

UC San Diego

UC San Diego Previously Published Works

Title

Introduction and Establishment of SARS-CoV-2 Gamma Variant in New York City in Early 2021

Permalink

<https://escholarship.org/uc/item/10t7903r>

Journal

The Journal of Infectious Diseases, 226(12)

ISSN

0022-1899

Authors

Vasylyeva, Tetyana I

Fang, Courtney E

Su, Michelle

et al.

Publication Date

2022-12-13

DOI

10.1093/infdis/jiac265

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

1 **Introduction and Establishment of SARS-CoV-2 Gamma Variant in New York City in Early 2021**

2 **Running title: Gamma Variant in New York City in 2021**

3
4 Tetyana I. Vasylyeva^{1,*}, Courtney E. Fang², Michelle Su², Jennifer L. Havens³, Edyth Parker⁴, Jade C.
5 Wang², Mark Zeller⁴, Anna Yakovleva⁵, Gabriel W. Hassler⁶, Moinuddin A. Chowdhury², Kristian G.
6 Andersen⁴, Scott Hughes², Joel O. Wertheim¹

7
8 ¹ Department of Medicine, University of California San Diego, La Jolla, CA, USA

9 ² New York City Public Health Laboratory, New York City Department of Health and Mental Hygiene, New
10 York, NY, USA

11 ³ Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA,
12 US

13 ⁴ Department of Immunology and Microbiology, Scripps Research, La Jolla, CA, US

14 ⁵ Medical Sciences Division, University of Oxford, Oxford, UK

15 ⁶ Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA, USA

16
17 *To whom correspondence should be addressed: Tetyana I. Vasylyeva (tvasylyeva@health.ucsd.edu)

18
19 **Funding**

20 T.I.V. is funded by a Branco Weiss Fellowship. J.O.W. acknowledges funding from the National
21 Institutes of Health (AI135992). G.W.H. acknowledges funding from the National Institutes of Health
22 (AI154824). K.G.A acknowledges funding from the National Institutes of Health (U01AI151812 and
23 U19AI135995) and CDC BAA contract 75D30120C09795. This work was supported (in part) by the
24 Epidemiology and Laboratory Capacity (ELC) for Infectious Diseases Cooperative Agreement (Grant

© The Author(s) 2022. Published by Oxford University Press on behalf of Infectious Diseases Society of America. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com This article is published and distributed under the terms of the Oxford University Press, Standard Journals Publication Model (https://academic.oup.com/journals/pages/open_access/funder_policies/chorus/standard_publication_model) 1

1 Number: ELC DETECT (6NU50CK000517-01-07) funded by the Centers for Disease Control and
2 Prevention (CDC). The findings and conclusions in this report are those of the authors and do not
3 necessarily represent the official position of the Centers for Disease Control and Prevention. Use of trade
4 names is for identification only and does not imply endorsement by the Centers for Disease Control and
5 Prevention.

6 **Acknowledgments**

7 We thank Mustapha Mustapha of the New York City Department of Health and Mental Hygiene for his
8 useful comments on the manuscript text.

9 **Declaration of interests**

10 K.G.A. has received consulting fees for advising on SARS-CoV-2, variants, and the COVID-19
11 pandemic. J.O.W. has received funding from the CDC (ongoing) via contracts and agreements to his
12 institution unrelated to this research.

13 **Corresponding author information**

14 Tetyana Vasylyeva, DPhil
15 Assistant Professor
16 Division of Infectious Diseases and Global Public Health
17 University of California San Diego
18 San Diego, California, USA
19 tvasylyeva@health.ucsd.edu
20 +1 (858) 766 1012

21 Alternate corresponding author
22 Joel Wertheim, PhD
23 Associate Professor
24 Division of Infectious Diseases and Global Public Health
25 University of California San Diego
26 San Diego, California, USA
27 jwertheim@health.ucsd.edu
28

1 **ABSTRACT**

2 Background. Monitoring the emergence and spread of SARS-CoV-2 variants is an important public health
3 objective. We investigated how the Gamma variant was established in New York City (NYC) in early 2021
4 in the presence of travel restrictions that aimed to prevent viral spread from Brazil, the country where the
5 variant was first identified.

6 Methods. We performed phylogeographic analysis on 15,967 Gamma sequences sampled between
7 March 10th through May 1st, 2021, to identify geographic sources of Gamma lineages introduced into
8 NYC. We identified locally circulating Gamma transmission clusters and inferred the timing of their
9 establishment in NYC.

10 Results. We identified 16 phylogenetically-distinct Gamma clusters established in NYC (cluster sizes
11 ranged 2-108 genomes); most of them were introduced from Florida and Illinois and only one directly from
12 Brazil. By the time the first Gamma case was reported by genomic surveillance in NYC on March 10th, the
13 majority (57%) of circulating Gamma lineages had already been established in the city for at least two
14 weeks.

15 Conclusions. Although travel from Brazil to the US was restricted from May 2020 through the end of the
16 study period, this restriction did not prevent Gamma from becoming established in NYC as most
17 introductions occurred from domestic locations.

18

19

20 *Keywords:* SARS-CoV-2, Gamma Variant, New York City, Travel restrictions, public health.

