Supplementary Methods

Data and Sample Collection

Participation was open to all English-speaking university affiliates whose primary residence was on-campus or within 100 miles of the university campus(1–3). Perspective participants were informed about the study through electronic (sent via email and posted on social media) and physical (flyers, signs, and posters) advertisements. At enrollment, participants completed an electronic questionnaire and indicated if they preferred to receive study communications via text or email. They also underwent baseline SARS-COV-2 testing. After enrollment, participants were sent daily attestation surveys. Throughout the study, participants were invited to test if they reported new symptoms, had a known exposure to a SARS-COV-2 case, or were members of a group experiencing an outbreak. Walk-in (on-demand) testing was also available for any reason, including return to campus after a university break. In addition, participants were invited to test from September 2020 to August 2021 following attendance at gatherings with >10 people, from September 2021 to July 2022 following out of state travel, and from August – September 2022 following report of a positive rapid test. Though testing was available to participants both by invitation and on-demand, participants were not required to test at any point in the study period. University students, faculty, and staff who did not wish to participate in HCT had access to free SARS-CoV-2 testing outside of the study through Seattle-King County Public Health programs. The UW IRB approved this study (#00011148). All participants gave informed consent or assent and parent/guardian consent for participants under 18 years of age.

Testing was available to participants in 3 formats. Participant self-swab collection observed by study personnel was performed at staffed testing kiosks on campus. Alternatively, participants could pick-up self-testing kits from staffed dropbox locations on campus. Self-testing kits could be returned either to a dropbox location or via courier. Staffing times were variable across kiosks and dropboxes sites. The main kiosk/dropbox site was staffed 7 days a week during the study period, except for a small number of holidays and several days of inclement weather.

Two swab types were used for observed swab collection at kiosks; US Cotton #3 Steripack Polyester Spun Swabs placed in a 10 mL tube were used at the beginning of the study with transition to RHINOsticTM RH_S000001 Automated Nasal Swabs placed in a MatrixTM 1.0 mL Thermo Fisher 3741 ScrewTop tube in 2021. Swabs returned to drop boxes were RHINOsticTM while those returned via courier were US Cotton #3 swabs.

University Covid Mitigation Strategies

After the beginning of the SARS-CoV-2 pandemic in spring 2020 (and prior to the start of the study period), the university implemented transmission mitigation interventions which included mask mandates in public spaces on campus, enhanced cleaning and disinfection protocols for public spaces, improvements in air filtration systems, limitations on the size of in-person gatherings, and a communications campaign to encourage handwashing and social distancing. On-campus housing was open throughout the study period. The percent of students living on-campus during year 1 (online instruction only) was approximately 20% increasing to approximately 30% during year 2 (online and in-person instruction). In compliance with CDC

recommendations, quarantine housing was provided to students with SARS-COV-2 infection who lived in on-campus. At the beginning of the study period, negative testing was required for clearance from quarantine. This was later changed to automatic clearance 10 days after the first positive test. Students, faculty, and staff living off-campus were provided with information about good quarantine practices and follow-up testing. Contact tracing for SARS-COV-2 cases in university students, faculty, and staff were conducted by university and/or county public health officials. Vaccination was strongly recommended for university students, faculty, and staff beginning in the Fall of 2021 and then mandated starting January 1, 2022. 99% of students, staff, and faculty had received the primary vaccine series by January 2022.

SARS-CoV-2 testing of samples

Sample aliquots of 5µL were used in four multiplexed RT-qPCR reactions. Two reactions used custom probe sets that targeted Orf1b and two used probes targeted the S-gene. Viral probe sets were multiplexed with a probe set for human RNase P. Specimens were positive for SARS-CoV-2 if viral gene targets and human RNase P were detected in at least three of four reactions.

Sequencing of SARS-CoV-2 positive samples

Magna Pure 96 kits (Roche) were used to extract nucleic acids from specimens and sequencing libraries were prepared using COVIDSeq kits (Illumina). Sequencing primers were updated at several points during the study period to account for emergence of new viral variants. Sequencing was performed using NextSeq2000 P200 kits (Illumina).

Sequence Analysis

Both HCT and GISAID genomes were screened for quality using Nextclade CLI(4). Nextclade was also used to assign sequences to Nextstrain clades and Pango lineages. Sequences given an unfavorable quality rating (quality control score of $\geq$ 100) by Nextclade (based on a sequence's complement of missing data, mixed sites, private mutations, mutation clusters, frameshifts, and premature stop codons), with a missing collection date, or for which Nextclade was unable to make a clade and/or lineage assignment were excluded from further analysis. We also excluded sequences from further analysis if they were missing base calls for $\geq$ 10% of genomic positions ($\geq$ 10% Ns). More information about Nextclade's sequence quality rating system can be found at https://docs.nextstrain.org/projects/nextclade/en/latest/user/algorithm/06-quality-control.html.

Identification of possible non-HCT descendants of HCT viruses

Each group of identical sequences were characterized by a set of mutations relative to the reference genome (Wuhan/Hu-1/2019, GenBank Accession MN908947). Any genome that carried the same mutations as a group of identical genomes plus additional mutations relative to the reference was categorized as a descendant of that group. Phylogenetic groups were identified as follows: a phylogenetic tree was constructed for each Nextstrain clade, which included all HCT and GISAID sequences from Washington State (WA) belonging to that clade during the study period. All terminal nodes in these trees were designated as HCT or non-HCT.

