Was New York City A Source Of Global Sars-Cov-2 Spread?

Cole Jensen
cole.jensen@yale.edu

Follow this and additional works at: https://elischolar.library.yale.edu/ysphtdl

Recommended Citation
https://elischolar.library.yale.edu/ysphtdl/2055

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Was New York City a Source of Global SARS-CoV-2 Spread?

A Thesis Submitted to the
Yale School of Public Health
In Partial Fulfillment of the Requirements for the
Degree of Master of Public Health

by
Cole Jensen
Master of Public Health, 2021

Under the advisement of
Dr. Nathan Grubaugh and Dr. Forrest Crawford

May 2021
Abstract

Coronavirus disease 2019 (COVID-19) was first diagnosed in the United States on January 15, 2020, in the state of Washington. Shortly after that, on February 29, 2020, the first case of COVID-19 was reported in New York City. From January 2020 through July 2020, there were 223,107 cases of COVID-19 that led to 18,787 COVID-19 related deaths in New York City alone. During this time, many left New York City to avoid contracting the virus or to return home. This travel seeded much of the transmission within the United States. We aimed to identify any international transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as well from New York City and quantify the number of lasting transmission events by combining traditional epidemiological methods with phylogenetic and genomic methods.

By subsampling more than 450,000 SARS-CoV-2 genomes publicly available up to 2020-11-02, we estimated the number of times SARS-CoV-2 was introduced from New York City to various geographic regions, not including the United States. On average, there were 6.7 different transmission events from New York City to other regions of the world between January and July of 2020. On average, Western Europe had the most introductions from New York City, followed by Northern Europe. These results were somewhat expected based on air-travel data where we saw that most travelers from New York City flew to the Caribbean, Northern Europe, and Western Europe. Overall, this study helps show that outbreaks within the United States, such as the New York City outbreak, helped fuel the international transmission events.
Acknowledgments

I would like to thank all those that sequence and upload genome sequences to GISAID for all the service they provided during the pandemic and the Centers for Disease Control and Prevention (CDC) for providing the OAG flight data. I would also like to thank Nathan Grubaugh, Anderson Brito, Guy Beale, and Forrest W. Crawford for all the support and guidance they provided during this study. I would also like to thank my wife, Rachel Jensen, for putting up with me during this project and listen to me babble on and on about graphs and trees. This work was funded by the Fast Grant from Emergent Ventures at the Mercatus Center at George Mason University (N.D.G.).
# Table of Contents

Abstract 2  
Acknowledgments 3  
List of Tables 5  
List of Figures 5  
Introduction 6  
Methods 7  
Results 9  
Discussion 14  
Conclusions 15  
References 16  
Supplement 24
List of Tables

Supplementary Table 1: Region allocations
Supplementary Table 2: Flight data containing the number of flight passengers from New York City
Supplementary Table 3: Transmission events captured in each subsample shown by region

List of Figures

Figure 1: Phylogenetic analysis results for transmission events from New York City internationally.
Figure 2: Locations of Transmission Events from New York City.
Figure 3: Flight and sequence availability.
Supplementary Figure 1: Flight passengers from New York City against the number of incident cases within New York City.
Introduction

In December 2019, it was determined that SARS-CoV-2 was the cause of a pneumonia outbreak in Wuhan, China (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses et al., 2020; Wu et al., 2020; Zhou et al., 2020; Fauver et al., 2020). As soon as January 13, 2020, travel-associated cases of COVID-19 were seen outside China (Fauver et al., 2020; WHO Situation Report-1, 2020; WHO Situation Report-63, 2020). Less than a week later, the first case of COVID-19 was diagnosed in the United States (CDC Travel Case, 2020). The virus has since spread across the globe to nearly every country (WHO Situation Report-63, 2020). As a result, the United States started to ban travel from various countries and regions starting on January 31, 2020 (Taylor, 2020). Roughly a month later, the first community-transmitted case of COVID-19 was diagnosed in the United States despite these bans (Moon et al., 2020).

