Detection and characterization of the SARS-CoV-2 lineage B.1.526 in New York

Anthony P. West, Jr.[1*], Joel O. Wertheim[2], Jade C. Wang[3], Tetyana I. Vasylyeva[2], Jennifer L. Havens[4], Moinuddin A. Chowdhury[3], Edimarlyn Gonzalez[3], Courtney E. Fang[3], Steve S. Di Lonardo[3], Scott Hughes[3], Jennifer L. Rakeman[3], Henry H. Lee[5,6] Christopher O. Barnes[1], Priyanthi N. P. Gnanapragasam[1], Zhi Yang[1], Christian Gaebler[7], Marina Caskey[7], Michel C. Nussenzweig[7,8], Jennifer R. Keeffe[1], Pamela J. Bjorkman[1]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

[2]Department of Medicine, University of California San Diego, La Jolla, CA 92093

[3]New York City Public Health Laboratory, New York City Department of Health and Mental Hygiene, New York, NY, 10016 USA

[4]Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093

[5]Pandemic Response Laboratory, Long Island City, NY 11101

[6]Department of Genetics, Harvard Medical School, Boston, MA 02115

[7]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA.

[8]Howard Hughes Medical Institute, The Rockefeller University, New York, NY, 10065 USA.

*Corresponding author: Anthony P. West, Jr., apwest@caltech.edu

Abstract

Wide-scale SARS-CoV-2 genome sequencing is critical to tracking viral evolution during the ongoing pandemic. Variants first detected in the United Kingdom, South Africa, and Brazil have spread to multiple countries. We developed the software tool, Variant Database (VDB), for quickly examining the changing landscape of spike mutations. Using VDB, we detected an emerging lineage of SARS-CoV-2 in the New York region that shares mutations with previously reported variants. The most common sets of spike mutations in this lineage (now designated as B.1.526) are L5F, T95I, D253G, E484K or S477N, D614G, and A701V. This lineage was first sequenced in late November 2020 when it represented <1% of sequenced coronavirus genomes that were collected in New York City (NYC). By February 2021, genomes from this lineage accounted for ~32% of 3288 sequenced genomes from NYC specimens. Phylodynamic inference confirmed the rapid growth of the B.1.526 lineage in NYC, notably the sub-clade defined by the spike mutation E484K, which has outpaced the growth of other variants in NYC. Pseudovirus neutralization experiments demonstrated that B.1.526 spike mutations adversely affect the neutralization titer of convalescent and vaccinee plasma, indicating the public health importance of this lineage.

Introduction

After the early months of the SARS-CoV-2 pandemic in 2020, the vast majority of sequenced genomes contained the spike mutation D614G (along with 3 separate nucleotide changes)[1]. Following a period of gradual change, the fourth quarter of 2020 witnessed the emergence of several variants containing multiple mutations, many within the spike gene[2–5]. Multiple lines of evidence support escape from antibody selective pressure as a driving force for the development of these variants[6–9].

Genomic surveillance of SARS-CoV-2 is now focused on monitoring the emergence of these variants and the functional impact that their mutations may have on the effectiveness of passive antibody therapies and the efficacy of vaccines to prevent mild or moderate COVID-19. While an increasing number of specimens are being sequenced, analysis of these genomes remains a challenge[10]. Here, we developed a simple and fast utility that permits rapid inspection of the mutational landscape revealed by genomic surveillance of SARS-CoV-2: Variant Database (**vdb**). With this tool, we uncovered several groups of recently sequenced genomes with mutations at critical antibody epitopes. Among this group is a new lineage emerging in NYC that has increased in frequency to now account for ~32% of sequenced genomes as of February 2021. We confirm the rapid spread of B.1.526 in NYC during early 2021 through phylodynamic inference. Furthermore, we evaluated the impact of the B.1.526 spike mutations on the neutralization titer of convalescent and vaccinee plasma.

Results

vdb

Phylogenetic analysis is critical to understand the relationships of viral genomes. However, other perspectives can be useful for detecting patterns in large numbers of sequences. We developed **vdb** as a utility to query the sets of spike mutations observed during genomic surveillance. Using the **vdb** tool to analyze SARS-CoV-2 sequences in the Global Initiative on Sharing Avian Influenza Data (GISAID) dataset[11,12], we detected several clusters of sequences distinct from variants B.1.1.7, B.1.351, B.1.1.248, and B.1.429[2–5] with spike mutations at sites known to be associated with resistance to antibodies against SARS-CoV-2[8,13] (**Table 1**). The **vdb** program can find clusters of virus sharing identical sets of spike mutations, and then these patterns can be used to find potentially related sequences.

Defining mutations of B.1.526

One notable cluster of genome sequences was collected from the New York region and represents a distinct lineage, now designated as B.1.526 (**Figure 1, Supplementary Figure 1**). This variant is found within the 20.C clade and is distinguished by 3 defining spike mutations: L5F, T95I, and D253G. Within B.1.526, the largest sub-clade is defined by E484K and two distinct sub-clades are each defined by S477N; both of these mutations located within the receptor-binding domain (RBD) of spike (**Figure 2 and Supplementary Table 1**). We note that the evolutionary history at spike position 701 varies depending on whether the tree is rooted using a molecular clock (**Figure 1**) versus its sister clade (characterized by an L452R mutation; **Supplementary Figure 2**), the latter of which posits a substitution A701V followed by a reversion V701A. Among the nucleotide mutations in lineage B.1.526, the most characteristic include A16500C (NSP13 Q88H), A22320G (spike D253G), and T9867C (NSP4_L438P). Another

notable feature of the B.1.526 lineage is the deletion of nucleotides 11288-11296 (NSP6 106-108), which also occurs in variants B.1.1.7, B.1.351, P.1, and B.1.525[14].

Regarding four of the spike mutations prevalent in this lineage: (1) E484K is known to attenuate neutralization of multiple anti-SARS-CoV-2 antibodies, particularly those found in class 2 anti-RBD neutralizing antibodies[13,15], and is also present in variants B.1.351[4] and P.1/B.1.1.248[2], (2) D253G has been reported as an escape mutation from antibodies against the N-terminal domain[16], (3) S477N has been identified in several earlier lineages[17], is near the epitopes of multiple antibodies[18], and has been implicated to increase viral infectivity through enhanced interactions with ACE2[19,20], and (4) A701V sits adjacent to the S2' cleavage site of the neighboring protomer and is shared with variant B.1.351[4]. The overall pattern of mutations in lineage B.1.526 (**Figure 2**) suggests that it arose in part in response to selective pressure from antibodies. Based on the dates of collection of these viruses, it appears that the frequency of this lineage has increased rapidly in New York (**Table 2**).

Trends in B.1.526 surveillance

As part of public health surveillance conducted by the New York City Public Health Laboratory (NYC PHL) and the Pandemic Response Lab (PRL) in New York, approximately 4.5 thousand SARS-CoV-2 genomes have been sequenced by NYC PHL and PRL from December 1, 2020 to February 28th, 2021. Of these genomes, approximately 25% are from lineage B.1.526. We separately analyzed these genomes, because viral genomic surveillance by PHL and PRL provides a less biased picture of viral diversity in NYC than genomes uploaded to GISAID. The proportion of B.1.526 genomes in NYC has steadily increased since this variant was first detected in NYC surveillance data in late 2020, and its weekly average exceeded 10% by 14

January 2021. From early January to early March, B.1.526 has been increasing by about 0.7% per day (segmented linear regression) and was at 43% the week prior to 03 March 2021 **(Figure 3A**). Around 54% (n=678) of the B.1.526 genomes contain the E484K mutation, which has also been rising in frequency since early 2021. The weekly average of B.1.526 genomes with E484K has been above 10% since 01 February 2021 and has been increasing around 0.4% per day (**Figure 3B**).

This increase in B.1.526 temporally coincides with the peak and subsequent decline of the second epidemic wave in NYC (**Figure 3C**). If we separate the approximated number of B.1.526 cases from the rest of second wave SARS-CoV-2, the non-B.1.526 virus has steadily declined since its peak in early January 2021. However, the increasing proportion of B.1.526 appears to have slowed the rate of decline in total COVID-19 case counts in NYC.

Geographic distribution of B.1.526 in NYC

The New York City Public Health Laboratory and the PRL in New York have sequenced 4538 SARS-CoV-2 genomes from December 2020 thru February 2021 (**Figure 4A**). Geographic case distribution of specimens received at PHL and PRL for SARS-CoV-2 diagnostic nucleic acid amplification testing (NAAT) are representative of citywide testing efforts. Those SARS-CoV-2 positive specimens with NAAT cross-threshold values below 32 were selected at random to be sequenced. On a month-to-month basis using data generated by NYC PHL and PRL, we have observed an increasing number of B.1.526 genomes identified throughout NYC. The geographic distribution of over 600 B.1.526 E484K cases is similar (**Figure 4B**). While the B.1.526 lineage is

not limited to NYC, almost 90% of genomes deposited to GISAID prior to March 2021, are from the New York region.

Phylodynamic analysis

Other SARS-CoV-2 variants of concern or interest (B.1.1.7, B.1.427, and B.1.429) have also been circulating in NYC contemporaneously with the rise of B.1.526 and have all risen in relative frequency during the second wave of the NYC pandemic (**Figure 3D**). To compare the relative growth rates of these variants during this time-period, we fitted an exponential population growth model[21] implemented in BEAST1.10[22] to the sequences that correspond to these lineages of interest. Specifically, we estimated the growth rate for the B.1.1.7, B.1.427, and B.1.429 variants and for two subsets of the B.1.526 clade sequences (with and without the E484K mutation).

The B.1.526 E484K clade experienced more rapid exponential growth compared with other lineages: 23.2 (95% highest posterior density [HPD]: 19.6–27.1). B.1.526 with E484 and B.1.1.7 experienced similar growth rates: 14.3 (95% HPD: 11.7–16.9) and 14.5 (95% HPD 11.6 – 17.8), respectively. The B.1.427 and B.1.429 lineages experienced lower growth rates that were significantly greater than zero: 3.8 (95% HPD: 0.7–7.0) and 5.2 (95% HPD: 2.1–8.3), respectively. We caution that these lineage growth rates do not distinguish between per-contact transmissibility or per-virion infectiousness and speak only to the relative number of people detected with these variants in NYC during late 2020 and early 2021.