Augur trait was used to assign HCT versus non-HCT states for all internal nodes and to provide a likelihood of each state assignment(5). All HCT terminal nodes that descended from the same internal node assigned a state of HCT with a likelihood of 95% or greater were grouped into a single phylogenetic cluster.

Transmission Modeling

Our transmission modeling analysis included SARS-CoV-2 sequences divided into 3 regions of origin: HCT, consisting of HCT sequences; KC, consisting of sequences from King County (KC), WA; and other, which consisted of contextual sequences from around the world to account for outside viral introductions. We employed an equal temporal subsampling scheme to enrich for under sampled time periods by randomly choosing a maximum of 400 total sequences per region (HCT, KC, and other) sampled equally per each calendar month via Augur filter(5), resulting in a set of 1137 total sequences, which were input into the model. Given the differential number of specimens in each region-year-month combination, not all demes included a total number of 400 sequences. We chose an equal temporal subsampling scheme based on recent work showing that maximizing spatiotemporal diversity reduces bias in MASCOT(6).

Using the compiled input sequence set, we employed a MASCOT-Skyline approach, which approximates the structured coalescent, to predict when the most recent common ancestor for each sequence pair in our input set existed and which of the three regions this ancestor would have existed in. To generate these predictions, we made assumptions about effective population sizes of the three regions and migration among the regions. To allow for population sizes to change over time, we modeled effective population sizes similar to the Skygrid approach for unstructured populations(7). We estimated the effective population size for each location between time *t*=0×tree height, …, *t*=1×tree height. Between each time point where we estimated *Ne*, we assumed exponential growth. *A priori*, we assumed that the effective population size at time *t*+1 is normally distributed with mean 0 and standard deviation σ, with σ being estimated. We assumed the migration rate to be constant forward-in-time, $m\ \frac{f}{zy}$, between states *y* and *z*. As the structured coalescent assumes backwards-in-time migration rates, we assumed that backwards-in-time rate of migration between state *y* and z, $m\frac{b}{yz} = m\ \frac{f}{zy} \times \frac{Ne(t)_z}{Ne(t)_y}$. To infer effective population sizes and migration rates over time, we employed an adaptable multivariate gaussian operator(8).

Parameter traces were visually evaluated for convergence using Tracer (v1.7.1)(9) and 30% burn-in was applied for all phylodynamic analyses. Output from our modeling analysis was a phylogenetic tree with internal nodes representing common ancestors of input sequence pairs. Tree plotting was performed with baltic (https://github.com/evogytis/baltic) and data visualizations were done using Altair(10). We summarized trees as maximum clade credibility trees using TreeAnnotator and visually inspected posterior tree distributions using IcyTree(11). Transmission between regions was calculated by measuring the number of migration jumps from HCT to KC and vice versa walking from tips to root in the posterior set of trees. Persistence time was measured by calculating the average number of days for a tip to leave its sampled location (HCT, KC, other), walking backwards up the phylogeny from tip up until node location was different from tip location(12).

References

1. Bennett JC, Luiten KG, O'Hanlon J, Han PD, McDonald D, Wright T, et al. Utilizing a university testing program to estimate relative effectiveness of monovalent COVID-19 mRNA booster vaccine versus two-dose primary series against symptomatic SARS-CoV-2 infection. Vaccine. 2024;42:1332–41.

2. Weil AA, Luiten KG, Casto AM, Bennett JC, O'Hanlon J, Han PD, et al. Genomic surveillance of SARS-CoV-2 Omicron variants on a university campus. Nat Commun. 2022;13:5240.

3. Weil AA, Sohlberg SL, O'Hanlon JA, Casto AM, Emanuels AW, Lo NK, et al. SARS-CoV-2 Epidemiology on a Public University Campus in Washington State. Open Forum Infect Dis. 2021;8:ofab464.

4. Aksamentov I, Roemer C, Hodcroft E, Neher R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. JOSS. 2021;6:3773.

5. Huddleston J, Hadfield J, Sibley TR, Lee J, Fay K, Ilcisin M, et al. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. J Open Source Softw. 2021;6:2906.

6. Layan M, Müller NF, Dellicour S, De Maio N, Bourhy H, Cauchemez S, et al. Impact and mitigation of sampling bias to determine viral spread: Evaluating discrete phylogeography through CTMC modeling and structured coalescent model approximations. Virus Evol. 2023;9:vead010.

7. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol Biol Evol. 2013;30:713–24.

8. Baele G, Lemey P, Rambaut A, Suchard MA. Adaptive MCMC in Bayesian phylogenetics: an application to analyzing partitioned data in BEAST. Bioinformatics. 2017;33:1798–805.

9. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018;67:901–4.

10. VanderPlas J, Granger B, Heer J, Moritz D, Wongsuphasawat K, Satyanarayan A, et al. Altair: Interactive Statistical Visualizations for Python. JOSS. 2018;3:1057.

11. Vaughan TG. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. Bioinformatics. 2017;33:2392–4.

12. Bedford T, Cobey S, Beerli P, Pascual M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). PLoS Pathog. 2010;6:e1000918.

