

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

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

<sup>1</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

<sup>2</sup>Department of Medicine, University of California San Diego, La Jolla, CA 92093

<sup>3</sup>New York City Public Health Laboratory, New York City Department of Health and Mental Hygiene, New York, NY, 10016 USA

<sup>4</sup>Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA 92093

<sup>5</sup>Pandemic Response Laboratory, Long Island City, NY 11101

<sup>6</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115

<sup>7</sup>Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA.

<sup>8</sup>Howard Hughes Medical Institute, The Rockefeller University, New York, NY, 10065 USA.

\*Corresponding author: Anthony P. West, Jr., [apwest@caltech.edu](mailto:apwest@caltech.edu)

## 1 Abstract

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

17

## 18 Introduction

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

25  
26 Genomic surveillance of SARS-CoV-2 is now focused on monitoring the emergence of these  
27 variants and the functional impact that their mutations may have on the effectiveness of  
28 passive antibody therapies and the efficacy of vaccines to prevent mild or moderate COVID-19.  
29 While an increasing number of specimens are being sequenced, analysis of these genomes  
30 remains a challenge<sup>10</sup>. Here, we developed a simple and fast utility that permits rapid  
31 inspection of the mutational landscape revealed by genomic surveillance of SARS-CoV-2:  
32 Variant Database (**vdb**). With this tool, we uncovered several groups of recently sequenced  
33 genomes with mutations at critical antibody epitopes. Among this group is a new lineage  
34 emerging in NYC that has increased in frequency to now account for ~32% of sequenced  
35 genomes as of February 2021. We confirm the rapid spread of B.1.526 in NYC during early 2021  
36 through phylodynamic inference. Furthermore, we evaluated the impact of the B.1.526 spike  
37 mutations on the neutralization titer of convalescent and vaccinee plasma.

## 38 Results

### 39 vdb

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

### 49 Defining mutations of B.1.526

50 One notable cluster of genome sequences was collected from the New York region and  
51 represents a distinct lineage, now designated as B.1.526 (**Figure 1, Supplementary Figure 1**).  
52 This variant is found within the 20.C clade and is distinguished by 3 defining spike mutations:  
53 L5F, T95I, and D253G. Within B.1.526, the largest sub-clade is defined by E484K and two distinct  
54 sub-clades are each defined by S477N; both of these mutations located within the receptor-  
55 binding domain (RBD) of spike (**Figure 2 and Supplementary Table 1**). We note that the  
56 evolutionary history at spike position 701 varies depending on whether the tree is rooted using  
57 a molecular clock (**Figure 1**) versus its sister clade (characterized by an L452R mutation;  
58 **Supplementary Figure 2**), the latter of which posits a substitution A701V followed by a  
59 reversion V701A. Among the nucleotide mutations in lineage B.1.526, the most characteristic  
60 include A16500C (NSP13 Q88H), A22320G (spike D253G), and T9867C (NSP4\_L438P). Another

61 notable feature of the B.1.526 lineage is the deletion of nucleotides 11288-11296 (NSP6 106-  
62 108), which also occurs in variants B.1.1.7, B.1.351, P.1, and B.1.525<sup>14</sup>.  
63  
64 Regarding four of the spike mutations prevalent in this lineage: (1) E484K is known to attenuate  
65 neutralization of multiple anti-SARS-CoV-2 antibodies, particularly those found in class 2 anti-  
66 RBD neutralizing antibodies<sup>13,15</sup>, and is also present in variants B.1.351<sup>4</sup> and P.1/B.1.1.248<sup>2</sup>, (2)  
67 D253G has been reported as an escape mutation from antibodies against the N-terminal  
68 domain<sup>16</sup>, (3) S477N has been identified in several earlier lineages<sup>17</sup>, is near the epitopes of  
69 multiple antibodies<sup>18</sup>, and has been implicated to increase viral infectivity through enhanced  
70 interactions with ACE2<sup>19,20</sup>, and (4) A701V sits adjacent to the S2' cleavage site of the  
71 neighboring protomer and is shared with variant B.1.351<sup>4</sup>. The overall pattern of mutations in  
72 lineage B.1.526 (**Figure 2**) suggests that it arose in part in response to selective pressure from  
73 antibodies. Based on the dates of collection of these viruses, it appears that the frequency of  
74 this lineage has increased rapidly in New York (**Table 2**).

#### 75 [Trends in B.1.526 surveillance](#)

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

84 January 2021. From early January to early March, B.1.526 has been increasing by about 0.7%  
85 per day (segmented linear regression) and was at 43% the week prior to 03 March 2021 (**Figure**  
86 **3A**). Around 54% (n=678) of the B.1.526 genomes contain the E484K mutation, which has also  
87 been rising in frequency since early 2021. The weekly average of B.1.526 genomes with E484K  
88 has been above 10% since 01 February 2021 and has been increasing around 0.4% per day  
89 (**Figure 3B**).

90  
91 This increase in B.1.526 temporally coincides with the peak and subsequent decline of the  
92 second epidemic wave in NYC (**Figure 3C**). If we separate the approximated number of B.1.526  
93 cases from the rest of second wave SARS-CoV-2, the non-B.1.526 virus has steadily declined  
94 since its peak in early January 2021. However, the increasing proportion of B.1.526 appears to  
95 have slowed the rate of decline in total COVID-19 case counts in NYC.

96  
97 [Geographic distribution of B.1.526 in NYC](#)  
98 The New York City Public Health Laboratory and the PRL in New York have sequenced 4538  
99 SARS-CoV-2 genomes from December 2020 thru February 2021 (**Figure 4A**). Geographic case  
100 distribution of specimens received at PHL and PRL for SARS-CoV-2 diagnostic nucleic acid  
101 amplification testing (NAAT) are representative of citywide testing efforts. Those SARS-CoV-2  
102 positive specimens with NAAT cross-threshold values below 32 were selected at random to be  
103 sequenced. On a month-to-month basis using data generated by NYC PHL and PRL, we have  
104 observed an increasing number of B.1.526 genomes identified throughout NYC. The geographic  
105 distribution of over 600 B.1.526 E484K cases is similar (**Figure 4B**). While the B.1.526 lineage is

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

108

### 109 [Phylodynamic analysis](#)

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

118

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

127

128 As part of the phylodynamic analysis, we inferred the time of most recent common ancestor  
129 (TMRCA) for the B.1.526 E484K clade to be 08 November 2020 (95% HPD: 22 October – 24  
130 November). The TMRCA for the rest of the B.1.526 clade was estimated to be 15 September  
131 2020 (95% HPD: 17 August – 08 October).

