

1 Spatiotemporal analyses illuminate the competitive advantage of a SARS-CoV-2
2 variant of concern over a variant of interest
3

4 Alexis Russell*¹, Collin O'Connor*^{2,3}, Erica Lasek-Nesselquist*^{4,5}, Jonathan Plitnick¹, John P.
5 Kelly⁶, Daryl M. Lamson¹, Kirsten St. George^{1,5#}
6

7 *These authors contributed equally to this work.
8

9 ¹ Laboratory of Viral Diseases, Wadsworth Center, New York State Department of Health,
10 Albany, NY 12208, USA

11 ² Division of Epidemiology, New York State Department of Health, Albany, NY 12242, USA

12 ³ Department of Geography, University at Buffalo, Buffalo, NY 14260

13 ⁴ Bioinformatics Core, Wadsworth Center, New York State Department of Health, Albany, NY
14 12208, USA

15 ⁵ Department of Biomedical Science, University at Albany, SUNY, Albany, NY 12208, USA

16 ⁶ Applied Genomic Technology Core, Wadsworth Center, New York State Department of
17 Health, Albany, NY 12208, USA
18

19 # Corresponding Author, ORCID # 0000-0001-5168-763X
20

21 Short title: Analysis of the competitive advantage of SARS-CoV-2 variants

22 Abstract:

23 The emergence of novel SARS-CoV-2 variants in late 2020 and early 2021 raised alarm
24 worldwide and prompted reassessment of the management, surveillance, and projected future of
25 COVID-19. Mutations that confer competitive advantages by increasing transmissibility or
26 immune evasion have been associated with the localized dominance of single variants. Thus,
27 elucidating the evolutionary and epidemiological dynamics among novel variants is essential for
28 understanding the trajectory of the COVID-19 pandemic. Here we show the interplay between
29 B.1.1.7 (Alpha) and B.1.526 (Iota) in New York (NY) from December 2020 to April 2021
30 through phylogeographic analyses, space-time scan statistics, and cartographic visualization. Our
31 results indicate that B.1.526 likely evolved in the Bronx in late 2020, providing opportunity for
32 an initial foothold in the heavily interconnected New York City (NYC) region, as evidenced by
33 numerous exportations to surrounding locations. In contrast, B.1.1.7 became dominant in regions
34 of upstate NY where B.1.526 had limited presence, suggesting that B.1.1.7 was able to spread
35 more efficiently in the absence of B.1.526. Clusters discovered from the spatial-time scan
36 analysis supported the role of competition between B.1.526 and B.1.1.7 in NYC in March 2021
37 and the outsized presence of B.1.1.7 in upstate NY in April 2021. Although B.1.526 likely
38 delayed the rise of B.1.1.7 in NYC, B.1.1.7 became the dominant variant in the Metro region by
39 the end of the study period. These results reveal the advantages endemicity may grant to a
40 variant (founder effect), despite the higher fitness of an introduced lineage. Our research
41 highlights the dynamics of inter-variant competition at a time when B.1.617.2 (Delta) is
42 overtaking B.1.1.7 as the dominant lineage worldwide. We believe our combined spatiotemporal
43 methodologies can disentangle the complexities of shifting SARS-CoV-2 variant landscapes at a
44 time when the evolution of variants with additional fitness advantages is impending.

45 Introduction:

46 The emergence of a novel SARS-CoV-2 variant B.1.1.7 (Alpha) in the United Kingdom
47 (UK) in late 2020 raised alarm worldwide and prompted major reassessment of the management,
48 surveillance, and projected future of COVID-19 (1,2). Evidence of increased transmissibility and
49 potential immune evasion prompted the World Health Organization to designate B.1.1.7 a
50 variant of concern (VOC) in December 2020 (3–6). Increased transmissibility of B.1.1.7 is likely
51 due to several mutations, including N501Y which confers antibody evasion (7,8) and increases
52 spike protein binding to the host cell (9), and del60-70 which enhances infectivity (10). The
53 emergence of B.1.1.7 and additional novel SARS-CoV-2 variants with competitive advantages
54 have resulted in localized dominance of single variants (11) and raised concern for increases in
55 COVID-19 incidence (12).

56 Around the time B.1.1.7 emerged and became dominant in the UK, novel variant B.1.526
57 (Iota) arose from within New York State (NY) (13,14). The proportion of B.1.526 quickly
58 increased in New York City (NYC) and led to a noticeable shift in lineage distribution during
59 early 2021 (13,14). The World Health Organization designated B.1.526 as a variant of interest
60 (VOI) due to its increase in prevalence coupled with mutations associated with immune evasion
61 (15). Spike mutation E484K, which is present in approximately 45% of B.1.526 genomes
62 sequenced in NY (GISAID.org) and is shared with the B.1.315 (Beta) and P.1 (Gamma) VOCs,
63 is associated with reduced neutralization by monoclonal antibodies (16). Further, 29% of
64 B.1.526 sequenced samples in NY possess the S477N spike mutation (GISAID.org), which
65 likely evolved twice independently within the lineage (13). S477N is hypothesized to confer
66 increased transmissibility due to its location at the ACE2 binding receptor in spike, which is

67 required for viral entry into the host cell. (17). Despite these concerns, an epidemiological
68 assessment of
69 B.1.526 in NYC from January through April 2021 found that the lineage did not cause more
70 severe disease and was not associated with increased risk of reinfection or vaccine breakthrough
71 (18).

72 Genomic surveillance of COVID-19 is a crucial tool to monitor and assess the
73 physiological and epidemiological characteristics of SARS-CoV-2 variants as they emerge. In
74 order to track the spread and impact of novel variants in NY, the New York State Department of
75 Health (NYSDOH) significantly expanded its genomic surveillance program in December 2020,
76 with the aim of sequencing a more representative subset of COVID-19 cases across the state. A
77 robust genomic surveillance system allows for the assessment of changes in variant distribution
78 over precise temporal and spatial scales. The co-circulation of B.1.526 and B.1.1.7 within NY in
79 late 2020 and early 2021 offers a unique opportunity to compare variant transmission dynamics
80 from a phylogeographic and spatial-epidemiological context. While phylogeographic analyses
81 are routinely employed to infer the timing and source of SARS-CoV-2 introductions around the
82 world (19–22), spatial statistics offer a complementary approach for examining COVID-19
83 incidence and the spatial spread of SARS-CoV-2 variants. Spatial statistics, including scan
84 statistics for detecting clusters, are often employed in an epidemiological context to assess
85 whether disease incidence is higher in one area than others (23). The discrete Poisson space-time
86 scan statistic is a spatial statistical method that has been adopted to search for geographic areas
87 with high incidence of COVID-19 (24,25). Similarly, the multinomial scan statistic is commonly
88 used for the purposes of detecting geographic areas with unusually high distributions of specific

89 disease or event types. The multinomial scan statistic has been used to clarify categorical
90 distributions in a variety of contexts, including the distribution of meningitis types (26),
91 multiresistance phenotypes of *Staphylococcus aureus* (27), and methods of suicide (28). To the
92 best of our knowledge, the use of multinomial scan statistic to quantitatively assess the
93 categorical distribution of SARS-CoV-2 variants among total COVID-19 cases has not been
94 reported. This study employs spatial scan statistics paired with phylogeographic analyses to
95 describe the shifting SARS-CoV-2 variant landscape in NY from December 2020 to April 2021,
96 particularly the inter-play between B.1.526 and B.1.1.7. We identify the sources of multiple
97 introductions of B.1.1.7 and B.1.526 in NY and track their spatiotemporal spread relative to
98 oneanother. As evidenced by the rapid dominance of the highly transmissible VOC Delta in
99 many locations around the world in the summer of 2021 (GISAID.org), inter-variant competition
100 will likely play a major role in the future public health response and epidemiology of COVID-
101 19. Our findings elucidate the dynamics of competing SARS-CoV-2 variants at a time when the
102 future variant landscape remains uncertain.