As part of the phylodynamic analysis, we inferred the time of most recent common ancestor (TMRCA) for the B.1.526 E484K clade to be 08 November 2020 (95% HPD: 22 October – 24 November). The TMRCA for the rest of the B.1.526 clade was estimated to be 15 September 2020 (95% HPD: 17 August – 08 October).

Neutralization activity of convalescent and vaccinee plasma against B.1.526

The identification of several mutations associated with resistance to anti-SARS-CoV-2 antibodies in B.1.526 sequences raises the question of the impact on SARS-CoV-2 immunity. We generated HIV-based pseudoviruses expressing SARS-CoV-2 spike protein containing either the most common B.1.526 mutation pattern (v.1: L5F, T95I, D253G, E484K, D614G, and A701V), the 2nd most common pattern (v.2: L5F, T95I, D253G, S477N, D614G, and Q957R), or only D614G. Pseudovirus neutralization titers were determined for human plasma samples from vaccinees [Moderna (mRNA-1273) or Pfizer-BioNTech(BNT162b2)][8] or convalescent plasma [at either 1.3[15] or 6.2 months[13] post-infection]. The E484K-containing B.1.526 pseudovirus had a statistically significant reduced neutralization titer compared to the D614G control: for vaccinee plasma, 4.5-fold reduced (p = 0.00005); for 1.3-month convalescent plasma, 6.0-fold reduced (p = 0.03); and for 6.2-month convalescent plasma, 4.8-fold reduced (p = 0.02) (**Figure 5a and Supplementary Table 2**). The smaller reduction of the titers in the 6.2-month convalescent plasma samples compared to the 1.3-month samples is consistent with the greater resistance of more matured anti-SARS-CoV-2 antibodies to viral escape mutations[23]. The S477N/Q957R-containing B.1.526 pseudovirus demonstrated a smaller effect on plasma neutralization (**Figure 5b**).

Discussion

Genomic surveillance is a critical tool to monitor the progression of the COVID-19 pandemic and modelling suggests that sequencing at least 5% of specimens that test positive for SARS-Cov-2 in a geographic region is necessary to reliably detect the emergence of novel variants at a lower prevalence limit of between 0.1% to 1%[24]. Through the combination of increased sequencing efforts and the use of the software utility described here, we were able to identify the B.1.526 lineage and to begin to characterize its phylogenetic and phylodynamic patterns in NYC in early 2021. Based on sequences in GISAID as of March 2021, the majority of cases with sequence data are in the NYC region, but it is expected that the prevalence B.1.526 variants will continue to increase beyond the NYC region. The B.1.526 variant has also been described in other recent studies[25,26].

Pseudovirus containing spike gene mutations associated with B.1.526 was significantly more resistant to neutralization by either convalescent or vaccinee plasma. The presence of E484K mutation likely plays a key role in facilitating increased viral transmission and reducing antibody neutralizing titers, as previously shown in other studies[7,27]. Continued monitoring for emerging variants with mutations such as E484K is important to maximize the impact of public health measures to mitigate the effects of the SARS-CoV-2 pandemic. For example, high frequencies of SARS-CoV-2 variants has potential impacts on selection of appropriate antibody therapeutics and vaccination strategies.

Methods

Variant Database Program

We developed a software tool named VDB (Variant Database). This tool consists of two Unix command line utilities: (1) **vdb**, a program for examining spike mutation patterns in a collection of sequenced viral genomes, and (2) **vdbCreate**, a program for generating a list of viral spike mutations from a multiple sequence alignment for use by **vdb**. The design goal for the query program **vdb** is to provide a fast, lightweight, and natural means to examine the landscape of SARS-CoV-2 spike mutations. These programs are written in Swift and are available for MacOS and Linux from the authors or from the Github repository: https://github.com/variant-database/vdb.

The **vdb** program implements a mutation pattern query language (see Supplemental Method) as a command shell. The first-class objects in this environment are a collection of viruses (a "cluster") and a group of spike mutations (a "pattern"). These objects can be assigned to variables and are the return types of various commands. Generally, clusters can be obtained from searches for patterns, and patterns can be found by examining a given cluster. Clusters can be filtered by geographical location, collection date, mutation count, or the presence or absence of a mutation pattern. The geographic or temporal distribution of clusters can be listed.

Results presented here are based on a multiple sequence alignment from GISAID[11,12] downloaded on February 10, 2021. Additional sequences downloaded from GISAID on February 22, 2021, were aligned with MAFFT v7.464[28].

Initial Phylogenetic Analysis

Multiple sequence alignments were performed with MAFFT v7.464[28]. The phylogenetic tree was calculated by IQ-TREE[29], and the tree diagram was generated using iTOL (Interactive Tree of Life)[30]. The Pango lineage nomenclature system[31] provides systematic names for SARS-CoV-2 lineages. The Pango lineage designation for B.1.526 was supported by the phylogenetic tree shown in **Supplementary Figure 1**.

Library preparation and sequencing

RNA was extracted from positive specimens collected at NYC PHL using the EZ1 (Qiagen, CA), NUCLISENS® easyMAG® (bioMérieux Inc., Netherlands), or Kingfisher™ Flex Purification System (Thermo Fisher Scientific, MA). RNA extracts were subjected to annealing reaction with random hexamers and dNTPs (New England Biolabs Inc., NEB, MA), and reverse transcribed with SuperScript IV Reverse Transcriptase at 42ºC for 50 min. The resulting cDNA was amplified using two separate multiplex PCRs with ARTIC V3 primer pools (Integrated DNA Technologies, IA) per sample in the presence of Q5 2X Hot Start Master Mix (NEB) at 98ºC for 30 secs, followed by 35 cycles of 98ºC for 15 secs and 65ºC for 5 min[32,33]. The resulting PCR products per sample were combined and purified using Agencourt Ampure XP magnetic beads (Beckman Coulter, IN), at a ratio of 1:1 sample to bead ratio and quantified using a Qubit 3.0 fluorometer (Thermo Fisher Scientific, MA). The PCR products were normalized to 90 ng as input for the NEBNext Ultra II Library Preparation Kit according to standard protocol (NEB): Briefly, the ARTIC PCR products were subjected to simultaneous end-repair, 5'-phosphorylation, and dA-tailing reaction at 20ºC for 30 min, followed by heat inactivation at 65ºC for 30 min. NEBNext Adaptor was then ligated at 25º for 30 min, and then cleaved by USER Enzyme at 37ºC for 15 min. This product was subjected to bead cleanup at a ratio of 0.6x sample to bed ratio. The eluted

product was amplified for 6 cycles using NEBNext Ultra II Q5 Master Mix in the presence of NEBNext Multiplex Oligos for Illumina (NEB). The PCR product was purified with Ampure XP beads at a 0.6x sample to bead ratio. The product was a barcoded library containing Illumina P5 and P7 adapters for sequencing on Illumina instruments. The individual libraries were quantified, normalized and pooled at equimolar concentration and loaded onto the Illumina MiSeq sequencing instrument using V3 600-cycle reagent kits and a V3 flow cell for 250-cycle paired end sequencing (Illumina, CA).

Genome Assembly

All raw paired end sequence reads are trimmed using Trim Galore version 0.6.4_dev[34] removing NEB adapters and quality score below 20 from ends of the reads. The trimmed reads were assembled using the Burrows-Wheeler Aligner MEM algorithm (BWA-MEM) version 0.7.12[35] with SARS-CoV-2 Wuhan-Hu-1 (GenBank accession number MN908947.3) as the reference sequence. Intrahost variant analysis of replicates (iVar)[36] tool was used to remove primer sequences from the amplicon-based sequencing data. Finally, the mutation calls and consensus genome were built using a combination of samtools mpileup[37] and iVar consensus, with a minimum quality score of 20, frequency threshold of 0.6, and minimum depth of 15 to optimize high quality variant calls. A sequence mapping quality control tool developed in-house was used to assess depth of coverage across all sequences, percent of ambiguous bases in the consensus genome and percent sequence mapped to the reference genome. Consensus genome with more than 3% ambiguous bases or less than 95% reference mapped were excluded from any further analyses.

Library preparation and sequencing (PRL)

Positive RNA specimens between cycle threshold of 15-30 were selected from all samples tested at Pandemic Response Labs, NYC and cDNA for each specimen was generated using LunaScript RT SuperMix (NEB, MA) according to manufacturer protocol. To target SARS-CoV-2 specifically, cDNA for each specimen was amplified in two separate pools, 28- and 30-plex respectively, to generate 1200bp of overlapping amplicons[38] using Q5 2x Hot-Start Master Mix (NEB, MA). The resulting pools are combined in equal volume and enriched for full length 1200 bp product using a SPRI-based magnetic bead cleanup. Enriched amplicons are tagmented (Illumina, CA) and barcoded (IDT, IA) and paired-end sequenced on an Illumina MiSeq or NextSeq 550.

Genome Assembly (PRL)

For each specimen, sequencing adapters are first trimmed using Trim Galore v0.6.6[34], then aligned to the SARS-CoV-2 Wuhan-Hu-1 reference genome (NCBI Nucleotide NC_045512.2) using BWA MEM 0.7.17-r1188[35]. Reads that are unmapped or those that have secondary alignments are discarded from the alignment. Consensus and mutations were called using samtools[37] and Intrahost variant analysis of replicates (iVar)[36] with a minimum quality score of 20, frequency threshold of 0.6 and a minimum read depth of 10x coverage. A consensus genome with ≥ 90% breath-of-coverage with ≤ 3000 ambiguous bases is considered a successful reconstruction (as per APHL recommendation).

Genome alignment

Complete genome sequences produced by the NYC PHL and the PRL with reported collection dates on or before 04 March 2021 were analyzed. We restricted our analysis to genomes produced by public health surveillance to NYC to reduce bias due to geography or preferential sequencing of viral variants by academic institutions. Genomes were aligned to the Wuhan-Hu-1 reference genome (GenBank Accession MN908947) using mafft v7.475 (mafft --6merpair --keeplength --addfragments)[28]. Pango lineage designations[31] for variants were assigned using Pangolin v2.3.2[39].

Segmented regression analysis

To estimate the timing and approximate linear slope of increase in B.1.526 and the E484K clade prevalence, we employed a segmented regression analysis (segmented package in R).