132

### 133 [Neutralization activity of convalescent and vaccinee plasma against B.1.526](#)

134 The identification of several mutations associated with resistance to anti-SARS-CoV-2  
135 antibodies in B.1.526 sequences raises the question of the impact on SARS-CoV-2 immunity. We  
136 generated HIV-based pseudoviruses expressing SARS-CoV-2 spike protein containing either the  
137 most common B.1.526 mutation pattern (v.1: L5F, T95I, D253G, E484K, D614G, and A701V), the  
138 2<sup>nd</sup> most common pattern (v.2: L5F, T95I, D253G, S477N, D614G, and Q957R), or only D614G.  
139 Pseudovirus neutralization titers were determined for human plasma samples from vaccinees  
140 [Moderna (mRNA-1273) or Pfizer-BioNTech(BNT162b2)]<sup>8</sup> or convalescent plasma [at either  
141 1.3<sup>15</sup> or 6.2 months<sup>13</sup> post-infection]. The E484K-containing B.1.526 pseudovirus had a  
142 statistically significant reduced neutralization titer compared to the D614G control: for vaccinee  
143 plasma, 4.5-fold reduced ( $p = 0.00005$ ); for 1.3-month convalescent plasma, 6.0-fold reduced ( $p$   
144 = 0.03); and for 6.2-month convalescent plasma, 4.8-fold reduced ( $p = 0.02$ ) (**Figure 5a and**  
145 **Supplementary Table 2**). The smaller reduction of the titers in the 6.2-month convalescent  
146 plasma samples compared to the 1.3-month samples is consistent with the greater resistance of  
147 more matured anti-SARS-CoV-2 antibodies to viral escape mutations<sup>23</sup>. The S477N/Q957R-  
148 containing B.1.526 pseudovirus demonstrated a smaller effect on plasma neutralization (**Figure**  
149 **5b**).

150



## 151 Discussion

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

162  
163 Pseudovirus containing spike gene mutations associated with B.1.526 was significantly more  
164 resistant to neutralization by either convalescent or vaccinee plasma. The presence of E484K  
165 mutation likely plays a key role in facilitating increased viral transmission and reducing antibody  
166 neutralizing titers, as previously shown in other studies<sup>7,27</sup>. Continued monitoring for emerging  
167 variants with mutations such as E484K is important to maximize the impact of public health  
168 measures to mitigate the effects of the SARS-CoV-2 pandemic. For example, high frequencies of  
169 SARS-CoV-2 variants has potential impacts on selection of appropriate antibody therapeutics  
170 and vaccination strategies.

## 171 Methods

### 172 Variant Database Program

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

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

189 Results presented here are based on a multiple sequence alignment from GISAID<sup>11,12</sup>  
190 downloaded on February 10, 2021. Additional sequences downloaded from GISAID on February  
191 22, 2021, were aligned with MAFFT v7.464<sup>28</sup>.

## 192 Initial Phylogenetic Analysis

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

## 198 Library preparation and sequencing

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

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

## 222 [Genome Assembly](#)

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

236 [Library preparation and sequencing \(PRL\)](#)

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

246

247 [Genome Assembly \(PRL\)](#)

248 For each specimen, sequencing adapters are first trimmed using Trim Galore v0.6.6<sup>34</sup>, then  
249 aligned to the SARS-CoV-2 Wuhan-Hu-1 reference genome (NCBI Nucleotide NC\_045512.2)  
250 using BWA MEM 0.7.17-r1188<sup>35</sup>. Reads that are unmapped or those that have secondary  
251 alignments are discarded from the alignment. Consensus and mutations were called using  
252 samtools<sup>37</sup> and Intrahost variant analysis of replicates (iVar)<sup>36</sup> with a minimum quality score of  
253 20, frequency threshold of 0.6 and a minimum read depth of 10x coverage. A consensus  
254 genome with  $\geq 90\%$  breath-of-coverage with  $\leq 3000$  ambiguous bases is considered a successful  
255 reconstruction (as per APHL recommendation).

256

## 257 [Genome alignment](#)

258 Complete genome sequences produced by the NYC PHL and the PRL with reported collection  
259 dates on or before 04 March 2021 were analyzed. We restricted our analysis to genomes  
260 produced by public health surveillance to NYC to reduce bias due to geography or preferential  
261 sequencing of viral variants by academic institutions. Genomes were aligned to the Wuhan-Hu-  
262 1 reference genome (GenBank Accession MN908947) using mafft v7.475 (mafft --6merpair --  
263 keeplength --addfragments)<sup>28</sup>. Pango lineage designations<sup>31</sup> for variants were assigned using  
264 Pangolin v2.3.2<sup>39</sup>.

265

## 266 [Segmented regression analysis](#)

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

269

## 270 [Maximum likelihood phylogenetic inference](#)

271 Maximum likelihood trees were inferred using IQTree2 for B.1.1.7, B.1.427, B.1.429, and  
272 B.1.526 genomes using a GTR+F+ $\Gamma_4$  substitution model<sup>40</sup>. Minimum branch length of 1e-9 was  
273 enforced and an expanded NNI search (--allnni) was employed to improve topology search.  
274 Preliminary molecular clock analyses were performed in TreeTime v0.8.1 using a fixed  
275 substitution rate of  $8 \times 10^{-4}$  substitutions/site/year and a skyline coalescent model<sup>41</sup>. This  
276 analysis identified 34 genomes whose root-to-tip genetic distance were flagged as problematic  
277 and excluded from subsequent phylodynamic analyses. TreeTime was also used to root and  
278 perform ancestral state reconstruction for a tree inferred from the 258 B.1.526 genomes  
279 sampled by the NYC PHL used to display the history of spike mutations in B.1.526 (**Figure 1**).

280

### 281 Bayesian phylodynamic inference

282 We performed population growth rate inference in coalescence-based framework using an  
283 exponential growth model in BEAST 1.10.4<sup>22</sup>. We used a strict molecular clock model with the  
284 fixed substitution rate of  $8 \times 10^{-4}$  substitutions/site/year. We applied a GTR+F+ $\Gamma_4$  substitution  
285 model and specified the following priors for the population growth model: OneOnX distribution  
286 prior for the population size parameter and Laplace distribution prior (mean = 0.0, scale = 1.0)  
287 for the growth rate prior. Markov chain Monte Carlo analyses were run for 100-300 million  
288 generations; the first 10% of samples were discarded as burn-in. Separate inference was  
289 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  
290 B.1.526 E484K (n=678). For the B.1.526 phylodynamic inference, we did not include two  
291 sequences most closely related to B.1.526 (hCoV-19/USA/NY-NYCPHL-001701/2020 and hCoV-  
292 19/USA/NY-NYCPHL-002542/2021).

293

### 294 Geocoding addresses

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

### 300 Point density method

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

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

308

#### 309 [Human plasma samples](#)

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

314

#### 315 [Pseudovirus neutralization by human plasma samples](#)

316 Human plasma samples were assayed for neutralization activity against lentiviruses  
317 pseudotyped with SARS-CoV-2 spike containing a 21-amino acid cytoplasmic tail deletion and  
318 either D614G or mutations corresponding to lineage B.1.526 (L5F, T95I, D253G, E484K, D614G,  
319 and A701V). Pseudotyped lentiviruses were generated and neutralizations assays were  
320 conducted as previously described<sup>43,44</sup>. Briefly, lentiviral particles were produced by co-  
321 transfecting the gene encoding SARS-CoV-2 spike protein (D614G or B.1.526) and Env-deficient  
322 HIV backbone expressing Luciferase-IRES-ZsGreen. Plasma samples were heat inactivated at  
323 56°C for 1 hour, then 3-fold serial diluted and incubated with SARS-CoV-2 pseudotyped virus for  
324 1 hour at 37°C. The virus/plasma mixture was added to 293T<sub>ACE2</sub> target cells, which were  
325 seeded the previous day on poly-L-lysine coated plates. After incubating for 48 hours at 37°C,



326 target cells were lysed with Britelite Plus (Perkin Elmer) and luciferase activity was measured as  
327 relative luminescence units (RLUs) and normalized to values derived from cells infected with  
328 pseudotyped virus in the absence of plasma. Data were fit to 2-parameter non-linear regression  
329 in Antibody database<sup>45</sup>.

