-2 reference sequence Wuhan-Hu-1 (MN908947.3). Vertical markers denote the location of consensus nucleotide differences between patient viruses and the reference. Purple vertical markers indicate identical virus sequences. B. A time-resolved phylogenetic tree built using Nextstrain tools with all Wisconsin sequences available as of 2021-01-15. Viruses involved in this case are denoted with thick branches and labeled tips. Color denotes clade.

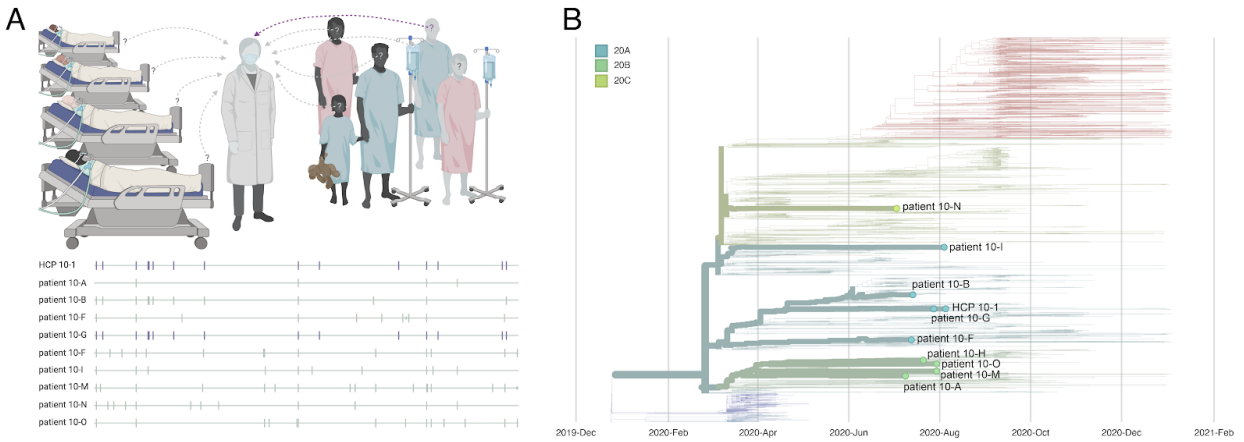


Figure 3. Graphical representation of case #10.

A. Virus sequences are aligned against SARS-CoV-2 reference sequence Wuhan-Hu-1 (MN908947.3). Vertical markers denote the location of consensus nucleotide differences between patient viruses and the reference. Purple vertical markers indicate identical virus sequences. B. A time-resolved phylogenetic tree built using Nextstrain tools with all Wisconsin sequences available as of 2021-01-15. Viruses involved in this case are denoted with thick branches and labeled tips. Color denotes clade.

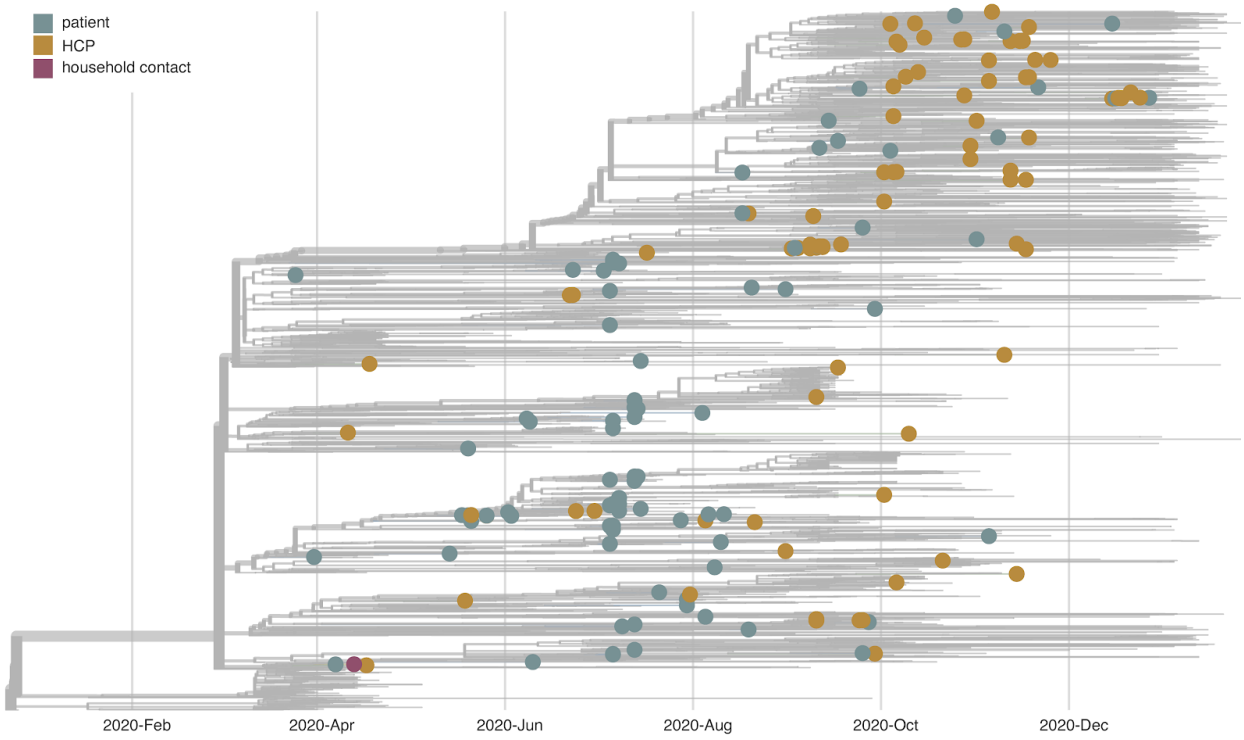


Figure 4. A time-resolved phylogenetic tree built using Nexstrain tools for all samples collected and shared in Wisconsin from March - December, 2020.

Healthcare-associated samples are denoted with enlarged tips and colored according to sample type. The grey tips reflect the community surveillance samples. It is likely additional HCP and patient sequences are represented in the community dataset, but we do not have access to sufficient metadata to make these designations. Laboratories responsible for obtaining and genetic sequence data included here, if not our own, are documented in **Supplementary File 2**.

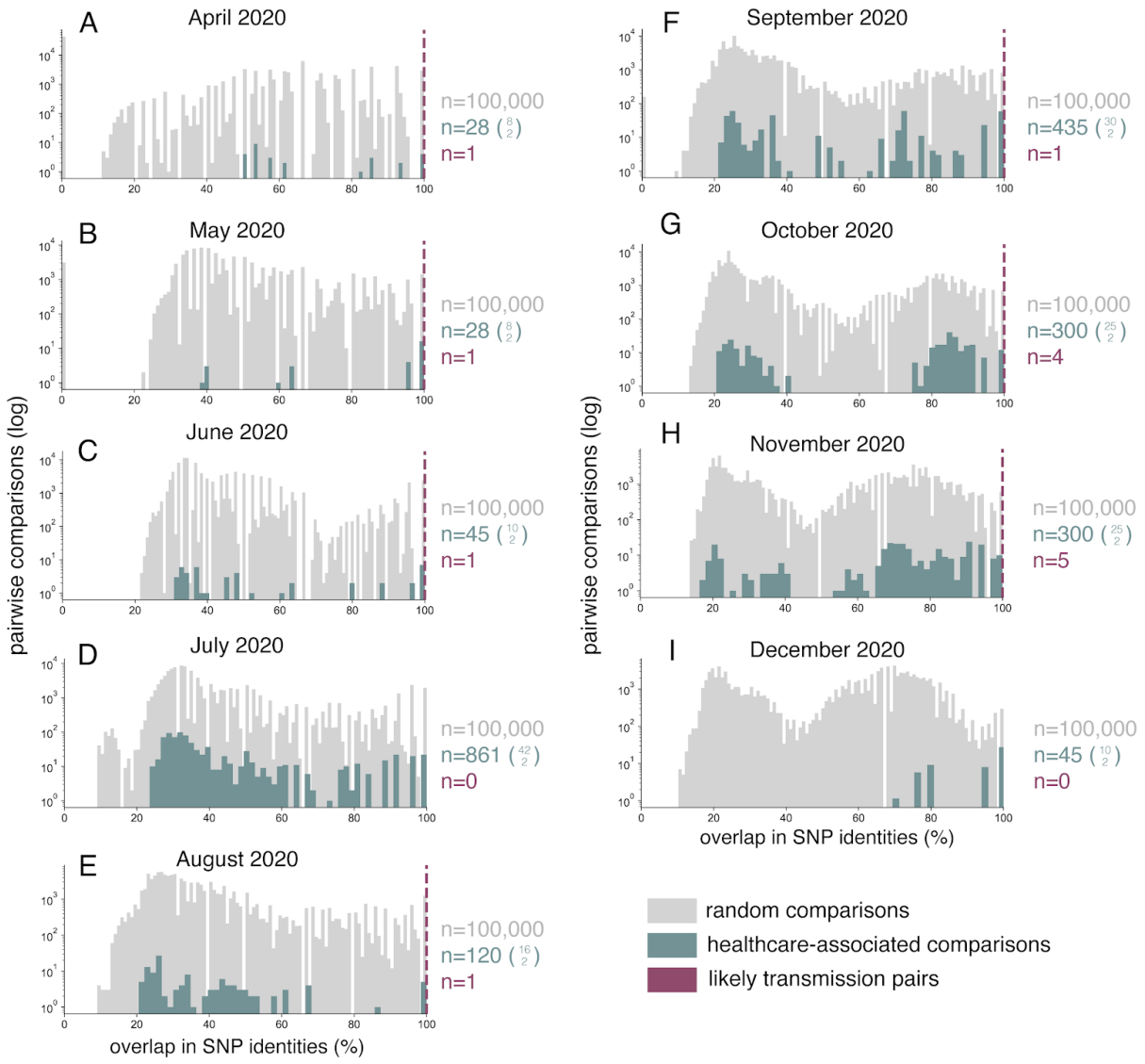


Figure 5.

Genetic diversity among pairwise comparisons of healthcare-associated viruses (HCP and patient samples) are generally similar to that of viruses circulating in the areas surrounding the academic medical center evaluated in this study. The grey distribution reflects 100,000 pairwise random comparisons of the community dataset per month (A-I). The turquoise distribution shows n-choose-2 comparisons from the healthcare-associated dataset per month where n is the total number of HCP and patient sequences

available within each month. The magenta dashed lines reflect the shared genetic diversity in healthcare-associated pairs where we have high confidence, based on epidemiological data, that these are true transmission pairs. The number of pairs represented in each magenta line is shown in magenta text to the right of each plot.

Likely source of infection in HCP	Number of cases
No evidence of healthcare-associated transmission	55 (57.9%)
Combined patient and employee cluster	12 (12.6%)
Inconclusive	11 (11.6%)
Employee source (via employee-employee interactions)	10 (10.5%)
Patient source (via employee-patient interactions)	4 (4.2%)
Outside community	3 (3.2%)
Total	95

Table 1.

Summary of the likely source of infection in the HCP evaluated in this study. Full definitions for each transmission bin can be found in **Supplemental File 1**. Briefly, “no evidence of healthcare-associated transmission” includes cases where available

sequences do not support transmission in the healthcare setting and “outside community” includes cases in which transmission outside the healthcare setting could be reasonably established. “Inconclusive” includes cases where no consensus sequence was available for the HCP and/or there were no appropriate comparator sequences.

Supplementary File 1

Likely sources of infection in HCP: definitions

Outside community. Among the sequences available for comparison, the likely source of infection was not a patient and was not a co-worker/employee.

Patient source. The most likely source of infection in the HCP was a patient source.

Employee source. The most likely source of infection in the HCP was a co-worker/employee.

Combined patient and employee cluster. A patient to HCP transmission event likely started this cluster and was followed by HCP-to-HCP transmission. However, we are unable to pinpoint the first HCP to become infected and/or are unable to distinguish ongoing sources of transmission as patient-to-HCP and HCP-to-HCP are both possible.

Inconclusive. No consensus sequence available, and/or there were no appropriate comparator sequences available, and/or epidemiological information were insufficient to interpret sequence data

Likely source of infection in HCP	Number of cases
Outside community	58 (60.4%)
Patient source (via employee-patient interactions)	4 (4.2%)
Employee source (via employee-employee interactions)	10 (10.4%)
Combined patient and employee cluster	12 (12.5)
Inconclusive	12 (12.5)
Total	96

Supplemental table 1.

Total number of patient comparator samples = 140 (96 consensus sequences).

Each case included in the above table is summarized below. For each case, we include the likely source of infection for all involved healthcare personnel (HCP). Next, we provide essential information for each associated sample in the form of a table, including sample collection date, GISAID identifier, Nextstrain clade, and Pangolin lineage. We report clades using the updated Nextstrain clade naming strategy as outlined by Bedford, Hodcroft, and Neher in Virological.org [1]. We report lineages according to the Pangolin nomenclature as outlined by Rambaut and colleagues [3]. A description of each Pangolin lineage can be found at cov-lineages [2]. Next, we provide a very brief overview of any known epidemiological interactions among the involved individuals. The level of epidemiological detail associated with each case is variable, but we have included all known information here. We include a simple alignment showing the consensus sequences mapped against the Wuhan-Hu-1 reference sequence.

Consensus-level differences amongst the reference and the sample sequences are denoted with a vertical black line. Particular variant identities for each sample can be found on the GitHub accompanying this manuscript. Finally, we include a time-resolved phylogenetic tree, built using Nextstrain algorithms, for each case. These trees include all sequences which are publicly available in the GISAID database from the state of Wisconsin. We highlight the samples involved in each case using bolded branches and nodes. An interactive view of this tree can be found here.

[1] https://virological.org/t/updated-nextstain-sars-cov-2-clade-naming-strategy/581

[2] https://cov-lineages.org/lineages.html

[3] Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020 Nov;5(11):1403-1407. doi: 10.1038/s41564-020-0770-5. Epub 2020 Jul 15. PMID: 32669681.

Report #1. 2020-04-15.

Likely source of HCP infection

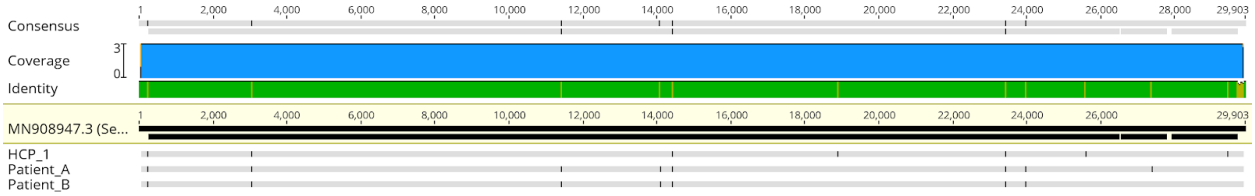
HCP 1. Outside community.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	April 2020	hCoV-19/USA/IA-UW-121/2020	20A	B.1
patient A	April 2020	hCoV-19/USA/WI-UW-120/2020	20A	B.1.19
patient B	April 2020	hCoV-19/USA/WI-UW-122/2020	20A	B.1.19

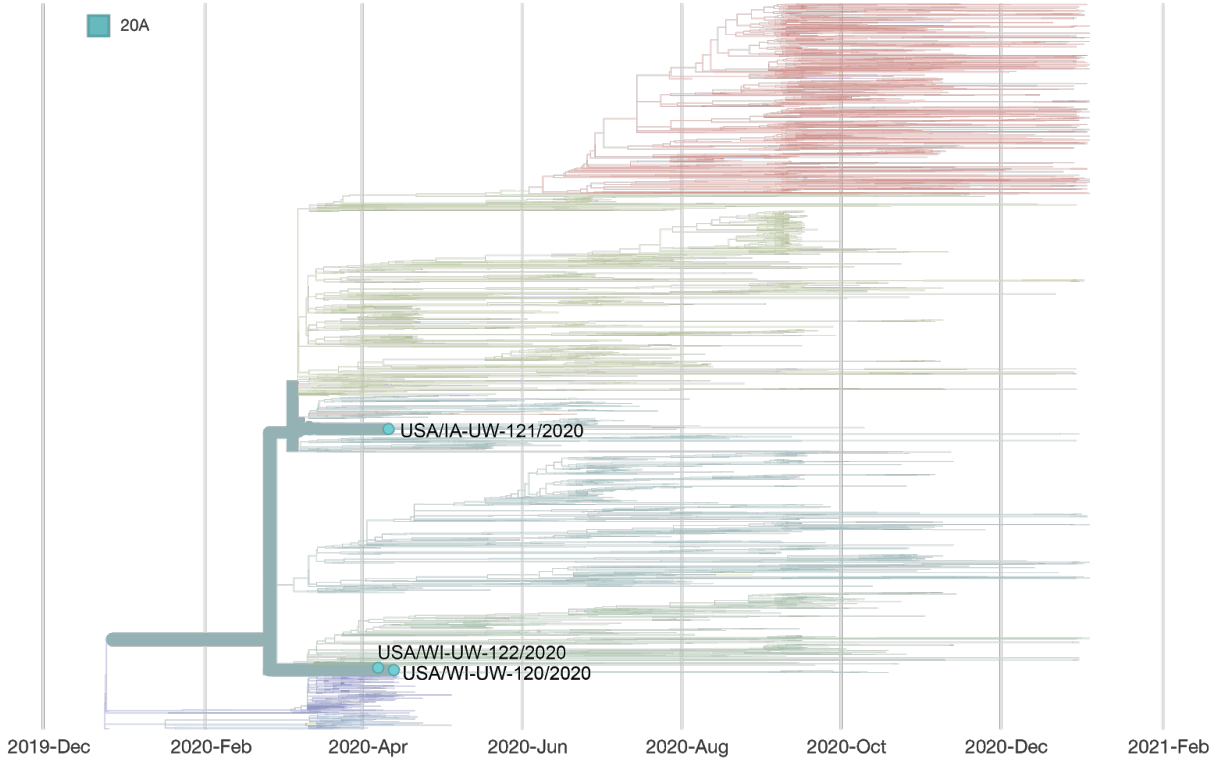
Epidemiological information

In the 14 days before symptom onset, HCP 1 interacted with patients A and B per comprehensive caregiver trace (see Methods, “Sample approvals and sample selection criteria” for further information). HCP 1 reported wearing appropriate personal protective equipment (PPE) while providing care to these patients

Alignment



Phylogeny



Report #2. 2020-04-16.

Likely source of HCP infection

HCP 1. Outside community (household contact).

HCP 2. Outside community.

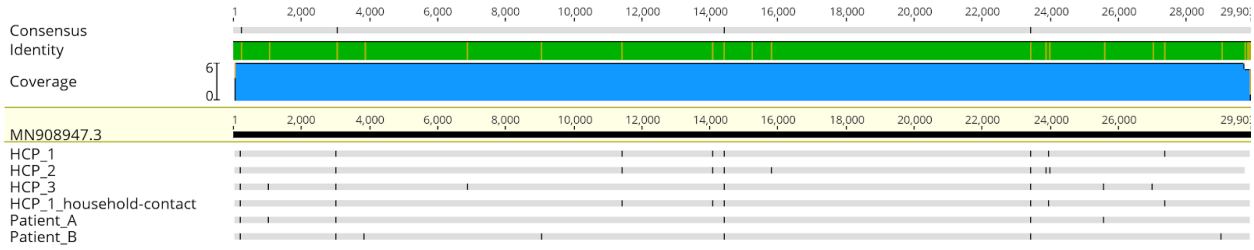
HCP 3. Outside community.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	April 2020	hCoV-19/USA/WI-UW-119/2020	20A	B.1.19
HCP 2	April 2020	hCoV-19/USA/WI-UW-259/2020	20A	B.1.19
HCP 3	April 2020	hCoV-19/USA/WI-UW-260/2020	20C	B.1
Household contact (HCP 1)	April 2020	hCoV-19/USA/WI-UW-120/2020	20A	B.1.19
patient A	March 2020	hCoV-19/USA/WI-UW-118/2020	20C	B.1
patient B	March 2020	hCoV-19/USA/WI-UW-110/2020	20A	B.1.139

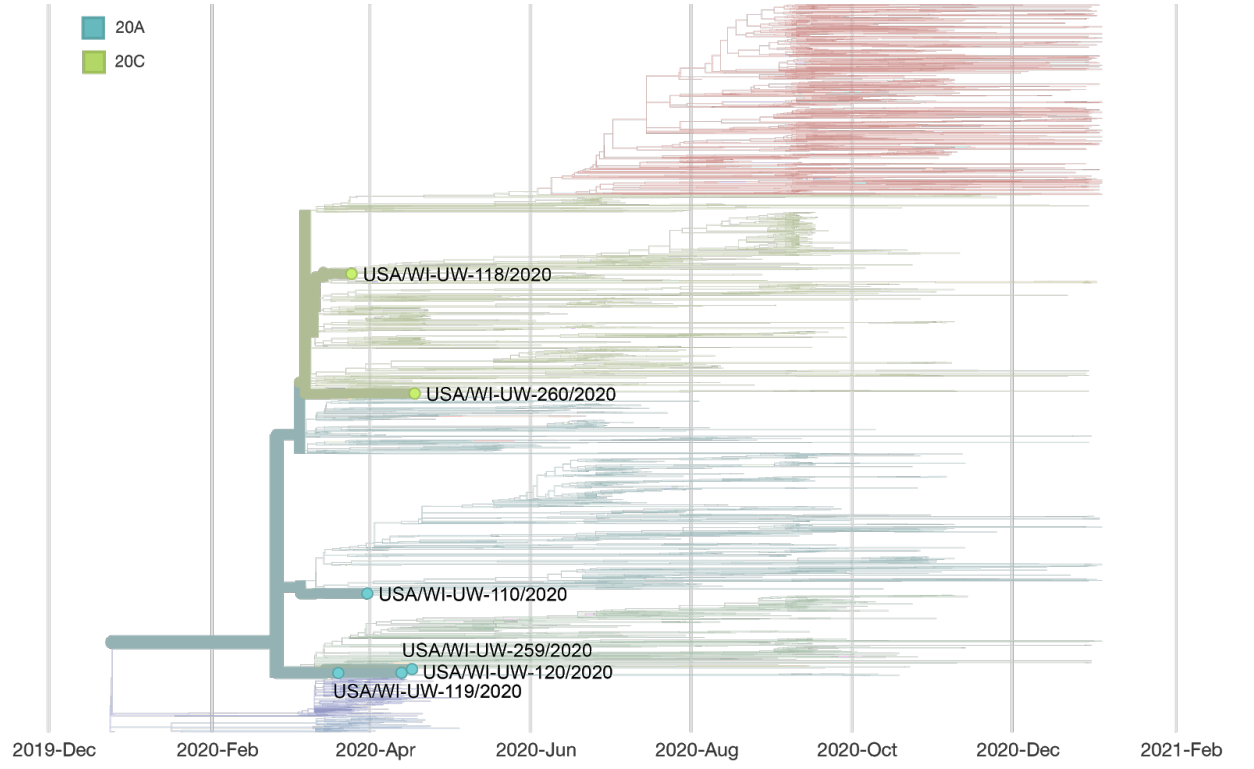
Epidemiological information

HCP 1 tested positive for SARS-CoV-2 after providing care for two SARS-CoV-2 positive patients, patients A and B. HCP 1 also had a household contact who tested positive for SARS-CoV-2 16 days before HCP 1. In the week following HCP 1's positive test, two additional HCP, HCP 2 and HCP 3 tested positive for SARS-CoV-2. HCPs 1, 2, and 3 all work in the same department, but we do not know whether these individuals had any high-risk contact with HCP 1 before their positive test results.

Alignment



Phylogeny



Notes

This case was published as an independent case report in Emerging Infectious Disease

224

Report #3. 2020-06-04.

Likely source of HCP infection

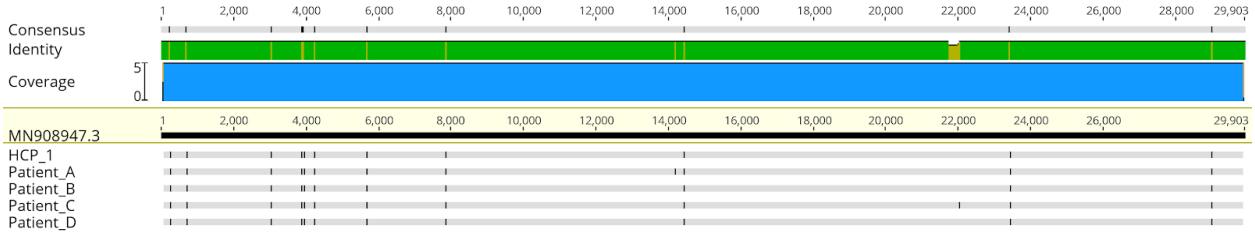
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	May 2020	hCoV-19/USA/WI-UW-388/2020	20A	B.1.139
HCP-C	Sample was not available			
patient A	May 2020	hCoV-19/USA/WI-UW-386/2020	20A	B.1.139
patient B	May 2020	hCoV-19/USA/WI-UW-390/2020	20A	B.1.139
patient C	May 2020	hCoV-19/USA/WI-UW-389/2020	20A	B.1.139
patient D	May 2020	hCoV-19/USA/WI-UW-387/2020	20A	B.1.139

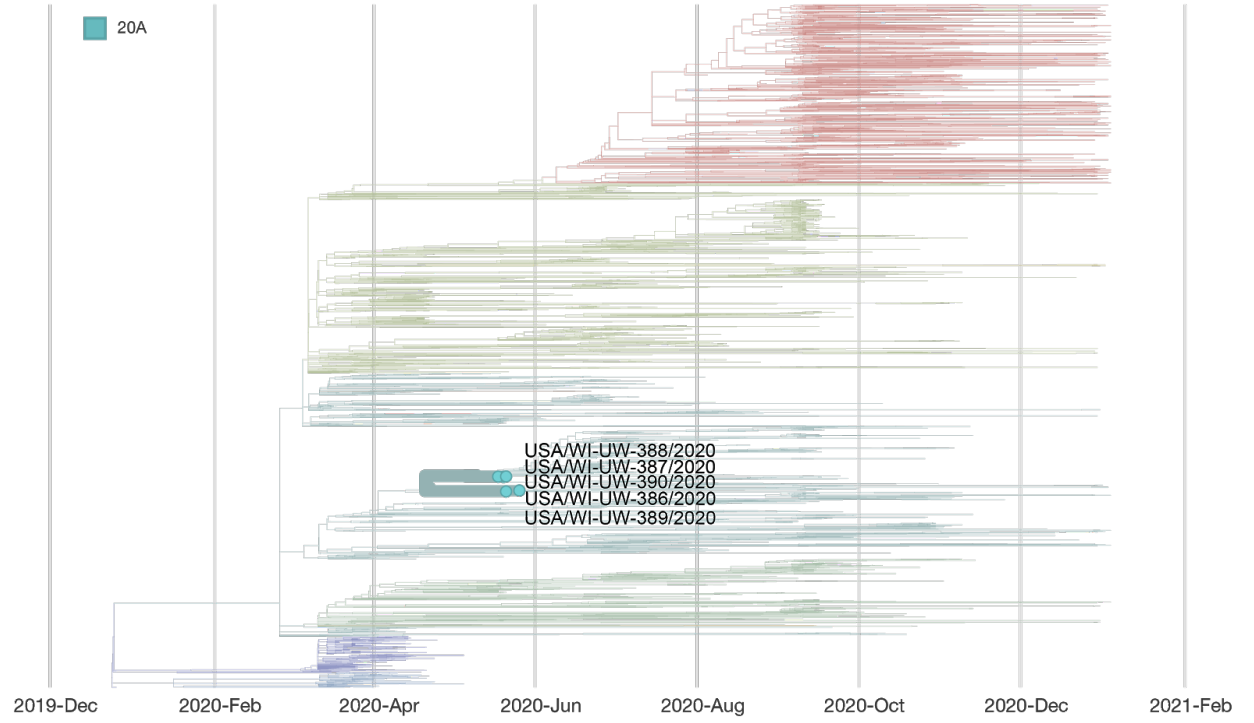
Epidemiological information

HCP 1 did not have direct contact with any of the patients included here. In this case, a household contact (HCP-C) of HCP 1 also works in healthcare, but in a different healthcare facility than HCP 1. The healthcare facility employing HCP-C was experiencing a COVID-19 outbreak at the time that HCP 1 tested positive. Patients A-D were patient samples collected from the HCP-C outbreak. A sample from the HCP-C was not available for comparison. Given the similarity in viral sequences between HCP 1 and all four patients from the outside healthcare facility, it is likely HCP 1 was exposed/infected via their household contact, who was likely exposed through patient contact.

Alignment



Phylogeny



Report #4. 2020-06-04.

Likely source of HCP infection

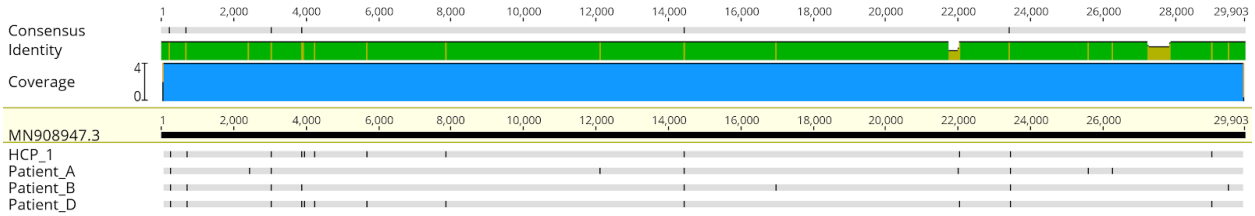
HCP 1. Patient source (patient D).

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	May 2020	hCoV-19/USA/WI-UW-391/2020	20A	B.1.139
patient A	May 2020	hCoV-19/USA/WI-UW-392/2020	20A	B.1.276
patient B	May 2020	hCoV-19/USA/WI-UW-393/2020	20A	B.1.139
patient C	May 2020	N/A - no consensus sequence		
patient D	May 2020	hCoV-19/USA/WI-UW-389/2020	20A	B.1.139

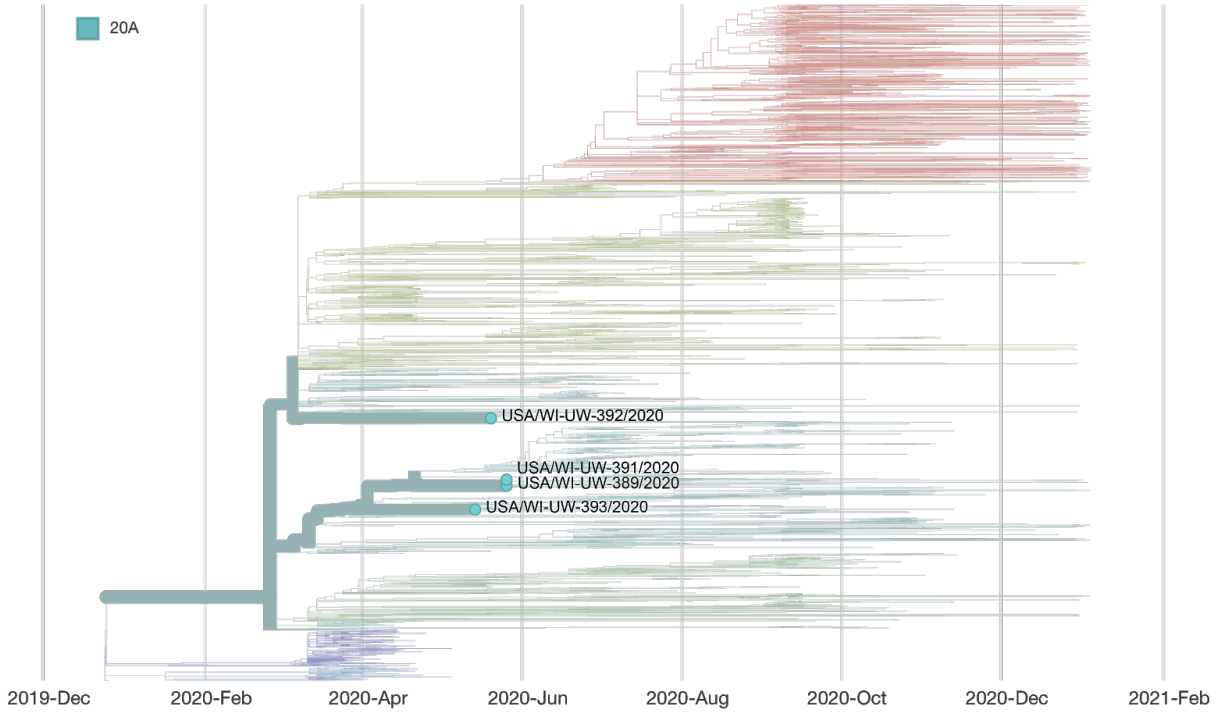
Epidemiological information

In the two weeks before symptom onset, HCP 1 provided direct care to patients A-D. HCP 1 wore appropriate PPE while providing care and was also present during patient D's treatment with the Aerobika nebulizer.

Alignment



Phylogeny



Notes

The Aerobika nebulizer was added to the list of aerosol-generating procedures requiring enhanced PPE based on this case.

Report #5. 2020-07-08.

Likely source of HCP infection

HCP 1. Outside community.

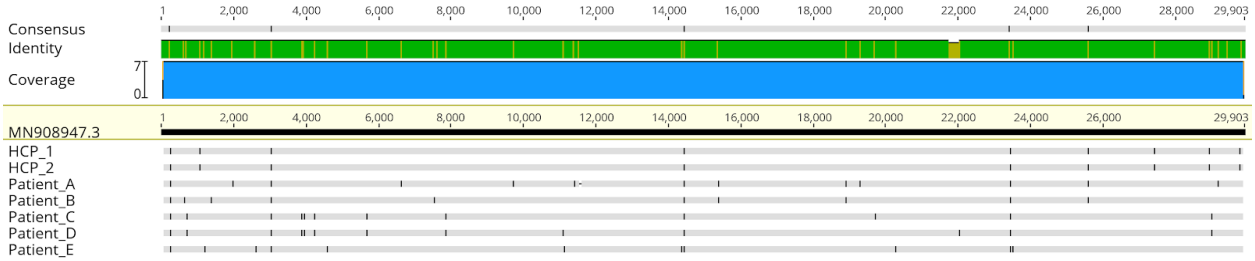
HCP 2. Outside community (likely source was HCP 1, but infection took place outside of the workplace).

