medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title: Combining genomic data and infection estimates to characterize the complex dynamics of SARS-CoV-2 Omicron variants in the United States

Authors: Rafael Lopes[1,*], Kien Pham[1], Fayette Klaassen[2], Melanie H. Chitwood[1], Anne M. Hahn[1], Seth Redmond[1], Nicole A. Swartwood[2], Joshua A. Salomon[3], Nicolas A. Menzies[2], Ted Cohen[1,*,†], Nathan D. Grubaugh[1,4,*,†]

[1] Department of Epidemiology of Microbial Diseases and Public Health Modeling Unit, Yale School of Public Health, New Haven, CT, USA
[2] Department of Global Health and Population, Harvard T. H. Chan School of Public Health, Boston, MA, USA
[3] Department of Health Policy, Stanford University School of Medicine, Stanford, CA, USA
[4] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA
† Co-senior authors
* Corresponding authors: rafael.lopes@yale.edu, theodore.cohen@yale.edu, nathan.grubaugh@yale.edu

Abstract: SARS-CoV-2 Omicron surged as a variant of concern (VOC) in late 2021. Subsequently, several distinct Omicron variants have appeared and overtaken each other. We combined variant frequencies from GISAID and infection estimates from a nowcasting model for each US state to estimate variant-specific infections, attack rates, and effective reproduction numbers (R_t). BA.1 rapidly emerged, and we estimate that it infected 47.7% of the US population between late 2021 and early 2022 before it was replaced by BA.2. We estimate that BA.5, despite a slower takeoff than BA.1, also infected 35.7% of the US population, persisting in circulation for nearly 6 months. Other Omicron variants - BA.2, BA.4, or XBB - infected 30.7% of the US population. We found a positive correlation between the state-level BA.1 attack rate and social vulnerability. Our findings reveal the complex interplay between viral evolution, population susceptibility, and social factors since Omicron emerged in the US.

One-Sentence Summary

For each US state, we estimate Omicron variant-specific infections, attack rates, and effective reproduction numbers.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Main Text:

Introduction:

Nearly four years since the World Health Organization declared the COVID-19 outbreak as a pandemic, SARS-CoV-2 caused more than 778 million confirmed cases globally and more than 6.9 million deaths [1]. The emergence of genetically distinct SARS-CoV-2 variants of concern (VOC) posed a major challenge for control programs and greatly extended the length and health impact of the pandemic.

Following the emergence of the first major VOC, Alpha, in late 2020 [2], new VOCs have arisen and resulted in successive waves of infection [3], [4]. Alpha co-circulated with both Beta and Gamma variants (first detected contemporaneously in late 2020 in South Africa and Brazil [5], [6], respectively); these variants were subsequently replaced after the emergence and spread of the Delta variant [7] in mid-2021. The emergence of the Omicron variant, first detected in South Africa and Botswana in November 2021 [8], [9] was followed by rapid global spread and the replacement of the Delta variant.

Large-scale genomic sequencing of SARS-CoV-2 isolates collected from individuals with detected COVID-19 disease has been instrumental in documenting the evolution of successive VOC in many settings [5], [7], [9]–[11]. However, a considerable fraction of SARS-CoV-2 infections do not result in documented disease [12]–[15], especially after the introduction of vaccines and the development of partial immunity associated with previous infection [16]–[18]. Understanding the dynamics of transmission and strain replacement requires methods to infer time trends in variant-specific infections. Here, we combine nationwide SARS-CoV-2 sequencing data from GISAID with infection estimates from a Bayesian nowcasting model to better characterize the rise and fall of Omicron variants in the United States (US) between late 2021 and March 2023.

Results:

Quantifying variant-specific infections by combining variant frequency and infection estimates

The emergence and spread of multiple SARS-CoV-2 variants has been a hallmark of the COVID-19 pandemic. Combining 3,103,250 SARS-CoV-2 genomic sequences (**Figs. S1-S3**) and infection estimates from a nowcasting model (*covidestim* [19]; **Fig. 1A**), we estimated daily infections by each major variant of Omicron (BA.1*, BA.2*, BA.4*, BA.5*, and XBB*) from each US state and the District of Columbia from December 2021 to March 2023 (**Fig. 1**).

