

# UC San Diego

## UC San Diego Previously Published Works

### Title

Genomic surveillance reveals dynamic shifts in the connectivity of COVID-19 epidemics.

### Permalink

<https://escholarship.org/uc/item/2bc8z2t0>

### Journal

Cell, 186(26)

### Authors

Matteson, Nathaniel

Hassler, Gabriel

Kurzban, Ezra

et al.

### Publication Date

2023-12-21

### DOI

10.1016/j.cell.2023.11.024

Peer reviewed



# HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2024 December 21.

Published in final edited form as:

Cell. 2023 December 21; 186(26): 5690–5704.e20. doi:10.1016/j.cell.2023.11.024.

## Genomic surveillance reveals dynamic shifts in the connectivity of COVID-19 epidemics

*A full list of authors and affiliations appears at the end of the article.*

### Summary

The maturation of genomic surveillance in the past decade has enabled tracking of the emergence and spread of epidemics at an unprecedented level. During the COVID-19 pandemic, for example, genomic data revealed that local epidemics varied considerably in the frequency of SARS-CoV-2 lineage importation and persistence, likely due to a combination of COVID-19 restrictions and changing connectivity. Here, we show that local COVID-19 epidemics are driven by regional transmission, including across international boundaries, but can become increasingly connected to distant locations following the relaxation of public health interventions. By integrating genomic, mobility, and epidemiological data, we find abundant transmission occurring between both adjacent and distant locations, supported by dynamic mobility patterns. We find that changing connectivity significantly influences local COVID-19 incidence. Our findings demonstrate a complex meaning of ‘local’ when investigating connected epidemics and emphasize the importance of collaborative interventions for pandemic prevention and mitigation.

### In Brief

Genomic surveillance, paired with mobility and epidemiological data, quantify the impact of local and international travel restrictions on SARS-CoV-2 transmission. Both phylogenetic and mobility analyses indicate that collaborative interventions are more effective than targeted border closures at reducing the transmission of SARS-CoV-2 between highly connected locations.

---

\*Corresponding authors: natem@scripps.edu (NLM), zellerm@scripps.edu (MZ), andersen@scripps.edu (KGA).

Author Contributions

Conceptualization, GWH, KG, JIL, EP, XJ, MW, KGA, SW, MZ; Methodology, NLM, GWH, KG, SW, MZ; Software, GWH, KG; Formal Analysis, NLM, GWH; Investigation, NLM; Resources, EK, MAS, SAP, DP, AH, PD, WC, AC, AR, AV, AR, CW, JH, JW, SNN, AP, ESL, EWS, ERE, RT, MC, NAB, PS, RAS, SA, TTN, TB, TO, RF, EHS, OEZ, IS, MS, JLM, AG, ARM, EM, JC, JDM, SS, SS, NM, AW, CA, ECH, JIL, KSR, KNN, KS, KMF, SLG, AB, DM, SK, NKM, RTS, AJN, HJR, JRC, MLG, JMC, DB, MI, NLW, WL, RSG, MAL, JA, BH, KJ, BO, GB, IL, JER, RF, SFK, AS, AP, BW, GM, MH, RWM, SLR, TW, IHM, RAF, LN, MMQ, AH, AM, OB, SHB, JO, DGR, MG, LCR, BJA, JG, AK, RK, LCL, GWY, AC, MZ; Data Curation, NLM, EK, MAS, SAP, OEZ, IS, MS, JLM, AG, NM, AW, CA, CMA, ECH, JIL, KSR, KNN, KS, KMF, AB, AJN, HJR, JRC, MLG, JMC, DB, MI, NLW, WL, MAL, JA, GB, JER, RAF, JO, DGR, MG, JG, AC, MZ; Writing - Original Draft, NLM; Writing - Review & Editing, NLM, KG, JIL, EP, KGA, SW, MZ; Visualization, NLM; Project Administration, EGS, RR; Funding Acquisition, KGA; All authors contributed to interpreting and reviewing the manuscript.

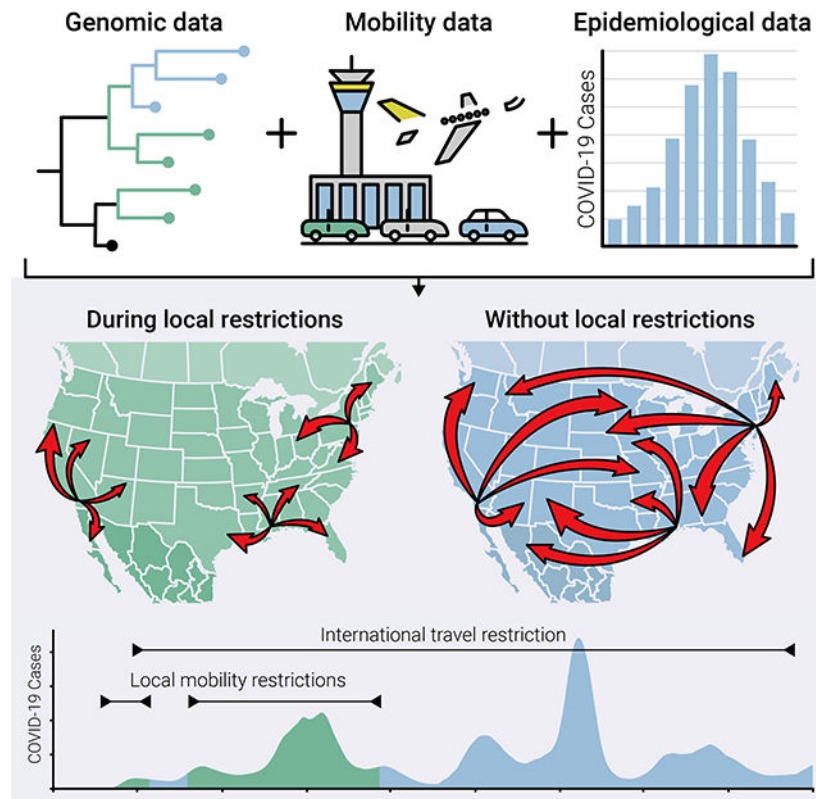
<sup>†</sup>Lead contact

**Publisher's Disclaimer:** This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Inclusion and Diversity

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. One or more of the authors of this paper self-identifies as living with a disability.

## Graphical Abstract



## Keywords

SARS-CoV-2; genomic epidemiology; phylogenetics; mobility; travel restrictions; viral sequencing

## Introduction

Human contact networks can help elucidate SARS-CoV-2 transmission dynamics. For example, it has been shown that the risk of infection for an individual increases with the number of contacts they have<sup>1</sup>, the locations they visit<sup>2</sup>, and the length of their visit<sup>3</sup>, and that the interactions of individuals between different locations hinders virus containment efforts<sup>4-7</sup>. As a result, we would expect SARS-CoV-2 transmission to be higher between geographic locations that have higher human connectivity<sup>8,9</sup>. Reconstructing the interaction networks and temporal dynamics between these “high connectivity” locations can illuminate the sources of emerging waves and help inform intervention strategies and at-risk populations<sup>10</sup>. Genomic data provides one way to measure these connectivity networks, as the geographic spread of rapidly evolving viruses like SARS-CoV-2 can be inferred from molecular data<sup>11,12</sup>.

Genomic surveillance programs, many of which were established during the COVID-19 pandemic, have generated large amounts of SARS-CoV-2 genomic data that has been

used to track the spread and evolution of the virus in near real-time<sup>13</sup>. While it is clear from genomic data that SARS-CoV-2 spreads between locations causing the reach of local outbreaks to overlap<sup>14–18</sup>, we know little about how to quantify this interaction, or the factors contributing to spread. In addition, it is unclear if the temporal dynamic of viral spread has changed as the pandemic transitions from the introductory and expansion phase into endemicity.

The initial spread of SARS-CoV-2 caused the implementation of social distancing policies and international travel restrictions. The former have been studied extensively, showing marked changes in individual behaviors and movements<sup>19,20</sup>. International travel restrictions, enacted by over 89 countries during the first five months of the pandemic<sup>21</sup>, stand in contrast to recommendations by the World Health Organization, which argued against the use of travel restrictions and border closures due to their substantial economic, social, and ethical effects, and a lack of evidence on their effectiveness<sup>22</sup>. Although some travel restrictions—including complete border closures, quarantines, and testing requirements—have since been shown to reduce imported cases<sup>17,23–26</sup>, the relative contribution of imported cases on local COVID-19 incidence before, during, and after restrictions is not fully understood. Additionally, connectivity across heavily-traveled and economically important land-borders, like the US-Mexico border, which was crossed 400 million times a year prior to the pandemic<sup>27</sup> and was closed to non-essential travel from March 19<sup>th</sup>, 2020–November 8<sup>th</sup>, 2021<sup>28,29</sup>, remains unstudied.

In this study, we characterized the connectivity of local COVID-19 epidemics over space and time, how travel restrictions affect this connectivity, and the impact of virus imports on epidemic growth. To reconstruct the dynamics of virus transmission from the beginning of the pandemic to the end of the first Omicron wave (March 2020–December 2022), we sequenced more than 82,000 SARS-CoV-2 samples as part of our routine genomic surveillance in San Diego, California and Baja California, Mexico, and compared them to SARS-CoV-2 genomes from North America and the rest of the world. By studying locations along international borders and comparing transmission in border regions to other US locations, we were able to investigate the effect of travel restrictions across the country and isolate the impact of restrictions on international borders. We found evidence for dynamic shifts in transmission of SARS-CoV-2 over the course of the pandemic regardless of location or border status. Our findings indicate that connectivity between locations plays an increasing role in maintaining local epidemics, which highlights the need for collaboration between regional and international governments to enact effective prevention strategies.

## Results

### North American locations experienced similar changes in connectivity over time

To understand how temporal and geographic changes in the connectivity of North American locations influenced SARS-CoV-2 transmission patterns, we investigated SARS-CoV-2 genomic data during the first five waves of the pandemic, until December 2022. We found that locations in North America became more connected over time, likely as a result of the easing of COVID-19 mandates.

We used phylogenetic similarity as a proxy for connectivity, since an increase in connectivity between locations would lead to an increased similarity of their viral populations due to frequent transmission between the locations<sup>30</sup>. We calculated phylogenetic similarity in the viral populations of epidemics in North American counties, states, or provinces, for each month of the pandemic (Figure 1A). To do this, we used the PhyloSor similarity<sup>31</sup> metric, which quantifies the similarity of viral populations as the proportion of branch lengths in a phylogenetic tree that are shared relative to the total branch lengths of both populations<sup>31</sup> (Supp. Figure 1). We found that in simulations, PhyloSor similarity recapitulated the number of contacts between communities (see methods; Supp. Figure 2) and was thus an appropriate metric to investigate connectivity.

Prior studies have found that local COVID-19 epidemics were regionally isolated during the beginning of the pandemic<sup>15–17</sup>, when travel restrictions and social-distancing measures were the most stringent. Our PhyloSor analysis showed that North America had highly similar virus populations at the onset of the pandemic, but we found that virus populations grew increasingly divergent, until they started to become more similar again after May 2020 (Figure 1A). This trend was negatively correlated with the stringency of COVID-19 mandates as measured by the Oxford COVID-19 Government Response Tracker<sup>32</sup>, suggesting that the relaxation of restrictions was associated with increased connectivity of North American locations (Pearson  $R = -0.68$  [95% CI:  $-0.43$  to  $-0.83$ ];  $P < 0.001$ ).

To determine whether this observed trend was consistent across most locations or driven by just a few specific locations, we calculated the heterogeneity in each location's temporal pairwise similarity to all other locations using the Gini index, a commonly used statistical measure of dispersion (Supp. Figure 3). We found that heterogeneity remained low throughout the entire duration of the pandemic, indicating that the trend in connectivity was experienced by most North American locations rather than being limited to a few influential locations. To better understand factors that explain the trend in connectivity, we also performed detailed analyses of SARS-CoV-2 diversity at a local scale, where transmission and mobility patterns could be more easily interpreted. We used results from our PhyloSor analysis to rank North American counties, states, and provinces by their average phylogenetic similarity to all other locations over time (Figure 1B). We found that metropolitan counties exhibited comparably high average phylogenetic similarity, without any notable outliers, further suggesting that the observed trend of connectivity in North America was not driven by a small number of influential locations.

### **Connectivity of local COVID-19 epidemics became more widespread over time**

To assess if travel restrictions and other COVID-19 mandates affected border locations in the same way, we examined locations along the US-Mexico border in more depth. We selected San Diego County (henceforth San Diego) because it is a popular vacation destination<sup>33</sup> and, along with Tijuana, Baja California, Mexico, contains the busiest international border crossing in the Western Hemisphere<sup>34</sup>. We found that the average phylogenetic similarity of San Diego to all other North America locations did not fall in either extreme (18<sup>th</sup> percentile) and was representative of the widespread trend in connectivity, suggesting that border locations were not uniquely impacted by travel restrictions (Figure 1B). To

investigate widespread trends in the diversity of SARS-CoV-2 at the local level, we generated and analyzed 80,323 and 1,950 SARS-CoV-2 genomes from San Diego and Baja California, respectively, making San Diego one of the most densely sampled locations in the United States (collection dates from March 25<sup>th</sup>, 2020 to December 13<sup>th</sup>, 2022; Supp. Figure 4).

More specifically, we quantified the phylogenetic similarity of SARS-CoV-2 from San Diego with other counties in California and North American states throughout the pandemic and found that SARS-CoV-2 genomes from San Diego were phylogenetically most similar to nearby locations (Figure 1C). The five locations with the highest median phylogenetic similarity to San Diego were Los Angeles, Orange, and Alameda counties, as well as Arizona and Nevada (Figure 1C). However, we found only a weak correlation between each location's median phylogenetic similarity to San Diego and its geographic proximity to San Diego, which is generally a proxy for human mobility<sup>35,36</sup> (Supp. Figure 5; Pearson p value = 0.003;  $R^2$  value = 0.10). We observed a correlation when we considered only US states (Pearson p value = 0.006;  $R^2$  = 0.14) or only California counties (Pearson p value = 0.156;  $R^2$  = 0.09), though it was not significant within California. Despite the lack of a strong correlation between geographic distance and phylogenetic similarity at any scale, high phylogenetic similarity between neighboring locations suggests that proximal locations across the US remained consistently connected throughout the entire pandemic regardless of COVID-19 restrictions.