103

104 Results:

105 A total of 8,517 SARS-CoV-2 specimens sequenced by Wadsworth with collection dates
106 between December 1, 2020 and April 30, 2021 were included in the study. Sequenced specimens
107 equated to 0.6% of new COVID-19 cases in the study period. Among the included specimens,
108 B.1.1.7 and B.1.526 constituted 1,107 (13%) and 904 (10.6%) of the samples, respectively.
109 Formerly defined parent lineage B.1.526 and sublineages B.1.526.1, B.1.526.2, and B.1.526.3
110 constituted 98, 388, 415, and 2 samples, respectively. The average patient age for all samples

111 was 46.5 years, which decreased from 49.6 years in December to 39.5 years in April. Patient age
112 and week of sample collection exhibited a negative association during the study ($\beta=0.765$, $t=$
113 16.31 , $p<0.0001$). The mean age of sampled patients also differed significantly between variants
114 ($df=2$, $F=81.39$, $p<0.0001$), where mean ages for B.1.1.7, B.1.526, and all other lineages were
115 39.3, 44.6, and 48.5 years, respectively.

116 The earliest B.1.1.7 samples sequenced by Wadsworth were collected on December 24,
117 2020 from a resident of Manhattan (Metro region) and an individual in Saratoga County (Capital
118 region). B.1.1.7 remained relatively rare among all samples through the end of January. The
119 Metro and Capital regions experienced the earliest increases in B.1.1.7, although the proportion
120 of B.1.1.7 did not exceed 15% through February. The proportion of B.1.1.7 increased in March
121 across all regions, most notably in the Western region where it constituted around 75% of all
122 samples by the end of March and continued to rise through April. The Metro region experienced
123 the most gradual increase in B.1.1.7, with the proportion not exceeding 40% until the end of
124 April.

125 The earliest B.1.526 sample sequenced by Wadsworth was collected on December 9,
126 2020 from a patient in the Bronx (Metro region). The proportion of B.1.526 increased in the
127 Metro region throughout December, reaching 10% of total samples by the end of the month. The
128 proportion of B.1.526 in the Metro region approached 40% by the end of January, peaked at
129 ~60% in mid-February to early March, and then plateaued around 50% through April. In
130 contrast, B.1.526 was not consistently detected in the other regions until February, with increases
131 after February in the Capital and Central regions, but the proportion generally remained under
132 40%. The Western region saw a minimal increase in B.1.526 after February, with the proportion

133 remaining under 15% through April. The combined proportion of all lineages other than B.1.1.7
134 and B.1.526 dropped below 20% in all NY regions by the end of April.

135 Maps of interpolated proportion of B.1.1.7 relative to all other lineages by ZCTA (Figure
136 1A) show a general trend of spread through the southern portion of NY in January, statewide
137 distribution by February, diffuse increase in proportion in March, and a sustained high
138 proportion throughout the state in April, with strong dominance in the Western region. In
139 contrast, maps of interpolated proportion of B.1.526 shows more constricted initial spread
140 focused around NYC and surrounding areas in January, with statewide distribution not achieved
141 until March, and a moderate proportion sustained mostly within the Metro region, without
142 notable increase in proportion from March to April (Figure 1B).

143 Maps of geographic mean centers of estimated B.1.1.7 and B.1.526 cases (Figure 2) show
144 that shifts in the SARS-CoV-2 variant landscape had an impact on the spatial distribution of
145 overall cases of COVID-19. The mean center of B.1.526 cases initially occurred near the NYC
146 area, then gradually moved slightly northwest as B.1.526 expanded modestly into upstate
147 regions. The mean center of B.1.1.7 cases was also in the NYC area at the beginning of the study
148 period, then moved northwest during March and April to a much greater degree than for B.1.526
149 cases, indicating more substantial expansion of B.1.1.7 than B.1.526 outside of NYC. The mean
150 centers of all COVID-19 cases showed a southeastern trajectory during December through
151 March, but then exhibited a large northwestern shift in April, suggesting that the spread of
152 B.1.1.7 in Upstate NY, especially within the Western region, resulted in a spatial shift of overall
153 COVID-19 cases.

154 Retrospective multinomial space-time scan analysis indicated six statistically significant
155 clusters with elevated relative risk (RR) of COVID-19 attributable to specific variants (Figure 3
156 and Table 1S). Two clusters of elevated RR of “Other” lineages were found in December, 2020
157 in the Metro/Capital regions as well as Long Island, reflecting the nearly non-existent risk for
158 B.1.1.7 and B.1.526. Of note, these clusters are likely to be limited in spatial extent due to
159 setting the maximum cluster size to 10% of the population-at-risk, as the presence of B.1.1.7 and
160 B.1.526 were also nearly non-existent statewide in December (Figure 1). Three clusters of
161 elevated RR of multiple combinations of B.1.1.7, B.1.526 parent lineage, B.1.526.1, and
162 B.1.526.2 were found in March, 2021 in the NYC/Long Island region. The sixth cluster exhibited
163 an elevated RR of greater than 7.0 for B.1.1.7, with a radius of 114.38 km centered in the Finger
164 Lakes area (Western and Central regions) during April. Additionally, the presence of clusters of
165 B.1.1.7 and B.1.526 in March and April coincide with a general statewide decrease in incidence
166 of COVID-19 (Figure 3).

167 The final B.1.526 dataset for phylogenetic reconstruction contained 980 genomes from
168 all regions of NY and various domestic locations (Bronx: 222, Hudson Valley: 128, Brooklyn:
169 39, Long Island: 78, Manhattan: 49, Queens: 81, Staten Island: 12, Upstate NY: 81, Domestic:
170 290). The final B.1.1.7 dataset contained 1,195 genomes from the NYC region (181), Finger
171 Lakes (239), Hudson Valley (78), Long Island (130), Western NY and the Southern Tier
172 (Southwestern NY: 56), Capital District, Mohawk Valley, Central NY and the North Country
173 (Northern NY: 149), as well as other states (Domestic: 362). Results from the phylogeographic
174 analysis indicated that B.1.526 emerged within the Bronx near the end of 2020, and that this
175 location was the major source of spread to other regions of NY and US (Domestic) in the

176 ensuing months (Figure 4A). Although sampling biases could have influenced the number of
177 introductions assigned to the Bronx, the Domestic category had greater representation in the
178 dataset but led to substantially fewer introductions (Table S2). Specifically, Domestic genomes
179 represented 29.5% of the dataset but this location was responsible for only 6.7% of all B.1.526
180 introductions, while the Bronx represented 22.7% of the dataset and led to 63.8% of all
181 introductions (Table S2). Excluding the Bronx, B.1.526 transmission between boroughs and
182 from these boroughs to other NY or Domestic locations was relatively infrequent. Further,
183 introductions were mostly uni-directional, with only 9% of the B.1.526 cases sequenced from the
184 Bronx due to re-introductions. This indicates the spread of B.1.526 in the Bronx was
185 overwhelmingly attributable to transmission within the Bronx itself and contrasts the trends for
186 other NY and Domestic locations, which showed more limited sustained transmission of B.1.526
187 as a function of the ratio of number of introductions to sample size.