Supplementary Note S1

Notes on Interpretation of Analyses of Transmission Dynamics between HCT and non-HCT populations

A) It should be noted that that reviewing WA sequences for possible HCT descendants permits detection of only a small fraction of spill-over events from HCT to WA. Due to a number of factors including geographically uneven sampling across the state and chronologically uneven sampling during the study period, it is difficult to estimate the size of this fraction. To provide some context, there was approximately one sequenced SARS-CoV-2 case for every 10 participants enrolled in HCT and one sequenced SARS-CoV-2 case for every 65 residents of WA. Assuming roughly equal per capita incidences of SARS-CoV-2 in HCT and WA over the study period, an HCT SARS-CoV-2 case would be about 6 times more likely to be detected and sequenced than a WA case. Consequently, if we had 6 sequenced HCT cases that all transmitted virus to one non-HCT WA resident, on average only one of the six transmitted WA viruses would be sequenced (assuming no further onward transmission for simplicity). In this scenario, we would not identify a WA descendant of the transmission event in 5 out of 6 cases. Under real world conditions, this estimate would further be affected by the number of WA cases "produced" by an HCT spill-over event and also by the methodology we used to detect possible WA descendants of HCT viruses, which is not 100% sensitive or specific.

B) Similarly, estimates produced by the modeling analysis of the number of viral introduction events into KC and HCT during the study period are not meant to reflect the total number of introductions into these populations. Rather these are estimates of the number of introduction events necessary to explain the cases represented by the 1137 sequences analyzed by the model and so represent a lower bound of the total number of introductions into these regions.

C) KC has population of 2.252 million while HCT enrolled 37,360 participants. If we imagine a SARS-CoV-2 transmission chain starting in KC, we expect a 1.7% chance of that transmission chain jumping into HCT due to differences in population size alone. Conversely, if the transmission chain started in HCT, we estimate a 98.3% chain of this chain infecting an individual in KC. This asymmetry corresponds to a 59.2 fold larger viral migration rate from HCT to KC than vice versa. The fact that this magnitude increase is similar but still less than the 85.7 fold difference in estimated forward migration rates suggests that the difference in migration rates can largely be explained by the difference in population sizes, further augmented by the presence of population structure.

Supplementary Note S2

Notes on Re-infection/Superinfection

A) For purposes of this study, we deemed an HCT participant to have been re-infected or superinfected with SARS-CoV-2 during the study period if they had at least two sequenced SARS-CoV-2 specimens that were determined to be of different clades and/or lineages. We did not observe any participants with 3 or more sequenced specimens each of different clades and/or lineages. It is possible that we excluded from the re-infected/superinfected participant group participants that experienced re-infection/superinfection with two closely related viruses (ie of the same clade and lineage). However, given the difficulty in distinguishing these situations from on-going shedding of the same virus, we chose to limit our list of re-infected/superinfected participants to those that had viruses of different clades and/or lineages.

B) Of the 46 individuals identified as re-infected or superinfected these, 36 were first infected by a non-Omicron SARS-CoV-2 variant followed by an Omicron variant, 9 were infected by two different Omicron variants, and one was infected with two different non-Omicron variants. One individual had two sequenced specimens of two different clades/lineages collected only 6 days apart (first specimen is 22B/BA.5 and second is 21L/BA.2.3). The sequencing reads for both specimens were reviewed and there was no evidence in either specimen of mixed infection. These two specimens were the participant's only SARS-CoV-2 positive specimens collected by HCT. Prior to testing positive the first time, the participant reported a known SARS-CoV-2 exposure but no recent travel. The participant denied having any history of previous SARS-CoV-2 infection prior to this first positive test. The participant reported experiencing symptoms for several days. The possibility of a specimen swap cannot be completely excluded, though there are numerous measures built into the study protocol to prevent such errors and a review of the documentation surrounding the collection of these specimens found no anomalies. We cannot exclude the possibility that another individual tested using the participant's identity, though it is less likely in this case as both specimens were collected via staff observed in-person testing.

Supplementary Note S3

Note on Vaccination in HCT Population

A) SARS-CoV-2 vaccines were introduced in the middle of year 1 of the study. As older individuals and those with medical co-morbidities were initially prioritized for vaccination, it is likely that average vaccination rates in KC were higher than among the relatively young and healthy HCT population in the second half of year 1. However, after vaccination was open to and recommended for all individuals, vaccine uptake rates were extremely high in the HCT population. By January 2022, 99% of students, faculty, and staff had received the primary vaccination series. Vaccination was strongly recommended for university students, faculty, and staff in the Fall of 2021 and then mandated starting January 1, 2022. In comparison, an estimated 79% of KC residents completed the initial vaccine series(1). Consequently, in year 2 of the study, it is likely that vaccination rates in the HCT population were higher than in the overall KC population. Given that there was a likely shift in relative vaccination rates between HCT and KC during the study period (from KC > HCT during the second half of year 1 to HCT > KC during year 2), it is possible that these vaccination rates impacted the relative flow of viral transmission between HCT and KC, which switched from KC to HCT during year 1 to HCT to KC during year 2.