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	June 2020	hCoV-19/USA/WI-UW-588/2020	20C	B.1
HCP 2	June 2020	hCoV-19/USA/WI-UW-610/2020	20C	B.1
patient A	June 2020	hCoV-19/USA/FL-UW-473/2020	20A	B.1.162
patient B	June 2020	hCoV-19/USA/WI-UW-469/2020	20A	B.1.162
patient C	June 2020	hCoV-19/USA/WI-UW-497/2020	20A	B.1.139
patient D	June 2020	hCoV-19/USA/WI-UW-493/2020	20A	B.1.139
patient E	June 2020	hCoV-19/USA/WI-UW-479/2020	20A	B.1.139
patient F	June 2020	N/A - no consensus sequence		

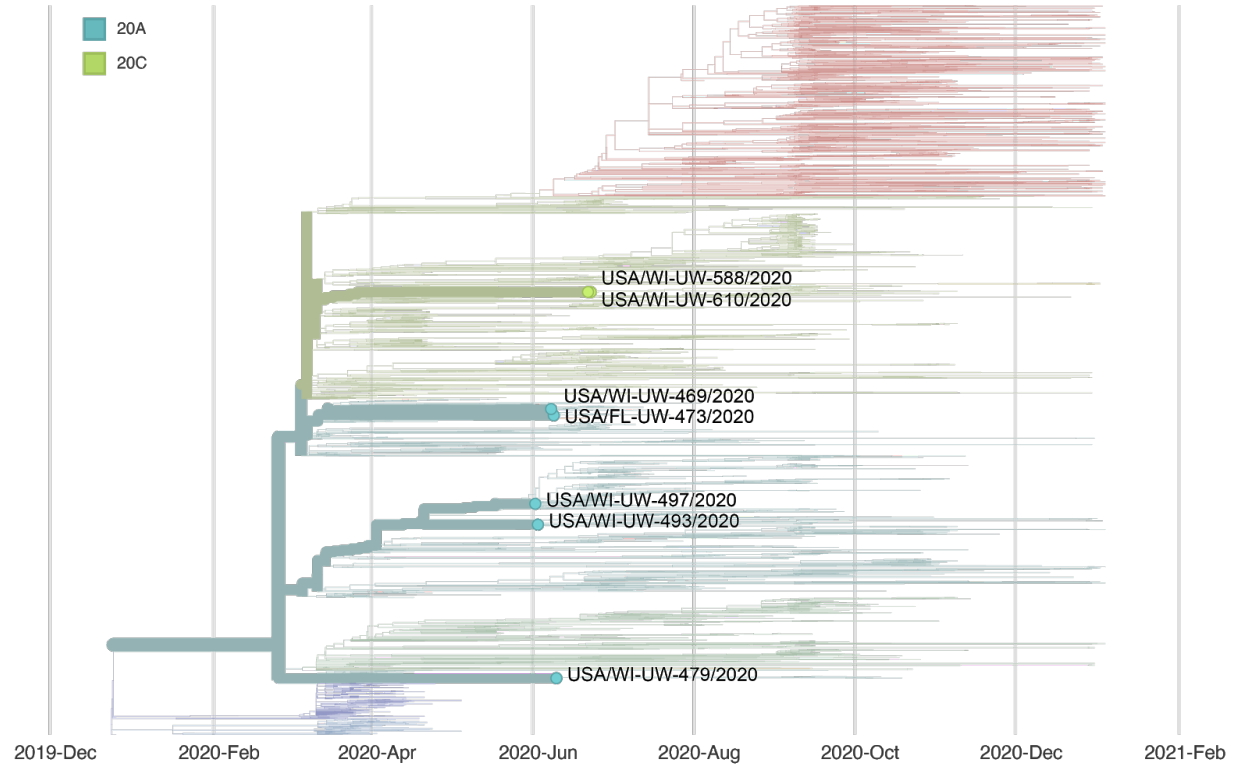
Epidemiological information

HCP 1 had direct contact while wearing appropriate PPE with patients A-E. HCP 1 and HCP 2 had unmasked interactions (>15 mins) with each other outside of the workplace. HCP 2 did not have contact with any SARS-CoV-2 positive patients in the 14 days before symptom onset.

Alignment



Phylogeny



Report #6. 2020-08-13.

Likely source of HCP infection

HCP 1. Outside community (based on epidemiological risk factors).

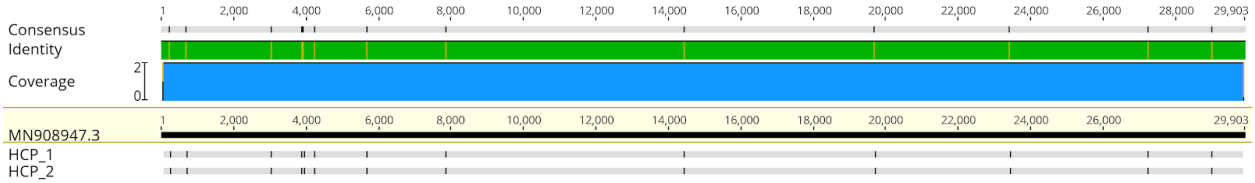
HCP 2. Employee source (likely source was HCP 1).

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	June 2020	hCoV-19/USA/WI-UW-520/2020	20A	B.1.139
HCP 2	June 2020	hCoV-19/USA/WI-UW-726/2020	20A	B.1.139

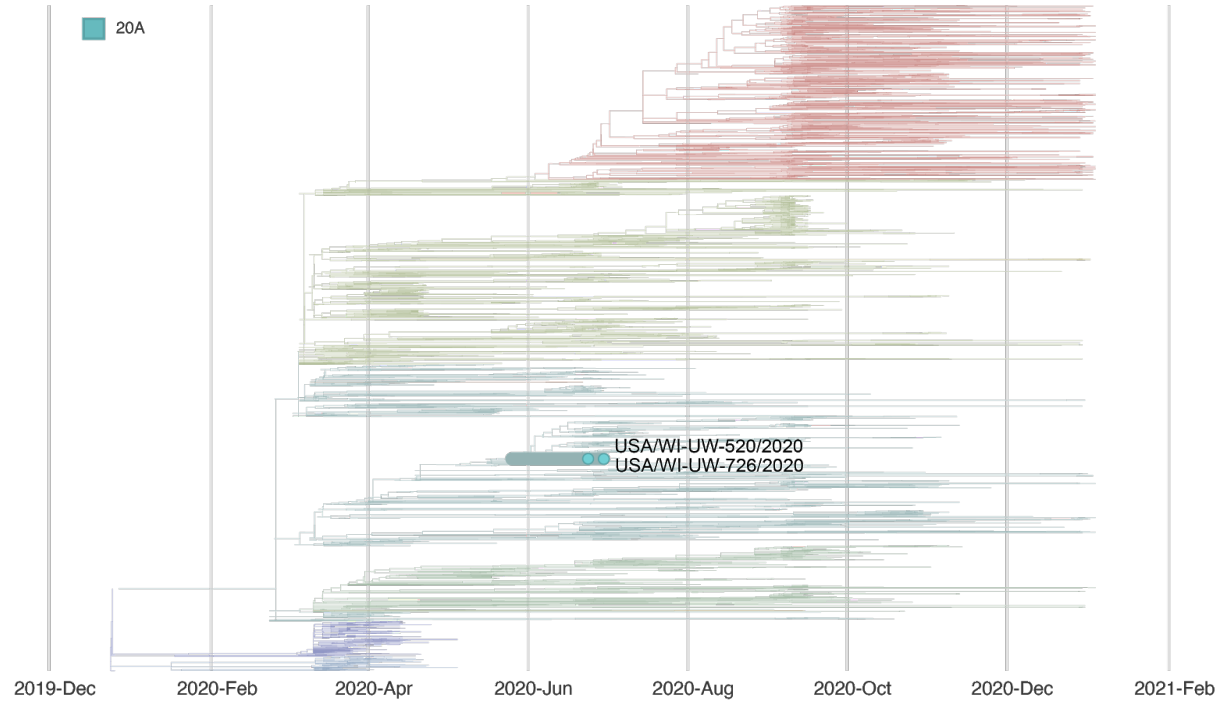
Epidemiological information

HCP 1 had a high-risk exposure event in the community before testing positive. This event was indoors, unmasked, and lasted longer than 15 minutes. HCP 1 works in the same department as HCP 2. Neither HCP 1 nor HCP 2 provided direct care to patients diagnosed with COVID-19 in the 14 days before their symptom onset. HCP 2 reported wearing a mask around all coworkers except while eating in the breakroom. HCP 2 reports removing their mask while eating, but maintaining a 6-foot physical distance from others during this time.

Alignment



Phylogeny



Report #7. 2020-08-04.

Likely source of HCP infection

HCP 1. Outside community.

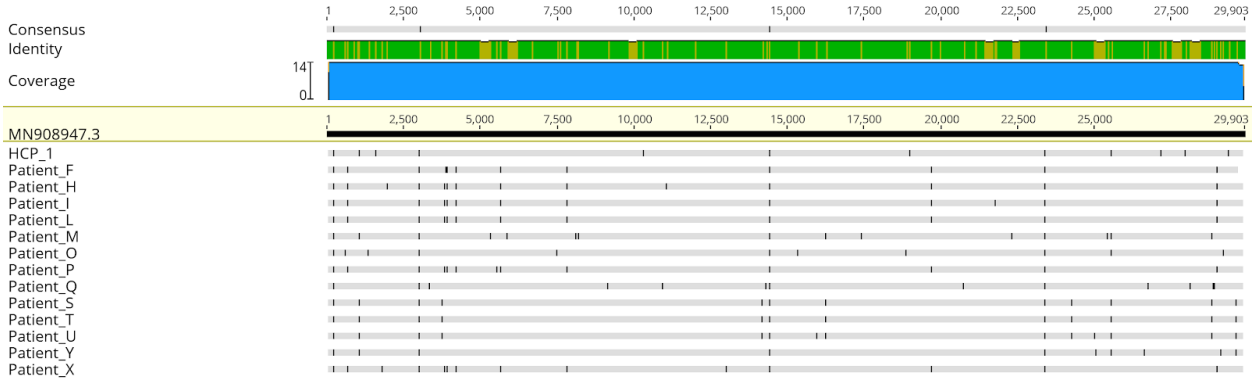
Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	July 2020	hCoV-19/USA/WI-UW-973/2020	20C	B.1.2
patient A	July 2020	N/A - no consensus sequence		
patient B	July 2020	N/A - no consensus sequence		
patient C	July 2020	N/A - no consensus sequence		
patient D	July 2020	N/A - no consensus sequence		
patient E	July 2020	N/A - no consensus sequence		
patient F	July 2020	hCoV-19/USA/WI-UW-852/2020	20A	B.1.139
patient G	July 2020	N/A - no consensus sequence		
patient H	July 2020	hCoV-19/USA/WI-UW-870/2020	20A	B.1.139
patient I	July 2020	hCoV-19/USA/WI-UW-847/2020	20A	B.1.139
patient J	July 2020	N/A - no consensus sequence		
patient K	July 2020	N/A - no consensus sequence		
patient L	July 2020	hCoV-19/USA/WI-UW-927/2020	20A	B.1.139
patient M	July 2020	hCoV-19/USA/WI-UW-878/2020	20C	B.1.369

patient N	July 2020	N/A - no consensus sequence		
patient O	July 2020	hCoV-19/USA/WI-UW-886/2020	20A	B.1.162
patient P	July 2020	hCoV-19/USA/WI-UW-880/2020	20A	B.1.139
patient Q	July 2020	hCoV-19/USA/WI-UW-894/2020	20B	B1.1
patient R	July 2020	N/A - no consensus sequence		
patient S	July 2020	hCoV-19/USA/WI-UW-895/2020	20C	B.1.1369
patient T	July 2020	hCoV-19/USA/WI-UW-876/2020	20C	B.1.369
patient U	July 2020	hCoV-19/USA/WI-UW-931/2020	20C	B.1.369
patient V	July 2020	N/A - no consensus sequence		
patient W	July 2020	N/A - no consensus sequence		
patient X	July 2020	hCoV-19/USA/WI-UW-898/2020	20A	B.1.369
patient Y	July 2020	hCoV-19/USA/WI-UW-883/2020	20C	B.1
patient Z	July 2020	N/A - no consensus sequence		
patient AA	July 2020	N/A - no consensus sequence		

Epidemiological information

HCP 1 collected nasopharyngeal specimens from patients with suspected COVID-19. HCP 1 wore appropriate PPE while collecting these specimens and reported no breach in PPE. Patients A-AA were collected in the 14 days before symptom onset in HCP 1.

Alignment



Phylogeny

Report #8. 2020-08-18.

Likely source of HCP infection

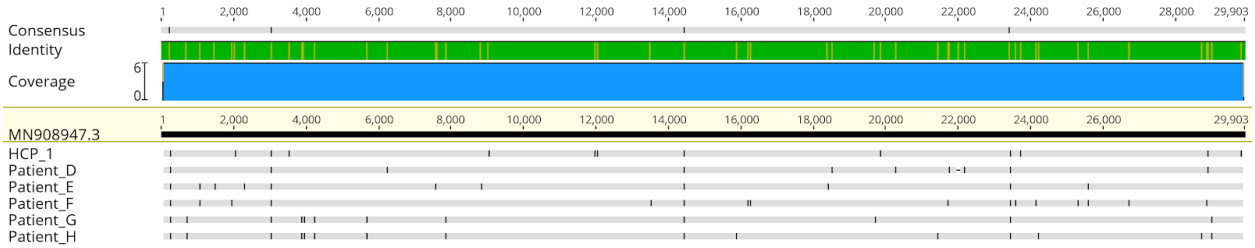
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	July 2020	hCoV-19/USA/WI-UW-1022/2020	20B	B.1.1.73
patient A	July 2020	N/A - no consensus sequence		
patient B	Unknown	N/A - no consensus sequence		
patient C	July 2020	N/A - no consensus sequence		
patient D	July 2020	hCoV-19/USA/WI-UW-923/2020	20A	B.1.240
patient E	June 2020	hCoV-19/USA/WI-UW-614/2020	20C	B.1.330
patient F	July 2020	hCoV-19/USA/WI-UW-875/2020	20C	B.1.369
patient G	July 2020	hCoV-19/USA/WI-UW-889/2020	20A	B.1.139
patient H	July 2020	hCoV-19/USA/WI-UW-1100/2020	20A	B.1.139

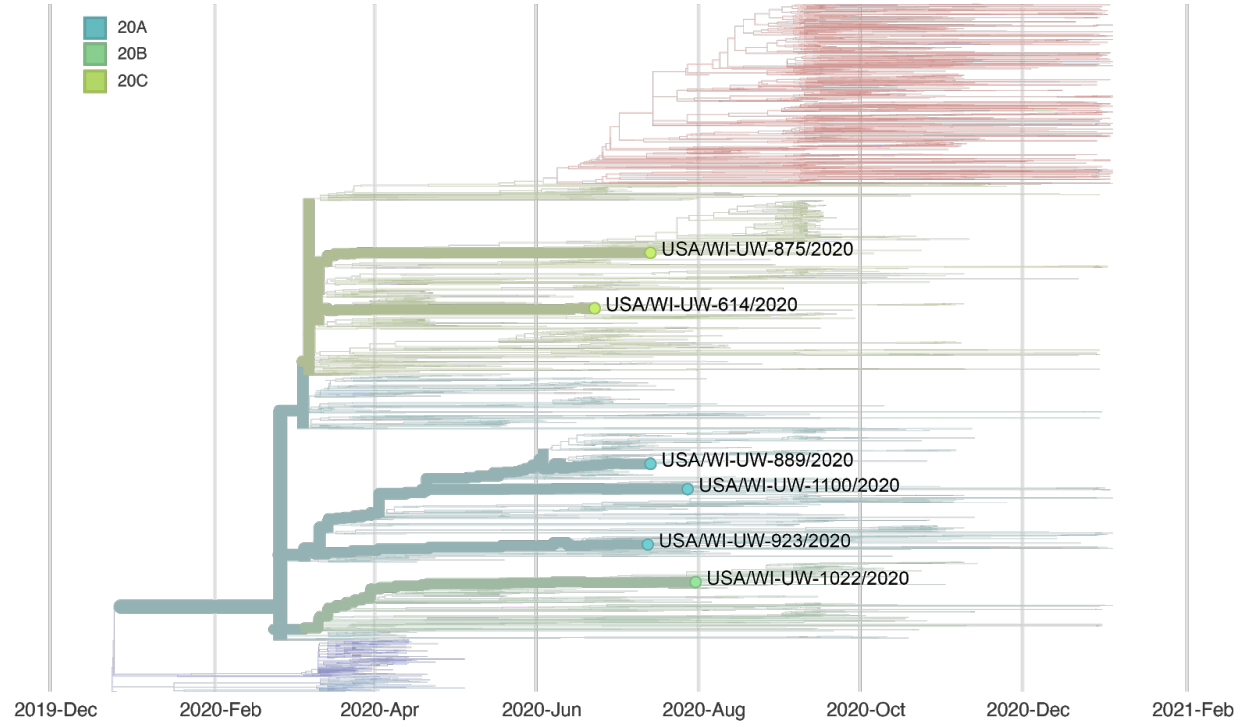
Epidemiological information

HCP 1 likely did not have direct interactions with any of the patients listed here. HCP 1 did, however, perform cleaning duties in the rooms of each of these patients in the 14 days before symptom onset.

Alignment



Phylogeny



Report #9. 2020-08-21.

Likely source of HCP infection

HCP 1. Outside community.

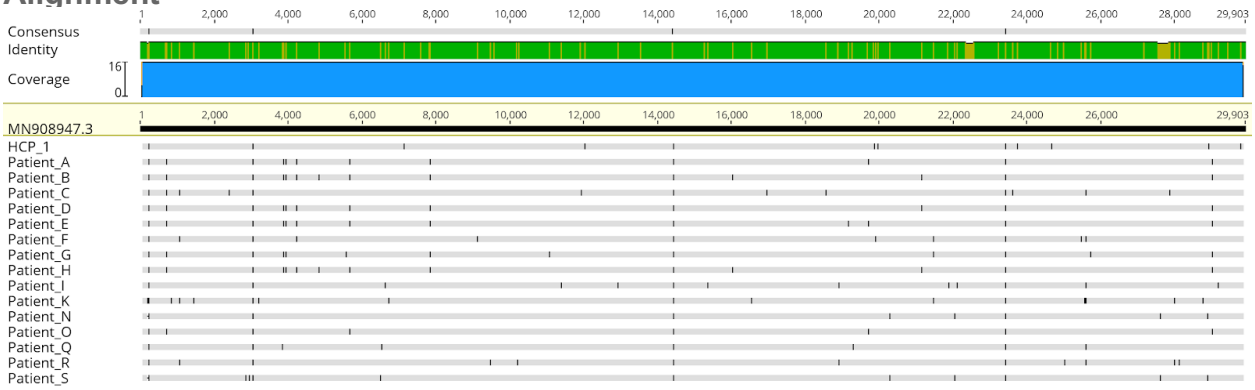
Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	July 2020	hCoV-19/USA/WI-UW-965/2020	20B	B.1.1.73
patient A	July 2020	hCoV-19/USA/WI-UW-765/2020	20A	B.1.139
patient B	July 2020	hCoV-19/USA/WI-UW-792/2020	20A	B.1.139
patient C	July 2020	hCoV-19/USA/WI-UW-761/2020	20C	B.1.3
patient D	July 2020	hCoV-19/USA/WI-UW-781/2020	20A	B.1.139
patient E	July 2020	hCoV-19/USA/WI-UW-794/2020	20A	B.1.139
patient F	July 2020	hCoV-19/USA/WI-UW-796/2020	20C	B.1.294
patient G	July 2020	hCoV-19/USA/WI-UW-790/2020	20A	B.1.139
patient H	July 2020	hCoV-19/USA/WI-UW-768/2020	20A	B.1.139
patient I	July 2020	hCoV-19/USA/WI-UW-803/2020	20A	B.1.162
patient J	July 2020	N/A - no consensus sequence		
patient K	July 2020	hCoV-19/USA/MO-UW-771/2020	20C	B.1.370
patient L	July 2020	N/A - no consensus sequence		
patient M	July 2020	N/A - no consensus sequence		

patient N	July 2020	hCoV-19/USA/WI-UW-784/2020	20A	B.1
patient O	July 2020	hCoV-19/USA/WI-UW-787/2020	20A	B.1.139
patient P	July 2020	N/A - no consensus sequence		
patient Q	July 2020	hCoV-19/USA/WI-UW-786/2020	20A	B.1.112
patient R	July 2020	hCoV-19/USA/WI-UW-766/2020	20C	B.1.370
patient S	July 2020	hCoV-19/USA/WI-UW-798/2020	20A	B.1

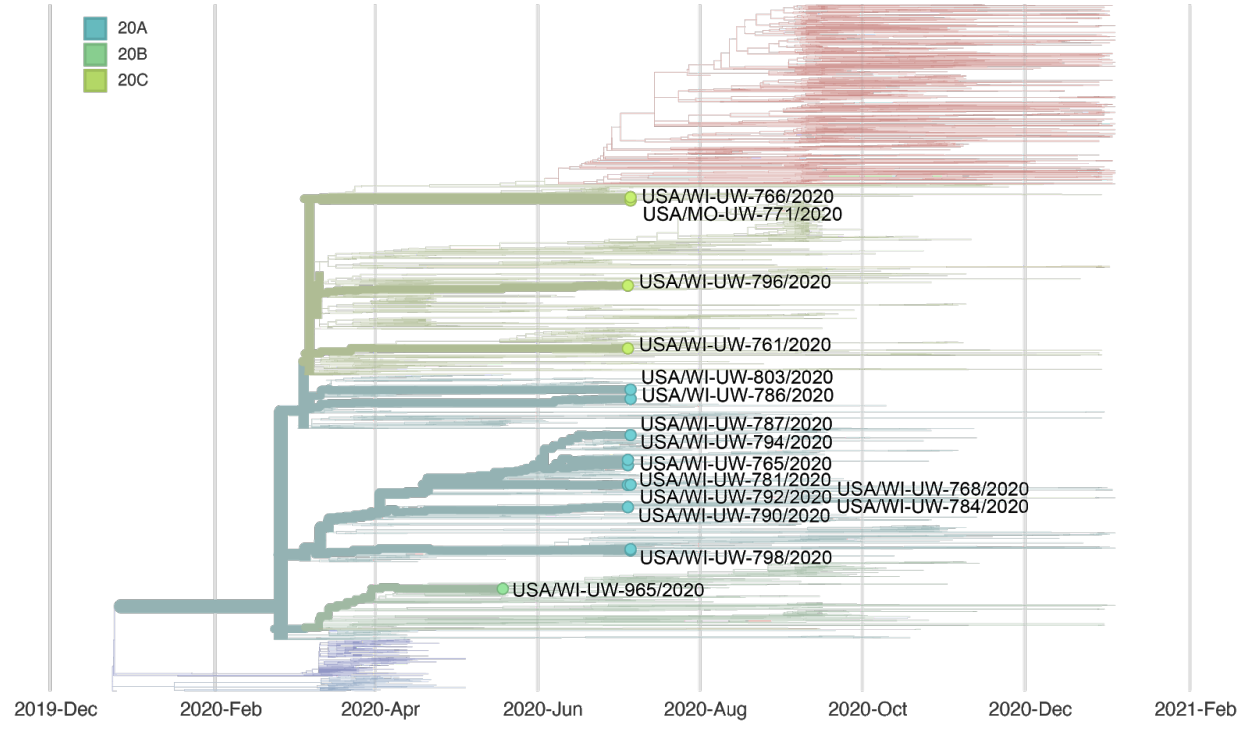
Epidemiological information

HCP 1 collected nasopharyngeal specimens from patients with suspected COVID-19. HCP 1 wore appropriate PPE while collecting these specimens and reported no breach in PPE. Patients A-S were collected in the 14 days before symptom onset in HCP 1.

Alignment



Phylogeny



Report #10. 2020-08-26.

Likely source of HCP infection

HCP 1. Patient source (patient G).

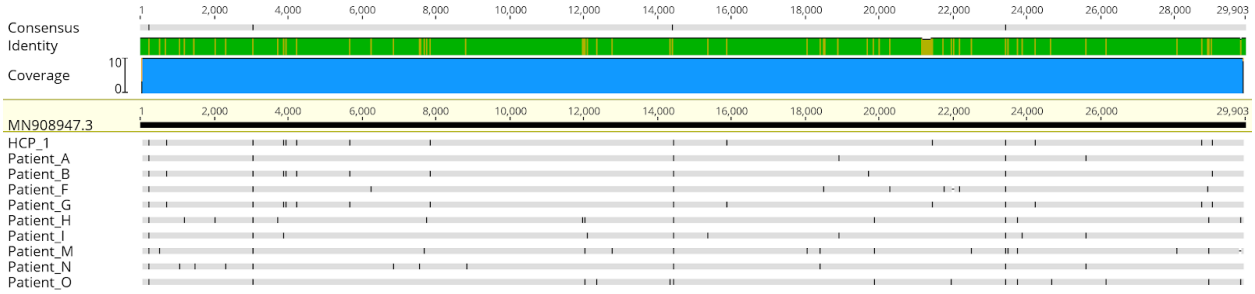
Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	August 2020	hCoV-19/USA/WI-UW-1125/2020	20A	B.1.139
patient A	August 2020	hCoV-19/USA/WI-UW-867/2020	20A	B.1.255
patient B	July 2020	hCoV-19/USA/WI-UW-889/2020	20A	B.1.139
patient C	July 2020	N/A - no consensus sequence		
patient D	July 2020	N/A - no consensus sequence		
patient E	August 2020	N/A - no consensus sequence		
patient F	July 2020	hCoV-19/USA/WI-UW-923/2020	20A	B.1.240
patient G	July 2020	hCoV-19/USA/WI-UW-1100/2020	20A	B.1.139
patient H	July 2020	hCoV-19/USA/WI-UW-998/2020	20B	B.1.1.73
patient I	August 2020	hCoV-19/USA/WI-UW-1035/2020	20A	B.1.162
patient J	July 2020	N/A - no consensus sequence		
patient K	July 2020	N/A - no consensus sequence		
patient L	July 2020	N/A - no consensus sequence		
patient M	July 2020	hCoV-19/USA/WI-UW-1054/2020	20B	B.1.1.73

patient N	July 2020	hCoV-19/USA/WI-UW-805/2020	20C	B.1.330
patient O	July 3030	hCoV-19/USA/WI-UW-1090/2020	20B	B.1.1.73

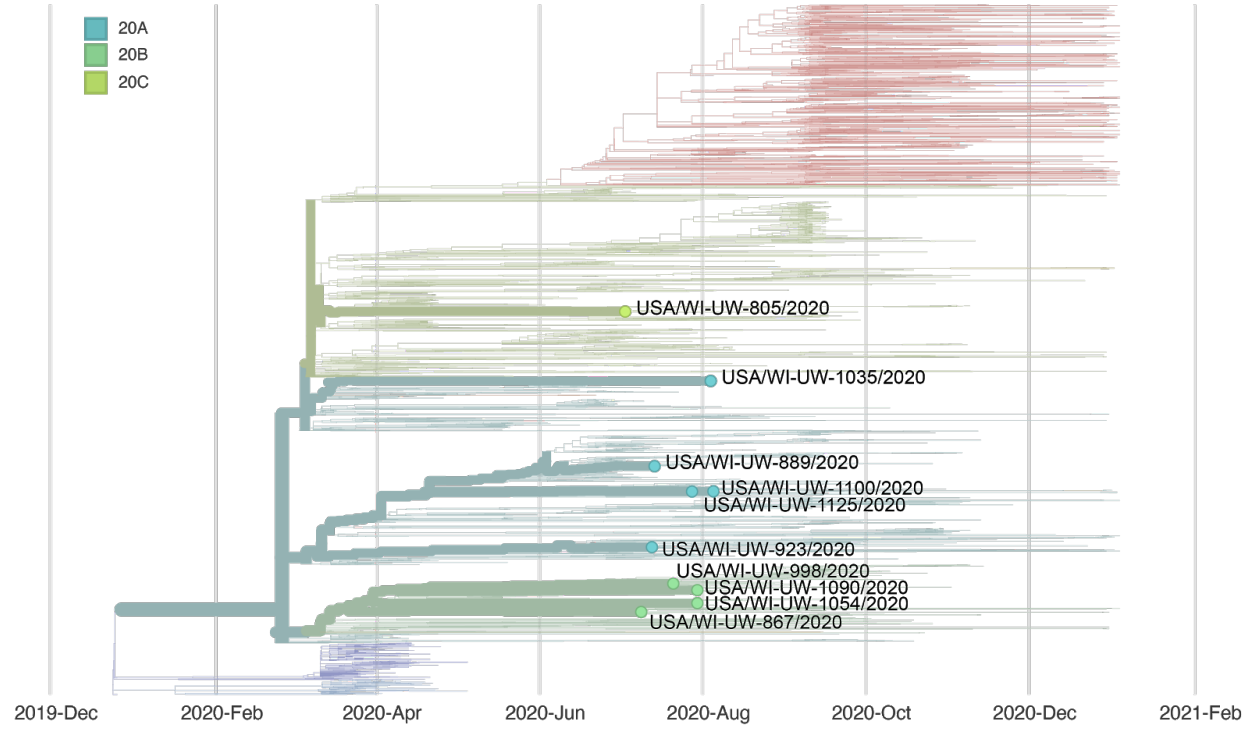
Epidemiological information

HCP 1 had direct contact while wearing appropriate PPE with patients A-O in the 14 days before symptom onset.

Alignment



Phylogeny



Report #11. 2020-09-11.

Likely source of HCP infection

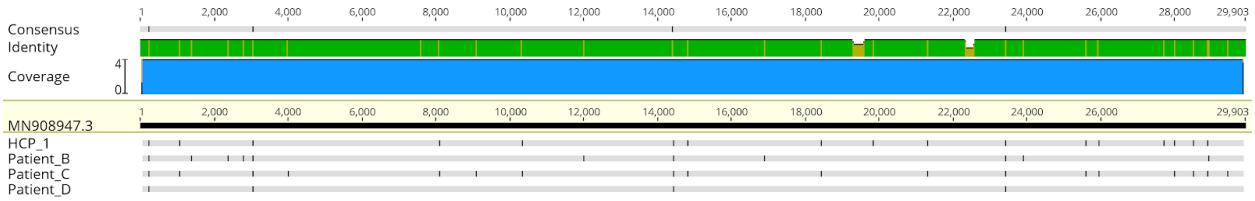
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	August 2020	hCoV-19/USA/WI-UW-1213/2020	20C	B.1.2
patient A	August 2020	N/A - no consensus sequence		
patient B	August 2020	hCoV-19/USA/WI-UW-1163/2020	20B	B.1.1.130
patient C	August 2020	hCoV-19/USA/WI-UW-1240/2020	20C	B.1.2
patient D	August 2020	hCoV-19/USA/WI-UW-1170/2020	20A	B.1.5

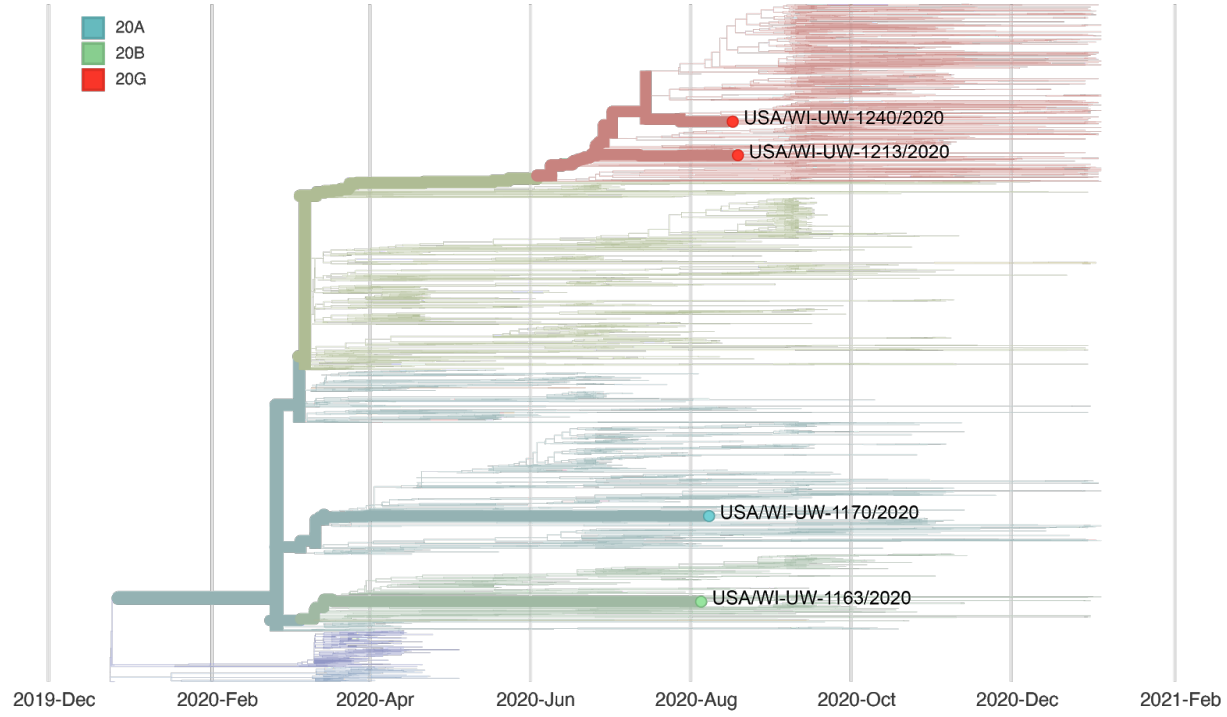
Epidemiological information

HCP 1 had direct contact while wearing appropriate PPE with patients A-D in the 14 days before symptom onset. HCP 1 denied any lapses in PPE and did not interact with coworkers without a surgical mask on.

Alignment



Phylogeny


20A
20B
20G
USA/WI-UW-1240/2020
USA/WI-UW-1213/2020
USA/WI-UW-1170/2020
USA/WI-UW-1163/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #12. 2020-09-11.

Likely source of HCP infection

HCP 1. Inconclusive.

