Spatiotemporal Analyses of 2 Co-Circulating SARS-CoV-2 Variants, New York, USA

Alexis Russell,[1] Collin O'Connor,[1] Erica Lasek-Nesselquist,[1]
Jonathan Plitnick, John P. Kelly, Daryl M. Lamson, Kirsten St. George

The emergence of novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants in late 2020 and early 2021 raised alarm worldwide because of their potential for increased transmissibility and immune evasion. Elucidating the evolutionary and epidemiologic dynamics among novel SARS-CoV-2 variants is essential for understanding the trajectory of the coronavirus disease pandemic. We describe the interplay between B.1.1.7 (Alpha) and B.1.526 (Iota) variants in New York State, USA, during December 2020–April 2021 through phylogeographic analyses, space-time scan statistics, and cartographic visualization. Our results indicate that B.1.526 probably evolved in New York City, where it was displaced as the dominant lineage by B.1.1.7 months after its initial appearance. In contrast, B.1.1.7 became dominant earlier in regions with fewer B.1.526 infections. These results suggest that B.1.526 might have delayed the initial spread of B.1.1.7 in New York City. Our combined spatiotemporal methodologies can help disentangle the complexities of shifting SARS-CoV-2 variant landscapes.

The emergence of a novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variant B.1.1.7 (Alpha) in the United Kingdom in late 2020 raised alarm worldwide and prompted major reassessment of the management, surveillance, and projected future of coronavirus disease (COVID-19) (*1*,*2*). Evidence of increased transmissibility and potential immune evasion prompted the World Health Organization to designate B.1.1.7 a variant of concern (VOC) in December 2020 (*3*–*5*; W.A. Haynes et al., unpub. data, https://doi.org/10.1101/2021.01.06.20248960). The emergence of B.1.1.7 and additional novel SARS-CoV-2 variants with competitive advantages has resulted in the localized dominance of single variants (E. Volz et al., unpub. data, https://doi.org/10.1101/2020.12.30.20249034) and raised concern for increases in COVID-19 incidence (*6*).

Novel variant B.1.526 (Iota) arose within New York State (NYS), USA, in late 2020 (E. Lasek-Nesselquist et al., unpub. data, https://doi.org/10.1101/2021.02.26.21251868) (*7*) and quickly increased in proportion throughout the state, leading to a noticeable shift in lineage distribution during early 2021 (E. Lasek-Nesselquist et al., unpub. data, https://doi.org/10.1101/2021.02.26.21251868) (*31*). The World Health Organization designated B.1.526 as a variant of interest (VOI) because of its increase in prevalence coupled with mutations associated with immune evasion (*8*). Despite these concerns, an epidemiologic assessment of B.1.526 in NYC during January–April 2021 found that the lineage did not cause more severe disease and was not associated with increased risk for reinfection or vaccine breakthrough (*9*). However, an epidemiologic study of NYS during late 2020–May 2021 concluded that B.1.526 was 35% more transmissible than non-VOC and non-VOI lineages circulating at the time (*10*).

Genomic surveillance of COVID-19 is a crucial tool to monitor and assess the physiologic and epidemiologic characteristics of SARS-CoV-2 variants as they emerge. The New York State Department of Health (NYSDOH) substantially expanded its genomic surveillance program in December 2020, with the aim of sequencing a more representative subset of COVID-19 cases across the state to track the spread and impact of novel variants. A robust genomic surveillance system enables assessment of changes in variant distribution over precise temporal and spatial scales.

Author affiliations: New York State Department of Health, Albany, New York, USA (A. Russell, C. O'Connor, E. Lasek-Nesselquist, J. Plitnick, J.P. Kelly, K. St. George); State University of New York at Buffalo, Buffalo, New York, USA (C. O'Connor); State University of New York at Albany, Albany (E. Lasek-Nesselquist, K. St. George)

DOI: https://doi.org/10.3201/eid2803.211972

[1]These authors contributed equally to this article.

This study employed spatial scan statistics paired with phylogeographic analyses to describe the shifting SARS-CoV-2 variant landscape in NYS during December 2020–April 2021, specifically the interplay between co-circulating B.1.526 and B.1.1.7 lineages. Our findings elucidate the dynamics of competing SARS-CoV-2 variants at a time when the highly transmissible VOC Delta had overtaken B.1.1.7 worldwide and future variant displacements were likely to occur.

Methods

Sample Acquisition and RNA Extraction

This study was approved by the NYSDOH Institutional Review Board, under study numbers 02-054 and 07-022. The NYSDOH Wadsworth Center coordinated with >30 clinical laboratories throughout NYS that routinely submitted respiratory swabs positive for SARS-CoV-2 for whole-genome sequencing (WGS). Specimens were required to have a real-time cycle threshold value <30. We performed nucleic acid extraction on a MagNAPure 96 with the Viral NA Small Volume Kit (Roche, https://www.roche.com) with 100 µL sample input and 100 µL eluate.

Sequencing and Bioinformatics Processing

We processed extracted RNA for WGS with a modified ARTIC V3 protocol (https://artic.network/ncov-2019) in the Applied Genomics Technology Core at the Wadsworth Center as previously described (*11*) (Appendix, https://wwwnc.cdc.gov/EID/article/28/3/21-1972-App1.pdf). We processed Illumina libraries with the ARTIC nextflow pipeline (https://github.com/connor-lab/ncov2019-artic-nf) as previously described (*14*) (Appendix).

Sample Inclusion Criteria

We included specimens with collection dates during December 2020–April 2021 with ZIP codes of patient addresses. We removed specimens that were prescreened for specific mutations or for clinical or epidemiologic criteria. For persons with multiple specimens collected, we included only the earliest specimen.

COVID Incidence Calculation

We obtained monthly COVID-19 case counts by ZIP code from online NYC COVID-19 data (https://github.com/nychealth/coronavirus-data) and from the NYSDOH Communicable Disease Electronic Surveillance System. We included reports with case status of confirmed or probable in the case count and assigned a month on the basis of diagnosis date. We converted ZIP code data to ZIP code tabulation area (ZCTA) and calculated incidence using population data from the 2019 1-year American Community Survey estimates (https://data.census.gov/cedsci).

Retrospective Multinomial Space-Time Scan Statistic

We used the retrospective multinomial space-time scan statistic in SaTScan version 9.6 and applied the nonordinal method (*12*,*13*). We calculated estimated SARS-CoV-2 variant data for each ZCTA-month aggregation by multiplying the proportion of either B.1.1.7, B.1.526, or other variants in our sample by the total number of COVID-19 cases.

We set the maximum cluster size parameter a priori to 10% of the population at risk (*14*) . Space-time cluster detection in SaTScan has a noted limitation where the size of clusters cannot change over time (*15*,*16*). Given that our data are aggregated to the temporal unit of months (December 2020–April 2021), we set the maximum temporal cluster size parameter to 1 month, to enable clusters to change their shape from month to month by being designated as new clusters (Appendix).

Inverse-Distance Weighted Interpolation and Spatial Average of SARS-CoV-2 Whole-Genome Sequencing

We used inverse-distance weighted (IDW) interpolation to visualize the spatiotemporal variation in the proportion of COVID-19 cases attributable to each SARS-CoV-2 variant in NYS and to provide estimates for these proportions in areas where we were missing data (*17*). We assigned the percentage of COVID-19 cases attributable to each variant per ZCTA to the ZCTA's centroid for the IDW calculation. IDW interpolation generated a continuous surface of values representing the percentage of total COVID-19 cases attributed to B.1.1.7 and B.1.526, which we then averaged over each ZCTA geometry.

We then multiplied the estimated percentage of each SARS-CoV-2 variant generated from IDW interpolation by the total number of COVID-19 cases for each ZCTA and month to estimate the total number of COVID-19 cases attributable to each variant. Estimated numbers of variant cases generated geographic mean centers for each month of the study period (*18*) (Appendix).