330

### 331 [Data availability](#)

332 The data analyzed as part of this project were obtained from the GISAID database and through  
333 a Data Use Agreement between NYC DOHMH and the University of California San Diego.  
334 Sequences analyzed by using the **vdb** tool were downloaded from GISAID. No personally  
335 identifying information were included as part of these analyses. SARS-CoV-2 genomes included  
336 in these analyses have been deposited in GISAID. See **Supplementary Data 1** for a list of  
337 genomes, including which genomes were excluded from the phylogenetic analysis.  
338 Data for **Figure 5** are provided in **Supplementary Table 2**.

339

### 340 [Code availability](#)

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

342 <https://github.com/variant-database/vdb>.

343

### 344 [Acknowledgments](#)

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

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

### 353 [Author Contributions](#)

354 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.  
355 performed genome sequencing and assembly. J.O.W. curated data. C.G., M. Caskey and M.C.N.  
356 provided clinical samples. P.N.P.G. and J.R.K. carried out experiments. A.P.W., C.O.B., Z.Y., S.H.,  
357 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  
358 manuscript with input from all co-authors. A.P.W., P.J.B., J.O.W., J.L.R. and S.H. supervised the  
359 study.

360

### 361 [Competing Interests](#)

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

367

368 **Tables**

369

370 **Table 1**

371 Mutation patterns of viruses with mutations at select Spike positions, excluding viruses related

372 to variants B.1.1.7, B.1.351, B.1.1.248, and B.1.429. Mutations included in this analysis were

373 E484K, N501Y, K417T, K417N, L452R, and A701V. In this table viruses are only included if their

374 spike mutation pattern exactly matches the given pattern. Note about P681H/P681R: variant

375 B.1.1.7 has P681H. Note about W152L: variant B.1.429 has W152C

376

377

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

385

386 [Table 2](#)

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

391  
392  
393 Viruses containing spike mutations T95I and D253G (earliest collection date Nov. 23, 2020)

394

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

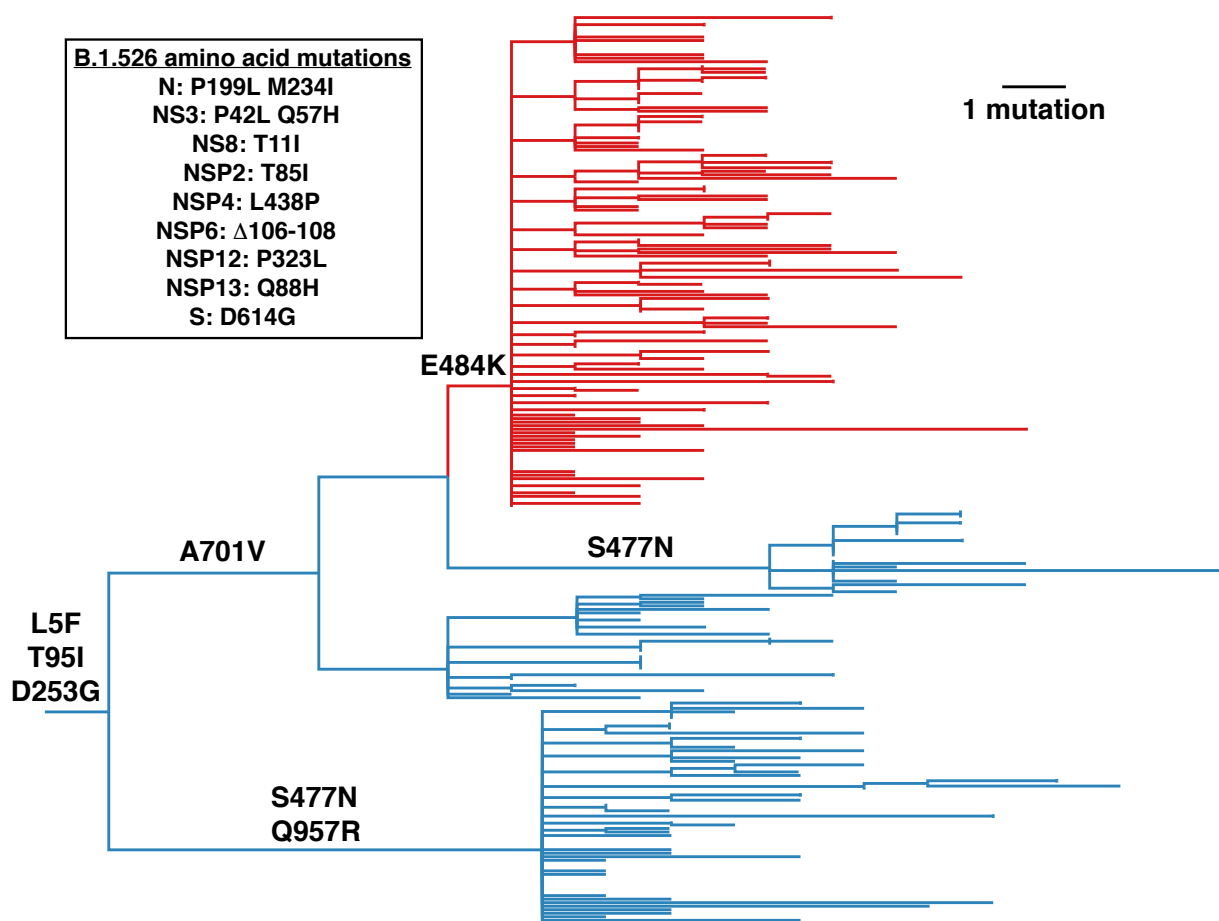
401  
402  
403  
404 Viruses containing spike mutations L5F, T95I, D253G, E484K, D614G, and A701V (earliest  
405 collection date Dec. 16, 2020)