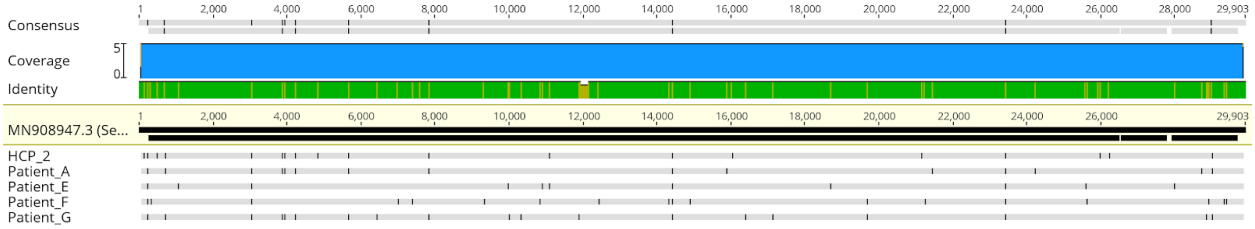
HCP 2. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	August 2020	N/A - no consensus sequence		
HCP 2	August 2020	hCoV-19/USA/WI-UW-1278/2020	20A	B.1.139
patient A	July 2020	hCoV-19/USA/WI-UW-1100/2020	20A	B.1.139
patient B	July 2020	N/A - no consensus sequence		
patient C	July 2020	N/A - no consensus sequence		
patient D	August 2020	N/A - no consensus sequence		
patient E	July 2020	hCoV-19/USA/WI-UW-850/2020	20C	B.1.2
patient F	August 2020	hCoV-19/USA/WI-UW-1248/2020	20B	B.1.1.244
patient G	August 2020	hCoV-19/USA/WI-UW-1204/2020	20A	B.1.139
patient H	August 2020	N/A - no consensus sequence		

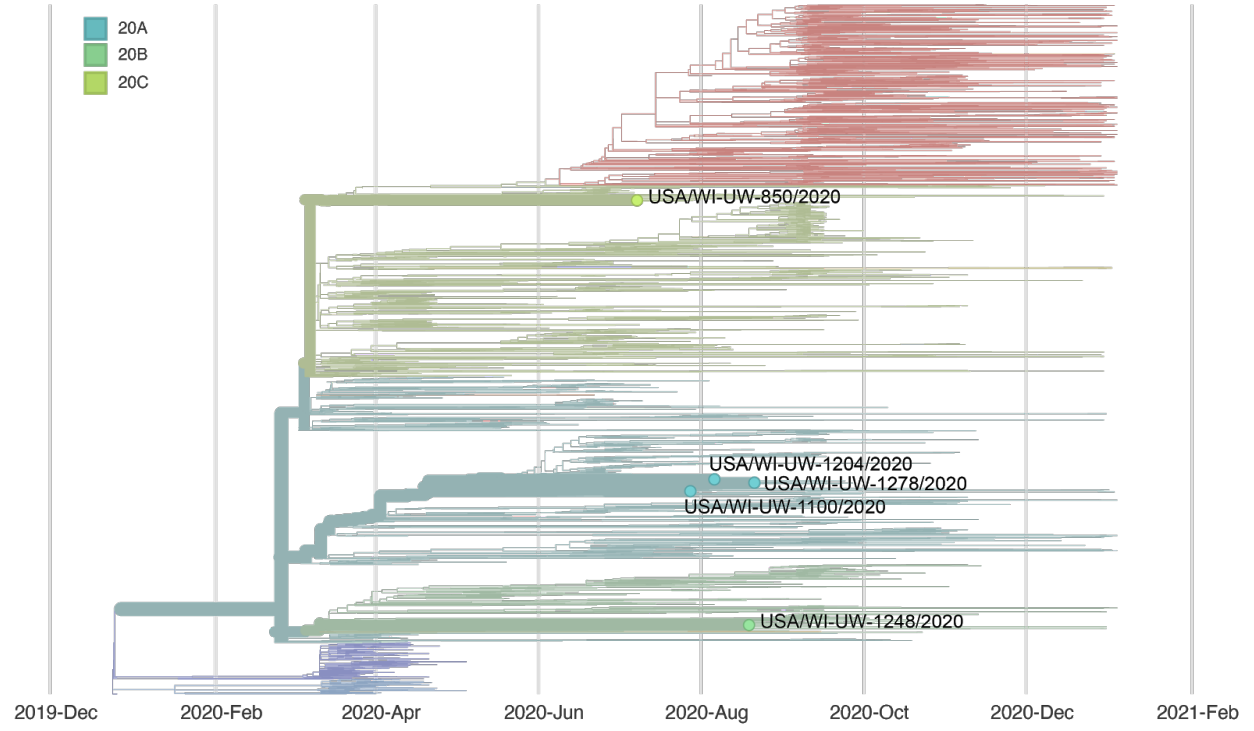
Epidemiological information

HCP 1 and HCP 2 are household contacts and one or both of these HCP provided direct patient care to patients A-H while wearing appropriate PPE in the two weeks before their symptom onset.

Alignment



Phylogeny



Report #13. 2020-09-14.

Likely source of HCP infection

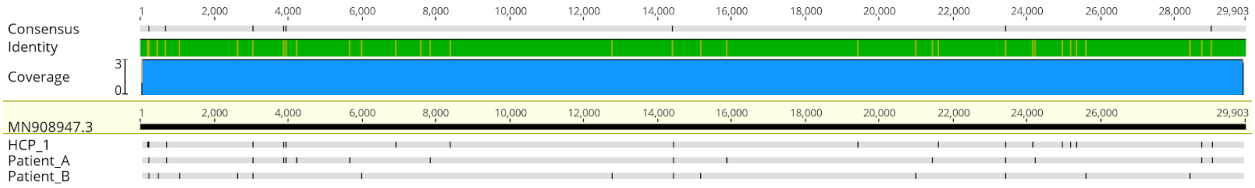
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	August 2020	hCoV-19/USA/WI-UW-1305/2020	20A	B.1.139
patient A	July 2020	hCoV-19/USA/WI-UW-1100/2020	20A	B.1.139
patient B	August 2020	hCoV-19/USA/WI-UW-1231/2020	20C	B.1.337
patient C	August 2020	N/A - no consensus sequence		
patient D	August 2020	N/A - no consensus sequence		

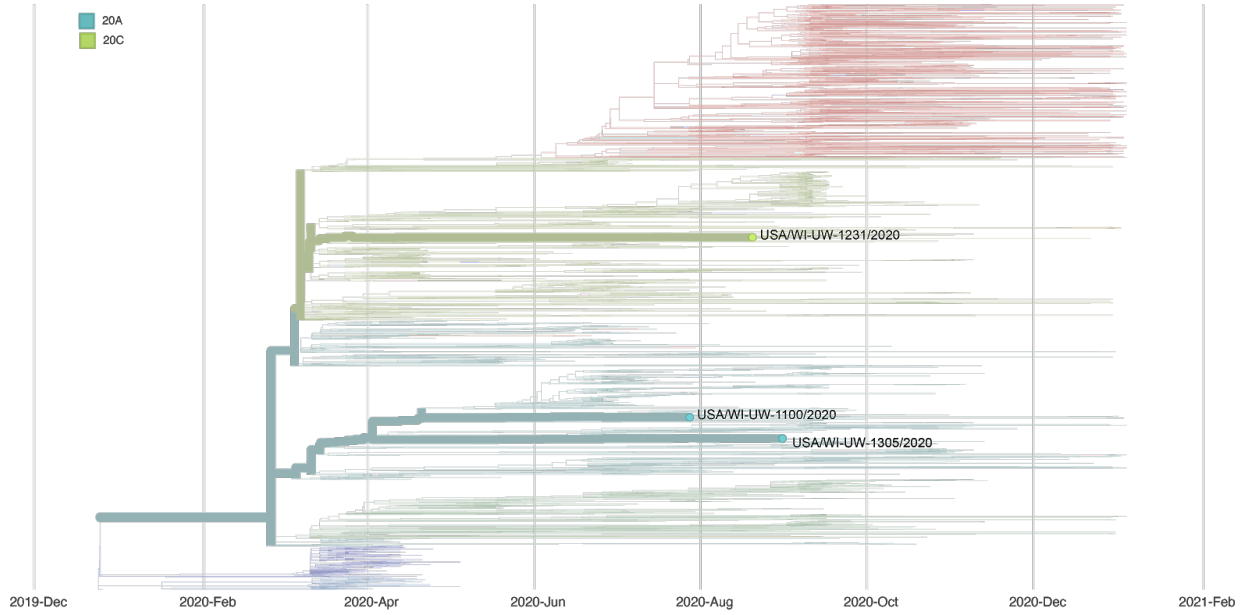
Epidemiological information

HCP 1 provided direct care while wearing appropriate PPE to patients A-D in the 14 days before symptom onset.

Alignment



Phylogeny



Report #14. 2020-09-22.

Likely source of HCP infection

HCP 1. Outside community.

HCP 2. Combined patient and employee cluster.

HCP 3. Inconclusive.

HCP 4. Combined patient and employee cluster.

HCP 5. Outside community.

HCP 6. Employee source (HCP 9 is a likely source of infection, although the HCP 6 → HCP 9 is also possible).

HCP 7. Outside community.

HCP 8. Inconclusive.

HCP 9. Outside community.

HCP 10. Inconclusive.

HCP 11. Combined patient and employee cluster.

HCP 12. Combined patient and employee cluster.

HCP 13. Combined patient and employee cluster.

HCP 14. Inconclusive.

HCP 15. Inconclusive.

HCP 16. Combined patient and employee cluster.

HCP 17. Combined patient and employee cluster.

HCP 18. Inconclusive.

HCP 19. Inconclusive.

HCP 20. Combined patient and employee cluster.

HCP 21. Inconclusive.

Patients A and C are the patients involved in this combined patient and employee cluster.

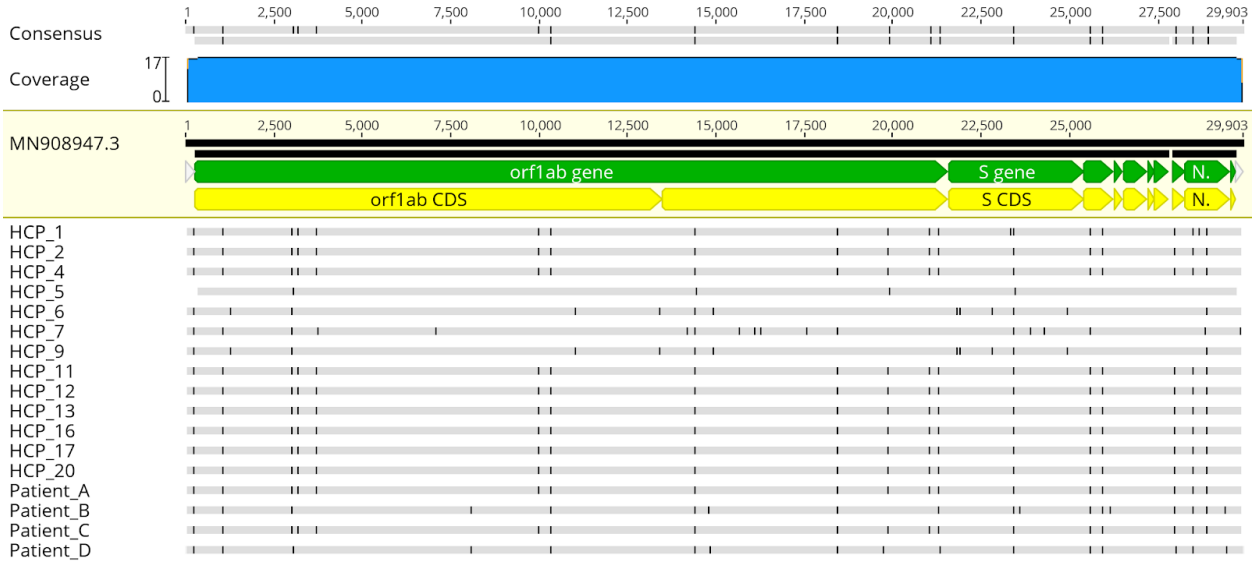
Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	September 2020	hCoV-19/USA/WI-UW-1348/2020	20G	B.1.2
HCP 2	September 2020	hCoV-19/USA/WI-UW-1344/2020	20G	B.1.2
HCP 3	September 2020	N/A - no consensus sequence		
HCP 4	September 2020	hCoV-19/USA/WI-UW-1412/2020	20G	B.1.2
HCP 5	September 2020	hCoV-19/USA/WI-UW-1325/2020	20G	B.1.5
HCP 6	September 2020	hCoV-19/USA/WI-UW-1480/2020	20B	B.1.1.251
HCP 7	September 2020	hCoV-19/USA/WI-UW-1477/2020	20G	B.1.369
HCP 8	September 2020	N/A - no consensus sequence		
HCP 9	September 2020	hCoV-19/USA/WI-UW-1475/2020	20B	B.1.1.251
HCP 10	September 2020	N/A - no consensus sequence		
HCP 11	September 2020	hCoV-19/USA/WI-UW-1441/2020	20G	B.1.2
HCP 12	September 2020	hCoV-19/USA/WI-UW-1445/2020	20G	B.1.2
HCP 13	September 2020	hCoV-19/USA/WI-UW-1452/2020	20G	B.1.2
HCP 14	September 2020	N/A - no consensus sequence		

HCP 15	September 2020	N/A - no consensus sequence		
HCP 16	September 2020	hCoV-19/USA/WI-UW-1459/2020	20G	B.1.2
HCP 17	September 2020	hCoV-19/USA/WI-UW-1436/2020	20G	B.1.2
HCP 18	September 2020	N/A - no consensus sequence		
HCP 19	September 2020	N/A - no consensus sequence		
HCP 20	September 2020	hCoV-19/USA/WI-UW-1461/2020	20G	B.1.2
HCP 21	September 2020	N/A - no consensus sequence		
patient A	September 2020	hCoV-19/USA/WI-UW-1406/2020	20G	B.1.2
patient B	September 2020	hCoV-19/USA/WI-UW-1499/2020	20G	B.1.2
patient C	September 2020	hCoV-19/USA/WI-UW-1900/2020	20G	B.1.2
patient D	September 2020	hCoV-19/USA/WI-UW-1941/2020	20G	B.1.2

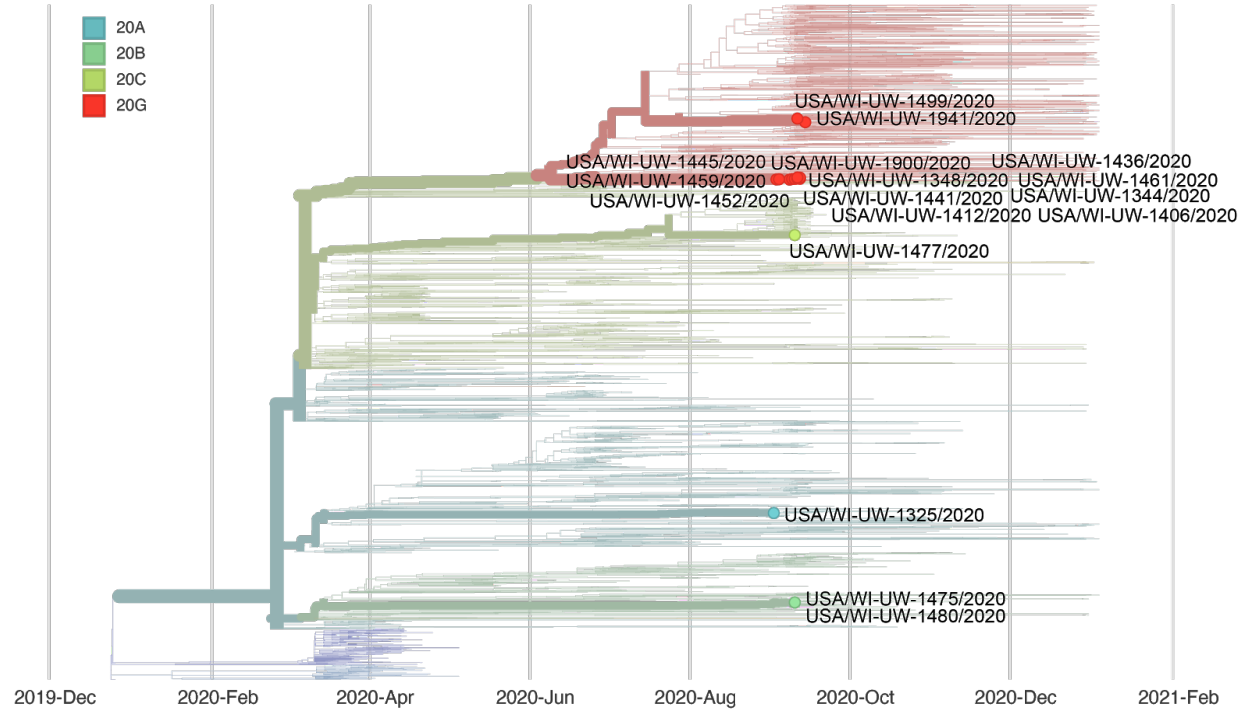
Epidemiological information

These HCP work in the same department. HCP 2, 4, 11, 12, 13, 16, and 20 provided direct care to patient A and/or patient C. HCP 17 and HCP 1 did not provide direct care to patient A or C.

Alignment



Phylogeny



Report #15. 2020-09-18.

Likely source of HCP infection

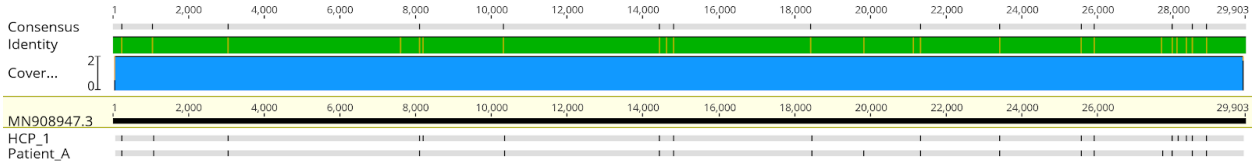
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	September 2020	hCoV-19/USA/WI-UW-1347/2020	20G	B.1.2
patient A	August 2020	hCoV-19/USA/WI-UW-1220/2020	20G	B.1.2

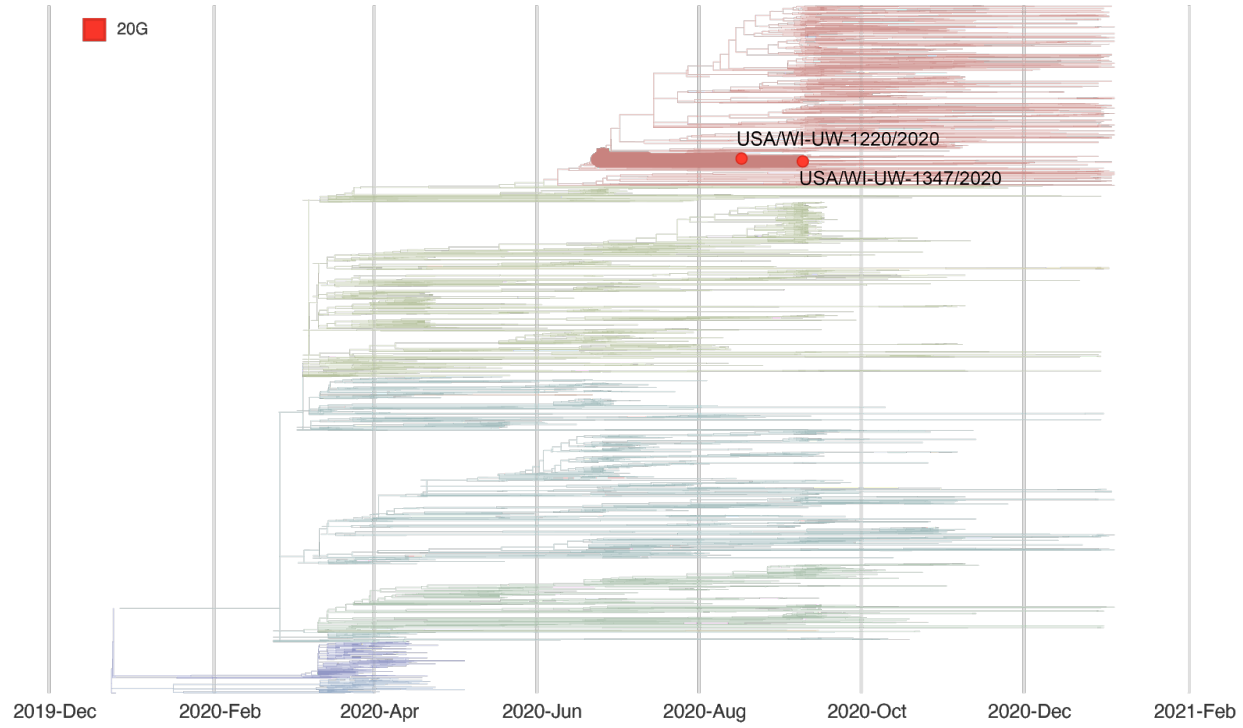
Epidemiological information

HCP 1 provided direct care to patient A while wearing appropriate PPE and with no reported lapses in PPE use.

Alignment



Phylogeny


20G
USA/WI-UW-1220/2020
USA/WI-UW-1347/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #16. 2020-10-29.

Likely source of HCP infection

HCP 1. Employee source (HCP 2).

HCP 2. Inconclusive.

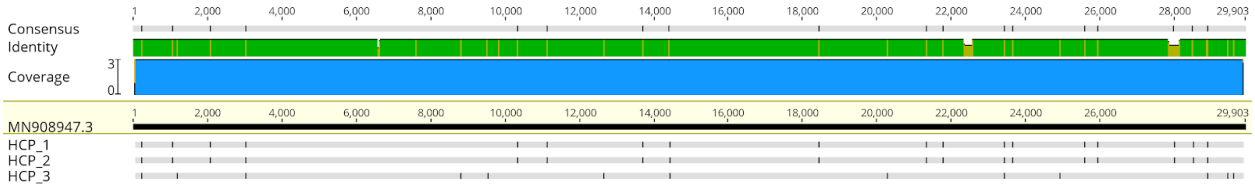
HCP 3. Outside community.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	September 2020	hCoV-19/USA/WI-UW-1901/2020	20G	B.1.2
HCP 2	September 2020	hCoV-19/USA/WI-UW-1350/2020	20G	B.1.2
HCP 3	September 2020	hCoV-19/USA/WI-UW-1898/2020	20A	B.1

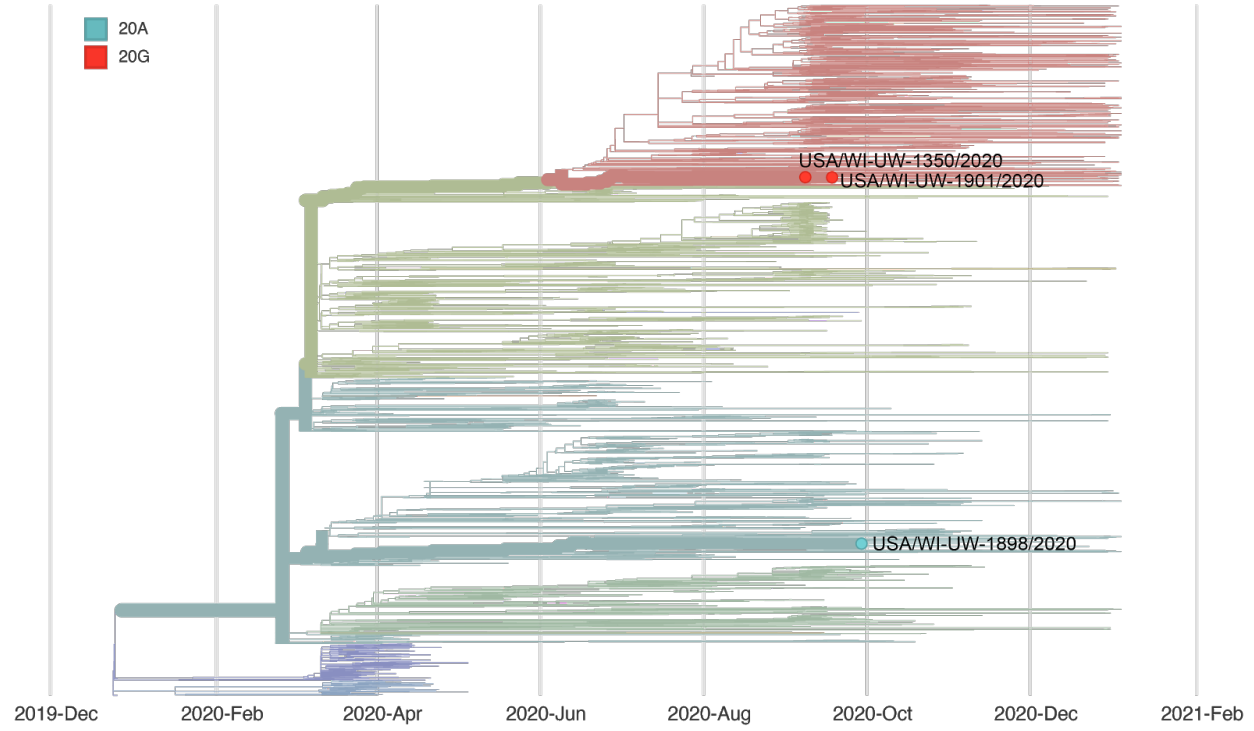
Epidemiological information

Contact tracing revealed HCP 2 worked for two days prior to symptom onset and may have had unmasked contact with HCP 1 during overlapping meal breaks. Contact tracing additionally revealed HCP 3 had a high-risk exposure even lasting >15 minutes in the outside community prior to testing positive.

Alignment



Phylogeny



Report #17. 2020-10-29.

Likely source of HCP infection

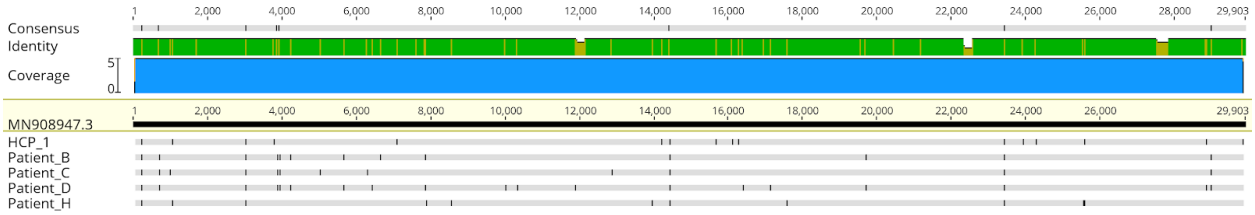
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	September 2020	hCoV-19/USA/WI-UW-1895/2020	20C	B.1.369
patient A	September 2020	N/A - no consensus sequence		
patient B	July 2020	hCoV-19/USA/WI-UW-774/2020	20A	B.1.139
patient C	August 2020	hCoV-19/USA/WI-UW-1198/2020	20A	B.1.139
patient D	August 2020	hCoV-19/USA/WI-UW-1166/2020	20A	B.1.139
patient E	August 2020	N/A - no consensus sequence		
patient F	September 2020	N/A - no consensus sequence		
patient G	August 2020	N/A - no consensus sequence		
patient H	August 2020	hCoV-19/USA/WI-UW-1301/2020	20C	B.1
patient I	August 2020	N/A - no consensus sequence		

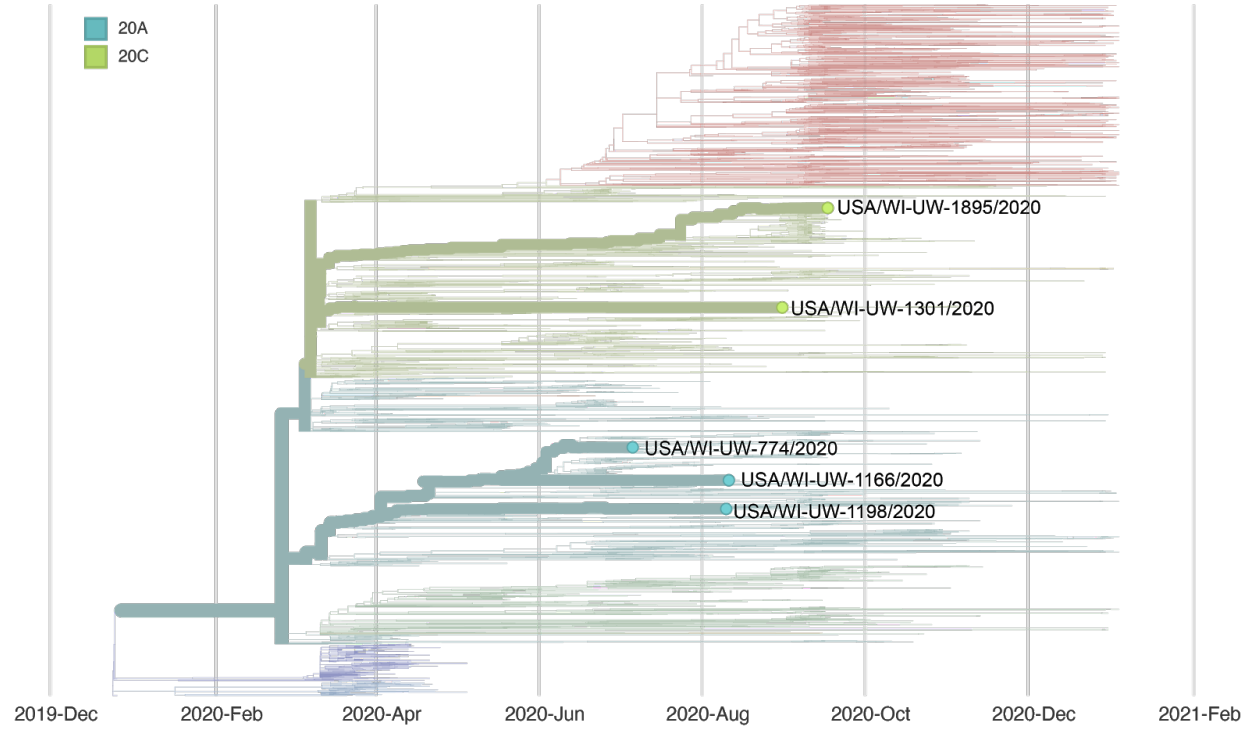
Epidemiological information

HCP 1 provided direct care to patients A-I while wearing appropriate PPE and with no reported lapses in PPE use. HCP 1 also had a household contact with confirmed SARS-CoV-2 infection who had symptoms onset prior to HCP 1’s symptom onset.

Alignment



Phylogeny



Report #18. 2020-10-29.

Likely source of HCP infection

HCP 1. Outside community.

HCP 2. Outside community.

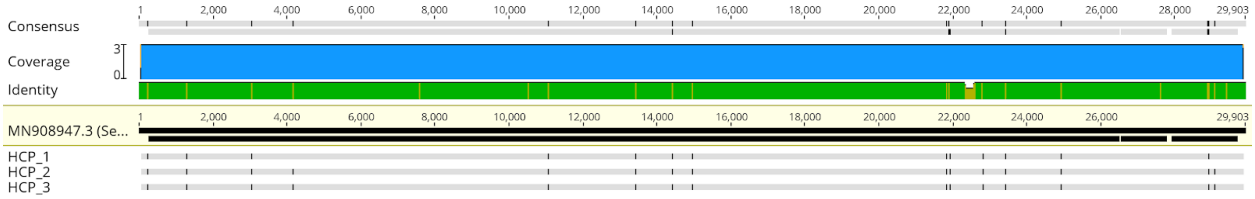
HCP 3. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	September 2020	hCoV-19/USA/WI-UW-1475/2020	20B	B.1.1.251
HCP 2	September 2020	hCoV-19/USA/WI-UW-1896/2020	20B	B.1.1.251
HCP 3	September 2020	hCoV-19/USA/WI-UW-1894/2020	20B	B.1.1.251

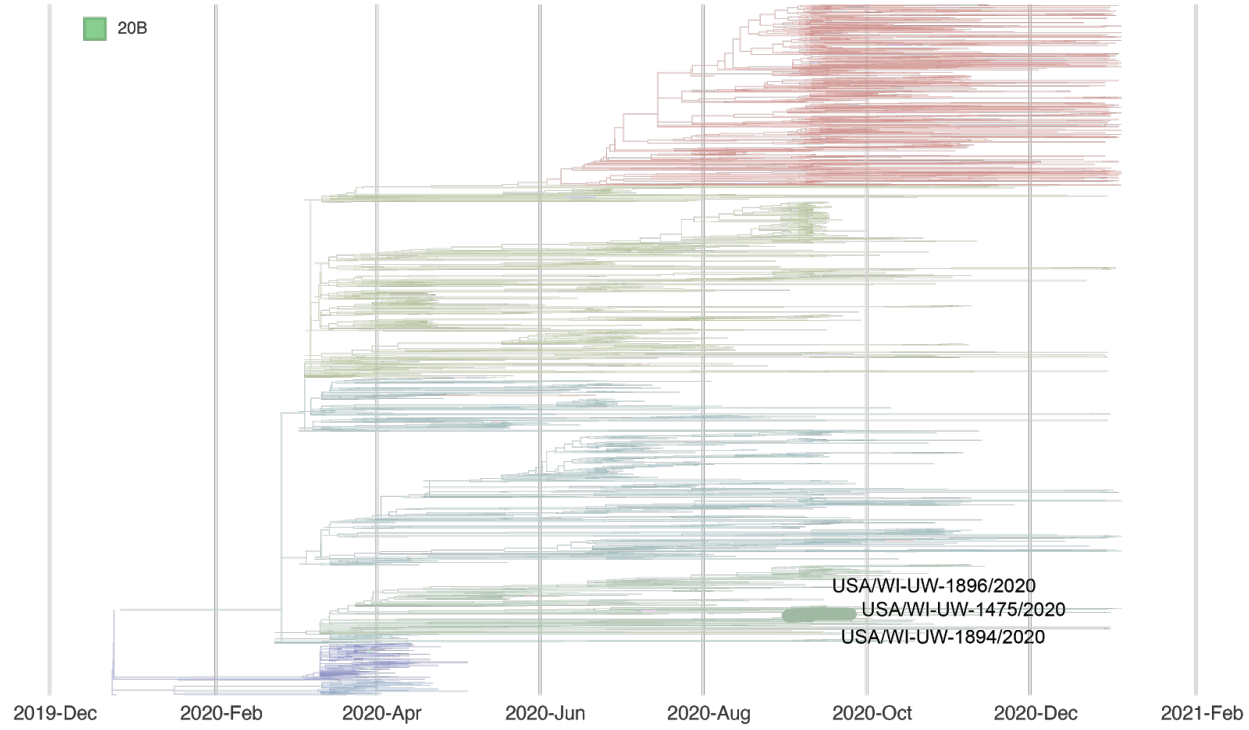
Epidemiological information

HCP 1 worked with HCP 3. HCP 3 was a household contact of HCP 2. None of these HCP had direct interactions with patients with known SARS-CoV-2 infection.