188 Multiple domestic introductions contributed to the initial presence of B.1.1.7 in NY (29),
189 with transmission occurring most frequently to the Finger Lakes and Northern NY (Figure 4B).
190 The Finger Lakes and Northern NY were well-represented in the dataset (20% and 12% of the
191 data, respectively) but contributed substantially less to the distribution of B.1.1.7 (accounting for
192 7% and 6% of the total number of introductions, respectively) than Domestic sites, which
193 represented 20% of the data and were responsible for the majority of introductions (~39%, Table
194 S3). Exchange between NYC and Long Island and NYC and the Hudson Valley was also
195 frequent, but transmission from these regions to Northern NY, Southwestern NY, and the Finger
196 Lakes was substantially more limited (Figure 4B, Table 3), possibly reflecting the effects of
197 interconnectivity in a metropolitan area dominated by commuter activity. The Finger Lakes

198 showed the lowest proportion of sequenced cases due to introductions but the largest sample size
199 in NY, suggesting more sustained transmission of B.1.1.7 within this region. The topologies of
200 the B.1.526 and B.1.1.7 phylogenetic trees highlight differences in transmission patterns as well.
201 Introductions of B.1.526 often represent single events not associated with additional transmission
202 (except for some introductions from the Bronx), which led to B.1.526 genomes from various
203 locations being distributed broadly throughout the tree. In contrast, introductions of B.1.1.7 often
204 led to large regional clusters, such as those in the Finger Lakes from Domestic sources,
205 producing a patchier distribution of genomes from the same region.

206

207 Discussion:

208 The repeated emergence of novel variants of SARS-CoV-2 has largely defined the
209 COVID-19 pandemic response in 2021. As vaccination rates, prior exposure levels, and
210 behavioral public health measures continuously change, so too does the strength of selection
211 (30). Given that selective pressures likely vary across regions, including between and within
212 states, it follows that the emergence and spread of SARS-CoV-2 variants are also regionally
213 dynamic. We combine spatial statistical, phylogeographic, and cartographic visualization
214 techniques to examine the spatiotemporal dynamics of the VOC B.1.1.7 (Alpha) and the VOI
215 B.1.526 (Iota) within NY from December 2020 through April 2021. Our study period captures a
216 time of substantial fluctuations in statewide COVID-19 incidence, with a major peak in late
217 December and early January followed by decline through February, and then a smaller peak in
218 late March and early April followed by a decline through April. The peak incidence in late 2020
219 and early 2021 represents the second major wave of COVID-19 cases in NYC, and the first

220 major wave for much of the rest of the state. Our study period also spans the launch of NY's
221 vaccination campaign in January 2021 and subsequent rise in immunization rates.

222 The concurrent spread of B.1.1.7 and B.1.526 offers a unique opportunity to compare the
223 dynamics of two competing variants of SARS-CoV-2 within a population. Shortly after its
224 appearance in the Bronx in late 2020, B.1.526 quickly became the most common lineage in NYC
225 and the surrounding region by mid-February. The rapid dominance of B.1.526 within NYC is
226 corroborated by time-calibrated phylogeny (Figure 4A), which depicts widespread initial
227 transmission within the Bronx, periodic introductions to neighboring boroughs, and later
228 introductions to the greater Metro region and other states, with few introductions leading to
229 sustained transmission outside the Bronx. The spread of B.1.526 appears to have been spatially
230 limited by the repeated introduction and apparent transmission advantage of B.1.1.7 outside of
231 NYC. Regions of NY where B.1.526 had not yet established experienced rapid dominance of
232 B.1.1.7 during March and April. This trend is most clearly seen in the near complete
233 displacement of all other lineages by B.1.1.7 in Western NY (Supplemental Figure 1A), resulting
234 in a large cluster of elevated RR for B.1.1.7 cases in the Finger Lakes region during April
235 (Figure 3). Phylogeographic analysis provides support for the sustained transmission of B.1.1.7
236 in the Finger Lakes, evidenced by the large regional clusters resolved in the tree. The
237 multinomial spatial scan detected three unique clusters in March 2021, all with increased RR for
238 B.1.1.7 and B.1.526. The values for RR within each NYC cluster details a distinct pattern:
239 clusters centered within the Bronx, Brooklyn, and Manhattan had higher RR for

240 B.1.526, while the cluster centered in east Queens and Long Island had a higher RR for B.1.1.7
241 (Table S1). Over the months following B.1.526's initial advantage in NYC, B.1.1.7 trends
242 towards become the majority variant in the Metro region. Despite the presence of both B.1.1.7
243 and B.1.526 in the Metro region in the early months of 2021, phylogeographic analyses indicated
244 that the parental lineage of B.1.526 was circulating in the Bronx by August 2020, with more
245 frequent transmission to Manhattan than Long Island. While B.1.1.7 was circulating in Long
246 Island by mid-October 2020, almost all B.1.526 introductions to this region occurred in 2021.
247 Thus, there are important distinctions between the spread of B.1.1.7 and B.1.526 even within a
248 geographically proximal area. The delayed dominance of B.1.1.7 in the Bronx compared to Long
249 Island and Queens suggests that B.1.526 was more difficult to displace than other lineages
250 circulating in these areas. The number of B.1.1.7 introductions into NYC was greater than into
251 the Finger Lakes, but sustained transmission was less prevalent, providing further evidence that
252 widespread circulation of B.1.526 in NYC stunted the initial spread of B.1.1.7 in this area.
253 Maps of the geographic mean centers of the estimated number of COVID-19 cases attributable
254 to each variant capture the rapid spread of B.1.1.7 out of NYC, as well as the relative inability of
255 B.1.526 to claim a foothold outside of the Metro region. The northwesterly shift in the trajectory
256 of overall COVID-19 cases in April indicates that the expansion of B.1.1.7, in particular
257 clustering in Western NY, had a measurable influence on the spatial spread of
258 COVID-19 cases overall.

259 There are several limitations of our study which primarily reflect the inherent limitations
260 of our genomic surveillance program. A degree of selection bias exists within our dataset given
261 that specimens were screened by cycle threshold value and were submitted by a selected group

262 of clinical and commercial labs that cannot perfectly represent all COVID-19 cases in NY. We
263 were unable to assess the demographic and clinical representativeness of our dataset because
264 these data were not available to us for many specimens. Additionally, the number of specimens
265 sequenced varied over the space and time of the study period, which created small sample sizes
266 within many ZCTA-months. This limitation extended to the multinomial scan statistic, which
267 was run with estimated values for COVID-19 cases attributable to B.1.1.7 and B.1.526, giving
268 all ZCTAs with samples equal weight. However, the spatial scan assesses data according to their
269 proximity to each other. In this context, ZCTAs are analyzed together rather than individually,
270 which has the potential to reduce bias. Another consequence of our limited sampling was that
271 our data exhibited zero samples from many ZCTAs for each month. We addressed this by using
272 IDW interpolation of the proportion of B.1.1.7 and B.1.526 sequenced samples at the ZCTA-
273 month level to visualize general patterns of variant proportions over geography.

274 Phylogeographic analyses were hampered by similar limitations: uneven sampling among
275 regions and the lack of global representation in our datasets could lead to incorrect trait
276 assignments. Smaller sample sizes for some regions might have caused an underestimation of
277 their contributions to variant transmission in NY while larger sample sizes might have inflated
278 the number of introductions assigned. However, we believe the results presented here largely
279 capture the transmission dynamics of B.1.526 and B.1.1.7 within NY, as smaller sample sizes
280 were consistent with lower incidence, while larger sample sizes did not always correspond to
281 regions with outsized contributions to the spread of either variant.

282 Our results are in general agreement with those of Petrone et al. 2021, who showed that

283 B.1.1.7 had a higher reproductive value and was able to spread faster than B.1.526 in
284 Connecticut, where both variants were introduced a similar number of times. The reproductive
285 value for B.1.1.7 in NYC was also higher than B.1.526 but the data were noisier, which supports
286 our conclusions that the clear advantage B.1.1.7 showed in Western NY and the Finger Lakes
287 was obfuscated in NYC by the initial dominating presence of B.1.526.