B) Changes in vaccination policy and vaccine availability during the roll-out of SARS-CoV-2 vaccines also likely affected relative vaccination rates within the HCT population. For instance, during the second half of year 1, the average age of HCT student participants who reported having completed their primary vaccination series was 22.9 versus an average age of 21.2 for those who reported not having completed their primary vaccination series (vaccination status and age was recorded at the time of testing for SARS-CoV-2). The difference in these two average values then narrowed to 21.6 versus 21.2 and 22.7 and 22.3 during the first and second halves of year 2, respectively. Because we would expect that vaccination decreases a participant's risk of re-infection, differences in vaccination rates by age within HCT may have been a factor in our observation that the average age at the time of infection was lower for those who experienced re-infection during the study than the average age at the time of infection for all participants with a sequenced virus. We also reported that younger HCT participants were more likely to have a virus that was closely related to other HCT viruses than older HCT participants. However, we do not think that relative vaccination rates within HCT impacted this observation. While it makes sense that vaccination status would influence a participant's overall risk of acquiring SARS-CoV-2, it is not clear to us how vaccination status alone would impact *where* an HCT participant with SARS-CoV-2 acquired their virus (on-campus versus from the community). It seems more likely to us that location of viral acquisition would be determined by place of residence, time spent on-campus, and other behavioral factors.

References

1. Summary of COVID-19 vaccination among King County residents. [Internet]. Communicable Diseases Epidemiology and Immunization Section, Seattle-King County Public Health; Available from: https://kingcounty.gov/en/dept/dph/health-safety/disease-illness/covid-19/data/vaccination

Supplementary Figure Legends

Supplementary Figure S1: Total number of sequenced HCT specimens collected across the study period by two week sliding window.

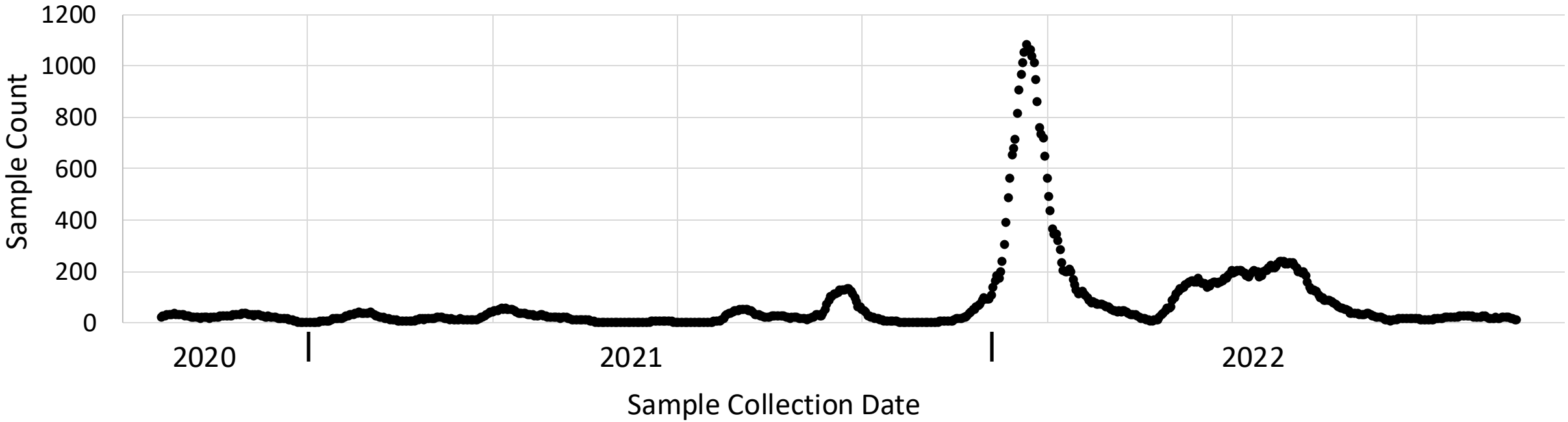
Supplementary Figure S2: Percent of WA clades/lineages observed among sequenced HCT specimens by study month.

Supplementary Figure S3: Date of first observation of clades and lineages in WA or KC and in HCT. Blue dots represent Nextstrain clades. Orange dots represent Pango lineages. A) Chart with date of first observation in HCT on the x-axis and first observation in WA on the y-axis. B) Chart with date of first observation in HCT on the x-axis and first observation in KC on the y-axis.