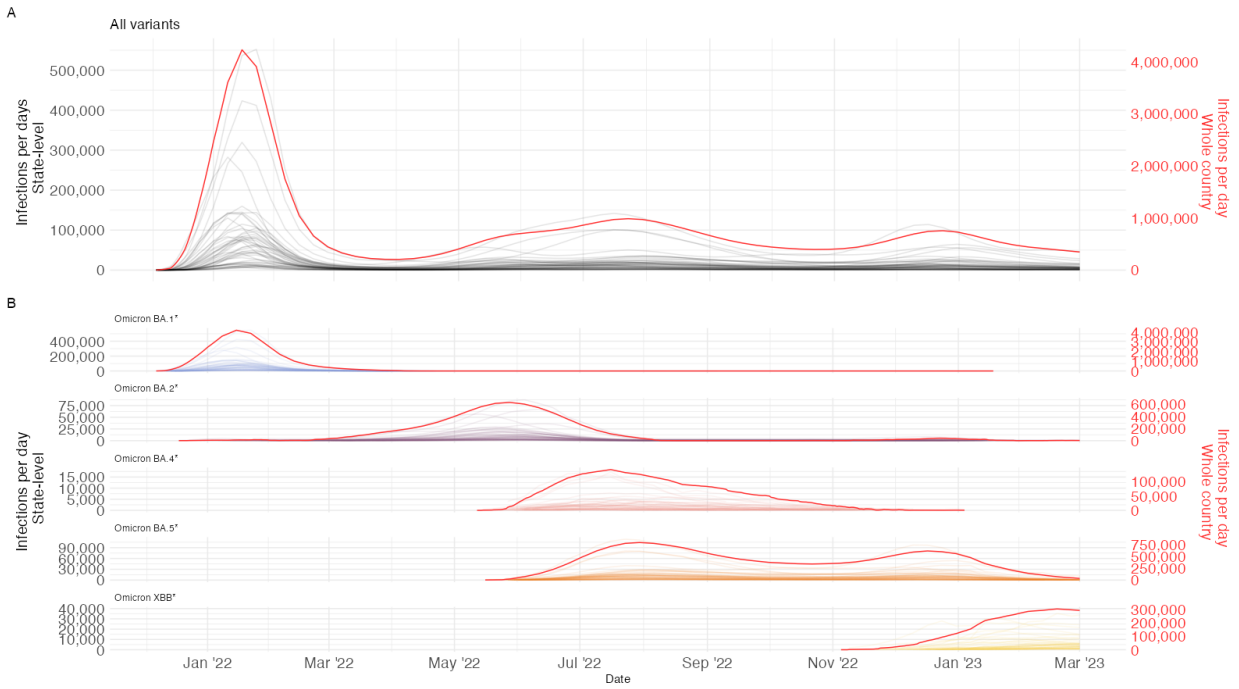
Reported cases, hospitalizations, and deaths provide an incomplete picture of the status of the COVID-19 pandemic since the majority of infections are asymptomatic. We address this by using infection estimates from *covidestim* [19], a nowcasting model that generates daily infection estimates while correcting for

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

under-reporting and notification delays (**Fig. 1A**). We then sorted the SARS-CoV-2 sequences for all 50 states and the District of Columbia and binned the lineages into variant categories - BA.1*, BA.2*, BA.4*, BA.5*, and XBB* (**Table S1**). Combining these two sets of analytic outputs, we calculated the daily frequencies of each Omicron variant (**Fig. S1**). We used this information to estimate the number of daily variant-specific infections via a spline interpolation (**Fig. 1B**). For more details see Materials and Methods.

We identified three peaks of infections in 2022 associated with the prevalence of distinct variants, one period in the winter, one in spring to early summer, and one in the late fall (**Fig. 1A**). The first Omicron period (BA.1*, December 2021 - January 2022) caused an estimated 4.2 million infections per day at its peak (about 1.25% of the US population) (**Fig. 1B, Tables 1** and **S2**). In total, we estimate that BA.1* caused approximately 160 million infections in the US during this wave (**Table 1**). The second Omicron period started in April 2022 (>2% frequency) with the emergence of Omicron BA.2* and lasted until November 2022 (<2% frequency) with the initial emergence of BA.4* and BA.5*. These variant-specific surges peaked at ~625,000 (BA.2*), ~140,000 (BA.4*), and ~800,000 (BA.5*) infections per day in the US. Finally, the third Omicron period, from November 2022 to March 2023, was driven by a resurgence of BA.5* and the emergence of the recombinant variant, XBB*, which peaked at ~500,000 and ~300,000 infections per day in the US, respectively.

