Multiple Introductions of SARS-CoV-2 Alpha and Delta Variants into White-Tailed Deer in Pennsylvania

Andrew D. Marques,[a] Scott Sherrill-Mix,[a] John K. Everett,[a] Hriju Adhikari,[a] Shantan Reddy,[a] Julie C. Ellis,[b] Haley Zeliff,[c] Sabrina S. Greening,[b] Carolyn C. Cannuscio,[d,e] Katherine M. Strelau,[d,e] ⓘRonald G. Collman,[f] ⓘBrendan J. Kelly,[g] Kyle G. Rodino,[h] ⓘFrederic D. Bushman,[a] Roderick B. Gagne,[b] Eman Anis[c]

[a]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[b]Department of Pathobiology, Wildlife Futures Program, University of Pennsylvania School of Veterinary Medicine, New Bolton Center, Kennett Square, Pennsylvania, USA

[c]Department of Pathobiology, University of Pennsylvania School of Veterinary Medicine, New Bolton Center, Kennett Square, Pennsylvania, USA

[d]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[e]Department of Family Medicine and Community Health, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[f]Pulmonary, Allergy and Critical Care Division; Department of Medicine; University of Pennsylvania Perelman School of Medicine; Philadelphia, Pennsylvania, USA

[g]Division of Infectious Diseases; Department of Medicine & Department of Biostatistics, Epidemiology, and Informatics; Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[h]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

ABSTRACT The SARS-CoV-2 pandemic began by viral spillover from animals to humans; today multiple animal species are known to be susceptible to infection. White-tailed deer, *Odocoileus virginianus*, are infected in North America at substantial levels, and genomic data suggests that a variant in deer may have spilled back to humans. Here, we characterize SARS-CoV-2 in deer from Pennsylvania (PA) sampled during fall and winter 2021. Of 123 nasal swab samples analyzed by RT-qPCR, 20 (16.3%) were positive for SARS-CoV-2. Seven whole genome sequences were obtained, together with six more partial spike gene sequences. These annotated as alpha and delta variants, the first reported observations of these lineages in deer, documenting multiple new jumps from humans to deer. The alpha lineage persisted in deer after its displacement by delta in humans, and deer-derived alpha variants diverged significantly from those in humans, consistent with a distinctive evolutionary trajectory in deer.

IMPORTANCE Coronaviruses have been documented to replicate in numerous species of vertebrates, and multiple spillovers of coronaviruses from animals into humans have founded human epidemics. The COVID-19 epidemic likely derived from a spillover of SARS-CoV-2 from bats into humans, possibly via an intermediate host. There are now several examples of SARS-CoV-2 jumping from humans into other mammals, including mink and deer, creating the potential for new animal reservoirs from which spillback into humans could occur. For this reason, data on formation of new animal reservoirs is of great importance for understanding possible sources of future infection. Here, we identify extensive infection in white-tailed deer in Pennsylvania, including what appear to be multiple independent transmissions. Data further suggests possible transmission among deer. These data thus help identify a potential new animal reservoir and provide background information relevant to its management.

Address correspondence to Eman Anis, eanis@vet.upenn.edu, Roderick B. Gagne, rgagne@vet.upenn.edu, or Frederic D. Bushman, bushman@pennmedicine.upenn.edu.

Multiple spillovers of coronaviruses from animals to humans have founded epidemics in humans. Examples include spillover of SARS from bats to humans, likely with civets as an intermediate host, and spillover of middle east respiratory

syndrome (MERS) from bats to camels to humans (1–3). For SARS-CoV-2, the epidemic likely began with spillover from *Rhinolophus* bats to humans, possibly via an intermediate host (4, 5).

SARS-CoV-2 infection has further "spilled back" from humans into numerous animal species (6–10), creating the risk of formation of new animal reservoirs. Examples of animals known to be infectible with SARS-CoV-2 include great apes, mice, cats, dogs, deer, mink, and hamsters (6, 8, 11–16). Infection of mink farms has been widely reported, and tracking via analysis of genome sequences suggests that a mink lineage has spilled back into humans (6).

The first indications of the establishment of a potential wildlife reservoir of concern have been observed in white-tailed deer, *Odocoileus virginianus*. White-tailed deer appear to have been widely infected in North America (11, 17, 18). Comparison of genomic sequences from the Alberta region suggest that the virus evolved a novel lineage during transmission in deer, and that this spilled back on at least one occasion to a human from the same region (19, 20).

Here, we report a study of infection and evolution of SARS-CoV-2 in the Pennsylvania region. White-tailed deer were sampled throughout Pennsylvania from 10/2/2021 to 12/27/2021 with the goal of characterizing this potential new animal reservoir. Samples were tested for the presence of SARS-CoV-2 by qPCR, and the viral lineages present assessed using viral whole genome sequencing after multiplex PCR. For samples with lower RNA amounts, sequencing was carried out on a single nested PCR amplicon encoding the spike receptor binding domain, which allows variant profiling. Results report the nature of this emerging reservoir in a heavily populated state.

RESULTS

Tracking SARS-CoV-2 in white-tailed deer using RT-qPCR. SARS-CoV-2 was detected by RT-qPCR in nasal swabs from 20 of 123 wild white-tailed deer sampled (16.3%; 95% confidence interval (CI): 10.4, 24.2) (described in Table S1 and S2). There was no significant difference in infection frequency associated with the cause of death (roadkill 7/24, and hunter-harvested 12/79; P-value = 0.14). There was no significant difference in infection rates between sexes (females 6/43, males 14/60; P-value = 0.46) and age groups (fawn 1/18, yearling 1/22, and adult 17/81; P-value = 0.093). No information was available on possible symptoms or disease for the animals sampled.

Virus positive deer were identified in 10 of the 31 Pennsylvania counties surveyed (Fig. 1). We used a Bayesian hierarchical binomial model to estimate positivity and predicted an average county had 9.9% positivity (95% CrI: 1.7-22%). Pike and Monroe counties were associated with higher apparent prevalence (Monroe: 7.5x higher odds (1.0-110x) of positivity; Pike: 9.3x higher odds [1.3-110x]). Grouping by region also showed a significant difference in infection rate (P-value = 0.0047), with the northeast region of PA showing a higher proportion of positive deer than the southeast region (Fisher's exact test with false discovery rate (FDR) correction; P-value = 0.017).

Assessing SARS-CoV-2 in white-tailed deer using viral whole genome sequencing. Eight samples had viral RNA concentrations that in human specimens typically yield successful whole genome sequencing (cycle threshold values from RT-qPCR less than 30), and high-quality genome sequences were recovered from 7 (Table S2). To confirm sample provenance, the non-viral sequence reads were analyzed for the deer samples and for human samples that were sequenced in parallel. All deer samples yielded reads mapping to the deer genome, and few or none mapping to the human genome. In contrast, reads from human samples overwhelmingly mapped to the human genome (Fig. S1), confirming the host species origin of our samples as deer.

Samples with lower SARS-CoV-2 RNA concentrations (Ct ≥30) were amplified using a single nested PCR amplicon targeting the spike coding region and sequenced, yielding spike sequences from six additional samples. Spike sequences were then used for lineage assignment (21) (Table S2).