Alignment



Phylogeny


20B
USA/WI-UW-1896/2020
USA/WI-UW-1475/2020
USA/WI-UW-1894/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #19. 2020-11-05.

Likely source of HCP infection

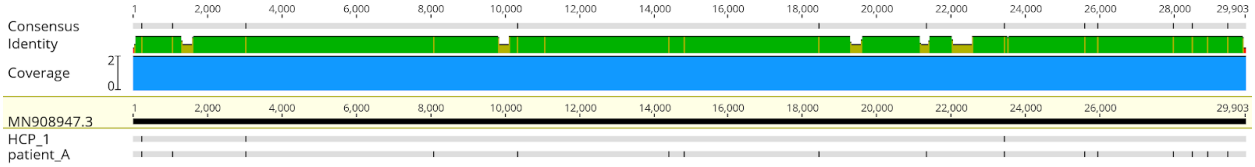
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-1928/2020	19A	B.1
patient A	September 2020	hCoV-19/USA/WI-UW-1930/2020	20G	B.1.2

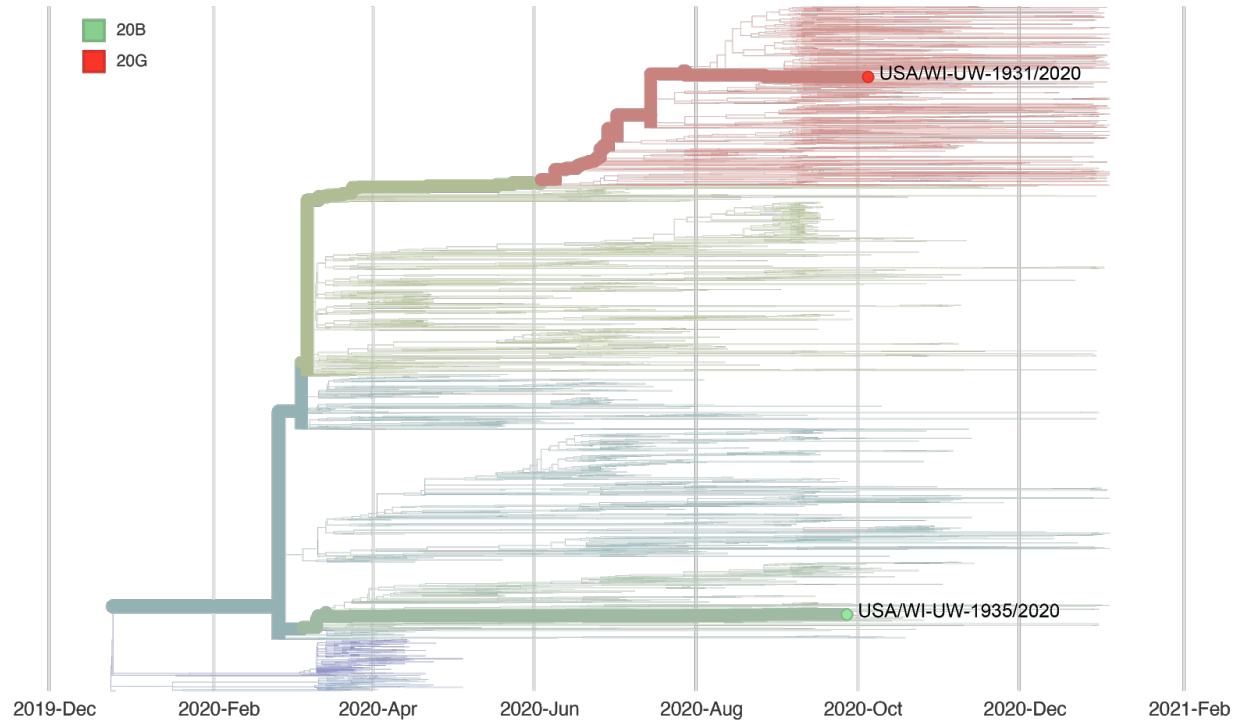
Epidemiological information

In the two weeks before symptom onset, HCP 1 provided direct care to patient A. HCP 1 wore appropriate PPE while providing care and reported no lapses in PPE use.

Alignment



Phylogeny


20B
20G
USA/WI-UW-1931/2020
USA/WI-UW-1935/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #20. 2020-11-05.

Likely source of HCP infection

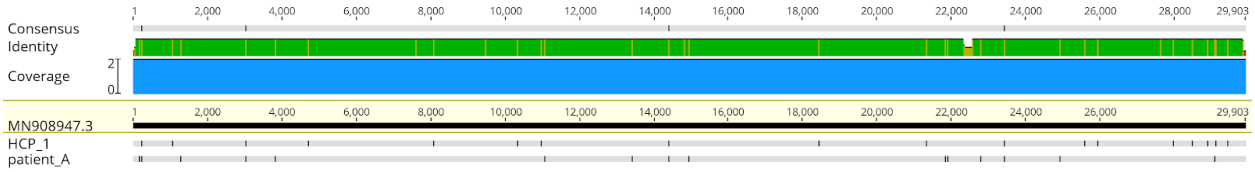
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-1931/2020	20G	B.1.2
patient A	September 2020	hCoV-19/USA/WI-UW-1935/2020	20A	B.1

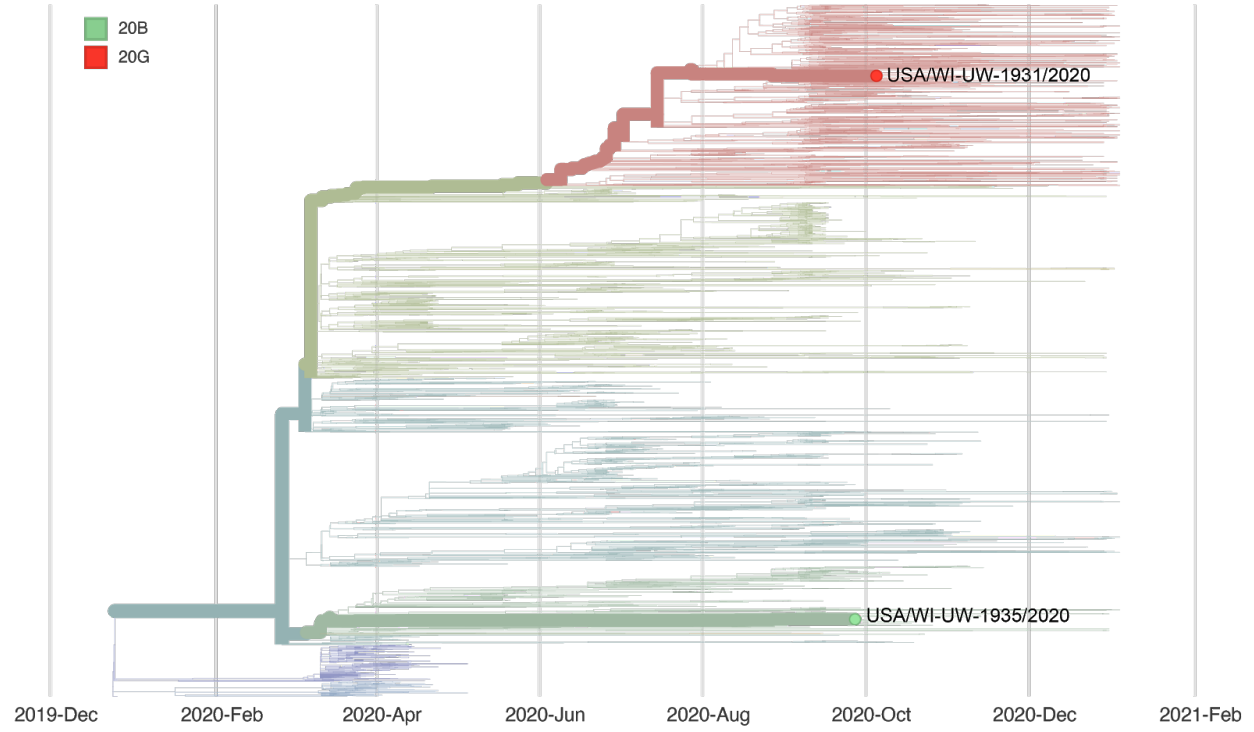
Epidemiological information

In the two weeks before symptom onset, HCP 1 provided direct care to patient A. HCP 1 wore appropriate PPE while providing care and reported no lapses in PPE use.

Alignment



Phylogeny



Report #21. 2020-11-05.

Likely source of HCP infection

HCP 1. Outside community.

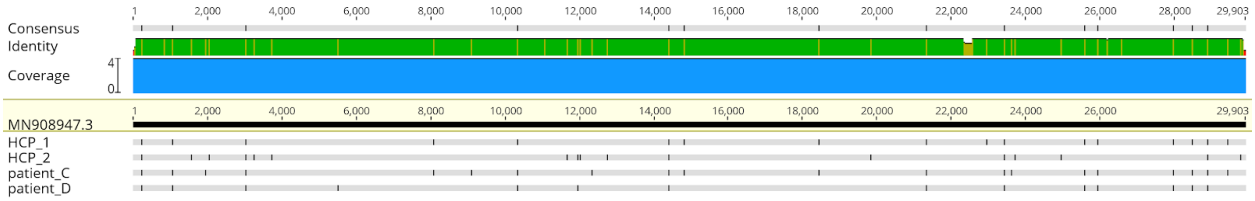
HCP 2. Outside community.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-1929/2020	20G	B.1.2
HCP 2	October 2020	hCoV-19/USA/WI-UW-1938/2020	20B	B.1.1.73
patient A	September 2020	N/A - no consensus sequence		
patient B	September 2020	N/A - no consensus sequence		
patient C	September 2020	hCoV-19/USA/WI-UW-1934/2020	20G	B.1.2
patient D	September 2020	hCoV-19/USA/IL-UW-1937/2020	20G	B.1.2
patient E	October 2020	N/A - no consensus sequence		

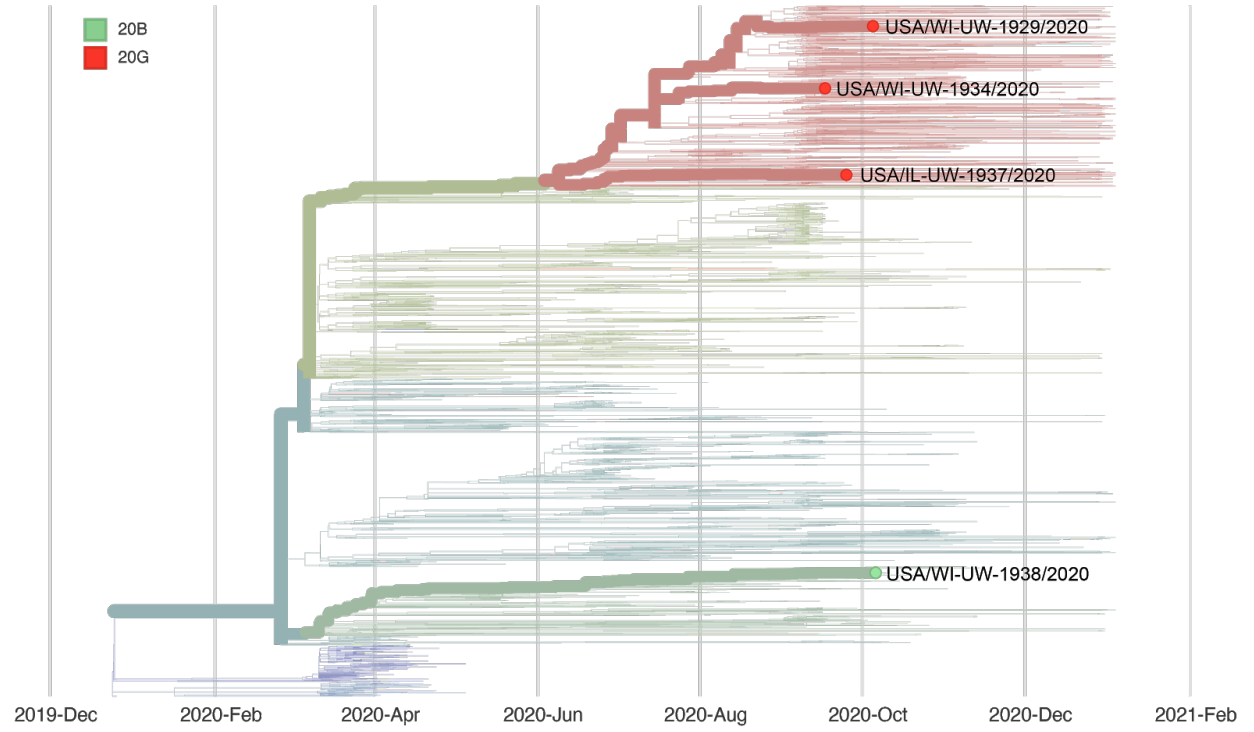
Epidemiological information

HCP 1 and 2 worked in the same department and both provided direct patient care to patients A-E in the two weeks before their onset of symptoms. HCP 1 and HCP 2 reported no lapses in PPE with each other or with their patients.

Alignment



Phylogeny



Report #22. 2020-11-05.

Likely source of HCP infection

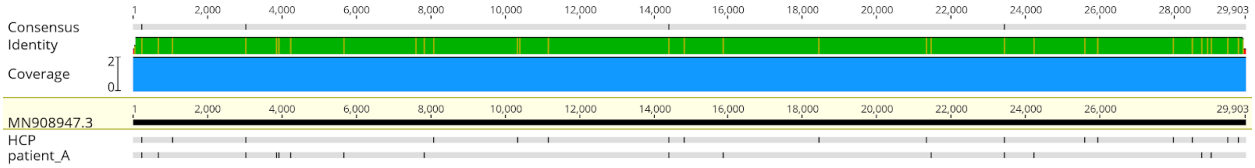
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-1936/2020	20G	B.1.2
patient A	July 2020	hCoV-19/USA/WI-UW-1100/2020	20A	B.1.139

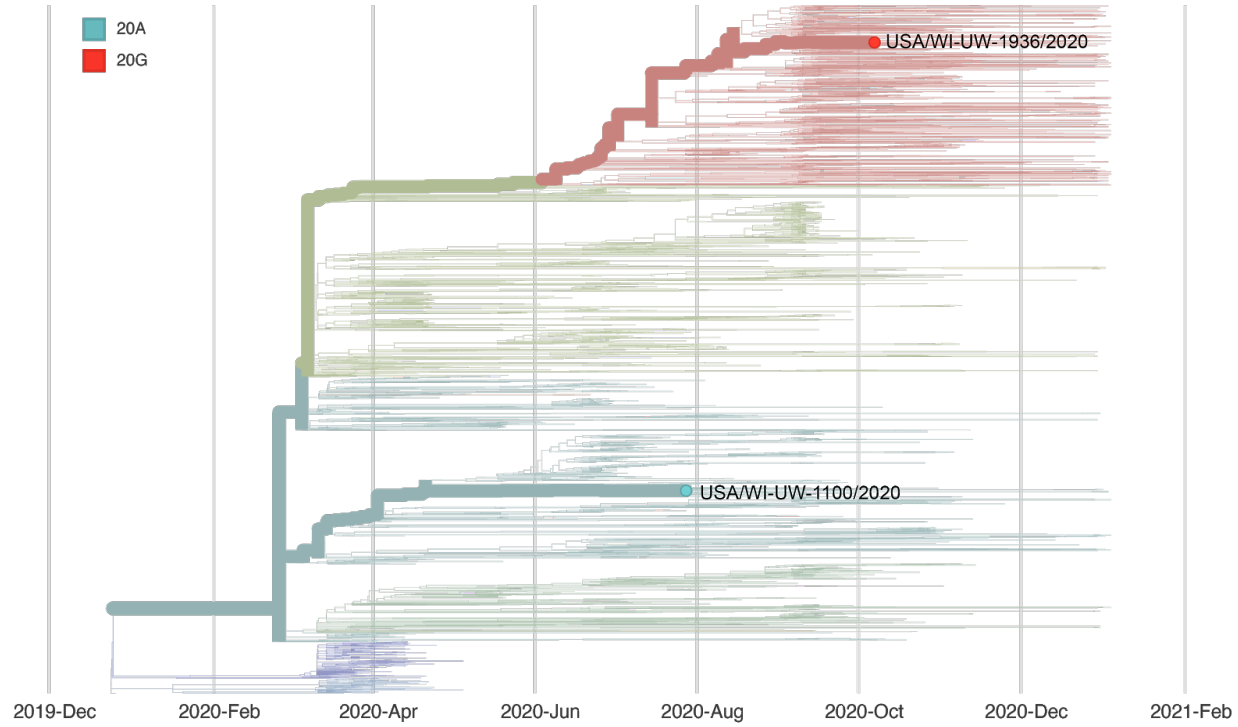
Epidemiological information

HCP 1 performed a postmortem examination on patient A. At the time the patient expired, they were known to have an active COVID-19 infection.

Alignment



Phylogeny


20A
20G
USA/WI-UW-1936/2020
USA/WI-UW-1100/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #23. 2020-11-05.

Likely source of HCP infection

HCP 1. Outside community (household contact).

HCP 2. Employee source.

HCP 3. Employee source.

HCP 4. Employee source.

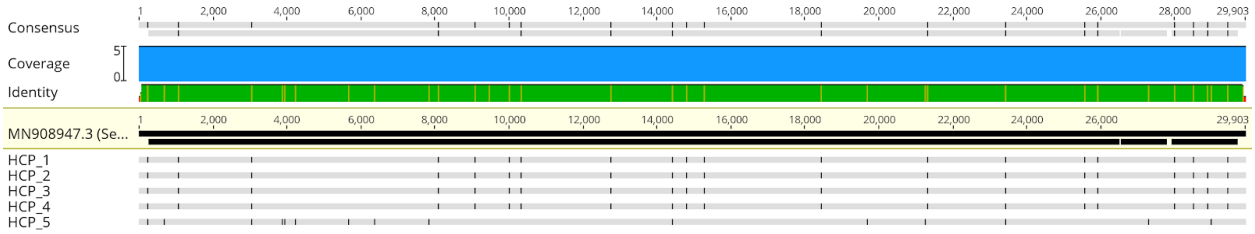
HCP 5. Outside community.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-1933/2020	20G	B.1.2
HCP 2	October 2020	hCoV-19/USA/WI-UW-1926/2020	20G	B.1.2
HCP 3	October 2020	hCoV-19/USA/WI-UW-1939/2020	20G	B.1.2
HCP 4	October 2020	hCoV-19/USA/WI-UW-1927/2020	20G	B.1.2
HCP 5	October 2020	hCoV-19/USA/WI-UW-1932/2020	20A	B.1.139

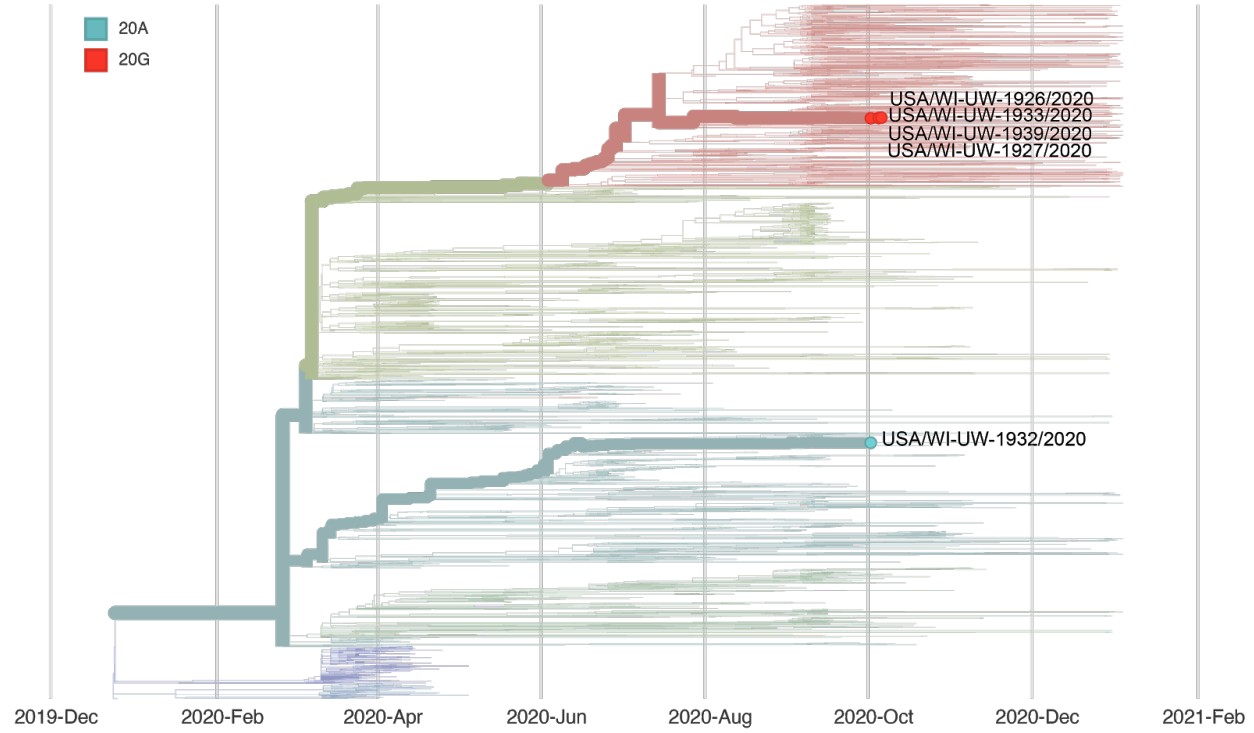
Epidemiological information

All of these HCP work in the same department, with the exception of HCP 1 who is a household contact of HCP 2. HCP 2-5 reported sharing an unmasked meal together prior to testing positive.

Alignment



Phylogeny



Report #25. 2020-12.08.

Likely source of HCP infection

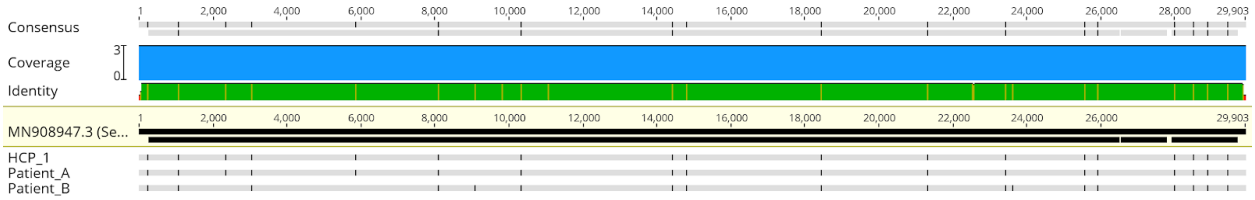
HCP 1. Patient source (patient A).

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-2326/2020	20C	B.1.2
patient A	October 2020	hCoV-19/USA/WI-UW-2362/2020	20C	B.1.2
patient B	October 2020	hCoV-19/USA/WI-UW-2389/2020	20C	B.1.2

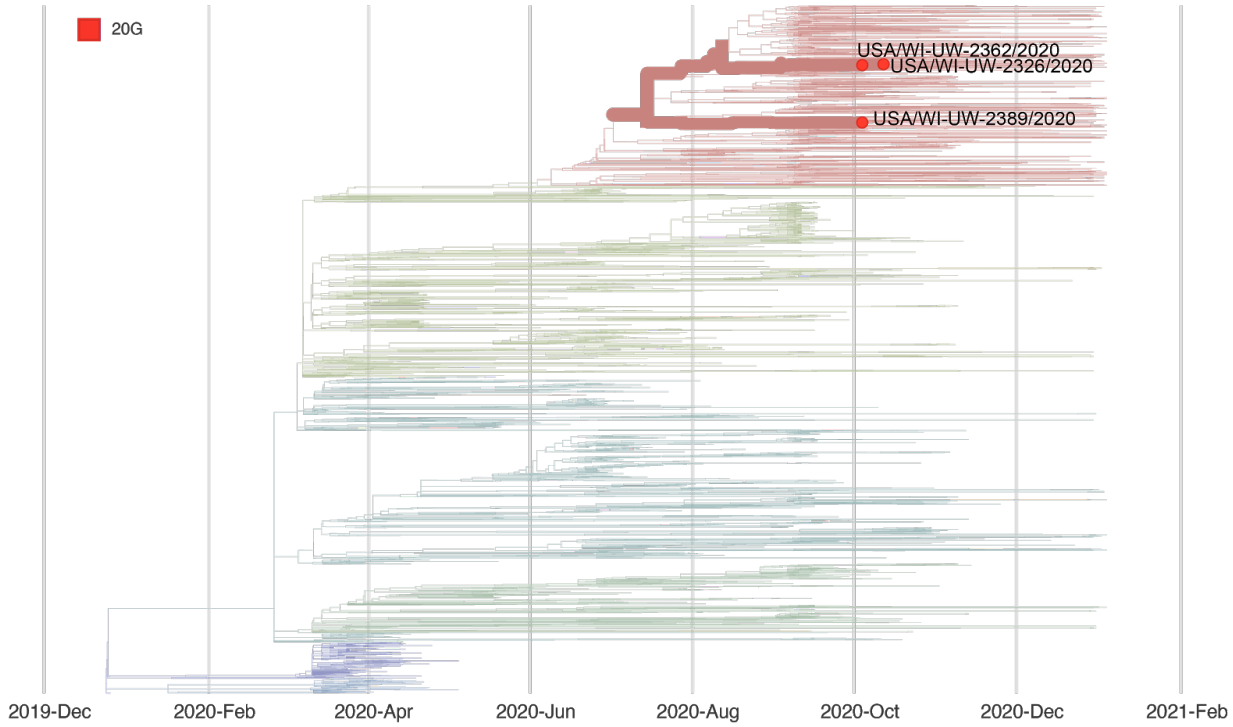
Epidemiological information

In the two weeks before symptom onset, HCP 1 provided direct care to patients A and B.

Alignment



Phylogeny


20G
USA/WI-UW-2362/2020
USA/WI-UW-2326/2020
USA/WI-UW-2389/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #27. 2020-12.09.

Likely source of HCP infection

HCP 1. Outside community.

HCP 2. Outside community.

HCP 3. Outside community.

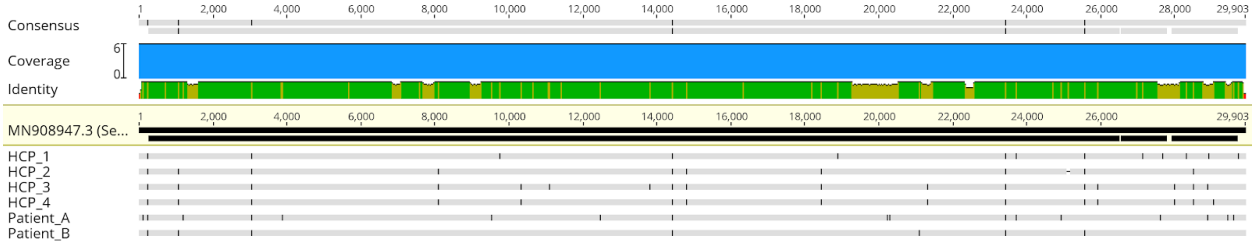
HCP 4. Outside community.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-2226/2020	20A	B.1
HCP 2	October 2020	N/A	20G	B.1.2
HCP 3	October 2020	hCoV-19/USA/WI-UW-2227/2020	20G	B.1.2
HCP 4	October 2020	hCoV-19/USA/WI-UW-2228/2020	20G	B.1.2
patient A	September 2020	hCov-19/USA/IL-UW-2327/2020	20A	B.1
patient B	September 2020	hCoV-19/USA/WI-UW-1972/2020	20G	B.1

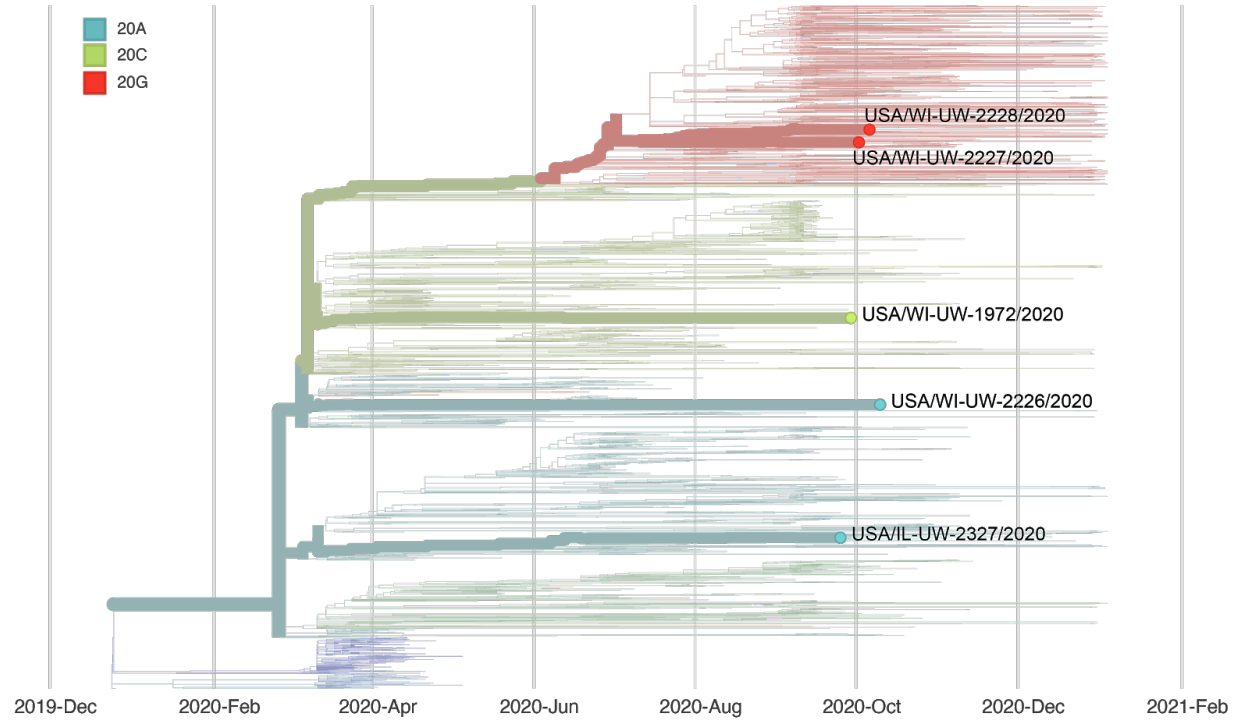
Epidemiological information

HCP 1-4 work in the same department. HCP 1-4 all had direct interactions with patients A and/or B during the 14 days prior to symptom onset.

Alignment



Phylogeny



Report #28. 2020-12.09.

Likely source of HCP infection

HCP 1. Outside community.

HCP 2. Outside community.

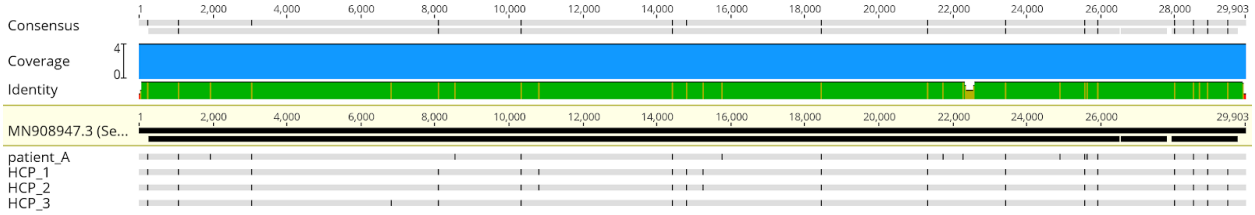
HCP 3. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
Patient A	November 2020	hCoV-19/USA/WI-UW-2265/2020	20G	B.1.2
HCP 1	October 2020	hCoV-19/USA/WI-UW-2266/2020	20G	B.1.2
HCP 2	October 2020	hCoV-19/USA/WI-UW-2267/2020	20G	B.1.2
HCP 3	October 2020	hCoV-19/USA/WI-UW-2268/2020	20G	B.1.2

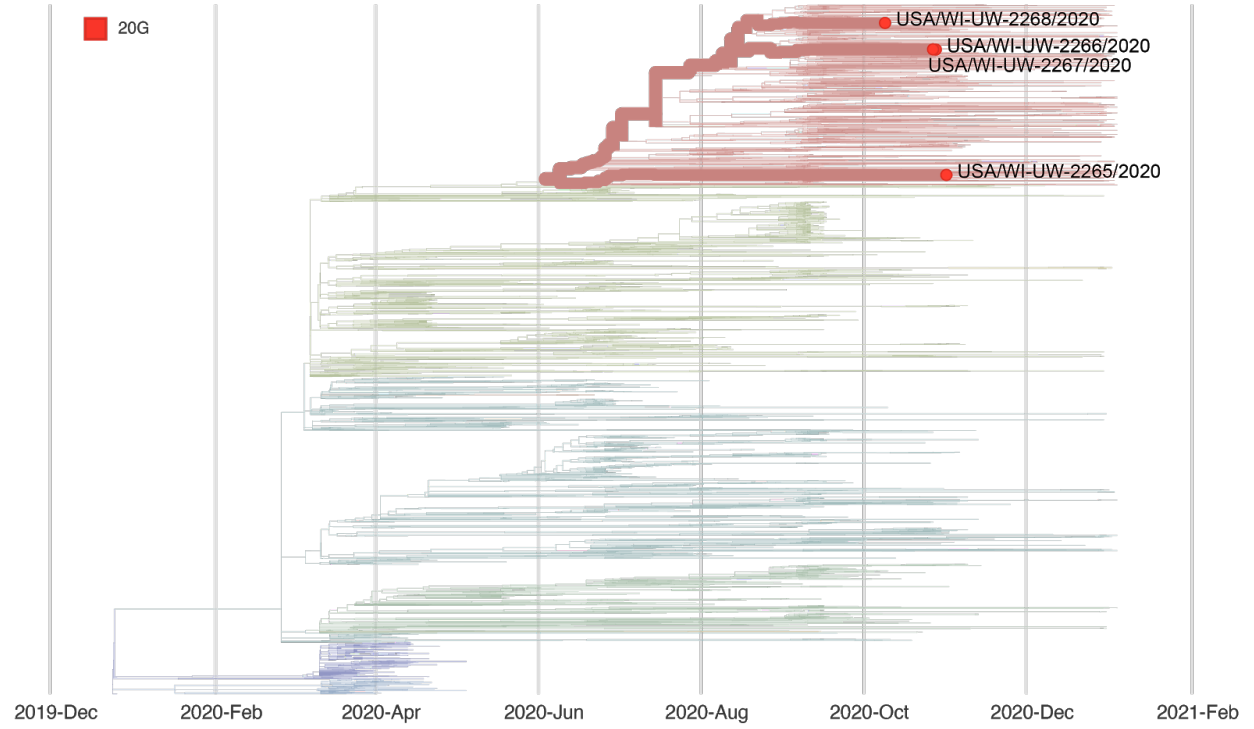
Epidemiological information

HCP 1-3 provided direct patient care to patient A in the 14 days before symptom onset. All HCP reported appropriate use of PPE with no lapses. HCP 1 and 2 had no known interactions prior to their infections. HCP 2 and HCP 3 both attended a high-risk community event for greater than 15 minutes together.