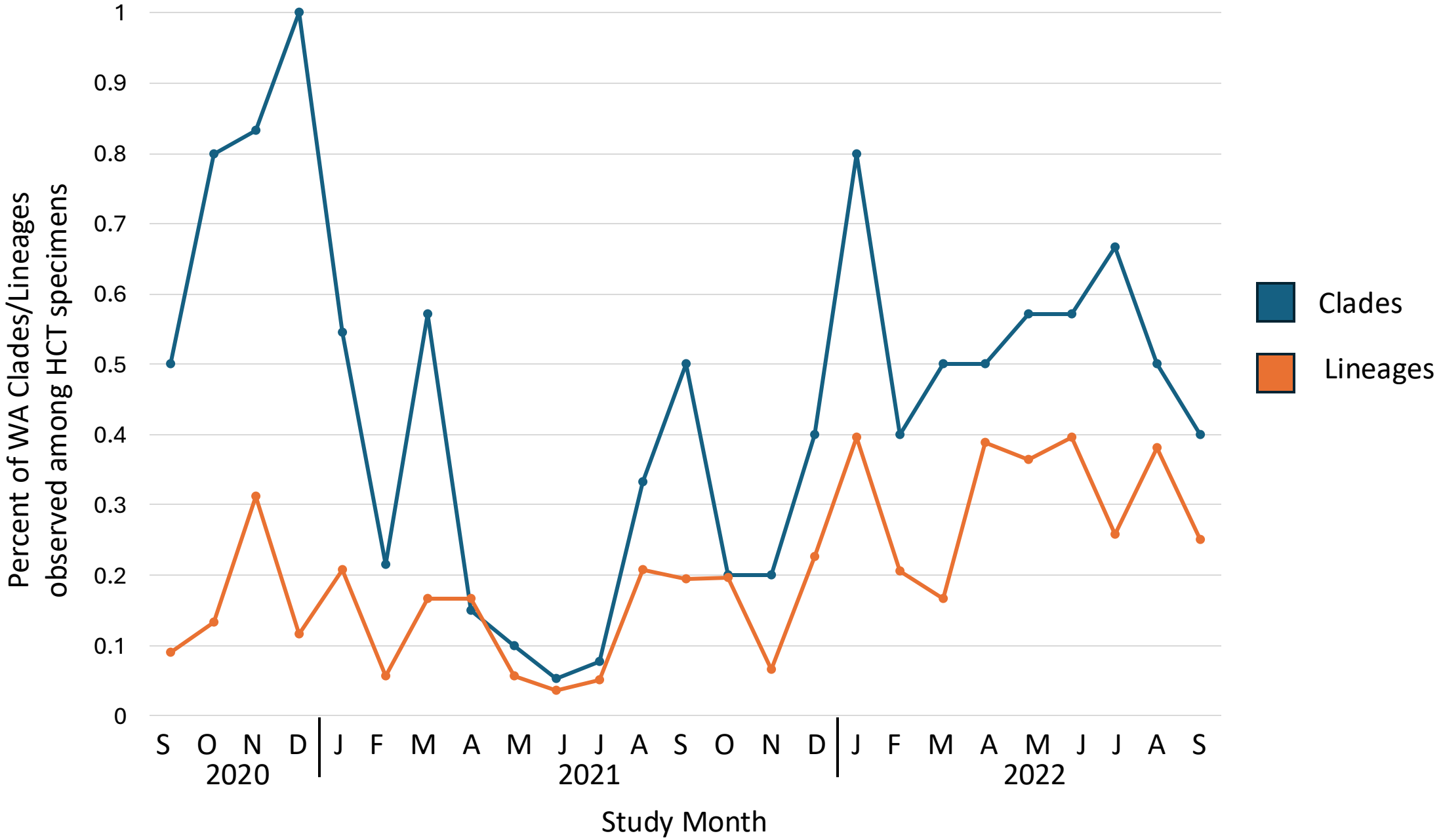
Supplementary Figure S4: Average size of HCT sequence clusters by month. A) Proportion of total sequenced HCT specimens represented by each Nextstrain clade by month. B) Average size of zero distance clusters of each Nextstrain clade by month. C) Average size of phylogenetic clusters of each Nextstrain clade by month.

Supplementary Figure S5: Demographic and epidemiologic characteristics of participants who experienced re-infection and of those with sequenced specimens in clusters. For each category (female, male, Latinx, Not Latinx, White, Asian, Other Race, Seattle campus, and other campuses), the mid-line (x = 0) represents the frequency of that category in the whole dataset. The purple, yellow, green, and blue dots give the percent difference between the frequency of that category among those who experienced re-infection and those with sequences in zero distance (groups of identical sequences), HCT-only zero distance (groups of identical sequences with haplotype unique to HCT), and phylogenetic clusters (groups of sequences that cluster phylogenetically), respectively, and the frequency of that category in the whole dataset. The intervals marked by black dots connected by a bar mark the 95% confidence interval of the values given by the purple, yellow, green, and blue dots.