21

22

23

1 BACKGROUND

2 Monitoring the emergence, introduction, and circulation of severe acute respiratory syndrome
3 coronavirus 2 (SARS-CoV-2) Variants of Concern (VOCs) is an important public health tool that is
4 essential to guiding an effective pandemic response [1]. Emergence and rapid spread of SARS-COV-2
5 variants has driven consecutive epidemic waves regionally and around the globe, starting with the Alpha
6 variant (B.1.1.7 lineage) identified in the UK in late 2020, followed by the Delta variant (B.1.617.2)
7 identified in India in early 2021, and the recent waves of Omicron variants infections the first of which was
8 identified in South Africa in late 2021. In the US and elsewhere, the response to the identification of these
9 variants has included restricting travel from locations where the variants were first identified; however,
10 widespread establishment of these variants nonetheless ensued.

11 Gamma VOC (P.1 lineage) was first identified in late 2020 in Manaus, Brazil, a region that previously
12 experienced a high COVID-19 burden and had an estimated 68% seroprevalence [2]. The rapid spread of
13 Gamma in a region with high pre-existing immunity immediately raised concerns about the transmission
14 potential of this variant. On January 25th, 2021, the first case of Gamma infection was documented in the
15 United States (US) in the state of Minnesota [3]. Travel from Brazil to the US was prohibited for non-US
16 citizens starting May 2020, and this was extended indefinitely on January 26th, 2021 due to concern about
17 Gamma spread [4]. Nevertheless, by May 2021, Gamma had been detected in at least 41 US states
18 (<https://outbreak.info/>).

19 Molecular epidemiology investigations into this lineage characterized 17 novel mutations relative to the
20 wild type virus (Hu-1 reference strain)—including ten mutations in the spike protein, three of which
21 (K417T, E484K, and N501Y) have been shown to increase spike ACE2 receptor binding affinity [5]—
22 leading to the lineage being designated a VOC by the CDC [6]. Gamma virus is poorly inhibited by
23 convalescent plasma and sera from vaccinated individuals [7], is partially resistant to monoclonal
24 antibodies used for COVID-19 treatment [8], and might be associated with increased COVID-19 severity
25 and mortality in younger adults and adults without comorbidities, compared to the wild type virus [9, 10].

1 New York City (NYC) was among the first locations in North America or Europe to experience a large
2 COVID-19 outbreak: by May 2020, the number of reported cases exceeded 150,000 [11]. The first
3 Gamma VOC case in NYC was identified on March 10th, 2021, in the middle of the second COVID-19
4 wave. The proportion of Gamma-attributed cases in NYC was growing but was outcompeted first by Iota
5 (B.1.526 lineage) and Alpha variants [12], and then by Delta in summer 2021. Importantly, NYC, unlike
6 many other locations in the US and globally, has a large-scale genomic surveillance system deployed by
7 the New York City Public Health Laboratory (NYC PHL) along with the establishment of the NYC
8 Pandemic Response Lab (PRL), which paved the way for 14,600 SARS-CoV-2 genomes to be
9 sequenced by NYC PHL and PRL between 1 January to 1 May 2021.

10 Here we used a phylogeographic approach to investigate patterns of introduction and local circulation of
11 the SARS-CoV-2 Gamma VOC in NYC in March-April 2021. Our analysis shows that nearly all circulating
12 Gamma transmission clusters were introduced from other US locations and not from outside the US,
13 indicating that whereas international travel restrictions might keep the number of introductions from the
14 country where that variant was first detected to a minimum, they do not prevent the lineage from
15 establishing in another location even if they are in place before the variant emerges. We also show that
16 even a comprehensive genomic surveillance system may not be able to detect imported novel variants
17 before they establish persistent transmission clusters.

18

19 **METHODS**

20 *Sequence data*

21 To characterize the emergence of the Gamma variant in NYC, we downloaded all complete high-
22 quality (<5% ambiguous sites) with an available date of sampling (day, month, year) SARS-CoV-2
23 Gamma genomes available through the GISAID database from March 10th to June 5th as of June 5th (N =
24 15,967) [13]. After removing duplicate sequences from all locations except for the location of interest
25 (NYC), we aligned the remaining Gamma sequences using MAFFT v.7.453 [14], trimmed the 5' and 3'
26 ends and masked all problematic sites in the alignment as suggested before [15].

1 IRB of UC San Diego waived ethical approval of this work as human subjects exempt.

2 *Phylogenetic analysis and downsampling*

3 To identify phylogenetic clusters comprising of lineages predominantly circulating in NYC, we first
4 inferred a maximum parsimony (MP) phylogenetic tree representing the Gamma global diversity using all
5 available unique sequences and IQTreev.2.0.3 software [16]. We then removed all clades that did not
6 have NYC sequences and were not ancestral to clades that contained NYC sequences. For large
7 polytomies, a random sub-sampling of genomes was performed. The resulting subsampled dataset was
8 further used to reconstruct a maximum likelihood (ML) phylogenetic tree under the GTR+F+I nucleotide
9 substitution model, collapsing polytomies, and enforcing a minimum branch length of $1e-9$.