Characteristics of SARS-CoV-2 genome sequences from deer in Pennsylvania. Deer whole genome sequences from Pennsylvania showed divergences relative to the
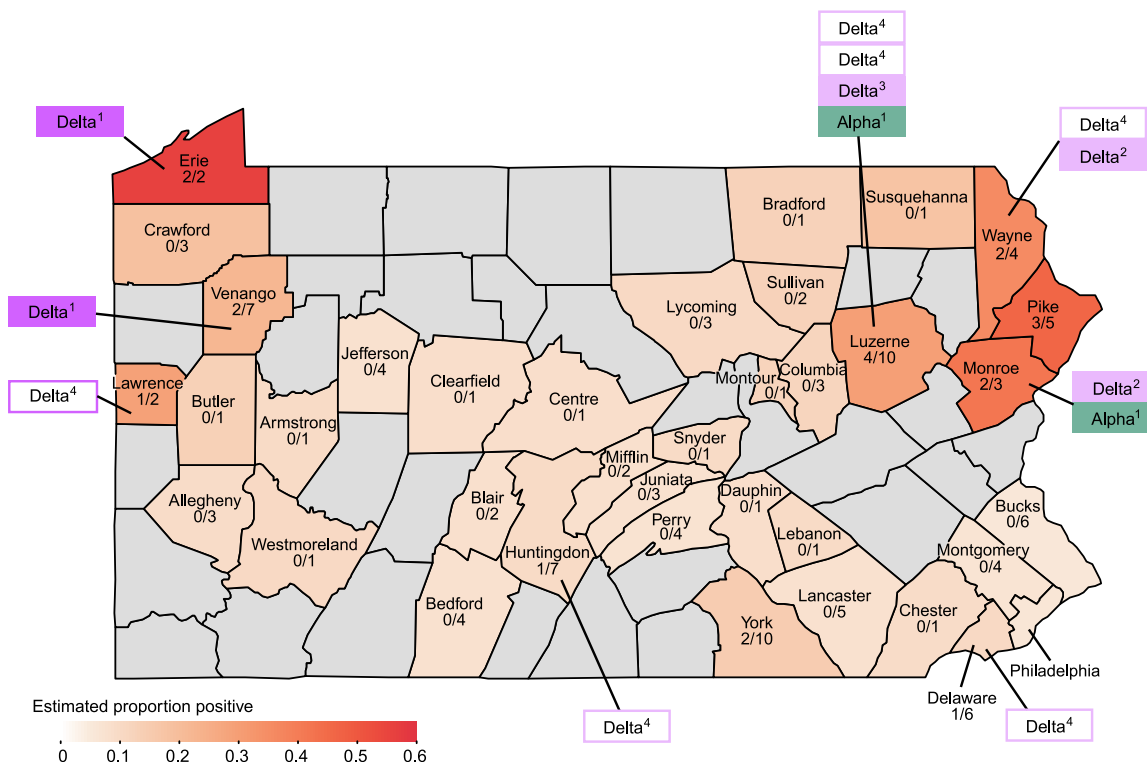
FIG 1 Map of Pennsylvania (PA), showing sampling sites and locations of SARS-CoV-2 positive deer. The counties comprising PA are outlined. The estimated proportion of positive samples is shown by the white-to-red color scale; counties that were not sampled are shown in gray. Total sample numbers and the number positive are written on each county sampled. The deer samples sequenced were assigned to variants as indicated by the rectangles outside the map; variant type is color-coded (teal for alpha/B1.1.7, purple/pink for delta/AY.#). The open boxes indicate sequences were available for the spike-only.

original Wuhan SARS-CoV-2 strain and previously reported human sequences, and were distinct from prior sequences reported from deer (Fig. 2A). A full list of deer substitutions is in Tables S3. The complete deer genome sequences reported here annotated as either alpha or delta variants (22), the first reported identification of these lineages in wild white-tailed deer. Previously the alpha variant was shown to be able to infect white-tailed deer that were experimentally inoculated (18). All the spike-only sequences harbored polymorphisms that place them in the delta lineage and were inconsistent with other common lineages.

The types of base changes associated with substitutions were not uniform. Changes from C in the Wuhan reference to U were found to be the most frequent (Fig. 2B). The frequency of C to U was even higher in deer than in humans (Pearson's Chi-squared test $P = 0.031$, followed by a *post hoc* analysis based on residuals yielding P-value $= 0.0010$ for C to U substitutions), consistent with a previous study (20), and suggestive of host-specific mutation rates.

To identify any deer-specific sequence polymorphisms, the alpha and delta SARS-CoV-2 sequences from deer determined here were compared to viral genomes from humans and deer reported previously. All deer genome sequences available from GISAID ($n = 108$) were downloaded (Table S4 and S5), corresponding to samples acquired from 9/28/2020 to 2/25/2021 in Iowa and Ohio and dominated by early pandemic lineages including B.1.2 and B.1.311 (11, 17). Local human isolates were derived from our SARS-CoV-2 sequencing program monitoring the Delaware Valley, including Philadelphia, PA (23, 24). This revealed several substitutions that were highly enriched or invariant in deer isolates, but rare or absent in human isolates. These include 3 silent mutations in ORF1ab, C7303U, C9430U, and C20259U. Mutation C7303U was found in 86% of these PA deer and 29% of previously published deer, but in only 0.04% of humans in Delaware Valley and 0.09% of global SARS-CoV-2 genomes reported by NextStrain (data from 1/26/2022). Mutation c9430t was found in 43% of PA deer and 56% of previously published deer, but 0.35% of Delaware
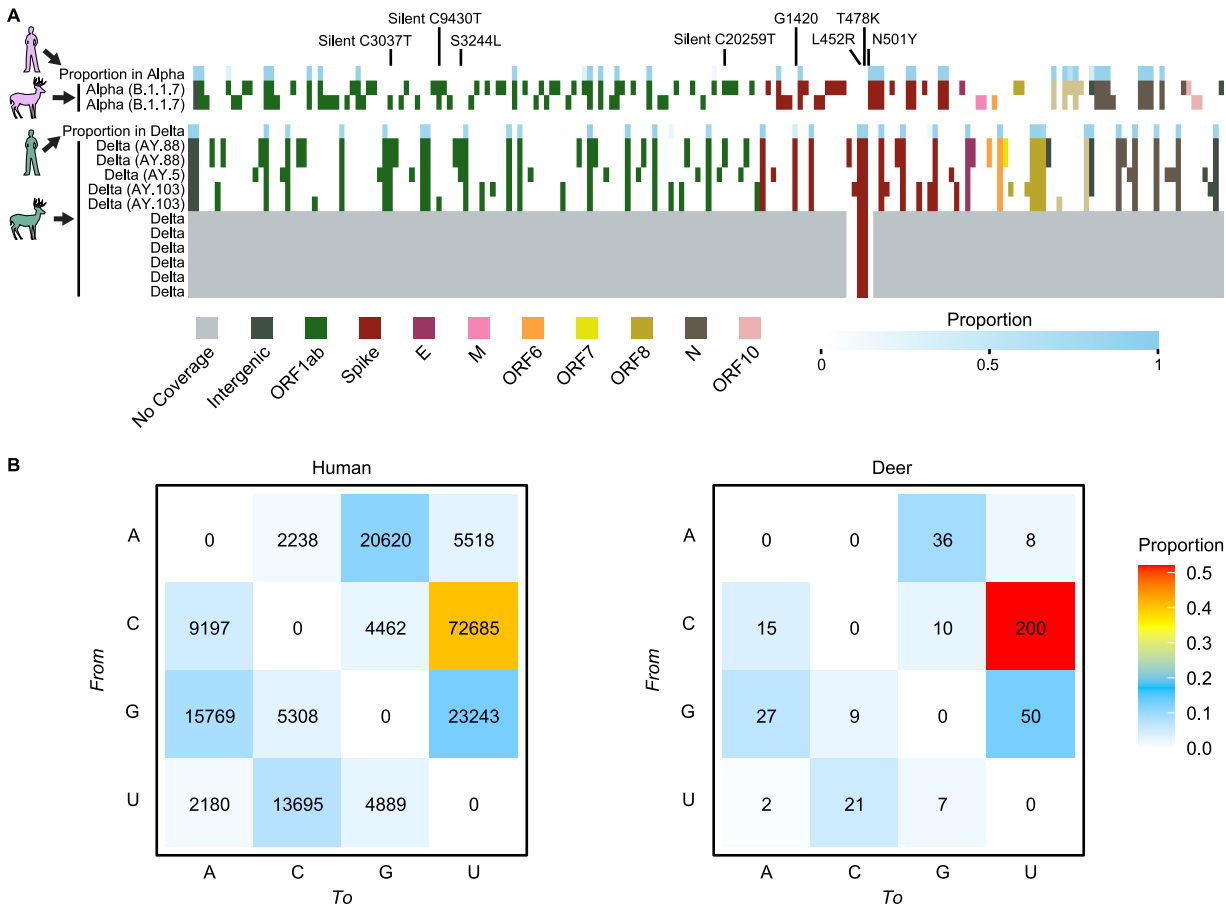
FIG 2 Base substitutions detected in sequences of deer from Pennsylvania. In (A) the proportions in humans sampled in the Delaware Valley is shown in blue along the top for alpha (upper) and delta (lower). Substitutions relative to the Wuhan strain references are shown by the colored boxes. The proportions in humans are shown by the blue shading in each box. Genes sampled are shown by the color code indicated along the bottom. Gray indicates lack of sampling. The bottom six rows of the delta samples indicate the spike-only amplicon sequences. (B) Types of base substitutions away from the Wuhan reference detected in humans (right) and deer (left). The proportions are shown by the color code at the left.