Maximum likelihood phylogenetic inference

Maximum likelihood trees were inferred using IQTree2 for B.1.1.7, B.1.427, B.1.429, and B.1.526 genomes using a GTR+F+Γ_4 substitution model[40]. Minimum branch length of 1e-9 was enforced and an expanded NNI search (--allnni) was employed to improve topology search. Preliminary molecular clock analyses were performed in TreeTime v0.8.1 using a fixed substitution rate of 8×10^{-4} substitutions/site/year and a skyline coalescent model[41]. This analysis identified 34 genomes whose root-to-tip genetic distance were flagged as problematic and excluded from subsequent phylodynamic analyses. TreeTime was also used to root and perform ancestral state reconstruction for a tree inferred from the 258 B.1.526 genomes sampled by the NYC PHL used to display the history of spike mutations in B.1.526 (**Figure 1**).

Bayesian phylodynamic inference

We performed population growth rate inference in coalescence-based framework using an exponential growth model in BEAST 1.10.4[22]. We used a strict molecular clock model with the fixed substitution rate of $8x10^{-4}$ substitutions/site/year. We applied a GTR+F+Γ_4 substitution model and specified the following priors for the population growth model: OneOnX distribution prior for the population size parameter and Laplace distribution prior (mean = 0.0, scale = 1.0) for the growth rate prior. Markov chain Monte Carlo analyses were run for 100-300 million generations; the first 10% of samples were discarded as burn-in. Separate inference was performed for B.1.1.7 (n=354), B.1.427 (n=35), B.1.429 (n=69), B.1.526 E484 (n=569), and B.1.526 E484K (n=678). For the B.1.526 phylodynamic inference, we did not include two sequences most closely related to B.1.526 (hCoV-19/USA/NY-NYCPHL-001701/2020 and hCoV-19/USA/NY-NYCPHL-002542/2021).

Geocoding addresses

To identify areas with the highest density of B.1.526 sequenced genomes in NYC from December 2020 to March 2021, patient addresses were geocoded to be visualized on a map[42]. Geocoding was performed using the NYC DOHMH's Geoportal application. Once geocoded, a map representing the point locations of individuals with sequenced B.1.526 genomes was created in ArcMap (v. 10.6.1) and exported as a point feature class.

Point density method

Point density maps of individuals with B.1.526 sequenced genomes were created by using the point density tool in ArcMap. Point density calculates the density-per-unit area from point

features (individuals with a SARS-CoV-2 B.1.526 sequenced genome) that fall within a defined neighborhood by totaling the number of points that fall within the neighborhood divided by the neighborhood area. Density calculations result in the observed gradient patterns. The point density map parameters were 4000 ft radius from the center of 250 square foot cells. The symbology class for point density classification was set at equal intervals of 5.

Human plasma samples

Human plasma samples were among those collected in previously reported studies[8,13,15]. The study visits and blood draws were performed in compliance with all relevant ethical regulations and the protocol for human participants was approved by the Institutional Review Board (IRB) of the Rockefeller University (protocol #DRO-1006).

Pseudovirus neutralization by human plasma samples

Human plasma samples were assayed for neutralization activity against lentiviruses pseudotyped with SARS-CoV-2 spike containing a 21-amino acid cytoplasmic tail deletion and either D614G or mutations corresponding to lineage B.1.526 (L5F, T95I, D253G, E484K, D614G, and A701V). Pseudotyped lentiviruses were generated and neutralizations assays were conducted as previously described[43,44]. Briefly, lentiviral particles were produced by co-transfecting the gene encoding SARS-CoV-2 spike protein (D614G or B.1.526) and Env-deficient HIV backbone expressing Luciferase-IRES-ZsGreen. Plasma samples were heat inactivated at 56ºC for 1 hour, then 3-fold serial diluted and incubated with SARS-CoV-2 pseudotyped virus for 1 hour at 37ºC. The virus/plasma mixture was added to $293T_{ACE2}$ target cells, which were seeded the previous day on poly-L-lysine coated plates. After incubating for 48 hours at 37ºC,

target cells were lysed with Britelite Plus (Perkin Elmer) and luciferase activity was measured as relative luminesce units (RLUs) and normalized to values derived from cells infected with pseudotyped virus in the absence of plasma. Data were fit to 2-parameter non-linear regression in Antibody database[45].

Data availability

The data analyzed as part of this project were obtained from the GISAID database and through a Data Use Agreement between NYC DOHMH and the University of California San Diego. Sequences analyzed by using the **vdb** tool were downloaded from GISAID. No personally identifying information were included as part of these analyses. SARS-CoV-2 genomes included in these analyses have been deposited in GISAID. See **Supplementary Data 1** for a list of genomes, including which genomes were excluded from the phylogenetic analysis.

Data for **Figure 5** are provided in **Supplementary Table 2**.

Code availability

The source code for the vdb program is available at the Github repository:

https://github.com/variant-database/vdb.

Acknowledgments

We thank the Global Initiative on Sharing Avian Influenza Data (GISAID) and the originating and submitting laboratories for sharing the SARS-CoV-2 genome sequences; see **Supplementary Table 3** for a list of sequence contributors. We thank Andrew Rambaut and Áine O'Toole for lineage designation. This work was supported by the Caltech Merkin Institute for Translational

Research (P.J.B.) and the Bill and Melinda Gates Foundation Collaboration for AIDS Vaccine Discovery (CAVD) (INV-002143). J.O.W. acknowledges funding from the National Institutes of Health (AI135992 and AI136056). T.I.V. is funded by a Branco Weiss Fellowship. M.C.N. is an HHMI Investigator.

Author Contributions

A.P.W., J.O.W., J.L.H., T.I.V., H.H.L., S.H. and J.C.W. analyzed data. J.C.W., M.A.C., E.G. and H.H.L. performed genome sequencing and assembly. J.O.W. curated data. C.G., M. Caskey and M.C.N. provided clinical samples. P.N.P.G. and J.R.K. carried out experiments. A.P.W., C.O.B., Z.Y., S.H., S.S.D., C.E.F. and J.O.W. prepared figures. A.P.W, J.O.W., T.I.V., C.O.B., J.C.W. and S.H. wrote the manuscript with input from all co-authors. A.P.W., P.J.B., J.O.W., J.L.R. and S.H. supervised the study.

Competing Interests

P.J.B. is a co-inventor on a provisional application from the California Institute of Technology for the use of mosaic nanoparticles as coronavirus immunogens. M.C.N., P.J.B., and C.O.B. are co-inventors on provisional applications for several anti-SARS-CoV-2 monoclonal antibodies. J.O.W. has received funding from Gilead Sciences, LLC (completed) and the CDC (ongoing) via grants and contracts to his institution unrelated to this research.

Tables

Table 1

Mutation patterns of viruses with mutations at select Spike positions, excluding viruses related to variants B.1.1.7, B.1.351, B.1.1.248, and B.1.429. Mutations included in this analysis were E484K, N501Y, K417T, K417N, L452R, and A701V. In this table viruses are only included if their spike mutation pattern exactly matches the given pattern. Note about P681H/P681R: variant B.1.1.7 has P681H. Note about W152L: variant B.1.429 has W152C

Pattern	Number of genomes	Top Locations	First collection date
L5F T95I D253G E484K D614G A701V	**243**	**US(240; NY 235)**	**12/16/2020**
E484K D614G V1176F	235	Brazil(132), US(40)	4/15/2020
W152L E484K D614G G769V	49	US(32)	11/1/2020
E484K D614G P681H	37	US(37; MD 27)	11/18/2020
R102I F157L V367F E484K Q613H P681R	36	England(35)	12/27/2020
Q52R A67V H69-V70- Y144- E484K D614G Q677H F888L	36	England(22)	12/15/2020

Table 2

Counts of virus genomes in lineage B.1.526 by month in New York State. The total number of sequenced genomes examined from GISAID from New York during these time periods is also listed. *Latest viral collection date was March 4, 2021. Note that geographic sampling may have varied over time as genome sequencing increased.

Viruses containing spike mutations T95I and D253G (earliest collection date Nov. 23, 2020)

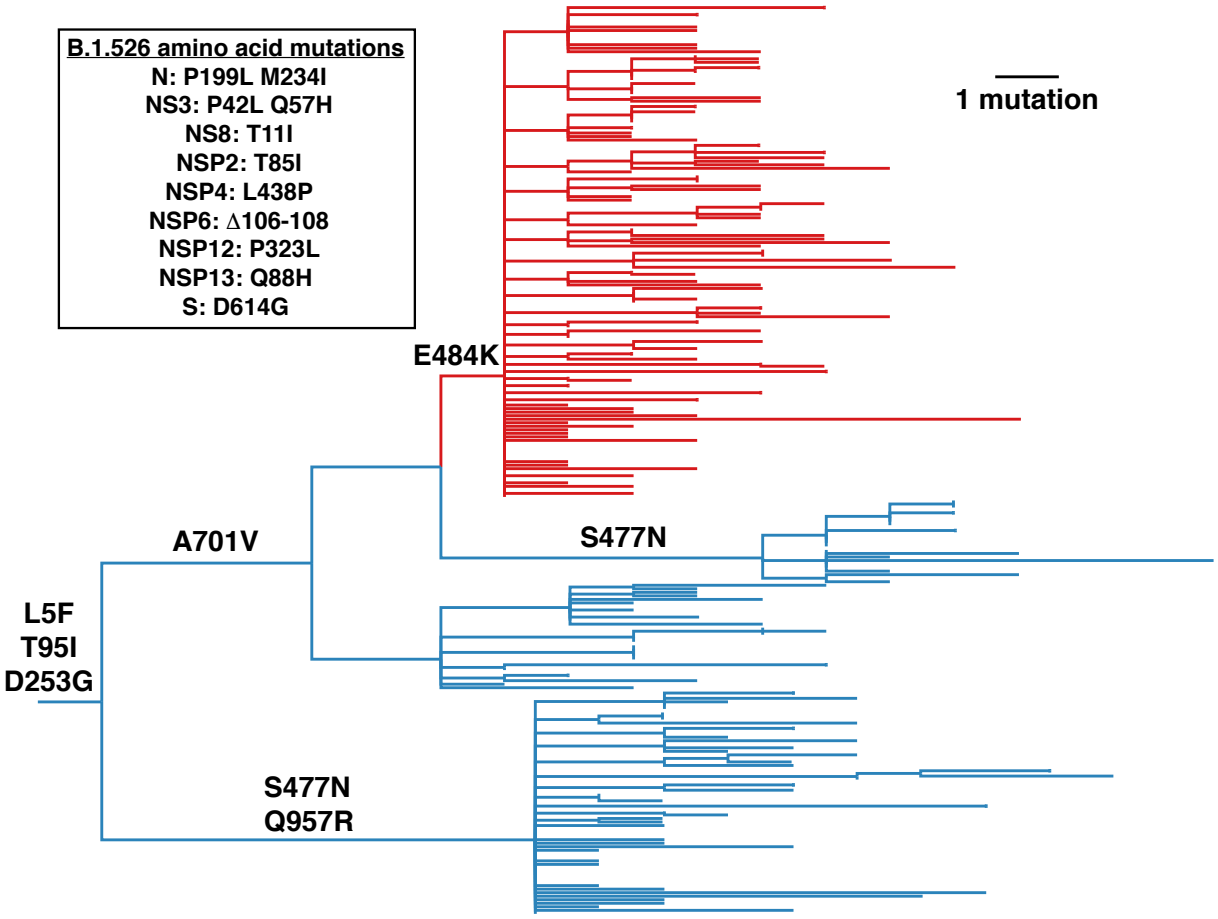
Month	count	total sequences	fraction
Nov. 2020	2	524	0.4%
Dec. 2020	46	2209	2.1%
Jan. 2021	201	3148	6.4%
Feb. 2021	1207	3868	31.2%
March 2021*	124	274	45.3%