406

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

413  
414

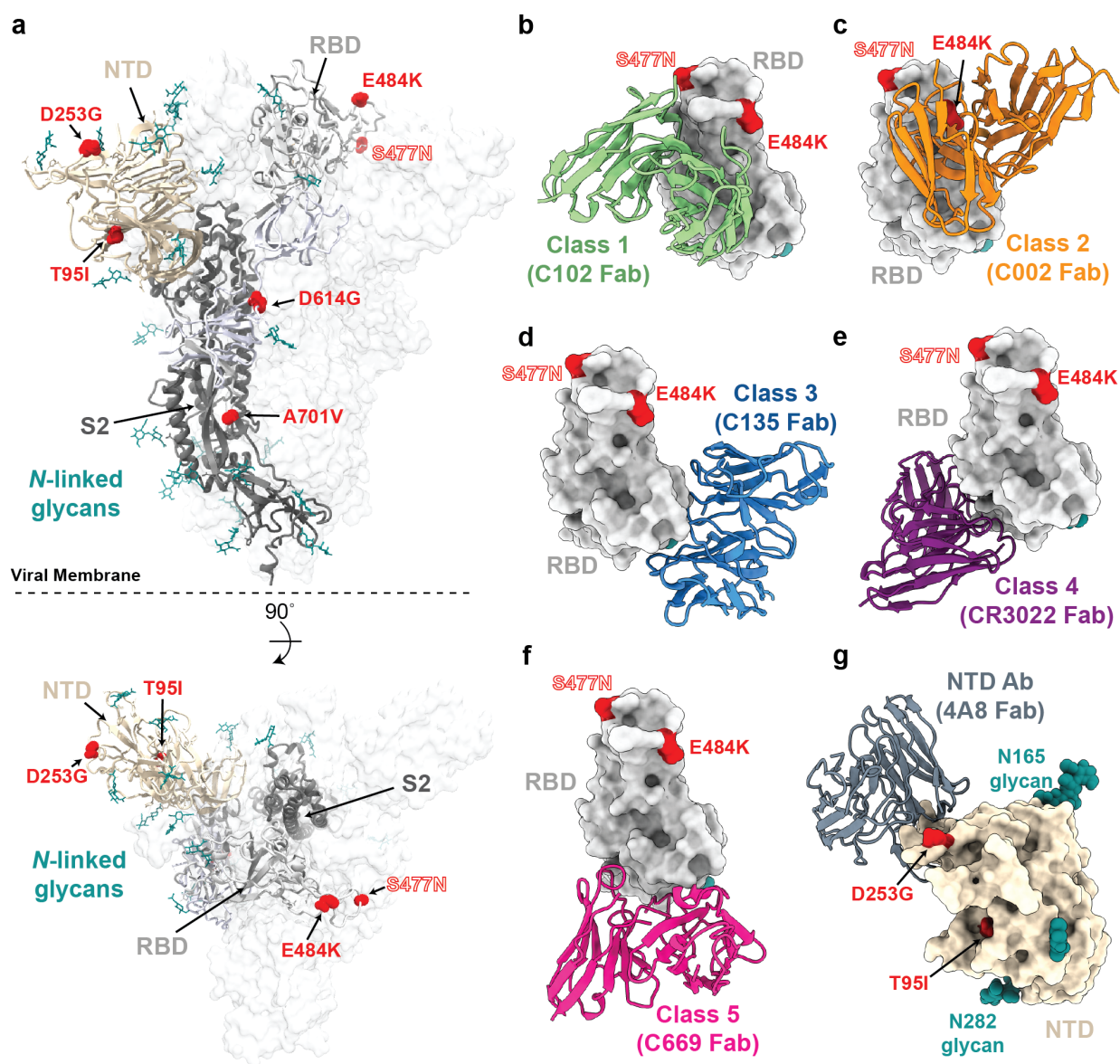
415 Figure 1.  
416 Phylogenetic tree of lineage B.1.526 indicating spike mutations. Maximum likelihood phylogeny  
417 of SARS-CoV-2 variant B.1.526 sampled by NYC PHL (n=258). Amino acid substitutions in the  
418 spike protein occurring on internal branches are labeled, including the three spike mutations  
419 characteristic of B.1.526. The B.1.526 clade defined by the E484K mutation is highlighted in red.  
420 Inset highlights non-spike amino acid substitutions and deletions differentiating the B.1.526  
421 clade from the Hu-1 reference genome. For display purposes, only NYC PHL genomes are  
422 shown.  
423



424

425

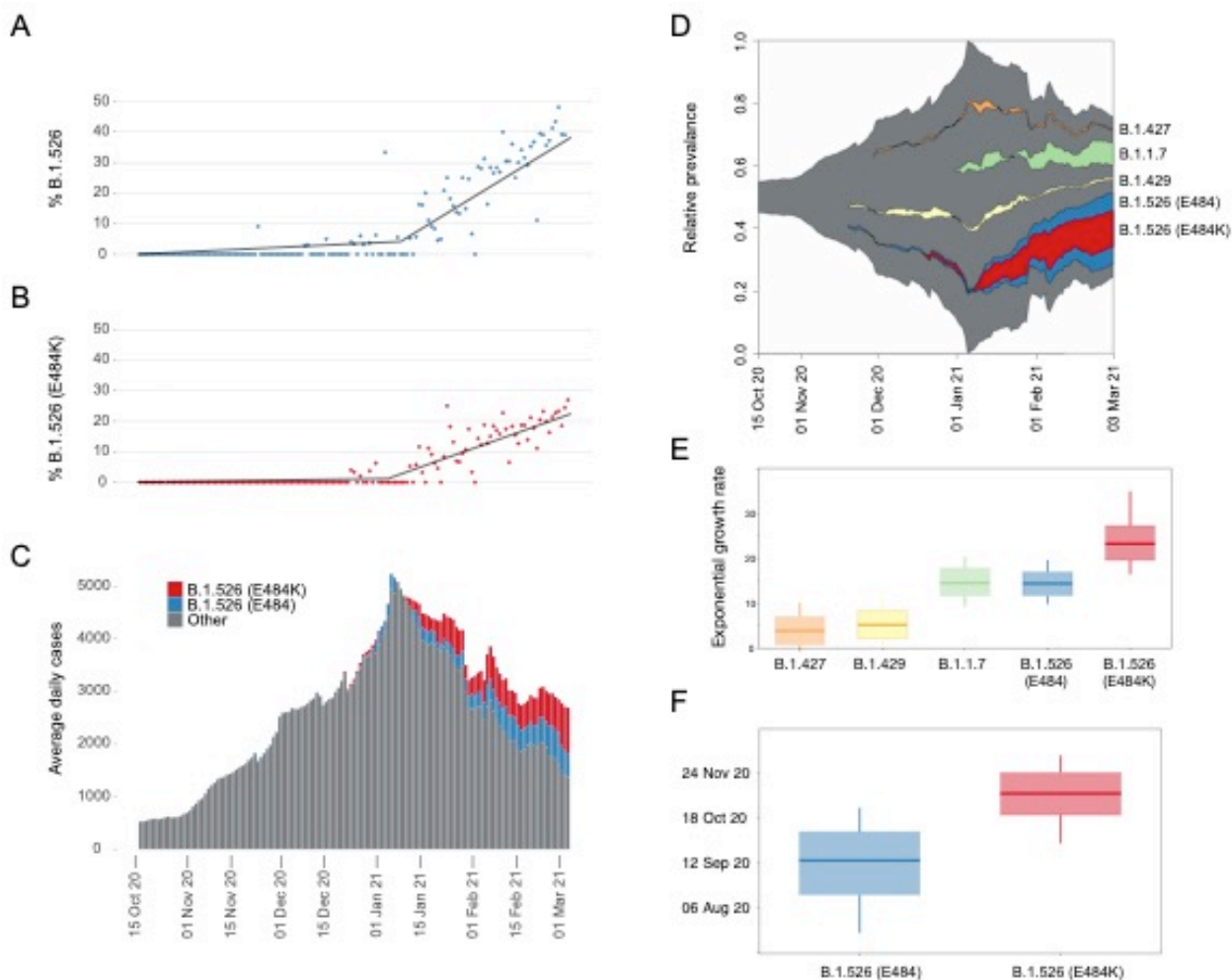
426 Figure 2.  
427 Structural locations of the spike mutations of lineage B.1.526.