Valley and 0.46% of global human SARS-CoV-2 genomes. Mutation C20259U was found in 43% of PA deer and 19% of previously published deer, whereas it was absent in Delaware Valley human subjects and present in 0.12% of global genomes. The enrichment of these mutations suggests possible functional interaction with deer-specific factors, which could influence RNA synthesis, RNA folding, or protein binding.

To assess relationships among deer and human-derived SARS-CoV-2 sequences, we used NextClade for sequence alignment and IQ-TREE to construct a maximum-likelihood phylogenetic tree for viral sequences from alpha (Fig. 3) and delta (Fig. 4) variants, in each case comparing PA deer to their 100 nearest genetic neighbors from NCBI's global human data set and contemporaneous human-derived sequences from the same variant in the Delaware Valley region. We also compared the deer lineages observed longitudinally to all contemporaneous human lineages sampled in PA (Fig. 5, and Table S5 and S6).

The 2 alpha sequences sampled from deer came from adjoining counties in northeastern PA. The isolates differed by 45 substitutions and were widely separated on the phylogenetic tree. A parsimonious explanation is that the two alpha lineages were introduced independently into deer and then diversified during transmission within deer. Molecular clock analysis yielded data consistent with a model in which alpha jumped from humans to deer as early as March/April of 2021 (95% credible interval [CrI] for VSP3574: 3/10/2021 to 3/30/2021; and for VSP3516: 4/23/2021 to 5/4/2021), during the alpha wave, and persisted in deer up until the sampling time 6 to
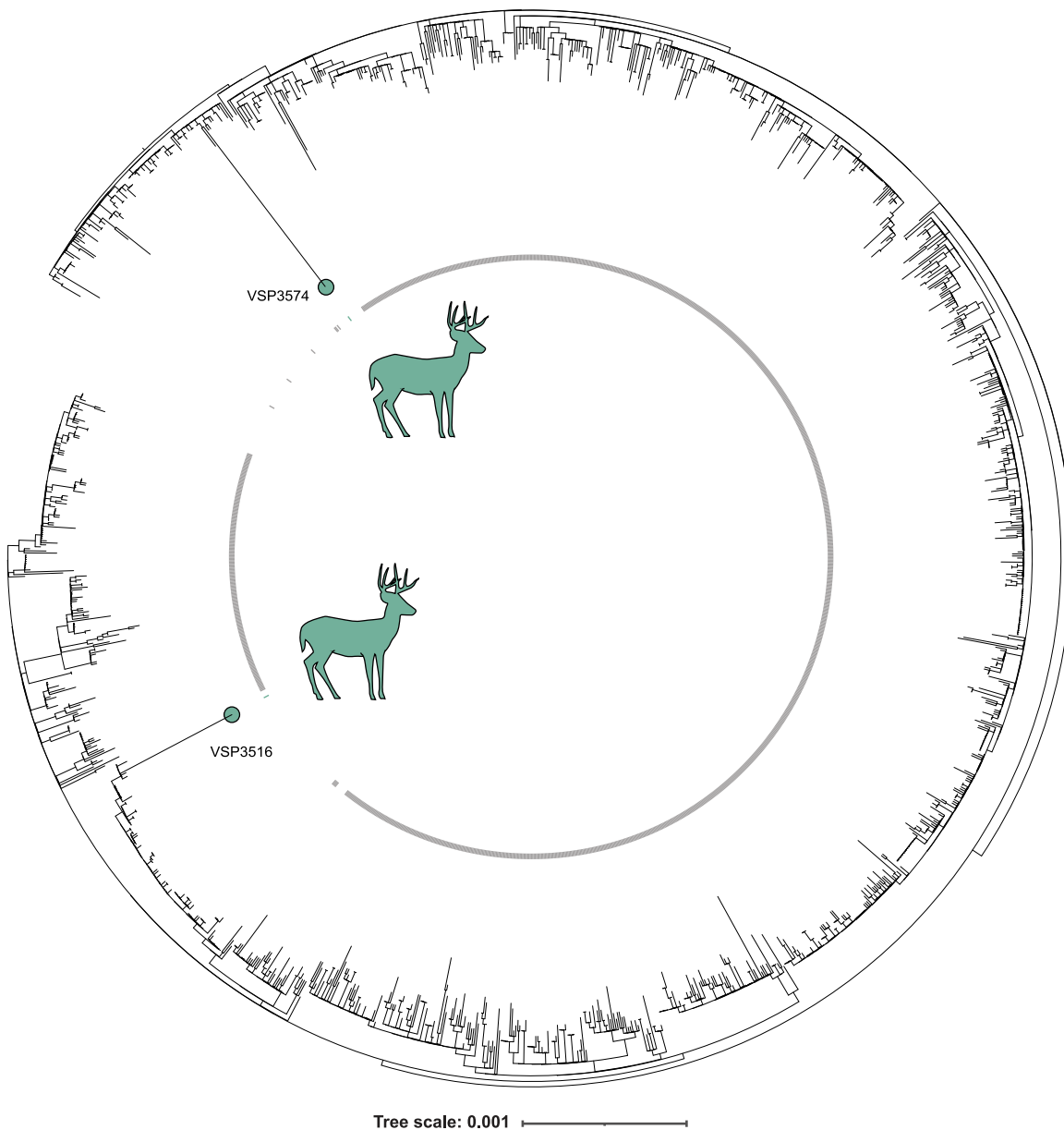
FIG 3 Analysis of alpha variant SARS-CoV-2 whole genome sequences from white-tailed deer, global nearest neighbor human-derived sequences, and Delaware Valley human-derived sequences. Phylogenetic analysis of alpha white-tailed deer-derived sequences, nearest global sequences, and contemporaneous human alpha lineage sequences from Delaware Valley, collected from 1/6/2021 to 11/16/2021(2 deer sequences, 200 global nearest neighbor human sequences, and 1239 contemporanious Delaware Valley human). Included in the contemporaneous Delaware Valley human samples are two examples of humans infected with the alpha variant during the delta wave as late as 10/17/2021 and 11/14/2021, although these samples are notably more similar to earlier human alpha sequences than the deer alpha sequences. Dear sequences are colored teal, contemporaneous Delaware Valley sequences are labeled gray, and global nearest neighbor sequences are uncolored.

7 months later in mid-November, when alpha was no longer circulating in the US human population (Fig. S2 and Table S5). The modeled estimates of the timing of spillover to deer have high uncertainties and provide only an early bound on spillover since it is unclear what proportion of viral evolution occurred in unsampled humans prior to transmission to deer and what occurred while circulating within deer.