Viruses containing spike mutations L5F, T95I, D253G, E484K, D614G, and A701V (earliest collection date Dec. 16, 2020)

Month	count	total sequences	fraction
Nov. 2020	0		
Dec. 2020	25	2209	1.1%
Jan. 2021	109	3148	3.5%
Feb. 2021	628	3868	16.2%
March 2021*	61	274	22.3%

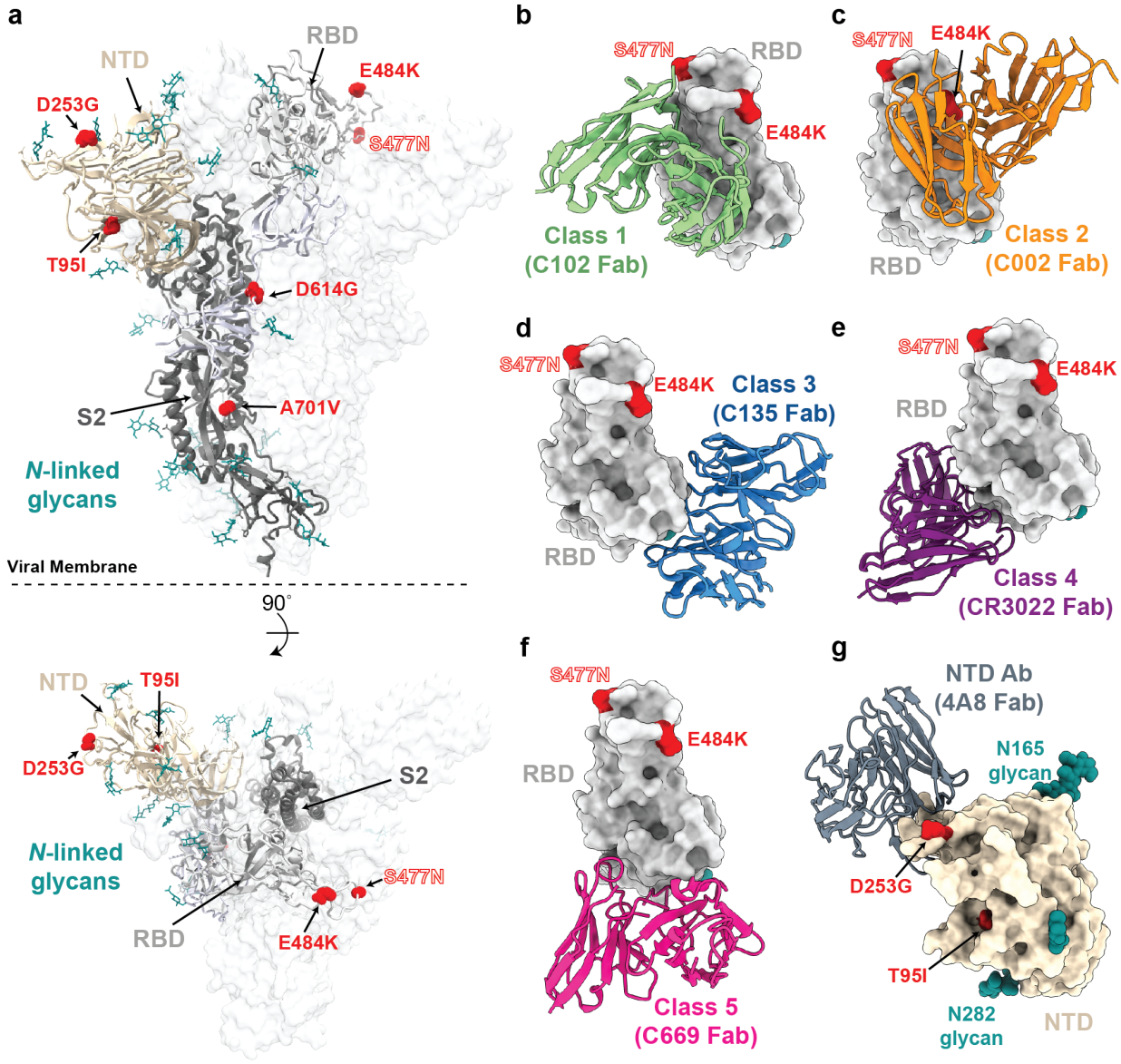
Figure 1.

Phylogenetic tree of lineage B.1.526 indicating spike mutations. Maximum likelihood phylogeny of SARS-CoV-2 variant B.1.526 sampled by NYC PHL (n=258). Amino acid substitutions in the spike protein occurring on internal branches are labeled, including the three spike mutations characteristic of B.1.526. The B.1.526 clade defined by the E484K mutation is highlighted in red. Inset highlights non-spike amino acid substitutions and deletions differentiating the B.1.526 clade from the Hu-1 reference genome. For display purposes, only NYC PHL genomes are shown.



426 Figure 2.

427 Structural locations of the spike mutations of lineage B.1.526.



428

429 **a**, Side and top views of the SARS-CoV-2 spike trimer (PDB 7JJI) with mutations of lineage

430 B.1.526 shown as spheres. **b-g**, Models of representative neutralizing antibodies (cartoon,

431 VH-VL domain only) bound to RBD (**b-f**, gray surface) or NTD (**g**, wheat surface). Sites for

432 B.1.526 lineage mutations are shown as red spheres. The S477N site is also shown for the

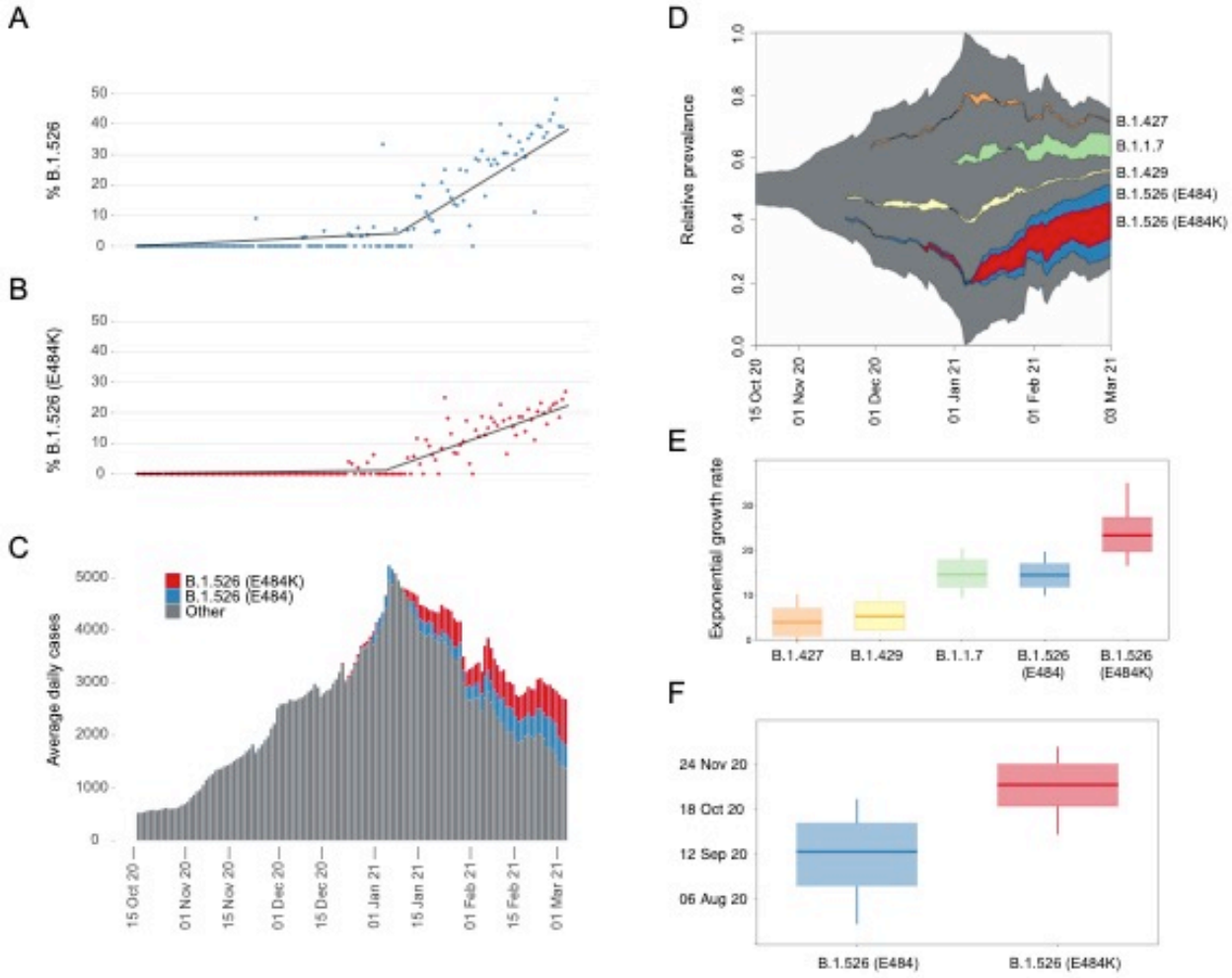
433 branch containing this mutation instead of the E484K mutation (see **Figure 1**); **b**, Class 1 (PDB

434 7K8M); **c**, Class 2 (PDB 7K8S); **d**, Class 3 (PDB 7K8Z); **e**, Class 4 (PDB 6W41); **f**, Class 5 [8] ; **g**, NTD-

435 specific antibody 4A8 (PDB 7C2L).