10

11 *Phylogeographic analysis*

12 We performed Bayesian phylogeographic reconstruction in BEAST1.10.4 [17]. First, we reconstructed
13 time-scaled phylogenetic trees based on the genomes in the subsampled dataset under an HKY85
14 nucleotide substitution model, assuming a strict molecular clock model (fixed to the value of 8×10^{-4}
15 nt/site/year) and a Bayesian skyline coalescent tree prior [18]. We ran 3 independent MCMC chains of
16 1×10^9 generations; for all BEAST analyses, we used Tracer [19] to assess convergence of MCMC chains
17 and to ensure ESS > 100 for all parameters. We used the LogCombiner package from BEAST to combine
18 the tree distributions from these three independent chains and resample them at a lower frequency to
19 obtain a distribution of 1,000 trees. The phylogeographic reconstruction of viral lineage movement was
20 then performed using the resulting tree distribution and assigning geographic trait value to the tips of the
21 phylogeny based on the sampling locations. Phylogeographic analysis assumes that location changes
22 follow the same process as sequence evolution; we used an asymmetric model that allows estimating
23 separate incoming and outgoing viral lineage flow rates for each location. We considered 10 locations in
24 our analysis: (1) NYC, (2) Other New York locations, (3) California, (4) Florida, (5) Illinois, (6) New Jersey,
25 (7) Other USA locations, (8) Brazil, (9) Other South America Locations, (10) Other locations in the world.
26 The specific states were selected based on the high number of Gamma lineages reported. Brazil was
27 selected as the country where the Gamma VOC was first identified [2]; similarly, other South American
28 countries were included as a separate location because of the close proximity to Brazil and the high

1 number of reported Gamma cases. We then ran another MCMC analysis of 5×10^6 generations which
2 allowed us to obtain posterior tree distributions annotated with Markov jumps indicating virus migration
3 events between the specified geographical locations.

4 Gamma clades circulating in NYC were then identified in a phylogeographic context, defined as
5 clades of ≥ 2 sequences for which the ancestral location was in NYC with 90% posterior probability
6 support. The ancestral location of NYC clades was then identified as the ancestral location of the
7 ancestral node that had the highest posterior probability support.

8 *Analysis of persisting lineages*

9 Lineage persistence analysis identifies whether lineages have been introduced into the studied
10 location within a defined time interval or they were already circulating in the given location before that
11 time period (i.e. persisting lineages) [20]. To describe the proportion of circulating Gamma strains versus
12 those that were being introduced from other locations, we ran lineage persistence analysis using
13 PersistenceSummarizer as described by Lemey et al [20]. This analysis allowed us to describe the
14 proportion of lineages being introduced to NYC from other locations within each 2-week period and the
15 proportion of lineages that continued circulating in NYC from before the start of the period (those being
16 “re-introduced” from within NYC). We split the time into 2-week intervals going backward from the most
17 recently sampled sequence (June 5th, 2021). Since the first Gamma sequence was identified in NYC on
18 March 10th, 2021, we analyzed the introduction and persistence of circulating lineages for six 2-week time
19 periods starting two weeks before the first week with a non-zero number of Gamma sequences (February
20 6th) and until the end of the second epidemic wave in NYC (May 1st).

21

22 **RESULTS**

23 *Genetic Sequence Data and phylogeographic analysis identified 16 NYC gamma lineage clusters*

24 As of June 5th, 2021, 15,967 unique Gamma genome sequences have been deposited in GISAID
25 from around the world, including 116 generated by NYC public health surveillance. We reconstructed a
26 maximum parsimony tree using these sequences and found that Gamma lineages circulating in NYC
27 were dispersed throughout the phylogeny, indicating multiple independent introductions (Suppl. Fig.1).

1 Bayesian phylogeographic analysis performed on a subset of Gamma sequences (N = 1,461),
2 selected as “relevant” to NYC infections based on the Maximum Parsimony phylogeny, allowed us to
3 identify 16 NYC transmission clusters, defined as clades with ancestral locations estimated to be in NYC
4 (Fig. 1). NYC clusters’ sizes ranged between 2 and 108 sequences (median = 4) and included between 1
5 and 41 NYC sequences (median = 2).

6 The majority of NYC transmission clusters were introduced from Florida (N=6), followed by Illinois
7 (N=4), and other locations in the US (N=4). Only one NYC Gamma transmission cluster (cluster 10, N =
8 2) was introduced from abroad, directly from Brazil [2]. The two largest clusters, cluster 2 (N = 41) and
9 cluster 7 (N = 108), included 16 and 41 NYC genomes and originated from the states of Florida and
10 California, respectively (Fig. 1).

11

12 *The majority of Gamma lineages were persisting in NYC since early March 2021*

13 To describe patterns of introduction and persistence of Gamma lineages in NYC, we ran a lineage
14 persistence analysis on the full Gamma dataset focusing on the period between February 6th (a month
15 before the first Gamma lineage was detected in NYC) and May 1st (the end of the second epidemic wave
16 in the city). This analysis allowed us to describe whether local lineages had been circulating in NYC more
17 than two weeks prior (persisting lineages) or if they were recently introduced from other locations.

18 In February 2021, less than 1% of NYC Gamma cases were associated with lineages imported from
19 abroad. Gamma lineages were mainly introduced from “Other” US states (states that were not identified
20 as separate discrete locations in our analyses, 59% of all introductions), followed by the introductions
21 from New Jersey (16%), other parts of the New York state (9%), and Florida (8%), with the proportion of
22 introductions from all other considered locations and local NYC persistence (lineages circulating in the
23 city since 2 weeks prior) <3% each (Fig.2). By March 20th, 57% of all Gamma cases were persisting in
24 NYC for at least two weeks; among the lineages that were newly introduced to NYC between the 6th and
25 20th of March 2021, the majority came from “Other” US states (22%), New Jersey (15%), other parts of
26 New York State (4%) and Florida (2%). By the end of the second wave of the epidemic, in the second half
27 of April 2021, 43% of Gamma lineages were persisting in NYC, 24% came from other parts of New York
28 State, 9% from New Jersey, 3% from Florida, and 21% from “Other” US states. Though the majority of

1 circulating NYC Gamma clusters identified through the phylogeographic analysis were introduced from
2 Florida and Illinois, only a small proportion (around 5% and 1%, respectively) of Gamma lineages
3 introduced in the observed period were from these states.