Phylogeographic Analyses

We incorporated into the analysis all NYS B.1.526 genomes generated by Wadsworth from the study period, barring a small fraction that did not pass quality control and those removed as redundant (Appendix). We downloaded all B.1.526 genomes from the United States and associated metadata (excluding NYS

sequences) from GISAID (https://www.gisaid.org) and randomly subsampled them proportionally to their overall frequency in the United States. The final dataset included B.1.526 genomes from domestic locations (Massachusetts, New Jersey, Pennsylvania, Connecticut, California, Florida, Maryland, Michigan, Minnesota, and North Carolina), the 5 boroughs of NYC (Bronx, Brooklyn, Queens, Staten Island, and Manhattan), Long Island, the Hudson Valley and upstate New York (Western NYS, the Finger Lakes, the Capital District, and Central NYS regions). We aligned genomes in mafft 7.475 (*19*), masking problematic sites (https://github.com/W-L/ProblematicSites_SARS-CoV2). We generated a maximum-likelihood phylogeny in IQTree 1.6.12 (*20*) with 1,000 ultrafast bootstrap replicates (*21*) and time-calibrated it in TreeTime 0.7.6 (*22*). This tree served as the fixed tree for ancestral state reconstruction in Beast 2.6.2 (*23*) to infer timing and source of B.1.526 introductions within NYS. We allowed the Bayesian analysis to run for >4 million generations and monitored it in Tracer 1.7.1 (*24*) until the effective sample size of all parameters ≥200 and the Markov chain Monte Carlo appeared to reach stationarity.

We conducted a B.1.1.7 phylogeographic analysis in the same manner with the states inferred for a fixed topology until all effective sample sizes reached ≥200. The final dataset included B.1.1.7 genomes from domestic locations (Massachusetts, New Jersey, Pennsylvania, Connecticut, California, and Florida) NYC, Long Island, Mid-Hudson, Finger Lakes, southwestern NYS (the Southern Tier and western regions of NYS) and Northern NYS (Capital District, Mohawk Valley, Central NYS, and the North Country).

We generated maximum clade credibility trees for B.1.526 and B.1.1.7 in TreeAnnotator 2.6.2 (*23*) with a 10% burn-in. We summarized the number of introductions between locations by using Baltic (https://github.com/evogytis/baltic), adopting the exploded tree script for Python 3. We considered only introductions with a posterior probability ≥0.7. We visualized and annotated trees in FigTree 1.5.5 (http://tree.bio.ed.ac.uk/software/figtree) and ggtree (*25*) for R 4.1.0 (http://www.R-project.org) (Appendix).

Results

Summary Statistics

We included in the study a total of 8,517 SARS-CoV-2 specimens sequenced by Wadsworth with collection dates during December 2020–April 2021. Among the included specimens, B.1.1.7 constituted 1,107 (13%) and B.1.526 constituted 904 (10.6%) of the samples.

The earliest B.1.1.7 samples sequenced by Wadsworth were collected on December 24, 2020, from a resident of Manhattan (Metro or NYC region) and a person in Saratoga County (Capital Region). B.1.1.7 remained relatively rare among all samples through the end of January. The Metro and Capital regions experienced the earliest increases in B.1.1.7, although the proportion of B.1.1.7 did not exceed 15% through February. The proportion of B.1.1.7 increased in March across all regions, most notably in the western region, where it constituted ≈75% of all samples by the end of March and continued to rise through April. The Metro Region experienced the most gradual increase in B.1.1.7; the proportion did not exceed 40% until the end of April.

The earliest B.1.526 sample sequenced by Wadsworth was collected on December 9, 2020, from a patient in the Bronx (Metro or NYC region). The proportion of B.1.526 increased in the Metro Region throughout December, reaching 10% of total samples by the end of the month. The proportion of B.1.526 in the Metro Region approached 40% by the end of January, peaked at ≈60% in mid-February to early March, and then plateaued at ≈50% through April. B.1.526 was not consistently detected in the other regions until February and its proportion generally remained <40%. The combined proportion of all lineages other than B.1.1.7 and B.1.526 dropped to <20% in all NYS regions by the end of April.

Cartographic Visualization

Maps of interpolated proportion of B.1.1.7 relative to all other lineages by ZCTA (Appendix Figure 1, panel A) show a general trend of spread through the southern portion of NYS in January, statewide distribution by February, diffuse increase in proportion in March, and a sustained high proportion throughout the state in April, with strong dominance in the western region. In contrast, maps of interpolated proportion of B.1.526 show more constricted initial spread focused around NYC and surrounding areas in January; statewide distribution was not achieved until March, and a moderate proportion was sustained mostly within the Metro Region (Appendix Figure 1, panel B).

Maps of geographic mean centers of estimated B.1.1.7 and B.1.526 cases (Figure 1) show that shifts in the SARS-CoV-2 variant landscape affected the spatial distribution of COVID-19 cases overall. In December 2020, the mean center of total COVID-19 cases and the mean center of the population of NYS were nearly spatially coincident, implying that COVID-19 cases were distributed in accordance with NYS's population. At the same time, the mean center of B.1.526 cases occurred near the NYC area, then gradually moved slightly northwest as

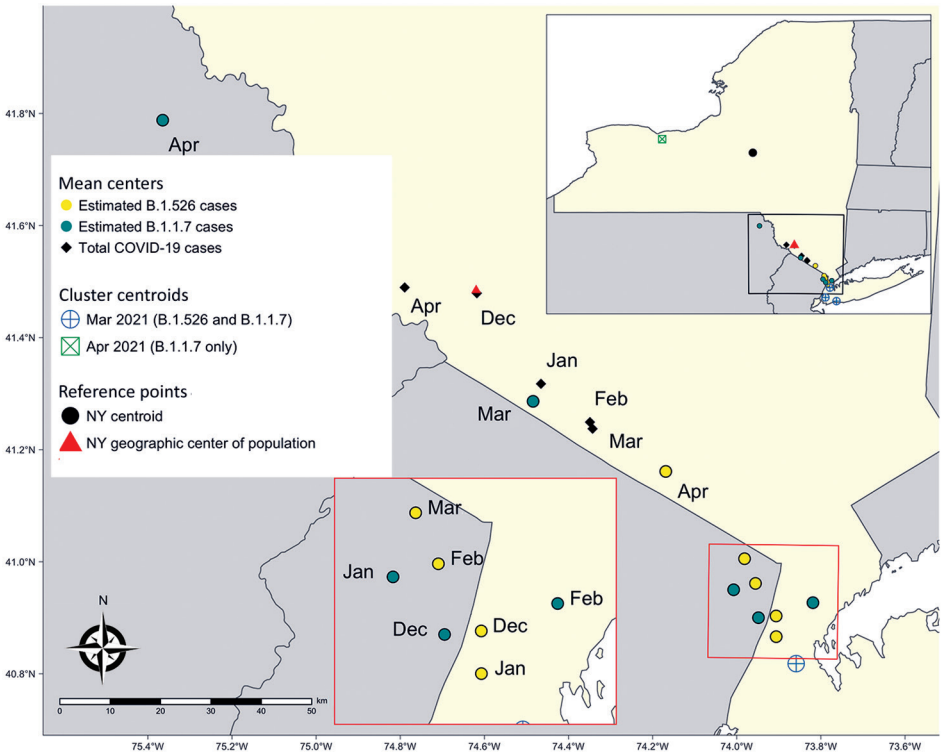


Figure 1. Geographically weighted mean centers of total and estimated coronavirus disease cases attributable to B.1.526 and B.1.1.7 variants, New York State, USA, December 2020–April 2021. Cluster centroids refer to the results of the multinomial space-time scan analysis (Figure 2). New York's centroid and geographic center of population are added as reference points.

B.1.526 expanded modestly into upstate regions. Similarly, the mean center of B.1.1.7 cases was located near NYC in December 2020, then moved northwest during March and April to a much greater degree than for B.1.526 cases, probably because of the B.1.1.7 cluster occurring in the Finger Lakes region. Consequently, the spread of B.1.1.7 in upstate NYS, especially within the western region, resulted in a northwesterly shift of the mean center of total COVID-19 cases by April 2021. The spatial shift pushed the mean center of COVID-19 cases northwest of NYS's population center, indicating that the April B.1.1.7 cluster had an outsized effect on the overall distribution of COVID-19 cases.

Retrospective Multinomial Space-Time Scan Statistic

Retrospective multinomial space-time scan analysis indicated 6 statistically significant clusters with elevated relative risk (RR) of COVID-19 attributable to specific variants (Figure 2; Appendix Table 1). Two clusters of elevated RR of other lineages were found in December, 2020 in the Metro and Capital regions as well as Long Island, reflecting the nearly nonexistent risk for B.1.1.7 and B.1.526 infection. Three clusters of elevated RR of multiple combinations of B.1.1.7, B.1.526, B.1.526.1, and B.1.526.2 were found in March 2021 in the NYC and Long Island regions. The sixth cluster exhibited an elevated RR of >7.0 for B.1.1.7, with a radius of 114.38 km centered in the Finger Lakes area (western and central regions) during April. In addition, the presence of B.1.1.7 and B.1.526 clusters in March and April coincide with a general statewide decrease in incidence of COVID-19 (Figure 2).