Alignment



Phylogeny



Report #29. 2020-12.09.

Likely source of HCP infection

HCP 1. Employee source (HCP 3).

HCP 2. Outside community.

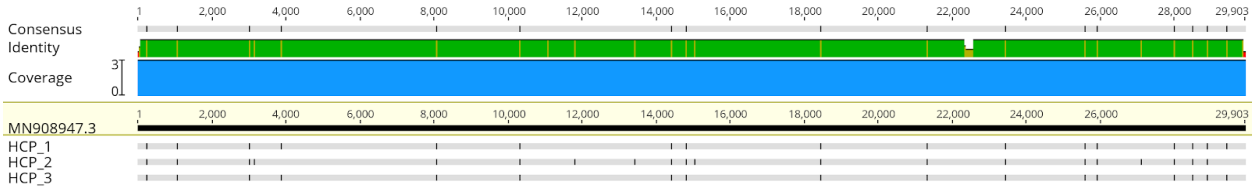
HCP 3. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	November 2020	hCoV-19/USA/WI-UW-2269/2020	20G	B.1.2
HCP 2	November 2020	hCoV-19/USA/WI-UW-2270/2020	20G	B.1.2
HCP 3	November 2020	hCoV-19/USA/WI-UW-2271/2020	20G	B.1.2

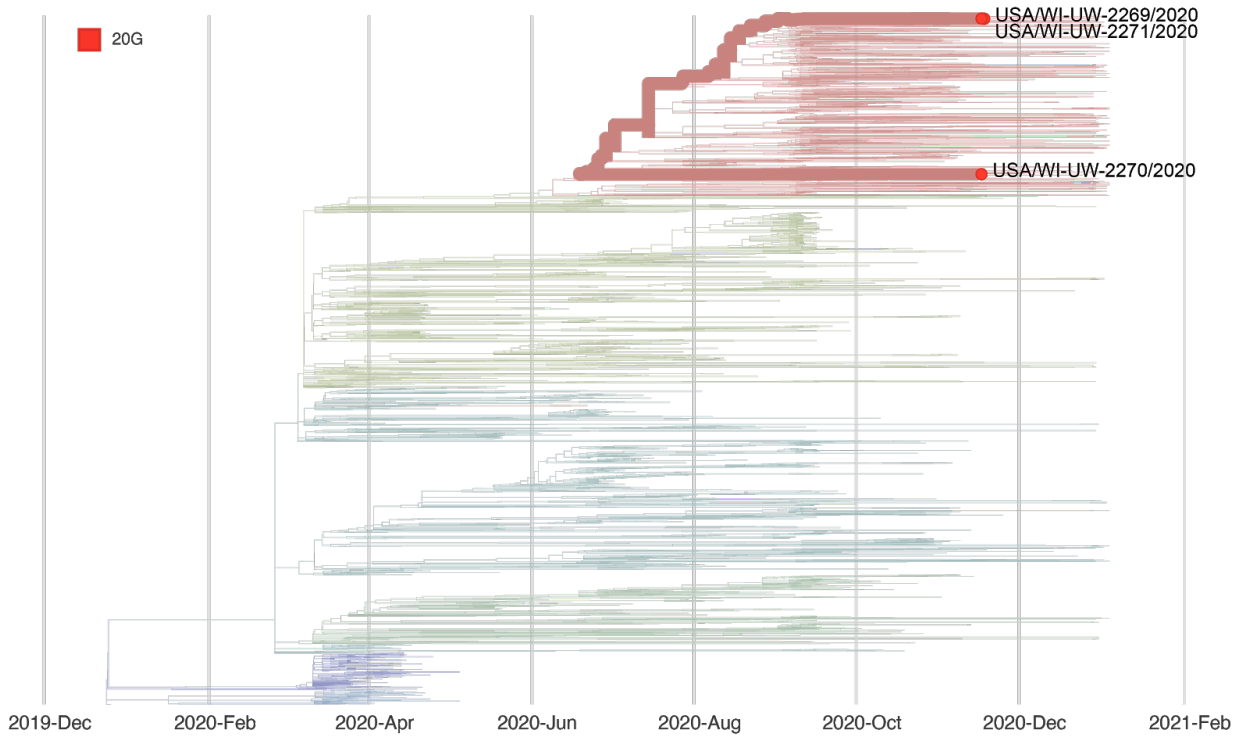
Epidemiological information

HCP 1-3 work in the same department. HCP 1 and 3 attended an in person meeting together and reported sitting 6-feet apart while wearing masks. HCP 2 did not attend this meeting. HCP 3 reported symptoms before HCP 1.

Alignment



Phylogeny


20G
USA/WI-UW-2269/2020
USA/WI-UW-2271/2020
USA/WI-UW-2270/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #30. 2020-12.09.

Likely source of HCP infection

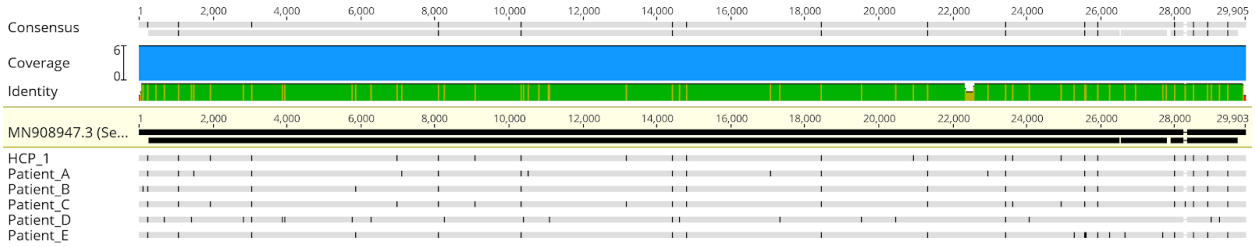
HCP 1. Patient source (patient C).

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	November 2020	hCoV-19/USA/WI-UW-2272/2020	20G	B.1.2
patient A	November 2020	hCov-19/USA/WI-UW-2324/2020	20G	B.1.2
patient B	October 2020	hCoV-19/USA/WI-UW-2273/2020	20G	B.1.2
patient C	November 2020	hCoV-19/USA/WI-UW-2274/2020	20G	B.1.2
patient D	November 2020	hCoV-19/USA/WI-UW-2275/2020	20A	B.1.139
patient E	November 2020	hCoV-19/USA/WI-UW-2276/2020	20G	B.1.2

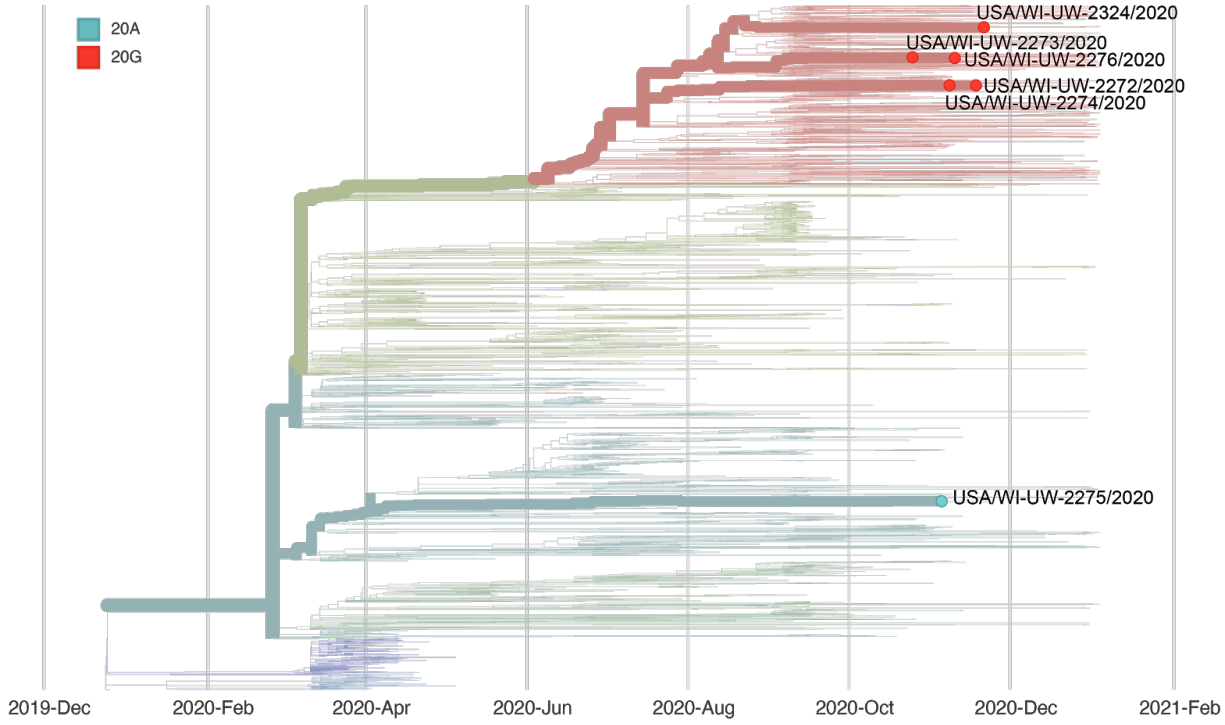
Epidemiological information

HCP 1 provided direct patient care to patients A-E and reported no lapses in PPE use.

Alignment



Phylogeny


20A
20G
USA/WI-UW-2324/2020
USA/WI-UW-2273/2020
USA/WI-UW-2276/2020
USA/WI-UW-2272/2020
USA/WI-UW-2274/2020
USA/WI-UW-2275/2020
2019-Dec
2020-Feb
2020-Apr
2020-Jun
2020-Aug
2020-Oct
2020-Dec
2021-Feb

Report #31. 2020-12.09.

Likely source of HCP infection

HCP 1. Outside community.

HCP 2. Outside community.

HCP 3. Outside community.

HCP 4. Employee source (HCP 2).

HCP 5. Employee source (HCP 2).

HCP 6. Outside community.

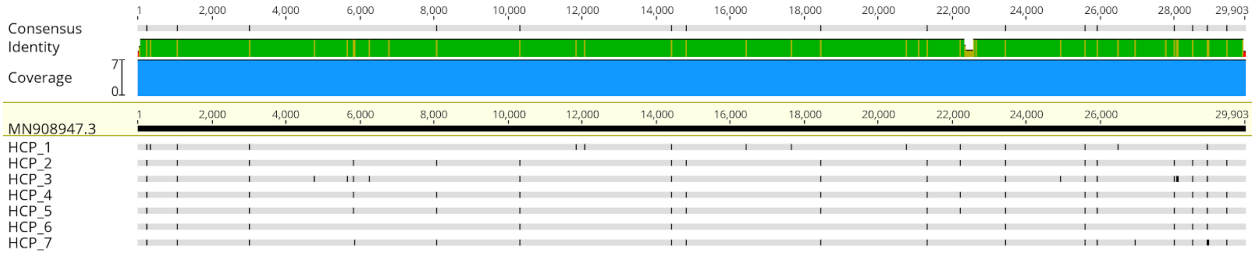
HCP 7. Outside community.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	November 2020	hCoV-19/USA/WI-UW-2312/2020	20C	B.1.363
HCP 2	November 2020	hCoV-19/USA/WI-UW-2277/2020	20G	B.1.2
HCP 3	November 2020	hCoV-19/USA/WI-UW-2278/2020	20G	B.1.2
HCP 4	November 2020	hCoV-19/USA/WI-UW-2279/2020	20G	B.1.2
HCP 5	November 2020	hCoV-19/USA/WI-UW-2280/2020	20G	B.1.2
HCP 6	November 2020	hCoV-19/USA/WI-UW-2281/2020	20G	B.1.2
HCP 7	November 2020	hCoV-19/USA/WI-UW-2282/2020	20G	B.1.2

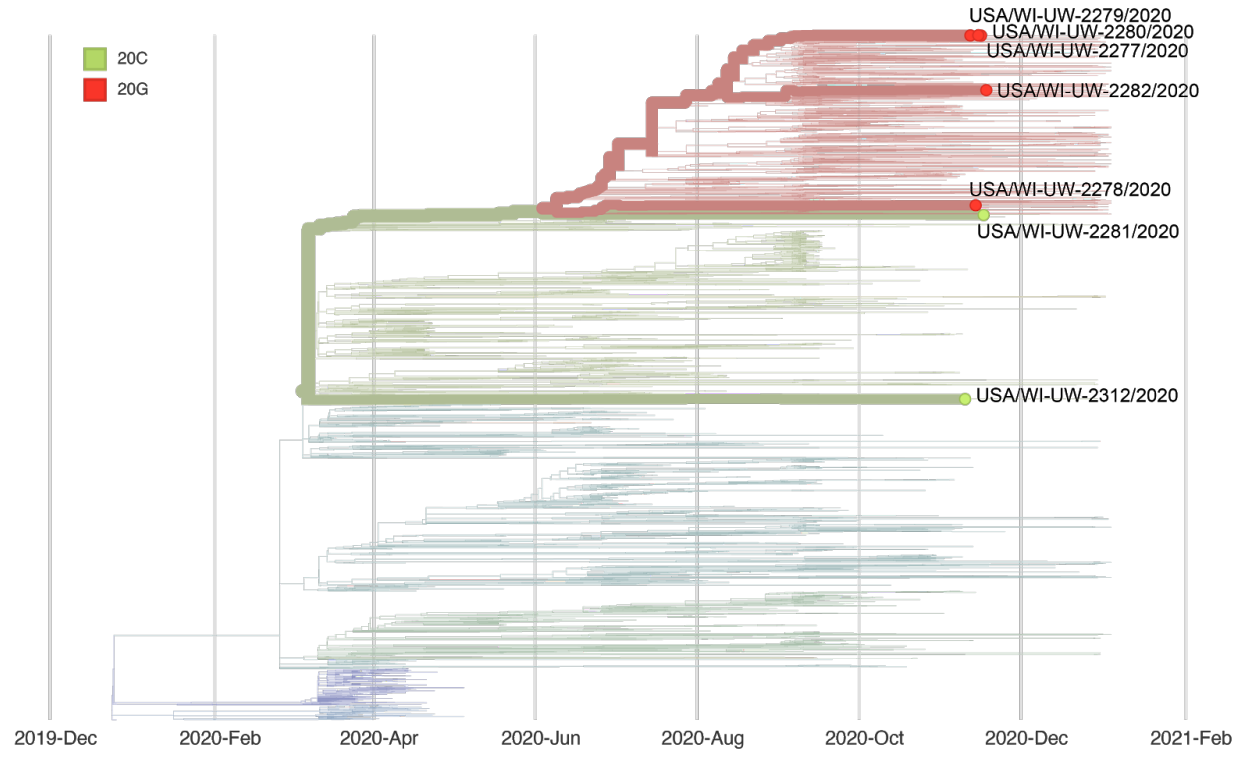
Epidemiological information

These HCP are all in the same department. Their level of interaction with each other is unclear.

Alignment



Phylogeny



Report #32. 2020-12.09.

Likely source of HCP infection

HCP 1. Outside community.

HCP 2. Outside community.

HCP 3. Outside community.

HCP 4. Outside community.

HCP 5. Outside community.

HCP 6. Outside community.

HCP 7. Outside community.

HCP 8. Outside community.

HCP 9. Outside community.

HCP 10. Outside community.

HCP 11. Outside community.

HCP 12. Outside community.

HCP 13. Inconclusive (could be HCP 8, but these samples were collected >14 apart).

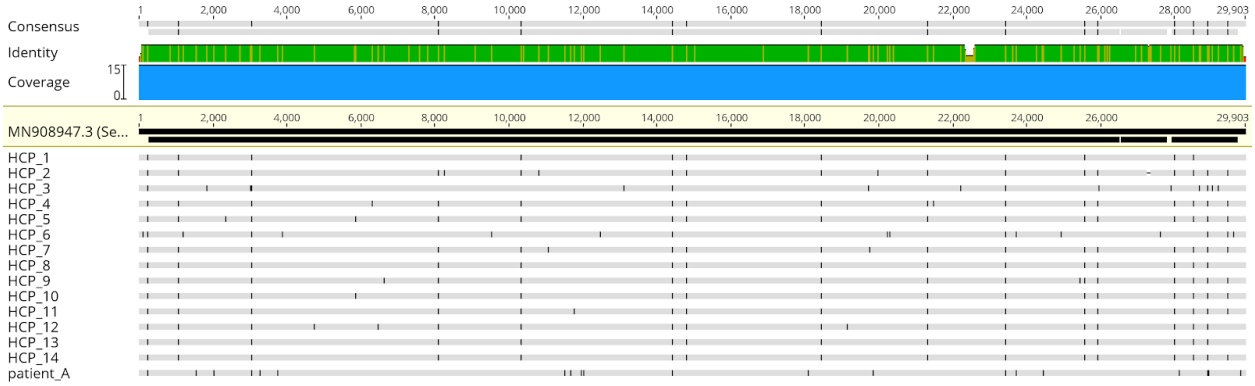
HCP 14. Employee source (HCP 8 or HCP 13).

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	October 2020	hCoV-19/USA/WI-UW-2283/2020	20G	B.1.370
HCP 2	October 2020	hCoV-19/USA/WI-UW-2284/2020	20G	B.1.2
HCP 3	October 2020	hCoV-19/USA/WI-UW-2285/2020	20A	B.1.216
HCP 4	October 2020	hCoV-19/USA/WI-UW-2286/2020	20G	B.1.2
HCP 5	October 2020	hCoV-19/USA/WI-UW-2390/2020	20G	B.1.2
HCP 6	October 2020	hCoV-19/USA/WI-UW-2391/2020	20G	B.1
HCP 7	November 2020	hCoV-19/USA/WI-UW-2287/2020	20G	B.1.2
HCP 8	November 2020	hCoV-19/USA/WI-UW-2288/2020	20G	B.1.2
HCP 9	November 2020	hCoV-19/USA/WI-UW-2289/2020	20G	B.1.2
HCP 10	November 2020	hCoV-19/USA/WI-UW-2313/2020	20G	B.1.2
HCP 11	November 2020	hCoV-19/USA/WI-UW-2290/2020	20G	B.1.2
HCP 12	November 2020	hCoV-19/USA/WI-UW-2291/2020	20G	B.1.2
HCP 13	November 2020	hCoV-19/USA/WI-UW-2292/2020	20G	B.1.2
HCP 14	November 2020	hCoV-19/USA/WI-UW-2293/2020	20G	B.1.2
patient A	November 2020	hCoV-19/USA/WI-UW-2325/2020	20B	B.1.1.73

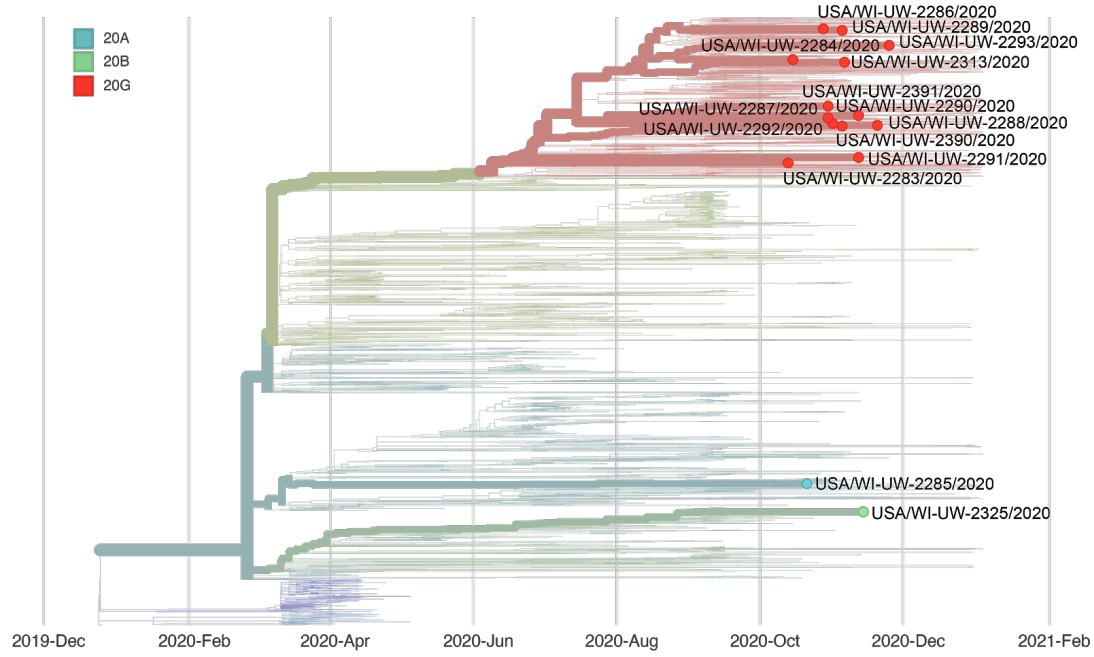
Epidemiological information

HCP 1-14 work in the same department together. One or more of these HCP provided direct patient care to patient A.

Alignment



Phylogeny



Report #33. 2021-01-03.

Likely source of HCP infection

HCP 1. Combined patient and employee cluster.

HCP 2. Combined patient and employee cluster.

HCP 3. Combined patient and employee cluster.

HCP 4. Outside community.

HCP 5. Combined patient and employee cluster.

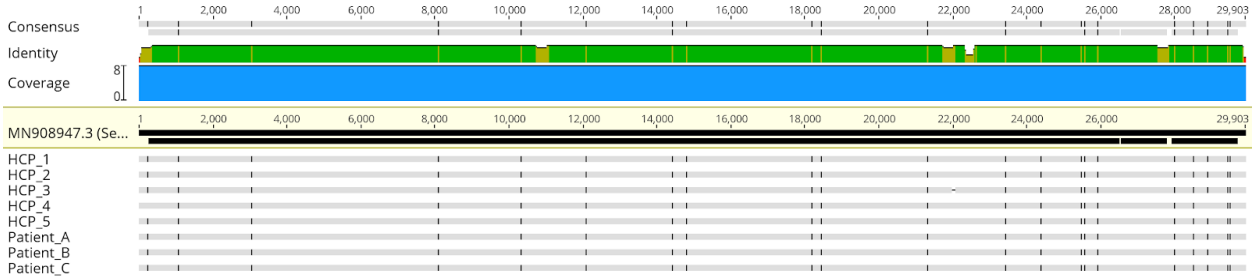
Patients A, B and C are the patients involved in this combined patient and employee cluster.

Sample type	**Sample collection date**	**GISAID identifier**	**Clade** (Nextstrain)	**Lineage** (Pangolin)
HCP 1	December 2020	hCoV-19/USA/WI-UW-2394/2020	20G	B.1.2
HCP 2	December 2020	hCoV-19/USA/WI-UW-2396/2020	20G	B.1.2
HCP 3	December 2020	hCoV-19/USA/WI-UW-2451/2020	20G	B.1.2
HCP 4	December 2020	hCoV-19/USA/WI-UW-2452/2020	20G	B.1.2
HCP 5	December 2020	hCoV-19/USA/WI-UW-2450/2020	20G	B.1.2
patient A	December 2020	hCoV-19/USA/WI-UW-2392/2020	20G	B.1.2
patient B	December 2020	hCoV-19/USA/WI-UW-2393/2020	20G	B.1.2
patient C	December 2020	hCoV-19/USA/IL-UW-2449/2020	20G	B.1.2

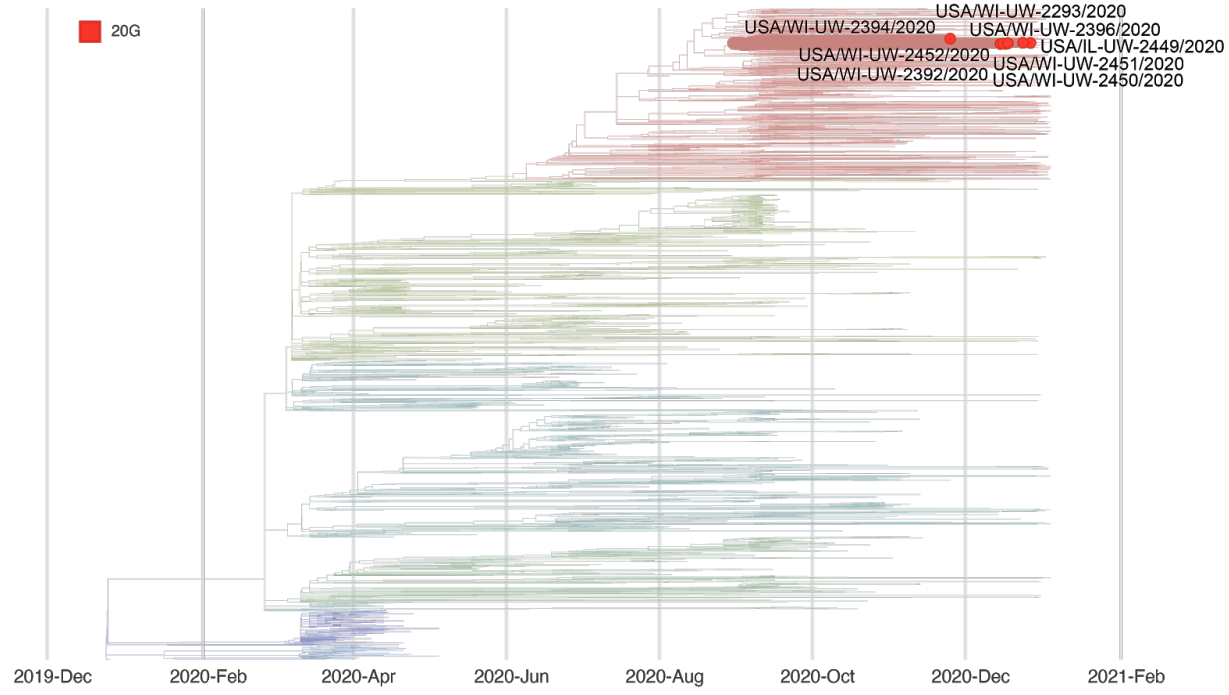
Epidemiological information

HCP 1-5 provided direct care to one or more of these patients, A-C. HCP 1-5 may have also interacted with each other. No lapses in PPE were reported. Patient A had the earliest reported symptom onset.

Alignment



Phylogeny



Report #34. 2021-01-03.

Likely source of HCP infection

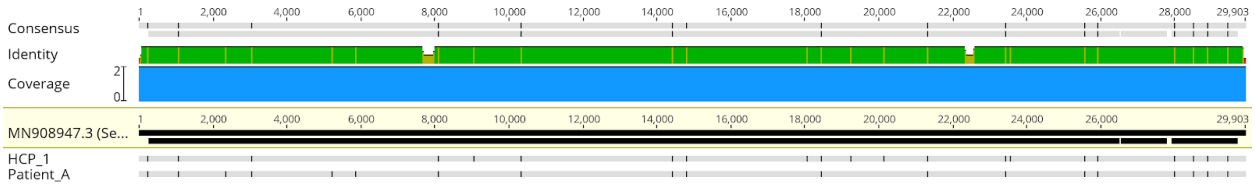
HCP 1. Outside community.

Sample type	Sample collection date	GISAID identifier	Clade (Nextstrain)	Lineage (Pangolin)
HCP 1	December 2020	hCoV-19/USA/WI-UW-2453/2020	20G	B.1.2
patient A	December 2020	hCoV-19/USA/WI-UW-2395/2020	20G	B.1.2

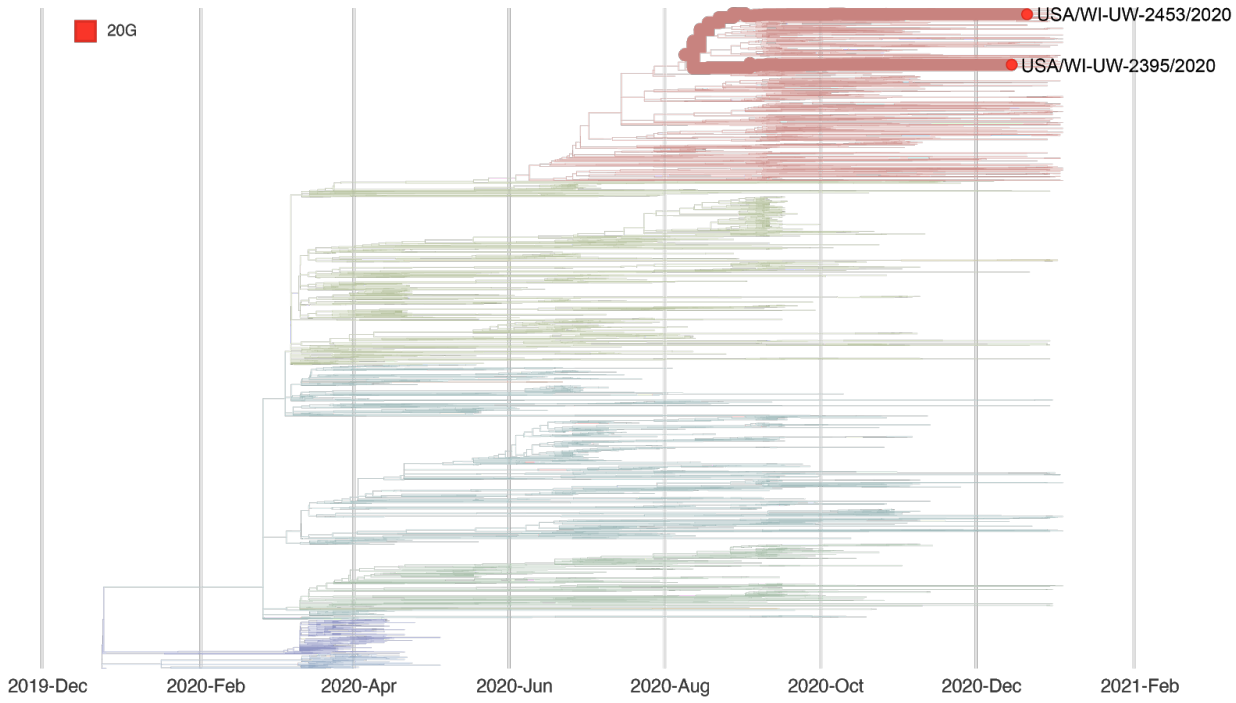
Epidemiological information

HCP 1 provided direct care to patient A during the 14 days prior to HCP 1's symptom onset. HCP 1 reported no lapse in PPE when providing care to patient A.

Alignment



Phylogeny



Supplemental file 2 can be accessed here.

Chapter 7:

Conclusions and future directions

The emergence of zoonotic viruses is one of the greatest threats to global health security. More than half of all known human pathogens can be traced to a zoonotic source [2,4,5]. Yet, the capacity for RNA viruses to rapidly adapt to new host environments and to respond to shifting selective pressures is not completely understood. In this dissertation, I investigated the evolutionary processes by which zoonotic RNA viruses adapt to mammalian hosts and I combine principles of viral evolution with epidemiology and population health to investigate the early patterns of SARS-CoV-2 spread in the state of Wisconsin.

Altogether, this work suggests the effects of randomness on viral populations within and between individual hosts are a previously underappreciated brake to the pace of viral adaptation and host-switching for influenza A virus (IAV) and SARS-CoV-2. Additionally, this work underscores the value of genomic epidemiology early in a pandemic to understand patterns of viral transmission in different populations and to assess the impact of public health guidelines and interventions on a rolling basis.

The effects of randomness contribute significantly to viral evolutionary dynamics within individual hosts

In **chapter 2**, we examined the within- and between-host viral dynamics of wildtype H7N9 viruses in a ferret model system. On its face, the fact that fully avian H7N9 viruses

efficiently replicate and transmit in a ferret model is somewhat alarming and raises the question: if wildtype H7N9 viruses can do this, what is stopping these viruses from causing a full-blown pandemic? In this chapter, we hypothesized that HPAI and LPAI avian H7N9 viruses would be under strong selective pressure to become more mammalian in the context of a ferret model system. To our surprise, we found no evidence to support this hypothesis. Instead, we found HPAI and LPAI H7N9 viruses in ferrets are under mild purifying selection and new mutations were generally being purged from the viral population, which is a signature classically associated with a virus that is at least moderately well adapted to its host. This is not to say the avian H7N9 viruses are optimally adapted to ferrets or mammals more generally, but they do replicate to high titer, transmit often, and are not subject to diversifying or positive selection to any measurable degree. I will expand on these nuances of these conclusions in the theoretical framework outlined below.

In this study we found HPAI and LPAI H7N9 viral diversity was dominated by low-frequency iSNVs in ferrets, which is consistent with purifying selection and population expansion. By leveraging the longitudinal sampling schema, we evaluated iSNV dynamics over time and observed some surprising patterns. Frequencies of specific mutations in donor ferrets, even in the case of known mammalian adaptive mutations, did not predict their frequencies post-transmission. We estimated a very narrow transmission bottleneck size for H7N9 virus in ferrets where a single virus (95% CI: 1-3) is likely responsible for initiating infection following transmission. Interestingly, while the H1N1 transmission bottlenecks were narrow (6, 95% CI: 3-11), they were not as stringent as

the H7N9 pairs. The vast majority of H7N9 iSNVs arising in ferret hosts are lost during transmission and because so few viruses found infection following transmission, any iSNV that happens to be present in a transmitting virus' genome will likely become fixed in the post-transmission viral population. Together, we speculate that purifying selection, randomness, and tight bottlenecks combine to severely constrain the ability of H7N9 viruses to efficiently adapt to mammalian hosts in typical spillover infections, even with onward airborne transmission.