Between January 21, 2020, and July 31, 2020, there have been 410,365 confirmed cases of COVID-19 and 27,600 COVID-19 related deaths in New York state. Of those confirmed cases and COVID-19 related deaths, 223,107 cases and 18,787 COVID-19 related deaths were reported from New York City alone (COVID Data Tracker, 2020). Many New Yorkers decided to leave the city as cases increased in the early wave of the pandemic. In early April, the daily incidence peaked in New York City at nearly 11,000 new cases a day (New York COVID Map, 2020). The post office received over 81,000 mail forwarding requests from New York City residents in that same month. The majority of those requests were for outside the city (Paybarah, Bloch, & Reinhard, 2020). Later in the year, it became apparent that infected travelers from New York City travel helped seed outbreaks across the United States (Carey & Glanz, 2020).

Despite this, local officials often don’t know where or how outbreaks start until they can retroactively analyze the outbreak. However, people often need a source of blame for the outbreak (Jaworsky & Oiaoan, 2020). Frequently, due to various reasons such as health policy and politics, local outbreaks are blamed on international travelers (Chinazzi et. al, 2020; Dickens et. al, 2020; DPH, 2020; Travelers, 2020; Jaworsky & Oiaoan, 2020). International travel is often listed as a potential source for a local outbreak or included in policies enacted to help slow the rate of new infections in hopes to protect “the well-being of America” (Liptak and Vazques, 2020; DPH, 2020; Governor Cumo, 2020; Governor Lamont, 2020). With this analysis, we hope to shift the focus from solely what international travelers might bring to the US to what travelers from the US might bring with them. By using phylogenetic methods to gain insight into one of the more significant outbreaks within the US, which helped seed numerous local outbreaks, we show that the New York City outbreak was also the origin of international spread of SARS-CoV-2. Therefore, showing that actions within the United States can lead to consequences elsewhere and that the US is just as involved in the further spread of the virus as other countries.
Methods

To investigate the transmission patterns of SARS-CoV-2 and possible transmission events from New York City internationally, we used phylogenetic analysis to model global spread. Due to the abundance of available SARS-CoV-2 genomes we decided that we needed subsampled genomes to ensure that only quality samples were included. Additionally, we subsampled to better infer reliable phylogenies on this data due to the large number of similar sequences and low number of mutations (Morel et al., 2020). To add further information to the claims that the phylogenetic analysis produced, we collected flight data from New York City that shows the number of passengers that flew from any of the airports in New York City (John F. Kennedy International Airport, LaGuardia Airport, and Newark International Airport). We then used the number of passengers who flew from New York City to their destination region to indicate where probable transmission events occurred.

Mortality Rates, Incidence, and Flight Data

Mortality rates were collected using the data publicly available on the Johns Hopkins Coronavirus Resource Center website (Mortality, 2020). We collected New York City and United States mortality rates from the Centers for Disease Control and Prevention (CDC), where they have separated total United States mortality and New York City's mortality rate to find individual mortality rates (United States, 2021). These rates were used to create a proportion of global mortality by country and multiplied by 1,350 to achieve the desired total genome amount from each country. The resulting number of genomes from each country were used in the phylogenetic analyses (all numbers rounding up to the nearest whole number).

To understand the spread of COVID-19 within New York City between December 2019 and July 2020, we collected daily COVID-19 incidence data from the CDC reflecting cases in New York City (United States, 2021). We also collected flight data from OAG Aviation Worldwide Limited and CDC to get an insight into where probably SARS-CoV-2 transmission events occurred. This data was aggregated and filtered to only include flights leaving New York City-based flights. It was then grouped by destination region. The region assignment was assigned using the region assignments described by the United Nations with the addition of the ‘New York City’ and ‘USA’ regions (Standard, 2021). The addition of these two regions allowed us to find introductions on a region level, while still allowing for the source to solely be New York City. In addition, it also allowed us to ignore introductions from New York City to other areas within the United States and separate those introduction events from the North America region. This region grouped method was used for the flight data as well as both the Bayesian and maximum-likelihood phylogenies. For a complete list of the number of flight passengers from New York City during the study period, see Supplementary Table 2.
Phylogenetic analyses

All publicly available complete SARS-CoV-2 genomes worldwide were based on data available from GISAID up to October 26, 2020. Since each country has different genome sequencing approaches and different infection rates, these genomes were then subsampled. All genomes used were filtered to be collected between December 1st, 2019, and July 31st, 2020, from a human host. Additionally, the genomes were then subsampled by country-level mortality rate, based on the mortality rate per 100,000 population for each country in the world and New York City. Using a subsampling scheme, we subsampled a proportion of the genomes from a country based on their mortality rate when compared to the global mortality rate. This process was repeated three different times, each with different genomes being selected, to solidify the number of introductions to each region. This allowed us to generate three different subsamples of about 1,350 genomes.