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<sup>8</sup>; **g**, NTD-  
435 specific antibody 4A8 (PDB 7C2L).  
436

437 Figure 3.



438

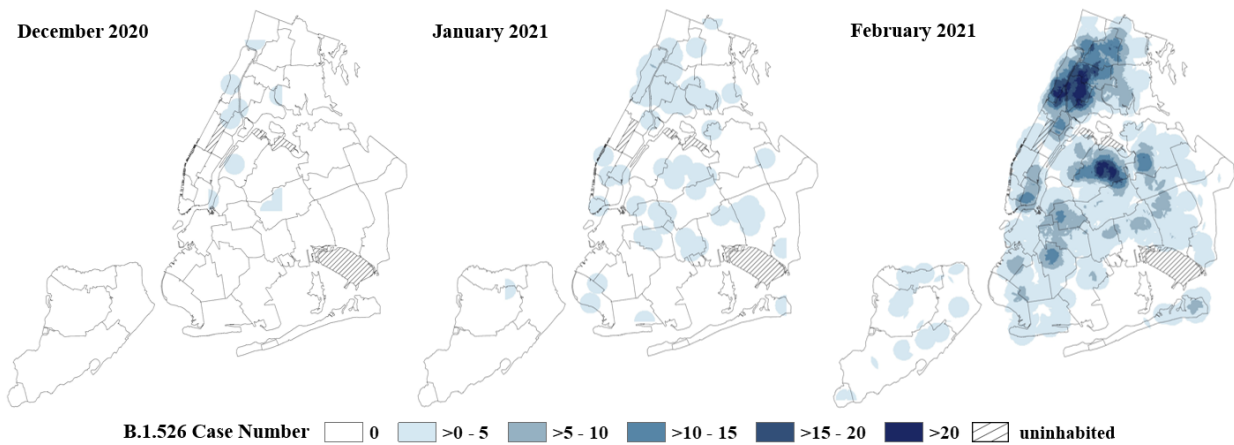
439 Rise of SARS-CoV-2 variants in New York City (NYC) in late-2020 and early 2021. (A) Relative  
 440 frequency of B.1.526. Segmented linear regression is shown as a solid black line. (B) Relative  
 441 frequency of B.1.526 with E484K mutation. Segmented linear regression is shown as a dashed  
 442 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

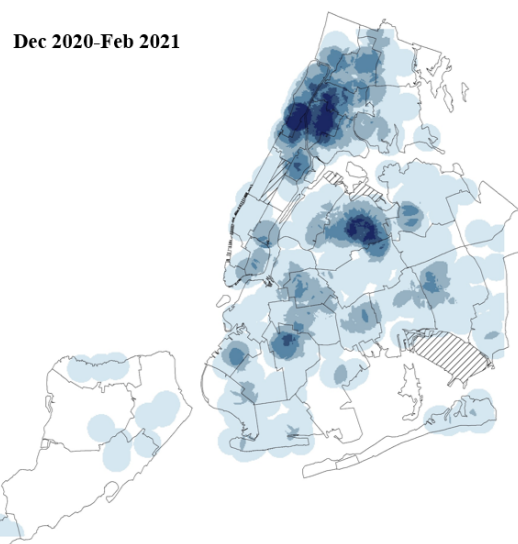
451 Figure 4.

452 **A**



453

454 **B**



455 B.1.526 E484K Case Number   0   >0 - 3   >3 - 6   >6 - 9   >9 - 12   >12   uninhabited

456

457 **(A)** Spatiotemporal increase of B.1.526 lineage in New York City (NYC). Point density of B.1.526

458 variants geo-located by case address overlaid on a map of NYC delineated by United Hospital

459 Fund areas. Data for each month is based on specimen collection date. The NYC PHL and the

460 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)

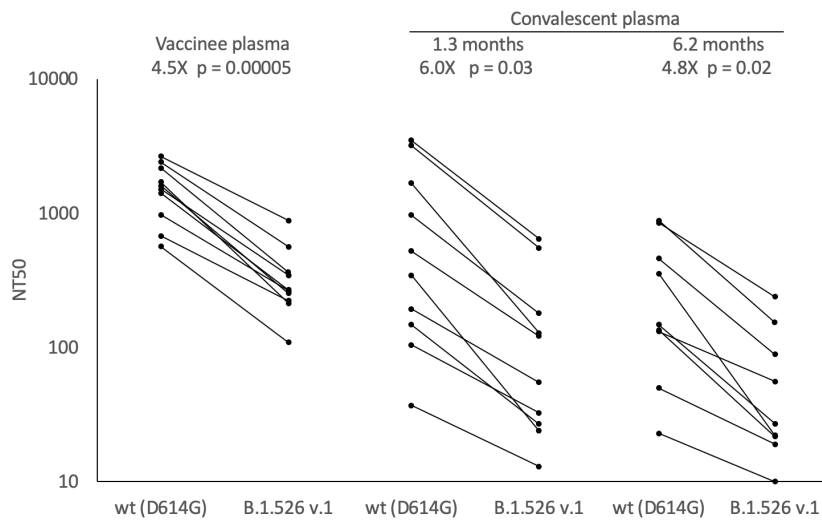
472 after infection. **(A)** Pseudovirus with spike mutations L5F, T95I, D253G, E484K, D614G

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

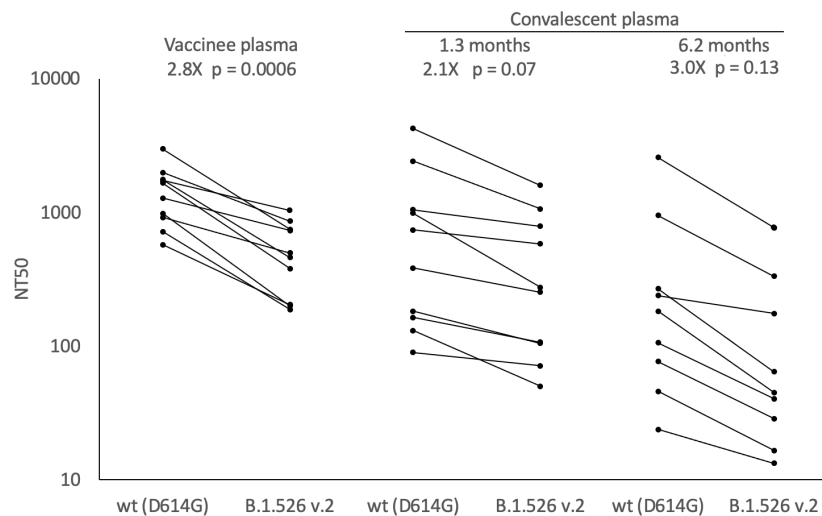
475 *t*-tests. Fold-differences of means are shown.

476 **A**



477

478 **B**



479

## 480 References

- 481 1. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases  
482 Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).
- 483 2. Faria, N. R. *et al.* Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus:  
484 preliminary findings. *virological.org* [https://virological.org/t/genomic-characterisation-of-](https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586)  
485 [an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586](https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586) (2021).
- 486 3. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage  
487 in the UK defined by a novel set of spike mutations. *virological.org*  
488 [https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-](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)  
489 [lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563](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).
- 490 4. Tegally, H. *et al.* *Emergence and rapid spread of a new severe acute respiratory syndrome-*  
491 *related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa.*  
492 <http://medrxiv.org/lookup/doi/10.1101/2020.12.21.20248640> (2020)  
493 doi:10.1101/2020.12.21.20248640.
- 494 5. Zhang, W. *et al.* Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA*  
495 (2021) doi:10.1001/jama.2021.1612.
- 496 6. Cele, S. *et al.* *Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma.*  
497 <http://medrxiv.org/lookup/doi/10.1101/2021.01.26.21250224> (2021)  
498 doi:10.1101/2021.01.26.21250224.
- 499 7. Greaney, A. J. *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-  
500 binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host &*  
501 *Microbe* **29**, 463-476.e6 (2021).