436

437 Figure 3.



438

439 Rise of SARS-CoV-2 variants in New York City (NYC) in late-2020 and early 2021. (**A**) Relative
440 frequency of B.1.526. Segmented linear regression is shown as a solid black line. (**B**) Relative
441 frequency of B.1.526 with E484K mutation. Segmented linear regression is shown as a dashed
442 gray line. (**C**) Rolling average number of total daily COVID-19 cases in NYC through time. Color
443 indicates the estimated proportions of B.1.526 (blue) and B.1.526 E484K (red) extrapolated
444 from a 7-day rolling average with an average of n=236 genomes sampled per week during this
445 time period. (**D**) Muller plot depicting sampling, with pseudocounts, of SARS-CoV-2 variants
446 scaled to the rolling average of total daily COVID-19 case counts. (**E**) Inferred exponential

447 growth rates for SARS-CoV-2 variants in NYC; the horizontal line indicates the median growth

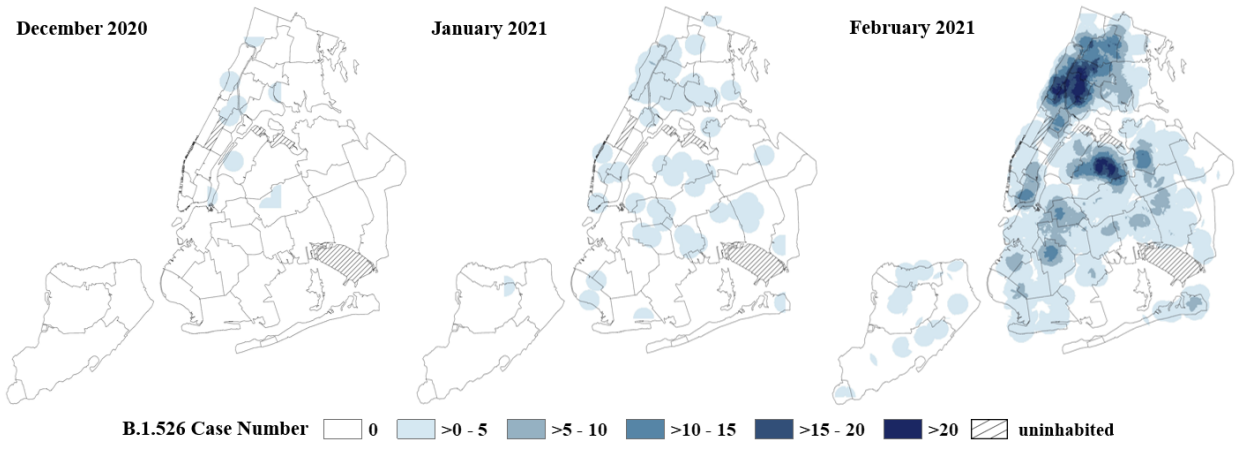
448 rate estimate, the box outlines the interquartile range. (**F**) Inferred time of most recent

449 common ancestor (TMRCA) estimates for B.1.526 (E484) and B.1.526 (E484K).

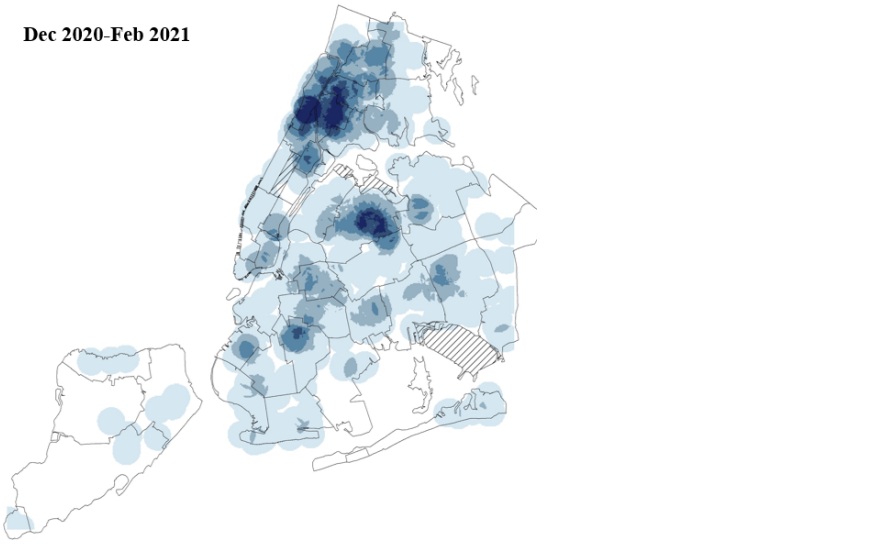
450

Figure 4.

A



B



(**A**) Spaciotemporal increase of B.1.526 lineage in New York City (NYC). Point density of B.1.526 variants geo-located by case address overlayed on a map of NYC delineated by United Hospital Fund areas. Data for each month is based on specimen collection date. The NYC PHL and the PRL in New York have sequenced 4538 SARS-CoV-2 genomes from December 2020 thru

461 February 2021. Data represents 11 B.1.526 variants out of 515 sequenced genomes in

462 December 2020, 80 B.1.526 variants out of 735 sequenced genomes in January and 1063

463 B.1.526 variants identified out of a total of 3288 sequenced genomes in February 2021. (**B**)

464 Distribution of B.1.526 E484K cases in NYC. Point density map of 608 B.1.526 E484K variant

465 cases in NYC. Data is based on specimen collection period from December 1, 2020 through

466 February 28th, 2021.

467

468 Figure 5.

469 Plasma neutralizing activity against pseudoviruses with B.1.526 lineage spike mutations. SARS-

470 CoV-2 pseudovirus neutralization assays were used to determine neutralization titer (NT50) for

471 COVID-19 vaccinee (n=10) and convalescent plasma at 1.3 months (n=10) and 6.2 months (n=9)

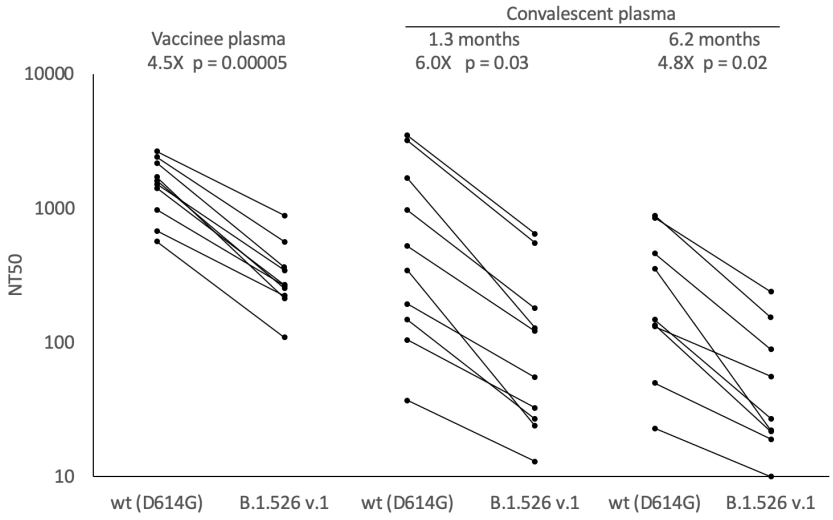
472 after infection. (**A**) Pseudovirus with spike mutations L5F, T95I, D253G, E484K, D614G

473 (B.1.526 v.1), and A701V, (**B**) Pseudovirus with spike mutations L5F, T95I, D253G, S477N,

474 D614G, and Q957R (B.1.526 v.2). Statistical significance was determined using paired two-tailed

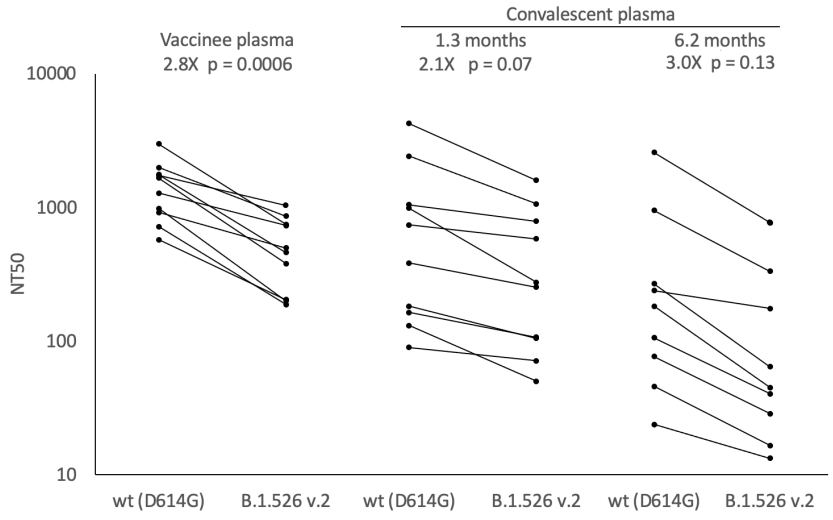
475 *t*-tests. Fold-differences of means are shown.

476 **A**



477

478 **B**



479

References

1. 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).
2. Faria, N. R. *et al.* Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *virological.org* https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (2021).
3. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *virological.org* 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).
4. Tegally, H. *et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa*. http://medrxiv.org/lookup/doi/10.1101/2020.12.21.20248640 (2020) doi:10.1101/2020.12.21.20248640.
5. Zhang, W. *et al.* Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA* (2021) doi:10.1001/jama.2021.1612.
6. Cele, S. *et al. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma*. http://medrxiv.org/lookup/doi/10.1101/2021.01.26.21250224 (2021) doi:10.1101/2021.01.26.21250224.
7. Greaney, A. J. *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe* **29**, 463-476.e6 (2021).

8. Wang, Z. *et al.* mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* (2021) doi:10.1038/s41586-021-03324-6.

9. Wibmer, C. K. *et al.* SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med* (2021) doi:10.1038/s41591-021-01285-x.

10. Hodcroft, E. B. *et al.* Want to track pandemic variants faster? Fix the bioinformatics bottleneck. *Nature* **591**, 30–33 (2021).

11. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health: Data, Disease and Diplomacy. *Global Challenges* **1**, 33–46 (2017).

12. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, (2017).

13. Gaebler, C. *et al.* Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639–644 (2021).

14. Martin, D. P. *et al. The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape*. http://medrxiv.org/lookup/doi/10.1101/2021.02.23.21252268 (2021) doi:10.1101/2021.02.23.21252268.

15. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).

16. McCallum, M. *et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2*. http://biorxiv.org/lookup/doi/10.1101/2021.01.14.426475 (2021) doi:10.1101/2021.01.14.426475.

17. Hodcroft, E. B. *et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020*. http://medrxiv.org/lookup/doi/10.1101/2020.10.25.20219063 (2020) doi:10.1101/2020.10.25.20219063.

18. Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).

19. Chen, J., Wang, R., Wang, M. & Wei, G.-W. Mutations Strengthened SARS-CoV-2 Infectivity. *Journal of Molecular Biology* **432**, 5212–5226 (2020).

20. Ou, J. *et al. Emergence of SARS-CoV-2 spike RBD mutants that enhance viral infectivity through increased human ACE2 receptor binding affinity*. http://biorxiv.org/lookup/doi/10.1101/2020.03.15.991844 (2020) doi:10.1101/2020.03.15.991844.

21. Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H. & Rambaut, A. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* **20**, 381–387 (2003).

22. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4**, vey016 (2018).

23. Muecksch, F. *et al. Development of potency, breadth and resilience to viral escape mutations in SARS-CoV-2 neutralizing antibodies*. http://biorxiv.org/lookup/doi/10.1101/2021.03.07.434227 (2021) doi:10.1101/2021.03.07.434227.

544 24. Vavrek, D. *et al. Genomic surveillance at scale is required to detect newly emerging strains*
545 *at an early timepoint*. http://medrxiv.org/lookup/doi/10.1101/2021.01.12.21249613 (2021)
546 doi:10.1101/2021.01.12.21249613.
547 25. Annavajhala, M. K. *et al. A Novel SARS-CoV-2 Variant of Concern, B.1.526, Identified in New*
548 *York*. http://medrxiv.org/lookup/doi/10.1101/2021.02.23.21252259 (2021)
549 doi:10.1101/2021.02.23.21252259.
550 26. Lasek-Nesselquist, E., Lapierre, P., Schneider, E., George, K. St. & Pata, J. *The localized rise of*
551 *a B.1.526 SARS-CoV-2 variant containing an E484K mutation in New York State*.
552 http://medrxiv.org/lookup/doi/10.1101/2021.02.26.21251868 (2021)
553 doi:10.1101/2021.02.26.21251868.
554 27. Wang, P. *et al.* Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *Nature*
555 (2021) doi:10.1038/s41586-021-03398-2.
556 28. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7:
557 Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780
558 (2013).
559 29. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective
560 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology*
561 *and Evolution* **32**, 268–274 (2015).
562 30. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree
563 display and annotation. *Bioinformatics* **23**, 127–128 (2007).
564 31. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist
565 genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).

566 32. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other
567 virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261–1276 (2017).
568 33. Tyson, J. R. *et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome*
569 *sequencing using nanopore*. http://biorxiv.org/lookup/doi/10.1101/2020.09.04.283077
570 (2020) doi:10.1101/2020.09.04.283077.
571 34. Krueger, F. *Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply*
572 *quality and adapter trimming to FastQ files*. (2015).
573 35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
574 *Bioinformatics* **25**, 1754–1760 (2009).
575 36. Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring
576 intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* **20**, 8 (2019).
577 37. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
578 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–
579 2993 (2011).
580 38. Freed, N. E., Vlková, M., Faisal, M. B. & Silander, O. K. Rapid and inexpensive whole-genome
581 sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid
582 Barcoding. *Biology Methods and Protocols* **5**, bpaa014 (2020).
583 39. O'Toole, Á., McCrone, J. T. & Scher, E. *Pangolin: lineage assignment in an emerging*
584 *pandemic as an epidemiological tool*. github.com/cov-lineages/pangolin (2020).
585 40. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference
586 in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

587 41. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic
588 analysis. *Virus Evol* **4**, vex042 (2018).
589 42. Wu, W. Y., Jiang, Q. & Di Lonardo, S. S. Poorly Controlled Diabetes in New York City:
590 Mapping High-Density Neighborhoods. *Journal of Public Health Management and Practice*
591 **24**, 69–74 (2018).
592 43. Cohen, A. A. *et al.* Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic
593 coronaviruses in mice. *Science* **371**, 735–741 (2021).
594 44. Crawford, K. H. D. *et al.* Protocol and Reagents for Pseudotyping Lentiviral Particles with
595 SARS-CoV-2 Spike Protein for Neutralization Assays. *Viruses* **12**, 513 (2020).
596 45. West, A. P. *et al.* Computational analysis of anti-HIV-1 antibody neutralization panel data to
597 identify potential functional epitope residues. *Proceedings of the National Academy of*
598 *Sciences* **110**, 10598–10603 (2013).
599
600

Supplementary Material

Supplementary Methods.

Commands for the program **vdb**, implementing a mutation pattern query language:

Notation

cluster = group of viruses < > = user input n = an integer
pattern = group of mutations [ ] = optional ( ) = explanation of command
"world" = all viruses in database -> result

To define a variable for a cluster or pattern: <name> = cluster or pattern
Set operations +, -, and * (intersection) can be applied to clusters or patterns
If no cluster is entered, all viruses will be used ("world")

Filter commands

<cluster> from <country or state> -> cluster
<cluster> containing [<n>] <pattern> -> cluster alias with (matches for >=n mutations)
<cluster> not containing <pattern> -> cluster alias without (considers whole pattern)
<cluster> before <date> -> cluster
<cluster> after <date> -> cluster
<cluster> > or < <n> -> cluster (filter by number of mutations)

Commands to find mutation patterns

consensus [for] <cluster or country or state> -> pattern
patterns [in] [<n>] <cluster> -> pattern (lists n patterns)

Listing commands

list [<n>] <cluster>
[list] countries [for] <cluster>
[list] states [for] <cluster>
[list] frequencies [for] <cluster> alias freq (frequency of individual mutations)
[list] monthly [for] <cluster> [<cluster2>] (number of viruses per month or week)
[list] weekly [for] <cluster> [<cluster2>] (as a fraction of number of viruses in cluster2)
[list] patterns (list built-in and user-defined patterns)
[list] clusters (list built-in and user-defined clusters)