Using an augur pipeline, we performed multiple sequence alignment using MAFFT (Hadfield et al., 2018; Katoh & Standley, 2013). Following this, we were able to mask the 5’ and 3’ ends along with any problematic sites (Maio et al., 2013). Then we were able to construct a maximum likelihood (ML) tree using IQ-Tree (Minh et al., 2020). This used a general time reversible (GTR) nucleotide substitution model (Guindon & Gascuel, 2003; Yang 1994). Additionally, IQ-Tree was used to get 1,000 UFBoot replicates during the tree inference (Minh et al., 2020). TimeTree 0.8.0 was then used to reconstruct ancestral states and infer divergence times (Sagulenko et al., 2018). The resulting trees (one for each subsample) were rooted on a SARS-CoV-2 virus sequence hCov-19/Wuhan/WH01/2019 (EPI_ISL_406798). This provided an insight into what we might expect from a Bayesian phylogenetic state reconstruction. During the creation of the maximum likelihood trees, region-level metadata was assembled using country-level metadata. Depending on the country of origin, a region label was assigned to each genome in the subsample (see Supplementary Table 1).

TempEst was then used to perform quality control measures. Using these trees as inputs for a root-to-tip analysis (Rambaut et al., 2016). We removed dozens of genomes (n=36,38,26) from each subsample to remove outliers with residuals above ± 0.00025 subs/site. These smaller datasets were then rerun through the pipeline.

Finally, we performed Bayesian inference of the ancestral states using BEAST v1.10 (Suchard et al., 2018). We used the time-scaled maximum-likelihood trees as fixed topologies, sampling every 1,000 generations, through 100,000,000 MCMC iterations, using the BEAGLE library (v3.1.0) to accelerate the computation (Ayres et al., 2019). We choose this approach to obtain the Bayesian inferred ancestral states without having to wait for the time needed to
compute a full Bayesian solution. This led to convergence and thorough mixing, with all parameters obtaining an ESS greater than 200 when evaluated with Tracer 1.7 (Rambauet et al., 2018). After discarding 20% of the sampled tree as burn-in, TreeAnnotator was used to get the final trees. These trees were identical to the TimeTree output but had BEAST inferred ancestral states. Then we combined the ML bootstrap support and Bayesian inference of ancestral states with the time-scaled maximum-likelihood trees. To identify any independent transmission events of SARS-CoV-2 from New York City, we found clades (UFBoot > 70; MRCA discrete state probability > 0.7) with at least three taxa (see Figure 1 D). These introductions were deemed lasting transmission (non-singleton) events, while clades with less than three taxa were considered to be singleton events (du Plessis et al., 2021). We did this as the assumption of a singleton as an independent introduction is questionable. This is particularly true when the sample is sparse or unrepresentative, and the proportion of sequencing is low (0.13% sequenced samples out of all cases) (du Plessis et al., 2021).

Results

SARS-CoV-2 was first detected in New York City on February 29th, 2020, with case numbers peaking in April of that year (Gonzalez-Reiche et al., 2020; United States, 2021). Studies have shown that SARS-CoV-2 was likely to have started circulating before February 29th in New York City, before the many US and non-US issued travel bans took place (Silverman, 2020; Chubb, 2020; Travelers, 2021). Due to the wide gaps in surveillance, continued travel, and the outbreak of COVID-19 in New York City, we hypothesized that travel from New York City would lead to transmission events of SARS-CoV-2 outside the United States between February 19, 2020, and June 25, 2020 (see Supplementary Figure 1).

To investigate whether SARS-CoV-2 was spread internationally from travel related to New York City, we used travel surveillance data to reveal where probable transmission events would have occurred (fig map/table airlines). Our data demonstrate that the majority of international travel from New York City between January 2020 and July 2020 indicates that the most probable regions for transmission to occur were the Caribbean, Northern Europe, and Western Europe (see Figure 3A).
Figure 1. Phylogenetic analysis results for transmission events from New York City internationally.