In **chapter 3**, we did a similar study, but instead of avian H7N9 in ferrets, we looked at SARS-CoV-2 in domestic cats. This study was among the first to evaluate the within- and between-host dynamics of SARS-CoV-2 in a mammalian model system so we did not know what to expect. In this study, we showed that SARS-CoV-2 genetic variation, like H7N9 viruses in ferrets, is predominantly influenced by genetic drift and purifying selection. Additionally, we found transmission bottlenecks were very narrow (combined estimate = 5, 99% CI 1-7), which was at odds with one [54] of the two other available studies that had also measured bottleneck sizes. Subsequent research that has been done since, including the work presented in **chapter 4**, supports narrow transmission bottlenecks in SARS-CoV-2. While we observed many similarities between SARS-CoV-2 evolutionary dynamics in domestic cats and H7N9 viruses in ferrets, there was one notable difference. A variant at amino acid position 655 in Spike (H655Y) arose rapidly in index cats, persisted at intermediate frequencies in the donor cats, and then became fixed following transmission in two of three pairs. This same variant has been shown to confer escape from human monoclonal antibodies and circulates in humans on multiple SARS-CoV-2

genetic lineages and is one of the defining mutations in the P.1. VOC. Although we did no functional studies to investigate the phenotypic impact of S H655Y in cats specifically, we speculated that S H655Y could have improved Spike fusion efficiency and host cell entry in cats. This was a subtle, but notable departure from the otherwise stochastic evolutionary dynamics of SARS-CoV-2 in a mammalian host.

In our third and final within- and between-host study, presented in **chapter 4**, we carefully characterized SARS-CoV-2 viral diversity in 133 natural infections in human hosts including 28 putative household transmission pairs and we took advantage of our large consensus-level surveillance dataset from the community where those individuals reside to compare within-host diversity to population-level diversity. We found most SARS-CoV-2 infections were characterized by very few iSNVs and the majority of these were low frequency. Most iSNVs detected in individuals were not detected in the local or global consensus genomes and were very rarely detected in downstream branches on the local and global phylogenetic trees. Even among putative household transmission pairs, iSNVs were shared very infrequently and the transmission bottleneck was narrow. We reasoned that the combination of low within-host diversity, narrow transmission bottlenecks, and infrequently propagation along transmission chains combines to slow the rate of novel variant emergence and the pace of viral evolution in typical, acute human infections.

Looking across these three within-host studies, viral diversity of H7N9 in ferrets and SARS-CoV-2 in cats and humans is subject to the combined effects of purifying selection and genetic drift with no evidence of diversifying selection. Airborne transmission in all of

these cases is characterized by a very narrow bottleneck where very few viruses make it out of a donor and into a recipient, which results in a dramatic reduction in viral diversity across the transmission event. While the similarities are obvious, there are some key differences as well. It is clear that within-host diversity is even more limited in SARS-CoV-2 infections than IAV infections. Additionally, the rapid outgrowth and preferential transmission of Spike H655Y in the cat transmission study was the only exception to the otherwise stochastic evolutionary dynamics driving SARS-CoV-2 and H7N9 in mammalian hosts. In an effort to unify these observations and the results of two previous studies led by past graduates of the Friedrich lab [204,205], I describe a theoretical model of respiratory virus evolution within and between hosts in the following section.

A proposed framework zoonotic respiratory RNA virus evolution within and between hosts

As I described in the introduction of this dissertation, positive selection appears to be a major driver of IAV and SARS-CoV-2 globally, but it does not drive evolution within hosts. Although the work in SARS-CoV-2 is more recent and limited, study after study has shown this to be true for seasonal IAV [122,133,376] and for avian IAV [135,239]. This has perplexed many virologists because RNA viruses appear to have many of the traits required for efficient and deterministic selection. RNA viruses generate plenty of genetic diversity, the substrate for evolution, through error-prone RNA polymerases, reassortment, and homologous recombination (in the case of SARS-CoV-2). Adaptive evolution is most efficient in large populations and RNA viruses often appear to achieve large population sizes via explosive replication within hosts. Even still, respiratory RNA viruses must

contend with genetic drift and narrow transmission bottlenecks and it appears that very often the effects of these stochastic processes dominate viral evolutionary trajectories within and between hosts.

If we accept that stochasticity drives evolution of RNA respiratory viruses at the level of the individual host, how can we explain IAV antigenic drift and positive selection of SARS-CoV-2 VOCs on a population scale? Well, I think RNA respiratory viruses might have taken a page out of Charles F. Kettering's playbook when he said "an inventor fails 999 times, and if he succeeds once, he's in. He treats his failures simply as practice shots." By this I mean that I imagine RNA viruses overcome the effects of stochasticity and randomness within individuals by pursuing large numbers of infections in a host population. RNA viruses are often able to escape humoral immunity, significantly improve receptor specificity, and bolster transmissibility with relatively few mutations. So, suppose an RNA virus infects millions (IAV) of hosts in a single year, even a single "jackpot event" involving the onward transmission of a novel escape variant or equivalent might be all that a virus needs to initiate a selective sweep across a host population. It follows that the pace of adaptive evolution of any RNA virus would generally correspond to the volume of recent viral infections and this is consistent with rapid global evolution of SARS-CoV-2 and limited global evolution of seasonal influenza. This comparison is not apples-to-apples because there was no preexisting human immunity to SARS-CoV-2, which is clearly not the case for IAV.

While the total number of infections a respiratory RNA virus achieves may be a key predictor of evolutionary pace, another key factor is the position of a virus on its "fitness landscape" which captures all of the complex fitness "peaks" and "valleys" resulting from the many combinations of virus and host genotypes/environments. **Figure 1** depicts a theoretical fitness landscape for IAV, which I will use to further illustrate these ideas. The y-axis of this figure represents virus genotypes as a virus transitions from avian to mammalian, the x-axis represents host genetics and conditions ranging from the avian reservoir to humans, and the z-axis (height) represents relative fitness at all possible x-y coordinates. The peaks and valleys shown here are conceptual and not derived from actual data. I have included five different IAV viruses in different host environments as colored dots and I will briefly discuss each of these below.

The dark red virus represents a seasonal H1N1 virus replicating in a human host, which I positioned on a fitness peak. The contours of this landscape are likely very dynamic and a new fitness peak might emerge after this seasonal H1N1 infects a large number of hosts and subsequently must contend with host immune responses. The orange virus is on a similar fitness peak, but is an avian-adapted virus in an avian host, like a H7N9 virus in a water fowl.

The yellow virus represents the avian H7N9 virus we investigated in ferret hosts in **chapter 2**. As outlined above, this virus appeared to be relatively well-adapted to ferret hosts so I placed them on a small fitness peak on this landscape. If a virus is on a fitness peak in this landscape, there are fewer possible genetic changes available to it that will

result in a fitness advantage (i.e. it is already towards the top of a fitness hill so it is hard for it move even further uphill). Consistent with this, we found no evidence of positive or adaptive evolution of these viruses in ferret hosts. Similarly, another recent study by Moncla et al. showed that while H5N1 spillover infections can generate low-frequency, mammalian-adaptive mutations, their spread is limited by purifying selection, genetic drift, and acute infection timeframes [239]. A fitness peak in the upper right-hand corner of this landscape is likely taller than the peak H7N9 and H5N1 viruses in mammalian spillover infections are located on, but making the move toward that peak is tough. I can imagine two routes for the yellow virus (an avian virus in a mammalian model or a spillover infection) to move to the red virus (a seasonal human IAV): (1) incremental changes via antigenic drift across this landscape, but this would require movement through fitness valleys so this is pretty unlikely unless the total number of infections is massive or (2) a giant jump, which a virus might achieve through a major reassortment event, as has been seen in all IAV pandemics.

The blue and green viruses represent genetically modified, partially avian and partially mammalian IAVs in two previous studies. The blue virus represents an avian H5 HA protein in the background of a human H1N1 virus [205]. The green virus represents a 1918-like avian virus with engineered mammalian-adaptive mutations [204,208]. Importantly, neither of these viruses exist in nature suggesting they are not fit and likely occupy valleys on this fitness landscape. While these viruses are subject to the same evolutionary constraints as the red, yellow and orange viruses (short infection times, genetic drift, and narrow transmission bottlenecks), the position of these viruses in a fitness valley means

that there are many possible genetic changes which will confer improved fitness. Diversifying selection and selective sweeps are much more likely to be detected in the context of viruses in this position. This is consistent with the results of this study, which were able to identify adaptive evolution and selective sweeps across the transmission bottleneck.

A similar fitness landscape likely exists for SARS-CoV-2, although its dimensionality may be even more complex given the diverse host range of this virus. I suspect SARS-CoV-2 in cats was positioned on a modest fitness valley, consistent with the preferential amplification and transmission of H655Y in Spike. I suspect SARS-CoV-2 in human hosts are already on a modest fitness peak, consistent with purifying selection within hosts and infrequent propagation of iSNVs along transmission chains. As SARS-CoV-2 continues to infect massive numbers of people, it is likely to discover even higher fitness peaks. Indeed, variants of interest and variants and concern with enhanced phenotypic changes began emerging beginning in fall 2020, so we are forced to wonder “has SARS-CoV-2 reached peak fitness?”. While SARS-CoV-2 continues to spread through the global population, it might seem like the virus is optimally adapted to humans, however convergent evolution of multiple lineages and the ongoing emergence of divergent lineages suggests that SARS-CoV-2 is becoming even more fit. Additionally, it also unfortunately suggests that if the global burden of infection continues, SARS-CoV-2 may be able to respond to the changing human immune landscape. Future evolution studies within and between hosts, across a variety of hosts and virus subtypes, are needed to understand the trajectory of SARS-CoV-2 evolution.

Future work on within and between host evolution of zoonotic respiratory RNA viruses

It is worth testing the hypothesis that the pace of adaptive evolution of IAV and SARS-CoV-2 in humans can be predicted by the prevalence of recent viral infections. If true, this supports the "jackpot event" model of evolution where a virus overcomes the stochastic forces driving within-host infections by infecting a large number of hosts and making incremental fitness jumps via relatively rare events.

Relatedly, there is an emerging hypothesis that prolonged IAV and SARS-CoV-2 infection allows for more time for selection to drive newly arising variants to a level where they can be detected and/or where they are more likely to transmit onward to a new host. Even a modest increase in frequency within a donor host enhances the likelihood of a beneficial variant becoming fixed following transmission in the setting of a narrow transmission bottleneck. It is also possible for selection to act during transmission such that some viruses harboring a particular mutation or group of mutations are preferentially transmitted [204]. Consistent with this model, Xue et al [136] identified strong evidence for positive selection in four immunocompromised hosts infected with IAV for a prolonged period. In this study, several mutations arose independently in these hosts and a subset of these mutations subsequently circulated globally at high frequency the following flu season. Similarly, the only time SARS-CoV-2 variants of concern have been detected sub-consensus within hosts are in rare cases of prolonged infection [265,266,377]. If true, this has implications for the importance of infection control and contact tracing of viral infections

in any cases of prolonged infection, particularly in immunocompromised hosts. Already, we have begun enrolling individuals with prolonged SARS-CoV-2 infection in a small case series. In the first individual enrolled in this case series we documented the emergence of Spike Δ142-144 and E484A. Interestingly, both of these mutations emerged independently in an immunocompromised individual in Boston [266]. Future studies could enroll a larger cohort of immunocompromised individuals or patients with prolong infection to characterize within-host viral evolution longitudinally and to document mutations emerging convergently across cases. Already it is apparent that collectively, prolonged infections pose a risk for the pace of the emergence of SARS-CoV-2 variants that escape immune recognition.

An exciting, though far-reaching, goal would be to work on uncovering the contours of these viral fitness landscapes using real data generated from a combination of experimental studies and modelling. Although studies would need to be designed with biosafety in mind, it would be valuable to investigate the evolutionary forces shaping SARS-CoV-2 in a wide range of host species, including the likely ancestral intermediate hosts, bats and pangolins. Any viruses sampled from the zoonotic reservoir or from humans are likely to be relatively fit viruses because they are currently replicating in nature. Therefore, to uncover the fitness valleys in this landscape, additional studies should evaluate SARS-CoV-2 genetically modified with bat-defining, pangolin-defining, and other zoonotic-defining mutations. Studying these genetically modified viruses which are not found in nature in mammalian systems might allow us to understand the routes that newly-emerging SARS-CoV-2 viruses might take in order to adapt to humans.

Additionally, novel sequencing strategies should be leveraged to more carefully evaluate viral population dynamics within and between individual hosts. Current SARS-CoV-2 sequencing strategies largely rely on short-amplicon sequencing in multiplexed PCRs. **Chapter 4** discusses the pitfalls of this sequencing approach and the caution that should be taken when making biological conclusions from low frequency iSNVs generated in similar studies. Novel sequencing approaches that use long-amplicon sequencing might be less prone to common method errors and would also preserve information about mutations that co-occur on a single molecule (linkage information). An additional promising approach for characterizing viral populations within and between hosts is to use libraries of molecularly-barcoded viruses in order to track individual members of the virus population. Though not included in the primary chapters of my thesis, I have worked closely with the Mehle Laboratory to develop sequencing approaches and bioinformatic tools to analyze populations of molecularly-barcoded IAV viruses. These methods could be applied to SARS-CoV-2 as a parallel method to quantify transmission bottlenecks and to investigate compartmentalization and movement of viruses within individual hosts.

Using viral sequencing to characterize patterns of viral spread in nearby communities and to evaluate the effectiveness of mitigation strategies

In chapter 5, I described our use of viral sequencing across individuals to evaluate disparate patterns of introduction and spread. In this study we sequenced viruses from two counties in Wisconsin which are less than 100km apart from each other, but have important demographic differences – Dane and Milwaukee counties. Other

contemporaneous studies look at patterns of SARS-CoV-2 introduction and spread in one city or geographic region [147–149,298,337,338,340], but none compared patterns of viral spread in nearby communities and contextualized these differences in the demographic and socioeconomic features of each community.

Working closely with the Milwaukee Health Department and other public health partners throughout the state, we were able to sequence a representative set of viruses circulating in Dane County as well as viruses circulating in Milwaukee County during spring 2020. We found Dane County's outbreak was defined by an early introduction of European-lineage viruses that contained the Spike D614G variant, whereas Milwaukee County's outbreak was defined by more frequent introductions of Asian-lineage viruses. Subsequent research showed Spike D614G confers increased transmissibility and its clade, 20G, outpaced growth in other clades during the summer and fall of 2020 [378–382]. Despite this, Dane County actually saw less community spread than Milwaukee County. We hypothesized that this difference in early viral spread was driven by human and behavioral factors as opposed to virus-specific factors. We were not able to directly test this hypothesis, but did compare the social vulnerability index in each county, which is a metric referring to the potential negative effects on communities caused by external stresses on human health (like a global pandemic) [383]. Milwaukee County has a higher social vulnerability index, meaning this community was more vulnerable, compared to Dane County (0.8268 vs 0.1974). We speculated that factors contributing to the SVI, like population density, access to financial, healthcare, and other support resources, as well

as race-based discrimination in healthcare systems, were likely at the root of these differences.

As part of this study, we sequenced viruses from before and after Wisconsin's state-wide "Safer-at-Home" order so we took advantage of this natural experiment to assess the effectiveness of this public health order. Amazingly, we showed evidence that viral spread in Dane and Milwaukee county fell by at least 40% following the "Safer-at-Home" order. These results were useful to state and public health officials who continued to advocate for state-wide distancing and density-reduction orders. In addition, the results of this work suggested that patterns of viral introduction and spread in very nearby communities can differ and careful characterization of these differences might allow public health interventions to be targeted to particular pattern of spread in a community – we often referred to this relatively novel concept as "precision public health".

The final chapter of this dissertation, **chapter 6**, describes our work using viral sequencing as an infection control tool in the setting of a large academic medical system in the upper midwestern United States. In accordance with CDC recommendations, this healthcare system implemented a number of policies and procedures aimed at protecting patients and healthcare personnel (HCP) from becoming infected with SARS-CoV-2 in the setting of the healthcare system [361]. To assess the effectiveness of these strategies, we sequenced viruses collected from HCP who became infected with SARS-CoV-2, their patient contacts, and cases circulating in the surrounding community. We found the majority of HCP infections could not be linked to a patient or co-worker and were more

likely to occur outside of the workplace. We found a smaller percentage could be traced to a coworker. The smallest proportion of HCP infections could be clearly traced to a patient source. These results suggested infection control measures in place at the institution evaluated in this case series were successfully working to protect HCP from healthcare-associated SARS-CoV-2 infections. To our knowledge, this was the first retrospective case series using viral sequencing to specifically investigate the source of SARS-CoV-2 infections in healthcare workers in the United States. A related study was published on similar work that was conducted in the Netherlands [374]. Like us, this group found little evidence for widespread nosocomial transmission of SARS-CoV-2.

Future directions on the application of viral sequencing to the health of humans and populations

Taken together, **chapters 5** and **6** illustrate that viral sequencing can augment traditional public health methods as well as infection control interventions. Already, seeing the value in viral sequencing to supplement track-and-trace efforts, Wisconsin's public health groups have increased their capacity to surveil viruses circulating in the state through high throughput viral sequencing and have created dashboards to publicly report the summary findings of these efforts. In addition, the infection control group has integrated viral sequencing into their investigation of all possible healthcare-associated transmission of SARS-CoV-2. Additionally, in collaboration with the infection control team, we have begun incorporating viral sequencing into investigations of SARS-CoV-2 breakthrough following infection and vaccination. The COVID-19 pandemic has shined light on the value of viral sequencing applied to public health datasets and even more importantly has

shined light on the value of collaborations between academic groups and public health partners.

Altogether, the work in this dissertation contributes to our understanding of IAV and SARS-CoV-2 evolution within and between individual hosts and in populations. The observations made here have contributed to funded grant proposals and additional ongoing studies, which are also likely to further advance our understanding of RNA virus evolution.

Figures, tables, and supplemental material

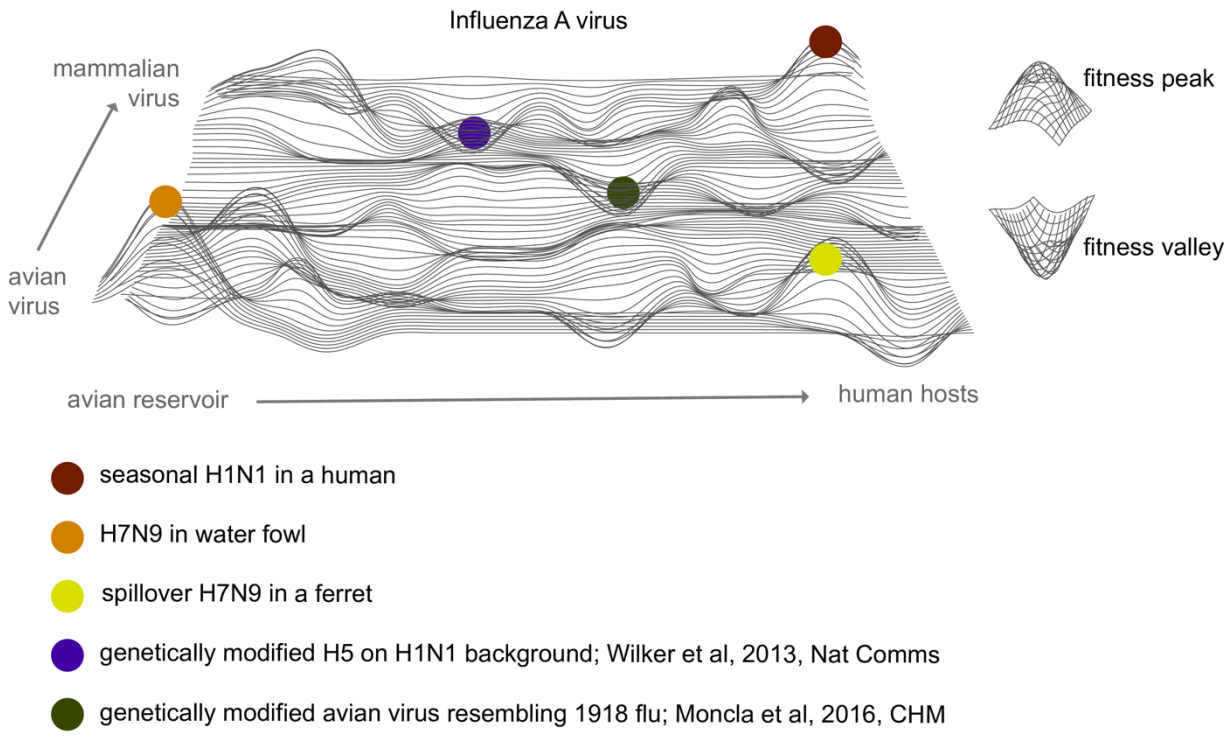


Figure 1 depicts a theoretical fitness landscape for IAV

The y-axis of this figure represents virus genotypes as a virus transitions from avian to mammalian, the x-axis represents host genetics and conditions ranging from the avian reservoir to humans, and the z-axis (height) represents relative fitness at all possible x-y coordinates. The peaks and valleys shown here are conceptual and not derived from actual data.

Appendix:

Contributions to co-authored manuscripts

An updated influenza A(H3N2) vaccine generates limited antibody responses to previously encountered antigens in children

Kelsey R. Florek[1,a], Luiza M. Campos[a], Katarina M. Braun[a], Huong Q. McLean[b], Jennifer P. King[b], Brendan Flannery[c], Edward A. Belongia[b,*], Thomas C. Friedrich[a,d,*]

[a]Department of Pathobiological Sciences, University of Wisconsin School of Veterinary Medicine, Madison, WI 53706, USA
[b]Center for Clinical Epidemiology and Population Health, Marshfield Clinic Research Institute, 1000 North Oak Ave, Marshfield 54449, WI, USA
[c]Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta 30333, GA, USA
[d]Wisconsin National Primate Research Center, Madison, WI 53715, USA

Abstract

Background

Influenza vaccination may provide a “back-boost” to antibodies against previously encountered strains. If the back-boost effect is common, this could allow more aggressive vaccine updates, as emerging variants would be expected to both elicit de-novo responses and boost pre-existing responses against recently circulating strains. Here we used the emergence of an antigenically novel A(H3N2) strain to determine whether an antigenically updated vaccine boosted antibodies against historical strains.

Methods

We performed hemagglutination-inhibition (HI) assays on pre- and post-vaccination sera from 124 children 5–17 years old who received 2015–2016 inactivated influenza vaccine, containing an antigenically updated A(H3N2) strain. We evaluated the mean fold increase in HI titer against both the 2015–2016 vaccine strain and representative strains from two prior antigenic clusters. Factors associated with post-vaccination titers against historical strains were evaluated using linear regression, adjusting for baseline titer.

Results

Geometric mean titers against each antigen examined increased significantly after vaccination ($P < .0001$). Mean fold increase was 3.29 against the vaccine strain and 1.22–1.46 against historical strains. Response to vaccine strain was associated with increased post-vaccination titers against historical strains.

Conclusions
A vaccine containing an antigenically novel A(H3N2) strain modestly boosted antibody responses against historical influenza strains in children.

Vaccine 36.5 (2018): 758-764.

Contributions
I performed hemagglutination-inhibition assays, assisted with data analysis, and edited the manuscript.

Influenza evolution with little host selection

Katarina M. Braun[1] & Thomas C. Friedrich[1]

[1]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, USA

Nat Ecol Evol 3, 159–160 (2019). https://doi.org/10.1038/s41559-018-0782-1

Abstract
Influenza viruses undergo rapid antigenic evolution. Analysis of a large dataset of influenza virus sequences, using host age as a proxy for immune experience, shows no evidence for immune positive selection driving antigenic evolution in individual infected humans.

Contributions
I reviewed the manuscript this commentary was based on (*Han, A.X., Maurer-Stroh, S. & Russell, C.A. Individual immune selection pressure has limited impact on seasonal influenza virus evolution. Nat Ecol Evol 3, 302–311 (2019). https://doi.org/10.1038/s41559-018-0741-x*) for Nature Ecology and Evolution and drafted this News and Views article with Dr. Friedrich.

African-lineage Zika virus replication dynamics and maternal-fetal interface infection in pregnant rhesus macaques

Chelsea M. Crooks[a], Andrea M. Weiler[b], Sierra L. Rybarczyk[b,*], Mason Bliss[b], Anna S. Jaeger[c], Megan E. Murphy[d,*], Heather A. Simmons[b], Andres Mejia[b], Michael K. Fritsch[e], Jennifer M. Hayes[b], Jens C. Eickhoff[f], Ann M. Mitzey[d], Elaina Razo[g], Katarina M. Braun[a], Elizabeth A. Brown[a], Keisuke Yamamoto[e,*], Phoenix M. Shepherd[e], Amber Possell[b], Kara Weaver[b], Kathleen M. Antony[h], Terry K. Morgan[i,j], Xiankun Zeng[k], Dawn M. Dudley[e], Eric Peterson[b], Nancy Schultz-Darken[b], David H. O'Connor[b,e], Emma L. Mohr[g], Thaddeus G. Golos[b,d,h], Matthew T. Aliota[c], and Thomas C. Friedrich[a,b]#

[a]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, USA
[b]Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI, USA
[c]Department of Veterinary and Biomedical Sciences, University of Minnesota,Twin Cities, St. Paul, MN, USA
[d]Department of Comparative Biosciences, University of Wisconsin-Madison, Madison, WI, USA
[e]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI, USA
[f]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA
[g]Department of Pediatrics, University of Wisconsin-Madison, Madison, WI, USA
[h]Department of Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, WI, USA
[i]Department of Pathology, Oregon Health and Science University, Portland, OR, USA
[j]Department of and Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, USA
[k]US Army Medical Research Institute of Infectious Disease, Fort Detrick, MD, USA

Accepted for publication in the Journal of Virology, 2021-05-06.

Abstract
Following the Zika virus (ZIKV) outbreak in the Americas, ZIKV was causally associated with microcephaly and a range of neurological and developmental symptoms, termed congenital Zika syndrome (CZS). The viruses responsible for this outbreak belonged to the Asian lineage of ZIKV. However, in-vitro and in-vivo studies assessing the pathogenesis of African-lineage ZIKV demonstrated that African-lineage isolates often replicated to high titer and caused more severe pathology than Asian-lineage isolates. To date, the pathogenesis of African-lineage ZIKV in a translational model, particularly during pregnancy, has not been rigorously characterized. Here we infected four pregnant rhesus macaques with a low-passage strain of African-lineage ZIKV and compared its pathogenesis to a cohort of four pregnant rhesus macaques infected with an Asian-lineage isolate and a cohort of mock-inoculated controls. Viral replication kinetics were not significantly different between the two experimental groups and both

groups developed robust neutralizing antibody titers above levels considered to be protective. There was no evidence of significant fetal head growth restriction or gross fetal harm at delivery (1-1.5 weeks prior to full term) in either group. However, a significantly higher burden of ZIKV vRNA was found in maternal-fetal interface tissues in the macaques exposed to an African-lineage isolate. Our findings suggest that ZIKV of any genetic lineage poses a threat to pregnant individuals and their infants.

Contributions

I performed placental dissections and assisted with tissue processing. I provided feedback on figure design and edited the manuscript.

Prior dengue immunity enhances Zika virus infection of the maternal-fetal interface in rhesus macaques

C. M. Crooks[1], A. M. Weiler[2], S. L. Rybarczyk[2,†], M. I. Bliss[2], A. S. Jaeger[3], M. E. Murphy[4,‡], H. A. Simmons[2], A. Mejia[2], M. K. Fritsch[5], J. M. Hayes[2], J. C. Eickhoff[6], A. M. Mitzey[4], E. Razo[7], K. M. Braun[1], E. A. Brown[1], K. Yamamoto[5,§], P. M. Shepherd[5], A. Possell[2], K. Weaver[2], K. M. Antony[8], T. K. Morgan[9,10], C. M. Newman[5], D. M. Dudley[5], N. Schultz-Darken[2], E. Peterson[2], L. C. Katzelnick[11,‖], A. Balmaseda[12], E. Harris[11], D. H. O'Connor[2,5], E. L. Mohr[7], T. G. Golos[2,4,8], T. C. Friedrich[1,2]*, and M. T. Aliota[3]*

[1]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, USA
[2]Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI, USA
[3]Department of Veterinary and Biomedical Sciences, University of Minnesota, Twin Cities, St. Paul, MN, USA
[4]Department of Comparative Biosciences, University of Wisconsin-Madison, Madison, WI, USA
[5]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI, USA
[6]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA
[7]Department of Pediatrics, University of Wisconsin-Madison, Madison, WI, USA
[8]Department of Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, WI, USA
[9]Department of Pathology, Oregon Health and Science University, Portland, OR, USA
[10]Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, USA
[11]Division of Infectious Diseases and Vaccinology, University of California Berkeley, Berkeley, CA, USA
[12]Sustainable Sciences Institute, Managua, Nicaragua

*Corresponding author, T.C. Friedrich, Email: tfriedri@wisc.edu
*Corresponding author, M.T. Aliota, Email: mtaliota@umn.edu

Preprint. bioRxiv.

Abstract
Concerns have arisen that pre-existing immunity to dengue virus (DENV) could enhance Zika virus (ZIKV) disease, due to the homology between ZIKV and DENV and the observation of antibody-dependent enhancement (ADE) among DENV serotypes. To date, no study has examined the impact of pre-existing DENV immunity on ZIKV pathogenesis during pregnancy in a translational non-human primate model. Here we show that prior DENV-2 exposure enhanced ZIKV infection of maternal-fetal interface tissues in macaques. However, pre-existing DENV immunity had no detectable impact on ZIKV replication kinetics in maternal plasma, and all pregnancies progressed to term

without adverse outcomes or gross fetal abnormalities detectable at delivery. Understanding the risks of ADE to pregnant women worldwide is critical as vaccines against DENV and ZIKV are developed and licensed and as DENV and ZIKV continue to circulate.

Contributions

I performed placental dissections and assisted with tissue processing. I provided feedback on figure design and edited the manuscript.

Characteristic and quantifiable COVID-19-like abnormalities in CT- and PET/CT-imaged lungs of SARS-CoV-2-infected crab-eating macaques (*Macaca fascicularis*)

Courtney L. Finch[1], Ian Crozier[2], Ji Hyun Lee[1], Russ Byrum[1], Timothy K. Cooper[1,4], Janie Liang[1], Kaleb Sharer[1], Jeffrey Solomon[2], Philip J. Sayre[1], Gregory Kocher[1,5], Christopher Bartos[1], Nina M. Aiosa[1,3], Marcelo Castro[1], Peter A. Larson[4], Ricky Adams[1,6], Brett Beitzel[4], Nicholas Di Paola[4], Jeffrey R. Kugelman[4], Jonathan R. Kurtz[1,7], Tracey Burdette[1], Martha C. Nason[5], Irwin M. Feuerstein[1], Gustavo Palacios[4,8], Marisa C. St. Claire[1], Matthew G. Lackemeyer[1], Reed F. Johnson[1], Katarina M. Braun[6,9], Mitchell D. Ramuta[7], Jiro Wada[1], Connie S. Schmaljohn[1], Thomas C. Friedrich[6,8*,10], David H. O'Connor[7,8,*], and Jens H. Kuhn[1,*]

[1]Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Fort Detrick, Frederick, MD 21702, USA
[2]Integrated Research Facility at Fort Detrick, Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research supported by the National Cancer Institute, Frederick, MD 21702, USA
[3]Center for Infectious Disease Imaging, Warren G Magnuson Clinical Center, National Institutes of Health, Bethesda, MD, 20814, USA
[4]United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21702, USA
[5]Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20892, USA
[6]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA
[7]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA
[8]Wisconsin National Primate Research Center, Madison, WI 53706, USA

*These authors contributed equally to this work

Preprint. bioRxiv. 2020;2020.05.14.096727. Published 2020 May 14.
doi:10.1101/2020.05.14.096727

Abstract
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is causing an exponentially increasing number of coronavirus disease 19 (COVID-19) cases globally. Prioritization of medical countermeasures for evaluation in randomized clinical trials is critically hindered by the lack of COVID-19 animal models that enable accurate, quantifiable, and reproducible measurement of COVID-19 pulmonary disease free from observer bias. We first used serial computed tomography (CT) to demonstrate that bilateral intrabronchial instillation of SARS-CoV-2 into crab-eating macaques (Macaca fascicularis) results in mild-to-moderate lung abnormalities qualitatively characteristic of subclinical or mild-to-moderate COVID-19 (e.g., ground-glass opacities with or without reticulation, paving, or alveolar consolidation, peri-bronchial thickening, linear opacities)

at typical locations (peripheral>central, posterior and dependent, bilateral, multi-lobar). We then used positron emission tomography (PET) analysis to demonstrate increased FDG uptake in the CT-defined lung abnormalities and regional lymph nodes. PET/CT imaging findings appeared in all macaques as early as 2 days post-exposure, variably progressed, and subsequently resolved by 6–12 days post-exposure. Finally, we applied operator-independent, semi-automatic quantification of the volume and radiodensity of CT abnormalities as a possible primary endpoint for immediate and objective efficacy testing of candidate medical countermeasures.

Contributions
I assisted with experimental design, figure generation, and editing of the manuscript.

Limited SARS-CoV-2 diversity within hosts and following passage in cell culture

Gage K. Moreno[1]*, Katarina M. Braun[2]*, Peter J. Halfmann[2,3], Trent M. Prall[1], Kasen K. Riemersma[2], Amelia K. Haj[1], Joseph Lalli[2], Kelsey R. Florek[3], Yoshihiro Kawaoka[2,4], Thomas C. Friedrich[2,4], David H. O'Connor[1,5, #]

* These authors contributed equally to this work

[1]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI, United States of America; [2]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, United States of America; [3]Wisconsin State Laboratory of Hygiene, Madison, WI, United States of America; [4]Influenza Research Institute, School of Veterinary Sciences, University of Wisconsin-Madison, Madison, WI, United States; [5]Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI, United States of America

Preprint. bioRxiv 2020.04.20.051011; doi: https://doi.org/10.1101/2020.04.20.051011

Abstract
Since the first reports of pneumonia associated with a novel coronavirus (COVID-19) emerged in Wuhan, Hubei province, China, there have been considerable efforts to sequence the causative virus, SARS-CoV-2 (also referred to as hCoV-19) and to make viral genomic information available quickly on shared repositories. As of 30 March 2020, 7,680 consensus sequences have been shared on GISAID, the principal repository for SARS-CoV-2 genetic information. These sequences are primarily consensus sequences from clinical and passaged samples, but few reports have looked at diversity of virus populations within individual hosts or cultures. Understanding such diversity is essential to understanding viral evolutionary dynamics. Here, we characterize within-host viral diversity from a primary isolate and passaged samples, all originally deriving from an individual returning from Wuhan, China, who was diagnosed with COVID-19 and subsequently sampled in Wisconsin, United States. We use a metagenomic approach with Oxford Nanopore Technologies (ONT) GridION in combination with Illumina MiSeq to capture minor within-host frequency variants ≥1%. In a clinical swab obtained from the day of hospital presentation, we identify 15 single nucleotide variants (SNVs) ≥1% frequency, primarily located in the largest gene – ORF1a. While viral diversity is low overall, the dominant genetic signatures are likely secondary to population size changes, with some evidence for mild purifying selection throughout the genome. We see little to no evidence for positive selection or ongoing adaptation of SARS-CoV-2 within cell culture or in the primary isolate evaluated in this study.