288 Our phylogeographic and spatiotemporal analyses offer a methodology for evaluating the
289 relative transmissibility and competitive advantages of co-circulating SARS-CoV-2 variants. We
290 demonstrate that the emergence of VOI B.1.526 slowed the rise of VOC B.1.1.7 within NYC,
291 while B.1.1.7 simultaneously became the majority variant in parts of NY devoid of B.1.526. In
292 this way, our study describes important dynamic interactions between variants with unequal
293 transmissibility and is potentially generalizable to interactions between any known variant and
294 the highly transmissible B.1.617.2 (Delta) variant and other variants to come. As B.1.617.2
295 continues to change the variant landscape on local and global levels, understanding its dynamic
296 interactions with other variants is increasingly important in the management of COVID-19.

297

298 Methods:

299 Sample acquisition and RNA extraction

300 The NYSDOH Wadsworth Center coordinated with over 30 clinical laboratories
301 throughout NY who routinely submitted respiratory swabs positive for SARS-CoV-2 for whole
302 genome sequencing. Specimens were required to have a real-time cycle threshold value less than
303 30. Nucleic acid extraction was performed on a Roche MagNAPure 96 with the Viral NA Small
304 Volume Kit (Roche, Indianapolis, IN) with 100 μ L sample input and 100 μ L eluate.

305 Sequencing and bioinformatics processing

306 Extracted RNA was processed for whole genome sequencing with a modified ARTIC
307 protocol (artic.network/ncov-2019) in the Applied Genomics Technology Core at the Wadsworth
308 Center as previously described (29); supplementary information). Illumina libraries were
309 processed with ARTIC nextflow pipelines
310 (github.com/connorlab/ncov2019/articnf/tree/illumina, last updated April 2020) as previously
311 described (Alpert et al., 2021; supplementary information)

312 Sample inclusion criteria

313 Specimens with collection dates between December 1, 2020 and April 30, 2021 were
314 included. Specimens that were sequenced as a result of pre-screening for specific mutations or
315 clinical/epidemiological criteria were removed from the analysis. In the case of duplicate
316 specimens from the same patient, the earliest collected specimen was included, and all other
317 specimens excluded from the analysis. Only specimens with ZIP code of patient address
318 available were included.

319 COVID incidence calculation

320 Monthly COVID case counts by ZIP code were obtained from
321 <https://github.com/nychealth/coronavirus-data> for NYC, and from the NYSDOH Communicable
322 Disease Electronic Surveillance System for the remainder of NY. Reports with case status of
323 ‘confirmed’ or ‘probable’ were included in the case count. Cases were assigned month based on
324 date of diagnosis. All ZIP code data was converted to ZIP code tabulation area (ZCTA).
325 Incidence was calculated using ZCTA-level population data from the 2019 1-year American
326 Community Survey estimates.

327 Retrospective multinomial space-time scan statistic

328 We utilize the retrospective multinomial space-time scan statistic in SaTScan version 9.6,
329 using the non-ordinal method (26,31). Estimated SARS-CoV-2 variant data used in the
330 multinomial scan statistic were calculated for each ZCTA-month aggregation by multiplying the
331 proportion of either B.1.1.7, B.1.526, or “Other” variants in our sample by the total number of
332 COVID-19 cases.

333 Maximum spatial and temporal cluster size parameters were set a priori for 10% of the
334 population at risk (24) and one month, respectively. Space-time cluster detection in SaTScan has
335 a noted limitation where the size of clusters cannot change over time (32,33). Given that our data
336 is aggregated to the temporal unit of months (December 2020 – April 2021), setting the
337 maximum temporal cluster size parameter to one month allows clusters to change their shape
338 from month to month by being designated as “new” clusters. A further description of settings for
339 the retrospective multinomial space-time scan statistic can be found in the supplemental
340 materials.

341
342 Inverse-distance weighted interpolation and spatial average of SARS-CoV-2 genome sequencing
343 We employed inverse-distance weighted (IDW) interpolation to better visualize how the
344 proportion of COVID-19 cases attributable to each SARS-CoV-2 variant spatiotemporally varies
345 in NY (34). The percentage of COVID-19 cases attributable to each variant in a ZCTA was
346 assigned each ZCTA’s corresponding centroid for the IDW calculation. IDW interpolation
347 generated a continuous surface of values representing the percentage of total COVID-19 cases
348 attributed to B.1.1.7 and B.1.526, which were then averaged over each ZCTA geometry.

349 The estimated percentage of each SARS-CoV-2 variant generated from IDW
350 interpolation was then multiplied by the total number of COVID-19 cases for each ZCTA and
351 month of our study period to estimate the total number of COVID-19 cases attributable to each
352 variant. Estimated numbers of variant cases were then used to generate geographic mean
353 centers for each month of the study period for visualization (35). Methods for geographic mean
354 center calculation can be found in the supplemental materials.

355 Phylogeographic analyses

356 All B.1.526 genomes from the United States (US) and associated metadata (excluding
357 NY sequences) were downloaded from GISAID (GISAID.org) and randomly subsampled to
358 approximately equal depth as the heaviest sampled NY region in our dataset, with the number of
359 genomes from each state sampled proportionally to their overall frequency in the US. The final
360 dataset included B.1.526 genomes from MA, NJ, PA, CT, CA, FL, MD, MI, MN, and NC,
361 aggregated as “Domestic”, genomes from the five boroughs of NYC (Bronx, Brooklyn, Queens,
362 Staten Island, Manhattan), Long Island, and the Hudson Valley and those from Western NY, the
363 Finger Lakes, the Capital District, and Central NY regions aggregated as “Upstate”. Genomes
364 were aligned in mafft v7.475 (36) with problematic sites masked according to
365 (https://github.com/W-L/ProblematicSites_SARS-CoV-2). A maximum likelihood phylogeny
366 was generated in IQTree v1.6.12 (37) with 1000 ultrafast bootstrap replicates (38) and
367 timecalibrated in TreeTime v0.7.6 (39). This tree served as the fixed tree for ancestral state
368 reconstruction in Beast v2.6.2 (40) to infer timing and source of B.1.526 introductions within the
369 state of NY. The Bayesian analysis was allowed to run for > 4 million generations and monitored
370 in Tracer until the effective sample size of all parameters \geq 200 and the MCMC chain appeared
371 to reach stationarity.

372 A B.1.1.7 phylogeographic analysis was conducted in the same manner with the states
373 inferred for a fixed topology over 6 million generations in BEAST2 under an exponential
374 coalescent model until all ESS reach ≥ 200 . The final dataset included B.1.1.7 genomes from
375 MA, PA, CT, NJ, CA, and FL grouped together as “Domestic”, the NYC, Long Island,
376 MidHudson and Finger Lakes regions of NY, “Southwest NY” composed of the Southern Tier
377 and Western regions of NY, and “Northern NY” comprised of the Capital District, Mohawk
378 Valley,
379 Central NY, and the North Country.

380 Maximum clade credibility trees for B.1.526 and B.1.1.7 were generated in
381 TreeAnnotator v.2.6.2 (40) with a 10% burn-in. The number of introductions between locations
382 was summarized by Baltic (<https://github.com/evogytis/baltic>) by adopting the exploded tree
383 script for Python 3. Only introductions with a posterior probability of $0.7 \geq$ were considered.
384 Trees were visualized in FigTree v1.5.5 (<http://tree.bio.ed.ac.uk/software/figtree/>) and ggtree
385 (41) for R v4.1.0 (<http://www.R-project.org>) (see supplementary information for additional
386 details).

387
388 Figure captions:

389
390 Figure 1: Inverse distance weighted interpolations of A) percent of B.1.1.7 and B) percent of
391 B.1.526 relative to all other lineages by ZIP code tabulation area.

392
393 Figure 2: Geographically weighted mean centers of total COVID-19 cases and estimated
394 COVID-19 cases attributable to B.1.526 and B.1.1.7, December 2020 through April 2021.