We hypothesized that the lack of a strong correlation between geographic distance and phylogenetic similarity resulted from an increase in the long range connectivity of locations following the relaxation of COVID-19 mandates. To test this using our data from San Diego, we calculated how the correlation between phylogenetic similarity and geographic distance to San Diego for all locations changed through each of the five waves of the COVID-19 pandemic and for the intervening troughs (Supp. Figure 6 & Figure 1D). We found that phylogenetic similarity was correlated with geographic distance during the first two waves, but not during the following three waves (Pearson p value < 0.05 for 1<sup>st</sup>-2<sup>nd</sup> waves;  $R^2$  > 0.15 for 1<sup>st</sup>-2<sup>nd</sup> waves; Figure 1D). This shows that the connectivity of locations increased in geographic reach after the first two waves of the pandemic.

In contrast to the waves of the pandemic, we noticed that a significant correlation between phylogenetic similarity and geographic distance persisted during the periods of low COVID-19 incidence (Figure 1D). This led us to investigate the association between connectivity and COVID-19 incidence. We found that, even after subsampling to correct for potential sampling biases (Supp. Figure 7), similarity between San Diego and other locations increased during periods of low COVID-19 incidence relative to the adjacent waves (Figure 1E–F). The effect was most pronounced in the locations with the highest phylogenetic similarity to San Diego, suggesting that increased viral diversity during periods of high incidence was responsible for the observed reduction in similarity between locations.

### **SARS-CoV-2 population similarity is driven by transmission frequency**

A key component of understanding transmission dynamics is determining if phylogenetic similarity between locations is due to transmission between them (“bidirectional

transmission”) or shared introduction sources. Prompted by the phylogenetic similarity of San Diego to both California (US) and Baja California (Mexico), we also considered whether these factors differed between domestic and international locations. To estimate the relative amount of SARS-COV-2 transmission between locations, we reconstructed the timing and number of geographic transitions (also called Markov jumps<sup>37</sup>) into and out of San Diego across the full posterior of a Bayesian phylogeographic reconstruction (Figure 2A). Despite all locations having similar numbers of genomes in the phylogeny (see Methods), we observed the most transitions between San Diego and neighboring and domestic locations (Figure 2B). Conversely, we found relatively few transitions between San Diego and international locations (excluding Baja California, which borders San Diego; Figure 2B), indicating that transmission frequency was consistent with viral similarity.

To investigate whether transmission was impacted by COVID-19 mandates, we next examined the temporal dynamics of bidirectional transmissions between San Diego and other locations. We used geographic transitions to do this rather than phylogenetic similarity, because of the risk that phylogenetic similarity might overestimate connectivity due to shared introduction sources. In fact, when we compared the percentage of transmission into each location that originated from each other location across the posterior distribution of trees (I.e. the introduction profile; Supp. Figure 8), we found that Los Angeles and San Diego had more similar introduction profiles than Baja California and San Diego (Los Angeles vs. San Diego introduction profile root-mean-square error [RMSE]: 6.5 percentage points [95% HPD 3.8-9.6 points]; Baja California vs. San Diego introduction profile RMSE: 16.8 points [95% HPD 13.9-19.8 points]; Figure 2C). This difference suggests that phylogenetic similarity can overstate connectivity due to shared introduction sources. Using geographic transitions from the same phylogeographic reconstruction, we found that bidirectional transmission involving San Diego was consistently present over time, but increased during five periods: (1) April–May 2020, (2) November–December 2020, (3) July–September 2021, (4) December 2021–February 2022, and (5) June–July 2022 (Figure 2D). The earliest period was dominated by transmission with Baja California, whereas during later periods transmission across the border was largely replaced by transmission with Los Angeles and other farther domestic locations (Figure 2D). This result agrees with findings from our PhyloSor analysis, suggesting that the frequency of transmission between more distant locations increased during the pandemic.

To examine how bidirectional transmission impacted local COVID-19 incidence, we investigated the temporal association of connectivity and local case numbers. We estimated the relative amount of incidence that could be attributed to connectivity as the percentage of viral lineages circulating in San Diego that could not be traced back to lineages circulating in San Diego at least two weeks earlier in our phylogeographic reconstruction. We found that, on average, half of all lineages in San Diego could be attributed to other locations, but the proportion decreased markedly during the waves of the epidemic relative to the adjacent troughs (Supp. Figure 9). This result is supported by contact tracing data, which indicated that, on average, 15% of all San Diego cases were directly associated with travel within the US or Mexico, including from Los Angeles and Baja California, and that the percentage of cases directly associated with travel was inversely correlated to COVID-19 incidence (Supp. Figure 10; Pearson  $R = -0.23$  [95% CI:  $-0.03$  to  $-0.42$ ];  $P = 0.03$ ).

Combined, these results indicate that connectivity between locations played a prominent role in maintaining local incidence, particularly during periods between epidemic waves, sustaining the COVID-19 pandemic.

### Temporal shifts in mobility impacted SARS-CoV-2 transmission risk

Our observation that connectivity to nearby locations impacted local COVID-19 cases prompted us to investigate factors driving transmission. Investigations of prior outbreaks have shown that a gravity model, which is widely used in economics to predict the flow of trade between different locations based on the population size and proximity of the locations, can be used to estimate transmission<sup>14,15,17</sup>. In the case of transmission, the spread of a virus from one location to another can be predicted based on the size and proximity of their viral populations, as approximated by the number of infections and the human mobility between them, respectively. We questioned if the spread of SARS-CoV-2 could also be explained by this model, and found that mobility was an important driver of SARS-CoV-2 transmission. We determined that changes in mobility corresponded to the increases in connectivity observed in previous analyses, and that mobility patterns in the border city of San Diego mirrored nationwide patterns.

To see if the gravity model explains SARS-CoV-2 spread, we examined whether transmission between locations with high connectivity followed mobility patterns estimated by travel surveys<sup>38</sup> and was correlated with the estimated number of infections at the origin. We reconstructed the transmission of SARS-CoV-2 into subcounty regions of San Diego using a discrete state phylogeographic analysis, and found that SARS-CoV-2 lineages transmitted to San Diego from Baja California were disproportionately more likely to be introduced into the South and Central regions of San Diego, which have large population centers close to the border (Figure 3A–C). This was the case whether transitions from Baja California to San Diego were compared to transitions from all other states into San Diego, or from Los Angeles into San Diego (Supp. Figure 11). This finding aligned with travel surveys conducted within San Diego prior to the pandemic that reported that visitors from Mexico traveled on average less than 30 miles in the US and typically remained within the border regions of San Diego<sup>38</sup>. Additionally, we found that transmission between San Diego and Los Angeles, and between San Diego and Baja California, was correlated with the estimated number of asymptomatic infections in that location that were able to travel from the source location (only infections that are pre-symptomatic or asymptomatic, see Methods;  $R^2 > 0.59$  for all pairs; Figure 3D–G). Our findings that there is a clear correlation between transmission magnitude and infection rate, together with the concordance between transmission across the US-Mexico border and known mobility patterns, indicate that the gravity model accurately describes SARS-CoV-2 transmission.

Using the gravity model, we next assessed how changes in mobility over time impacted SARS-CoV-2 transmission. To examine the relationship between mobility and transmission, we analyzed weekly land and air travel data collected by SafeGraph<sup>41</sup>. We found that neighboring California counties (Riverside, Los Angeles, and Orange), states (Arizona), and countries (Mexico; of which 99% originates from Baja California<sup>38</sup>) consistently dominated mobility into San Diego (Figure 4A). Further, we found that there was a



moderate correlation between the number of travelers arriving from a location and the median phylogenetic similarity of that location to San Diego (Pearson  $r = 0.44$  [95% CI, 0.24 to 0.60];  $P < 0.001$ ; Supp. Figure 12). Consistent with our previous findings that connectivity to more distant locations increased over the course of the pandemic, we found that mobility from more distant locations also increased over time. (Figure 4B–C). Travelers from Riverside County, Los Angeles County, Orange County, Arizona, and Mexico accounted for 75% of travelers into San Diego during early 2020, but less than 50% of travelers from June 2021 onwards (Figure 4C). This latter amount is similar to the proportion of travelers arriving from these locations during 2019, before responses to the COVID-19 pandemic affected mobility (Figure 4C).

We then considered the infection rate at the transmission source location, the other component of the gravity model<sup>42</sup>, by estimating the number of COVID-19 infected travelers arriving into San Diego from each location (“import risk”). We calculated this as the product of the number of travelers arriving in San Diego from each source location as determined by SafeGraph data and the estimated COVID-19 infection rate at the source location (Figure 4D). We found that the relative import risk from neighboring locations was low at the beginning of the pandemic, peaked during the spring of 2020, and decreased steadily from then until the end of available data (74.8% in April 2020 to 36.9% in July 2021; Figure 4E). This trend parallels changes in mobility and transmission estimates (Figure 2D & Figure 4C), supporting our findings that the connectivity and major sources of imports into local epidemics shifted from neighboring locations to more distant domestic locations as a result of the relaxation of COVID-19 mandates, such as stay-at-home orders, curfews, and business closures.

Finally, we assessed whether the trend in connectivity we observed in San Diego, wherein connectivity to more distant locations reduced sharply at the onset of the pandemic but slowly recovered thereafter, occurred nationwide. To investigate changes in connectivity across the US, we examined mobility between all US counties, and found reductions in the frequency and distance of travel at the beginning of the pandemic, comparable to those observed in San Diego. The mean number of travelers between each county in the US and the mean distance traveled decreased markedly in early 2020 and slowly recovered to pre-pandemic levels over the course of the next two years (Figure 5A–B), suggesting that shifts in connectivity were widespread.

To determine whether the reduction and slow recovery in connectivity explained the observed variation in phylogenetic similarity between local epidemics during the COVID-19 pandemic, we determined the correlation between mobility and phylogenetic similarity for each pair of locations included in our PhyloSor analysis. We found a positive correlation between mobility and phylogenetic similarity for 84.4% of all locations, suggesting that changes in connectivity resulted in the observed similarity between viral populations of local epidemics (Figure 5C). Likewise, our prior results indicating that bidirectional transmission explained only a portion of the phylogenetic similarity between locations, is consistent with the presence of the small subset of location pairs lacking a positive correlation (Figure 2B–C). To determine whether the widespread trend in connectivity was the result of COVID-19 mandates implemented by the US government (Figure 1A), we calculated the correlation

between the stringency of COVID-19 mandates and the frequency and distance of travel in the US. We found that trends in mobility were partly explained by the implementation and relaxation of COVID-19 mandates (Pearson  $R^2$  for mean pairwise trips = 0.36 [95% CI 0.18–0.53],  $p < 0.001$ ;  $R^2$  for mean distance traveled = 0.57 [95% CI 0.40–0.71],  $p < 0.001$ ; Figure 5D), suggesting that COVID-19 mandates were effective at reducing transmission.

### **US-Mexico border closure was ineffective in preventing imported cases**

While we found that San Diego is representative of widespread trends in connectivity, the position of the county along the US-Mexico border uniquely enabled us to directly investigate COVID-19 mandates targeting border travel. Our analyses indicate that there was significant transmission between San Diego and nearby locations, including Baja California, during the entire duration of the pandemic. However, the US-Mexico border was closed to non-essential travel from March 19<sup>th</sup>, 2020 to November 8<sup>th</sup>, 2021<sup>28,29</sup>, prompting us to evaluate whether this restriction was effective at preventing cross-border transmission. To do so, we measured how changes in mobility resulting from the border closure impacted import risk into San Diego.

Using mobility data from SafeGraph, we found that the number of northbound travelers across the US-Mexico border was 23.1% less during the partial closure (March 2020 to July 2021) than in 2019 (Supp. Figure 13). We additionally found that import risk across the border was reduced by 22.8% when we compared the observed import risk from Mexico to import risk calculated using mobility estimates from 2019 (Supp. Figure 13). However, given that most travelers into San Diego arrived from locations other than Mexico (Figure 4A), this reduction only amounted to a 3.1% reduction in the total import risk into San Diego, indicating the impact of the non-essential closure was ineffective (Figure 6A). However, we observed a general reduction in the number of travelers visiting San Diego from most locations beginning in March 2020, which were affected by stay-at-home orders and travel hesitancy rather than official travel restrictions<sup>3,24</sup> (Figure 4A). This led us to study whether the reduction in mobility from Mexico to San Diego was more or less impactful than general reductions in mobility.

To investigate this, we compared the observed import risk to San Diego with import risk calculated using mobility estimates from 2019 for each location (Figure 6B). While we found that Mexico had the 3<sup>rd</sup> largest reduction in import risk, behind Los Angeles County and Texas, the reduction was only a small fraction of the total import risk from Mexico (22.8% reduction). Specifically, we found that the relative reduction in import risk from Mexico was less than 80.0% of all other locations and half of the median reduction in import risk (40.1%; Figure 6C). This indicates that the closure of the US-Mexico border to non-essential travel was significantly less impactful at reducing imports than the reductions in travel associated with stay-at-home orders and decreases in mobility resulting from hesitancy<sup>3,24</sup>.

## **Discussion**

In this study, we determined a widespread shift in the connectivity of local COVID-19 epidemics during the pandemic, by integrating genomic surveillance data with

epidemiological and mobility data. Focusing on the first five waves of the pandemic, we found that the implementation of COVID-19 mandates, such as travel restrictions and stay-at-home orders, contained the spread of SARS-CoV-2 locally at the beginning of the pandemic. However, the lifting of mandates enabled the virus to spread further as travel increased. We found that travel-associated infections accounted for half of the incidence during some periods of the pandemic, indicating that local outbreaks were largely affected by epidemics in other locations. By estimating cross-border transmission of SARS-CoV-2, we show that closures to non-essential travel minimally reduced transmission, and were less effective at reducing transmission than non-targeted restrictions.