- 502 8. Wang, Z. *et al.* mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants.  
503 *Nature* (2021) doi:10.1038/s41586-021-03324-6.
- 504 9. Wibmer, C. K. *et al.* SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19  
505 donor plasma. *Nat Med* (2021) doi:10.1038/s41591-021-01285-x.
- 506 10. Hodcroft, E. B. *et al.* Want to track pandemic variants faster? Fix the bioinformatics  
507 bottleneck. *Nature* **591**, 30–33 (2021).
- 508 11. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID’s innovative  
509 contribution to global health: Data, Disease and Diplomacy. *Global Challenges* **1**, 33–46  
510 (2017).
- 511 12. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to  
512 reality. *Eurosurveillance* **22**, (2017).
- 513 13. Gaebler, C. *et al.* Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639–644  
514 (2021).
- 515 14. Martin, D. P. *et al.* *The emergence and ongoing convergent evolution of the N501Y lineages*  
516 *coincides with a major global shift in the SARS-CoV-2 selective landscape.*  
517 <http://medrxiv.org/lookup/doi/10.1101/2021.02.23.21252268> (2021)  
518 doi:10.1101/2021.02.23.21252268.
- 519 15. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent  
520 individuals. *Nature* **584**, 437–442 (2020).
- 521 16. McCallum, M. *et al.* *N-terminal domain antigenic mapping reveals a site of vulnerability for*  
522 *SARS-CoV-2.* <http://biorxiv.org/lookup/doi/10.1101/2021.01.14.426475> (2021)  
523 doi:10.1101/2021.01.14.426475.

- 524 17. Hodcroft, E. B. *et al.* *Emergence and spread of a SARS-CoV-2 variant through Europe in the*  
525 *summer of 2020.* <http://medrxiv.org/lookup/doi/10.1101/2020.10.25.20219063> (2020)  
526 doi:10.1101/2020.10.25.20219063.
- 527 18. Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic  
528 strategies. *Nature* **588**, 682–687 (2020).
- 529 19. Chen, J., Wang, R., Wang, M. & Wei, G.-W. Mutations Strengthened SARS-CoV-2 Infectivity.  
530 *Journal of Molecular Biology* **432**, 5212–5226 (2020).
- 531 20. Ou, J. *et al.* *Emergence of SARS-CoV-2 spike RBD mutants that enhance viral infectivity*  
532 *through increased human ACE2 receptor binding affinity.*  
533 <http://biorxiv.org/lookup/doi/10.1101/2020.03.15.991844> (2020)  
534 doi:10.1101/2020.03.15.991844.
- 535 21. Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H. & Rambaut, A. The  
536 epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian  
537 coalescent approach. *Mol Biol Evol* **20**, 381–387 (2003).
- 538 22. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST  
539 1.10. *Virus Evol* **4**, vey016 (2018).
- 540 23. Muecksch, F. *et al.* *Development of potency, breadth and resilience to viral escape*  
541 *mutations in SARS-CoV-2 neutralizing antibodies.*  
542 <http://biorxiv.org/lookup/doi/10.1101/2021.03.07.434227> (2021)  
543 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](https://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

## 601 Supplementary Material

602

### 603 Supplementary Methods.

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

605

#### 606 Notation

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

610

611 To define a variable for a cluster or pattern: <name> = cluster or pattern

612 Set operations +, -, and \* (intersection) can be applied to clusters or patterns

613 If no cluster is entered, all viruses will be used ("world")

614

#### 615 Filter commands

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

622

#### 623 Commands to find mutation patterns

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

626

#### 627 Listing commands

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

636

#### 637 Other commands

638 sort <cluster>                                (by date)

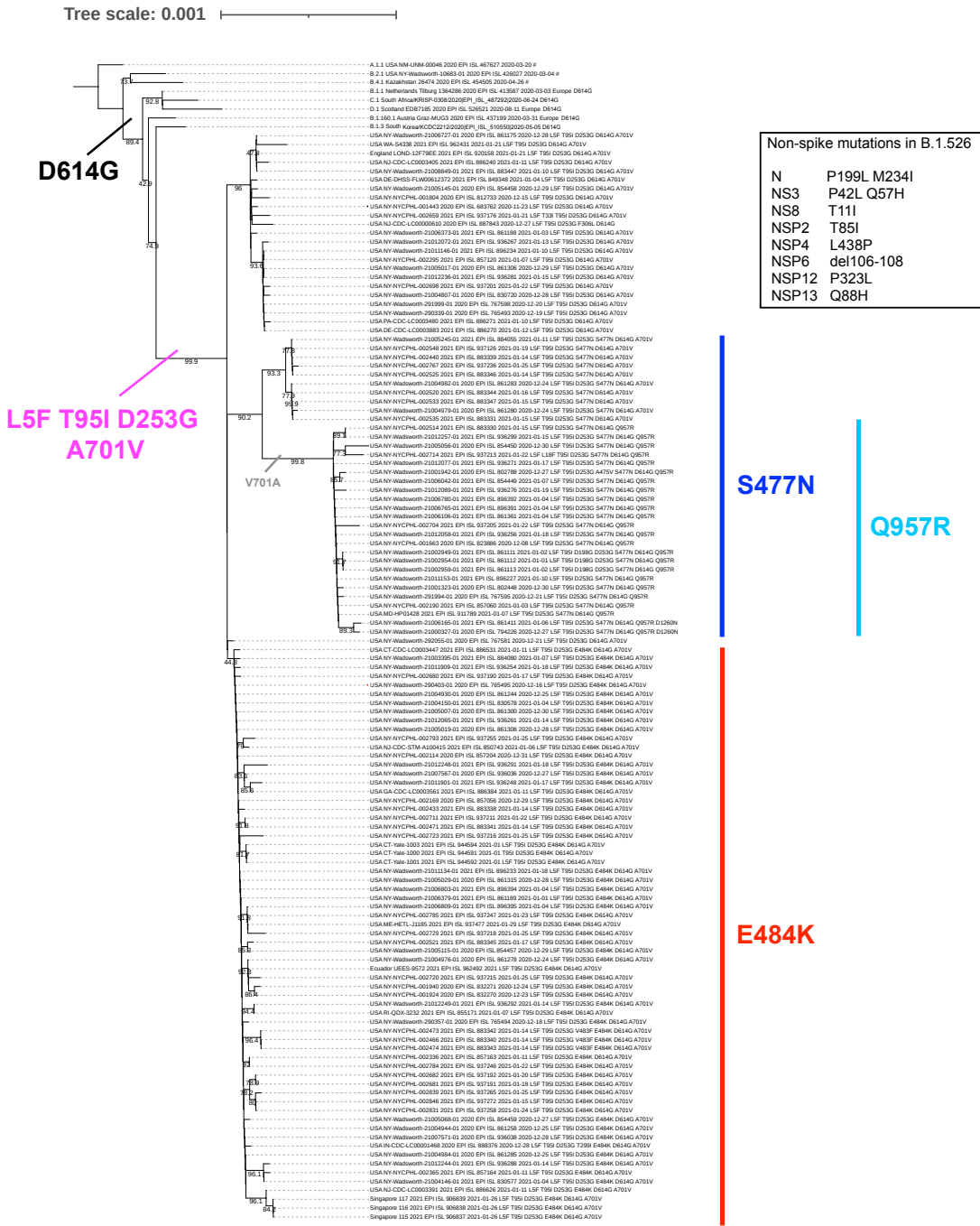
639 help

640 history

641 quit

642

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

647 **Supplementary Figure 2.**

648 **Maximum likelihood phylogenetic tree of the B.1.526 lineage in relation to a sister clade**

649 defined by an L452R spike mutation and the 20C ancestral virus (both shown in gray). Tree was

650 rooted using the clade 20C ancestral viruses. Amino acid substitutions in the spike protein