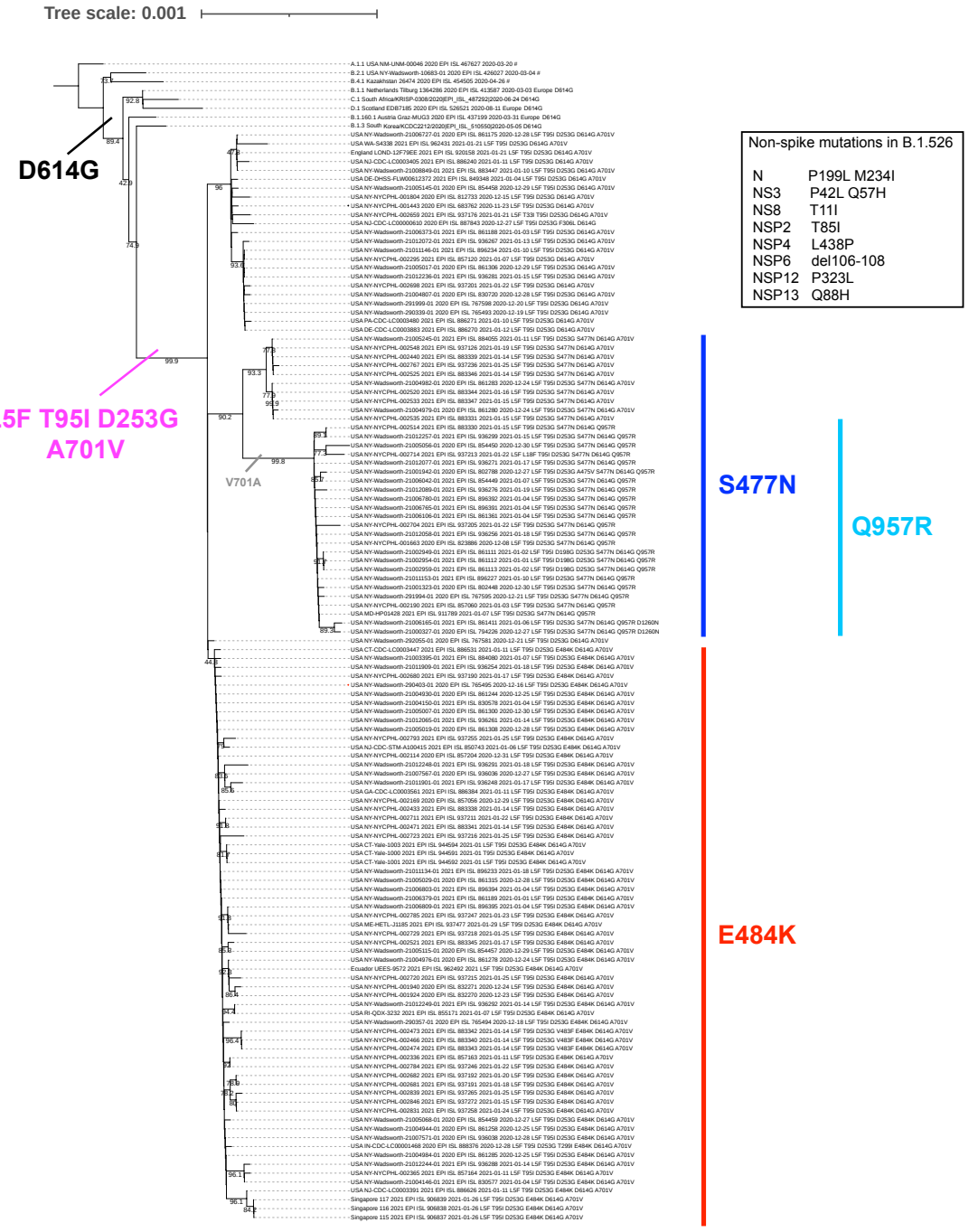
Other commands

sort <cluster> (by date)
help
history
quit

643 Supplementary Figure 1.

644 Phylogenetic tree of lineage B.1.526 indicating spike mutations. The inset lists non-spike

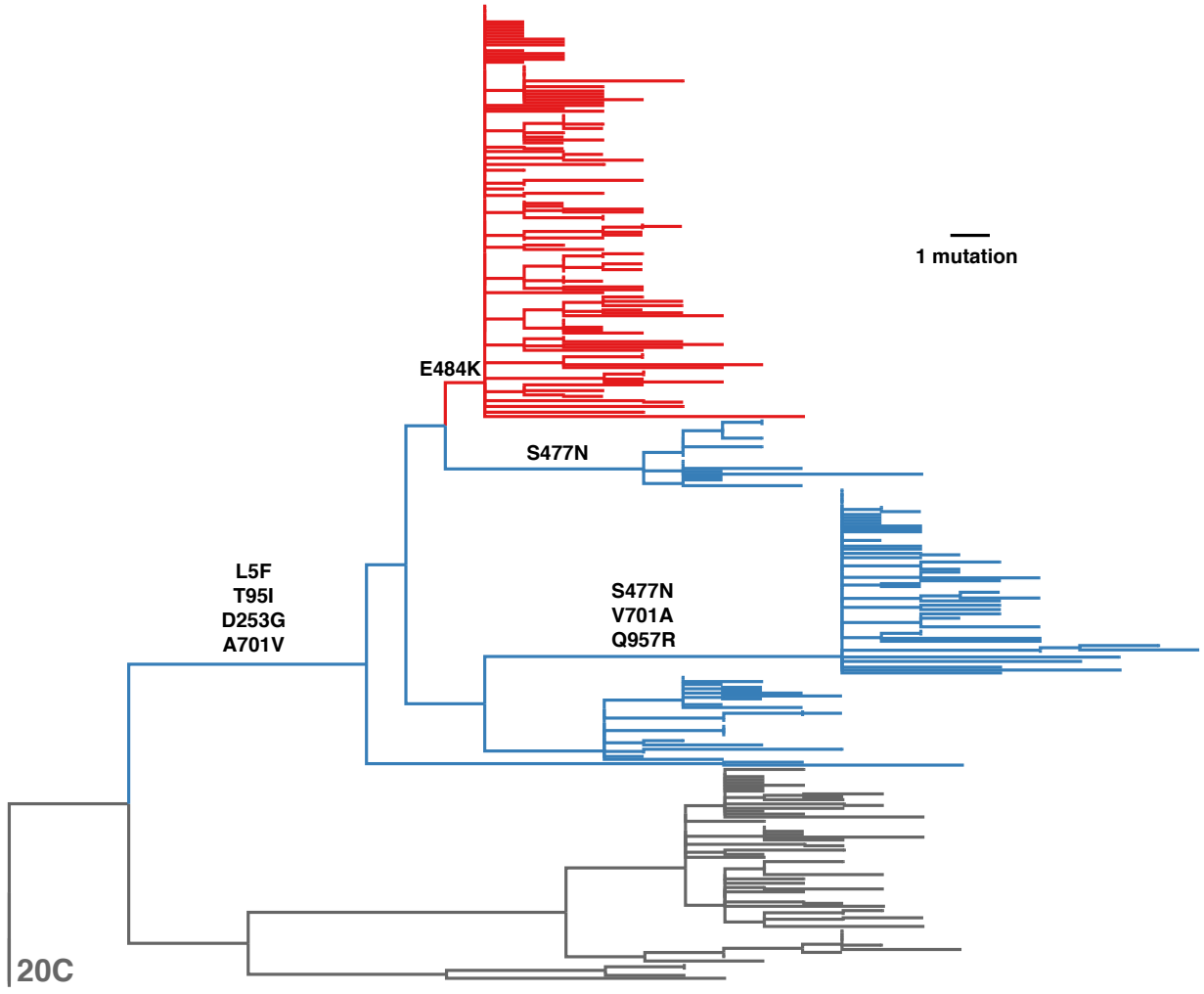
645 mutations common in this lineage.



646

Supplementary Figure 2.

Maximum likelihood phylogenetic tree of the B.1.526 lineage in relation to a sister clade defined by an L452R spike mutation and the 20C ancestral virus (both shown in gray). Tree was rooted using the clade 20C ancestral viruses. Amino acid substitutions in the spike protein occurring on internal branches are labeled leading to and within B.1.526 are labeled. The B.1.526 lineage is colored blue, except for the clade defined by the E484K mutation, which is highlighted in red. The most common pattern of spike mutations in the sister clade is D80G, ΔY144, F157S, L452R, D614G, T859N, and D950H.



Supplementary Table 1.

List of 124 viral genomes (with their accession number, location, collection date, and spike mutations) in lineage B.1.526. Mutations E484K, S477N, Q957R are highlighted in red, blue, and cyan, respectively.

EPI_ISL_683762, USA/NY-NYCPHL-001443/2020-11-23 : L5F T95I D253G D614G A701V
EPI_ISL_823886, USA/NY-NYCPHL-001663/2020-12-08 : L5F T95I D253G S477N D614G Q957R
EPI_ISL_812733, USA/NY-NYCPHL-001804/2020-12-15 : L5F T95I D253G D614G A701V
EPI_ISL_765495, USA/NY-Wadsworth-290403-01/2020-12-16 : L5F T95I D253G E484K D614G A701V
EPI_ISL_765494, USA/NY-Wadsworth-290357-01/2020-12-18 : L5F T95I D253G E484K D614G A701V
EPI_ISL_765493, USA/NY-Wadsworth-290339-01/2020-12-19 : L5F T95I D253G D614G A701V
EPI_ISL_767598, USA/NY-Wadsworth-291999-01/2020-12-20 : L5F T95I D253G D614G A701V
EPI_ISL_767581, USA/NY-Wadsworth-292055-01/2020-12-21 : L5F T95I D253G D614G A701V
EPI_ISL_767595, USA/NY-Wadsworth-291994-01/2020-12-21 : L5F T95I D253G S477N D614G Q957R
EPI_ISL_832270, USA/NY-NYCPHL-001924/2020-12-23 : L5F T95I D253G E484K D614G A701V
EPI_ISL_832271, USA/NY-NYCPHL-001940/2020-12-24 : L5F T95I D253G E484K D614G A701V
EPI_ISL_861278, USA/NY-Wadsworth-21004976-01/2020-12-24 : L5F T95I D253G E484K D614G A701V
EPI_ISL_861280, USA/NY-Wadsworth-21004979-01/2020-12-24 : L5F T95I D253G S477N D614G A701V
EPI_ISL_861283, USA/NY-Wadsworth-21004982-01/2020-12-24 : L5F T95I D253G S477N D614G A701V
EPI_ISL_861258, USA/NY-Wadsworth-21004944-01/2020-12-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_861285, USA/NY-Wadsworth-21004984-01/2020-12-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_861244, USA/NY-Wadsworth-21004930-01/2020-12-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_802788, USA/NY-Wadsworth-21001942-01/2020-12-27 : L5F T95I D253G A475V S477N D614G Q957R
EPI_ISL_794226, USA/NY-Wadsworth-21000327-01/2020-12-27 : L5F T95I D253G S477N D614G Q957R D1260N
EPI_ISL_936036, USA/NY-Wadsworth-21007567-01/2020-12-27 : L5F T95I D253G E484K D614G A701V
EPI_ISL_854459, USA/NY-Wadsworth-21005068-01/2020-12-27 : L5F T95I D253G E484K D614G A701V
EPI_ISL_887843, USA/NJ-CDC-LC00000610/2020-12-27 : L5F T95I D253G F306L D614G
EPI_ISL_861315, USA/NY-Wadsworth-21005029-01/2020-12-28 : L5F T95I D253G E484K D614G A701V
EPI_ISL_888376, USA/IN-CDC-LC00001468/2020-12-28 : L5F T95I D253G T299I E484K D614G A701V
EPI_ISL_830720, USA/NY-Wadsworth-21004807-01/2020-12-28 : L5F T95I D253G D614G A701V
EPI_ISL_861308, USA/NY-Wadsworth-21005019-01/2020-12-28 : L5F T95I D253G E484K D614G A701V
EPI_ISL_861175, USA/NY-Wadsworth-21006727-01/2020-12-28 : L5F T95I D253G D614G A701V
EPI_ISL_936038, USA/NY-Wadsworth-21007571-01/2020-12-28 : L5F T95I D253G E484K D614G A701V
EPI_ISL_854458, USA/NY-Wadsworth-21005145-01/2020-12-29 : L5F T95I D253G D614G A701V
EPI_ISL_854457, USA/NY-Wadsworth-21005115-01/2020-12-29 : L5F T95I D253G E484K D614G A701V
EPI_ISL_857056, USA/NY-NYCPHL-002169/2020-12-29 : L5F T95I D253G E484K D614G A701V
EPI_ISL_861306, USA/NY-Wadsworth-21005017-01/2020-12-29 : L5F T95I D253G D614G A701V
EPI_ISL_802448, USA/NY-Wadsworth-21001323-01/2020-12-30 : L5F T95I D253G S477N D614G Q957R
EPI_ISL_861300, USA/NY-Wadsworth-21005007-01/2020-12-30 : L5F T95I D253G E484K D614G A701V
EPI_ISL_854450, USA/NY-Wadsworth-21005056-01/2020-12-30 : L5F T95I D253G S477N D614G Q957R
EPI_ISL_857204, USA/NY-NYCPHL-002114/2020-12-31 : L5F T95I D253G E484K D614G A701V
EPI_ISL_861112, USA/NY-Wadsworth-21002954-01/2021-01-01 : L5F T95I D198G D253G S477N D614G Q957R
EPI_ISL_861189, USA/NY-Wadsworth-21006379-01/2021-01-01 : L5F T95I D253G E484K D614G A701V
EPI_ISL_944591, USA/CT-Yale-1000/2021-01-01 : T95I D253G E484K D614G A701V
EPI_ISL_944592, USA/CT-Yale-1001/2021-01-01 : L5F T95I D253G E484K D614G A701V
EPI_ISL_944594, USA/CT-Yale-1003/2021-01-01 : L5F T95I D253G E484K D614G A701V
EPI_ISL_962492, Ecuador/UEES-9572/2021-01-01 : L5F T95I D253G E484K D614G A701V
EPI_ISL_962493, Ecuador/UEES-9602/2021-01-01 : L5F T95I D253G E484K D614G A701V