Phylogeographic Analyses

The final B.1.526 dataset for phylogenetic reconstruction contained 980 genomes from all regions of NYS and various domestic locations (Bronx, 222; Hudson Valley, 128; Brooklyn, 39; Long Island, 78; Manhattan, 49; Queens, 81; Staten Island, 12; upstate NYS, 81; domestic, 290). The final B.1.1.7 dataset contained 1,195 genomes from the NYC region (181), Finger Lakes (239), Hudson Valley (78), Long Island (130), Western NYS and the Southern Tier (southwestern NYS, 56), Capital District, Mohawk Valley, Central NYS, and the north country (Northern NYS, 149), as well as other states (domestic, 362). Results from the phylogeographic analysis indicated that B.1.526 emerged within the NYC area near the end of 2020 and that the Bronx was a major source of spread to other regions of NYS and the United States (domestic) (Figure 3). Although sampling biases could have influenced the number of introductions assigned to the Bronx, the domestic category had greater representation in the dataset but led to substantially fewer introductions (Appendix Table 2). Domestic genomes represented 29.5% of the dataset but this location was responsible

for only 6.7% of all B.1.526 introductions, whereas the Bronx represented 22.7% of the dataset and led to 63.8% of all introductions (Appendix Table 2). Excluding the Bronx, B.1.526 transmission between boroughs and from these boroughs to other locations was relatively infrequent. We used subsampling strategies to investigate the strength of our results from the full B.1.526 phylogeographic analysis. These strategies included evenly sampling each region or borough, sampling evenly across time (except for December, which had very few B.1.526 cases compared with other months), sampling proportionally to the total incidence of SARS-CoV-2 per region or borough, and downsampling high-incidence regions or boroughs to the mean of B.1.526 cases per month. We performed each subsampling analysis in triplicate. Despite the different subsampling strategies, the root of the tree consistently fell within NYC, and the Bronx continued to serve as a major source of B.1.526 (data not shown).

Multiple domestic introductions contributed to the initial presence of B.1.1.7 in NYS (*11*) (Figure 4), with transmission occurring most frequently in the Finger Lakes and Northern NYS (Figure 4). The Finger Lakes and Northern NYS were well-represented in the dataset (32% of genomes) but contributed substantially less to the distribution of B.1.1.7 (accounting for 13% of the total number of introductions) than domestic sites, which represented 20% of the data and were responsible for the highest percentage of introductions (≈39%) (Appendix Table 3). The Finger Lakes showed the lowest proportion of sequenced cases attributable to introductions but the largest sample size in NYS, suggesting more sustained transmission of B.1.1.7 in this region.

Discussion

The repeated emergence of novel variants of SARS-CoV-2 has largely defined the COVID-19 pandemic response in 2021. As vaccination rates, prior exposure levels, and behavioral public health measures continuously change, so too will selective pressures (*26*). Given that selective pressures likely vary across regions, it follows that the emergence and spread of SARS-CoV-2 variants are also regionally dynamic. We combined spatial statistical, phylogeographic, and cartographic visualization techniques to examine the spatiotemporal dynamics of the VOC B.1.1.7 (Alpha) and the VOI B.1.526 (Iota) in NYS during December 2020–April 2021.

The concurrent spread of B.1.1.7 and B.1.526 offers a unique opportunity to compare the dynamics of competing variants of SARS-CoV-2 within a population during a period of substantial fluctuations in statewide COVID-19 incidence and the implementation of a vaccination campaign in January 2021. Shortly after its appearance in the Bronx in late 2020, B.1.526 quickly became the most common lineage in NYC and the surrounding region. The rapid dominance of B.1.526 in NYC is corroborated by our phylogeographic results (Figure 3), which depict widespread initial transmission within the Bronx, periodic introductions to neighboring boroughs, and later introductions to the greater Metro Region and

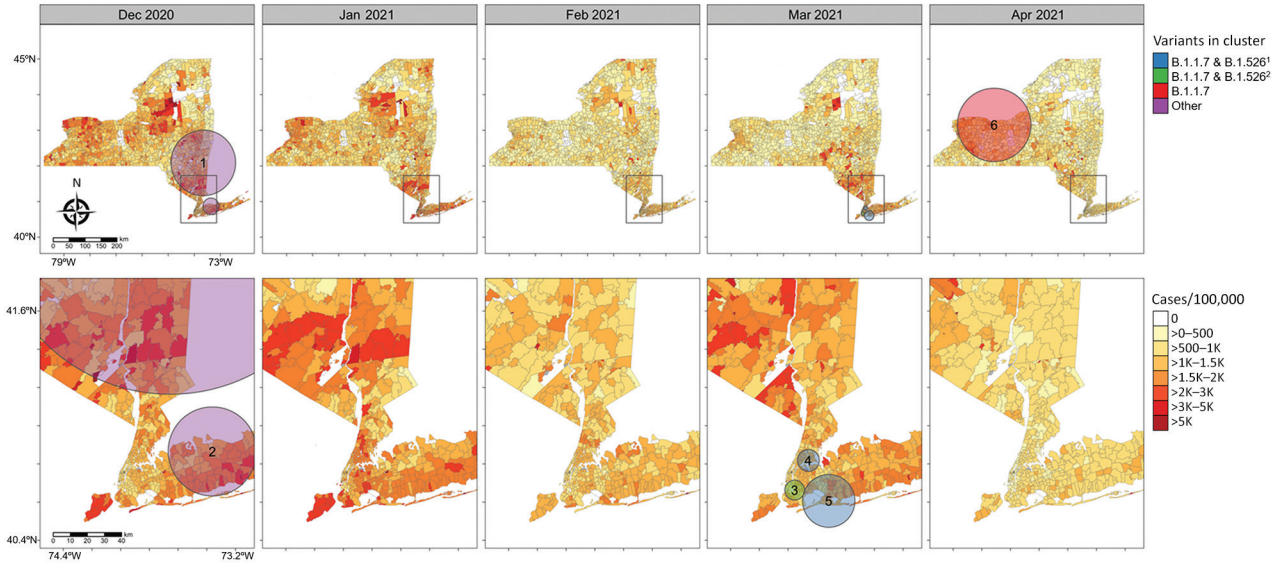


Figure 2. Severe acute respiratory syndrome coronavirus 2 variant clusters identified from retrospective multinomial space-time scan analysis and coronavirus disease incidence by ZIP code tabulation area, New York State, USA, December 2020–April 2021. Circles indicate clusters with relative risk >1. 1, variant includes B.1.526, B.1.526.1, and B.1.526.2; 2, variant includes B.1.526 and B.1.526.2.