For the delta variant (Fig. 4), 5 deer genomes annotated as delta, and all 6 spike sequences showed patterns also consistent with delta (Table S3). Two deer genomes were assigned to the AY.103 clade, 2 genomes to AY.88, and 1 to AY.5. The delta lineages from deer were not as diverged from related human sequences as the alpha
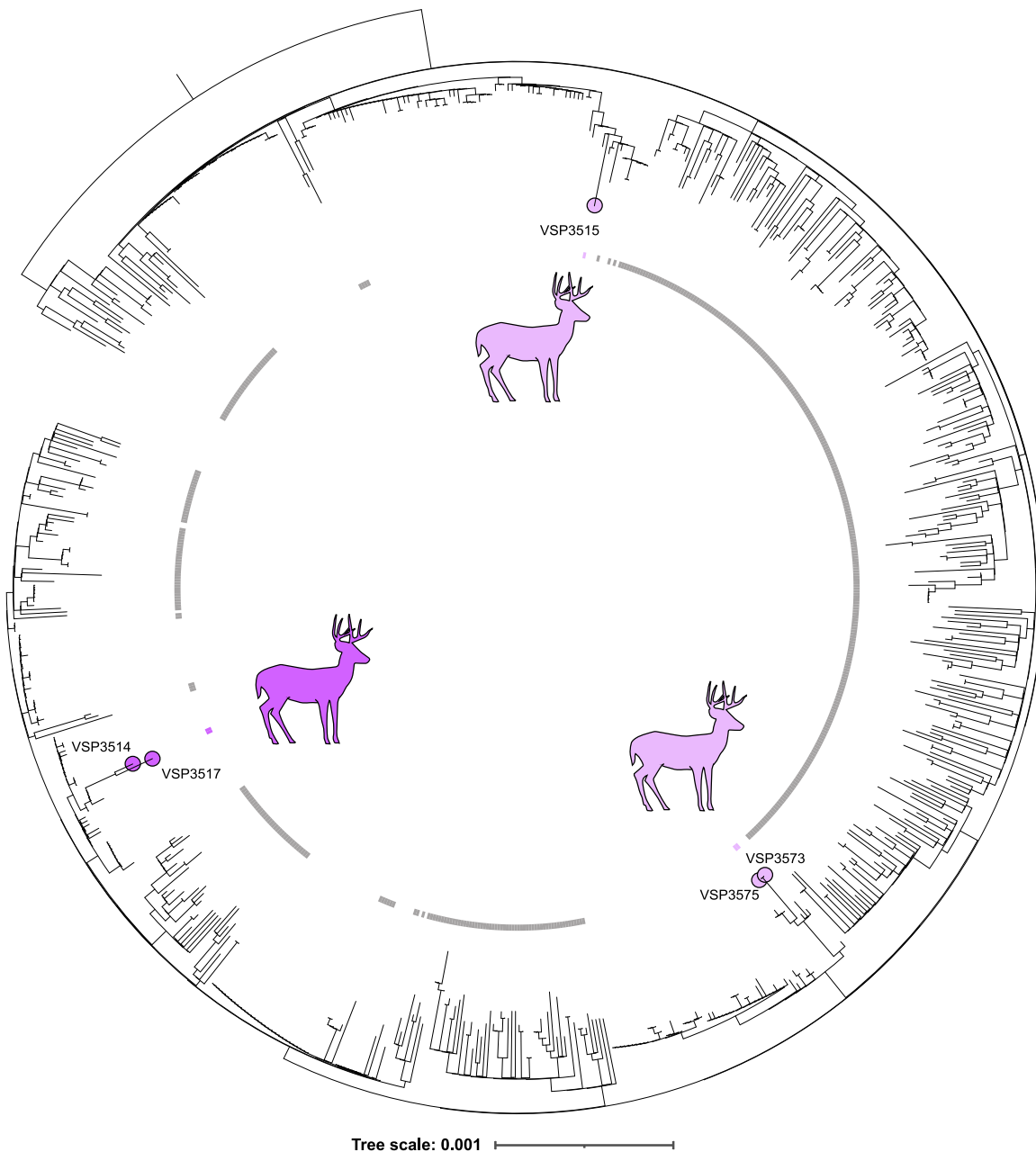
FIG 4 Phylogeny of delta white-tailed deer sequences, nearest global human sequences, and human delta lineage sequences from Delaware Valley, collected from 10/14/2021 to 11/28/2021 (spanning the period of genomes obtained from deer). Included are 5 deer, 300 global nearest neighbor human, and 440 Delaware Valley human sequences. Deer sequences are marked in purple, Delaware Valley human sequences are marked in gray, and global nearest neighbor sequences are unmarked. Time resolved trees for each spillover are in Fig. S2.

lineages, consistent with more recent introductions and less long-term circulation in deer. This is as expected since delta became widespread in the US only shortly before the time of sampling. Molecular clock analysis suggested that the closely related AY.103 pair of genomes likely entered the deer population as early as early September 2021 (95% CrI: 8/8/2021 to 10/4/2021); the AY.5 lineage entered as early as mid-August 2021 (95% CrI: 7/31/2021 to 8/27/2021); and the AY.88 pair entered deer as early as mid-October 2021 (95% CrI: 9/30/2021 to 10/31/2021) (Fig. S2). For the delta cases, the proposed jumps from deer to humans likely occurred during the height of the delta wave. The 6 samples sequenced as single spike amplicons also displayed mutations consistent with delta variants and were collected at times and locations similar to the
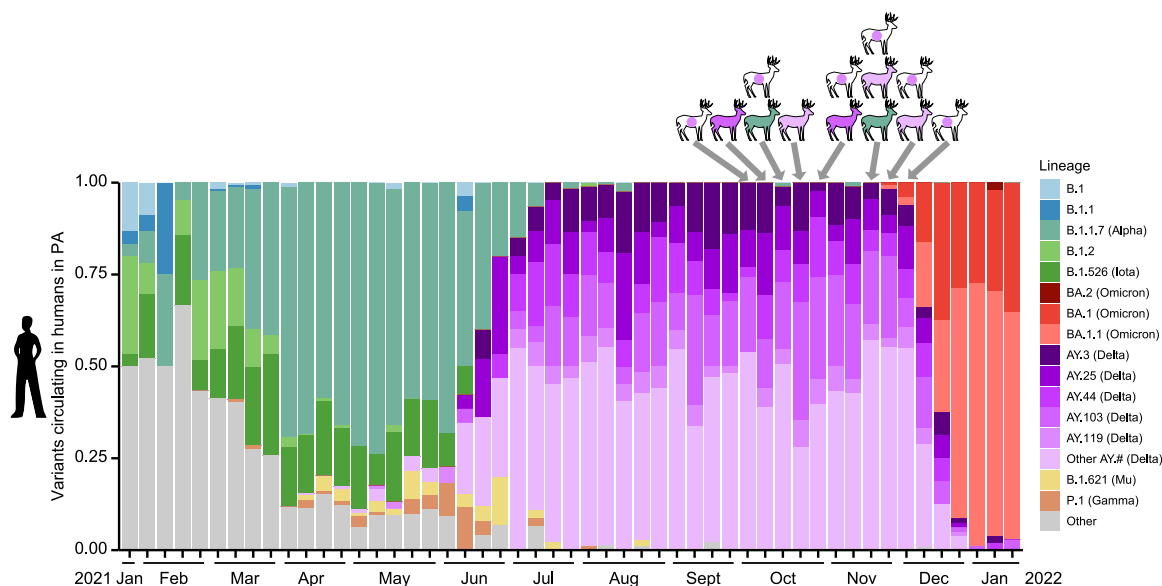
FIG 5 Longitudinal comparison of deer variants and human variants. The bar plots show the progression of SARS-CoV-2 variants detected by genome sequencing in humans in eastern Delaware Valley from 1/31/2021 to 1/3/2022. The variants are color-coded according to the key to the right. The variants from white-tailed deer sequences are shown at the top of the figure, with the arrows showing the times of sampling. The deer isolates are color-coded on the deer icons as in the key; purple dots on deer icons indicate sequences identified as delta lineage by spike amplicon sequences but without whole genome sequences to further identify clade.

fully sequenced delta variant genomes from deer. The small size of these amplicons prevents assignment to clade and identification of transmission events but confirms the extensive presence of delta in deer.