Contributions
I generated the deep sequencing dataset. I analyzed the data, and created the figure displays. I wrote the first draft of the manuscript and edited the manuscript.

Using Virus Sequencing to Determine Source of SARS-CoV-2 Transmission for Healthcare Worker

Nasia Safdar[1,2], Gage K. Moreno[3], Katarina M. Braun[3], Thomas C. Friedrich[3], and David H. O'Connor[3]

[1]University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA; [2]William S. Middleton Memorial Veterans Hospital, Madison; [3]University of Wisconsin–Madison, Madison

Emerging Infectious Diseases. 2020 Aug 6;26(10). PMID: 32758345

Abstract
Whether a healthcare worker's severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is community or hospital acquired affects prevention practices. We used virus sequencing to determine that infection of a healthcare worker who cared for 2 SARS-CoV-2–infected patients was probably community acquired. Appropriate personal protective equipment may have protected against hospital-acquired infection.

Contributions
I generated and analyzed the sequence dataset. I wrote the first draft of the manuscript, and assisted in manuscript editing.

Circulation of a Spike ∆69/70 SARS-CoV-2 lineage (B.1.375) in the US and evidence of reinfection

Brendan B. Larsen[1*], Gage K. Moreno[2*], Katarina M. Braun[3*], Anderson Brito4[*], Tara Alpert[4], Derly A. Molina[5], Juan Carlos F. Cadena[5], Mallery Breban[4], Bernardo G. Granja[6], Oliver G. Pybus[6], Nathan D. Grubaugh[4‡], Thomas C. Friedrich[3‡], David H. O'Connor[2‡], Joseph R. Fauver[4‡], Michael Worobey[‡1]

* These authors contributed equally to this article
‡Corresponding authors (nathan.grubaugh@yale.edu, tfriedri@wisc.edu, dhoconno@wisc.edu, joseph.fauver@yale.edu, worobey@arizona.edu)

bioRxiv and Virological.org

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona; [2]Department of Pathology and Laboratory Medicine, University of Wisconsin - Madison, Madison, Wisconsin, USA; [3]Department of Pathobiological Sciences, University of Wisconsin - Madison, Madison, Wisconsin, USA; [4]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut, USA; [5]Universidad de Especialidades Espíritu Santo, Guayaquil, Ecuador; [6]Department of Zoology, University of Oxford, Oxford, UK

Abstract
We identified a new SARS-CoV-2 lineage with Spike deletion ∆69/70 (designated 'B.1.375') that has been circulating in the United States since September 2020. We characterize the timing and geographic spread of this lineage and report a possible case of reinfection with B.1.375 virus.

Contributions
I generated the deep sequencing dataset. I assisted in analyzing the data and provided feedback on data visualization. I additionally assisted in writing the first draft of the manuscript, and manuscript editing.

Rapid evolution of enhanced Zika virus virulence during direct vertebrate transmission chains

Kasen K. Riemersma[a], Anna S. Jaeger[b], Chelsea M. Crooks[a], Katarina M. Braun[2], James Weger-Lucarelli[c], Gregory D. Ebel[3], Thomas C. Friedrich[a], Matthew T. Aliota[b]

[a]University of Wisconsin–Madison, Madison, Wisconsin, USA
[b]University of Minnesota, Twin Cities, St. Paul, Minnesota, USA
[c]Virginia Tech, Blacksburg, Virginia, USA
[d]Colorado State University, Fort Collins, Colorado, USA

Journal of Virology Mar 2021, 95 (8) e02218-20; DOI: 10.1128/JVI.02218-20

Abstract
Zika virus (ZIKV) has the unusual capacity to circumvent natural alternating mosquito-human transmission and be directly transmitted human to human via sexual and vertical routes. The impact of direct transmission on ZIKV evolution and adaptation to vertebrate hosts is unknown. Here, we show that molecularly barcoded ZIKV rapidly adapted to a mammalian host during direct transmission chains in mice, coincident with the emergence of an amino acid substitution previously shown to enhance virulence. In contrast, little to no adaptation of ZIKV to mice was observed following chains of direct transmission in mosquitoes or alternating host transmission. Detailed genetic analyses revealed that ZIKV evolution in mice was generally more convergent and subjected to more relaxed purifying selection than that in mosquitoes or alternate passages. These findings suggest that prevention of direct human transmission chains is paramount to resist gains in ZIKV virulence.

Contributions
I created and validated the wet-lab and bioinformatic protocols used to sequence barcoded ZIKV viruses in this project: unique molecular identifier (UMI) sequencing for barcoded viruses.

Gender can influence student experiences in MD-PhD Training

Anna S Heffron[1,2]*, Katarina M Braun[1,3]*, Cora Allen-Savietta[4], Amarette Filut[5], Chelsea Hanewall [1], Anna Huttenlocher[1,6], Jo Handelsman[7], Molly Carnes[5,8,9,10]

[1]Medical Scientist Training Program, University of Wisconsin-Madison, Madison, Wisconsin, USA.
[2]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA.
[3]Department of Pathobiological Sciences, and University of Wisconsin-Madison, Madison, Wisconsin, USA.
[4]Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin, USA.
[5]Center for Women's Health Research, University of Wisconsin-Madison, Madison, Wisconsin, USA.
[6]Department of Medical Microbiology & Immunology and Pediatrics, University of Wisconsin-Madison, Madison, Wisconsin, USA.
[7]Department of Plant Pathology, Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, Wisconsin, USA.
[8]Department of Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA.
[9]Department of Psychiatry, and University of Wisconsin-Madison, Madison, Wisconsin, USA.
[10]Department of Industrial & Systems Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA.

Star: * indicates co-first authorship

J Womens Health (Larchmt). 2021 Jan;30(1):90-102. doi: 10.1089/jwh.2019.8094. Epub 2020 Apr 28. PMID: 32349608; PMCID: PMC7826439.

Abstract

Background:
Female physician-scientists have led major advances in medicine broadly and particularly in women's health. Women remain underrepresented in dual MD-PhD degree programs that train many physician-scientists despite gender parity among medical and biomedical research students.

Materials and Methods:
To explore how the training environment might be experienced differently for male and female students in one MD-PhD program, the authors analyzed gender differences in annual symposium speakers with exact binomial tests, student participation as question-askers at a weekly seminar with logistic regression, and number of publications with quasi-Poisson generalized linear models. They compared male and female students' perceptions of gender-based discrimination using a survey, including qualitative analysis of free text responses. The program consisted of 71 total students in the 2017-2018 and 2018-2019 academic years. Female students comprised 42.0% (81/191) of program matriculants from 1997 to 2019.

Results:
Male and female students were equally likely to present at the annual program symposium, but faculty ($p = 0.001$) and keynote ($p = 0.012$) presenters were more likely to be male. Compared with male counterparts, female students asked fewer seminar questions ($p < 0.005$) and female speakers received more questions ($p = 0.03$). Female students perceived less support and differed from men in reasons for asking or not asking seminar questions. Free text responses described repeated small acts of discrimination toward women with cumulative impact. Positive program changes followed presentation of findings to program leaders and students.

Conclusions:
The authors identified several aspects of one MD-PhD program that could discourage career or training persistence of female students. Increasing awareness of these issues was temporally related to positive programmatic changes.

Contributions
I designed the study, collected the data, analyzed the data, generated figures, wrote and edited the manuscript, and took this project through the review process.

Initial Outbreak of COVID-19 on a University Campus and Interventions to Disrupt Transmission, Wisconsin, August - October, 2020

Dustin W. Currie[1,2]*, Gage K. Moreno[3]*, Miranda J. Delahoy[1,2], Ian W. Pray[1,2,4], Amanda Jovaag[5], Katarina M. Braun[6], Devlin Cole[4,7], Todd Shechter[5], Geroncio C. Fajardo[1], Carol Griggs[5], Brian S. Yandell[8,9], Steve Goldstein[9], Dena Bushman[1,2], Hannah Segaloff,[1,2,4] G. Patrick Kelly[5], Collin Pitts[5,7], Christine Lee[1,10], Katarina M. Grande[11], Amanda Kita-Yarbro[11], Brittany Grogan[11], Sara Mader[11], Jake Baggott[5], Allen C. Bateman[12], Ryan P.Westergaard[4,7], Jacqueline E. Tate[1], Thomas C. Friedrich[6], Hannah L. Kirking[1], David H. O'Connor[3], Marie E. Killerby[1]

* These authors contributed equally to this work.

[1]CDC COVID-19 Response Team, Atlanta, GA
[2]Epidemic Intelligence Service, Centers for Disease Control and Prevention, Atlanta, GA; [3]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI
[4]Wisconsin Department of Health Services, Madison, WI
[5]University Health Services, University of Wisconsin-Madison, Madison, WI
[6]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI
[7]School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI
[8]Department of Statistics, University of Wisconsin-Madison, Madison, WI
[9]American Family Insurance Data Science Institute, University of Wisconsin-Madison, Madison, WI
[10]Laboratory Leadership Service, CDC
[11]Public Health Madison and Dane County, Madison, WI
[12]Wisconsin State Laboratory of Hygiene, Madison, WI

Under review.

Abstract

University settings have demonstrated potential for COVID-19 outbreaks, as they can combine congregate living, substantial social activity, and a young population predisposed to mild illness. Using genomic and epidemiologic data, we describe a COVID-19 outbreak at the University of Wisconsin (UW)–Madison. During August – October 2020, 3,485 students tested positive, including 856/6,162 students living in residence halls. Case counts began rising during move-in week for on-campus students (August 25-31, 2020), then rose rapidly during September 1-11, 2020. UW-Madison initiated multiple prevention efforts, including quarantining two residence halls; a subsequent decline in cases was observed. Genomic surveillance of cases from Dane County, where UW-Madison is located, did not find evidence of transmission from a large cluster of cases in the two residence halls quarantined during the outbreak.

Coordinated implementation of prevention measures can effectively reduce SARS-CoV-2 spread in university settings and may limit spillover to the community surrounding the university.

Contributions

I generated the deep sequencing dataset. I assisted in analyzing the data and provided feedback on data visualization. I additionally assisted in writing the first draft of the manuscript and provided edits on manuscript revisions.

SARS-CoV-2 transmission in intercollegiate athletics not fully mitigated with daily antigen testing

Gage K. Moreno[1]*, Katarina M. Braun[2]*, Ian W. Pray[3,4], Hannah E. Segaloff[3,4], Ailam Lim[5], Keith Poulsen[5], Jonathan Meiman[3], James Borcher[6], Ryan P. Westergaard[3,7], Michael K. Moll[8], Thomas C. Friedrich[2], David H. O'Connor[1]

* These authors contributed equally

[1] Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison USA 53711
[2] Department of Pathobiological Sciences, University of Wisconsin-Madison USA 53711
[3] Wisconsin Department of Health Services, USA 53703
[4] Epidemic Intelligence Service, Centers for Disease Control and Prevention USA 30333
[5] Wisconsin Veterinary Diagnostic Laboratory, University of Wisconsin-Madison USA 53711
[6] Department of Family Medicine, Division of Sports Medicine, Ohio State University USA 43210
[7] Department of Medicine, University of Wisconsin-Madison, USA 53711
[8] Athletic Department, University of Wisconsin-Madison USA 53711

Accepted for publication at Clinical Infectious Diseases.

Abstract

High-frequency, rapid-turnaround SARS-CoV-2 testing continues to be proposed as a way of efficiently identifying and mitigating transmission in congregate settings. However, two SARS-CoV-2 outbreaks occurred among intercollegiate university athletic programs during the fall 2020 semester despite mandatory directly observed daily antigen testing. During the fall 2020 semester, athletes and staff in both programs were tested daily using Quidel's Sofia SARS Antigen Fluorescent Immunoassay (FIA), with positive antigen results requiring confirmatory testing with real-time reverse transcription polymerase chain reaction (RT-PCR). We used genomic sequencing to investigate transmission dynamics in these two outbreaks. In Outbreak 1, 32 confirmed cases occurred within a university athletics program after the index patient attended a meeting while infectious despite a negative antigen test on the day of the meeting. Among isolates sequenced from Outbreak 1, 24 (92%) of 26 were closely related, suggesting sustained transmission following an initial introduction event. In Outbreak 2, 12 confirmed cases occurred among athletes from two university programs that faced each other in an athletic competition despite receiving negative antigen test results on the day of the competition. Sequences from both teams were closely related and distinct from viruses circulating in Team 1's community, suggesting transmission during intercollegiate competition in Team 2's community. These findings suggest that antigen testing alone, even when mandated and directly observed, may not be sufficient as an intervention to prevent SARS-CoV-2 outbreaks in congregate settings, and highlight the importance of supplementing serial antigen testing with appropriate mitigation strategies to prevent SARS-CoV-2 outbreak in congregate settings.

Contributions

I generated the deep sequencing dataset. I analyzed the data and performed data visualization. I wrote the first draft of the manuscript, performed manuscript editing, and carried the manuscript through peer review.

Evidence of Early Household Transmission of SARS-CoV-2 Involving a School-aged Child – Dane County, Wisconsin, March 2020

Jonathan L. Temte, MD, PhD[1]; Shari Barlow, BA[1]; Emily Temte, BA[1]; Maureen Goss, MPH[1]; Kelsey Florek, PhD,MPH[2]; Katarina M. Braun, MD-PhD candidate[3]; Thomas C. Friedrich, PhD[3]; Erik Reisdorf, MS[2]; Allen C. Bateman, PhD, MPH[2]; Amra Uzicanin, MD, MPH[4]

[1] University of Wisconsin, School of Medicine and Public Health, Department of Family Medicine and Community Health, Madison, WI
[2] Wisconsin State Laboratory of Hygiene, Madison, WI
[3] University of Wisconsin, School of Veterinary Medicine, Department of Pathobiological Sciences, Madison, WI
[4] U.S. Centers for Disease Control and Prevention, Atlanta, GA

This paper has been submitted, but is not yet under review.

Abstract

Importance:
Little is known about the role of school-aged children and household transmission in the early phases of the SARS-CoV-2 pandemic in the United States.

Objective:
To evaluate the prevalence of SARS-CoV-2 in a low-risk population of school-aged children and assess possible household transmission.

Design:
Longitudinal, community-based influenza and other respiratory virus surveillance study.
Setting: Oregon School District, Dane County, Wisconsin from December 1, 2019 through June 30, 2020.

Participants:
School-aged children, aged 4–18 years, with acute respiratory infections, and household members participating in a household influenza transmission sub-study.

Intervention:
None

Main Outcome Measures:
Detection of SARS-CoV-2 using RT-PCR in 567 archived nasal swab specimens. Evaluation of virus lineage using whole genome sequencing.

Results:

Very low prevalence (0.2% [95% CI: 0.03–0.99]) of SARS-CoV-2 was detected in this population of school-aged children during the analysis period. The single case detected in March 2020 was associated with SARS-CoV-2 detection in all other household members. All sequences were identical or near-identical to Clade 17B (A.4 lineage).

Conclusions and Relevance:
In the very early phases of the SARS-CoV-2 pandemic, infection in school-aged children was associated with strong evidence of household transmission. Such unrecognized transmission likely contributed to wide seeding across populations.

Contributions
I generated the and analyzed the deep sequence data included in this manuscript.

Molecular barcodes for high-resolution characterization of the bottlenecks governing influenza A virus infection and transmission

Katherine A Amato[1]*, Katarina M Braun[2]*, Luis A Haddock III[2], Grace A Schaack[1], Christina A Higgins[1], Emma C Boehm[2], Gabrielle L Barry[2], Victoria Meliopoulos[3], Brandi Livingston[3], Rebekah Honce[3], Stacey Schultz-Cherry[3], Thomas C Friedrich[2,4], Andrew Mehle[1]

* These authors contributed equally

[1]Department of Medical Microbiology & Immunology, University of Wisconsin-Madison, Madison, WI 53706, USA
[2]Department of Pathobiological Sciences, University of Wisconsin School of Veterinary Medicine, Madison, WI 53706, USA
[3]St. Jude CHildren's ResearchHospital, Department of Infectious Diseases, Memphis TN 38112, USA
[4]Wisconsin National Primate Research Center, Madison, WI 53715, USA

Abstract
Transmission of influenza A virus (IAV) is constrained by numerous physical and biological barriers to infection.IAV overcomes these barriers in part by low-fidelity replication that drives the emergence of new variants, but bottlenecks limit population-level diversity and shape evolutionary pathways taken during infection and transmission. High-resolution characterization of the nature and stringency of bottlenecks on IAV infection has not been previously possible. To characterize viral populations and quantify bottlenecks, we created two molecularly barcoded IAV libraries, each with ~60,000 individual members identifiable by deep sequencing. Our barcoded libraries captured selective sweeps as adaptive mutations arose in HA during tissue culture passage. The barcode revealed that the adaptive variant arose only once in the diverse starting population followed by a selective sweep to dominate as ~70% of all HA genotypes. Using a similar technical approach, we produced two highly diverse and evenly distributed barcoded libraries with ~20,000 barcodes on the HA segment. Upon infection in mice, these barcoded viruses caused weight loss as well as viral titers at 3 and 6 days post-infection that were indistinguishable from the parental IAV. Whereas mice were inoculated with the entire library, barcode sequencing demonstrated an effective inoculation bottleneck size of (TBD) post inoculation. We then applied this quantitative approach to study the movement of viruses within infected animals by initiating site-specific infections in ferrets. We detected large bottlenecks during inoculation, where less than 10% of the viruses in our starting population initiated infection. While the majority of our virus established compartmentalized infections, small populations moved between upper and lower respiratory tract and made detectable contributions to the genetic diversity in each location. The use of large populations with uniquely quantifiable members allows us to measure population dynamics during IAV infection and identify major reductions in genetic diversity before and after transmission events.

Contributions

I worked closely with Katie Amato to characterize the barcoded viral libraries. I developed and validated sequencing approaches to perform amplicon sequencing of the barcoded regions and I developed bioinformatic tools sets to analyze and deploy clustering algorithms to the libraries of barcoded viruses. I processed, sequenced, and analyzed data from the mouse and ferret experiments outlined in this paper.

Bibliography

1. Hassell, J. M., Begon, M., Ward, M. J. & Fèvre, E. M. Urbanization and Disease Emergence: Dynamics at the Wildlife–Livestock–Human Interface. *Trends Ecol. Evol.* **32**, 55–67 (2017).
2. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* vol. 451 990–993 (2008).
3. Holmes, E. C. The Evolutionary Genetics of Emerging Viruses. (2009) doi:10.1146/annurev.ecolsys.110308.120248.
4. Cleaveland, S., Laurenson, M. K. & Taylor, L. H. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 991–999 (2001).
5. Taylor, L. H., Latham, S. M. & Woolhouse, M. E. Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **356**, 983–989 (2001).
6. Mazet, J. A. K. *et al.* A 'One Health' Approach to Address Emerging Zoonoses: The HALI Project in Tanzania. *PLoS Med.* **6**, e1000190 (2009).
7. Bardosh, K. *One Health: Science, politics and zoonotic disease in Africa*. (Routledge, 2016).
8. Tran, P. D. *et al.* Strengthening global health security capacity--Vietnam demonstration project, 2013. *MMWR Morb. Mortal. Wkly. Rep.* **63**, 77–80 (2014).
9. Borchert, J. N. *et al.* Rapidly building global health security capacity--Uganda demonstration project, 2013. *MMWR Morb. Mortal. Wkly. Rep.* **63**, 73–76 (2014).
10. Bird, B. H. & Mazet, J. A. K. Detection of Emerging Zoonotic Pathogens: An Integrated One Health Approach. *Annu Rev Anim Biosci* **6**, 121–139 (2018).

11. Krauss, S. & Webster, R. G. Avian influenza virus surveillance and wild birds: past and present. *Avian Dis.* **54**, 394–398 (2010).
12. Mehle, A. Unusual influenza A viruses in bats. *Viruses* **6**, 3438–3449 (2014).
13. Lu, Lu, L., Lycett, S. J. & Leigh Brown, A. J. Reassortment patterns of avian influenza virus internal segments among different subtypes. *BMC Evolutionary Biology* vol. 14 16 (2014).
14. Marshall, N., Priyamvada, L., Ende, Z., Steel, J. & Lowen, A. C. Influenza Virus Reassortment Occurs with High Frequency in the Absence of Segment Mismatch. *PLoS Pathog.* **9**, e1003421 (2013).
15. Steel, J. & Lowen, A. C. Influenza A virus reassortment. *Curr. Top. Microbiol. Immunol.* **385**, 377–401 (2014).
16. Tao, H., Li, L., White, M. C., Steel, J. & Lowen, A. C. Influenza A Virus Coinfection through Transmission Can Support High Levels of Reassortment. *J. Virol.* **89**, 8453–8461 (2015).
17. Rith, S. *et al.* Natural co-infection of influenza A/H3N2 and A/H1N1pdm09 viruses resulting in a reassortant A/H3N2 virus. *J. Clin. Virol.* **73**, 108–111 (2015).
18. Vijaykrishna, D., Mukerji, R. & Smith, G. J. D. RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. *PLoS Pathog.* **11**, e1004902 (2015).
19. Dugan, V. G. *et al.* The Evolutionary Genetics and Emergence of Avian Influenza Viruses in Wild Birds. *PLoS Pathogens* vol. 4 e1000076 (2008).
20. Taubenberger, J. K. & Kash, J. C. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **7**, 440–451 (2010).

21. Chen, R. & Holmes, E. C. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol. Biol. Evol.* **23**, 2336–2341 (2006).
22. Influenza (Seasonal). https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal).
23. CDC. Disease burden of influenza. https://www.cdc.gov/flu/about/burden/index.html (2020).
24. This year's flu season was virtually nonexistent. That could be bad news for next year. https://www.advisory.com/en/daily-briefing/2021/03/30/flu-season.
25. de Jong, J. C., Claas, E. C., Osterhaus, A. D., Webster, R. G. & Lim, W. L. A pandemic warning? *Nature* **389**, 554 (1997).
26. WHO | Cumulative number of confirmed human cases of avian influenza A(H5N1) reported to WHO. (2020).
27. Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–1897 (2013).
28. FAO H7N9 situation update - Avian Influenza A(H7N9) virus - FAO Emergency Prevention System for Animal Health (EMPRES-AH). http://www.fao.org/ag/againfo/programmes/en/empres/h7n9/situation_update.html.
29. Corman, V. M., Muth, D., Niemeyer, D. & Drosten, C. Hosts and Sources of Endemic Human Coronaviruses. *Adv. Virus Res.* **100**, 163–188 (2018).
30. Viruses, C. S. G. of T. I. C. on T. of & Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* vol. 5 536–544 (2020).

31. Gaunt, E. R., Hardie, A., Claas, E. C. J., Simmonds, P. & Templeton, K. E. Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J. Clin. Microbiol.* **48**, 2940–2947 (2010).
32. Aldridge, R. *et al.* Dataset: Seasonality and immunity to laboratory-confirmed seasonal coronaviruses (HCoV-NL63, HCoV-OC43, and HCoV-229E): results from the Flu Watch cohort study. (2020) doi:10.14324/000.ds.10093909.
33. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
34. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/.
35. Transmission of SARS-CoV-2: implications for infection prevention precautions. https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.
36. Greenhalgh, T. *et al.* Ten scientific reasons in support of airborne transmission of SARS-CoV-2. *Lancet* (2021) doi:10.1016/S0140-6736(21)00869-2.
37. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**, 507–513 (2020).
38. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
39. Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020).

40. Guan, W.-J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
41. Vaira, L. A., Salzano, G., Deiana, G. & De Riu, G. Anosmia and ageusia: Common findings in COVID-19 patients. *The Laryngoscope* vol. 130 1787 (2020).
42. Giacomelli, A. *et al.* Self-reported Olfactory and Taste Disorders in Patients With Severe Acute Respiratory Coronavirus 2 Infection: A Cross-sectional Study. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* vol. 71 889–890 (2020).
43. CDC. COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/ (2020).
44. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239–1242 (2020).
45. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
46. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2021).
47. Wölfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020).
48. Ziegler, C. G. K. *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* **181**, 1016–1035.e19 (2020).
49. Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives

Development of COVID-19. *Cell* **181**, 1036–1045.e9 (2020).

50. Shang, J. *et al.* Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 11727–11734 (2020).
51. Belouzard, S., Millet, J. K., Licitra, B. N. & Whittaker, G. R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **4**, 1011–1033 (2012).
52. Heald-Sargent, T. & Gallagher, T. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* **4**, 557–580 (2012).
53. Tang, T. *et al.* Proteolytic Activation of SARS-CoV-2 Spike at the S1/S2 Boundary: Potential Role of Proteases beyond Furin. *ACS Infect Dis* **7**, 264–272 (2021).
54. Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci. Transl. Med.* **12**, (2020).
55. Moreno, G. K. *et al.* Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread. *Nat. Commun.* **11**, 5558 (2020).
56. Valesano, A. L. *et al.* Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. *bioRxiv* (2021) doi:10.1101/2021.01.19.427330.
57. Braun, K. M. *et al.* Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. *bioRxiv* (2020) doi:10.1101/2020.11.16.384917.
58. Wang, H., Pipes, L. & Nielsen, R. Synonymous mutations and the molecular evolution of SARS-CoV-2 origins. *Virus Evol* **7**, veaa098 (2021).
59. Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L. & Denison, M. R. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J.*

Virol. **81**, 12135–12144 (2007).

60. Rota, P. A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399 (2003).
61. Ma, Y. *et al.* Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9436–9441 (2015).
62. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**, 17–37 (2006).
63. Xu, X., Cox, N. J., Bender, C. A., Regnery, H. L. & Shaw, M. W. Genetic variation in neuraminidase genes of influenza A (H3N2) viruses. *Virology* **224**, 175–183 (1996).
64. Shu, L. L., Bean, W. J. & Webster, R. G. Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *J. Virol.* **67**, 2723–2729 (1993).
65. Nobusawa, E. & Sato, K. Comparison of the mutation rates of human influenza A and B viruses. *J. Virol.* **80**, 3675–3678 (2006).
66. Callaway, E. The coronavirus is mutating - does it matter? *Nature* vol. 585 174–177 (2020).
67. Koyama, T., Platt, D. & Parida, L. Variant analysis of SARS-CoV-2 genomes. *Bull. World Health Organ.* **98**, 495–504 (2020).
68. Samuel, C. E. Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology* **411**, 180–193 (2011).
69. Boni, M. F., Zhou, Y., Taubenberger, J. K. & Holmes, E. C. Homologous recombination is very rare or absent in human influenza A virus. *J. Virol.* **82**, 4807–4811 (2008).

70. Lai, M. M. C. & Cavanagh, D. The Molecular Biology of Coronaviruses. in *Advances in Virus Research* (eds. Maramorosch, K., Murphy, F. A. & Shatkin, A. J.) vol. 48 1–100 (Academic Press, 1997).
71. Ganyani, T. *et al.* Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill.* **25**, (2020).
72. Baccam, P., Beauchemin, C., Macken, C. A., Hayden, F. G. & Perelson, A. S. Kinetics of influenza A virus infection in humans. *J. Virol.* **80**, 7590–7599 (2006).
73. Rouzine, I. M., Rodrigo, A. & Coffin, J. M. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol. Mol. Biol. Rev.* **65**, 151–185 (2001).
74. Lauring, A. S. Within-Host Viral Diversity: A Window into Viral Evolution. *Annu Rev Virol* **7**, 63–81 (2020).
75. Ghedin, E. *et al.* Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* **437**, 1162–1166 (2005).
76. Rambaut, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615–619 (2008).
77. Isabel, S. *et al.* Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Sci. Rep.* **10**, 14031 (2020).
78. França, M., Stallknecht, D. E. & Howerth, E. W. Expression and distribution of sialic acid influenza virus receptors in wild birds. *Avian Pathol.* **42**, 60–71 (2013).
79. Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A. & Klenk, H.-D. Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4620–4624 (2004).

80. Thompson, C. I., Barclay, W. S., Zambon, M. C. & Pickles, R. J. Infection of human airway epithelium by human and avian strains of influenza a virus. *J. Virol.* **80**, 8060–8068 (2006).
81. Stevens, J. *et al.* Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* **355**, 1143–1155 (2006).
82. de Vries, R. P. *et al.* Three mutations switch H7N9 influenza to human-type receptor specificity. *PLoS Pathog.* **13**, e1006390 (2017).
83. Tumpey, T. M. *et al.* A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* **315**, 655–659 (2007).
84. Huang, K.-Y. A. *et al.* Structure-function analysis of neutralizing antibodies to H7N9 influenza from naturally infected humans. *Nat Microbiol* **4**, 306–315 (2019).
85. Yan, H. *et al.* ACE2 receptor usage reveals variation in susceptibility to SARS-CoV and SARS-CoV-2 infection among bat species. *Nat Ecol Evol* **5**, 600–608 (2021).
86. Damas, J. *et al.* Broad Host Range of SARS-CoV-2 Predicted by Comparative and Structural Analysis of ACE2 in Vertebrates. doi:10.1101/2020.04.16.045302.
87. Zhao, X. *et al.* Broad and Differential Animal Angiotensin-Converting Enzyme 2 Receptor Usage by SARS-CoV-2. *J. Virol.* **94**, (2020).
88. Sasaki, T. *et al.* Long lasting immunity in chickens induced by a single shot of influenza vaccine prepared from inactivated non-pathogenic H5N1 virus particles against challenge with a highly pathogenic avian influenza virus. *Vaccine* vol. 27 5174–5177 (2009).
89. Massin, P., van der Werf, S. & Naffakh, N. Residue 627 of PB2 is a determinant of

cold sensitivity in RNA replication of avian influenza viruses. *J. Virol.* **75**, 5398–5404 (2001).

90. Mehle, A. & Doudna, J. A. An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. *Cell Host Microbe* **4**, 111–122 (2008).
91. Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **67**, 1761–1764 (1993).
92. Moncorgé, O., Mura, M. & Barclay, W. S. Evidence for avian and human host cell factors that affect the activity of influenza virus polymerase. *J. Virol.* **84**, 9978–9986 (2010).
93. Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* **293**, 1840–1842 (2001).
94. Munster, V. J. *et al.* The molecular basis of the pathogenicity of the Dutch highly pathogenic human influenza A H7N7 viruses. *J. Infect. Dis.* **196**, 258–265 (2007).
95. Salomon, R. *et al.* The polymerase complex genes contribute to the high virulence of the human H5N1 influenza virus isolate A/Vietnam/1203/04. *J. Exp. Med.* **203**, 689–697 (2006).
96. Zhang, H. *et al.* The PB2 E627K mutation contributes to the high polymerase activity and enhanced replication of H7N9 influenza virus. *J. Gen. Virol.* **95**, 779–786 (2014).
97. Smith, W., Andrewes, C. H. & Laidlaw, P. P. A virus obtained from influenza patients. *Lancet* **222**, 66–68 (1933).
98. Doud, M. B. & Bloom, J. D. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. *Viruses* **8**, (2016).

99. Krammer, F. The human antibody response to influenza A virus infection and vaccination. *Nat. Rev. Immunol.* **19**, 383–397 (2019).

100. Koel, B. F. *et al.* Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* **342**, 976–979 (2013).

101. Wang, W. *et al.* A mutation in the receptor binding site enhances infectivity of 2009 H1N1 influenza hemagglutinin pseudotypes without changing antigenicity. *Virology* **407**, 374–380 (2010).

102. Kim, T. S., Sun, J. & Braciale, T. J. T cell responses during influenza infection: getting and keeping control. *Trends Immunol.* **32**, 225–231 (2011).

103. Montaner, A. E., Escribano Montaner, A. & de Juanes Pardo, J. R. Infection by Influenza Virus in Childhood: A Call for Broader Influenza Vaccination. *Archivos de Bronconeumología ((English Edition))* vol. 40 231–235 (2004).

104. Webster, R. G. Original antigenic sin in ferrets: the response to sequential infections with influenza viruses. *J. Immunol.* **97**, 177–183 (1966).

105. Butler, D. The ghost of influenza past and the hunt for a universal vaccine. *Nature* vol. 560 158–160 (2018).

106. Gostic, K. M. *et al.* Childhood immune imprinting to influenza A shapes birth year-specific risk during seasonal H1N1 and H3N2 epidemics. *PLoS Pathog.* **15**, e1008109 (2019).

107. Shah, V. K., Firmal, P., Alam, A., Ganguly, D. & Chattopadhyay, S. Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. *Front. Immunol.* **11**, 1949 (2020).

108. Meyer, B., Drosten, C. & Müller, M. A. Serological assays for emerging

coronaviruses: challenges and pitfalls. *Virus Res.* **194**, 175–183 (2014).

109. Harvey, R. A. *et al.* Real-world data suggest antibody positivity to SARS-CoV-2 is associated with a decreased risk of future infection. doi:10.1101/2020.12.18.20248336.

110. Daniel, W., Nivet, M., Warner, J. & Podolsky, D. K. Early Evidence of the Effect of SARS-CoV-2 Vaccine at One Medical Center. *N. Engl. J. Med.* (2021) doi:10.1056/NEJMc2102153.

111. Hu, J. *et al.* Emerging SARS-CoV-2 variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. *Cell. Mol. Immunol.* **18**, 1061–1063 (2021).

112. CDC. SARS-CoV-2 Variant Classifications and Definitions. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html (2021).

113. Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. doi:10.1101/2020.12.24.20248822.

114. Doud, M. B., Hensley, S. E. & Bloom, J. D. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* **13**, e1006271 (2017).

115. Archetti, I. & Horsfall, F. L., Jr. [Persistent antigenic variations of influenza A virus after incomplete neutralization in ovo with heterologous immune serum]. *Rend. Ist. Sup. Sanit.* **14**, 909–936 (1951).

116. Foll, M. *et al.* Influenza virus drug resistance: a time-sampled population genetics perspective. *PLoS Genet.* **10**, e1004185 (2014).

117. Bedford, T. *et al.* Integrating influenza antigenic dynamics with molecular

evolution. *Elife* **3**, e01914 (2014).

118. Bedford, T. *et al.* Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523**, 217–220 (2015).

119. Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **3**, 1470–1476 (2007).

120. 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).

121. Debbink, K. *et al.* Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLoS Pathog.* **13**, e1006194 (2017).

122. McCrone, J. T. *et al.* Stochastic processes constrain the within and between host evolution of influenza virus. *Elife* **7**, (2018).

123. Sobel Leonard, A. *et al.* Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intrahost Genetic Diversification. *J. Virol.* **90**, 11247–11258 (2016).