We focused our genomic surveillance on the epidemic in and around San Diego and the US-Mexico border. The frequency of both domestic and international travel in the region made it an ideal choice to study, and provided a unique opportunity to compare the impacts of domestic COVID-19 mandates and international travel restrictions. While caution should be taken in applying our conclusions to less-populated or less-connected locations, comparisons between San Diego and other North American counties, states, and provinces using both phylogenetic (Figure 1) and mobility data (Figure 4–5) provide independent lines of evidence that the epidemic in San Diego was representative of most locations. Other locations and continents underwent distinct patterns of restriction implementation and relaxation<sup>43–45</sup>, and thus further work can determine whether these patterns differentially impacted connectivity.

Our finding that local COVID-19 epidemics are highly interconnected highlight the importance of collaborative and inclusive public health measures. The beginning of the pandemic led to a substantial reduction in long distance travel, with little to no impact on local connections (Figure 4A). This reduction was also observed in the national mobility networks of France, Italy, and the UK following their respective lockdowns<sup>46</sup>. The local connections of San Diego extended to neighboring epidemics throughout the pandemic, supporting previous evidence that connectivity between adjacent locations precludes virus containment. For instance, the dispersal of SARS-CoV-2 from Wuhan progressed mainly to adjacent cities<sup>47</sup>, and B.1.1.7 was seen to spread from Kent and Greater London to other locations in the UK at a rate proportional to their mobility with Kent and Greater London<sup>48</sup>. San Diego's connectivity to Mexico, provides further evidence that state and international borders did not act as barriers to the spread of the virus, and indicates that both domestic and international collaborations are necessary to control the spread of pathogens.

When we evaluated enacted control measures, we found that the closure of the US-Mexico border was ineffective at reducing cross-border transmission. Preventing transmission requires completely halting travel whereas the border closure only restricted non-essential travel. Correspondingly, even in the month of the pandemic with the fewest travelers, the US Department of Transportation found that more than two million people crossed between Mexico and San Diego<sup>27</sup>. While our data provides only limited examples of any other official travel restrictions, we found that the US-Mexico border closure (22.8% reduction in import risk) was less effective than the total border closure in Jordan<sup>17</sup> (65% reduction in import risk), the mandatory 14-day quarantine enacted in Hong Kong<sup>25</sup> (94% reduction in imported cases), and the ban in Wuhan on all outgoing travel<sup>24</sup> (74% reduction in exported

cases). Additionally, whereas other non-pharmaceutical interventions caused individuals to take shorter, less complex trips and reduce person-to-person contacts, our finding that the border closure did not result in any changes in the destination of travelers crossing the border from Mexico into the US (Figure 3B), suggests that the closure had a limited effect on behavior<sup>3</sup>. As a result, the enacted border closure would have had to be much more stringent to be as effective as other travel restrictions in reducing imports of the virus. However, our finding that outbreaks are increasingly interconnected adds to a growing body of evidence that targeted travel restrictions have limited practical value<sup>5,49,50</sup>.

Our ability to detect cross-border transmission was due to the pooling of resources and collaboration between academic laboratories, public health laboratories, and hospitals on both sides of the border. A recently developed framework for identifying transmission lineages using limited sequencing resources, indicates that only 0.5% of infections need to be sequenced to detect 95% of transmission lineages with a frequency of at least 2% in the population<sup>51</sup>. Current sequencing efforts in Baja California and San Diego, with a sampling fraction of 1% and 10%, respectively, appear to surpass this recommendation. However, the high lag observed between infections in San Diego and transitions to Baja California (Figure 3F), suggests that we detected Baja California transmission lineages later than would be expected given our sampling rate. The difference in sampling rates between San Diego and Baja California, as well as the absence of estimates for the number of southbound travelers crossing the border, prevent any conclusion on the directionality of cross-border transmission<sup>52</sup>. Additionally, other well-traveled border crossings along the US-Mexico border have sequenced much less than the recommended 0.5% of cases, limiting the region's ability to monitor disease spread. Considering these locations are large sources of potential infections, it is critical that regional surveillance capacity is strengthened in these areas.

As the COVID-19 pandemic progresses to endemicity, durable genomic surveillance systems will be critical to provide insights into the continued spread and evolution of SARS-CoV-2. We show that the information produced by epidemiological and genomic surveillance can be integrated with mobility data to quantifiably provide estimates on the sources of introductions and local transmission, and how they change over time. We found that for well-connected locations such as San Diego, transmission resulting from connectivity blurred the division between seemingly separated local epidemics, particularly when cases were low in any of the locations. Consequently, it is a necessity that the international community equitably distribute surveillance infrastructure and enact travel restrictions collaboratively. It is vital that the effects of such restrictions, particularly when they are not equally experienced, be carefully weighed against their quantitative benefit.

### Limitations of the Study

In this study, we assess temporal differences in the connectivity of North American locations using genomic, epidemiological, and mobility data. While our results are robust across data types, each data source comes with its own caveats. First, our analysis used publicly available genomes and epidemiological data which our analyses assume are representative of local cases and collected uniformly within and between locations. Although this assumption would be violated if, for instance, samples were collected primarily as a result of contact

tracing, guidelines set out by the WHO, CDC, and ECDC suggest that in most cases samples were collected without preference for certain groups of individuals. Additionally, mobility data is a reliable estimate of the magnitude of human movement between locations<sup>2,24,46,53</sup>, but does not cover all populations (for example, children under 13, adults without cell phones, etc.) and is insensitive to behavioral changes which impact SARS-CoV-2 transmission risk, including wearing face-masks, handwashing, quarantining, maintaining physical distance, and reducing travel duration<sup>3,54,55</sup>.

## STAR Methods

### Resource availability

**Lead Contact**—Further information and requests for reagents may be directed to the lead contact, Mark Zeller (mzeller@scripps.edu)

**Materials Availability**—This study did not generate new unique reagents, but raw data and code generated as part of this research can be found in the supplemental files, as well as on public resources as specified in the Data and code availability section below. Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

**Code and Data Availability**—Code for all analyses and figure generation, XMLs and log file for BEAST analyses, and configs for simulations are available at: [https://github.com/andersen-lab/project\\_2023\\_SARS-CoV-2\\_Connectivity](https://github.com/andersen-lab/project_2023_SARS-CoV-2_Connectivity). All genomes used in this analysis can be downloaded from GISAID. Sequencing data, including consensus sequences and raw data, is available on NCBI under the BioProject accession ID PRJNA612578. Raw sequencing data is also available on our [Google Cloud](#).

### Experimental model and subject details

**Ethical Statement**—Sample collection, RNA extraction, and viral sequencing was evaluated by the Institutional Review Board (IRB) at Scripps Health (IRB-21-7739). All samples were de-identified before receipt by the study investigators. Aggregated contact tracing data was publicly available prior to the initiation of the study.

### Method details

**SARS-CoV-2 Amplicon Sequencing**—SARS-CoV-2 RNA samples were collected from routine diagnostic tests performed by SEARCH in San Diego, and by Salud Digna, Centro de Diagnóstico COVID-19, Institute of Epidemiological Diagnosis and Reference, Genómica Lab Molecular, and Infectolab in Baja California, Mexico. SARS-CoV-2 was sequenced using PrimalSeq-Nextera XT. This protocol is based on the ARTIC PrimalSeq protocol<sup>56</sup>, except that amplicon sizes were reduced to enable 2x150 read length requirements. The ARTIC network nCoV-2019 V4 primer scheme uses two multiplexed primer pools to create overlapping 250 bp amplicon fragments in two PCR reactions. Full details of the protocol can be found here: [Protocol for HCoV-19 sequencing: PrimalSeq-LibAmp](#). Briefly, SARS-CoV-2 RNA (2 mL) was reverse transcribed with LunaScript RT (New England Biolabs). The virus cDNA was amplified in two multiplexed PCR reactions

(one reaction per primer pool; custom primer scheme can be found here: [Primers for SARS-CoV-2 PrimalSeq-LibAmp](#)) using Q5 DNA High-fidelity Polymerase (New England Biolabs). Following an AMPureXP bead (Beckman Coulter) purification of the combined PCR products, sequencing adaptors containing sample specific indexes were added using a step-out PCR reaction using Q5 DNA High-fidelity Polymerase. The libraries were purified with AMPureXP beads and quantified using the Qubit High Sensitivity DNA assay kit (Invitrogen) and TapeStation D5000 tape (Agilent). The individual libraries were normalized and pooled in equimolar amounts at 1.5 nM. The 2 nM library pool was sequenced on an Illumina NovaSeq 6000 (300 cycles kit). Consensus sequences were deposited on GISAID and Raw reads were deposited under BioProject accession ID PRJNA612578.

**SARS-CoV-2 Genomic Data.**—We queried the GISAID EpiCoV database for all SARS-CoV-2 genomes collected up to January 4<sup>th</sup>, 2023<sup>13</sup>. We removed genomes that (1) were less than 20,000 nucleotides in length, (2) had greater than 12.5% ambiguous nucleotides, (3) had an incomplete or incorrect year-month-day sampling date reported, (4) had a sampling country that could not be interpreted, (5) were not collected from a human infection, (6) had less than 50% agreement with Hu-1 (GenBank Accession ID: NC\_045512.2), or (7) had greater than 500 discrete indels. The final dataset contained 13,722,590 genomes.

**PhyloSor Analysis.**—We used the global SARS-CoV-2 phylogeny provided by GISAID as of January 5<sup>th</sup>, 2023 (also called Audacity). The phylogeny contains all SARS-CoV-2 genomes available on GISAID that were marked as both ‘complete’ and ‘high-coverage’ by GISAID, were longer than 28,000 nucleotides, contained less than 1,000 ambiguous nucleotides, were not identified as being on a long branch, were not manually identified as questionable, and were included in the genomic dataset described previously. For each location pair, the phylogeny was pruned to taxa present in the genomic dataset that were collected from either location. Using the pruned tree, for each month in the period of January 2020 to December 2022, the PhyloSor metric was calculated using only sequences collected in that month<sup>31</sup>. Briefly, the PhyloSor metric is calculated as the ratio of branch lengths (in units of per-site substitution rate) that are shared by two sets of tips ( $BL_{Both}$  compared to the total branch length that is unique to each set of tips (Supp. Figure 2).

$$PhyloSor = \frac{BL_{Both}}{(BL_A + BL_B) * 0.5}$$

Where  $BL_A$  and  $BL_B$  indicate the total branch lengths of either the first set (A) or the second set (B).

To limit the impact of low sampling, we only compared locations that sampled at least 1000 total sequences and collected a sequence in at least 75% of the epidemiological weeks between March 2020 and December 2022. Additionally, within these comparisons we only considered months where at least 30 sequences were included from each location. Here location refers to counties within California, and states in the rest of Canada, Mexico, and the US.

To assess differences in PhyloSor similarity resulting from unequal sampling fractions, we compared San Diego's similarity to all suitably sampled locations under two different subsampling schemes. In the first, a constant number of San Diego sequences were sampled for each month in the analysis, equal to the number of sequences available for San Diego from the month with the least number of sequences greater than 30. This number was 149. In the second scheme, a number of San Diego sequences were sampled such that they represented a constant fraction of cases. 2.5% was selected as it was the 10<sup>th</sup> percentile of the sampling fraction of all months. Ten replicates of each subsampling scheme were performed, and the median PhyloSor similarity of San Diego to all other locations was compared between the subsampling schemes and the analysis performed with the non-downsampled dataset.

**PhyloSor Validation.**—To validate the use of PhyloSor in measuring the temporal connectivity between locations, we conducted epidemic simulations using FAVITES V1.1.35<sup>57</sup>. First, we generated static contact networks in FAVITES using a modified Barabási-Albert algorithm<sup>58</sup>. We generated two separate 20,000 member communities using the Barabási-Albert algorithm with a mean value of 8 contacts per day. For each community, we calculated intra-community connectivity as the fraction of all possible contacts that were made. Inter-community edges were sampled by randomly deciding for each pair of nodes in different communities if they should be connected by an edge or not. The probability of connecting two nodes in two communities was calculated as a fraction of the average intra-community connectivity. We called this term inter-community connectivity. Ten contact networks were generated using inter-community connectivity values between 0.5 and 0.001 were simulated.

We then simulated a transmission network over each contact network using a Susceptible-Infected-Recovered model. The simulation sampled a single viral lineage from each infected individual at a random point during their infectious period to represent viral genome sequencing, and a virus phylogeny in units of time (years) was constructed under a coalescent model using the VirusTreeSimulator package embedded in FAVITES. Based on Tonkin-Hill et al., we assumed that a constant coalescent model was representative of the within-host effective virus population size during the infectious period<sup>59</sup>. All parameters for the transmission network simulation and viral lineage sample were identical to parameters used in Worobey et al. 2020<sup>7</sup>. Ultimately, the virus phylogeny in units of time was converted to units of per-site mutation rate by multiplying the branch length by a constant  $1.1 \times 10^{-3}$  subs/site/year, consistent with Duchene et al. 2020<sup>60</sup>. PhyloSor similarity between the two communities for the first month of the simulation was calculated using the phylogeny. We detected a strong correlation between PhyloSor similarity and inter-community connectivity (Pearson  $R = 0.89$  [95% CI: 0.82-0.93];  $P < 0.001$ ).

**Network Analysis.**—For each month in the period of January 2020–November 2022, we considered a complete weighted undirected graph, where nodes are locations in North America and edges weights are the PhyloSor similarity between locations. For this analysis, location refers to the county-level in the US, and state-level in Canada and Mexico. However, where counties did not meet the inclusion criteria (greater than 1000 sequences

and at least one sequence in 75% of the epidemiological weeks between March 2020 and December 2022), their sequences were assigned to the state-level. We calculated the average pairwise similarity between all locations as the global efficiency of the graph, which takes into consideration the multiple pathways between locations in the graph. Global efficiency is a network measure that describes how easily information is exchanged over the network and can be defined as the average shortest path length between each pair of nodes in the network<sup>61</sup>. Low global efficiency indicates a network has few strong connections, while high efficiency indicates that most locations are strongly connected. Given a weighted network  $G$  with  $n$  nodes, global efficiency can be calculated as:

$$E_{glob} = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

where  $d_{ij}$  is the distance of the shortest path length between nodes  $i$  and  $j$ . The shortest path length is the smallest sum of weights throughout all the possible paths in the network from  $i$  to  $j$ . In our case, because our edge weights represent a similarity metric rather than a distance metric, we used the reciprocal of edge weights to calculate the shortest path length between nodes.