Supplementary Figure S1

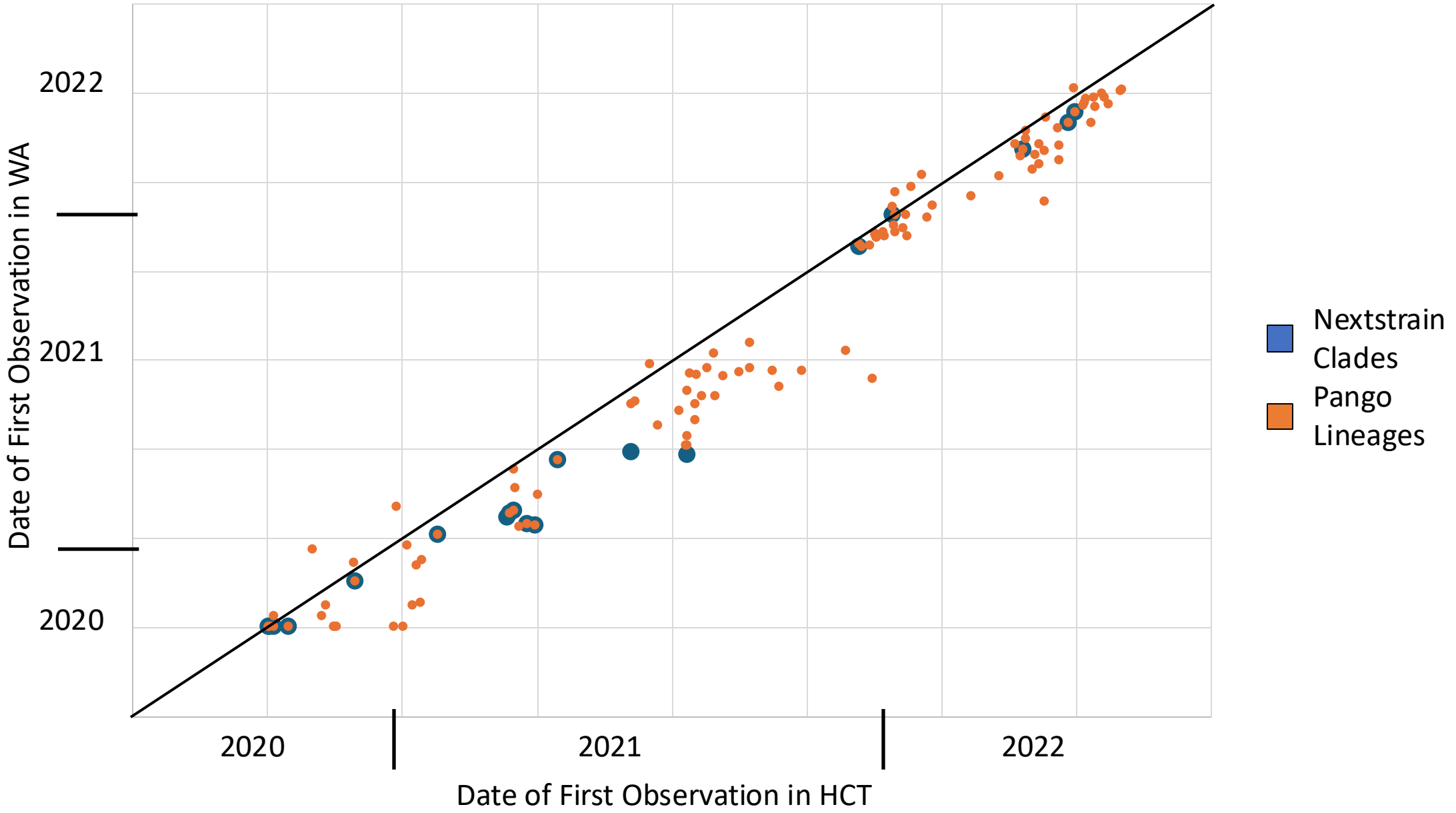


Supplementary Figure S2

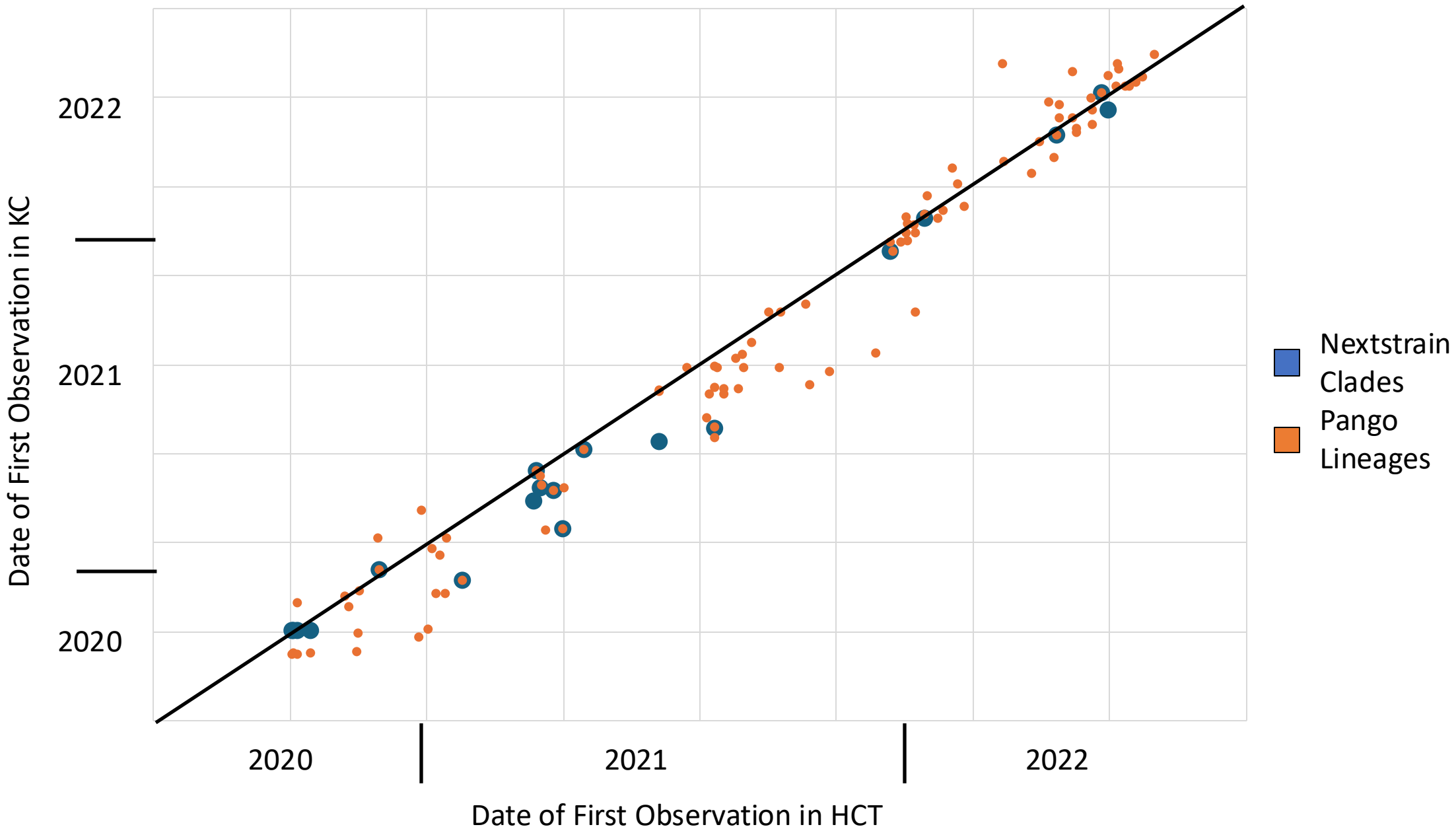


A

Supplementary Figure S3

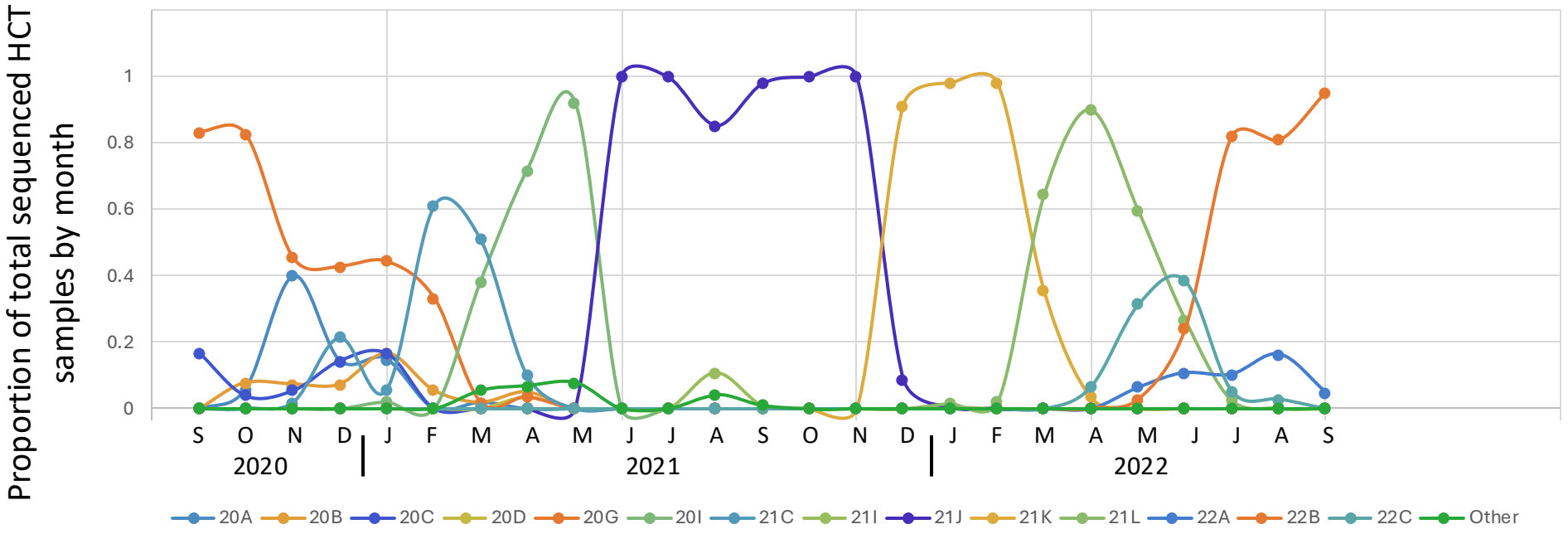


B

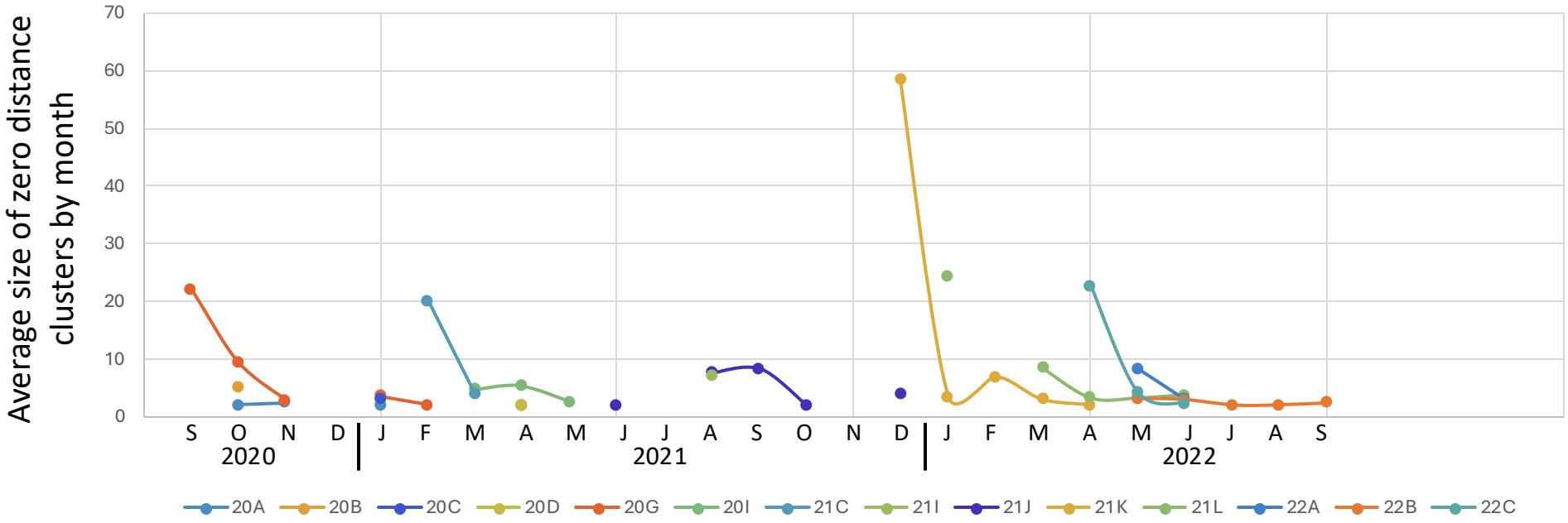

2022
2021
2020
Date of First Observation in KC