706 EPI_ISL_861113, USA/NY-Wadsworth-21002959-01/2021-01-02 : L5F T95I D198G D253G S477N D614G Q957R
707 EPI_ISL_861111, USA/NY-Wadsworth-21002949-01/2021-01-02 : L5F T95I D198G D253G S477N D614G Q957R
708 EPI_ISL_857060, USA/NY-NYCPHL-002190/2021-01-03 : L5F T95I D253G S477N D614G Q957R
709 EPI_ISL_861188, USA/NY-Wadsworth-21006373-01/2021-01-03 : L5F T95I D253G D614G A701V
710 EPI_ISL_896394, USA/NY-Wadsworth-21006803-01/2021-01-04 : L5F T95I D253G E484K D614G A701V
711 EPI_ISL_849348, USA/DE-DHSS-FLW00612372/2021-01-04 : L5F T95I D253G D614G A701V
712 EPI_ISL_896391, USA/NY-Wadsworth-21006765-01/2021-01-04 : L5F T95I D253G S477N D614G Q957R
713 EPI_ISL_896392, USA/NY-Wadsworth-21006780-01/2021-01-04 : L5F T95I D253G S477N D614G Q957R
714 EPI_ISL_830577, USA/NY-Wadsworth-21004146-01/2021-01-04 : L5F T95I D253G E484K D614G A701V
715 EPI_ISL_896395, USA/NY-Wadsworth-21006809-01/2021-01-04 : L5F T95I D253G E484K D614G A701V
716 EPI_ISL_861361, USA/NY-Wadsworth-21006106-01/2021-01-04 : L5F T95I D253G S477N D614G Q957R
717 EPI_ISL_830578, USA/NY-Wadsworth-21004150-01/2021-01-04 : L5F T95I D253G E484K D614G A701V
718 EPI_ISL_861411, USA/NY-Wadsworth-21006165-01/2021-01-06 : L5F T95I D253G S477N D614G Q957R D1260N
719 EPI_ISL_850743, USA/NJ-CDC-STM-A100415/2021-01-06 : L5F T95I D253G E484K D614G A701V
720 EPI_ISL_854449, USA/NY-Wadsworth-21006042-01/2021-01-07 : L5F T95I D253G S477N D614G Q957R
721 EPI_ISL_911789, USA/MD-HP01428/2021-01-07 : L5F T95I D253G S477N D614G Q957R
722 EPI_ISL_857120, USA/NY-NYCPHL-002295/2021-01-07 : L5F T95I D253G D614G A701V
723 EPI_ISL_855171, USA/RI-QDX-3232/2021-01-07 : L5F T95I D253G E484K D614G A701V
724 EPI_ISL_884080, USA/NY-Wadsworth-21003395-01/2021-01-07 : L5F T95I D253G E484K D614G A701V
725 EPI_ISL_896234, USA/NY-Wadsworth-21011146-01/2021-01-10 : L5F T95I D253G D614G A701V
726 EPI_ISL_883447, USA/NY-Wadsworth-21008849-01/2021-01-10 : L5F T95I D253G D614G A701V
727 EPI_ISL_896227, USA/NY-Wadsworth-21011153-01/2021-01-10 : L5F T95I D253G S477N D614G Q957R
728 EPI_ISL_886271, USA/PA-CDC-LC0003480/2021-01-10 : L5F T95I D253G D614G A701V
729 EPI_ISL_886626, USA/NJ-CDC-LC0003391/2021-01-11 : L5F T95I D253G E484K D614G A701V
730 EPI_ISL_857164, USA/NY-NYCPHL-002365/2021-01-11 : L5F T95I D253G E484K D614G A701V
731 EPI_ISL_886531, USA/CT-CDC-LC0003447/2021-01-11 : L5F T95I D253G E484K D614G A701V
732 EPI_ISL_857163, USA/NY-NYCPHL-002336/2021-01-11 : L5F T95I D253G E484K D614G A701V
733 EPI_ISL_886240, USA/NJ-CDC-LC0003405/2021-01-11 : L5F T95I D253G D614G A701V
734 EPI_ISL_884055, USA/NY-Wadsworth-21005245-01/2021-01-11 : L5F T95I D253G S477N D614G A701V
735 EPI_ISL_886384, USA/GA-CDC-LC0003561/2021-01-11 : L5F T95I D253G E484K D614G A701V
736 EPI_ISL_886270, USA/DE-CDC-LC0003883/2021-01-12 : L5F T95I D253G D614G A701V
737 EPI_ISL_936267, USA/NY-Wadsworth-21012072-01/2021-01-13 : L5F T95I D253G D614G A701V
738 EPI_ISL_936292, USA/NY-Wadsworth-21012249-01/2021-01-14 : L5F T95I D253G E484K D614G A701V
739 EPI_ISL_936261, USA/NY-Wadsworth-21012065-01/2021-01-14 : L5F T95I D253G E484K D614G A701V
740 EPI_ISL_883338, USA/NY-NYCPHL-002433/2021-01-14 : L5F T95I D253G E484K D614G A701V
741 EPI_ISL_883342, USA/NY-NYCPHL-002473/2021-01-14 : L5F T95I D253G V483F E484K D614G A701V
742 EPI_ISL_883339, USA/NY-NYCPHL-002440/2021-01-14 : L5F T95I D253G S477N D614G A701V
743 EPI_ISL_883340, USA/NY-NYCPHL-002466/2021-01-14 : L5F T95I D253G V483F E484K D614G A701V
744 EPI_ISL_936288, USA/NY-Wadsworth-21012244-01/2021-01-14 : L5F T95I D253G E484K D614G A701V
745 EPI_ISL_883343, USA/NY-NYCPHL-002474/2021-01-14 : L5F T95I D253G V483F E484K D614G A701V
746 EPI_ISL_883346, USA/NY-NYCPHL-002525/2021-01-14 : L5F T95I D253G S477N D614G A701V
747 EPI_ISL_883341, USA/NY-NYCPHL-002471/2021-01-14 : L5F T95I D253G E484K D614G A701V
748 EPI_ISL_883330, USA/NY-NYCPHL-002514/2021-01-15 : L5F T95I D253G S477N D614G Q957R
749 EPI_ISL_883347, USA/NY-NYCPHL-002533/2021-01-15 : L5F T95I D253G S477N D614G A701V
750 EPI_ISL_936299, USA/NY-Wadsworth-21012257-01/2021-01-15 : L5F T95I D253G S477N D614G Q957R
751 EPI_ISL_883331, USA/NY-NYCPHL-002535/2021-01-15 : L5F T95I D253G S477N D614G A701V
752 EPI_ISL_937272, USA/NY-NYCPHL-002846/2021-01-15 : L5F T95I D253G E484K D614G A701V
753 EPI_ISL_936281, USA/NY-Wadsworth-21012236-01/2021-01-15 : L5F T95I D253G D614G A701V
754 EPI_ISL_883344, USA/NY-NYCPHL-002520/2021-01-16 : L5F T95I D253G S477N D614G A701V
755 EPI_ISL_937190, USA/NY-NYCPHL-002680/2021-01-17 : L5F T95I D253G E484K D614G A701V
756 EPI_ISL_883345, USA/NY-NYCPHL-002521/2021-01-17 : L5F T95I D253G E484K D614G A701V
757 EPI_ISL_936248, USA/NY-Wadsworth-21011901-01/2021-01-17 : L5F T95I D253G E484K D614G A701V
758 EPI_ISL_936271, USA/NY-Wadsworth-21012077-01/2021-01-17 : L5F T95I D253G S477N D614G Q957R

EPI_ISL_936256, USA/NY-Wadsworth-21012058-01/2021-01-18 : L5F T95I D253G S477N D614G Q957R
EPI_ISL_936254, USA/NY-Wadsworth-21011909-01/2021-01-18 : L5F T95I D253G E484K D614G A701V
EPI_ISL_936291, USA/NY-Wadsworth-21012248-01/2021-01-18 : L5F T95I D253G E484K D614G A701V
EPI_ISL_896233, USA/NY-Wadsworth-21011134-01/2021-01-18 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937191, USA/NY-NYCPHL-002681/2021-01-18 : L5F T95I D253G E484K D614G A701V
EPI_ISL_936276, USA/NY-Wadsworth-21012089-01/2021-01-19 : L5F T95I D253G S477N D614G Q957R
EPI_ISL_937126, USA/NY-NYCPHL-002548/2021-01-19 : L5F T95I D253G S477N D614G A701V
EPI_ISL_937192, USA/NY-NYCPHL-002682/2021-01-20 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937176, USA/NY-NYCPHL-002659/2021-01-21 : L5F T33I T95I D253G D614G A701V
EPI_ISL_920158, England/LOND-12F79EE/2021-01-21 : L5F T95I D253G D614G A701V
EPI_ISL_962431, USA/WA-S4338/2021-01-21 : L5F T95I D253G D614G A701V
EPI_ISL_937205, USA/NY-NYCPHL-002704/2021-01-22 : L5F T95I D253G S477N D614G Q957R
EPI_ISL_937211, USA/NY-NYCPHL-002711/2021-01-22 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937201, USA/NY-NYCPHL-002698/2021-01-22 : L5F T95I D253G D614G A701V
EPI_ISL_937213, USA/NY-NYCPHL-002714/2021-01-22 : L5F L18F T95I D253G S477N D614G Q957R
EPI_ISL_937246, USA/NY-NYCPHL-002784/2021-01-22 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937247, USA/NY-NYCPHL-002785/2021-01-23 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937258, USA/NY-NYCPHL-002831/2021-01-24 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937236, USA/NY-NYCPHL-002767/2021-01-25 : L5F T95I D253G S477N D614G A701V
EPI_ISL_937218, USA/NY-NYCPHL-002729/2021-01-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937216, USA/NY-NYCPHL-002723/2021-01-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937265, USA/NY-NYCPHL-002839/2021-01-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937215, USA/NY-NYCPHL-002720/2021-01-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937255, USA/NY-NYCPHL-002793/2021-01-25 : L5F T95I D253G E484K D614G A701V
EPI_ISL_906839, Singapore/117/2021-01-26 : L5F T95I D253G E484K D614G A701V
EPI_ISL_906838, Singapore/116/2021-01-26 : L5F T95I D253G E484K D614G A701V
EPI_ISL_906837, Singapore/115/2021-01-26 : L5F T95I D253G E484K D614G A701V
EPI_ISL_937477, USA/ME-HETL-J1185/2021-01-29 : L5F T95I D253G E484K D614G A701V