395 Cluster centroids refer to the results of the multinomial space-time scan analysis (Figure 3).

396

397 Figure 3: SARS-CoV-2 variant clusters identified from retrospective multinomial space-time
398 scan analysis and COVID-19 incidence by ZIP code tabulation area, December 2020 through
399 April 2021.

400

401 Figure 4:

402 A) Time-calibrated phylogeny of SARS-CoV-2 variant B.1.526. Left panel represents a
403 maximum likelihood phylogeny of 908 genomes from New York and other states
404 generated in IQ-tree with timescale inferred by TreeTime and ancestral state
405 reconstruction performed in BEAST. Faceted panels indicate the source of B.1.526
406 introductions into different region of New York or other states (Domestic). Only
407 introductions supported by an ancestral state probability of ≥ 0.7 are shown. Right panel
408 shows locations sampled and sample sizes.

409

408 B) Time-calibrated phylogeny of SARS-CoV-2 variant B.1.1.7. Left panel represents a
409 maximum likelihood phylogeny of 1,195 genomes from New York and other states
410 generated in IQ-Tree with timescale inferred by TreeTime and ancestral state
411 reconstruction performed in BEAST. The tree was rooted with a P.1 genomes (not
412 shown). Faceted panels indicate the source of B.1.1.7 introductions into different regions
413 of New York and to other states (Domestic). Only introductions supported by an
414 ancestral state probability of ≥ 0.7 are shown. Right panel shows locations
sampled and sample
415 sizes.
416
417

418 IRB Approval

419 This work was approved by the New York State Department of Health Institutional Review
420 Board, under study numbers 02-054 and 07-022.

421

422

423 Acknowledgements

424 The authors gratefully acknowledge the Advanced Genomic Technologies Core of the
425 Wadsworth Center where all next generation sequencing was performed. We also graciously
426 thank the New York State clinical laboratories that submitted SARS-CoV-2 positive specimens
427 to Wadsworth for sequence analysis, all originating and submitting laboratories for their SARS-
428 CoV-2 sequence contributions to the GISAID database, Wadsworth Center's Virology
429 Laboratory for initial processing of specimens and the Bioinformatics Core for sequence
430 processing and analysis.

431

432

433 References:

- 434 1. Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M. Public health actions to
435 control new SARS-CoV-2 variants. *Cell*. 2021 Mar;184(5):1127–32.
- 436 2. Lauring AS, Hodcroft EB. Genetic Variants of SARS-CoV-2—What Do They Mean?
437 *JAMA*. 2021 Feb 9;325(6):529.
- 438 3. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al. Genomic
439 characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London,
440 UK: a whole-genome sequencing and hospital-based cohort study. *The Lancet Infectious
441 Diseases*. 2021 Apr;S1473309921001705.
- 442 4. Haynes WA, Kamath K, Lucas C, Shon J, Iwasaki A. Impact of B.1.1.7 variant mutations
443 on antibody recognition of linear SARS-CoV-2 epitopes [Internet]. *Infectious Diseases
444 (except HIV/AIDS)*; 2021 Jan [cited 2021 Jun 18]. Available from:

- 445 <http://medrxiv.org/lookup/doi/10.1101/2021.01.06.20248960>
- 446 5. Kidd M, Richter A, Best A, Cumley N, Mirza J, Percival B, et al. S-Variant SARS-CoV-2
447 Lineage B.1.1.7 Is Associated With Significantly Higher Viral Load in Samples Tested by
448 TaqPath Polymerase Chain Reaction. *The Journal of Infectious Diseases*. 2021 May
449 28;223(10):1666–70.
- 450 6. Planas D, Bruel T, Grzelak L, Guivel-Benhassine F, Staropoli I, Porrot F, et al. Sensitivity
451 of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nat*
452 *Med*. 2021 May;27(5):917–24.
- 453 7. Peters MH, Bastidas O, Kokron DS, Henze CE. Transformations, Lineage Comparisons,
454 and Analysis of Down to Up Protomer States of Variants of the SARS-CoV-2 Prefusion
455 Spike Protein Including the UK Variant B.1.1.7 [Internet]. *Biophysics*; 2021 Feb [cited
456 2021 Jun 18]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2021.02.09.430519>
- 457 8. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, et al.
458 SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma.
459 *Nat Med*. 2021 Apr;27(4):622–5.
- 460 9. Tian F, Tong B, Sun L, Shi S, Zheng B, Wang Z, et al. Mutation N501Y in RBD of Spike
461 Protein Strengthens the Interaction between COVID-19 and its Receptor ACE2 [Internet].
462 *Biophysics*; 2021 Feb [cited 2021 Jun 18]. Available from:
463 <http://biorxiv.org/lookup/doi/10.1101/2021.02.14.431117>
- 464 10. Meng B, Kemp SA, Papa G, Datir R, Ferreira IATM, Marelli S, et al. Recurrent emergence
465 of SARS-CoV-2 spike deletion H69/V70 and its role in the variant of concern lineage
466 B.1.1.7. *Cell Reports*. 2021 Jun;109292.
- 467 11. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Transmission of
468 SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and
469 genetic data [Internet]. *Infectious Diseases (except HIV/AIDS)*; 2021 Jan [cited 2021 Jan
470 12]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.12.30.20249034>
- 471 12. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al.
472 Emergence and rapid spread of a new severe acute respiratory syndrome-related
473 coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa
474 [Internet]. *Epidemiology*; 2020 Dec [cited 2021 Jun 15]. Available from:
475 <http://medrxiv.org/lookup/doi/10.1101/2020.12.21.20248640>
- 476 13. Lasek-Nesselquist E, Lapierre P, Schneider E, George KSt, Pata J. The localized rise of a
477 B.1.526 SARS-CoV-2 variant containing an E484K mutation in New York State [Internet].
478 *Epidemiology*; 2021 Mar [cited 2021 Jul 26]. Available from:
479 <http://medrxiv.org/lookup/doi/10.1101/2021.02.26.21251868>
- 480 14. West AP, Wertheim JO, Wang JC, Vasylyeva TI, Havens JL, Chowdhury MA, et al.

- 481 Detection and characterization of the SARS-CoV-2 lineage B.1.526 in New York
482 [Internet].
483 Bioinformatics; 2021 Feb [cited 2021 Jun 18]. Available from:
484 <http://biorxiv.org/lookup/doi/10.1101/2021.02.14.431043>
485 15. World Health Organization. Epidemiological update: Variants of SARS-CoV-2 in the
486 Americas. [Internet]. 2021 Mar. Available from: <https://iris.paho.org/handle/10665.2/53382>
- 487 16. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al.
488 SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*. 2021
489 Jul;19(7):409–24.
- 490 17. Hodcroft EB, Zuber M, Nadeau S, Crawford KHD, Bloom JD, Veessler D, et al. Emergence
491 and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv*.
492 2020 Jan 1;2020.10.25.20219063.
- 493 18. Thompson CN, Hughes S, Ngai S, Baumbartner J, Wang JC, McGibbon E, et al. Rapid
494 Emergence and Epidemiologic Characteristics of the SARS-CoV-2 B.1.526 Variant - New
495 York City, New York, January 1 - April 5, 2021. *MMWR*. 2021 May 14;70(19):712–6.
- 496 19. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et
497 al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science*.
498 2020 Jul 17;369(6501):297.
- 499 20. Maurano MT, Ramaswami S, Zappile P, Dimartino D, Boytard L, Ribeiro-dos-Santos AM,
500 et al. Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York
501 City Region. *medRxiv*. 2020 Jan 1;2020.04.15.20064931.
- 502 21. Rambaut, Andrew. Phylodynamic Analysis | 176 genomes [Internet].
503 <https://virological.org>.
504 2020 [cited 2020 Nov 17]. Available from: [https://virological.org/t/phylodynamic-](https://virological.org/t/phylodynamic-analysis176-genomes-6-mar-2020/356)
505 [analysis176-genomes-6-mar-2020/356](https://virological.org/t/phylodynamic-analysis176-genomes-6-mar-2020/356)
- 506 22. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, et al. The emergence of
507 SARS-CoV-2 in Europe and North America. *Science*. 2020 Oct 30;370(6516):564.
- 508 23. Kulldorff M, Nagarwalla N. Spatial disease clusters: Detection and inference. *Statist Med*.
509 1995 Apr 30;14(8):799–810.
- 510 24. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United
511 States using a prospective space-time scan statistic: Detecting and evaluating emerging
512 clusters. *Applied Geography*. 2020 May;118:102202.
- 513 25. Rosillo N, Del-Águila-Mejía J, Rojas-Benedicto A, Guerrero-Vadillo M, Peñuelas M,
514 Mazagatos C, et al. Real time surveillance of COVID-19 space and time clusters during the
515 summer 2020 in Spain. *BMC Public Health*. 2021 Dec;21(1):961.