4

5 *Gamma transmission clusters were established in NYC in February 2021*

6 Phylodynamic analysis of the times to most recent common ancestor (TMRCA) for the identified
7 Gamma transmission clusters showed that the majority of locally circulating clusters had been established
8 in February 2021. The cluster TMRCA ranged between January 5th to May 18th, 2021 (Fig. 3). The
9 TMRCA for the largest circulating clusters, clusters 2 and 7, were February 19th, 2021 (95% highest
10 posterior density [HPD] February 9 – 28, 2021) and February 25th, 2021 (95% HPD February 15th – March
11 9th, 2021), respectively. The TMRCA for the only transmission cluster introduced from Brazil was
12 estimated to be May 3rd, 2021 (95% HPD April 24th and May 5th, 2021).

13

14 *Gamma variant in neighborhoods with high pre-existing seropositivity in NYC in March – April 2021*

15 Between March 6th and May 1st, 2021, NYC PHL and PRL sequenced 11,385 SARS-CoV-2 genomes
16 and identified 116 of them as Gamma using Pangolin [21]. The two NYC neighborhoods (as defined by
17 the United Hospital Fund) with the highest number of identified Gamma genomes were in Queens and
18 Brooklyn (West Queens and East New York, Fig.4A). While the average proportion of SARS-CoV-2
19 sequences identified as Gamma remained low throughout city neighborhoods between March 6 and May
20 1 (0.7%, 95% CI 0.45 – 0.95 [22]), it was >2% in Chelsea, Greenpoint, Bayside Meadows, South East
21 Queens, and East New York in that period. Some of these areas, including South East Queens and East
22 New York, also showed some of the highest seropositivity rates observed in NYC as of March 11th, 2021 (
23 [22], Fig. 4B), suggesting a more effective Gamma transmission in the presence of other circulating
24 lineages in these areas with pre-existing immunity.

25 The first three Gamma lineages were found in Greenpoint, Bayside-Meadows, and Southeast
26 Queens before March 19th, 2021 (Suppl. Fig.2). In the second half of March, at which point most Gamma
27 lineages were already established locally and persisting in NYC, they became prevalent in West Queens
28 and Coney Island – the areas where the majority of SARS-CoV-2 lineages were circulating in March-April

1 2021. In late April 2021, when the number of SARS-CoV-2 infections dropped in NYC, the Gamma
2 lineages were most prevalent in Jamaica. Upon further investigation of the geographical spread of the
3 Gamma lineages identified in the two largest circulating NYC clusters, clusters 2 and 7, no localized
4 geographical patterns were observed: infections from both clusters were found throughout the city (Suppl.
5 Fig.2).

6

7 **DISCUSSION**

8 The emergence and swift spread of SARS-CoV-2 variants have threatened pandemic control efforts.
9 To date, countries with well-developed SARS-CoV-2 genomic surveillance systems that have enabled
10 early identification of newly emergent VOCs, such as South Africa and the United Kingdom, have been
11 subjected to travel restrictions by other countries after alerting the international community of the new
12 variant [4]. Such measures might have substantial socio-political and economic consequences, but have
13 not prevented the global dissemination of the variants and the subsequent infections growth [23]. Here we
14 show how the introduction and spread of the Gamma variant in NYC illustrates that although travel
15 restrictions might reduce introductions from the origin locations, they failed to prevent variant
16 establishment, as local NYC transmission clusters were seeded by domestic transmission.

17 Using a phylogeographic approach, we showed that despite the limited travel from Brazil - the country
18 where Gamma lineage was first identified in late 2020 [2] - being enforced since May 2020, most Gamma
19 lineages arrived in NYC and established locally circulating clusters in February 2021. All but one identified
20 Gamma transmission clusters circulating in NYC were introduced from US locations, suggesting that the
21 variant has been widely circulating in the US at the time and travel restrictions with Brazil did not prevent
22 the variant from establishing in NYC. Only one of the Gamma transmission clusters circulating in NYC
23 resulted from an introduction from outside of the US when the virus was imported from Brazil, but this
24 introduction (in May 2021) was preceded by multiple introductions from domestic locations. Similarly,
25 previous analyses of genomic, epidemiological, and travel data showed minimal evidence of direct SARS-
26 CoV-2 transmission to NYC from China, the country where the virus was first identified, in early 2020; all
27 other identified introductions that resulted in local NYC transmission were from other locations in the US
28 [11]. International travel restrictions imposed on China in early 2020 likely reduced the number of virus

1 importation events but only modestly affected epidemic trajectory globally; their effect could have been
2 greater if combined with local interventions that reduce transmission [24]. Other evidence suggests that
3 limiting travel might only have an impact on epidemic dynamics in countries with low COVID-19 incidence
4 [25]: the high daily number of COVID-19 cases in the US in early 2021 [26] might have contributed to a
5 limited role of international introductions in the Gamma transmission. We show that once introduced, a
6 large proportion of Gamma lineages persisted in NYC, also suggesting local mitigation strategies may
7 have a much more substantial impact on variant spread than international travel restrictions.