2020
2021
2022
Date of First Observation in HCT
Nextstrain Clades
Pango Lineages

Supplementary Figure S4

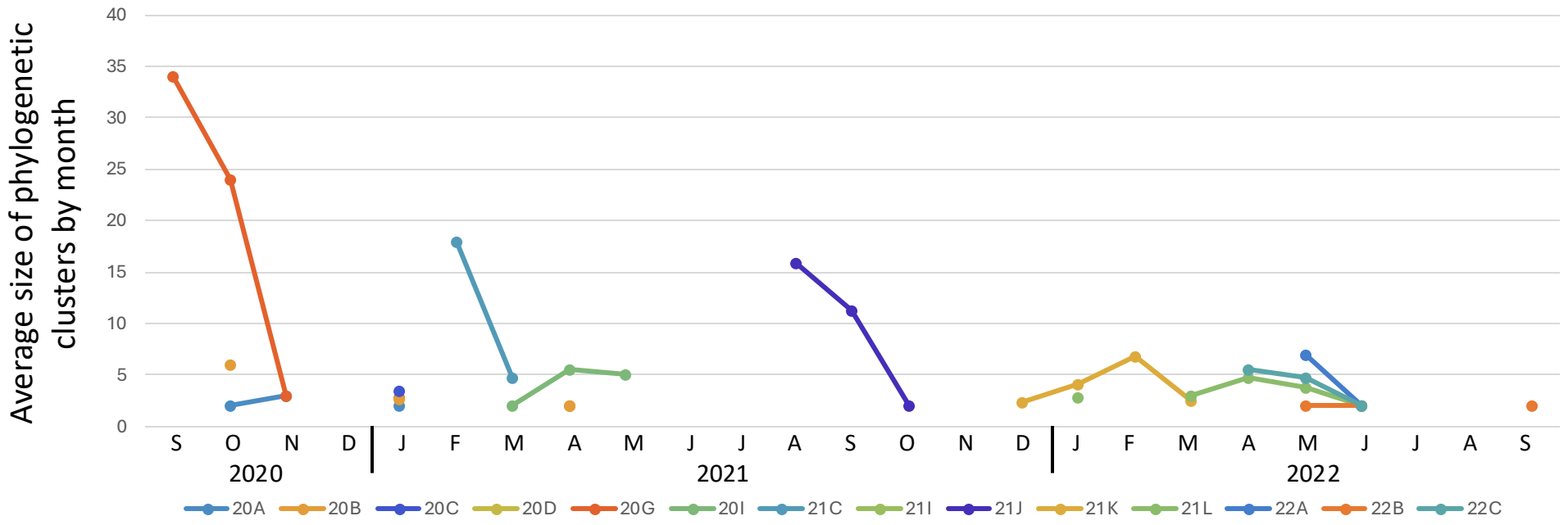
A



B


Average size of zero distance clusters by month
70
60
50
40
30
20
10
0
S O N D J F M A M J J A S O N D J F M A M J J A S
2020
2021
2022
20A 20B 20C 20D 20G 20I 21C 21I 21J 21K 21L 22A 22B 22C

C


Average size of phylogenetic clusters by month
40
35
30
25
20
15
10
5
0
S O N D J F M A M J J A S O N D J F M A M J J A S
2020
2021
2022
20A 20B 20C 20D 20G 20I 21C 21I 21J 21K 21L 22A 22B 22C

Supplementary Figure S5