We calculated heterogeneity in each nodes' contribution to global efficiency using the Gini index<sup>62</sup>. The contribution of each node to global efficiency, also called nodal efficiency, is the average PhyloSor similarity between it and all other nodes. In our network, locations with high nodal efficiency are phylogenetically similar to a greater portion of North American locations. It can be calculated as:

$$e_i = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}}$$

To summarize across the networks of all months, the nodal efficiencies of all locations for a given month were min-max normalized and the median normalized nodal efficiencies were reported.

**Stringency of US Response to COVID-19.**—To summarize the strictness of the US government's response to COVID-19, we used the Stringency Index as calculated by the Oxford Coronavirus Government Response Tracker<sup>32</sup>. Briefly, the index is a composite metric which considers school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, and international travel controls (see citation for full calculation details). The metric was calculated daily for the US and returns a value between 0 and 100; a higher score indicating a stricter response. We additionally calculated the mean stringency index for each month in the period of January 2020 to November 2022.

**Genomic Dataset Generation.**—The massive amount of sequencing data produced during the COVID-19 pandemic prevented us from including all data in our phylogenetic



analyses. In order to limit the computational burden of the phylogeographic analysis, we subsampled 2500 genomes from our SARS-CoV-2 genomic dataset. To focus the analysis on the region around San Diego County, we allocated 500 genomes each to San Diego County, Los Angeles County, and Baja California.

The remaining 1000 genomes were allocated to all other locations proportionally to their distance to San Diego and the total number of flights connecting the location and San Diego in 2019. Here, location refers to a state (or first administration level) in the US and Mexico, and country everywhere else. Geographic distance was calculated as the centroid-centroid distance to San Diego County, rescaled to have unit scale, and inverted, so that nearby locations had the greatest value. Total number of flights into San Diego was obtained from the OpenSky Network<sup>63</sup>, and also rescaled to have unit scale (for more details see following methods section Travel and Mobility Data). The sum of these two values proportional to all other locations was the proportion of the 1000 contextual genomes allocated to that location. In order to sample virus diversity in each location equally, sequences were randomly sampled proportional to the location-specific incidence data binned by epidemiological week.

To estimate a root with a reasonable date and location state in our phylogenetic inference, we also included the 50 earliest SARS-CoV-2 genomes in our dataset. To accurately infer the timing and geographic state of the lineages responsible for widespread epidemiological waves, we included the 10 earliest sequences assigned to Alpha, Delta, BA.1 (Omicron) and BA.2 (Omicron). Lastly, to assess the accuracy of the timing of the basal structure of our phylogeny, we included genomes from three outbreaks with well-described introductions<sup>7,14,64</sup>. A list of all included sequences, their GISAID accession IDs, and the compartment they filled is shown in Supplemental Table 1.

**Phylogenetic Analysis.**—We aligned the sequence dataset to reference genome Hu-1 (GISAID ID: EPI\_ISL\_402125) using minimap2 v2.17 and [gofasta v0.0.6](#)([virus-evolution/gofasta](#))<sup>65</sup>. We masked the 3' and 5' UTRs as well as sites that may confound phylogenetic inference of SARS-CoV-2 genomes<sup>66</sup>. We constructed a maximum likelihood phylogenetic tree for the dataset using IQ-TREE2 and an HKY substitution model<sup>67,68</sup>. We rooted the resulting phylogeny on Hu-1 and time-resolved it using TreeTime v0.7.4 with a strict clock rate of 0.00091 substitutions/site/year, pruning taxa that were more than three interquartile ranges from the clock-rate regression<sup>69</sup>. Lastly, we randomly resolved polytomies in the tree by adding 0 length branches with [gotree](#)<sup>70</sup>.

We reconstructed the time-resolved phylogeny using BEAST v1.10.5<sup>71</sup>. We used the HKY substitution model with gamma distributed rate variation among all sites. We fixed the clock rate at  $9.1 \times 10^{-4}$  substitutions/site/year and used an exponential growth coalescent tree prior. We also fixed the root of the tree on November 20th, 2019<sup>72</sup>. We combined two independent MCMC chains of 200 million states ran with the BEAGLE computational library<sup>71</sup>. Parameters and trees were sampled every 10,000 and 100,000 steps, respectively, with 20-60% of steps discarded as burn-in (depending on the chain). Convergence and mixing of the MCMC chains were assessed with Tracer v.1.7.2, and all estimated parameters were determined<sup>73</sup> to have effective sample sizes of greater than 100.

**Phylogeographic Reconstruction.**—We performed two discrete state ancestral reconstructions on geographic states using BEAST. This analysis reconstructed location-transition history across an empirical distribution of 2000 time-calibrated trees sampled from the posterior tree distribution estimated above. In the first analysis the discrete states used were (1) San Diego County, (2) Los Angeles County, (3) USA (not including either California county), (4) Baja California, (5) Mexico (not including Baja California), and (6) a final state corresponding to all remaining locations. The second analysis assigned San Diego County taxa into the County of San Diego Health and Human Services (HHSA) region they were collected in based on the ZIP code they were collected from. The ZIP code to HHSA region table we used was retrieved from HHSA website ([https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/sd\\_community\\_action\\_partnership/26%20HHSA%20sdcnty\\_zipcode.pdf](https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/sd_community_action_partnership/26%20HHSA%20sdcnty_zipcode.pdf)). We assumed that geographic transitions rates were reversible and used a symmetric substitution model for both analyses. We used Bayesian stochastic search variable selection to infer non-zero migration rates<sup>37</sup>. We used the TreeMarkovJumpHistoryAnalyzer from the pre-release version of BEAST v1.10.5 to obtain the Markov jump estimates and their timings from the posterior tree distribution, and assumed that they are a suitable proxy for the transmission between two locations<sup>37,74,75</sup>. We used TreeAnnotator v1.10 to construct a maximum clade credibility (MCC) tree which we visualized with baltic (<https://github.com/evogytis/baltic>). We examined the sensitivity of our results to whether we assumed symmetric or asymmetric transition rates and found that our conclusion regarding the proportion of Markov jumps between San Diego and the other discrete states was robust between the two discrete state models (Supp. Figure 14).

**Persistence Analysis.**—We used the PersistenceAnalyzer from the pre-release version of BEAST v1.10.5 to summarize the relative contribution of independent introductions on local circulating lineages in San Diego across the posterior tree distribution labeled with Markov jumps. Briefly, for each two week period represented in the phylogeny, we identified the number of lineages circulating in San Diego at the end of the period and determined whether they resulted from a lineage that was estimated to be circulating in San Diego at the beginning of the period or from a unique introduction during the period. Persistent lineages are lineages that could be traced back to locally circulating lineages.

**Contact Tracing Data.**—Contact tracing data was obtained from the Epidemiology and Immunization Service Branch of the County of San Diego. Up until March 2022, contact tracers in San Diego interviewed between 40-60% of all confirmed cases in the county and asked, among other questions, whether there was travel within the US (excluding California), Mexico, or internationally during the 2-16 days prior to onset of symptoms (or positive test date if asymptomatic). Uncertainty in the proportion of interviewed cases that were travel-related was assessed by bootstrapping interviews for each week 100 times.

**Travel and Mobility Data.**—We followed Zeller et al.<sup>14</sup> in calculating travel into San Diego County, using the weekly patterns data from SafeGraph, a data company that aggregates anonymized data from numerous applications to provide insights about physical places, via the Placekey Community (See Zeller et al. citation for full calculation details).

SafeGraph estimates human movements using cell phone tracking, which has been shown to capture both land and air travel at a variety of distance scales<sup>2,76</sup>. Briefly, we estimated the true number of travelers for a given week ( $w$ ) between a source and destination location ( $S$  and  $D$ ;  $travelers_{w,S<D}$ ) using the raw number of devices that traveled from the source to the destination location ( $devices_{w,S>D}$ ), the total number of devices detected at the destination ( $totalDevices_{w,D}$ ) and the total population of the destination location ( $population_D$ ), according to:

$$travelers_{w,S>D} = \frac{devices_{w,S>D}}{totalDevices_{w,D}} population_D$$

We note that because of the EU General Data Protection Regulation, SafeGraph was not able to provide mobility data from countries within the European Economic Area (see <https://www.safegraph.com/privacy-policy>). Therefore, EU countries are excluded from our mobility analyses. However, the San Diego Tourism Authority reported that international travelers (excluding those from Canada and Mexico) accounted for 2.9% of the visitors to San Diego in 2019<sup>33</sup>, suggesting that the impact of this exclusion is slight.

We also note that SafeGraph does not provide mobility data finer than the country-level for international locations, particularly Mexico. However, independent travel surveys indicate that it is reasonable to assume that 99% of travel into San Diego from Mexico originated in Baja California<sup>38</sup>.

We noticed that travel from international locations from 2020 onwards increased uniformly relative to data from 2019. This was not consistent with independent sources of mobility. For example, we observed no increase in monthly inbound crossings at the US-Mexico border into San Diego collected by the Department of Transportation (<https://explore.dot.gov/#/views/BorderCrossingData/Monthly>). To correct this artifact, we normalized mobility data from January 1<sup>st</sup>, 2020 onward by multiplying it by the ratio of the mean mobility between January 1<sup>st</sup> and March 1<sup>st</sup> in 2019 compared to 2020. March 1<sup>st</sup> was chosen because it was generally before any reductions in mobility occurred as a result of the spread of SARS-CoV-2 in the US. The impact of this correction should be slight as our conclusions rely on the relative, rather than absolute mobility, into San Diego.

We calculated the mean number of travelers traveling between each county in the US using a network where nodes were US counties and edges weighted by the estimated number of travelers between them. We constructed a network for each week between January 2019 and July 2021. The mean number of travelers was equivalent to the global efficiency of this network, using the inverse of travelers as the distance term.

We also obtained weekly air travel flight data into San Diego International Airport (KSAN) from the OpenSky Network<sup>63</sup>. We filtered data for flights with complete origin and destination airport ICAO codes. ICAO codes were matched to country, state, and county location using an airport database (<https://github.com/mwgg/Airports>). Using the dataset, we counted the total flight counts from each US state and non-US country in 2019 to use as an input for the genomic database generation.

**Epidemiological Data and Estimated import Risk.**—We calculate import risk according to du Plessis et al.<sup>15</sup>, with some modifications. Briefly, we estimated the number of infected travelers arriving each day into San Diego from each source location as the product of the number of asymptotically infectious individuals in each source location on that day and the number of travelers arriving in San Diego from the source location as estimated from the SafeGraph mobility data. SafeGraph mobility data utilizes cell phone tracking data so both air and land travel are included. Like du Plessis et al., we conservatively estimate that only asymptomatic infections contributed to importation risk, as symptomatic infections would not travel. Therefore, the asymptomatic infection rate is derived only from the number of pre-symptomatic and asymptomatic infections. We estimated the asymptomatic infectious rate for each location by back-extrapolating the death time series assuming the same estimates for the latent and incubation period, infectious duration, symptom-onset-to-death, asymptomatic proportion, and infection fatality rate as du Plessis et al.<sup>15</sup> (See citation for full details). To back-calculate infections from deaths, we specifically assumed the infection fatality rate of COVID-19 was 1%, which is consistent with independent studies in China, France and aboard the Diamond Princess during the first year of the pandemic<sup>77–79</sup>. Critically, we assumed that the infection fatality rate remained constant over our study period, even though it likely varied between locations and across time<sup>80</sup>. For instance, death ascertainment rates were known to be lower in Baja California<sup>81</sup> than in California<sup>82</sup>. Consequently, we primarily focused on the temporal dynamics of infections, and limited our analysis of absolute infection numbers to the period of the pandemic prior to widespread vaccine use, as the impact of vaccinations on the infectious fatality rate is most significant. Specifically, we used May 5<sup>th</sup>, 2021 as the cutoff point, as it marked the time when at least 50% of San Diego’s population had received at least one dose of a SARS-CoV-2 vaccine<sup>39,40</sup>.

Because travel surveys indicated that 99% of all Mexican travelers visiting San Diego originated in Baja California, we used Baja California’s asymptotically infectious individuals in place of Mexico’s for estimating import risk<sup>38</sup>. We obtained the time series of reported deaths from each California county, US state, and county from the outbreak.info R package which provides data from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University<sup>83, 84</sup>. We additionally obtained the time series of reported deaths from each Mexican state directly from the Mexican Department of Health (<https://datos.covid-19.conacyt.mx/>), as they were more complete than other sources.