DISCUSSION

These data support 5 independent transmissions of SARS-CoV-2 from humans to deer in the samples from Pennsylvania analyzed. These include one jump each for the 2 alpha sequences, and 3 for the delta sequences (Fig. S2). Other scenarios are also possible, involving more independent jumps and convergent evolution in deer, or fewer jumps and divergent evolution along trajectories matching trajectories in humans.

Mechanisms of infection and transmission in deer are incompletely understood. Studies of experimental infections show efficient transmission between deer, potentially involving close interactions such as touching noses and grooming (25). The mechanism of transmission from humans to deer remains obscure, though evidence here of multiple transmission events suggests it is not a rare event.

This study has several limitations. The sample size is modest, and whole genome sequence acquisition was limited by viral RNA concentrations in samples. Our sample size limits the interpretation of epidemiological findings (e.g., differences in infection rates between sample types), though does not undermine the primary findings of this work (e.g., divergence of the alpha lineage in deer, and evidence for multiple spillovers). The background human data is incomplete, with only an estimated 1.7% percent of all human cases in PA subjected to viral whole genome sequencing, limiting our ability to identify and correctly time spillover and spillback events. Each evolutionary analysis of sequences requires assumptions on the structures of background populations that are not fully investigated. In addition, we were not able to obtain serum samples from the deer studied and so could not investigate immune responses.

In summary, we report a survey of SARS-CoV-2 in 123 deer in Pennsylvania over the fall-winter of 2021, which showed 16% of the deer to be positive. Prior surveys carried out over the fall and winter of 2020 also showed high point prevalence of infection in Iowa (33%) (17) and Ohio (36%) (11), and later extensive infection at other locations (20, 26–28). We report the first examples of the alpha and delta lineages in wild white-

tail deer and identify five likely independent spillovers from humans to deer among seven fully sequenced genomes. Given that there are estimated to be 30 deer per square mile in PA, and over a million deer total, this suggests an enormous number of spillovers and infected deer in the state (29, 30). Our findings of alpha persistence in deer after replacement of alpha by delta in humans, and the divergence seen between our deer and human alpha genomes, are all consistent with long-term persistence and spread of the alpha variant in deer. Yet, there is no evidence for spillback of the deer lineages identified here into humans; ongoing efforts to characterize human and deer SARS-CoV-2 lineages are valuable to maintain surveillance for such events.

MATERIALS AND METHODS

Collection of samples from white-tailed deer. Samples were collected from hunter-harvested deer and injured deer that were euthanized by state game wardens. Nasal swabs were taken within hours of death, placed in phosphate-buffered saline (PBS) and stored in commercial refrigerators in field offices. Samples were shipped to the University of Pennsylvania within 1 week of collection and stored at −80°C until RNA extraction. Fisher's Exact tests were performed using R Statistical software (v4.0.5) (31) (R Core Development Team 2021) to test for differences in proportions of virus positive and negative deer by sex, age, and cause of death (i.e., hunter-harvested, road-killed, or all other causes of death).

RNA extraction and detection of SARS-CoV-2 RNA by RT-qPCR. Nucleic acid was extracted with a QIAamp viral RNA minikit (Qiagen) according to the manufacturer's instructions. The presence of SARS-CoV-2 RNA was assessed by a RT-qPCR targeting 2 regions of the viral nucleocapsid gene as described previously (32).

SARS-CoV-2 whole genome sequencing. Viral genomes were sequenced using the POLAR protocol (33). The sample RNA (5 μL) was mixed with Random Hexamers (0.5 μL of 50 μM, Thermo Fisher), dNTPs Mix (0.5 μL of 10 mM each, Thermo Fisher), and nuclease-free water (1 μL). This mixture was incubated (5 min at 65°C). Subsequently, reverse transcription was performed adding 6.5 μL from the previous reaction to SuperScript III Reverse Transcriptase (0.5 μL, Thermo Fisher), 5X First-Strand Buffer (2 μL, Thermo Fisher), DTT (0.5 μL of 0.1M, Thermo Fisher), and RNaseOut (0.5 μL, Thermo Fisher). This mixture was incubated (50 min at 42°C, then 10 min at 70°C). For cDNA amplification, the previous mixture (2.5 μL) was added to Q5 Hot Start DNA polymerase (0.25 μL, NEB), 5X Q5 Reaction Buffer (5 μL, NEB), dNTPs mix (0.5 μL of 10 mM each, NEB), and pooled Artic-ncov2019 v4 primers (reaction set 1 used 4.0 μL of pooled primer set 1, reaction set 2 used 3.98 μL of pooled primer set 2, IDT) and water (volume reaction brought to 25 μL). The samples were brought to 98°C for 30 s then cycled from 98°C for 15 s to 65°C for 5 min for 25 cycles, followed by 98°C for 15 s and 65°C for 5 min. The 2 reactions (one for each primer set) were pooled together. Tagmentation was completed with the Nextera XT Library Preparation Kit (Illumina). IDT for Illumina DNA/RNA UD Indexes were used for barcoding (Illumina). Quantification of DNA was performed using Quant-iT PicoGreen dsDNA quantitation assay kit (Invitrogen). The pooled libraries were quantified using the Qubit1X dsDNA HS assay kit (Invitrogen) and sequenced using the Illumina NextSeq platform.

SARS-CoV-2 spike amplicon sequencing. Amplicon sequencing was carried out using nested PCR (34) to target the spike coding region. Samples with genomic viral load deemed too low for whole genome sequencing, or samples that failed whole genome sequencing, underwent nested PCR targeting the spike's receptor binding domain sequence. Primers used are provided in Table S7. PCR 1 includes reverse transcription and initial DNA amplification using the Superscript IV One-Step RT-PCR System (Thermo Fisher Scientific,12594100). The 25 μL reaction consists of 5 μL of extracted viral RNA (extracted as described in the whole genome sequencing section), 0.25 μL of Superscript IV, 12.5 μL of 2X Platinum SuperFi RT-PCR mix, 1.25 μL of each PCR 1 primers (10 μM), 4.75 μL molecular grade water. The cycle conditions are as follows: 25°C for 2 min, 50°C for 20 min, 95°C for 2 min, 25 cycles of amplification (95°C for 2 min, 55°C for 30 s, 70°C for 1 min). PCR 2 is a nested PCR and addition of sequencing adapters for the amplicon product from reaction 1. The 25 μL mixture contains 5 μL of amplicon product from reaction 1, 12.5 μL 2x Q5 hot start master mix (NEB, M0494S), 0.5 μL of 10 mM dNTPs (NEB, N0447S), 1.25 μL of each PCR 2 primers (10 μM), and 7.5 μL water. The conditions are as follows: 95°C for 2 min, 20 cycles of amplification (95°C for 15 s, 55°C for 30 s, 72°C for 1 min). PCR 3 adds the unique dual indexes and clustering adapters. The 50 μL reaction contains 25 μL of amplicon product from PCR 2, 0.5 μL of Phusion polymerase (NEB, M0530L), 10 μL of 5x Phusion buffer, 1 μL of 10 mM dNTPs (NEB, N0447S), 5 μL of IDT for Illumina DNA/RNA UD Indexes set A (Illumina, 20027213), and 8.5 μL of water. The reaction conditions are as follows: 98°C for 3 min, 7 cycles of amplification (95°C for 15 s, 50°C for 30 s, 72°C for 30 s), and 72°C for 7 min. 10 μL of the barcoded PCR 3 mixture was pooled together, AMPure purified, and quantified by Qubit1X dsDNA and TAPE station before being sequenced on an Illumina MiSeq instrument using a 600 cycle v3 standard flow cell 290/10/10/290 protocol (Illumina, MS-102-3003). After sequencing, BWA v0.7.17-r1188 was used to generate bam files, Samtools v1.10 (using htslib 1.10.2-3) was used to filter reads for lengths between 280 and 300 bp, filter reads by quality using a PHRED threshold of 30, sort the reads, index the reads, and generate a pileup. Samples must meet a minimum mapping quality PHRED score of 30, have 95% 50-fold coverage for the 550 bp amplicon target to pass additional filtering. Substitutions were called with a greater than 0.67 allele frequency for a given position.