8 As expected, in the presence of other dominant SARS-COV-2 strains such as the Iota and the Alpha
9 variants, Gamma spread more effectively in areas with high seroprevalence than in other neighborhoods:
10 NYC areas with some of the highest seroprevalence levels were also those with the highest number of
11 Gamma cases. Vaccine rollout undoubtedly slowed down the spread of Delta [27], but its impact on
12 Gamma spread in NYC was unlikely as pronounced. Even though vaccination rates were accelerating in
13 early 2021 (the number of people who received at least one dose of the vaccines grew from 7% on
14 February 1st to 18.8% on March 10th to 45.4% on May 1st [22]), these efforts were delayed relative to the
15 time of Gamma introduction and establishment in the city. Thus, the most successful Gamma lineages
16 that resulted in large locally-circulating clusters were widespread in the city and were not constrained to
17 specific neighborhoods.

18 NYC dramatically increased its sequencing capacity in January-February 2021 with the
19 implementation of sequencing at PRL: in February 2021, in the middle of the second epidemic wave, 4-
20 6% of all SARS-CoV-2 infections in NYC were genotyped [22], which is much higher than average
21 (1.75%) in North America [1]. Although these efforts allowed for timely identification of the rapidly-growing
22 Iota variant that originated in NYC [12], they did not help to substantially reduce a lag between the
23 introduction and identification of Gamma lineages, which were present at much lower frequencies, in
24 NYC. This lag was similar to the one recorded in the early days of the pandemic in the absence of
25 available widespread testing when the first SARS-CoV-2 case in NYC was registered on February 29th,
26 2020 but phylodynamic analysis later revealed that the virus has likely been introduced to NYC as early
27 as January [11, 28]. Importantly, laboratory surveillance includes a necessary lag from the sample
28 collection to laboratory analysis and quality checks before a newly emerging variant can be reported. In

1 NYC, the public health laboratory is continuously improving systems and processes to minimize this lag
2 and encourages timely reporting of genomic data by hospitals and academic laboratories that led to
3 substantial improvement in volume and timeliness of genomic data in each subsequent SARS-CoV-2
4 pandemic wave.

5 Patterns of introduction from within the US showed that a small number of introductions can result in
6 a large number of locally circulating clusters. Even though <5% of the circulating NYC Gamma lineages,
7 which includes circulating Gamma clusters and singletons, were introduced from the state of Florida in
8 February-April 2021, introductions from Florida seeded the majority of circulating NYC Gamma
9 transmission clusters. Similarly, even though the overall proportion of Gamma lineages introduced from
10 Illinois was <1% at all times during the observation period, these introductions seeded 4 of the 16 (25%)
11 identified circulating transmission clusters. Such differences might stem from certain behavioral patterns
12 displayed by travelers from Florida and Illinois that resulted in more onward transmission upon arrival.
13 These patterns are similar to those previously reported for Alpha VOC (B.1.1.7): the lineage first identified
14 in the UK was initially introduced into three locations in the US (California, Florida, and Georgia) and then
15 spread through unmitigated transmission between different US locations, with Florida state playing an
16 important role in that transmission [29].

17 Similar to any such analysis, our work is affected by the disproportional sampling and sequencing
18 efforts in various US locations and abroad. As our analysis only focused on the introductions and
19 establishment of Gamma into NYC, we did not investigate importations of Gamma into other domestic
20 locations. Furthermore, our subsampling approach was based on maximum parsimony phylogenetic
21 reconstruction which could have affected the representation of different geographic locations in our final
22 dataset but was necessary to create a dataset that could be analyzed in a Bayesian framework. Both the
23 development of methods that can analyze larger datasets to avoid down-sampling, and further
24 improvement in genomic surveillance capacity proportionally across the US locations and worldwide can
25 help improve similar investigations in the future.

26 The analysis of the introduction and spread of Gamma lineages in NYC, a location with large-scale
27 genomic surveillance, shows that even in the presence of effective genomic monitoring efforts, low-
28 prevalence variants with potential public health importance can circulate undetected amidst an ongoing
29 epidemic. Even though US travel restrictions might be effective at reducing the number of direct
30 introductions from the international locations where the variant was first identified, they do not preclude
31 the establishment of local transmission clusters if the variant has already established itself elsewhere
32 domestically.

33

1 **FIGURE LEGENDS**

2 **Figure 1.** Molecular clock phylogenetic tree inferred from the subsampled Gamma lineages. Gamma
3 clusters circulating in NYC in March and April 2021 are highlighted and color-coded by ancestral location.

4
5 **Figure 2.** The proportion of Gamma lineages introduced (from New Jersey (NJ), Florida (FL), Illinois (IL),
6 California (CA), New York State (excluding NYC; NYS), other domestic sources (Other US), and Brazil)
7 and persisting (NYC) from various location in the US and abroad between February 6th and May 1st 2021.
8 The NYC portion represents lineages persisting in NYC for more than two weeks since their introduction.

9
10 **Figure 3.** TMRCAs for Gamma transmission clusters circulating in NYC in early 2021. The red dotted line
11 corresponds to March 10th, when the first Gamma sequence was identified in NYC. The blue and green
12 shaded areas correspond to the six two-week intervals in the analysis of persisting lineages. Orange
13 squares highlight TMRCAs of the two largest clusters (N = 41 and N = 108 genomes).