Counterfactual import risk was calculated as above, except that for dates from March 1<sup>st</sup>, 2020 onward the number of travelers arriving in San Diego on a given day were replaced with the number of travelers arriving in San Diego for that day in 2019.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Authors

Nathaniel L. Matteson<sup>1,43,\*</sup>, Gabriel W. Hassler<sup>2,43</sup>, Ezra Kurzban<sup>1,44</sup>, Madison A. Schwab<sup>1,44</sup>, Sarah A. Perkins<sup>1,44</sup>, Karthik Gangavarapu<sup>3,1</sup>, Joshua I. Levy<sup>1</sup>, Edyth Parker<sup>1</sup>, David Pride<sup>4,5</sup>, Abbas Hakim<sup>6,7,8</sup>, Peter De Hoff<sup>6,7,8</sup>, Willi Cheung<sup>6,7,8</sup>, Anelizze Castro-Martinez<sup>6,7,9</sup>, Andrea Rivera<sup>6</sup>, Anthony Veder<sup>6</sup>, Ariana Rivera<sup>6</sup>, Cassandra Wauer<sup>6</sup>, Jacqueline Holmes<sup>6</sup>, Jedediah Wilson<sup>6</sup>, Shayla N. Ngo<sup>6</sup>, Ashley Plascencia<sup>6</sup>, Elijah S. Lawrence<sup>6</sup>, Elizabeth W. Smoot<sup>6</sup>, Emily R. Eisner<sup>6</sup>, Rebecca Tsai<sup>6</sup>, Marisol Chacón<sup>6</sup>, Nathan A. Baer<sup>6</sup>, Phoebe Seaver<sup>6</sup>, Rodolfo A. Salido<sup>6</sup>, Stefan Aigner<sup>6</sup>, Toan T. Ngo<sup>6</sup>, Tom Barber<sup>6</sup>, Tyler Ostrander<sup>6</sup>, Rebecca Fielding-Miller<sup>10,11</sup>, Elizabeth H. Simmons<sup>12</sup>, Oscar E. Zazueta<sup>13</sup>, Idanya Serafin-Higuera<sup>14</sup>, Manuel Sanchez-Alavez<sup>14,15</sup>, Jose L. Moreno-Camacho<sup>16</sup>, Abraham García-Gil<sup>16</sup>, Ashleigh R. Murphy Schafer<sup>17</sup>, Eric McDonald<sup>17</sup>, Jeremy Corrigan<sup>17</sup>, John D. Malone<sup>17</sup>, Sarah Stous<sup>17</sup>, Seema Shah<sup>17</sup>, Niema Moshiri<sup>18</sup>, Alana Weiss<sup>1</sup>, Catelyn Anderson<sup>1</sup>, Christine M. Aceves<sup>1</sup>, Emily G. Spencer<sup>1</sup>, Emory C. Hufbauer<sup>1</sup>, Justin J. Lee<sup>1</sup>, Alison J. King<sup>1</sup>, Karthik S. Ramesh<sup>1</sup>, Kelly N. Nguyen<sup>1</sup>, Kieran Saucedo<sup>1</sup>, Refugio Robles-Sikisaka<sup>1</sup>, Kathleen M. Fisch<sup>7,19</sup>, Steven L. Gonias<sup>4</sup>, Amanda Birmingham<sup>20</sup>, Daniel McDonald<sup>20</sup>, Smruthi Karthikeyan<sup>20</sup>, Natasha K. Martin<sup>11</sup>, Robert T. Schooley<sup>11</sup>, Agustin J. Negrete<sup>21</sup>, Horacio J. Reyna<sup>21</sup>, Jose R. Chavez<sup>21</sup>, Maria L. Garcia<sup>21</sup>, Jose M. Cornejo-Bravo<sup>22</sup>, David Becker<sup>23</sup>, Magnus Isaksson<sup>23</sup>, Nicole L. Washington<sup>23</sup>, William Lee<sup>23</sup>, Richard S. Garfein<sup>10</sup>, Marco A. Luna-Ruiz Esparza<sup>24</sup>, Jonathan Alcántar-Fernández<sup>24</sup>, Benjamin Henson<sup>25</sup>, Kristen Jepsen<sup>25</sup>, Beatriz Olivares-Flores<sup>26</sup>, Gisela Barrera-Badillo<sup>26</sup>, Irma Lopez-Martínez<sup>26</sup>, José E. Ramírez-González<sup>26</sup>, Rita Flores-León<sup>26</sup>, Stephen F. Kingsmore<sup>27</sup>, Alison Sanders<sup>28</sup>, Allorah Pradenas<sup>28</sup>, Benjamin White<sup>28</sup>, Gary Matthews<sup>28</sup>, Matt Hale<sup>28</sup>, Ronald W. McLawhon<sup>28</sup>, Sharon L. Reed<sup>28</sup>, Terri Winbush<sup>28</sup>, Ian H. McHardy<sup>29</sup>, Russel A. Fielding<sup>29</sup>, Laura Nicholson<sup>29</sup>, Michael M. Quigley<sup>29</sup>, Aaron Harding<sup>30</sup>, Art Mendoza<sup>30</sup>, Omid Bakhtar<sup>30</sup>, Sara H. Browne<sup>31,11</sup>, Jocelyn Olivas Flores<sup>32,22</sup>, Diana G. Rincon Rodríguez<sup>32,33</sup>, Martin Gonzalez Ibarra<sup>32,33</sup>, Luis C. Robles Ibarra<sup>32,34</sup>, Betsy J. Arellano Vera<sup>32,35</sup>, Jonathan Gonzalez Garcia<sup>32,36</sup>, Alicia Harvey-Vera<sup>37</sup>, Rob Knight<sup>20,18,38</sup>, Louise C. Laurent<sup>6,7,9</sup>, Gene W. Yeo<sup>6,9,39</sup>, Joel O. Wertheim<sup>5</sup>, Xiang Ji<sup>40</sup>, Michael Worobey<sup>41</sup>, Marc A. Suchard<sup>42,45</sup>, Kristian G. Andersen<sup>1,45,\*f</sup>, Abraham Campos-Romero<sup>24,45</sup>, Shirlee Wohl<sup>1,45</sup>, Mark Zeller<sup>1,45,\*</sup>

## Affiliations

<sup>1</sup>Department of Immunology and Microbiology, Scripps Research, La Jolla, CA, USA

<sup>2</sup>Department of Computational Medicine, University of California, Los Angeles, CA, USA

<sup>3</sup>Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA

<sup>4</sup>Department of Pathology, University of California San Diego, La Jolla, CA, USA

<sup>5</sup>Department of Medicine, University of California San Diego, La Jolla, CA, USA

<sup>6</sup>Expedited COVID Identification Environment (EXCITE) Laboratory, Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

<sup>7</sup>Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California San Diego, La Jolla, CA, USA

<sup>8</sup>COVID-19 Detection, Investigation, Surveillance, Clinical, and Outbreak Response, California Department of Public Health, Richmond, CA, USA

<sup>9</sup>Sanford Consortium of Regenerative Medicine, University of California San Diego, La Jolla, CA, USA

<sup>10</sup>Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, USA

<sup>11</sup>Division of Infectious Disease and Global Public Health, University of California San Diego, La Jolla, CA, USA

<sup>12</sup>Academic Affairs, University of California San Diego, La Jolla, CA, USA

<sup>13</sup>Department of Epidemiology. Secretaria de Salud de Baja California, Tijuana, B.C., Mexico

<sup>14</sup>Centro de Diagnostico COVID-19 UABC Tijuana, B.C., Mexico

<sup>15</sup>Department of Molecular Medicine, Scripps Research, La Jolla, CA, USA

<sup>16</sup>Clinical Laboratory Department, Salud Digna, A.C, Tijuana, B.C., Mexico

<sup>17</sup>County of San Diego Health and Human Services Agency, San Diego, CA, USA

<sup>18</sup>Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA

<sup>19</sup>Center for Computational Biology and Bioinformatics, University of California San Diego, La Jolla, CA, USA

<sup>20</sup>Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

<sup>21</sup>Facultad de Ciencias de la Salud Universidad Autonoma de Baja California Valle de Las Palmas, Tijuana, B.C., Mexico

<sup>22</sup>Facultad de Ciencias Quimicas e Ingenieria, Universidad Autonoma de Baja California, Mexico

<sup>23</sup>Helix, San Mateo, CA, USA

<sup>24</sup>Innovation and Research Department, Salud Digna, A.C, Tijuana, B.C., Mexico

<sup>25</sup>Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

<sup>26</sup>Instituto de Diagnóstico y Referencia Epidemiológicos (InDRE), Ciudad de México, CDMX, Mexico

<sup>27</sup>Rady Children's Institute for Genomic Medicine, San Diego, CA, USA

<sup>28</sup>Return to Learn, University of California San Diego, La Jolla, CA, USA

- <sup>29</sup>Scripps Health, San Diego, La Jolla, CA, USA
- <sup>30</sup>Sharp Healthcare, San Diego, CA, USA
- <sup>31</sup>Specialist in Global Health, Encinitas, CA, USA
- <sup>32</sup>UniHealthMx, Tijuana, B.C., Mexico
- <sup>33</sup>Facultad de Medicina, Universidad Xochicalco, Tijuana, B.C., Mexico
- <sup>34</sup>Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado, Tijuana, B.C., Mexico
- <sup>35</sup>Instituto Mexicano del Seguro Social, Tijuana, B.C., Mexico
- <sup>36</sup>SIMNSA, Tijuana, B.C., Mexico
- <sup>37</sup>University of California San Diego, La Jolla, CA, USA
- <sup>38</sup>Department of Bioengineering, University of California San Diego, La Jolla, CA, USA
- <sup>39</sup>Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA
- <sup>40</sup>Department of Mathematics, School of Science and Engineering, Tulane University, New Orleans, LA, USA
- <sup>41</sup>Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA
- <sup>42</sup>Department of Human Genetics, University of California, Los Angeles, CA, USA
- <sup>43</sup>These authors contributed equally
- <sup>44</sup>These authors contributed equally
- <sup>45</sup>Senior author

## Acknowledgements

We thank the administrators of the GISAID database for supporting rapid and transparent sharing of genomic data during the COVID-19 pandemic and all our colleagues sharing data on GISAID. The research leading to these results has received funding from the National Institutes of Health (grants U19AI135995, U01AI151812, UL1TR002550, F32AI154824, and R01AI153044) and the CDC contract 7530122C14843. We also gratefully acknowledge support from NVIDIA Corporation and Advanced Micro Devices, Inc., with the donation of parallel computing resources used for this research.

## Declaration of Interests

KGA has received consulting fees and compensated expert testimony on SARS-CoV-2 and the COVID-19 pandemic.

## References

1. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, Tam C, and Dickens BL (2020). Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect. Dis* 20, 678–688. [PubMed: 32213332]

2. Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, and Leskovec J (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 82–87. [PubMed: 33171481]
3. Lucchini L, Centellegher S, Pappalardo L, Gallotti R, Privitera F, Lepri B, and De Nadai M (2021). Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures. *Sci. Rep* 11, 24452. [PubMed: 34961773]
4. Hill V, Plessis D, Peacock TP, Aggarwal D, Colquhoun R, Carabelli AM, Ellaby N, Gallagher E, Groves N, Jackson B, et al. (2022). The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evol* 8, veac080. [PubMed: 36533153]
5. McCrone JT, Hill V, Bajaj S, Pena RE, Lambert BC, Inward R, Bhatt S, Volz E, Ruis C, Dellicour S, et al. (2022). Context-specific emergence and growth of the SARS-CoV-2 Delta variant. *Nature* 610, 154–160. [PubMed: 35952712]
6. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, Anyaneji UJ, Bester PA, Boni MF, Chand M, et al. (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 603, 679–686. [PubMed: 35042229]
7. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA, Wertheim JO, and Lemey P (2020). The emergence of SARS-CoV-2 in Europe and North America. *Science* 370, 564–570. [PubMed: 32912998]
8. Craft ME (2015). Infectious disease transmission and contact networks in wildlife and livestock. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 370. 10.1098/rstb.2014.0107.
9. Specht I, Sani K, Loftness BC, Hoffman C, Gionet G, Bronson A, Marshall J, Decker C, Bailey L, Siyanbade T, et al. (2022). Analyzing the impact of a real-life outbreak simulator on pandemic mitigation: An epidemiological modeling study. *PATTER* 3. 10.1016/j.patter.2022.100572.
10. Firth JA, Hellewell J, Klepac P, Kissler S, CMMID COVID-19 Working Group, Kucharski AJ, and Spurgin LG (2020). Using a real-world network to model localized COVID-19 control strategies. *Nat. Med* 26, 1616–1622. [PubMed: 32770169]
11. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, Sun J, François S, Kraemer MUG, Faria NR, et al. (2020). Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell* 181, 997–1003.e9. [PubMed: 32359424]
12. Ypma RJF, van Ballegooijen WM, and Wallinga J (2013). Relating phylogenetic trees to transmission trees of infectious disease outbreaks. *Genetics* 195, 1055–1062. [PubMed: 24037268]
13. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, Akite N, Ho J, Lee RT, Yeo W, et al. (2021). GISAIID's Role in Pandemic Response. *China CDC Wkly* 3, 1049–1051. [PubMed: 34934514]
14. Zeller M, Gangavarapu K, Anderson C, Smither AR, Vanchiere JA, Rose R, Snyder DJ, Dudas G, Watts A, Matteson NL, et al. (2021). Emergence of an early SARS-CoV-2 epidemic in the United States. *Cell* 184, 4939–4952.e15. [PubMed: 34508652]
15. du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, Raghwan J, Ashworth J, Colquhoun R, Connor TR, et al. (2021). Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* 371, 708–712. [PubMed: 33419936]
16. McLaughlin A, Montoya V, Miller RL, Mordecai GJ, Worobey M, Poon AFY, Joy JB, and Canadian COVID-19 Genomics Network (CanCOGen) Consortium (2022). Genomic epidemiology of the first two waves of SARS-CoV-2 in Canada. *eLife* 11. 10.7554/elife.73896.
17. Parker E, Anderson C, Zeller M, Tibi A, Havens JL, Laroche G, Benlarbi M, Ariana A, Robles-Sikisaka R, Latif AA, et al. Regional connectivity drove bidirectional transmission of SARS-CoV-2 in the Middle East during travel restrictions. 10.1101/2022.01.27.22269922.
18. Hodcroft EB, Zuber M, Nadeau S, Vaughan TG, Crawford KHD, Althaus CL, Reichmuth ML, Bowen JE, Walls AC, Corti D, et al. (2021). Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* 595, 707–712. [PubMed: 34098568]
19. Courtemanche C, Garuccio J, Le A, Pinkston J, and Yelowitz A (2020). Strong social distancing measures in the United States reduced the COVID-19 growth rate. *Health Aff.* 39, 1237–1246.
20. Kishore N, Kahn R, Martinez PP, De Salazar PM, Mahmud AS, and Buckee CO (2021). Lockdowns result in changes in human mobility which may impact the epidemiologic dynamics of SARS-CoV-2. *Sci. Rep* 11, 6995. [PubMed: 33772076]