Sequence data processing. Sequence data were processed as previously described (23). BWA aligner tool (v0.7.17) was used to align viral sequences to the Wuhan reference sequence (NC_045512.2) (35). Samtools package (v1.10) was used to filter alignments (36). Variants were called using Pangolin lineage software (3.1.17 with the PangoLEARN 2021-12-06 release) (22, 37). NVSL's vSNP pipeline was also run on deer samples and results compared to those generated by our published pipeline; findings from both platforms deviated only for thresholds used to call low abundance substitutions. All variant calls remained the same between platforms.

Host sequence analysis. The proportion of host sequences was inferred using a sampling of raw reads from all samples on any sequencing batch performed with deer specimen. 1,000 raw reads for each sample were blasted (38–40) against a database constructed of a SARS-CoV-2 genome (NC_045512.2), human genome (GCF_000001405.39), and white-tailed deer genome (GCF_002102435.1). Reads were tallied for each genome to which they were closely matched.

Mutation analysis. One hundred eight deer-derived SARS-CoV-2 genomes were downloaded from GISAID and compared with the 7 deer sequenced in this publication and the human data set (Tables S4 and S5). The previously published genomes were downloaded on 12/29/2021. Mutations were called in reference to the Wuhan strain (NC_045512.2).

Bayesian analysis of county proportions. To account for the variable sampling between counties and potential similarities between neighboring counties, we estimated the underlying proportion of deer testing positive within each county of the m counties using a Bayesian conditional autoregressive model. The number of positive tests, y_i, out of n_i total tests within each county i was modeled as:

$$y_i \sim \text{Binomial}(p_i, n_i)$$

where p_i is the proportion of deer expected to test positive in that county and:

$$p_i = \text{logit}^{-1}(\alpha + \beta_i)$$

Here, α represents the average proportion positive for a county and the vector of differences from this average for each county, β, is distributed as a multivariate normal:

$$\beta \sim \text{Normal}_{\text{prec}}(0, \frac{1}{\sigma}(D - \theta A))$$

where D is a $m \times m$ matrix with 0s on the off diagonal and the diagonal element on each row i, $D_{i,i}$, equal to the number of counties that are adjacent to county i, A is a $m \times m$ adjacency matrix with element $A_{i,j}$ is 1 if county i neighbors county j and 0 otherwise and the diagonal set to 0 and $Normal_{prec}(x, y)$ is a multivariate normal distribution with means x and precision matrix y. For priors, θ was given a uniform prior between 0 and 1, $\sigma \sim \text{Gamma}(1, 1)$ and $\alpha \sim \text{Normal}(-2, 10)$.

Posterior probability distributions were estimated using Markov chain Monte Carlo sampling using Stan v2.21.0 (41).

County adjacency data was obtained from the US Census Bureau (https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html).

Time-scale Bayesian maximum clade credibility tree. A phylogenetic approach using a time scale Bayesian maximum clade credibility (MCC) tree was used to estimate the introduction of SARS-CoV-2 into PA white-tailed deer. We subsampled genomes based on nearest human neighbor to deer sequences using NCBI's global nucleotide data set consisting of 943,071 complete SARS-CoV-2 genomes. To determine the nearest human neighbors, a BLASTn search was used. If a deer's nearest hits included another deer, then both were analyzed in the same tree. Sequences without exact dates of collection were excluded. Sample sizes were chosen as the 100 nearest neighbor whole genome sequence isolates from humans. Each data set was aligned using NextClade with Wuhan-Hu-1 as a reference (NC_045512.2). A Markov chain Monte Carlo (MCMC) method was used to generate time-scales Bayesian molecular clocks using BEAST v.1.10.4 (42). Several parameters were assessed using path sampling and stepping-stone sampling of marginal likelihood estimation (Table S6). The best performing settings were a general time reversible substitution model with gamma-distributed rate variation among sites (Yang96) (43, 44) with an uncorrelated relaxed lognormal clock to allow for branch-specific variation in evolutionary rates (45–47) and a Bayesian Skyline tree prior with a group size of 10 (48). A previously published SARS-CoV-2 paper examining infection of white-tailed deer also converged on similar parameters (11) and similar parameters performed well in a comparison of potential models of SARS-CoV-2 evolution (49). MCMC sampling was run for 100 million iterations with subsampling every thousand iterations and 10 million iterations discarded as burn-in. The BEAGLE 3 library was used to improve computational performance (50). Tracer v.1.7.1 was used to visually assess convergence, TreeAnnotator v.1.10.4 (42) to summarize the MCC tree and FigTree v.1.4.4 (51) to visualize the tree. All tree analyses were repeated 3 times independently and visually checked for convergence and all effective sample sizes were confirmed to be greater than 200 using Tracer v1.7.2 (52). Maximum-likelihood trees were generated by IQ-TREE v1.6.12 (53) using the samples selected for the BEAST analysis and the root-to-tip genetic distances were compared to collection dates using a rooting maximizing correlation with TempEst v1.5.3 (54) to assess sample selection and evolutionary rates (Fig. S3).

Human subjects. Human sequences newly determined here were collected as follows. For most samples, the University of Pennsylvania Institutional Review Board (IRB) reviewed the research protocol and deemed the limited data elements extracted to be exempt from human subject research per 45 CFR 46.104, category 4 (IRB #848605). For hospitalized subjects, following informed consent (IRB protocol #823392), patients were sampled by collection of saliva, oropharyngeal and/or nasopharyngeal swabs, or endotracheal aspirates if intubated, as previously described (23). Further samples were collected from asymptomatic

subjects detected in a screening program at the Perelman School of Medicine at the University of Pennsylvania and symptomatic subjects tested throughout the PennMedicine clinical network under IRB protocols #843565 and #848608. Human samples were sequenced as described for deer samples.

Phylogenetic analysis. Deer and human viral sequences are available at GenBank (Table S2 and S5). Samples used include i) deer viral sequences of the same variant, ii) 100 nearest human viral sequences from a global search, and iii) all human viral sequences of the same variant from our local Delaware River Valley data set. NextClade was used to align sequences to the Wuhan reference (53). IQ-TREE (v1.6.12) was used to infer a tree using maximum-likelihood methods using 1,000 bootstrap replicates (55). Visualization of the inferred tree was performed using iTOL v6.

Materials. Key materials and resources are compiled in Table S7.

Data availability. Sequence accession numbers for deer-derived SARS-CoV-2 genomes can be found in Table S2 (OM570187-OM570193 & ON350842-ON350847). Accession numbers for human SARS-CoV-2 genomes can be found in Table S5. Sequence processing and analysis code are deposited at https://doi.org/10.5281/zenodo.4046252 and https://doi.org/10.5281/zenodo.6842232. BEAST/BEAUti xml files used for the Bayesian maximum credibility trees are available at https://zenodo.org/record/6761298#.YrnKZXbMKUk.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.1 MB.
FIG S2, PDF file, 0.3 MB.
FIG S3, PDF file, 0.9 MB.
TABLE S1, PDF file, 0.4 MB.
TABLE S2, XLSX file, 0.03 MB.
TABLE S3, XLSX file, 0.02 MB.
TABLE S4, XLSX file, 0.3 MB.
TABLE S5, XLSX file, 0.3 MB.
TABLE S6, PDF file, 0.4 MB.
TABLE S7, PDF file, 0.4 MB.