A. A maximum likelihood phylogeny colored by region. B. This is the clade within the above phylogeny that highlights the transmission events from New York City. C. Time-informed maximum likelihood phylogeny clades showing where transmission events from New York City can be found. D. Where the average number of transmission events were introduced. The size of the region is a proportion of total average introduction events.
However, based on the phylogenetic analysis, we found different results. Using the cutoff values of UFBoot > 70 and MRCA discrete state probability > 0.7 which were found to best emulate Bayesian inference, we estimate that there were only 6.67 transmission events from New York City, resulting in sustained transmission chains (Alpert et al., 2021). Additionally, there were two transmission events into Western Europe, one transmission event into Northern Europe, and no transmission events into the Caribbean, unlike what we expected based on travel data. Interestingly, based on available genomic data, our analyses have not detected any transmission events from New York City into Southeastern Asia, Southern Asia, Central Asia, Eastern Europe, Middle Africa, Eastern Africa, Southern Africa, North America, and the Caribbean (see Figure 2 A). Each subsample found varying transmission events by region except for transmission events into Northern Europe (see Figure 2 B-D). The largest range of captured transmission events between the subsamples was two. In addition to capturing the transmission events from New York City to various geographic regions, we were able to estimate that these events happened both before and after travel restrictions were imposed both by the United States and by other countries. For values captured by each subsample, see supplementary table 3.
When we looked into each geographic region, we found that, with the exception of the Caribbean, there was a slight positive correlation (0.528) between the number of available sequences and the number of passengers on flights from New York City to that region (see Figure 3 B). The Caribbean had the most flight passengers from New York City, but it also had the least amount of publicly available genomes (see Figures 3 A and 3 C). This limited the chances of detecting sustained transmissions from New York City to that region. When we looked into the correlation between the average number of transmission events from New York City and the number of flight passengers, and only accounted for regions with more than one thousand genomes, we also found a positive correlation of 0.632 (see Figure 3 D). However, there were still regions that we didn’t capture any transmission events that still met this criterion.

Figure 2. Locations of Transmission Events from New York City.
A. A map and figure legend for where transmission events from New York City were captured. The arrow width is the number of average transmission events captured to that region. B-D. The transmission events captured in the various subsamples. E. Box plots that show the spread of the events captured between subsamples by geographic region.
Figure 3. Flight and sequence availability.
A. The proportion of flight passengers that flew from New York City. Their end destination is considered the region to which they travelled. B. A scatterplot of the number of publicly available genomes by geographic region at the time of subsampling by the number of flight passengers. The total number of genomes here are show on a log scale. C. A bar graph that shows the relationship between the number of publicly available genomes by geographic location on a log scale. The numbers above each column represent the total number of available genomes at the time of subsampling in that region. D. A scatterplot that shows the relationship between the average number of transmission events captured and the number of flight passengers by region. Only regions with one thousand or more publicly available genomes are shown.
Discussion

By shifting the focus from what international travelers might bring to the US and focusing on what US travelers might bring to other regions, we found evidence that transmission events of SARS-CoV-2 from New York City to geographic regions outside the United States occurred. During our study period of December 2019 through July 2020, New York City had varying amounts of restrictions. During the beginning of the study period, there were little to no COVID-19 related restrictions in New York City until early March, despite the area becoming a hot spot of transmission (Governor, 2020; New York State, 2021; NYC, 2020). During this period, many residents and tourists of the area were unknowingly transmitting the virus. Using flight passenger data, we estimated that New York City originating transmission events would most likely occur in the Caribbean, Northern Europe, and Western Europe. Our study provides that there were independent COVID-19 introductions from New York City to multiple geographic regions. Moreover, we found that all transmission events captured, regardless of region, were non-singleton introductions.

By only looking at one US city, we were able to see multiple transmission events, including numerous lasting transmission events, across the globe. These findings, therefore, highlight the need to view US travelers as a potential giver of SARS-CoV-2 to other regions. The US has been the recipient of virus transmission during this pandemic, but through our research we can see that it has also dealt its fair share of transmission.

To ensure that we were capturing transmission events, despite subsampling the data, we created three different subsamples of genomic data and created phylogenetic models for each dataset. It was also compelling to see that each subsample performed very similarly. Each model began with about 1,350 genomes, and after quality control measures, each ended with roughly 1,320 genomes. After constructing the phylogenetic trees, each subsample also captured lasting transmission events to the same regions. Additionally, the number of singleton introductions was roughly the same. Different subsamples help us highlight the variability of phylogenetic models while also ensuring that true transmission events were captured.