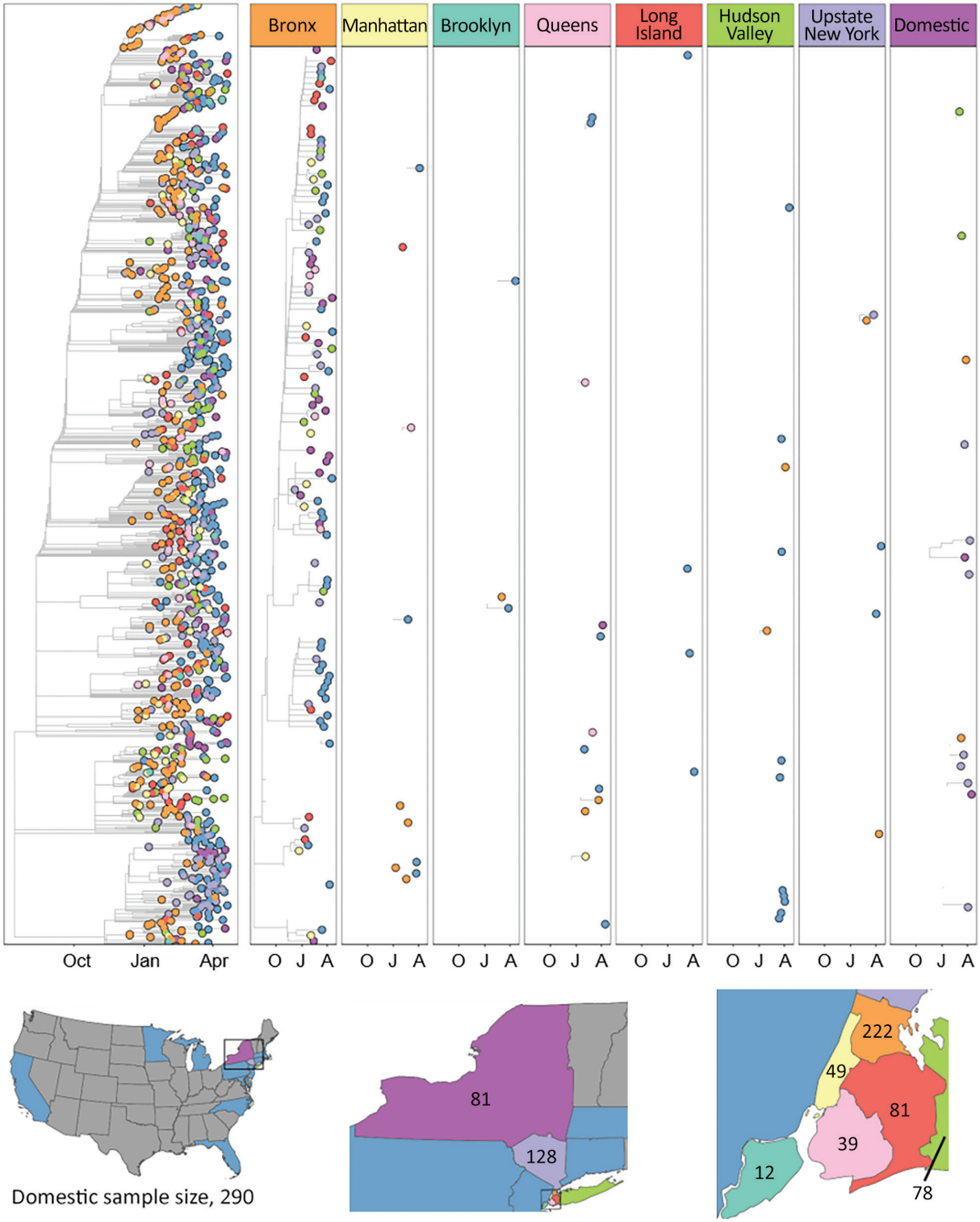


Figure 3. Time-calibrated phylogeny of severe acute respiratory syndrome coronavirus 2 variant B.1.526, New York and other states, USA, December 2020–April 2021. Left panel represents a maximum-likelihood phylogeny of 980 genomes from New York and other US states generated in IQTree 1.6.12 (*20*) with timescale inferred by TreeTime 0.7.6 (*22*) and ancestral state reconstruction performed in BEAST 2.6.2 (*23*). Faceted panels indicate the source of B.1.526 introductions into different regions of New York and other states (domestic). Only introductions supported by an ancestral state probability of ≥0.7 are shown. Bottom panel shows locations sampled and sample sizes. A, April; J, January; O, October.

other states. The spread of B.1.526 appears to have been spatially limited by the repeated introduction and transmission of B.1.1.7 outside NYC. However, behaviors such as differences in travel rates in NYC and between NYC and other regions probably contributed to the dynamics we observed. Similarly, the regional success of either variant may depend on the seroprevalence of the population. B.1.351 (Beta) was predicted to dominate in populations with a high degree of naturally acquired immunity because of immune evasion conferred by the E484K mutation in Spike (C.L. Althaus et al., unpub. data, https://doi.org/10.1101/2021.06.10.21258468), a mutation also observed in most NYS B.1.526 genomes. The spike mutations of B.1.526 (including the E484K mutation) were shown to reduce neutralization activity of convalescent-phase plasma and several antibodies (*10*). Thus, founder effects in an area of high transmission combined with high levels of prior exposure to SARS-CoV-2 might have provided B.1.526 its initial growth advantages in the NYC area, which was also the initial epicenter of the pandemic in the United States. Regions of NYS where B.1.526 had not yet established experienced rapid dominance of B.1.1.7 during March and April. This trend is most clearly seen in the near complete displacement of all other lineages by B.1.1.7 in Western NYS (Appendix Figure 2, panel A), resulting in a large cluster of elevated RR for B.1.1.7 cases in the Finger Lakes region during April (Figure 2). This finding is consistent with the enhanced transmissibility of B.1.1.7 in comparison to non-VOCs and non-VOIs and the conclusion that B.1.1.7 will dominate in populations with lower seroprevalence (C.L. Althaus et al., unpub. data, https://doi.org/10.1101/2021.06.10.21258468), such as those outside the NYC area.

The multinomial spatial scan detected 3 unique clusters in March 2021, all with increased RR for B.1.1.7 and B.1.526. The values for RR in each NYC cluster detail a distinct pattern: clusters centered within the Bronx, Brooklyn, and Manhattan had higher RR for B.1.526, whereas the cluster centered in east Queens and Long Island had a higher RR for B.1.1.7 (Appendix Table 1). During the months after B.1.526's initial advantage in NYC, B.1.1.7 trends toward becoming the major variant in the Metro Region. Given the elevated reproductive number of B.1.1.7 and B.1.526 in comparison to other non-VOC/VOIs lineages (*27*) and the delayed dominance of B.1.1.7 in the Bronx compared with Long Island and Queens, we hypothesize that B.1.526 was more difficult to displace than other lineages circulating at the time. Almost no difference can be observed in the average global reproductive number of B.1.526 compared with B.1.1.7, although differences exist on a country level (*27*). This finding supports the idea that B.1.526 was generally more competitive against B.1.1.7 than background lineages, but other factors, including the location examined, probably influence our results.

Similarly, maps of the geographic mean centers of the estimated number of COVID-19 cases attributable to each variant capture the rapid spread of B.1.1.7 out of NYC and the relative inability of B.1.526 to claim a foothold outside of the Metro Region. The northwesterly shift in the trajectory of overall COVID-19 cases in April indicates that the expansion of B.1.1.7, in particular clustering in western NYS, had a measurable influence on the spatial spread of COVID-19 cases overall.

There are some limitations to our study. A degree of selection bias exists within the dataset, given that specimens were screened by cycle threshold value and were submitted by a selected group of clinical and commercial laboratories that cannot perfectly represent all COVID-19 cases in NYS. We were unable to assess the demographic and clinical representativeness of our dataset because these data were not available to us. In addition, the number of specimens sequenced varied over the space and time of the study period, which created small sample sizes within many ZCTA-months. This limitation extended to the multinomial scan statistic, which was run with estimated values for COVID-19 cases attributable to B.1.1.7 and B.1.526, giving all ZCTAs with samples equal weight. However, the spatial scan assesses data according to their proximity to each other. In this context, ZCTAs are analyzed together rather than individually, which has the potential to reduce bias. Another consequence of our limited sampling was that our data exhibited zero samples from many ZCTAs for each month, which we addressed by using IDW interpolation of the proportion of B.1.1.7 and B.1.526 sequenced samples at the ZCTA-month level to visualize general patterns of variant proportions over geography. Phylogeographic analyses were hampered by similar limitations; uneven sampling among regions and the lack of global representation in our datasets could lead to incorrect trait assignments. Smaller sample sizes for some regions might have caused an underestimation of their contributions to variant transmission in NYS, whereas larger sample sizes might have inflated the number of introductions assigned. However, we believe our results largely capture the transmission dynamics of B.1.526 and B.1.1.7 in NYS, given that larger sample