REFERENCES

1. Temmam S, Vongphayloth K, Salazar EB, Munier S, Bonomi M, Régnault B, Douangboubpha B, Karami Y, Chretien D, Sanamxay D, Xayaphet V, Paphaphanh P, Lacoste V, Somlor S, Lakeomany K, Phommavanh N, Pérot P, Donati F, Bigot T, Nilges M, Rey F, Werf Svd, Brey P, Eloit M. 2021. Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. Res Square https://doi.org/10.21203/rs.3.rs-871965/v1.

2. Zhou H, Ji J, Chen X, Bi Y, Li J, Wang Q, Hu T, Song H, Zhao R, Chen Y, Cui M, Zhang Y, Hughes AC, Holmes EC, Shi W. 2021. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2

and related viruses. Cell 184:4380–4391. https://doi.org/10.1016/j.cell
.2021.06.008.

3. Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou J-J, Li N, Guo Y, Li X, Shen X, Zhang Z, Shu F, Huang W, Li Y, Zhang Z, Chen R-A, Wu Y-J, Peng S-M, Huang M, Xie W-J, Cai Q-H, Hou F-H, Chen W, Xiao L, Shen Y. 2020. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature 583:286–289. https://doi.org/10.1038/s41586-020-2313-x.

4. Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, Anthony SJ, Barclay WS, Boni MF, Doherty PC, Farrar J, Geoghegan JL, Jiang X, Leibowitz JL, Neil SJD, Skern T, Weiss SR, Worobey M, Andersen KG, Garry RF, Rambaut A. 2021. The origins of SARS-CoV-2: a critical review. Cell 184:4848–4856. https://doi.org/10.1016/j.cell.2021.08.017.

5. Singh D, Yi SV. 2021. On the origin and evolution of SARS-CoV-2. Exp Mol Med 53:537–547. https://doi.org/10.1038/s12276-021-00604-z.

6. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, van der Spek A, Tolsma P, Rietveld A, Brouwer M, Bouwmeester-Vincken N, Harders F, Hakze-van der Honing R, Wegdam-Blans MCA, Bouwstra RJ, GeurtsvanKessel C, van der Eijk AA, Velkers FC, Smit LAM, Stegeman A, van der Poel WHM, Koopmans MPG. 2021. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science 371:172–177. https://doi.org/10.1126/science.abe5901.

7. Lu L, Sikkema RS, Velkers FC, Nieuwenhuijse DF, Fischer EAJ, Meijer PA, Bouwmeester-Vincken N, Rietveld A, Wegdam-Blans MCA, Tolsma P, Koppelman M, Smit LAM, Hakze-van der Honing RW, van der Poel WHM, van der Spek AN, Spierenburg MAH, Molenaar RJ, Rond J d, Augustijn M, Woolhouse M, Stegeman JA, Lycett S, Oude Munnink BB, Koopmans MPG. 2021. Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. Nat Commun 12:6802. https://doi.org/10.1038/s41467-021-27096-9.

8. Oreshkova N, Molenaar RJ, Vreman S, Harders F, Munnink BBO, Hakze-van der Honing RW, Gerhards N, Tolsma P, Bouwstra R, Sikkema RS, Tacken MGJ, de Rooij MMT, Weesendorp E, Engelsma MY, Bruschke CJM, Smit LAM, Koopmans M, van der Poel WHM, Stegeman A. 2020. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. Euro Surveill 25. https://doi.org/10.2807/1560-7917.ES.2020.25.23.2001005.

9. Salajegheh Tazerji S, Magalhães Duarte P, Rahimi P, Shahabinejad F, Dhakal S, Singh Malik Y, Shehata AA, Lama J, Klein J, Safdar M, Rahman MT, Filipiak KJ, Rodríguez-Morales AJ, Sobur MA, Kabir F, Vazir B, Mboera L, Caporale M, Islam MS, Amuasi JH, Gharieb R, Roncada P, Musaad S, Tilocca B, Koohi MK, Taghipour A, Sait A, Subbaram K, Jahandideh A, Mortazavi P, Abedini MA, Hokey DA, Hogan U, Shaheen MNF, Elaswad A, Elhaig MM, Fawzy M. 2020. Transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to animals: an updated review. J Transl Med 18:358. https://doi.org/10.1186/s12967-020-02534-2.

10. Murphy HL, Ly H. 2021. Understanding the prevalence of SARS-CoV-2 (COVID-19) exposure in companion, captive, wild, and farmed animals. Virulence 12:2777–2786. https://doi.org/10.1080/21505594.2021.1996519.

11. Hale VL, Dennis PM, McBride DS, Nolting JM, Madden C, Huey D, Ehrlich M, Grieser J, Winston J, Lombardi D, Gibson S, Saif L, Killian ML, Lantz K, Tell RM, Torchetti M, Robbe-Austerman S, Nelson MI, Faith SA, Bowman AS. 2021. SARS-CoV-2 infection in free-ranging white-tailed deer. Nature 602. https://doi.org/10.1038/s41586-021-04353-x.

12. Lenz OC, Marques AD, Kelly BJ, Rodino KG, Cole SD, Perera RAPM, Weiss SR, Bushman FD, Lennon EM. 2022. SARS-CoV-2 Delta variant (AY.3) in the feces of a domestic cat. Viruses 14:421. https://doi.org/10.3390/v14020421.

13. Sit THC, Brackman CJ, Ip SM, Tam KWS, Law PYT, To EMW, Yu VYT, Sims LD, Tsang DNC, Chu DKW, Perera RAPM, Poon LLM, Peiris M. 2020. Infection of dogs with SARS-CoV-2. Nature 586:776–778. https://doi.org/10.1038/s41586-020-2334-5.

14. Sia SF, Yan L-M, Chin AWH, Fung K, Choy K-T, Wong AYL, Kaewpreedee P, Perera RAPM, Poon LLM, Nicholls JM, Peiris M, Yen H-L. 2020. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. Nature 583:834–838. https://doi.org/10.1038/s41586-020-2342-5.

15. Yen H-L, Sit THC, Brackman CJ, Chuk SSY, Gu H, Tam KWS, Law PYT, Leung GM, Peiris M, Poon LLM, HKU-SPH study team. 2022. Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study. Lancet 399:1070–1078. https://doi.org/10.1016/S0140-6736(22)00326-9.

16. Nagy A, Stará M, Vodička R, Černíková L, Jiřincová H, Křivda V, Sedlák K. 2022. Reverse-zoonotic transmission of SARS-CoV-2 lineage alpha (B.1.1.7) to great apes and exotic felids in a zoo in the Czech Republic. Arch Virol 167:1681–1685. https://doi.org/10.1007/s00705-022-05469-9.

17. Kuchipudi SV, Surendran-Naira M, Ruden RM, Yona M, Nissly RH, Vandegrift KJ, Nelli RK, Lingling L, Jayarao BM, Maranas CD, Levine N, Willgert K, Conlan

AJK, Olsen RJ, Davis JJ, Musser JM, Hudson PJ, and Kapur V. 2022. Multiple spillovers from humans and onward transmission of SARS-CoV-2 in white-tailed deer. Proc Natl Acad Sci U S A 119:e2121644119. https://doi.org/10.1073/pnas.2121644119.

18. Cool K, Gaudreault NN, Morozov I, Trujillo JD, Meekins DA, McDowell C, Carossino M, Bold D, Mitzel D, Kwon T, Balaraman V, Madden DW, Artiaga BL, Pogranichniy RM, Roman-Sosa G, Henningson J, Wilson WC, Balasuriya UBR, García-Sastre A, Richt JA. 2022. Infection and transmission of ancestral SARS-CoV-2 and its alpha variant in pregnant white-tailed deer. Emerg Microbes Infect 11:95–112. https://doi.org/10.1080/22221751.2021.2012528.