- 516 26. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. *Statist Med.*
517 2010 Aug 15;29(18):1910–8.
- 518 27. Stelling J, Read JS, Peters R, Clark A, Bokhari M, O’Brien TF. Staphylococcus aureus
519 antimicrobial susceptibility trends and cluster detection in Vermont: 2012-2018. *Expert*
520 *Review of Anti-infective Therapy.* 2021 Jun 3;19(6):777–85.
- 521 28. Ngamini Ngui A, Apparicio P, Moltchanova E, Vasiliadis H-M. Spatial analysis of suicide
522 mortality in Québec: Spatial clustering and area factor correlates. *Psychiatry Research.*
523 2014 Dec;220(1–2):20–30.
- 524 29. Alpert T, Brito AF, Lasek-Nesselquist E, Rothman J, Valesano AL, MacKay MJ, et al.
525 Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States.
526 *Cell.* 2021 May 13;184(10):2595-2604.e13.
- 527 30. Neher R. The virus is under increasing selection pressure [Internet]. 2021 [cited 2021 Aug
528 10]. Available from: <https://www.mpg.de/16371358/coronavirus-variants>
- 529 31. Kulldorff M. Software for the spatial and space-time scan statistics [Internet]. Information
530 Management Services, Inc.; 2018. Available from: <http://www.satscan.org/>
- 531 32. Iyengar VS. Space-time clusters with flexible shapes. *MMWR Suppl.* 2005 Aug 26;54:71–
532 6.
- 533 33. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for
534 disease outbreak detection and monitoring. *Int J Health Geogr.* 2008;7(1):14.
- 535 34. Shepard D. A two-dimensional interpolation function for irregularly-spaced data. In:
536 *Proceedings of the 1968 23rd ACM national conference on -* [Internet]. Not Known: ACM
537 Press; 1968 [cited 2021 Jun 8]. p. 517–24. Available from:
538 <http://portal.acm.org/citation.cfm?doid=800186.810616>
- 539 35. Rogerson PA, Plane DA. *Geographical Analysis of Population: With Applications to*
540 *Planning and Business.* Internat.ed. John Wiley and Sons Ltd; 1994. 416 p.
- 541 36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
542 improvements in performance and usability. *Mol Biol Evol.* 2013 Apr;30(4):772–80.
- 543 37. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
544 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol.*
545 2015 Jan 1;32(1):268–74.
- 546 38. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast Approximation for Phylogenetic
547 Bootstrap. *Molecular Biology and Evolution.* 2013 Feb 15;30(5):1188–95.
- 548 39. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis.
549 *Virus Evolution* [Internet]. 2018 Jan 8 [cited 2020 Jul 15];4(vex042). Available from:
550 <https://doi.org/10.1093/ve/vex042>

- 551 40. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et
552 al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS
553 Computational Biology. 2019 Apr 8;15(4):e1006650.

552 41. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and
553 annotation of phylogenetic trees with their covariates and other associated data. *Methods in* 554
Ecology and Evolution. 2017 Jan 1;8(1):28–36.