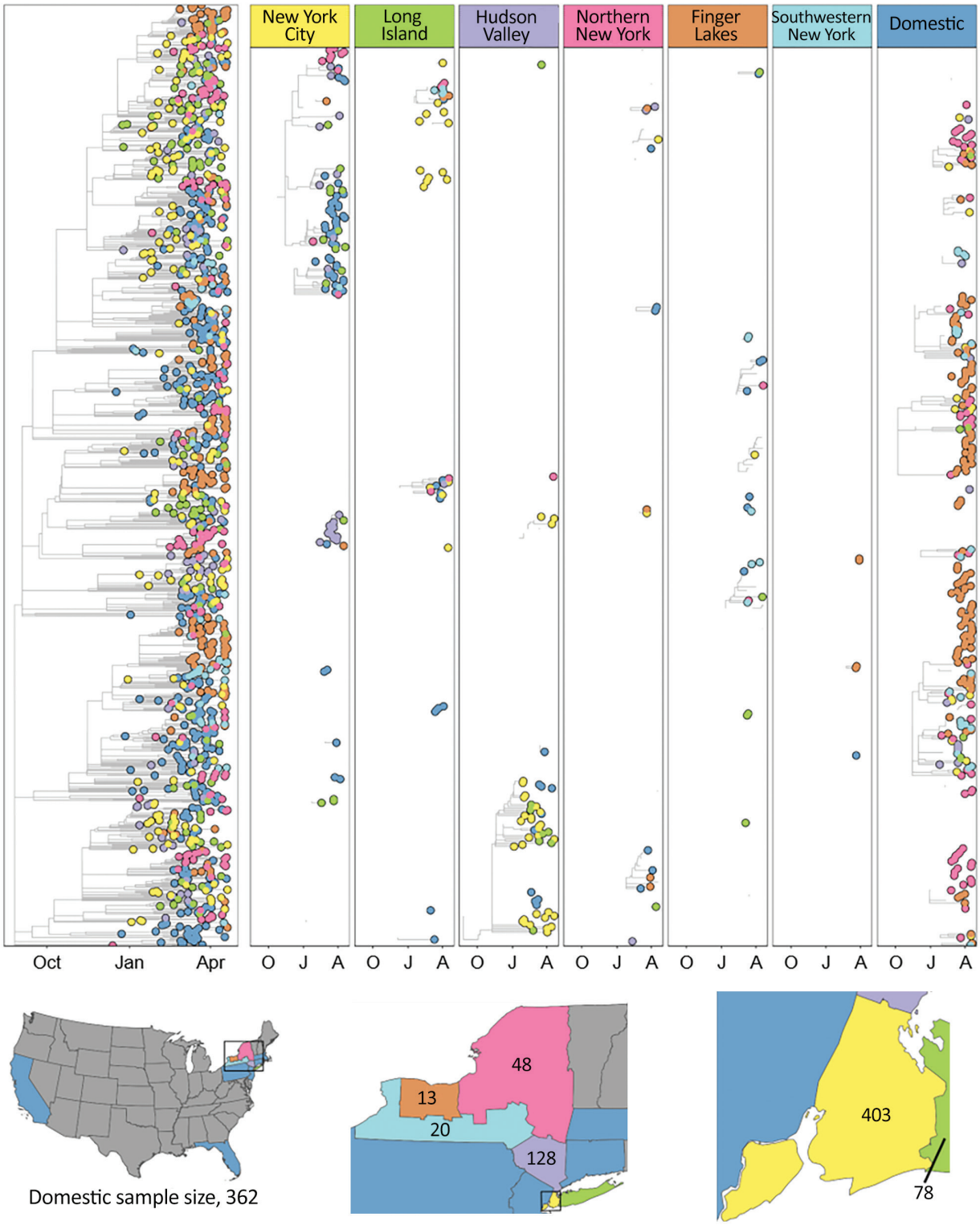


Figure 4. Time-calibrated phylogeny of severe acute respiratory syndrome coronavirus 2 variant B.1.1.7, New York and other states, USA, December 2020–April 2021. Left panel represents a maximum-likelihood phylogeny of 1,195 genomes from New York and other states generated in IQTree 1.6.12 (*20*) with timescale inferred by TreeTime 0.7.6 (*22*) and ancestral state reconstruction performed in BEAST 2.6.2 (*23*). The tree was rooted with a P.1 genome (not shown). Faceted panels indicate the source of B.1.1.7 introductions into different regions of New York and other states (domestic). Only introductions supported by an ancestral state probability of >0.7 are shown. Bottom panel shows locations sampled and sample sizes. A, April; J, January; O, October.

sizes did not always correspond to regions with outsized contributions to the spread of either variant, and subsampling the B.1.526 dataset consistently showed NYC as the dominant source of introductions.

Our phylogeographic and spatiotemporal analyses offer a method for evaluating the competitive advantages of co-circulating SARS-CoV-2 variants. We believe the emergence of VOI B.1.526 contributed to the slower rise of VOC B.1.1.7 as the dominant lineage in NYC compared with regions devoid of B.1.526. In this way, our study describes important dynamic interactions between variants with unequal transmissibility and is potentially generalizable to interactions between any known variants, including the highly transmissible Delta and Omicron variants and other variants to come.

Acknowledgments

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

Initial funding for sequencing was generously provided by the New York Community Trust. The work for this publication was also funded by the Centers for Disease Control and Prevention (cooperative agreement no. NU50CK000516). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

A.R., C.O., E.L.-N., J.P., J.P.K., and D.M.L. declare no conflicts of interest. K.S.G. receives research support from ThermoFisher for the evaluation of new assays for the diagnosis and characterization of viruses, and has a royalty-generating collaborative agreement with Zeptometrix.

Author contributions: A.R. and C.O. conceived of the manuscript. A.R., C.O., and E.L.-N. analyzed data, interpreted results, and wrote and revised the manuscript. A.R., J.P., D.M.L., and K.S.G. contributed to the acquisition and processing of specimens. J.K. generated sequence data. D.M.L. and K.S.G. contributed conceptual insight and revised the manuscript. All authors approved of the final version.

About the Author

Ms. Russell is a research scientist in the Special Projects Unit of the Laboratory of Viral Diseases at the Wadsworth Center, New York State Department of Health, Albany, New York. Her primary research interests are in infectious disease epidemiology, with a particular focus on innovative applications of molecular surveillance methods.

References

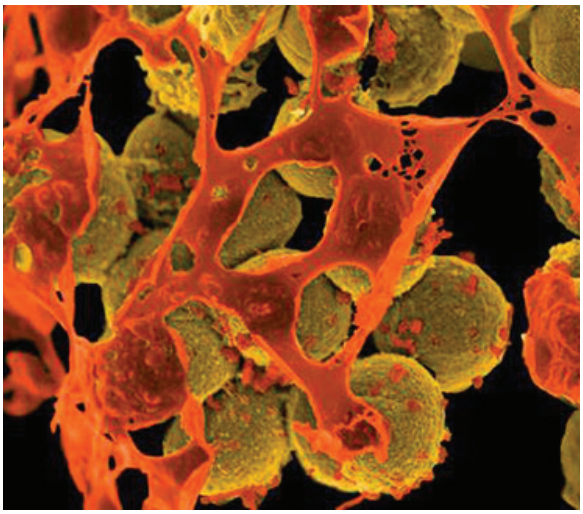
1. Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M. Public health actions to control new SARS-CoV-2 variants. Cell. 2021;184:1127–32. https://doi.org/10.1016/j.cell.2021.01.044
2. Lauring AS, Hodcroft EB. Genetic variants of SARS-CoV-2—what do they mean? JAMA. 2021;325:529–31. https://doi.org/10.1001/jama.2020.27124
3. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. Lancet Infect Dis. 2021;21:1246–56. https://doi.org/10.1016/S1473-3099(21)00170-5
4. Kidd M, Richter A, Best A, Cumley N, Mirza J, Percival B, et al. S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral load in samples tested by TaqPath polymerase chain reaction. J Infect Dis. 2021;223:1666–70. https://doi.org/10.1093/infdis/jiab082
5. Planas D, Bruel T, Grzelak L, Guivel-Benhassine F, Staropoli I, Porrot F, et al. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. Nat Med. 2021;27:917–24. https://doi.org/10.1038/s41591-021-01318-5
6. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature. 2021;592:438–43. https://doi.org/10.1038/s41586-021-03402-9
7. West AP Jr, Wertheim JO, Wang JC, Vasylyeva TI, Havens JL, Chowdhury MA, et al. Detection and characterization of the SARS-CoV-2 lineage B.1.526 in New York. Nat Commun. 2021;12:4886. https://doi.org/10.1038/s41467-021-25168-4
8. World Health Organization. Epidemiological update: variants of SARS-CoV-2 in the Americas. 2021 Mar [cited 2021 Dec 9]. https://iris.paho.org/handle/10665.2/53382
9. Thompson CN, Hughes S, Ngai S, Baumgartner J, Wang JC, McGibbon E, et al.; PhD1. PhD1. Rapid emergence and epidemiologic characteristics of the SARS-CoV-2 B.1.526 variant—New York City, New York, January 1–April 5, 2021. MMWR Morb Mortal Wkly Rep. 2021;70:712–6. https://doi.org/10.15585/mmwr.mm7019e1
10. Annavajhala MK, Mohri H, Wang P, Nair M, Zucker JE, Sheng Z, et al. Emergence and expansion of SARS-CoV-2 B.1.526 after identification in New York. Nature. 2021;597:703–8. https://doi.org/10.1038/s41586-021-03908-2
11. Alpert T, Brito AF, Lasek-Nesselquist E, Rothman J, Valesano AL, MacKay MJ, et al. Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. Cell. 2021;184:2595–2604.e13. https://doi.org/10.1016/j.cell.2021.03.061
12. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. Stat Med. 2010;29:1910–8. https://doi.org/10.1002/sim.3951