To create our phylogenetic trees, we took a nontraditional approach. Instead of using purely Bayesian phylogenetic methods, such as a traditional BEAST analysis, we instead used a combination of methods and meshed them together. This allowed for various things to occur, such as quick results, the ability to set thresholds, and the potential to compare future traditional Bayesian results to this analysis to check validity. By using BEAST only to determine ancestral state reconstruction instead of performing the complete analysis, we were able to produce results in days rather than weeks or months. In addition, by using a bootstrapped IQ tree with branch lengths determined by TreeTime, we could set cutoff thresholds. This allowed us to set thresholds that resemble previous traditional BEAST results without the wait. This meshed approach could be useful for near real-time genomic epidemiology to produce robust and meaningful results without the wait time associated with Bayesian phylogenetics.
Our study does have several limitations. First, the full extent of transmission events, both lasting and singleton, were likely not captured. It’s important to note that due to the variability among sequencing capacity, various regions and countries do not have the same ratio of publicly available genomes as other geographic regions. Therefore, due to dependence on genomes, we may not have captured all of the transmission events from New York City or all of the lasting transmission events. For example, most air passengers during the study period flew from New York City to the Caribbean. However, we didn’t find any transmission events occurring from New York City to that region. This is largely due to the lack of publicly available genomes from the Caribbean. While there are hundreds of thousands of publicly available SARS-CoV-2 genomes, not all regions have the same amount of sequences available. This may have influenced the results of other regions as well. Additionally, our analysis does not account for behavior beyond air travel. In reality, individual behavior, local health policies, and local conditions play important roles in transmission patterns and could help explain why New York City originating transmission events were successful in various regions while others were not. Finally, by subsampling the genomes and by not being able to observe all infections, we may have also failed to capture transmission events. But the lack of available genomic data made the inclusion of more genomes from highly sequenced regions difficult. As more sequencing data becomes available, efforts to capture transmission events should be revisited.

Notwithstanding these limitations, our analysis provides insights into the introduction, exportation, and establishment of SARS-CoV-2. This was made possible by linking travel data, epidemiological data, and genomic data together. Public health authorities should be operating with the assumption that the United States is playing a role in international transmission and establishment of SARS-CoV-2. Therefore, the use of genomic epidemiology and traditional epidemiological methods should be prioritized to help understand and slow the transmission of SARS-CoV-2.

Conclusions

By combining travel data with genomic epidemiology, we estimated transmission events from New York City to different world regions. All of these transmission events lead to non-singleton transmission within the region. This debunks the notion that the United States was only a recipient of virus introduction and not the giver of virus transmission. By looking at the outbreak in New York City, we were able to see how the US helped further international outbreaks. However, it is critical to acknowledge that due to the limited amount of publicly available genomes from various geographic regions and the sample size of the analysis, the strength of the conclusion drawn was impacted. It is therefore essential that results from genomic epidemiology studies are validated with epidemiological data.
References


Baltic. Retrieved from https://pypi.org/project/baltic


Chubb, H. (2020). These are the 26 countries and territories open to american tourists as travel ban amended. Retrieved from [https://people.com/travel/these-26-countries-are-open-to-american-tourists-as-state-dept-amends-do-not-travel-order/](https://people.com/travel/these-26-countries-are-open-to-american-tourists-as-state-dept-amends-do-not-travel-order/)


DPH issues COVID-19 alert to danbury community after significant spike in new cases(2020).

[Abstract]. Connecticut State Department of Public Health, Retrieved from


First travel-related case of 2019 novel coronavirus detected in the united states. (2020). Centers for Disease Control and Prevention, Retrieved from


Liptak, K., & Vazquez, M. (2020, Mar 11,). Trump says he's suspending travel from Europe to US, though citizens and others are exempt. CNN Wire Service. Retrieved from https://search.proquest.com/docview/2375977271


Moon, S., Yan, H., Christensen, J., & Maxouris, C. (2020, Feb 27,). The CDC has changed its criteria for testing patients for coronavirus after the first case of unknown origin was confirmed. *CNN Wire Service* Retrieved from https://search.proquest.com/docview/2364917123


https://www1.nyc.gov/site/coronavirus/index.page#:~:text=Halls%20are%20restricted%20to%2050,patrons%20must%20be%20installed.