555

Figure 1A

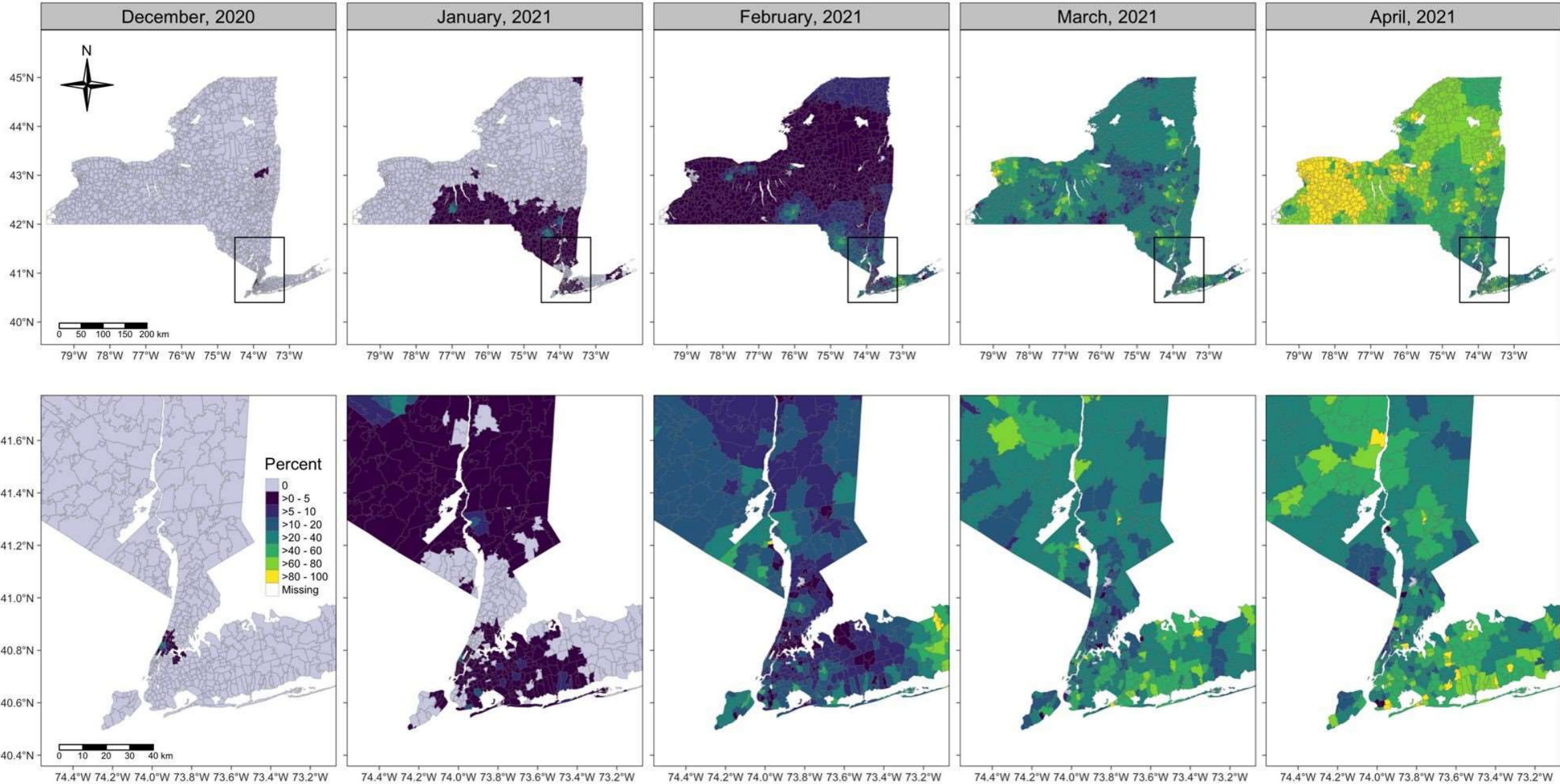


Figure 1B

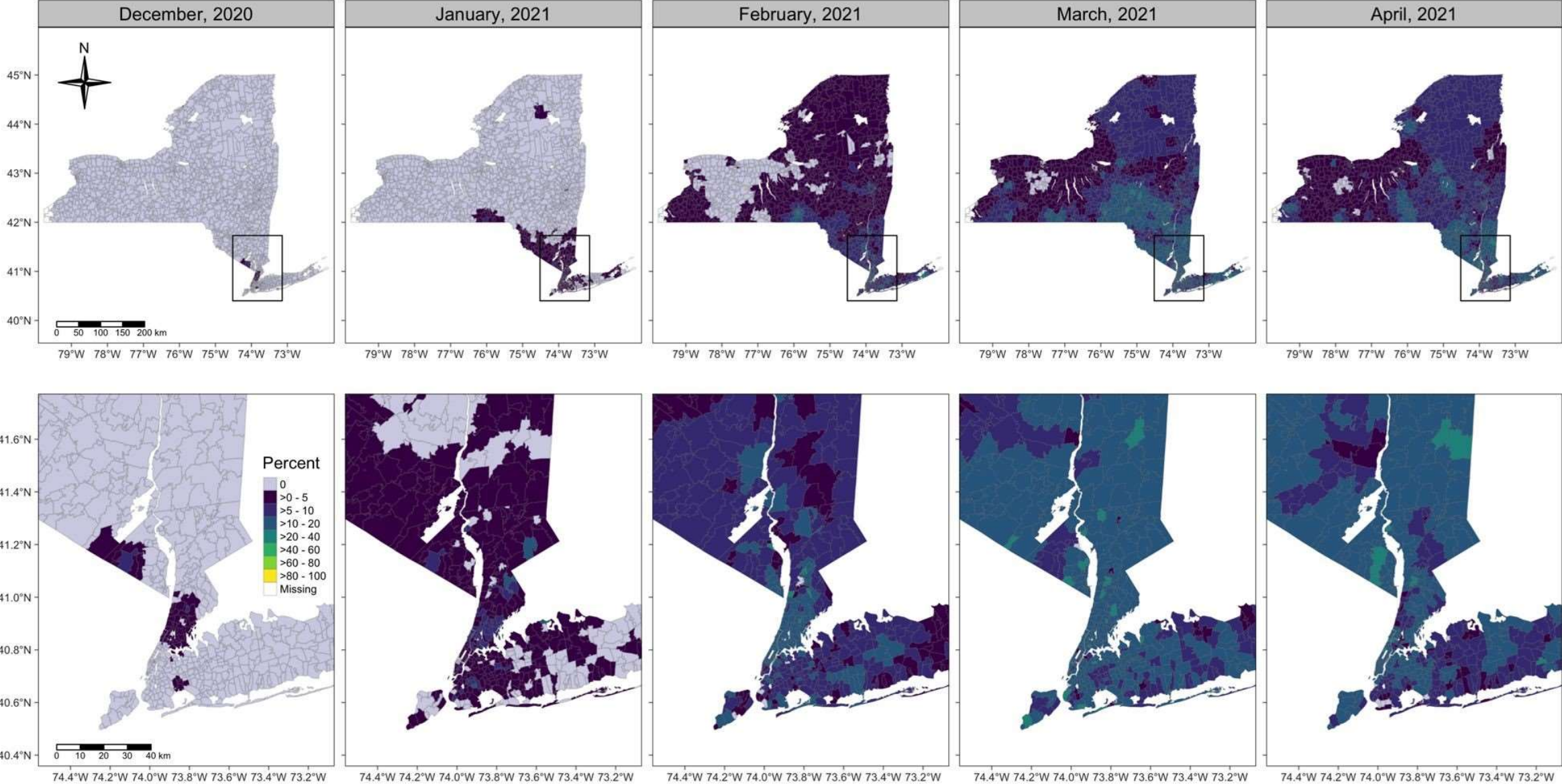


Figure 2