At the state level, we estimated that the daily BA.1* infections peaked at ~548,000, ~422,000, ~318,000, and ~281,000, for California, Texas, Florida, and New York, respectively. Similar to our national estimates, these daily peaks of infection represent over 1% of the state population. We summarize the total and peak daily infections for each Omicron variant for all 50 states, the District of Columbia, and the whole country **(Tables 1, S2,** and **S3)**.



medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Fig. 1. Time series of daily Omicron variant infections across the entire United States.
The left y-axis, in black, is the state-level, and the right y-axis, in red, is the national scale. Note that the scale of the y-axis differs between time series.
A) Time series of infection estimates for all variants. The gray lines are infection estimates per state and the red line is the mean infection estimates per day for the whole US.
B) Time series of infection estimates for each variant. The gray lines are infection estimates per state and the red lines are the mean of infection estimates per day for the whole US. The scales differ by each variant subplot, as each variant had a different size of total infection per day.

Table 1. Variant-specific Attack Rate, Peak, and Total Infections for the United States

	Variant Categories					Total
	Omicron BA.1*	**Omicron BA.2***	**Omicron BA.4***	**Omicron BA.5***	**Omicron XBB***	
Attack Rate mean (max, min)						
	47.7% (57.2 - 37.9)%	17.2% (29.6 - 6.4)%	4.4% (6.7 - 2.3)%	35.7% (47.9 - 24.3)%	9.1% (15.6 - 3.4)%	
Peak Daily Infections						
	4,204,319	625,656	140,772	799,728	303,653	
Total Infections						
	161,762,895	52,112,475	14,179,483	119,878,658	28,635,157	376,568,668

Omicron variant attack rates for each state

We used the daily infections to calculate the percent of the population estimated to have been ever infected during each variant wave (variant-specific attack rates) for each US state (**Figs. 2** and **S4**). During the BA.1* wave, states with the highest attack rates - Kentucky (57.2%), Alabama (56.5%), and Louisiana (56.3%) - were concentrated in the southeast, while we estimate the lowest attack rates from Iowa (38.3%), South Dakota (38.0%), and Idaho (42.1%). The highest and lowest state attack rates for the other Omicron variants were as follows: BA.2* highest in Hawaii (30%), lowest in South Dakota (6%); BA.4* highest in North Carolina (6.8%), lowest in Vermont (2.3%); BA.5* highest in Kentucky (48%), lowest in Vermont (24%); XBB* highest in Rhode Island (15.6%), lowest is Arkansas (3.4%; **Fig. 2B, S4**). While Kentucky often had high attack rates and Vermont and South Dakota generally had lower attack rates, we did not detect consistent geographical patterns for each Omicron variant.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

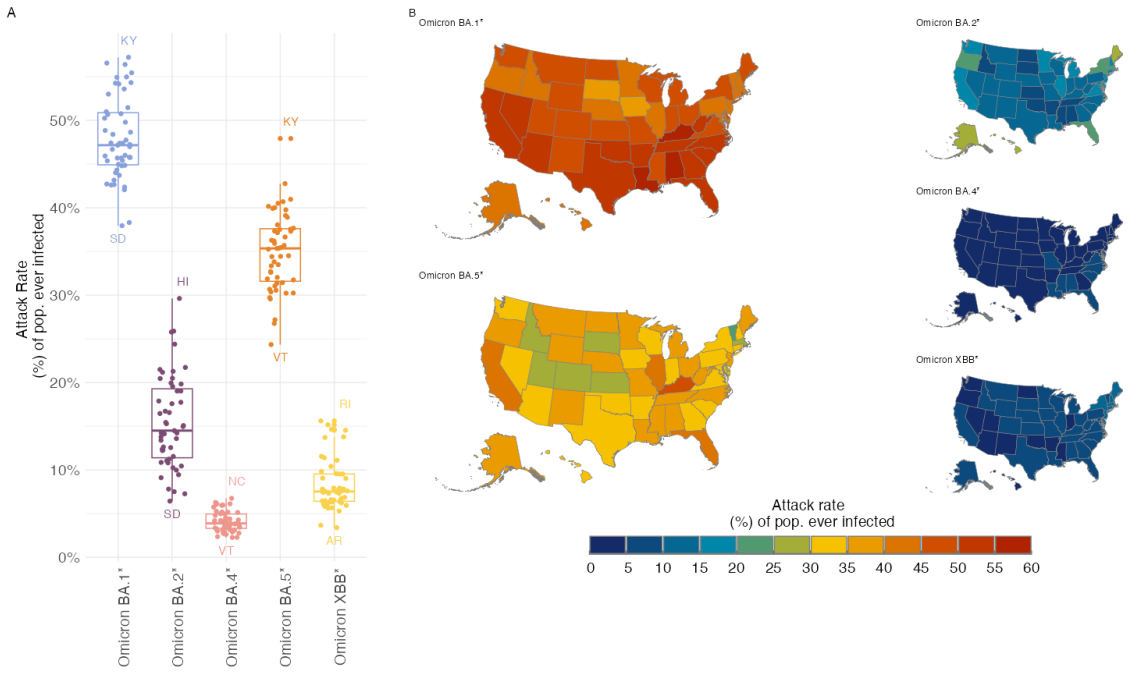


Fig. 2. Distribution of attack rate estimates across the United States for each Omicron variant.
A) Attack rate distribution and state-level attack rate estimates. Each dot is a state attack rate estimate, and the boxplots show the distribution of attack rate values across all states.
B) Maps of the attack rate estimates. For all the Omicron variants we show the US map, with Alaska and Hawaii placed below. Color on the state map indicates the state-level attack rate value of each variant.