Novel coronavirus (2019-nCoV)

SITUATION REPORT - 1. (2020). World Health Organization, Retrieved from


Silverman, H. Travelers will face new restrictions and cancellations as coronavirus cases grow in the US. (2020, Mar 2,). CNN Wire


Viruses, Coronaviridae Study Group Of The International Committee On Taxonomy Of,

(2020). Coronaviridae study doi:10.1038/s41564-020-0695-z

579(7798), 265-269. doi:10.1038/s41586-020-2008-3

Yale SARS-CoV-2 genomic surveillance initiative. (2021). Retrieved from
https://covidtrackerct.com/portfolio/current/

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with
variable rates over sites: Approximate methods. Journal of Molecular Evolution, 39(3),
306-314. doi:10.1007/BF00160154

outbreak associated with a new coronavirus of probable bat origin. Nature (London),
579(7798), 270-273. doi:10.1038/s41586-020-2012-7
Supplement

To ensure that we were capturing introductions from New York City to different geographic regions, we needed a way to ensure that each genome that belonged to a specific region was grouped within that region. As such, we used the standard country methodology given by the United Nations statistics division with the addition of the ‘USA’ and ‘New York City.’ These two additional regions were included for our analysis purposes to see where transmission events from New York City on a region level occurred and to make sure that transmission events captured from New York City to other locations within the United States were able to be differentiated from transmission events to other North American countries.

**Supplementary Table 1: Region allocations.** Below shows each geographic region used in the analysis. Each row contains the region name and what countries or locations were grouped into that region. For this analysis, the United States and New York City were listed as individual regions outside North America. Countries with multiple spellings are denoted as country name (secondary country name).

<table>
<thead>
<tr>
<th>Region</th>
<th>Countries Within the Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oceania</td>
<td>Australia, Christmas Island, Cocos, Heard Island and McDonald Islands, New Zealand, Norfolk Island, Melanesia, Fiji, New Caledonia, Papua New Guinea, Solomon Islands, Vanuatu, Micronesia, Guam, Kiribati, Marshall Islands, Micronesia, Nauru, Northern Mariana Islands, Palau, Polynesia, American Samoa, Cook Islands, French Polynesia, Niue, Pitcairn Islands, Samoa, Tokelau, Tonga, Tuvalu, Wallis and Futuna</td>
</tr>
<tr>
<td>Eastern Asia</td>
<td>China, Hong Kong, Macau, Japan, Mongolia, North Korea, South Korea, Taiwan</td>
</tr>
<tr>
<td>Southeastern Asia</td>
<td>Brunei, Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar(Burma), Philippines, Singapore, Thailand, Timor-Leste, Vietnam</td>
</tr>
<tr>
<td>Southern Asia</td>
<td>Afghanistan, Bangladesh, Bhutan, India, Iran, Maldives, Nepal, Pakistan, Sri Lanka</td>
</tr>
<tr>
<td>Central Asia</td>
<td>Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan</td>
</tr>
<tr>
<td>Western Asia</td>
<td>Armenia, Azerbaijan, Bahrain, Cyprus, Georgia, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, State of Palestine, Syria, Turkey, United Arab Emirates, Yemen</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>Åland Islands, Denmark, Estonia, Faroe Islands, Finland, Guernsey, Iceland, Ireland, Isle of Man, Jersey, Latvia, Lithuania, Norway, Sark, Svalbard and Jan Mayen, Sweden, United Kingdom</td>
</tr>
<tr>
<td>Western Europe</td>
<td>Austria, Belgium, France, Germany, Liechtenstein, Luxembourg, Monaco, Netherlands, Switzerland</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>Belarus, Bulgaria, Czech Republic (Czechia), Hungary, Moldova, Poland, Romania, Russia, Slovakia, Ukraine</td>
</tr>
</tbody>
</table>
Supplementary Table 2: Flight data containing the number of flight passengers from New York City. Listed are the number of passengers flying from one of the New York City airports internationally. The number of flights is shown by month and separated by region of destination.