21. Han X, Xu Y, Fan L, Huang Y, Xu M, and Gao S (2021). Quantifying COVID-19 importation risk in a dynamic network of domestic cities and international countries. *Proc. Natl. Acad. Sci. U. S. A* 118. 10.1073/pnas.2100201118.
22. Updated WHO recommendations for international traffic in relation to COVID-19 outbreak <https://www.who.int/news-room/articles-detail/updated-who-recommendations-for-international-traffic-in-relation-to-covid-19-outbreak>.
23. Douglas J, Mendes FK, Bouckaert R, Xie D, Jiménez-Silva CL, Swanepoel C, de Ligt J, Ren X, Storey M, Hadfield J, et al. Phylodynamics reveals the role of human travel and contact tracing in controlling the first wave of COVID-19 in four island nations. 10.1101/2020.08.04.20168518.
24. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Pastore Y Piontti A, Mu K, Rossi L, Sun K, et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368, 395–400. [PubMed: 32144116]
25. Zhang N, Jia W, Wang P, Dung C-H, Zhao P, Leung K, Su B, Cheng R, and Li Y (2021). Changes in local travel behaviour before and during the COVID-19 pandemic in Hong Kong. *Cities* 112, 103139. [PubMed: 33589850]
26. Bart SM, Smith TC, Guagliardo SAJ, Walker AT, Rome BH, Li SL, Aichele TWS, Stein R, Ernst ET, Morfino RC, et al. (2023). Effect of Predeparture Testing on Postarrival SARS-CoV-2-Positive Test Results Among International Travelers - CDC Traveler-Based Genomic Surveillance Program, Four U.S. Airports, March-September 2022. *MMWR Morb. Mortal. Wkly. Rep* 72, 206–209. [PubMed: 36821719]
27. Border crossing/entry data Bureau of Transportation Statistics, <https://www.bts.gov/browse-statistical-products-and-data/border-crossing-data/border-crossingentry-data>.
28. U.S. Customs and Border Protection (2020). Notification of Temporary Travel Restrictions Applicable to Land Ports of Entry and Ferries Service Between the United States and Mexico. *Federal Register* 85, 16547–16548.
29. U.S. Customs and Border Protection (2021). Notification of the Lifting of Temporary Travel Restrictions Applicable to Land Ports of Entry and Ferries Service Between the United States and Mexico for Certain Individuals Who Are Fully Vaccinated Against COVID-19 and Can Present Proof of COVID-19 Vaccination Status. *Federal Register* 86, 72843–72844.
30. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, and Andersen KG (2019). Tracking virus outbreaks in the twenty-first century. *Nat Microbiol* 4, 10–19. [PubMed: 30546099]
31. Bryant JA, Lamanna C, Morion H, Kerkhoff AJ, Enquist BJ, and Green JL (2008). Microbes on mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences* 105, 11505–11511. 10.1073/pnas.0801920105.
32. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, Webster S, Cameron-Blake E, Hallas L, Majumdar S, et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* 5, 529–538. [PubMed: 33686204]
33. Industry research <https://www.sandiego.org/about/industry-research.aspx>.
34. U.S. General Services Administration (2019). San Ysidro Land Port of Entry Fact Sheet.
35. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, and Tomasini M (2018). Human mobility: Models and applications. *Phys. Rep* 734, 1–74.
36. Simini F, González MC, Maritan A, and Barabási A-L (2012). A universal model for mobility and migration patterns. *Nature* 484, 96–100. [PubMed: 22367540]
37. Lemey P, Rambaut A, Drummond AJ, and Suchard MA (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol* 5, e1000520. [PubMed: 19779555]
38. True North Research (2020). Cross-Border Travel Behavior Survey Summary Report (San Diego Association of Governments).
39. COVID-19 Forecasting Team (2022). Variation in the COVID-19 infection-fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. *Lancet* 399, 1469–1488. [PubMed: 35219376]

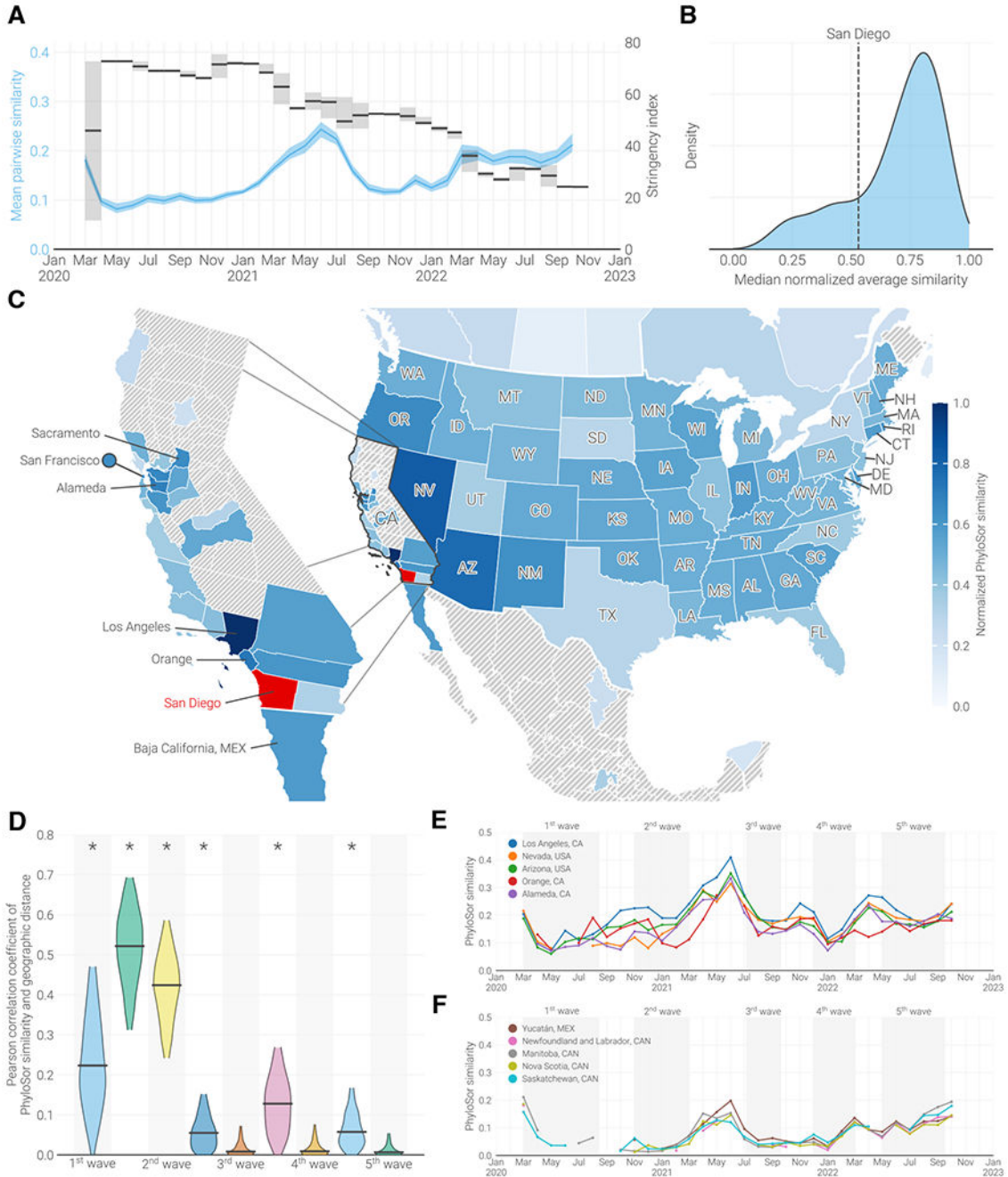
40. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, Giattino C, and Rodés-Guirao L (2021). A global database of COVID-19 vaccinations. *Nat Hum Behav* 5, 947–953. [PubMed: 33972767]
41. Weekly Patterns (2021). SafeGraph. <https://docs.safegraph.com/docs/weekly-patterns>.
42. Truscott J, and Ferguson NM (2012). Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Comput. Biol* 8, e1002699. [PubMed: 23093917]
43. Tegally H, Khan K, Huber C, de Oliveira T, and Kraemer MUG (2022). Shifts in global mobility dictate the synchrony of SARS-CoV-2 epidemic waves. *J. Travel Med* 29. 10.1093/jtm/taac134.
44. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Whittaker C, Zhu H, Berah T, Eaton JW, et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 584, 257–261. [PubMed: 32512579]
45. Liebig J, Najeebullah K, Jurdak R, Shoghri AE, and Pains D (2021). Should international borders re-open? The impact of travel restrictions on COVID-19 importation risk. *BMC Public Health* 27, 1573.
46. Galeazzi A, Cinelli M, Bonaccorsi G, Pierri F, Schmidt AL, Scala A, Pammolli F, and Quattrocchi W (2021). Human mobility in response to COVID-19 in France, Italy and UK. *Sci. Rep* 77, 13141.
47. Feng Y, Li Q, Tong X, Wang R, Zhai S, Gao C, Lei Z, Chen S, Zhou Y, Wang J, et al. (2020). Spatiotemporal spread pattern of the COVID-19 cases in China. *PLoS One* 75, e0244351.
48. Kraemer MUG, Hill V, Ruis C, Dellicour S, Bajaj S, McCrone JT, Baele G, Parag KV, Battle AL, Gutierrez B, et al. (2021). Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science* 373, 889–895. [PubMed: 34301854]
49. Tegally H, Wilkinson E, Martin D, Moir M, Brito A, Giovanetti M, Khan K, Huber C, Bogoch II, San JE, et al. (2022). Global Expansion of SARS-CoV-2 Variants of Concern: Dispersal Patterns and Influence of Air Travel. *medRxiv*. 10.1101/2022.11.22.22282629.
50. Tsui JL-H, Lambert B, Bajaj S, McCrone JT, Inward RPD, Bosetti P, Hill V, Pena RE, Zarebski AE, Peacock TP, et al. (2023). Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1. *medRxiv*. 10.1101/2023.01.02.23284109.
51. Brito AF, Semenova E, Dudas G, Hassler GW, Kalinich CC, Kraemer MUG, Ho J, Tegally H, Githinji G, Agoti CN, et al. (2021). Global disparities in SARS-CoV-2 genomic surveillance. *medRxiv*. 10.1101/2021.08.21.21262393.
52. Villabona-Arenas CJ, Hanage WP, and Tully DC (2020). Phylogenetic interpretation during outbreaks requires caution. *Nat Microbiol* 5, 876–877. [PubMed: 32427978]
53. Buckee CO, Balsari S, Chan J, Crosas M, Dominici F, Gasser U, Grad YH, Grenfell B, Halloran ME, Kraemer MUG, et al. (2020). Aggregated mobility data could help fight COVID-19. *Science* 368, 145–146.
54. Howard J, Huang A, Li Z, Tufekci Z, Zdimar V, van der Westhuizen H-M, von Delft A, Price A, Fridman L, Tang L-H, et al. (2021). An evidence review of face masks against COVID-19. *Proc. Natl. Acad. Sci. U. S. A* 778. 10.1073/pnas.2014564118.
55. Talic S, Shah S, Wild H, Gasevic D, Maharaj A, Ademi Z, Li X, Xu W, Mesa-Eguiagaray I, Rostron J, et al. (2021). Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: systematic review and meta-analysis. *BMJ* 375, e068302. [PubMed: 34789505]
56. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc* 72, 1261.
57. Moshiri N, Ragonnet-Cronin M, Wertheim JO, and Mirarab S (2019). FAVITES: simultaneous simulation of transmission networks, phylogenetic trees and sequences. *Bioinformatics* 35, 1852–1861. [PubMed: 30395173]
58. Cabrera B, Ross B, Röcher D, Brünker F, and Stieglitz S (2021). The influence of community structure on opinion expression: an agent-based model. *J. Bus. Econ. Manage* 91, 1331–1355.

59. Tonkin-Hill G, Martincorena I, Amato R, Lawson ARJ, Gerstung M, Johnston I, Jackson DK, Park N, Lensing SV, Quail MA, et al. (2021). Patterns of within-host genetic diversity in SARS-CoV-2. *Elite* 10. 10.7554/eLife.66857.
60. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, and Baele G (2020). Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol* 6, veaa061. [PubMed: 33235813]
61. Latora V, and Marchiori M (2001). Efficient behavior of small-world networks. *Phys. Rev. Lett* 87, 198701. [PubMed: 11690461]
62. Gini C (1921). Measurement of Inequality of Incomes. *Econ J* 31, 124–125.
63. Schäfer M, Strohmeier M, Lenders V, Martinovic I, and Wilhelm M (2014). Bringing up OpenSky: A large-scale ADS-B sensor network for research. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pp. 83–94.
64. New Confirmed COVID-19 Cases in SDSU student population <https://president.sdsu.edu/from-the-president/statements/page/new-confirmed-covid-19-cases-in-sdsu-student-population>.
65. Li H (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 10.1093/bioinformatics/btab705.
66. De Maio N, Walker C, Borges R, Weilguny L, Slodkowitz G, and Goldman N (2020). Issues with SARS-CoV-2 sequencing data.
67. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, and Lanfear R (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol* 37, 1530–1534. [PubMed: 32011700]
68. Hasegawa M, Kishino H, and Yano T (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol* 22, 160–174. [PubMed: 3934395]
69. Sagulenko P, Puller V, and Neher RA (2018). TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol* 4, vex042. [PubMed: 29340210]
70. Lemoine F, and Gascuel O (2021). Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genom Bioinform* 3, Iqab075.
71. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, and Rambaut A (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4, vey016. [PubMed: 29942656]
72. Pekar JE, Magee A, Parker E, Moshiri N, Izhikevich K, Havens JL, Gangavarapu K, Malpica Serrano LM, Crits-Christoph A, Matteson NL, et al. (2022). The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science* 377, 960–966. [PubMed: 35881005]
73. Rambaut A, Drummond AJ, Xie D, Baele G, and Suchard MA (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol* 67, 901–904. [PubMed: 29718447]
74. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, Russell CA, Smith DJ, Pybus OG, Brockmann D, et al. (2014). Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog*. 10, e1003932. [PubMed: 24586153]
75. Lemey P, Ruktanonchai N, Hong SL, Colizza V, Poletto C, Van den Broeck F, Gill MS, Ji X, Levasseur A, Oude Munnink BB, et al. (2021). Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature* 595, 713–717. [PubMed: 34192736]
76. Kraemer MUG, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott DM, open COVID-19 data working group, du Plessis L, Faria NR, Li R, et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *medRxiv*. 10.1101/2020.03.02.20026708.
77. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PGT, Fu H, et al. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis* 20, 669–677. [PubMed: 32240634]
78. Roques L, Klein EK, Papalix J, Sar A, and Soubeyrand S (2020). Using Early Data to Estimate the Actual Infection Fatality Ratio from COVID-19 in France. *Biology* 9. 10.3390/biology9050097.
79. Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, CMMID COVID-19 working group, Flasche S, Eggo RM, Edmunds WJ, et al. (2020). Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data