13. Kulldorff M. Software for the spatial and space-time scan statistics. 2018 [2021 Dec 9]. http://www.satscan.org
14. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. Appl Geogr. 2020;118:102202. https://doi.org/10.1016/j.apgeog.2020.102202
15. Iyengar VS. Space-time clusters with flexible shapes. MMWR Suppl. 2005;54:71–6.
16. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. Int J Health Geogr. 2008;7:14. https://doi.org/10.1186/1476-072X-7-14
17. Shepard D. A two-dimensional interpolation function for irregularly-spaced data. 1968 Jan 1 [cited 2021 Jun 8]. http://portal.acm.org/citation.cfm?doid=800186.810616
18. Rogerson PA, Plane DA. Geographical analysis of population: with applications to planning and business. International edition. Hoboken (New Jersey): John Wiley and Sons Ltd; 1994.
19. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010
20. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300
21. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30:1188–95. https://doi.org/10.1093/molbev/mst024
22. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042. https://doi.org/10.1093/ve/vex042
23. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLOS Comput Biol. 2019;15:e1006650. https://doi.org/10.1371/journal.pcbi.1006650
24. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67:901–4. https://doi.org/10.1093/sysbio/syy032
25. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. https://doi.org/10.1111/2041-210X.12628
26. Neher R. The virus is under increasing selection pressure. Max-Planck-Gesellschaft. 2021 [cited 2021 Aug 10]. https://www.mpg.de/16371358/coronavirus-variants
27. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill. 2021;26:2100509. https://doi.org/10.2807/1560-7917.ES.2021.26.24.2100509

Address for correspondence: Kirsten St. George, Laboratory of Viral Diseases, Wadsworth Center, New York State Department of Health, David Axelrod Institute, 120 New Scotland Ave, Albany, New York, NY 12208, USA; email: Kirsten.St.George@health.ny.gov

EID Podcast

Livestock, Phages, MRSA, and People in Denmark



Methicillin-resistant *Staphylococcus aureus*, better known as MRSA, is often found on human skin. But MRSA can also cause dangerous infections that are resistant to common antimicrobial drugs. Epidemiologists carefully monitor any new mutations or transmission modes that might lead to the spread of this infection.

Approximately 15 years ago, MRSA emerged in livestock. From 2008 to 2018, the proportion of infected pigs in Denmark rocketed from 3.5% to 90%.

What happened, and what does this mean for human health?

In this EID podcast, Dr. Jesper Larsen, a senior researcher at the Statens Serum Institut, describes the spread of MRSA from livestock to humans.

Visit our website to listen:
https://go.usa.gov/x74Jh

EMERGING INFECTIOUS DISEASES®

Article DOI: https://doi.org/10.3201/eid2803.211972

Spatiotemporal Analyses of 2 Co-Circulating SARS-CoV-2 Variants, New York State, USA

Appendix

Geographic Mean Center Calculation

The geographic mean center of total cases and estimated variant cases of COVID-19 were calculated using the following equation.

$$\frac{\sum_i^N x_i z_i}{\sum_i^N z_i}, \frac{\sum_i^N y_i z_i}{\sum_i^N z_i}$$

Where x_i and y_i denote latitude and longitude values of a ZCTA centroid, respectively, and zi denotes the number of cases recorded or estimated for a ZCTA. Centroid calculation, spatial averaging, IDW methods and maps were performed using the 'sf', 'raster', 'gstat', and 'tmap' packages in RStudio version 4.0.2, respectively (*1–5*).

Retrospective Multinomial Space-Time Scan Statistic

The procedure for the multinomial scan statistic implemented in SaTScan from (*6*) is described below, using terms that apply to our research questions. The multinomial scan statistic assesses the null hypothesis of no clustering by globally testing whether the probability of acquiring a specific variant of SARS-CoV-2 relative to all variants of SARS-CoV-2 is the same in all parts of the study area. The rejection of the global null hypothesis permits for the scanning of a specific region and regions while testing the same null hypothesis locally. Specifically, the space-time scan procedure operates by searching for clusters in a "moving cylinder" fashion, such that the base of the cylinder is the spatial scan, while the height of the cylinder indicates the temporal scan. As the cylinder moves throughout the spatiotemporal study region, the test statistic is calculated for each scanning window, and the window that maximizes the likelihood ratio test statistic is selected as the most likely cluster. For specific details on the likelihood function and test statistic (*6*).

The moving cylinder method employed by SaTScan presents a key limitation for use examining disease outbreaks. The geometry of a cylinder does not allow for the change in the spatial extent of a cluster throughout time, as would be expected for a disease cluster that is spreading (*7*). Methodologies have been proposed to alleviate this problem, including the "square pyramid" method and the "flexible space-time scan statistic" (*7*,*8*). Neither the square pyramid nor the flexible space-time scan statistics were available in the SaTScan software, thus, we elected to reduce our maximum temporal cluster size to be equivalent to our time precision. Additionally, adjusting the population at risk parameter when using the multinomial scan statistic sets an upper bound for the size of a cluster according to the number of cases it will include, rather than the population at risk. In this way, clusters resulting from our analysis will not include more than 10% of the total cases during our specific time aggregation units of one month.

Illumina Library Preparation and Sequencing

Extracted RNA was processed for whole genome sequencing with a modified ARTIC protocol (artic.network/ncov-2019) in the Applied Genomics Technology Core at the Wadsworth Center. Briefly, cDNA was synthesized with SuperScript™ IV reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and random hexamers. Amplicons were generated by pooled PCR with two premixed ARTIC V3 primer tools (Integrated DNA Technologies, Coralville, IA, USA). Additional primers to supplement those showing poor amplification efficiency (github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV2019) were added separately to the pooled stocks. PCR conditions were 98°C for 30 seconds, 24 cycles of 98°C for 15 seconds/63°C for 5 minutes, and a final 65°C extension for 5 minutes. Amplicons from pool 1 and pool 2 reactions were combined and purified by AMPure XP beads (Beckman Coulter, Brea, CA, USA) with a 1X bead-to-sample ratio and eluted in 10mM Tris-HCl (pH 8.0). The amplicons were quantified using Quant-IT™ dsDNA Assay Kit on an ARVO™ X3 Multimode Plate Reader (Perkin Elmer, Waltham, MA, USA). Illumina sequencing libraries were generated using the Nextera DNA Flex Library Prep Kit with Illumina Index Adaptors and sequencing on a MiSeq instrument (Illumina, San Diego, CA, USA).

Bioinformatics Processing

Illumina libraries were processed with ARTIC nextflow pipelines (github.com/connor-lab/ncov2019/articnf/tree/illumine, last updated April 2020) as previously described (*9*). Reads

were trimmed with TrimGalore (github.com/FelixKrueger/TrimGalore) and aligned to the reference assembly MN908947.3 (Wuhan-1) by BWA (*10*). Primers were trimmed with iVar (*11*) and variants were called with samtools mpileup function (*12*), the output of which was used by iVar to generate consensus sequences. Positions were required to be covered by a minimum depth of 50 reads and variants were required to be present at a frequency ≥0.75.

Lineages were determined by GISAID using Pangolin software 29, last updated May 27, 2021 (*13*). At the time of this analysis, B.1.526 was divided into a B.1.526 parent lineage and sublineages B.1.526.1, B.1.526.2, and B.1.526.3, which we analyzed separately in the multinomial scan analysis. Pangolin has since collapsed the sublineages and reassigned all to B.1.526.

Phylogeographic Analyses

All B.1.526 genomes from the United States (US) and associated metadata (excluding NY sequences) were downloaded from GISAID (GISAID.org) and randomly subsampled, with the number of genomes from each state sampled proportionally to their overall frequency in the US. Genomes were aligned in mafft v7.475 (*14*) with problematic sites masked according to (https://github.com/W-L/ProblematicSites_SARS-CoV2). Putative transmission clusters were identified by TreeCluster v1.0.3 (*15*) with a threshold free approach and only one representative genome was selected from each cluster if 1) all genomes derived from the same state within a one week time period or 2) all genomes derived from the same NY county within a one week time period to reduce the size of the dataset. After generating an initial ML tree in IQTree v1.6.12 (*16*) under a GTR+G substitution model, it became apparent that most states contributed minimally or not at all to the number of B.1.526 introductions into NY. It also appeared that most B.1.526 viral circulation occurred between NY and geographically proximal states (M.E. Petrone et al., unpub. data, https://doi.org/10.1101/2021.07.01.21259859). As the focus of our paper was mainly to document the spread of B.1.526 within NY as compared to B.1.1.7, we further reduced our dataset to include only states with the greatest number of sequenced B.1.526 cases and neighboring states to NY. Temporal signal was confirmed by TempEst v1.5.3 (*17*) and genomes with residuals > 0.005 were removed. The final dataset included B.1.526 genomes from MA, NJ, PA, CT, CA, FL, MD, MI, MN, and NC, aggregated as "Domestic". Because B.1.526 likely originated within the Metro region (Appendix Figure 2, panel B), we elected to keep the five boroughs of NYC (Bronx, Brooklyn, Queens, Staten Island, Manhattan) as well as Long

Island and Hudson Valley as distinct to infer the geographic origin of B.1.526 and determine transmission dynamics in this epicenter. The other regions of NY had either no or a considerably lower number of sequenced cases of B.1.526, which is consistent with the incidence of the variant in those regions. Thus, Western NY, the Finger Lakes, the Capital District, and Central NY regions were aggregated as "Upstate". A second ML tree was generated for this reduced dataset in IQTree under a GTR +G4 substitution model with 1000 ultrafast bootstrap replicates (*18*). This tree was then input into TreeTime v0.7.6 (*19*) to estimate a molecular clock (inferred as ~4.0E-04 substitutions per site per year) and re-root the tree with the least-squares method. The time-calibrated tree was input as the fixed tree for discrete ancestral state reconstruction (a method previously validated by Alpert et al., 2021([9]) in BEAST2 v2.6.2 (*20*,*21*), using a symmetric substitution model and strict clock. All tree priors were removed from the final XML document output by BEAUTI and the BSSVS operator was turned on before running in BEAST. The Bayesian analysis was allowed to run for > 4 million generations and monitored in Tracer until the effective sample size of all parameters >= 200 and the MCMC chain appeared to reach stationarity.