14
15 **Figure 4.** A) Geographical distribution of Gamma sequences sampled in NYC between March 10th and
16 May 1st, 2021. B) NYC city-wide seropositivity as of March 11th, 2021.

17

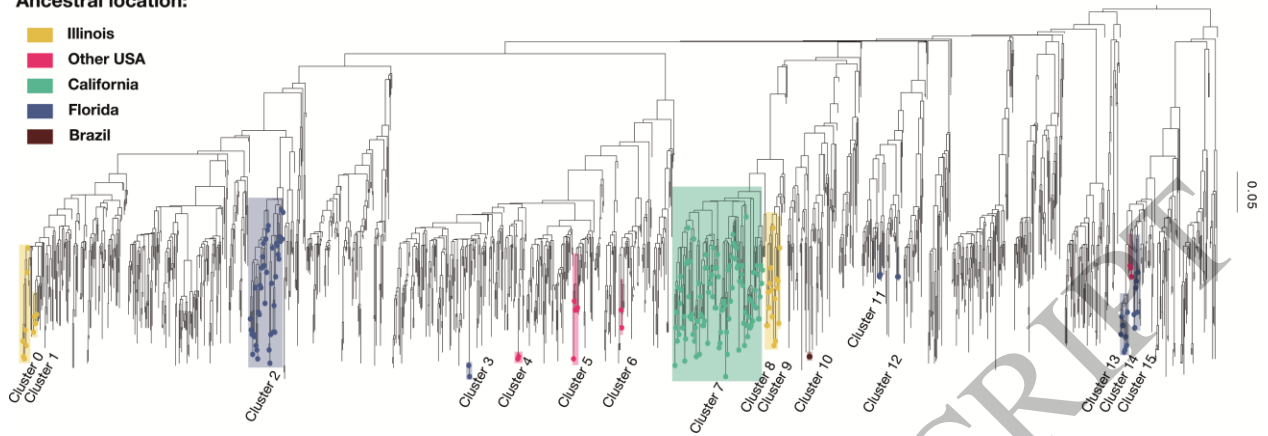
1 **Literature**

- 2 1. Oude Munnink, B.B., et al., *The next phase of SARS-CoV-2 surveillance: real-time*
3 *molecular epidemiology*. Nat Med, 2021. **27**(9): p. 1518-1524.
- 4 2. Faria, N.R., et al., *Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus,*
5 *Brazil*. Science, 2021.
- 6 3. Firestone, M.J., et al., *First Identified Cases of SARS-CoV-2 Variant P.1 in the United*
7 *States - Minnesota, January 2021*. MMWR Morb Mortal Wkly Rep, 2021. **70**(10): p. 346-
8 347.
- 9 4. KPMG, *Global Mobility COVID-19 Global Tracker*. 2022.
- 10 5. Dejnirattisai, W., et al., *Antibody evasion by the P.1 strain of SARS-CoV-2*. Cell, 2021.
11 **184**(11): p. 2939-2954 e9.
- 12 6. CDC, *Science Brief: Emerging SARS-CoV-2 Variants*. 2021.
- 13 7. Gidari, A., et al., *Cross-neutralization of SARS-CoV-2 B.1.1.7 and P.1 variants in*
14 *vaccinated, convalescent and P.1 infected*. J Infect, 2021. **83**(4): p. 467-472.
- 15 8. Hoffmann, M., et al., *SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing*
16 *antibodies*. Cell, 2021.
- 17 9. Nonaka, C.K.V., et al., *SARS-CoV-2 variant of concern P.1 (Gamma) infection in young*
18 *and middle-aged patients admitted to the intensive care units of a single hospital in*
19 *Salvador, Northeast Brazil, February 2021*. Int J Infect Dis, 2021. **111**: p. 47-54.
- 20 10. Freitas, A.R.R., et al., *The emergence of novel SARS-CoV-2 variant P.1 in Amazonas*
21 *(Brazil) was temporally associated with a change in the age and sex profile of COVID-19*
22 *mortality: A population based ecological study*. Lancet Reg Health Am, 2021. **1**: p.
23 100021.
- 24 11. Gonzalez-Reiche, A.S., et al., *Introductions and early spread of SARS-CoV-2 in the New*
25 *York City area*. Science, 2020. **369**(6501): p. 297-301.
- 26 12. West, A.P., Jr., et al., *Detection and characterization of the SARS-CoV-2 lineage B.1.526*
27 *in New York*. Nat Commun, 2021. **12**(1): p. 4886.
- 28 13. Shu, Y.L. and J. McCauley, *GISAID: Global initiative on sharing all influenza data - from*
29 *vision to reality*. Eurosurveillance, 2017. **22**(13): p. 2-4.
- 30 14. Katoh, K., G. Asimenos, and H. Toh, *Multiple alignment of DNA sequences with MAFFT*.
31 *Methods Mol Biol*, 2009. **537**: p. 39-64.
- 32 15. De Maio, N., Walker, C., Borges, R., Weilguny, L., Slodkowicz, G., Goldman, N., *Masking*
33 *strategies for SARS-CoV-2 alignments*, in *Virological.org*. 2020.
- 34 16. Minh, B.Q., et al., *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic*
35 *Inference in the Genomic Era*. Mol Biol Evol, 2020. **37**(5): p. 1530-1534.