124. Rogers, M. B. *et al.* Intrahost dynamics of antiviral resistance in influenza A virus reflect complex patterns of segment linkage, reassortment, and natural selection. *MBio* **6**, (2015).

125. Ghedin, E. *et al.* Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *J. Infect. Dis.* **203**, 168–174 (2011).

126. Gubareva, L. V., Kaiser, L., Matrosovich, M. N., Soo-Hoo, Y. & Hayden, F. G. Selection of influenza virus mutants in experimentally infected volunteers treated

with oseltamivir. *J. Infect. Dis.* **183**, 523–531 (2001).

127. Illingworth, C. J. R. & Mustonen, V. Components of selection in the evolution of the influenza virus: linkage effects beat inherent selection. *PLoS Pathog.* **8**, e1003091 (2012).

128. Hamilton, M. B. *Population Genetics*. (John Wiley & Sons, 2021).

129. Nelson, M. I. *et al.* Stochastic processes are key determinants of short-term evolution in influenza a virus. *PLoS Pathog.* **2**, e125 (2006).

130. Voronin, Y., Holte, S., Overbaugh, J. & Emerman, M. Genetic drift of HIV populations in culture. *PLoS Genet.* **5**, e1000431 (2009).

131. Brown, R. J. P. *et al.* Hepatitis C virus envelope glycoprotein fitness defines virus population composition following transmission to a new host. *J. Virol.* **86**, 11956–11966 (2012).

132. Gutiérrez, S., Michalakis, Y. & Blanc, S. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr. Opin. Virol.* **2**, 546–555 (2012).

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

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

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

136. Xue, K. S. & Bloom, J. D. Reconciling disparate estimates of viral genetic diversity during human influenza infections. *Nature genetics* vol. 51 1298–1301

(2019).

137. Leonard, A. S., Weissman, D. B., Greenbaum, B., Ghedin, E. & Koelle, K. Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. *J. Virol.* **91**, (2017).

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

139. McCrone, J. T. & Lauring, A. S. Genetic bottlenecks in intraspecies virus transmission. *Curr. Opin. Virol.* **28**, 20–25 (2018).

140. 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).

141. COVID-19 Genomic Epidemiology Toolkit. https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html (2021).

142. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228–232 (2016).

143. Faria, N. R. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. *Science* **352**, 345–349 (2016).

144. Faria, N. R. *et al.* Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* **361**, 894–899 (2018).

145. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

146. Seemann, T. *et al.* Tracking the COVID-19 pandemic in Australia using genomics. *Nat. Commun.* **11**, 4376 (2020).

147. Bedford, T. *et al.* Cryptic transmission of SARS-CoV-2 in Washington State.

medRxiv (2020) doi:10.1101/2020.04.02.20051417.

148. Deng, X. *et al.* Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* **369**, 582–587 (2020).

149. Miller, D. *et al.* Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *bioRxiv* (2020) doi:10.1101/2020.05.21.20104521.

150. Volz, E. M. & Siveroni, I. Bayesian phylodynamic inference with complex models. *PLoS Comput. Biol.* **14**, e1006546 (2018).

151. Grubaugh, N. D. *et al.* Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* **546**, 401–405 (2017).

152. Thézé, J. *et al.* Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. *Cell Host Microbe* **23**, 855–864.e7 (2018).

153. Duchene, S. *et al.* Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol* **6**, veaa061 (2020).

154. Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat. Rev. Genet.* **19**, 9–20 (2018).

155. Pybus, O. G. The Epidemic Behavior of the Hepatitis C Virus. *Science* vol. 292 2323–2325 (2001).

156. Baele, G., Suchard, M. A., Rambaut, A. & Lemey, P. Emerging Concepts of Data Integration in Pathogen Phylodynamics. *Syst. Biol.* **66**, e47–e65 (2017).

157. Lipsitch, M. *et al.* Viral factors in influenza pandemic risk assessment. *Elife* **5**, (2016).

158. Russell, C. A. *et al.* The potential for respiratory droplet-transmissible A/H5N1

influenza virus to evolve in a mammalian host. *Science* **336**, 1541–1547 (2012).

159. Neumann, G. & Kawaoka, Y. Predicting the Next Influenza Pandemics. *J. Infect. Dis.* **219**, S14–S20 (2019).

160. Neher, R. A. & Bedford, T. nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* **31**, 3546–3548 (2015).

161. Du, X., King, A. A., Woods, R. J. & Pascual, M. Evolution-informed forecasting of seasonal influenza A (H3N2). *Sci. Transl. Med.* **9**, (2017).

162. Morris, D. H. *et al.* Predictive Modeling of Influenza Shows the Promise of Applied Evolutionary Biology. *Trends Microbiol.* **26**, 102–118 (2018).

163. Zhang, F. *et al.* Human infections with recently-emerging highly pathogenic H7N9 avian influenza virus in China. *The Journal of infection* vol. 75 71–75 (2017).

164. Shen, Y. & Lu, H. Global concern regarding the fifth epidemic of human infection with avian influenza A (H7N9) virus in China. *Biosci. Trends* **11**, 120–121 (2017).

165. Wang, X. *et al.* Epidemiology of avian influenza A H7N9 virus in human beings across five epidemics in mainland China, 2013-17: an epidemiological study of laboratory-confirmed case series. *Lancet Infect. Dis.* **17**, 822–832 (2017).

166. Zhou, L. *et al.* Sudden increase in human infection with avian influenza A(H7N9) virus in China, September-December 2016. *Western Pac Surveill Response J* **8**, 6–14 (2017).

167. Ke, C. *et al.* Human Infection with Highly Pathogenic Avian Influenza A(H7N9) Virus, China. *Emerg. Infect. Dis.* **23**, 1332–1340 (2017).

168. Sutton, T. C. The Pandemic Threat of Emerging H5 and H7 Avian Influenza Viruses. *Viruses* **10**, (2018).

169. Taft, A. S. *et al.* Identification of mammalian-adapting mutations in the polymerase complex of an avian H5N1 influenza virus. *Nat. Commun.* **6**, 7491 (2015).

170. Imai, M. *et al.* A Highly Pathogenic Avian H7N9 Influenza Virus Isolated from A Human Is Lethal in Some Ferrets Infected via Respiratory Droplets. *Cell Host Microbe* **22**, 615–626.e8 (2017).

171. Qi, W. *et al.* Emergence and Adaptation of a Novel Highly Pathogenic H7N9 Influenza Virus in Birds and Humans from a 2013 Human-Infecting Low-Pathogenic Ancestor. *J. Virol.* **92**, (2018).

172. Yu, D. *et al.* The re-emergence of highly pathogenic avian influenza H7N9 viruses in humans in mainland China, 2019. *Euro Surveill.* **24**, (2019).

173. King, J. *et al.* Novel HPAIV H5N8 Reassortant (Clade 2.3.4.4b) Detected in Germany. *Viruses* **12**, (2020).

174. Watanabe, T. *et al.* Characterization of H7N9 influenza A viruses isolated from humans. *Nature* **501**, 551–555 (2013).

175. Zhang, Q. *et al.* H7N9 influenza viruses are transmissible in ferrets by respiratory droplet. *Science* **341**, 410–414 (2013).

176. Zhou, L. *et al.* Clusters of Human Infection and Human-to-Human Transmission of Avian Influenza A(H7N9) Virus, 2013-2017. *Emerg. Infect. Dis.* **24**, (2018).

177. Richard, M. *et al.* Limited airborne transmission of H7N9 influenza A virus between ferrets. *Nature* **501**, 560–563 (2013).

178. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420–

428 (2012).

179. Galloway, S. E., Reed, M. L., Russell, C. J. & Steinhauer, D. A. Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. *PLoS Pathog.* **9**, e1003151 (2013).

180. Linster, M. *et al.* Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. *Cell* **157**, 329–339 (2014).

181. Imai, M. *et al.* Influenza A variants with reduced susceptibility to baloxavir isolated from Japanese patients are fit and transmit through respiratory droplets. *Nat Microbiol* **5**, 27–33 (2020).

182. Liu, S. *et al.* Control of avian influenza in China: Strategies and lessons. *Transbound. Emerg. Dis.* **67**, 1463–1471 (2020).

183. Kandeil, A. *et al.* Efficacy of commercial vaccines against newly emerging avian influenza H5N8 virus in Egypt. *Sci. Rep.* **8**, 9697 (2018).

184. Briggs, H. Coronavirus: WHO developing guidance on wet markets. *BBC* (2020).

185. Reaser, J. K., Witt, A., Tabor, G. M., Hudson, P. J. & Plowright, R. K. Ecological Countermeasures for Preventing Zoonotic Disease Outbreaks: When Ecological Restoration is a Human Health Imperative. *Restor. Ecol.* e13357 (2021).

186. Zhu, W. *et al.* Biological characterisation of the emerged highly pathogenic avian influenza (HPAI) A(H7N9) viruses in humans, in mainland China, 2016 to 2017. *Euro Surveill.* **22**, (2017).

187. Itoh, Y. *et al.* In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. *Nature* **460**, 1021–1025 (2009).

188. Neumann, G. & Kawaoka, Y. Transmission of influenza A viruses. *Virology* **479-480**, 234–246 (2015).

189. McCrone, J. T. & Lauring, A. S. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. *J. Virol.* **90**, 6884–6895 (2016).

190. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).

191. Teiling, C. BaseSpace: Simplifying metagenomic analysis. doi:10.26226/morressier.56d5ba2ed462b80296c9509d.

192. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

193. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

194. Li, H. seqtk Toolkit for processing sequences in FASTA/Q formats. *GitHub* **767**, 69 (2012).

195. Koboldt, D. C. *et al.* VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283–2285 (2009).

196. Robasky, K., Lewis, N. E. & Church, G. M. The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Genet.* **15**, 56–62 (2014).

197. 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).

198. *SNPGenie*. (Github).

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

200. Hughes, A. L. & of Biological Sciences Austin L Hughes. *Adaptive Evolution of Genes and Genomes*. (Oxford University Press, 1999).

201. Braun, K. *katarinabraun - Overview*. (Github).

202. Enkirch, T. & von Messling, V. Ferret models of viral pathogenesis. *Virology* **479-480**, 259–270 (2015).

203. Zhao, L. & Illingworth, C. J. R. Measurements of intrahost viral diversity require an unbiased diversity metric. *Virus Evol* **5**, vey041 (2019).

204. 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).

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

206. Furuse, Y., Suzuki, A., Kamigaki, T. & Oshitani, H. Evolution of the M gene of the influenza A virus in different host species: large-scale sequence analysis. *Virol. J.* **6**, 67 (2009).

207. Belser, J. A. *et al.* A Guide for the Use of the Ferret Model for Influenza Virus Infection. *Am. J. Pathol.* **190**, 11–24 (2020).

208. Watanabe, T. *et al.* Circulating avian influenza viruses closely related to the 1918

virus have pandemic potential. *Cell Host Microbe* **15**, 692–705 (2014).

209. Muller, H. J. THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. *Mutat. Res.* **106**, 2–9 (1964).

210. Sanjuán, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8396–8401 (2004).

211. Visher, E., Whitefield, S. E., McCrone, J. T., Fitzsimmons, W. & Lauring, A. S. The Mutational Robustness of Influenza A Virus. *PLoS Pathog.* **12**, e1005856 (2016).

212. Lyons, D. M. & Lauring, A. S. Mutation and Epistasis in Influenza Virus Evolution. *Viruses* **10**, (2018).

213. Short, K. R. *et al.* One health, multiple challenges: The inter-species transmission of influenza A virus. *One Health* vol. 1 1–13 (2015).

214. Buhnerkempe, M. G. *et al.* Mapping influenza transmission in the ferret model to transmission in humans. *Elife* **4**, (2015).

215. Shi, J. *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016–1020 (2020).

216. Rockx, B. *et al.* Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. *Science* **368**, 1012–1015 (2020).

217. 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).

218. Shan, C. *et al.* Infection with Novel Coronavirus (SARS-CoV-2) Causes

Pneumonia in the Rhesus Macaques. doi:10.21203/rs.2.25200/v1.

219. Oreshkova, N. *et al.* SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro Surveill.* **25**, (2020).

220. Lakdawala, S. S. & Menachery, V. D. The search for a COVID-19 animal model. *Science* vol. 368 942–943 (2020).

221. Richard, M. *et al.* SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nat. Commun.* **11**, 3496 (2020).

222. Oude Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**, 172–177 (2021).

223. 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).

224. 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).

225. 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).

226. 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).

227. Wang, D. *et al.* Population Bottlenecks and Intra-host Evolution during Human-to-Human Transmission of SARS-CoV-2. doi:10.1101/2020.06.26.173203.

228. Lythgoe, K. A. *et al.* SARS-CoV-2 within-host diversity and transmission. *Science* **372**, (2021).

229. Graudenzi, A., Maspero, D., Angaroni, F., Piazza, R. & Ramazzotti, D. Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. *iScience* **24**, 102116 (2021).

230. Sapoval, N. *et al.* Hidden genomic diversity of SARS-CoV-2: implications for qRT-PCR diagnostics and transmission. *bioRxiv* (2020) doi:10.1101/2020.07.02.184481.

231. Karamitros, T. *et al.* SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. doi:10.1101/2020.03.27.009480.

232. Shen, Z. *et al.* Genomic diversity of SARS-CoV-2 in coronavirus disease, patients. *Clin. Infect. Dis.* (2019).

233. Ramazzotti, D. *et al.* VERSO: A comprehensive framework for the inference of robust phylogenies and the quantification of intra-host genomic diversity of viral samples. *Patterns (N Y)* **2**, 100212 (2021).

234. Halfmann, P. J. *et al.* Transmission of SARS-CoV-2 in Domestic Cats. *N. Engl. J. Med.* **383**, 592–594 (2020).

235. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).

236. Quick, J. nCoV-2019 sequencing protocol v2 v1 (protocols.io.bdp7i5rn). *protocols.io* doi:10.17504/protocols.io.bdp7i5rn.

237. Gu, H. *et al.* Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine

efficacy. *Science* **369**, 1603–1607 (2020).

238. Braun, K. *SARSCoV2_transmission_in_domestic_cats*. (Github).

239. Moncla, L. H. *et al.* Quantifying within-host diversity of H5N1 influenza viruses in humans and poultry in Cambodia. *PLoS Pathog.* **16**, e1008191 (2020).

240. Hodcroft, E. B. *et al.* Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* (2020) doi:10.1101/2020.10.25.20219063.

241. Molecular Population Genetics. *Genetics* **213**, 721–722 (2019).

242. Rausch, J. W., Capoferri, A. A., Katusiime, M. G., Patro, S. C. & Kearney, M. F. Low genetic diversity may be an Achilles heel of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America* vol. 117 24614–24616 (2020).

243. Guallar, M. P. *et al.* Inoculum at the time of SARS-CoV-2 exposure and risk of disease severity. *Int. J. Infect. Dis.* **97**, 290–292 (2020).

244. Gandhi, M., Beyrer, C. & Goosby, E. Masks Do More Than Protect Others During COVID-19: Reducing the Inoculum of SARS-CoV-2 to Protect the Wearer. *J. Gen. Intern. Med.* **35**, 3063–3066 (2020).

245. Tanaka, Y. & Mizokami, M. Genetic diversity of hepatitis B virus as an important factor associated with differences in clinical outcomes. *The Journal of infectious diseases* vol. 195 1–4 (2007).

246. Tahamtan, A., Askari, F. S., Bont, L. & Salimi, V. Disease severity in respiratory syncytial virus infection: Role of host genetic variation. *Rev. Med. Virol.* **29**, e2026 (2019).

247. Xiao, Y. *et al.* Poliovirus intrahost evolution is required to overcome tissue-

specific innate immune responses. *Nature Communications* vol. 8 (2017).

248. Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344–348 (2006).

249. Memoli, M. J. *et al.* In vivo evaluation of pathogenicity and transmissibility of influenza A(H1N1)pdm09 hemagglutinin receptor binding domain 222 intrahost variants isolated from a single immunocompromised patient. *Virology* vol. 428 21–29 (2012).

250. Biswas, S. K. & Mudi, S. R. Genetic variation in SARS-CoV-2 may explain variable severity of COVID-19. *Med. Hypotheses* **143**, 109877 (2020).

251. Zhang, X. *et al.* Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583**, 437–440 (2020).

252. Baum, A. *et al.* Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **369**, 1014–1018 (2020).

253. Dieterle, M. E. *et al.* A Replication-Competent Vesicular Stomatitis Virus for Studies of SARS-CoV-2 Spike-Mediated Cell Entry and Its Inhibition. *Cell Host Microbe* **28**, 486–496.e6 (2020).

254. Chan, J. F.-W. *et al.* Simulation of the Clinical and Pathological Manifestations of Coronavirus Disease 2019 (COVID-19) in a Golden Syrian Hamster Model: Implications for Disease Pathogenesis and Transmissibility. *Clinical Infectious Diseases* (2020) doi:10.1093/cid/ciaa325.

255. Yang, X., Dong, N., Chan, E. W.-C. & Chen, S. Identification of super-transmitters of SARS-CoV-2. doi:10.1101/2020.04.19.20071399.

256. 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).

257. Koel, B. F. *et al.* Disparate evolution of virus populations in upper and lower airways of mechanically ventilated patients. *bioRxiv* 509901 (2019) doi:10.1101/509901.

258. Yan, J. *et al.* Infectious virus in exhaled breath of symptomatic seasonal influenza cases from a college community. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1081–1086 (2018).

259. SARS-CoV, W. H. O. mink-associated variant strain—Denmark. 2020. (2AD).

260. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).

261. auspice. https://nextstrain.org/groups/neherlab/ncov/S.H655Y?c=gt-S_655&p=grid&r=country.

262. Hanada, K., Gojobori, T. & Li, W.-H. Radical amino acid change versus positive selection in the evolution of viral envelope proteins. *Gene* **385**, 83–88 (2006).

263. Public Health England. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01. https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201 (2020).

264. arambaut, garmstrong & isabel. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations.

https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).

265. Kemp, S. A. *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**, 277–282 (2021).

266. Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* **383**, 2291–2293 (2020).

267. Baang, J. H. *et al.* Prolonged Severe Acute Respiratory Syndrome Coronavirus 2 Replication in an Immunocompromised Patient. *The Journal of Infectious Diseases* vol. 223 23–27 (2021).

268. Truong, T. T. *et al.* Persistent SARS-CoV-2 infection and increasing viral variants in children and young adults with impaired humoral immunity. *medRxiv* (2021) doi:10.1101/2021.02.27.21252099.

269. Choudhary, M. C., Crain, C. R., Qiu, X., Hanage, W. & Li, J. Z. SARS-CoV-2 sequence characteristics of COVID-19 persistence and reinfection. *bioRxiv* (2021) doi:10.1101/2021.03.02.21252750.

270. Xue, K. S. *et al.* Parallel evolution of influenza across multiple spatiotemporal scales. *Elife* **6**, (2017).

271. van Beek, J. *et al.* Chronic norovirus infection among solid organ recipients in a tertiary care hospital, the Netherlands, 2006-2014. *Clin. Microbiol. Infect.* **23**, 265.e9–265.e13 (2017).

272. Avanzato, V. A. *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901–1912.e9 (2020).

273. Debbink, K. *et al.* Within-Host Evolution Results in Antigenically Distinct GII.4 Noroviruses. *Journal of Virology* vol. 88 7244–7255 (2014).

274. Rhee, C., Kanjilal, S. & Baker, M. Duration of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infectivity: when is it safe to discontinue isolation? *Clin. Infect. Dis.* (2020).

275. Morris, D. H. *et al.* Asynchrony between virus diversity and antibody selection limits influenza virus evolution. *Elife* **9**, (2020).

276. Bullard, J. *et al.* Predicting Infectious Severe Acute Respiratory Syndrome Coronavirus 2 From Diagnostic Samples. *Clin. Infect. Dis.* **71**, 2663–2666 (2020).

277. Tonkin-Hill, G. *et al.* Patterns of within-host genetic diversity in SARS-CoV-2. *bioRxiv* 2020.12.23.424229 (2020) doi:10.1101/2020.12.23.424229.

278. James, S. E. *et al.* High Resolution analysis of Transmission Dynamics of Sars-Cov-2 in Two Major Hospital Outbreaks in South Africa Leveraging Intrahost Diversity. *medRxiv* (2020) doi:10.1101/2020.11.15.20231993.

279. Martin, M. A. & Koelle, K. Reanalysis of deep-sequencing data from Austria points towards a small SARS-COV-2 transmission bottleneck on the order of one to three virions. *bioRxiv* (2021).

280. CDC. CDC Diagnostic Tests for COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/lab/testing.html (2020).

281. Panther Fusion® SARS-CoV-2 Assay. https://www.hologic.com/package-inserts/diagnostic-products/panther-fusionr-sars-cov-2-assay.

282. Hologic. https://www.hologic.com/package-inserts/diagnostic-products/aptimar-sars-cov-2-assay-pantherr-system.

283. Quick, J. nCoV-2019 sequencing protocol. (2020) doi:10.17504/protocols.io.bbmuik6w.

284. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* **4**, vex042 (2018).

285. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).

286. Richard, D., Owen, C. J., van Dorp, L. & Balloux, F. No detectable signal for ongoing genetic recombination in SARS-CoV-2. *bioRxiv* (2020).

287. Dudas, G. *baltic*. (Github).

288. Moncla, L. H. *et al.* Repeated introductions and intensive community transmission fueled a mumps virus outbreak in Washington State. *bioRxiv* (2020) doi:10.1101/2020.10.19.20215442.

289. Kinganda-Lusamaki, E. *et al.* Operationalizing genomic epidemiology during the Nord-Kivu Ebola outbreak, Democratic Republic of the Congo. *bioRxiv* (2020) doi:10.1101/2020.06.08.20125567.

290. Shen, Z. *et al.* Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* vol. 71 713–720 (2020).

291. De Maio, Conor Walker, Rui Borges, Lukas Weilguny, Greg Slodkowicz, Nick Goldman, Nicola. Issues with SARS-CoV-2 sequencing data. https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473 (2020).

292. Lythgoe, K. A. *et al.* Within-host genomics of SARS-CoV-2. *Cold Spring Harbor*

Laboratory 2020.05.28.118992 (2020) doi:10.1101/2020.05.28.118992.

293. CDC. Science brief: Emerging SARS-CoV-2 variants. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html (2021).

294. US Coronavirus cases & deaths by state. https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/ (2021).

295. Fauver, J. R. *et al.* Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell* **181**, 990–996.e5 (2020).

296. Lu, J. *et al.* Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* **181**, 997–1003.e9 (2020).

297. Oude Munnink, B. B. *et al.* Rapid SARS-CoV-2 whole genome sequencing for informed public health decision making in the Netherlands. *Cold Spring Harbor Laboratory* 2020.04.21.050633 (2020) doi:10.1101/2020.04.21.050633.

298. Candido, D. S. *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. *bioRxiv* (2020) doi:10.1101/2020.06.11.20128249.

299. Pollett, S. *et al.* Genomic Epidemiology as a Public Health Tool to Combat Mosquito-Borne Virus Outbreaks. *J. Infect. Dis.* **221**, S308–S318 (2020).

300. Wisconsin Department of Transportation Traffic Count Map. https://wisdot.maps.arcgis.com/apps/webappviewer/index.html?id=2e12a4f051de4ea9bc865ec6393731f8.

301. Wisconsin Public Health Profiles - Dane County.

302. Wisconsin Public Health Profiles - Milwaukee County.

303. Evers, T. Order for Statewide School Closure. (2020).

304. Evers, T. Order Prohibiting Mass Gatherings of 50 People or More. (2020).

305. Evers, T. Emergency Order #5, Prohiniting Mass Gatherings of 10 People or More. (2020).

306. Executive Orders. https://evers.wi.gov/Pages/Newsroom/Executive-Orders.aspx.

307. Evers, T. Emergency Order #28 Safer at Home Order. (2020).

308. Kafetzopoulou, L. E. *et al.* Assessment of metagenomic Nanopore and Illumina sequencing for recovering whole genome sequences of chikungunya and dengue viruses directly from clinical samples. *Euro Surveill.* **23**, (2018).

309. Moreno, G. & David. Sequence-Independent, Single-Primer Amplification of RNA viruses v3 (protocols.io.bckxiuxn). *protocols.io* (2020) doi:10.17504/protocols.io.bckxiuxn.

310. Quick, J. nCoV-2019 sequencing protocol v1 (protocols.io.bbmuik6w). *protocols.io* (2020) doi:10.17504/protocols.io.bbmuik6w.

311. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

312. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

313. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

314. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

315. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* **9**, 90–95 (2007).

316. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

317. Geidelberg, L. *et al.* Genomic epidemiology of a densely sampled COVID-19 outbreak in China. *bioRxiv* (2020) doi:10.1101/2020.03.09.20033365.

318. Bi, Q. *et al.* Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020).

319. Endo, A., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott, S., Kucharski, A. J. & Funk, S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res* **5**, 67 (2020).

320. Unwin, H. *et al. Report 23: State-level tracking of COVID-19 in the United States*. http://spiral.imperial.ac.uk/handle/10044/1/79231 (2020) doi:10.25561/79231.

321. Moreno, G. *SARS-CoV-2-in-Southern-Wisconsin*. (Github).

322. ArcGIS Dashboards. https://cityofmadison.maps.arcgis.com/apps/opsdashboard/index.html#/e22f5ba4f1f94e0bb0b9529dc82db6a3.

323. County, M. Milwaukee County. *Hos Diphtheria Antitoxin Division* (2020).

324. Data and Statistics. https://www.dhs.wisconsin.gov/stats/index.htm (2014).

325. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33–46 (2017).

326. Lee, S., Meyler, P., Mozel, M., Tauh, T. & Merchant, R. Asymptomatic carriage and transmission of SARS-CoV-2: What do we know? *Can. J. Anaesth.* **67**, 1424–1430 (2020).

327. Furukawa, N. W., Brooks, J. T. & Sobel, J. Evidence Supporting Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 While Presymptomatic or Asymptomatic. *Emerg. Infect. Dis.* **26**, (2020).

328. Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and the US. *bioRxiv* (2020) doi:10.1101/2020.05.21.109322.

329. Chan, J. F.-W. *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **9**, 221–236 (2020).

330. Althouse, B. M. *et al.* Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. *arXiv [q-bio.PE]* (2020).

331. Frieden, T. R. & Lee, C. T. Identifying and Interrupting Superspreading Events-Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 1059–1066 (2020).

332. Quinn, S. C. & Kumar, S. Health inequalities and infectious disease epidemics: a challenge for global health security. *Biosecur. Bioterror.* **12**, 263–273 (2014).

333. Quinn, S. C. *et al.* Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am. J. Public Health* **101**, 285–293 (2011).

334. Kumar, S., Quinn, S. C., Kim, K. H., Daniel, L. H. & Freimuth, V. S. The impact of workplace policies and other social factors on self-reported influenza-like illness

incidence during the 2009 H1N1 pandemic. *Am. J. Public Health* **102**, 134–140 (2012).

335. Zipfel, C. M. & Bansal, S. Health inequities in influenza transmission and surveillance. *bioRxiv* (2020) doi:10.1101/2020.03.30.20048017.

336. Yuen, K.-S., Ye, Z.-W., Fung, S.-Y., Chan, C.-P. & Jin, D.-Y. SARS-CoV-2 and COVID-19: The most important research questions. *Cell Biosci.* **10**, 40 (2020).

337. Filipe, A. da S. *et al.* Genomic epidemiology of SARS-CoV-2 spread in Scotland highlights the role of European travel in COVID-19 emergence. *bioRxiv* (2020) doi:10.1101/2020.06.08.20124834.

338. Preliminary analysis of SARS-CoV-2 importation & establishment of UK transmission lineages. https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507 (2020).

339. Moreno, G. K. *et al.* Distinct patterns of SARS-CoV-2 transmission in two nearby communities in Wisconsin, USA. *medRxiv* (2020) doi:10.1101/2020.07.09.20149104.

340. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).

341. COVID-19: County Data. https://www.dhs.wisconsin.gov/covid-19/county.htm (2020).

342. Guan, W.-J. *et al.* Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur. Respir. J.* **55**, (2020).

343. de Lusignan, S. *et al.* Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary

care network: a cross-sectional study. *Lancet Infect. Dis.* **20**, 1034–1042 (2020).

344. Mays, V. M., Cochran, S. D. & Barnes, N. W. Race, race-based discrimination, and health outcomes among African Americans. *Annu. Rev. Psychol.* **58**, 201–225 (2007).

345. Millett, G. A. *et al.* Assessing differential impacts of COVID-19 on black communities. *Ann. Epidemiol.* **47**, 37–44 (2020).

346. CDC's Social Vulnerability Index. https://svi.cdc.gov/index.html.

347. Nayak, A. *et al.* Impact of Social Vulnerability on COVID-19 Incidence and Outcomes in the United States. *medRxiv* (2020) doi:10.1101/2020.04.10.20060962.

348. Daniloski, Z. *et al.* The D614G mutation in SARS-CoV-2 Spike increases transduction of multiple human cell types. *bioRxiv* (2020) doi:10.1101/2020.06.14.151357.

349. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812–827.e19 (2020).

350. Zhang, L. *et al.* The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* (2020) doi:10.1101/2020.06.12.148726.

351. Korber, B. *et al.* Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *Cold Spring Harbor Laboratory* 2020.04.29.069054 (2020) doi:10.1101/2020.04.29.069054.

352. COVID-19 Health Alert # 9: Required data collection for COVID-19 disease testing; New guidelines for COVID-19 testing at WI public health laboratories. https://content.govdelivery.com/accounts/WIDHS/bulletins/28bddb5.

353. Scotch, M. *et al.* Enhancing phylogeography by improving geographical information from GenBank. *J. Biomed. Inform.* **44 Suppl 1**, S44–7 (2011).

354. Nguyen, L. H. *et al.* Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Health* **5**, e475–e483 (2020).

355. Chou, R. *et al.* Epidemiology of and Risk Factors for Coronavirus Infection in Health Care Workers: A Living Rapid Review. *Ann. Intern. Med.* **173**, 120–136 (2020).

356. Lai, J. *et al.* Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. *JAMA Netw Open* **3**, e203976 (2020).

357. Cheng, V. C.-C., Wong, S.-C. & Yuen, K.-Y. Estimating Coronavirus Disease 2019 Infection Risk in Health Care Workers. *JAMA network open* vol. 3 e209687 (2020).

358. Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1271 (2020).

359. Arpacioglu, S., Gurler, M. & Cakiroglu, S. Secondary Traumatization Outcomes and Associated Factors Among the Health Care Workers Exposed to the COVID-19. *International Journal of Social Psychiatry* 002076402094074 (2020) doi:10.1177/0020764020940742.

360. for Disease Control, C., Prevention & Others. Interim infection prevention and control recommendations for healthcare personnel during the coronavirus disease

2019 (COVID-19) pandemic. *Accessed July* **22**, (2020).

361. Lepak, A. J., Shirley, D. K., Buys, A., Stevens, L. & Safdar, N. Implementation of infection control measures to prevent healthcare-associated transmission of severe acute respiratory coronavirus virus 2 (SARS-CoV-2). *Infect. Control Hosp. Epidemiol.* 1–4 (2020).

362. Houldcroft, C. J. *et al.* Use of Whole-Genome Sequencing of Adenovirus in Immunocompromised Pediatric Patients to Identify Nosocomial Transmission and Mixed-Genotype Infection. *The Journal of Infectious Diseases* vol. 218 1261–1271 (2018).

363. Greninger, A. L. *et al.* Rapid Metagenomic Next-Generation Sequencing during an Investigation of Hospital-Acquired Human Parainfluenza Virus 3 Infections. *J. Clin. Microbiol.* **55**, 177–182 (2017).

364. Houlihan, C. F. *et al.* Use of Whole-Genome Sequencing in the Investigation of a Nosocomial Influenza Virus Outbreak. *J. Infect. Dis.* **218**, 1485–1489 (2018).

365. Deurenberg, R. H. *et al.* Application of next generation sequencing in clinical microbiology and infection prevention. *J. Biotechnol.* **243**, 16–24 (2017).

366. Tang, P., Croxen, M. A., Hasan, M. R., Hsiao, W. W. L. & Hoang, L. M. Infection control in the new age of genomic epidemiology. *Am. J. Infect. Control* **45**, 170–179 (2017).

367. Safdar, N., Moreno, G. K., Braun, K. M., Friedrich, T. C. & O'Connor, D. H. Using Virus Sequencing to Determine Source of SARS-CoV-2 Transmission for Healthcare Worker. *Emerg. Infect. Dis.* **26**, 2489–2491 (2020).

368. Sikkens, J. J. *et al.* Serologic Surveillance and Phylogenetic Analysis of SARS-

CoV-2 Infection in Hospital Health Care Workers. *medRxiv* (2021).

369. Biek, R., Pybus, O. G., Lloyd-Smith, J. O. & Didelot, X. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.* **30**, 306–313 (2015).

370. Rai, B., Shukla, A. & Dwivedi, L. K. Estimates of serial interval for COVID-19: A systematic review and meta-analysis. *Clin Epidemiol Glob Health* **9**, 157–161 (2021).

371. auspice. https://nextstrain.org/ncov/north-america.

372. Ladner, J. T., Grubaugh, N. D., Pybus, O. G. & Andersen, K. G. Precision epidemiology for infectious disease control. *Nat. Med.* **25**, 206–211 (2019).

373. Braun, K. *SARSCoV2_sequencing_healthcare-association_infections*. (Github).

374. Sikkema, R. S. *et al.* COVID-19 in health-care workers in three hospitals in the south of the Netherlands: a cross-sectional study. *Lancet Infect. Dis.* **20**, 1273–1280 (2020).

375. Jacob, J. T. *et al.* Risk Factors Associated With SARS-CoV-2 Seropositivity Among US Health Care Personnel. *JAMA Netw Open* **4**, e211283 (2021).

376. Xue, K. S., Moncla, L. H., Bedford, T. & Bloom, J. D. Within-Host Evolution of Human Influenza Virus. *Trends Microbiol.* **26**, 781–793 (2018).

377. Baang, J. H. *et al.* Prolonged Severe Acute Respiratory Syndrome Coronavirus 2 Replication in an Immunocompromised Patient. *J. Infect. Dis.* **223**, 23–27 (2021).

378. Hou, Y. J. *et al.* SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).

379. Zhou, B. *et al.* SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* **592**, 122–127 (2021).

380. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).

381. Volz, E. *et al.* Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* **184**, 64–75.e11 (2021).

382. Daniloski, Z. *et al.* The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife* **10**, (2021).

383. CDC/ATSDR's Social Vulnerability Index (SVI). https://www.atsdr.cdc.gov/placeandhealth/svi/index.html (2021).