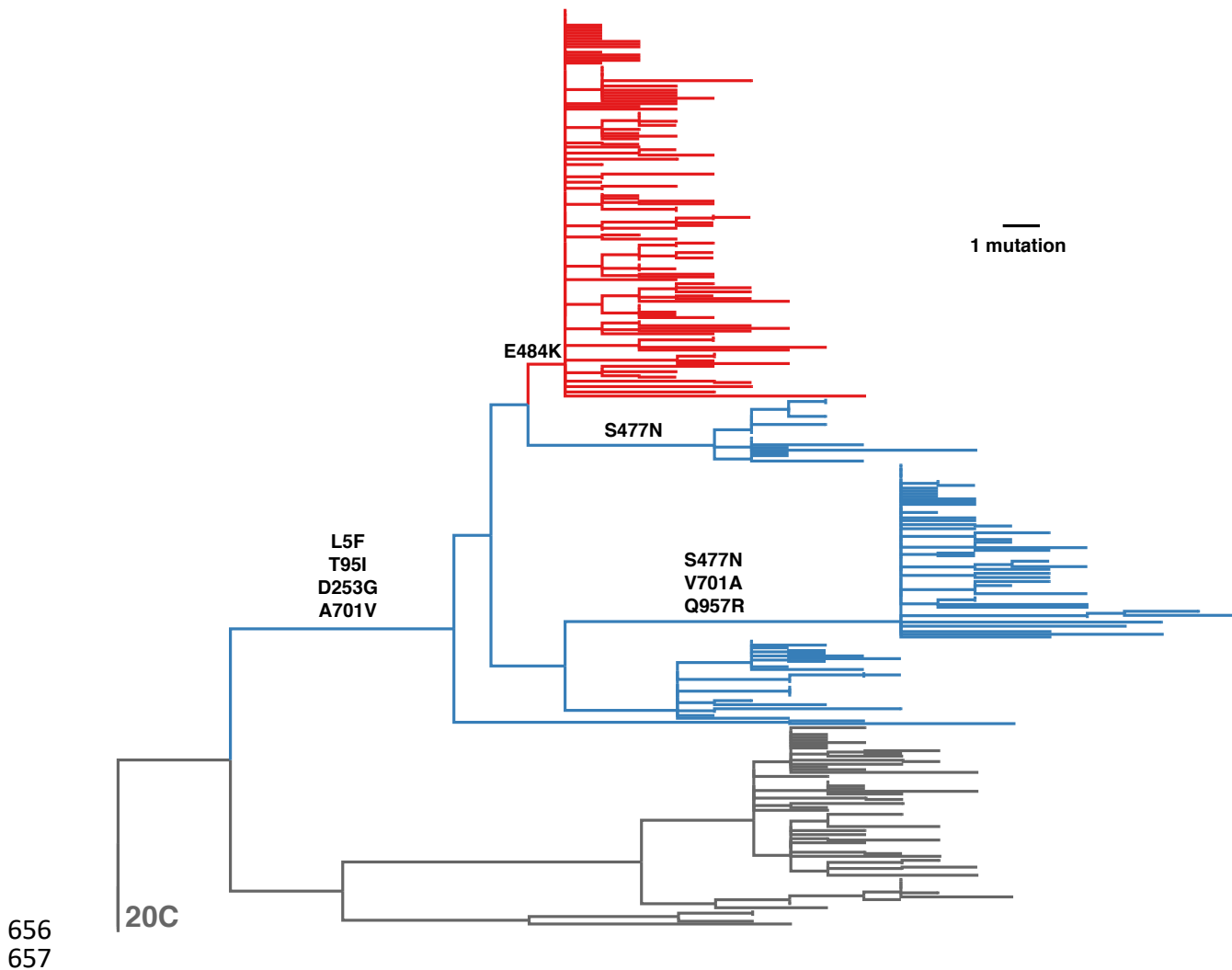
651 occurring on internal branches are labeled leading to and within B.1.526 are labeled. The

652 B.1.526 lineage is colored blue, except for the clade defined by the E484K mutation, which is

653 highlighted in red. The most common pattern of spike mutations in the sister clade is D80G,

654  $\Delta$ Y144, F157S, L452R, D614G, T859N, and D950H.

655



658 [Supplementary Table 1.](#)

659 List of 124 viral genomes (with their accession number, location, collection date, and spike

660 mutations) in lineage B.1.526. Mutations E484K, S477N, Q957R are highlighted in red, blue, and

661 cyan, respectively.