- from the outbreak on the Diamond Princess cruise ship, February 2020. *Euro Surveill.* 25. 10.2807/1560-7917.ES.2020.25.12.2000256.
80. Chapman LAC, Barnard RC, Russell TW, Abbott S, van Zandvoort K, Davies NG, and Kucharski AJ (2022). Unexposed populations and potential COVID-19 hospitalisations and deaths in European countries as per data up to 21 November 2021. *Euro Surveill.* 27. 10.2807/1560-7917.ES.2022.27.1.2101038.
81. Dahal S, Luo R, Swahn MH, and Chowell G (2021). Geospatial Variability in Excess Death Rates during the COVID-19 Pandemic in Mexico: Examining Socio Demographic, Climate and Population Health Characteristics. *Int. J. Infect. Dis* 113, 347–354. [PubMed: 34678505]
82. Chen Y-H, Riley AR, Duchowny KA, Aschmann HE, Chen R, Kiang MV, Mooney AC, Stokes AC, Glymour MM, and Bibbins-Domingo K (2022). COVID-19 mortality and excess mortality among working-age residents in California, USA, by occupational sector: a longitudinal cohort analysis of mortality surveillance data. *Lancet Public Health* 7, e744–e753. [PubMed: 36057273]
83. Gangavarapu K, Latif AA, Mullen JL, Alkuzweny M, Hufbauer E, Tsueng G, Haag E, Zeller M, Aceves CM, Zaiets K, et al. (2022). Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *medRxiv*, 2022.01.27.22269965.
84. Dong E, Du H, and Gardner L (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis* 20, 533–534. [PubMed: 32087114]

### Highlights

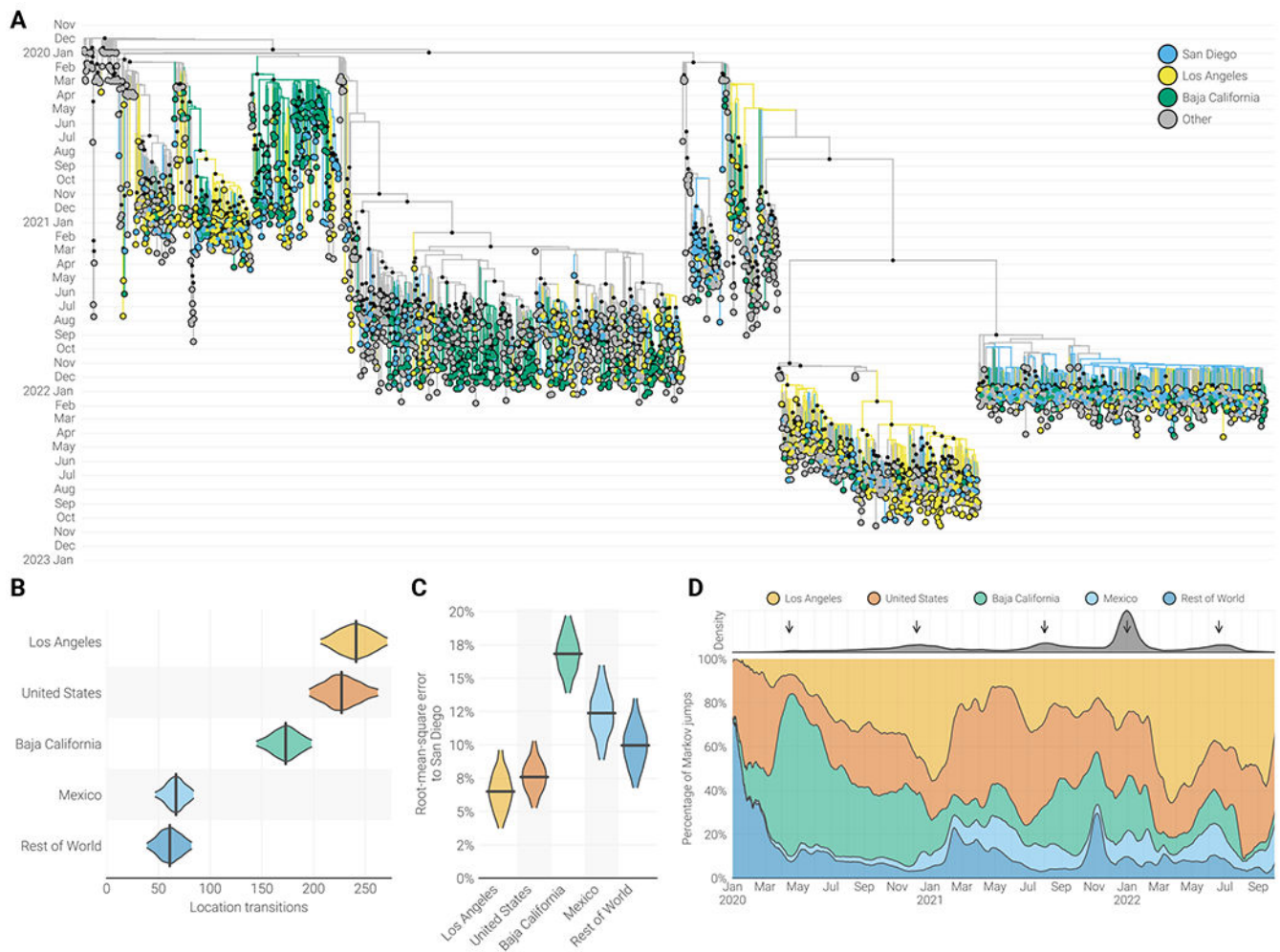
- Phylogenetic similarity of virus populations suggests connectivity between locations.
- COVID-19 mandates contained the spread of SARS-CoV-2 in the US.
- The lifting of mandates enabled SARS-CoV-2 to spread further as travel increased.
- Border closures to non-essential travel minimally impacted cross-border transmission.



**Figure 1. Regional similarity of SARS-CoV-2 genomes over time.**

(A) Primary axis, in blue, indicates temporal trends in the mean pairwise PhyloSor similarity of North American locations. Shaded region indicates 95% confidence interval as calculated by bootstrapping locations 100 times. Secondary axis, in black shows the mean stringency of the US government’s response to COVID-19. Higher values indicate a stricter response. Shaded area refers to the range of stringency values observed in a given month. (B) Distribution of median min-max normalized average PhyloSor similarity for all locations in North America. The median normalized phylogenetic similarity of San Diego to all

other locations is indicated by the dashed vertical line. **(C)** Map showing each location's median min-max normalized PhyloSor similarity to San Diego for the period of March 2020–August 2022. Here location refers to the county level within California and the state level in the rest of the United States, Canada, and Mexico. Each location is colored by their median value, and locations which were not included in the analysis are hashed out in gray. California is outlined in black and shown in greater detail in the inset on the left. San Diego is indicated in red. Some parts of Canada, Mexico, and the United States are excluded for clarity. **(D)** Pearson correlation coefficient between median PhyloSor similarity to San Diego and log-normalized centroid-centroid distance to San Diego for each period of the pandemic. Waves of cases are indicated by a gray box, while troughs reside between successive waves. Wave definitions can be found in Supp. Figure 6. Confidence intervals calculated by bootstrapping 1000 times. An asterisk indicates that the p-value of correlation is less than 0.05. **(E-F)** Temporal differences in PhyloSor similarity to San Diego for the 5 locations with the highest (E) and lowest (F) median normalized PhyloSor similarity to San Diego.

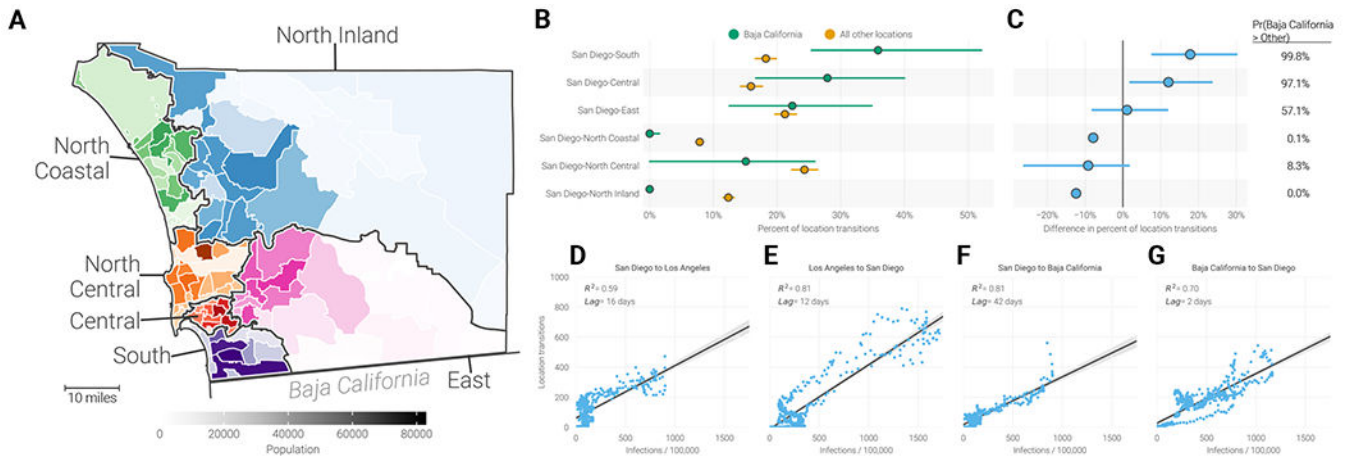


**Figure 2. Phylogenetic analysis of SARS-CoV-2 in the Californias.**

(A) Maximum clade credibility tree of whole genome SARS-CoV-2 sequences sampled from Baja California, Los Angeles County, San Diego, and the rest of the world. Black circles at internal nodes indicate posterior support greater than 0.5. (B) Median

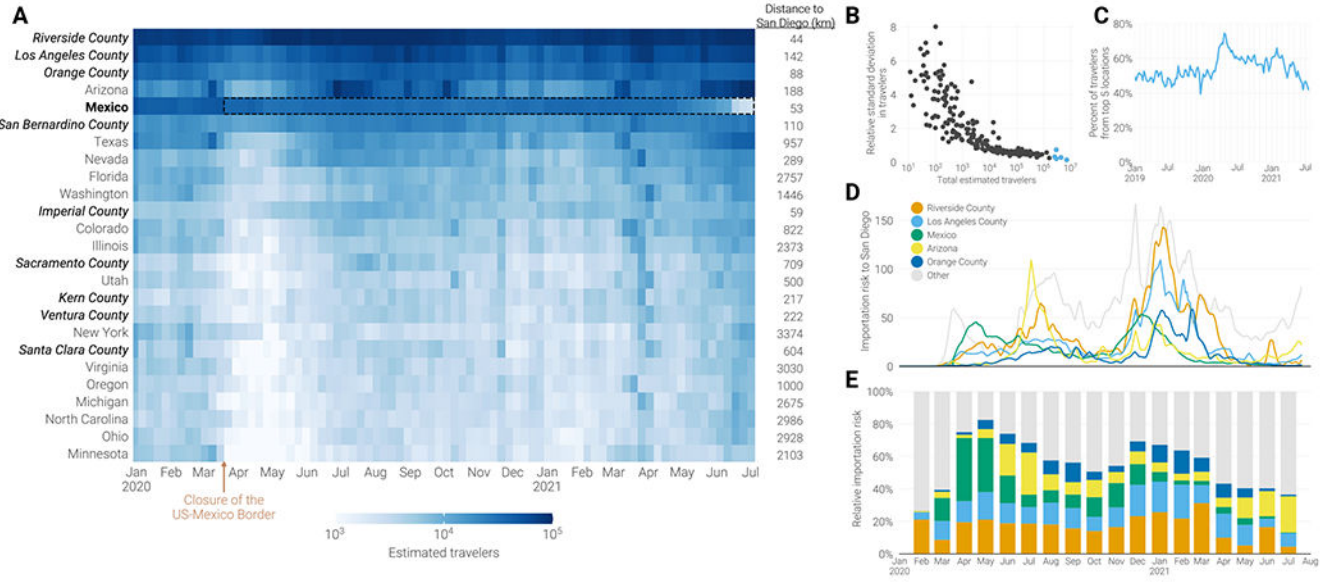
number of transitions between each location and San Diego inferred by phylogeographic reconstruction. Black bar indicates the median value. (C) Root-mean-square error between the estimated source composition of introductions into each location compared to San Diego. (D) Proportion of location transitions between San Diego and all other locations in the discrete state analysis. Top facet indicates the temporal density of location transitions across the posterior distribution of trees. Arrows are used to show periods of increased location transitions.





**Figure 3. Dynamics of cross-border transmission.**

(A) Boundary of approximate Health and Human Service Agency (HHSAs) regions within San Diego. ZIP codes are colored according to the HHSAs region they are in and their opacity is determined by their population (darker colors indicate a larger population). Scale bar indicates a distance of 10 miles. (B) Percentage of location transitions from either Baja California (in green) or all other locations (in orange) into San Diego that were inferred to land in each of the county's HHSAs regions. Dots indicate the median value while bars show the 95% highest posterior density interval. (C) Relative difference in percentage of location transitions originating in either Baja California or outside Baja California for each of San Diego's HHSAs region indicated in panel B. Probability refers to the percentage of trees in the posterior in which the proportion of location transitions from Baja California is greater than the proportion from all other locations combined. (D-G) Correlation between the magnitude of location transitions and the estimated number of infections at the origin for each location pair indicated.  $R^2$  was determined using ordinary least squares regression. In order to limit the impact of vaccines on our infection calculations, we only show the correlation for dates prior to May 5<sup>th</sup>, 2021 when at least 50% of San Diego's population received at least one dose of a SARS-CoV-2 vaccine<sup>39,40</sup>.