- 1 17. Suchard, M.A., et al., *Bayesian phylogenetic and phylodynamic data integration using*
2 *BEAST 1.10*. *Virus Evol*, 2018. **4**(1): p. vey016.
- 3 18. Drummond, A.J., et al., *Bayesian coalescent inference of past population dynamics from*
4 *molecular sequences*. *Mol Biol Evol*, 2005. **22**(5): p. 1185-92.
- 5 19. Rambaut, A., et al., *Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7*.
6 *Syst Biol*, 2018. **67**(5): p. 901-904.
- 7 20. Lemey, P., et al., *Untangling introductions and persistence in COVID-19 resurgence in*
8 *Europe*. *Nature*, 2021. **595**(7869): p. 713-717.
- 9 21. O'Toole, A., et al., *Assignment of epidemiological lineages in an emerging pandemic*
10 *using the pangolin tool*. *Virus Evol*, 2021. **7**(2): p. veab064.
- 11 22. DOH, N., *COVID-19: Data*. 2022.
- 12 23. Abou-Setta, A.M., Lam, O.L.T., Kasireddy, V., Askin, N., Tricco, A.C., *Border closure and*
13 *travel restrictions to control the spread of COVID-19: an update to a Cochrane review*
14 *MedRXiv*, 2022.
- 15 24. Chinazzi, M., et al., *The effect of travel restrictions on the spread of the 2019 novel*
16 *coronavirus (COVID-19) outbreak*. *Science*, 2020. **368**(6489): p. 395-400.
- 17 25. Russell, T.W., et al., *Effect of internationally imported cases on internal spread of COVID-*
18 *19: a mathematical modelling study*. *Lancet Public Health*, 2021. **6**(1): p. e12-e20.
- 19 26. CDC, *COVID Data Tracker*. 2022.
- 20 27. Shoukat, A., et al., *Lives saved and hospitalizations averted by COVID-19 vaccination in*
21 *New York City: a modeling study*. *Lancet Reg Health Am*, 2022. **5**: p. 100085.
- 22 28. Maurano, M.T., et al., *Sequencing identifies multiple early introductions of SARS-CoV-2*
23 *to the New York City region*. *Genome Res*, 2020. **30**(12): p. 1781-1788.
- 24 29. Washington, N.L., et al., *Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the*
25 *United States*. *Cell*, 2021. **184**(10): p. 2587-2594 e7.
- 26
- 27

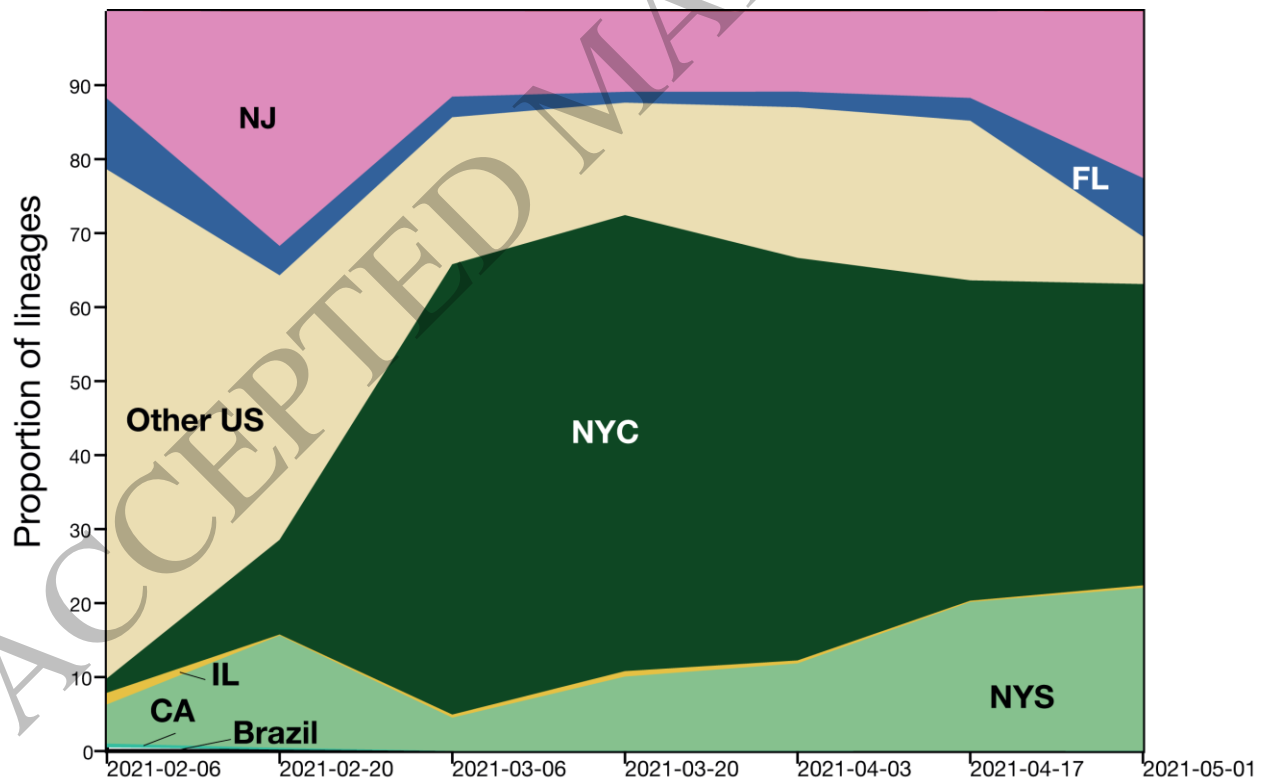
Ancestral location:

- Illinois
- Other USA
- California
- Florida
- Brazil



1
2
3
4
5
6

Figure 1
165x58 mm (.38 x DPI)



7
8
9
10

Figure 2
165x106 mm (.38 x DPI)

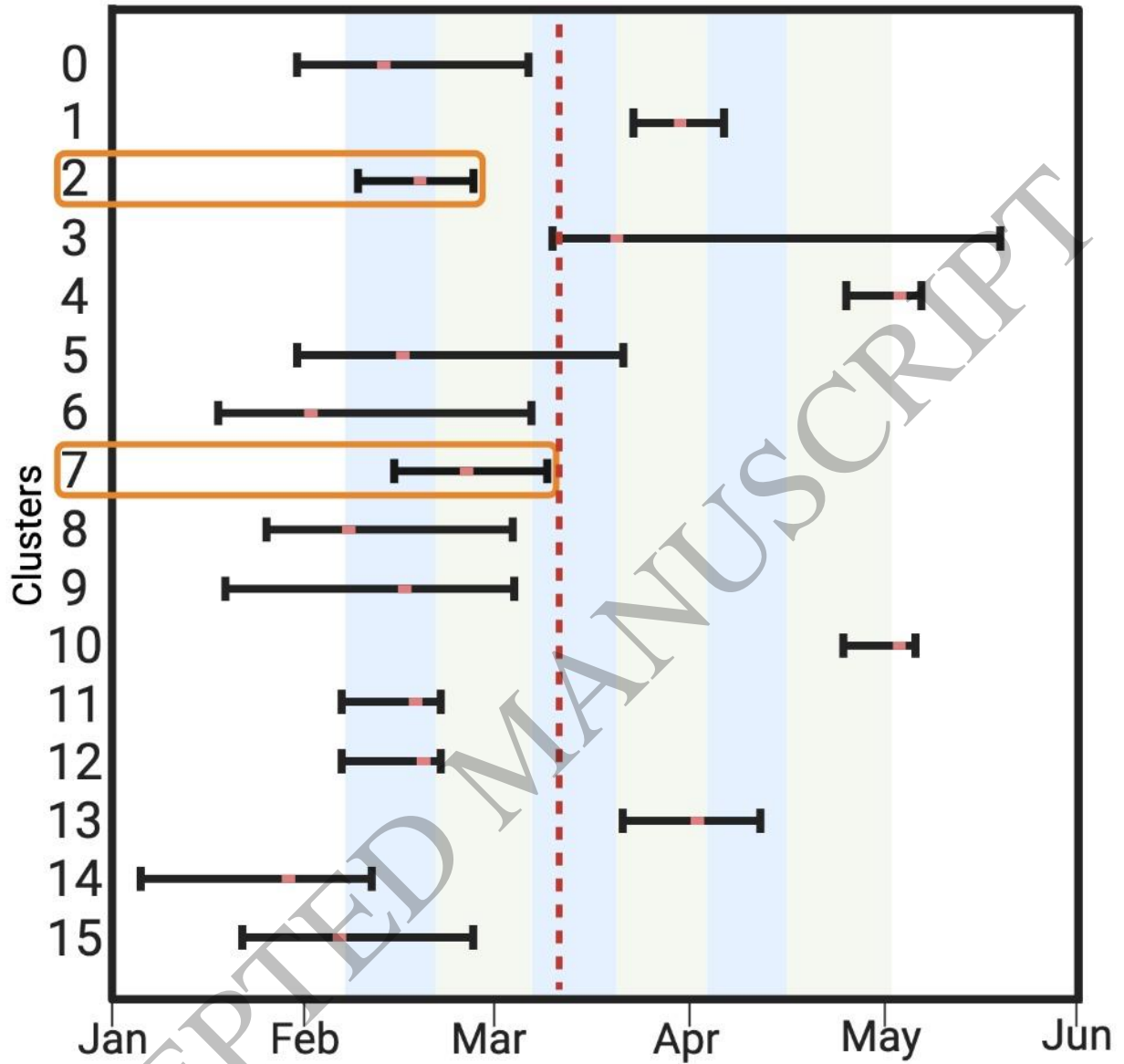
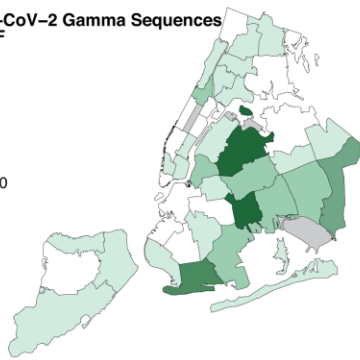
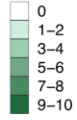


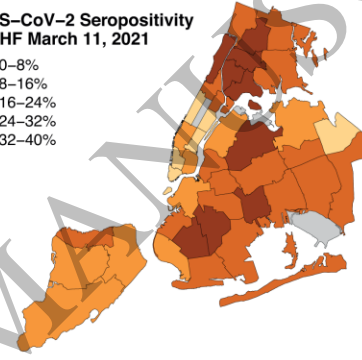
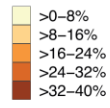
Figure 3
165x156 mm (.38 x DPI)

1
2
3
4

**SARS-CoV-2 Gamma Sequences
by UHF
n=80**



**SARS-CoV-2 Seropositivity
by UHF March 11, 2021**



1
2
3

Figure 4
165x124 mm (.38 x DPI)