662

663 EPI\_ISL\_683762, USA/NY-NYCPHL-001443/2020-11-23 : L5F T95I D253G D614G A701V  
664 EPI\_ISL\_823886, USA/NY-NYCPHL-001663/2020-12-08 : L5F T95I D253G **S477N** D614G **Q957R**  
665 EPI\_ISL\_812733, USA/NY-NYCPHL-001804/2020-12-15 : L5F T95I D253G D614G A701V  
666 EPI\_ISL\_765495, USA/NY-Wadsworth-290403-01/2020-12-16 : L5F T95I D253G **E484K** D614G A701V  
667 EPI\_ISL\_765494, USA/NY-Wadsworth-290357-01/2020-12-18 : L5F T95I D253G **E484K** D614G A701V  
668 EPI\_ISL\_765493, USA/NY-Wadsworth-290339-01/2020-12-19 : L5F T95I D253G D614G A701V  
669 EPI\_ISL\_767598, USA/NY-Wadsworth-291999-01/2020-12-20 : L5F T95I D253G D614G A701V  
670 EPI\_ISL\_767581, USA/NY-Wadsworth-292055-01/2020-12-21 : L5F T95I D253G D614G A701V  
671 EPI\_ISL\_767595, USA/NY-Wadsworth-291994-01/2020-12-21 : L5F T95I D253G **S477N** D614G **Q957R**  
672 EPI\_ISL\_832270, USA/NY-NYCPHL-001924/2020-12-23 : L5F T95I D253G **E484K** D614G A701V  
673 EPI\_ISL\_832271, USA/NY-NYCPHL-001940/2020-12-24 : L5F T95I D253G **E484K** D614G A701V  
674 EPI\_ISL\_861278, USA/NY-Wadsworth-21004976-01/2020-12-24 : L5F T95I D253G **E484K** D614G A701V  
675 EPI\_ISL\_861280, USA/NY-Wadsworth-21004979-01/2020-12-24 : L5F T95I D253G **S477N** D614G A701V  
676 EPI\_ISL\_861283, USA/NY-Wadsworth-21004982-01/2020-12-24 : L5F T95I D253G **S477N** D614G A701V  
677 EPI\_ISL\_861258, USA/NY-Wadsworth-21004944-01/2020-12-25 : L5F T95I D253G **E484K** D614G A701V  
678 EPI\_ISL\_861285, USA/NY-Wadsworth-21004984-01/2020-12-25 : L5F T95I D253G **E484K** D614G A701V  
679 EPI\_ISL\_861244, USA/NY-Wadsworth-21004930-01/2020-12-25 : L5F T95I D253G **E484K** D614G A701V  
680 EPI\_ISL\_802788, USA/NY-Wadsworth-21001942-01/2020-12-27 : L5F T95I D253G A475V **S477N** D614G **Q957R**  
681 EPI\_ISL\_794226, USA/NY-Wadsworth-21000327-01/2020-12-27 : L5F T95I D253G **S477N** D614G **Q957R** D1260N  
682 EPI\_ISL\_936036, USA/NY-Wadsworth-21007567-01/2020-12-27 : L5F T95I D253G **E484K** D614G A701V  
683 EPI\_ISL\_854459, USA/NY-Wadsworth-21005068-01/2020-12-27 : L5F T95I D253G **E484K** D614G A701V  
684 EPI\_ISL\_887843, USA/NJ-CDC-LC00000610/2020-12-27 : L5F T95I D253G F306L D614G  
685 EPI\_ISL\_861315, USA/NY-Wadsworth-21005029-01/2020-12-28 : L5F T95I D253G **E484K** D614G A701V  
686 EPI\_ISL\_888376, USA/IN-CDC-LC00001468/2020-12-28 : L5F T95I D253G T299I **E484K** D614G A701V  
687 EPI\_ISL\_830720, USA/NY-Wadsworth-21004807-01/2020-12-28 : L5F T95I D253G D614G A701V  
688 EPI\_ISL\_861308, USA/NY-Wadsworth-21005019-01/2020-12-28 : L5F T95I D253G **E484K** D614G A701V  
689 EPI\_ISL\_861175, USA/NY-Wadsworth-21006727-01/2020-12-28 : L5F T95I D253G D614G A701V  
690 EPI\_ISL\_936038, USA/NY-Wadsworth-21007571-01/2020-12-28 : L5F T95I D253G **E484K** D614G A701V  
691 EPI\_ISL\_854458, USA/NY-Wadsworth-21005145-01/2020-12-29 : L5F T95I D253G D614G A701V  
692 EPI\_ISL\_854457, USA/NY-Wadsworth-21005115-01/2020-12-29 : L5F T95I D253G **E484K** D614G A701V  
693 EPI\_ISL\_857056, USA/NY-NYCPHL-002169/2020-12-29 : L5F T95I D253G **E484K** D614G A701V  
694 EPI\_ISL\_861306, USA/NY-Wadsworth-21005017-01/2020-12-29 : L5F T95I D253G D614G A701V  
695 EPI\_ISL\_802448, USA/NY-Wadsworth-21001323-01/2020-12-30 : L5F T95I D253G **S477N** D614G **Q957R**  
696 EPI\_ISL\_861300, USA/NY-Wadsworth-21005007-01/2020-12-30 : L5F T95I D253G **E484K** D614G A701V  
697 EPI\_ISL\_854450, USA/NY-Wadsworth-21005056-01/2020-12-30 : L5F T95I D253G **S477N** D614G **Q957R**  
698 EPI\_ISL\_857204, USA/NY-NYCPHL-002114/2020-12-31 : L5F T95I D253G **E484K** D614G A701V  
699 EPI\_ISL\_861112, USA/NY-Wadsworth-21002954-01/2021-01-01 : L5F T95I D198G D253G **S477N** D614G **Q957R**  
700 EPI\_ISL\_861189, USA/NY-Wadsworth-21006379-01/2021-01-01 : L5F T95I D253G **E484K** D614G A701V  
701 EPI\_ISL\_944591, USA/CT-Yale-1000/2021-01-01 : T95I D253G **E484K** D614G A701V  
702 EPI\_ISL\_944592, USA/CT-Yale-1001/2021-01-01 : L5F T95I D253G **E484K** D614G A701V  
703 EPI\_ISL\_944594, USA/CT-Yale-1003/2021-01-01 : L5F T95I D253G **E484K** D614G A701V  
704 EPI\_ISL\_962492, Ecuador/UEES-9572/2021-01-01 : L5F T95I D253G **E484K** D614G A701V  
705 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

759 EPI\_ISL\_936256, USA/NY-Wadsworth-21012058-01/2021-01-18 : L5F T95I D253G S477N D614G Q957R  
760 EPI\_ISL\_936254, USA/NY-Wadsworth-21011909-01/2021-01-18 : L5F T95I D253G E484K D614G A701V  
761 EPI\_ISL\_936291, USA/NY-Wadsworth-21012248-01/2021-01-18 : L5F T95I D253G E484K D614G A701V  
762 EPI\_ISL\_896233, USA/NY-Wadsworth-21011134-01/2021-01-18 : L5F T95I D253G E484K D614G A701V  
763 EPI\_ISL\_937191, USA/NY-NYCPHL-002681/2021-01-18 : L5F T95I D253G E484K D614G A701V  
764 EPI\_ISL\_936276, USA/NY-Wadsworth-21012089-01/2021-01-19 : L5F T95I D253G S477N D614G Q957R  
765 EPI\_ISL\_937126, USA/NY-NYCPHL-002548/2021-01-19 : L5F T95I D253G S477N D614G A701V  
766 EPI\_ISL\_937192, USA/NY-NYCPHL-002682/2021-01-20 : L5F T95I D253G E484K D614G A701V  
767 EPI\_ISL\_937176, USA/NY-NYCPHL-002659/2021-01-21 : L5F T33I T95I D253G D614G A701V  
768 EPI\_ISL\_920158, England/LOND-12F79EE/2021-01-21 : L5F T95I D253G D614G A701V  
769 EPI\_ISL\_962431, USA/WA-S4338/2021-01-21 : L5F T95I D253G D614G A701V  
770 EPI\_ISL\_937205, USA/NY-NYCPHL-002704/2021-01-22 : L5F T95I D253G S477N D614G Q957R  
771 EPI\_ISL\_937211, USA/NY-NYCPHL-002711/2021-01-22 : L5F T95I D253G E484K D614G A701V  
772 EPI\_ISL\_937201, USA/NY-NYCPHL-002698/2021-01-22 : L5F T95I D253G D614G A701V  
773 EPI\_ISL\_937213, USA/NY-NYCPHL-002714/2021-01-22 : L5F L18F T95I D253G S477N D614G Q957R  
774 EPI\_ISL\_937246, USA/NY-NYCPHL-002784/2021-01-22 : L5F T95I D253G E484K D614G A701V  
775 EPI\_ISL\_937247, USA/NY-NYCPHL-002785/2021-01-23 : L5F T95I D253G E484K D614G A701V  
776 EPI\_ISL\_937258, USA/NY-NYCPHL-002831/2021-01-24 : L5F T95I D253G E484K D614G A701V  
777 EPI\_ISL\_937236, USA/NY-NYCPHL-002767/2021-01-25 : L5F T95I D253G S477N D614G A701V  
778 EPI\_ISL\_937218, USA/NY-NYCPHL-002729/2021-01-25 : L5F T95I D253G E484K D614G A701V  
779 EPI\_ISL\_937216, USA/NY-NYCPHL-002723/2021-01-25 : L5F T95I D253G E484K D614G A701V  
780 EPI\_ISL\_937265, USA/NY-NYCPHL-002839/2021-01-25 : L5F T95I D253G E484K D614G A701V  
781 EPI\_ISL\_937215, USA/NY-NYCPHL-002720/2021-01-25 : L5F T95I D253G E484K D614G A701V  
782 EPI\_ISL\_937255, USA/NY-NYCPHL-002793/2021-01-25 : L5F T95I D253G E484K D614G A701V  
783 EPI\_ISL\_906839, Singapore/117/2021-01-26 : L5F T95I D253G E484K D614G A701V  
784 EPI\_ISL\_906838, Singapore/116/2021-01-26 : L5F T95I D253G E484K D614G A701V  
785 EPI\_ISL\_906837, Singapore/115/2021-01-26 : L5F T95I D253G E484K D614G A701V  
786 EPI\_ISL\_937477, USA/ME-HETL-J1185/2021-01-29 : L5F T95I D253G E484K D614G A701V  
787