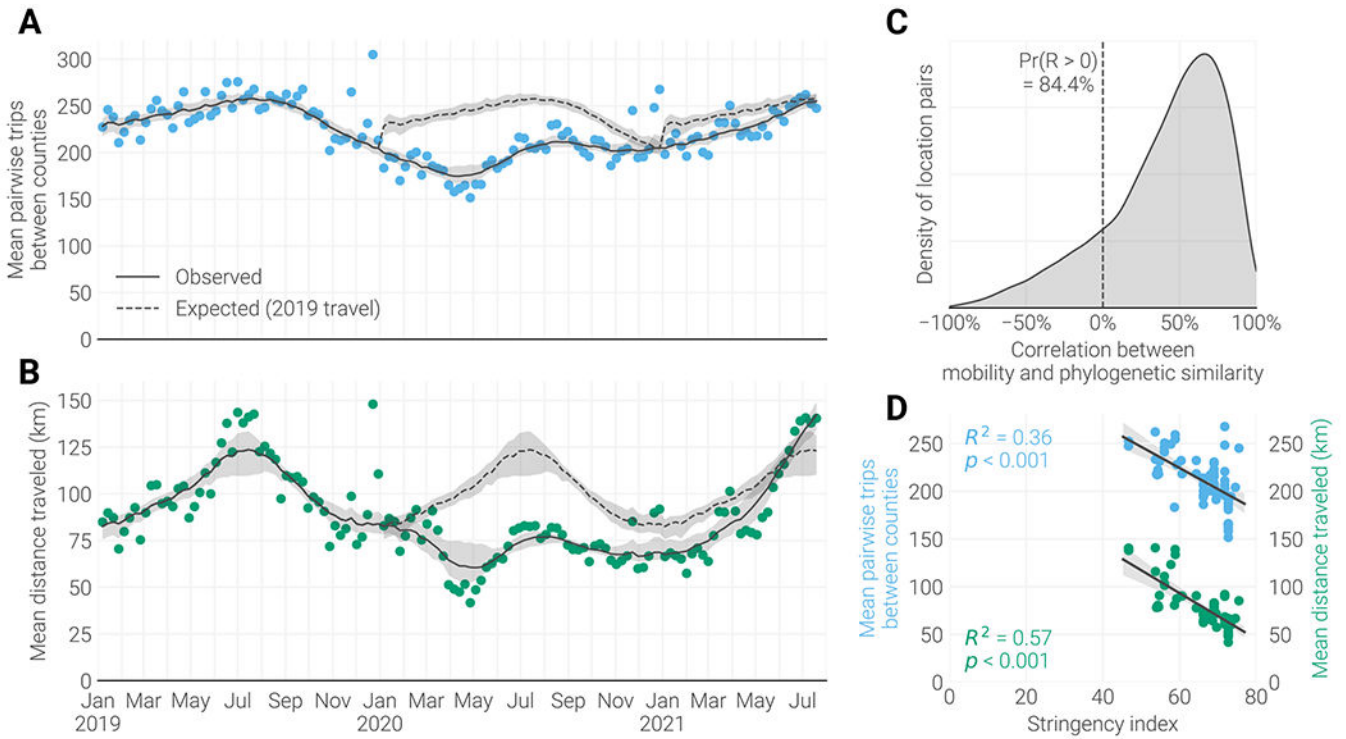
**Figure 4. SARS-CoV-2 import risk into San Diego.** (A) Weekly estimated number of travelers arriving into San Diego from January 2020-June 2021. Locations are sorted by the total number of estimated visitors over this period and only the top 25 are shown. Location names are styled depending on their administrative level: California counties are italicized, countries are bolded, and US states weighted normally. The centroid-centroid distance of each location to San Diego is listed to the right of the plot. A black dashed box indicates the closure of the US-Mexico border to non-essential travel from March 19<sup>th</sup>, 2020 to November 8<sup>th</sup>, 2021. (B) Scatter plot of each locations’ total estimated travelers into San Diego and the relative standard deviation in estimated travelers for the period indicated by panel A. The five locations with the greatest total estimated travelers into San Diego are highlighted in blue. (C) Proportion of travelers arriving into San Diego from the five locations with the greatest total estimated travelers (top-most five locations in panel A). (D) Import risk into San Diego. Import risk was estimated based on the number of infectious travelers relative to the population size and the total number of travelers at the origin. Only the five locations with the greatest total import risk into San Diego are shown. All other locations are colored in gray. (E) Relative import risk into San Diego. Locations are colored as in panel D, with gray representing all locations outside the top five locations.

Author Manuscript

Author Manuscript

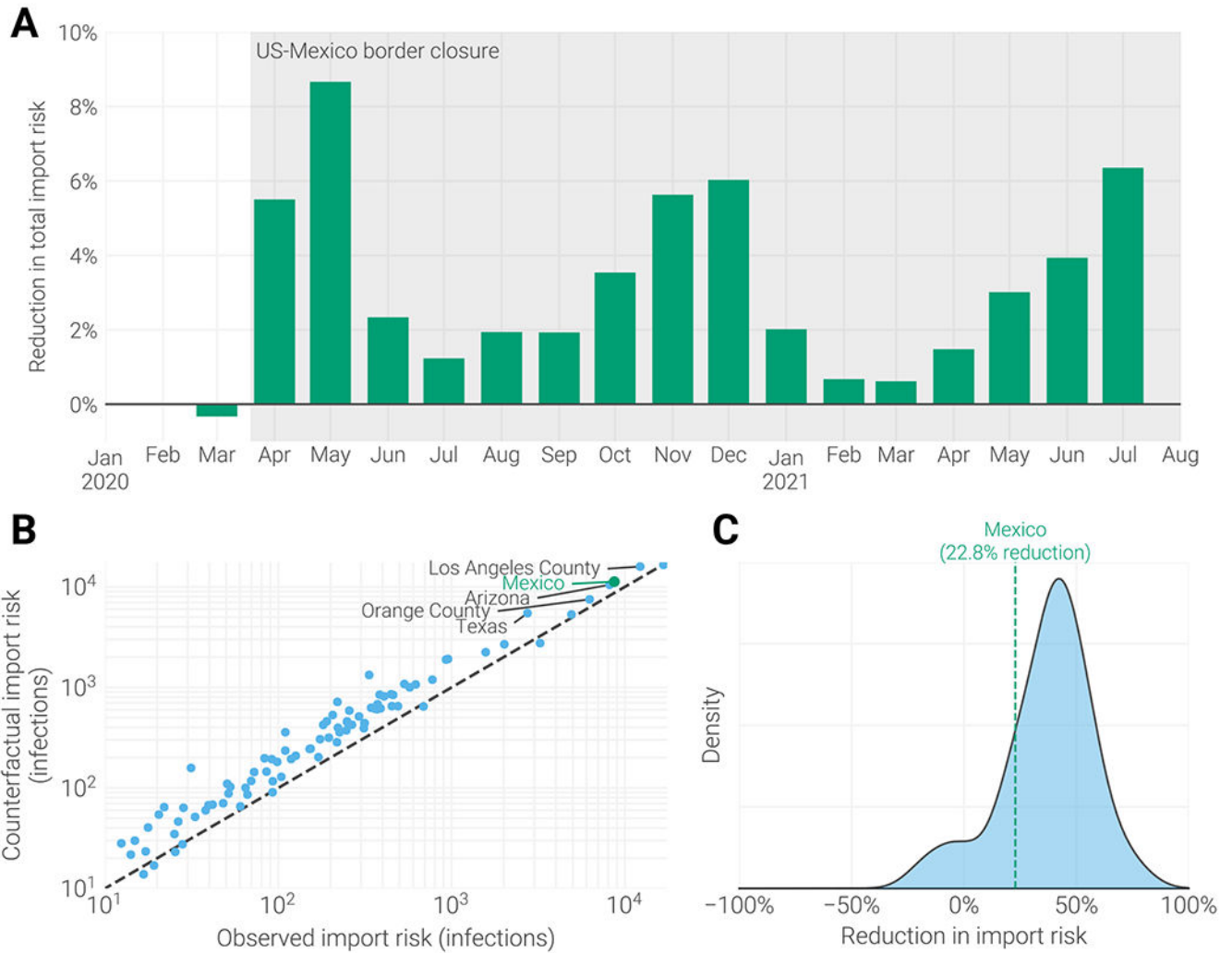
Author Manuscript

Author Manuscript



**Figure 5. Mobility changes in North America.**

(A) Mean weekly number of travelers traveling between each county in the US. Scatter points, in blue, indicate raw measurements. Temporal trend and 95% confidence intervals, indicated by the solid black line and shaded area, were calculated by bootstrapping LOESS regression 1000 times. Temporal trend from 2019 is transposed to 2020-2021, dashed line and shading, in order to visualize changes independent of season. (B) Mean weekly distance traveled by travelers in the US. Scatter points indicate raw measurements, while temporal trend and 95% confidence intervals were calculated by bootstrapping LOESS regression 1000 times. As with A, temporal trend from 2019 is transposed to 2020-2021, dashed line and shading, in order to visualize changes independent of season. (C) Distribution in Pearson correlation coefficients between mobility and PhyloSor similarity for each US location pair included in the PhyloSor analysis (see Methods). Dashed line labels the percentage of location pairs with a correlation coefficient greater than 0 (84.4%). (D) Primary axis, in blue, shows the weekly correlation between the stringency of the US government’s response to COVID-19 and the mean number of pairwise trips between US counties. Secondary axis, in green, shows the correlation between stringency and the mean distance traveled. For both axes, strength of correlation was determined using Pearson correlation coefficient. Because the stringency index aggregates a number of response indicators, many of which have little effect on mobility, correlation was only determined for dates after the first official stay-at-home order (March 15<sup>th</sup>, 2020)<sup>3</sup>.



**Figure 6. Impact of US-Mexico border closure.**

(A) Percentage reduction in the total import risk into San Diego when travel from Mexico is held at 2019-levels compared to observed travel. (B) Plot of total counterfactual import risk (calculated using mobility estimates from 2019) vs. observed import risk. To limit the impact of variability resulting from low traveler counts, only locations with an absolute import risk greater than 10 infected travelers are shown (accounting for 45% of all locations and 99.8% of the total import risk into San Diego). The five locations with the greatest difference between the counterfactual and observed import risk are labeled, import risk from Mexico is indicated with a green point. (C) Distribution of the relative reduction from the counterfactual to the observed import risk for each location in panel B. Mexico's relative reduction of 22.8% is indicated by the dashed vertical bar.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
Omega BioTek MagBind Viral DNA/RNA Kit	Omega Biotek	Cat#M6246-03
LunaScript RT	New England Biolabs	Cat#76509-480
Q5 DNA High-fidelity Polymerase	New England Biolabs	Cat#M0492L
AMPureXP beads	Beckman Coulter	Cat#A63882
Qubit High Sensitivity DNA assay kit	Invitrogen	Cat#Q32851
Tapestation D5000 tape	Agilent	Cat#5067-5588
Illumina NextSeq with 500/550 Mid Output Kit v2.5	Illumina	Cat#20024908
KingFisher Flex Purification System	ThermoFisher Scientific	Cat#5400630
Deposited Data		
SARS-CoV-2 reference genome	NCBI	NCBI: NC_045512.2
SARS-CoV-2 consensus sequences	GISAID	Table S1
SARS-CoV-2 raw data	NCBI	BioProject ID: PRJNA612578
BEAST XML and log files	This paper	<a href="https://github.com/andersen-lab/project_2023_SARS-CoV-2_Connectivity">https://github.com/andersen-lab/project_2023_SARS-CoV-2_Connectivity</a>
Epidemiological data	Outbreak.info	<a href="https://outbreak.info/">https://outbreak.info/</a>
Mobility data	SafeGraph	<a href="https://www.safegraph.com/covid-19-data-consortium">https://www.safegraph.com/covid-19-data-consortium</a>
Air travel flight data	OpenSky	<a href="https://zenodo.org/record/7923702">10.5281/zenodo.7923702</a>
COVID-19 policy stringency	Oxford Covid-19 Government Response Tracker	<a href="https://www.bsg.ox.ac.uk/covid-19">10.1038/s41562-021-01079-8</a>
Oligonucleotides		
ARTIC Network n-CoV-19 V3 primers	ARTIC Network	<a href="https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V3">https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V3</a>
Software and Algorithms		
Pangolin v2.0	Rambaut et al., 2020	<a href="https://github.com/cov-lineages/pangolin">https://github.com/cov-lineages/pangolin</a>
FAVITES V1.1.35	Moshiri et al., 2018	<a href="https://github.com/niemasd/FAVITES">https://github.com/niemasd/FAVITES</a>
minimap2 v2.17	Li, 2018	<a href="https://github.com/lh3/minimap2">https://github.com/lh3/minimap2</a>
Gofasta v0.0.6	Jackson, 2022	<a href="https://github.com/virus-evolution/gofasta">https://github.com/virus-evolution/gofasta</a>
IQ-TREE2	Nguyen et al., 2015	<a href="https://github.com/iqtree/iqtree2">https://github.com/iqtree/iqtree2</a>
TreeTime v0.7.4	Sagulenko et al., 2018	<a href="https://github.com/ve/vex042">10.1093/ve/vex042</a>
BEAST v1.10.5	Rambaut et al., 2021	<a href="https://github.com/beast-dev/beast-mcmc/tree/v1.10.5pre_thorney_v0.1.0">https://github.com/beast-dev/beast-mcmc/tree/v1.10.5pre_thorney_v0.1.0</a>
BEAGLE	Ayres et al., 2019	<a href="https://faculty.washington.edu/browning/beagle/beagle.html#download">https://faculty.washington.edu/browning/beagle/beagle.html#download</a>
Tracer v1.7.2	Rambaut et al., 2028	<a href="https://github.com/beast-dev/tracer/releases/tag/v1.7.2">https://github.com/beast-dev/tracer/releases/tag/v1.7.2</a>
Baltic	GitHub	<a href="https://github.com/evogytis/baltic">https://github.com/evogytis/baltic</a>
Snakemake	Köster and Rahmann, 2012	<a href="https://snakemake.readthedocs.io/en/stable/">https://snakemake.readthedocs.io/en/stable/</a>