19. Bashor L, Gagne RB, Bosco-Lauth AM, Bowen RA, Stenglein M, Vande-Woude S. 2021. SARS-CoV-2 evolution in animals suggests mechanisms for rapid variant selection. Proc Natl Acad Sci U S A 118:e2105253118. https://doi.org/10.1073/pnas.2105253118.

20. Pickering B, Lung O, Maguire F, Kruczkiewicz P, Kotwa JD, Buchanan T, Gagnier M, Guthrie JL, Jardine CM, Marchand-Austin A, Massé A, McClinchey H, Nirmalarajah K, Aftanas P, Blais-Savoie J, Chee H-Y, Chien E, Yim W, Goolia M, Suderman M, Pinette M, Smith G, Sullivan D, Rudar J, Adey E, Nebroski M, Côté M, Laroche G, McGeer AJ, Nituch L, Mubareka S, Bowman J. 2022. Highly divergent white-tailed deer SARS-CoV-2 with potential deer-to-human transmission. bioRxiv https://doi.org/10.1101/2022.02.22.481551.

21. O'Toole A, Pybus OG, Abram ME, Kelly EJ, Rambaut A. 2022. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. BMC Genomics 23:121. https://doi.org/10.1186/s12864-022-08358-2.

22. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5:1403–1407. https://doi.org/10.1038/s41564-020-0770-5.

23. Everett J, Hokama P, Roche AM, Reddy S, Hwang Y, Kessler L, Glascock A, Li Y, Whelan JN, Weiss SR, Sherrill-Mix S, McCormick K, Whiteside SA, Graham-Wooten J, Khatib LA, Fitzgerald AS, Collman RG, Bushman F. 2021. SARS-CoV-2 genomic variation in space and time inhHospitalized patients in Philadelphia. mBio 12:e03456-20. https://doi.org/10.1128/mBio.03456-20.

24. Marques AD, Sherrill-Mix S, Everett JK, Reddy S, Hokama P, Roche AM, Hwang Y, Glascock A, Whiteside SA, Graham-Wooten J, Khatib LA, Fitzgerald AS, Moustafa AM, Bianco C, Rajagopal S, Helton J, Deming R, Denu L, Ahmed A, Kitt E, Coffin SE, Newbern C, Mell JC, Planet PJ, Badjatia N, Richards B, Wang Z-X, Cannuscio CC, Strelau KM, Jaskowiak-Barr A, Cressman L, Loughrey S, Ganguly A, Feldman MD, Collman RG, Rodino KG, Kelly BJ, Bushman FD. 2022. SARS-CoV-2 variants associated with vaccine breakthrough in the Delaware Valley through summer 2021. mBio 13:e0378821. https://doi.org/10.1128/mbio.03788-21.

25. Hirth DH. 1977. Social behavior of white-tailed deer in relation to habitat. Wildlife Monographs 53:3–55.

26. Mallapaty S. 2022. COVID is spreading in deer. What does that mean for the pandemic? Nature 604:612–615. https://doi.org/10.1038/d41586-022-01112-4.

27. Roundy CM, Nunez CM, Thomas LF, Auckland LD, Tang W, Richison JJ, Green BR, Hilton CD, Cherry MJ, Pauvolid-Corrêa A, Hamer GL, Cook WE, Hamer SA. 2022. High seroprevalence of SARS-CoV-2 in white-tailed deer (Odocoileus virginianus) at one of three captive cervid facilities in Texas. Microbiol Spectr 10:e0057622. https://doi.org/10.1128/spectrum.00576-22.

28. Palermo PM, Orbegozo J, Watts DM, Morrill JC. 2022. SARS-CoV-2 neutralizing antibodies in white-tailed deer from Texas. Vector Borne Zoonotic Dis 22:62–64. https://doi.org/10.1089/vbz.2021.0094.

29. National Deer Association Staff. An annual report on the status of white-tailed deer - the foundation of the hunting industry in North America. National Deer Association, 2022. https://www.deerassociation.com/wp-content/uploads/2022/01/NDA-DR2022-Final.pdf.

30. Rosenberry CR, Fleegle JT, Wallingford BD. Monitoring deer populations in Pennsylvania. Bureau of Wildlife Management, Pennsylvania Game Commission 2011. https://www.pgc.pa.gov/Wildlife/WildlifeSpecies/White-tailedDeer/Documents/PASAK_Documentation.pdf.

31. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing. Accessed 16 April 2021.

32. Anis E, Turner G, Ellis JC, Di Salvo A, Barnard A, Carroll S, Murphy L. 2021. Evaluation of a real-time RT-PCR panel for detection of SARS-CoV-2 in bat guano. J Vet Diagn Invest 33:331–335. https://doi.org/10.1177/1040638721990333.

33. Hilaire BGS, Durand NC, Mitra N, Pulido SG, Mahajan R, Blackburn A, Colaric ZL, Theisen JWL, Weisz D, Dudchenko O, Gnirke A, Rao S, Kaur P, Aiden EL, Aiden AP . 2020. A rapid, low cost, and highly sensitive SARS-

CoV-2 diagnostic based on whole genome sequencing. bioRxiv. https://doi.org/10.1101/2020.04.25.061499.

34. Smyth DS, Trujillo M, Gregory DA, Cheung K, Gao A, Graham M, Guan Y, Guldenpfennig C, Hoxie I, Kannoly S, Kubota N, Lyddon TD, Markman M, Rushford C, San KM, Sompanya G, Spagnolo F, Suarez R, Teixeiro E, Daniels M, Johnson MC, Dennehy JJ. 2022. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. Nat Commun 13:635. https://doi.org/10.1038/s41467-022-28246-3.

35. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

37. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, Yeats C, du Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A. 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 7. https://doi.org/10.1093/ve/veab064.

38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Molecular Biology 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

39. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.

40. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

41. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A . 2017. STAN: a probabilisitic programming language. J Stat Software 76:1–32. https://doi.org/10.18637/jss.v076.i01.

42. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016. https://doi.org/10.1093/ve/vey016.

43. Yang Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J Mol Evol 42:587–596. https://doi.org/10.1007/BF02352289.

44. Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol Evol 11:367–372. https://doi.org/10.1016/0169-5347(96)10041-0.

45. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol 4:e88. https://doi.org/10.1371/journal.pbio.0040088.

46. Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. Syst Biol 56:453–466. https://doi.org/10.1080/10635150701420643.

47. Lepage T, Bryant D, Philippe H, Lartillot N. 2007. A general comparison of relaxed molecular clock models. Mol Biol Evol 24:2669–2680. https://doi.org/10.1093/molbev/msm193.

48. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185–1192. https://doi.org/10.1093/molbev/msi103.

49. Tay JH, Porter AF, Wirth W, Duchene S. 2022. The emergence of SARS-CoV-2 variants of concern is driven by acceleration of the substitution rate. Mol Biol Evol 39. https://doi.org/10.1093/molbev/msac013.

50. Ayres DL, Cummings MP, Baele G, Darling AE, Lewis PO, Swofford DL, Huelsenbeck JP, Lemey P, Rambaut A, Suchard MA. 2019. BEAGLE 3: improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. Syst Biol 68:1052–1061. https://doi.org/10.1093/sysbio/syz020.

51. Rambaut A. 2007. FigTree. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 30 December 2021.

52. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 67:901–904. https://doi.org/10.1093/sysbio/syy032.

53. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. 2021. Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software 6. https://doi.org/10.21105/joss.03773.

54. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2. https://doi.org/10.1093/ve/vew007.

55. Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst Biol 65:997–1008. https://doi.org/10.1093/sysbio/syw037.