<table>
<thead>
<tr>
<th>Region</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oceania</td>
<td>18,177</td>
<td>12,539</td>
<td>7,291</td>
<td>282</td>
<td>380</td>
<td>446</td>
<td>1,586</td>
</tr>
</tbody>
</table>

New York City

United States of America
<table>
<thead>
<tr>
<th>Region</th>
<th>First Subsample</th>
<th>Second Subsample</th>
<th>Third Subsample</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oceania</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>South East Asia</td>
<td>124,386</td>
<td>78,801</td>
<td>54,191</td>
<td>11,586</td>
</tr>
<tr>
<td>South East Asia</td>
<td>50,130</td>
<td>29,469</td>
<td>15,749</td>
<td>1,041</td>
</tr>
<tr>
<td>South Asia</td>
<td>73,043</td>
<td>57,603</td>
<td>35,340</td>
<td>2,092</td>
</tr>
<tr>
<td>Central Asia</td>
<td>3,157</td>
<td>2,435</td>
<td>1,358</td>
<td>20</td>
</tr>
<tr>
<td>Western Asia</td>
<td>73,979</td>
<td>61,964</td>
<td>44,121</td>
<td>4,795</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>168,356</td>
<td>159,366</td>
<td>62,589</td>
<td>8,829</td>
</tr>
<tr>
<td>Western Europe</td>
<td>142,773</td>
<td>128,006</td>
<td>58,582</td>
<td>6,909</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>34,236</td>
<td>25,800</td>
<td>16,628</td>
<td>3,849</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>117,185</td>
<td>94,962</td>
<td>43,464</td>
<td>2,865</td>
</tr>
<tr>
<td>Northern Africa</td>
<td>14,914</td>
<td>12,390</td>
<td>8,322</td>
<td>131</td>
</tr>
<tr>
<td>Western Africa</td>
<td>11,054</td>
<td>8,477</td>
<td>6,927</td>
<td>134</td>
</tr>
<tr>
<td>Middle Africa</td>
<td>723</td>
<td>530</td>
<td>286</td>
<td>1</td>
</tr>
<tr>
<td>Eastern Africa</td>
<td>13,403</td>
<td>11,680</td>
<td>5,946</td>
<td>311</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>8,565</td>
<td>6,472</td>
<td>4,584</td>
<td>22</td>
</tr>
<tr>
<td>South America</td>
<td>113,540</td>
<td>102,576</td>
<td>61,460</td>
<td>2,700</td>
</tr>
<tr>
<td>Central America</td>
<td>142,359</td>
<td>125,227</td>
<td>70,698</td>
<td>6,279</td>
</tr>
<tr>
<td>North America*</td>
<td>109,705</td>
<td>100,636</td>
<td>74,599</td>
<td>2,913</td>
</tr>
<tr>
<td>Caribbean</td>
<td>372,642</td>
<td>333,469</td>
<td>136,756</td>
<td>14,025</td>
</tr>
</tbody>
</table>

*Flights from New York City to destinations within the United States are not included in this table.
<table>
<thead>
<tr>
<th>Region</th>
<th>1</th>
<th>2</th>
<th>1</th>
<th>1.33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Asia</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1.33</td>
</tr>
<tr>
<td>South Eastern Asia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Southern Asia</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td>Central Asia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Western Asia</td>
<td>9</td>
<td>4</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>11</td>
<td>13</td>
<td>19</td>
<td>14.33</td>
</tr>
<tr>
<td>Western Europe</td>
<td>24</td>
<td>18</td>
<td>20</td>
<td>20.67</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2.67</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2.33</td>
</tr>
<tr>
<td>Northern Africa</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Western Africa</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td>Middle Africa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Eastern Africa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>South America</td>
<td>17</td>
<td>10</td>
<td>16</td>
<td>14.33</td>
</tr>
<tr>
<td>Central America</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>7.67</td>
</tr>
<tr>
<td>North America*</td>
<td>8</td>
<td>14</td>
<td>15</td>
<td>12.33</td>
</tr>
<tr>
<td>Caribbean</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Transmission events from New York City that spread within the United States are not included in this table.
Supplementary Figure 1. Flight passengers from New York City against the number of incident cases within New York City. Each line corresponds with a geographic region that was travelled to from New York City between January 2020 and the end of July 2020. The bar graph in the background is the number of incident cases within the city by epiweek in the same time period. It wasn’t until March that New York city was officially reporting cases of COVID-19. Cases peaked in the end of March and the beginning of April during this study period.