A B.1.1.7 phylogeographic analysis was conducted in the same manner with the following exceptions: the tree was initially rooted with a P.1 (Gamma) representative as B.1.1.7 cases in NY had multiple origins, the five boroughs of NYC were included as the same region as it has been established that B.1.1.7 was introduced several times from non-NYC locations, the Capital District, Mohawk Valley, Central NY, and the North Country were aggregated as "Northern NY" given their proximity to each other, Western NY and its neighboring region, the Southern Tier, were grouped together as "Southwestern NY", the Finger Lakes, the Hudson Valley, and Long Island remained distinct. B.1.1.7 locations required different coding than B.1.526 due to the substantial differences in sample sizes. For example, genomes from the Finger Lakes accounted for over 25% of the B.1.1.7 data but less than 2% of the data for B.1.526. NY regions were sampled proportionally to their contribution to the total number of B.1.1.7 cases for the state. MA, PA, CT, NJ, CA, and FL were grouped together as "Domestic" sources of B.1.1.7. Ancestral states were inferred for a fixed topology over 6 million generations in BEAST2 until all ESS reach >= 200. Maximum clade credibility trees for B.1.526 and B.1.1.7 were generated in TreeAnnotator v.2.6.2 (*20*) with a 10% burn-in. The number of introductions between locations was summarized by Baltic (https://github.com/evogytis/baltic) by adopting the

exploded tree script for Python 3. Only introductions with a posterior probability of 0.7 >= were considered. Trees were visualized in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and ggtree (*22*) for R v4.1.0 (http://www.R-project.org).

References

1. Hijmans RJ. raster: geographic data analysis and modeling. 2020 [cited 2021 Dec 9]. https://CRAN.R-project.org/package=raster

2. Pebesma E. Multivariable geostatistics in S: the gstat package. Comput Geosci. 2004;30:683–91. https://doi.org/10.1016/j.cageo.2004.03.012

3. Pebesma E. Simple features for R: standardized support for spatial vector data. 2018 [cited 2021 Dec 9]. https://journal.r-project.org/archive/2018/RJ-2018-009/RJ-2018-009.pdf

4. RStudio Team. RStudio: integrated development environment for R. RStudio. 2018 [cited 2021 Dec 9]. https://www.rstudio.com/categories/integrated-development-environment

5. Tennekes M. tmap: thematic maps in R. J Stat Softw. 2018:84 [cited 2021 Dec 9]. https://www.jstatsoft.org/article/view/v084i06

6. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. Stat Med. 2010;29:1910–8. PubMed https://doi.org/10.1002/sim.3951

7. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. Int J Health Geogr. 2008;7:14. PubMed https://doi.org/10.1186/1476-072X-7-14

8. Iyengar VS. Space-time clusters with flexible shapes. MMWR Suppl. 2005;54:71–6. PubMed

9. Alpert T, Brito AF, Lasek-Nesselquist E, Rothman J, Valesano AL, MacKay MJ, et al. Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. Cell. 2021;184:2595–2604.e13. PubMed https://doi.org/10.1016/j.cell.2021.03.061

10. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95. PubMed https://doi.org/10.1093/bioinformatics/btp698

11. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20:8. PubMed https://doi.org/10.1186/s13059-018-1618-7

12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. PubMed https://doi.org/10.1093/bioinformatics/btp352

13. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 2020 [cited 2021 Dec 9]. 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

14. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. PubMed https://doi.org/10.1093/molbev/mst010

15. Balaban M, Moshiri N, Mai U, Jia X, Mirarab S. TreeCluster: Clustering biological sequences using phylogenetic trees. PLoS One. 2019;14:e0221068. PubMed https://doi.org/10.1371/journal.pone.0221068

16. Nguyen TH, Nguyen HL, Nguyen TY, Vu SN, Tran ND, Le TN, et al. Field evaluation of the establishment potential of wMelPop Wolbachia in Australia and Vietnam for dengue control. Parasit Vectors. 2015;8:563. PubMed https://doi.org/10.1186/s13071-015-1174-x

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

18. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30:1188–95. PubMed https://doi.org/10.1093/molbev/mst024

19. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042. PubMed https://doi.org/10.1093/ve/vex042

20. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Comput Biol. 2019;15:e1006650. PubMed https://doi.org/10.1371/journal.pcbi.1006650

21. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLOS Comput Biol. 2009;5:e1000520. PubMed https://doi.org/10.1371/journal.pcbi.1000520

22. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36. https://doi.org/10.1111/2041-210X.12628

Appendix Table 1. Multinomial cluster analysis cluster-specific relative risks*

		Lineage					
Cluster Number	Month	B.1.1.7	B.1.526	B.1.526.1	B.1.526.2	B.1.526.3	Other
1	December	0	0	0	0	0	**1.56**
2	December	0	0	0	0	0	**1.56**
3	March	**2.83**	**4.11**	0.36	**1.48**	0	0.13
4	March	**1.32**	**2.82**	**4.44**	**3.77**	0	0.18
5	March	**4.59**	**1.66**	**1.12**	**1.71**	0	0.19
6	April	**7.49**	0.29	0.27	0.54	0	0.11

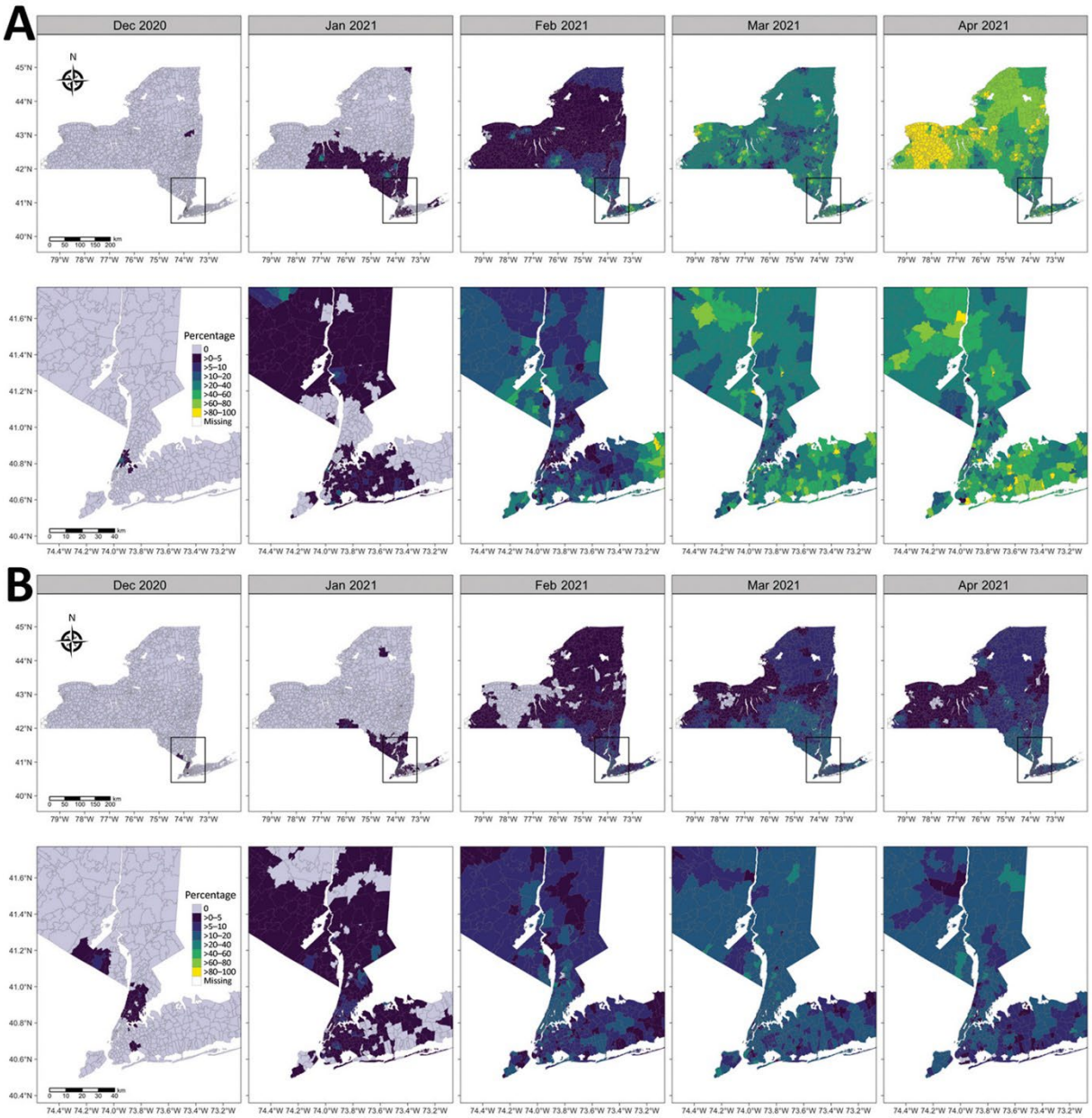
*Relative risk (RR) greater than 1.0 is bolded.