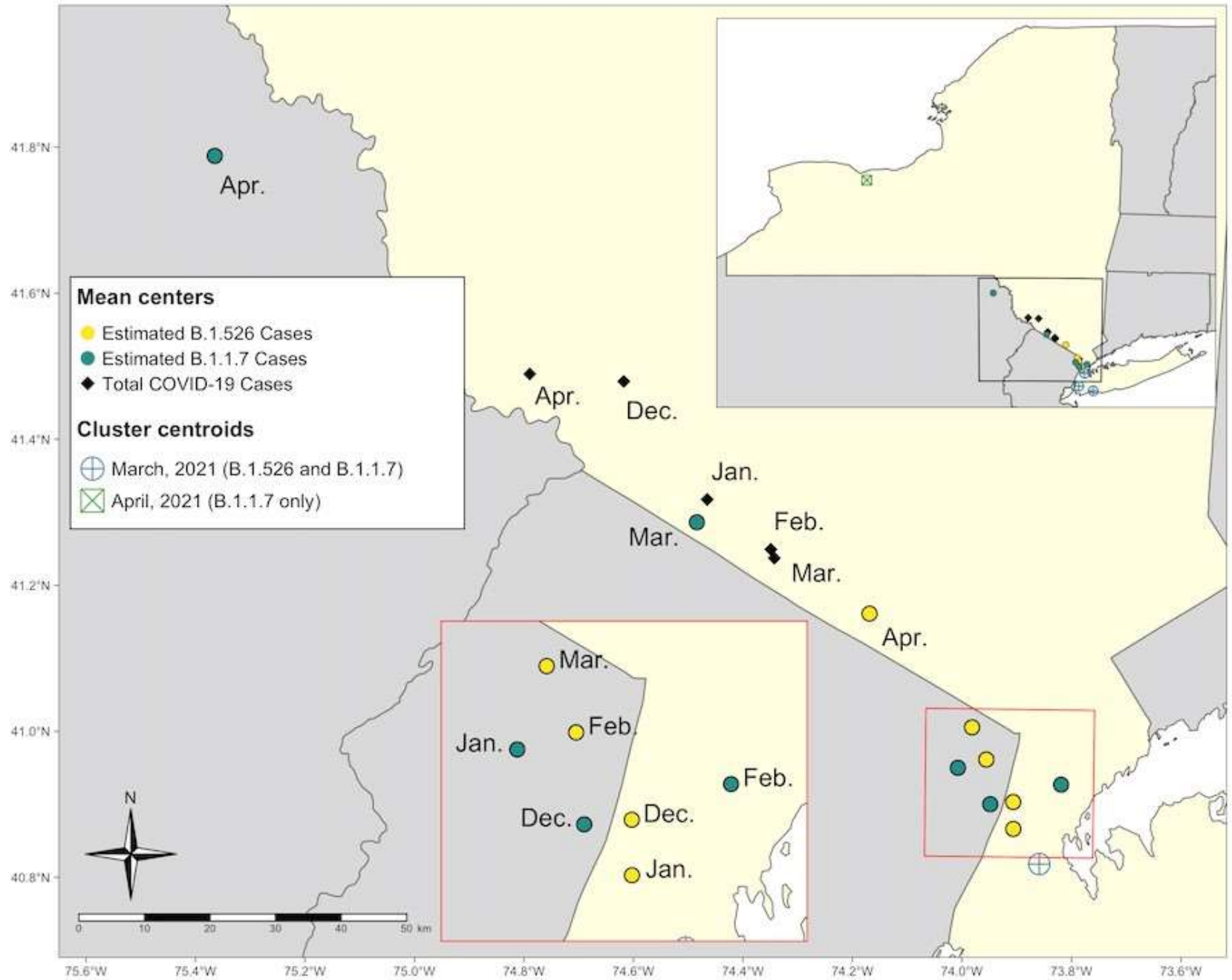


Figure 3

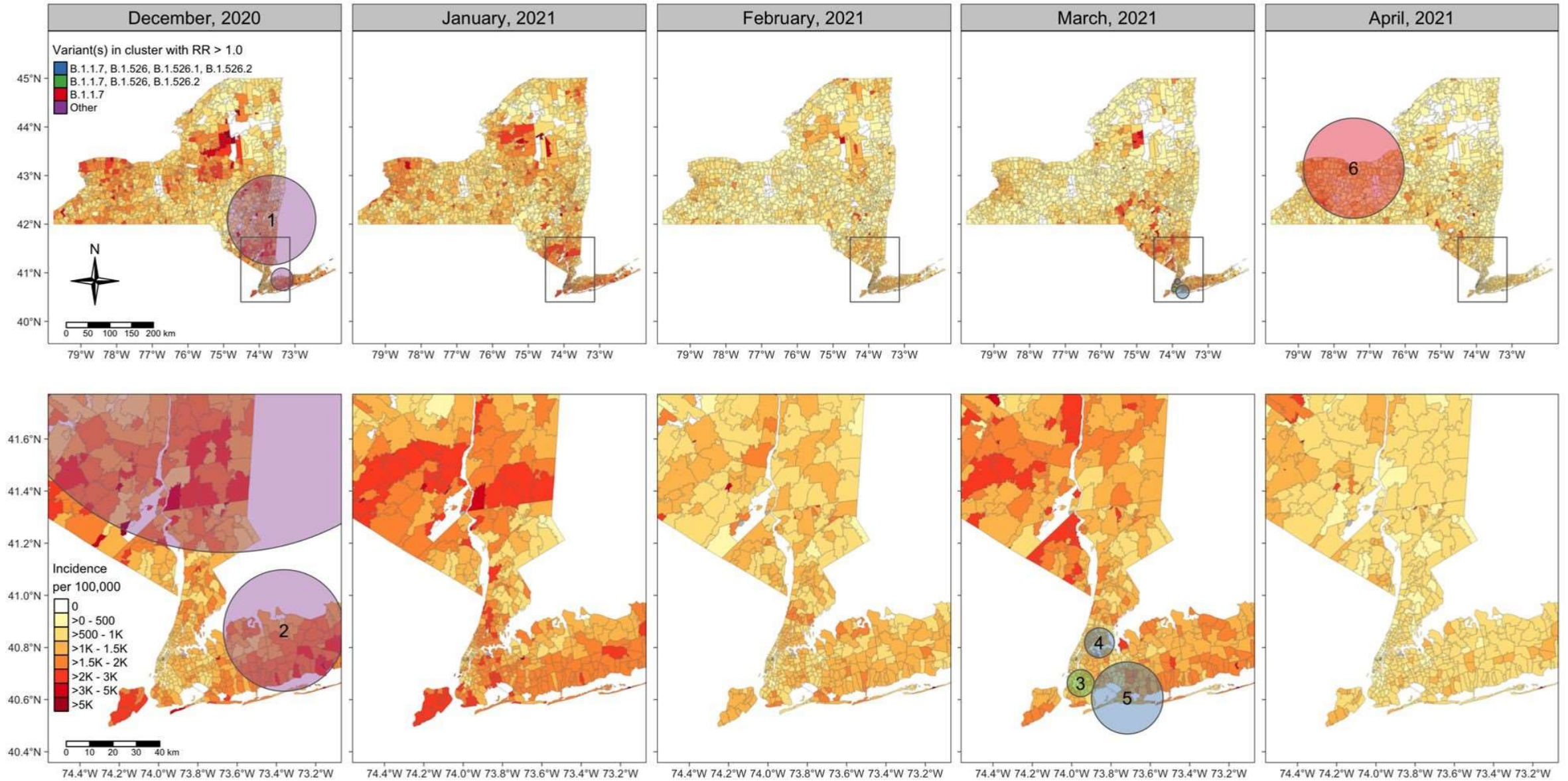


Figure 4A

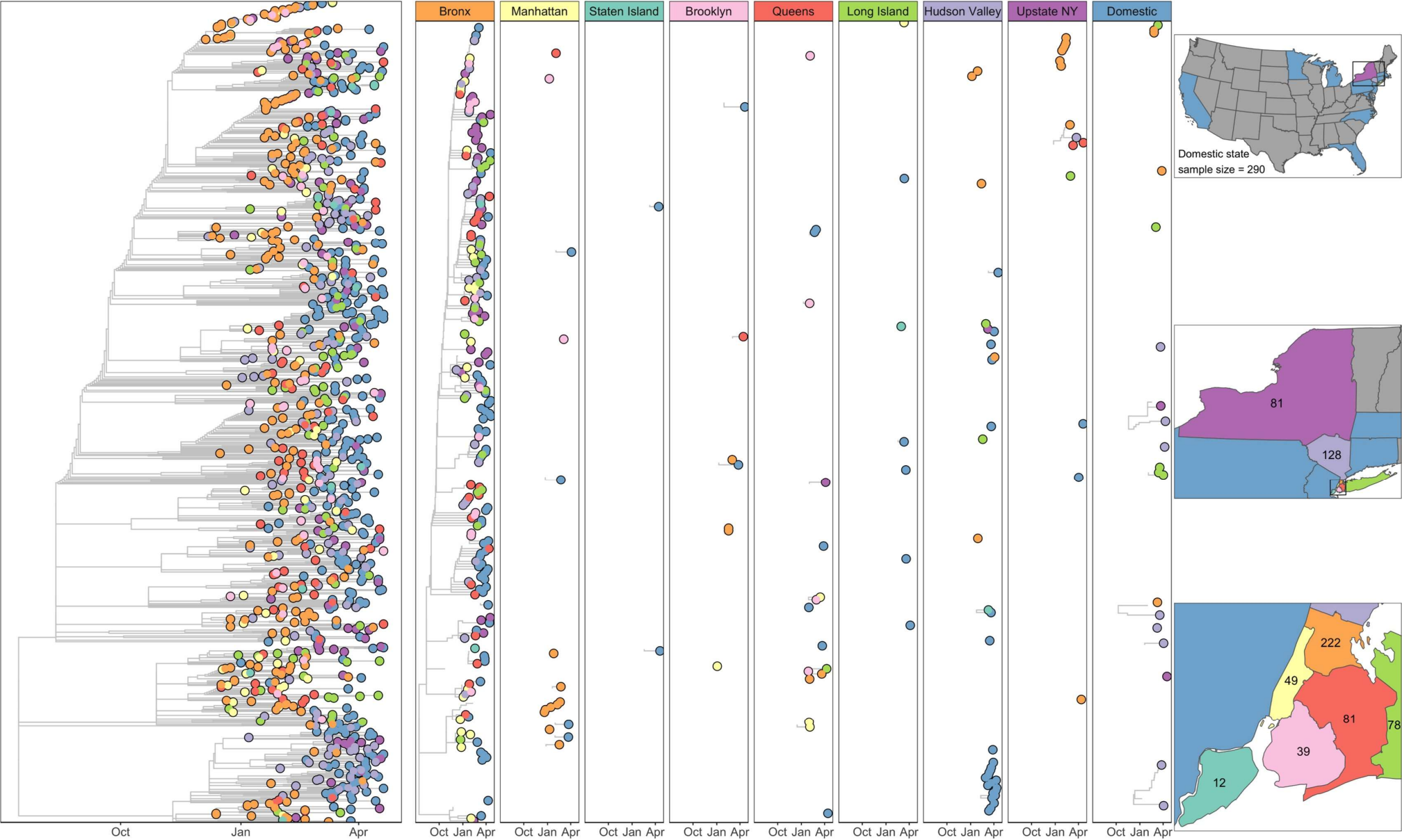


Figure 4B