Appendix Table 2. Number of B.1.526 introductions with posterior probability >0.7

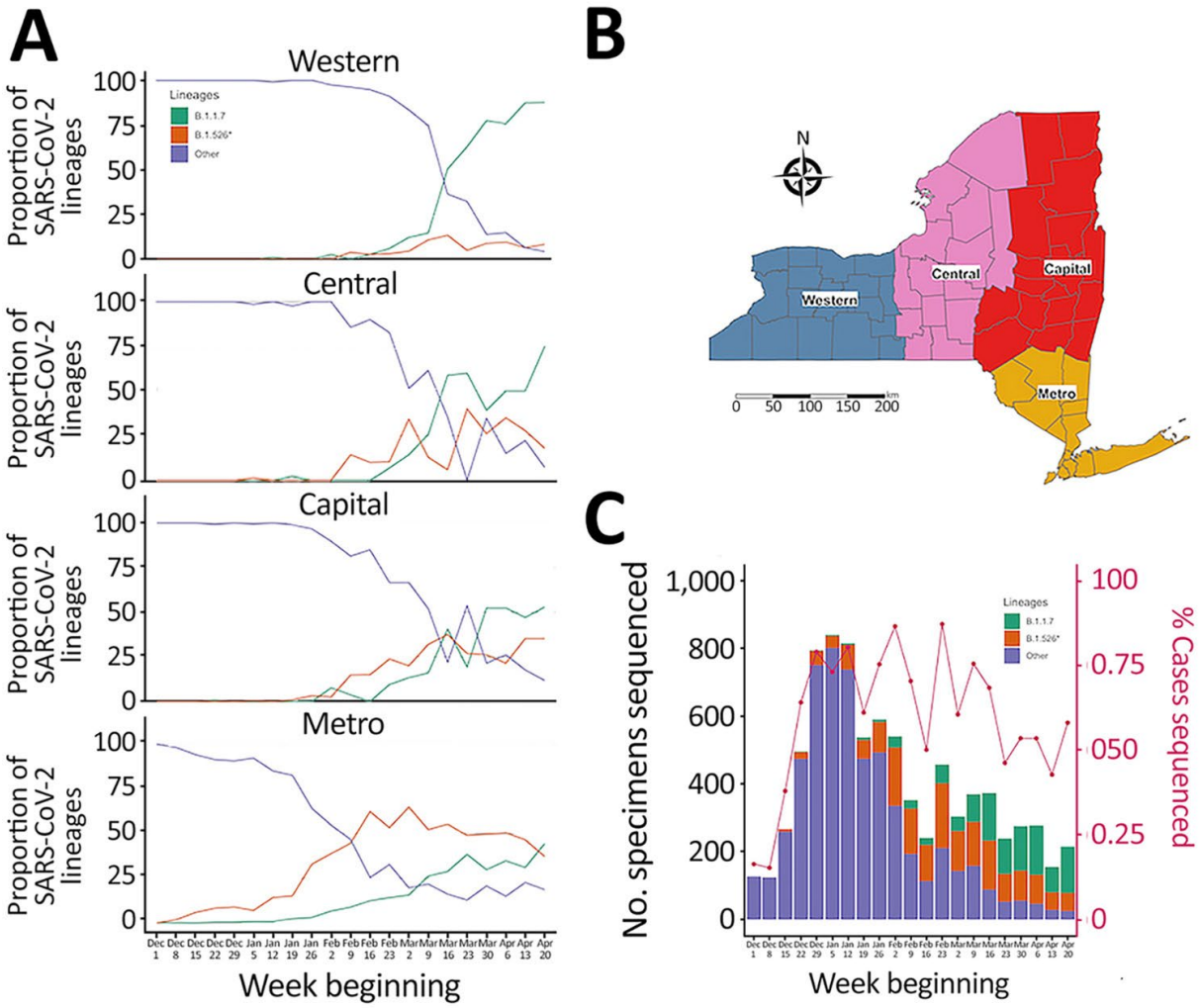
From	To	Introductions
Bronx	Domestic	50
Bronx	Hudson Valley	24
Bronx	Brooklyn	11
Bronx	Long Island	17
Bronx	Manhattan	18
Bronx	Queens	19
Bronx	Staten Island	2
Bronx	Upstate	21
Domestic	Bronx	3
Domestic	Hudson Valley	8
Domestic	Long Island	4
Domestic	Upstate	2
Hudson Valley	Bronx	5
Hudson Valley	Domestic	15
Hudson Valley	Long Island	2
Hudson Valley	Staten Island	1
Hudson Valley	Upstate	1
Brooklyn	Bronx	2
Brooklyn	Domestic	2
Brooklyn	Manhattan	1
Brooklyn	Queens	1
Long Island	Domestic	5
Long Island	Manhattan	1
Long Island	Staten Island	1
Manhattan	Bronx	5
Manhattan	Domestic	4
Manhattan	Brooklyn	2
Manhattan	Queens	1
Queens	Bronx	2
Queens	Domestic	5
Queens	Brooklyn	4
Queens	Long Island	1
Queens	Manhattan	3
Queens	Upstate	1
Staten Island	Domestic	2
Upstate	Bronx	3
Upstate	Domestic	2
Upstate	Hudson Valley	1
Upstate	Long Island	1
Upstate	Queens	1

Appendix Table 3. Number of B.1.1.7 introductions with posterior probability >0.7

From	To	Introductions
Capital	Domestic	6
Capital	Finger Lakes	4
Capital	Hudson Valley	3
Capital	NYC	2
Capital	Long Island	1
Domestic	Finger Lakes	30
Domestic	Capital	26
Domestic	NYC	15
Domestic	Southwestern	14
Domestic	Hudson Valley	9
Domestic	Long Island	5
Finger Lakes	Domestic	7
Finger Lakes	Long Island	4
Finger Lakes	Southwestern	4
Finger Lakes	Capital	2
Finger Lakes	NYC	1
Hudson Valley	NYC	24
Hudson Valley	Long Island	7
Hudson Valley	Domestic	5
Hudson Valley	Capital	1
Long Island	NYC	11
Long Island	Domestic	8
Long Island	Capital	3
Long Island	Southwestern	2
Long Island	Finger Lakes	1
Long Island	Hudson Valley	1
NYC	Domestic	26
NYC	Long Island	12
NYC	Hudson Valley	10
NYC	Capital	4
NYC	Finger Lakes	2
NYC	Southwestern	1
Southwestern	Finger Lakes	3
Southwestern	Domestic	1



Appendix Figure 1. Inverse distance weighted interpolations of percentage of severe acute respiratory syndrome coronavirus 2 infections attributable to B.1.1.7 (A) and B.1.526 (B) variants compared with all other lineages, by ZIP code tabulation area, New York State, USA, December 2020–April 2021.


A
Western
Central
Capital
Metro
Proportion of SARS-CoV-2 lineages
Lineages
B.1.1.7
B.1.526*
Other
Week beginning
B
N
Western
Central
Capital
Metro
0 50 100 150 200
C
No. specimens sequenced
% Cases sequenced
Week beginning