Variant-specific effective reproduction numbers estimated from across the US

We estimated Omicron variant-specific effective reproductive numbers (R_t) for each state to gain insight into variant transmission (**Fig. 3**). We produced variant-specific estimates of R_t across all states by applying the *EpiEstim* R package [17] functions to our variant-specific daily infection estimates (**Fig. 1B**). For Omicron BA.1*, the median R_t across all states started as high as 3 (1.5, 3) (**Table S5**), while the R_t estimates for the other variants were smaller. We found similar longitudinal R_t estimates for BA.4* and BA.5*, indicating that they were generating similar numbers of secondary cases in the US and thus able to co-exist for several months. This observation suggests that there are variant-specific factors that can impact their relative transmissibility (e.g. immune escape, infectivity), but there are important population factors that also impact infection incidence.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

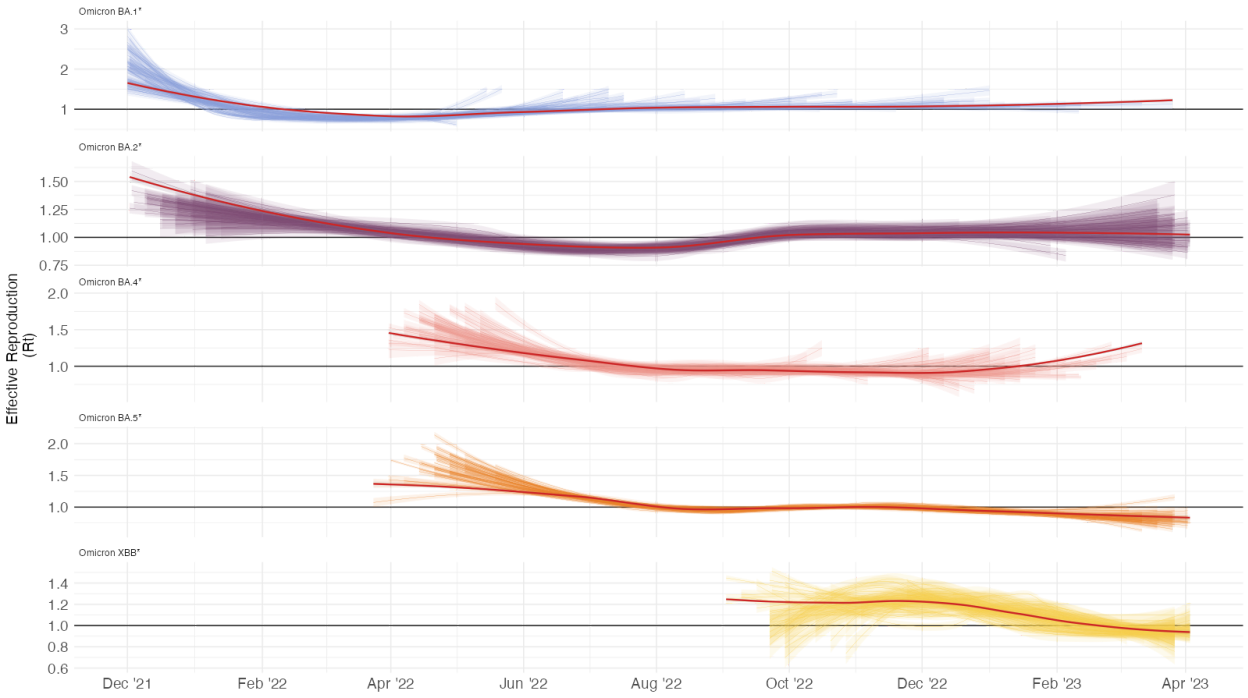


Fig. 3. Time series of variant-specific effective reproductive numbers across all states.
On each facet is depicted the time series for all US states and its confidence interval to the R_t estimate. The red line is the national average overall states. To help the visualization we apply over each state R_t time series a locally estimated scatterplot smoothing function (LOESS). The y-axes showing the R_t values are independently scaled for each variant to highlight changes over time.

Variant-specific associations between attack rates and social vulnerability

To investigate whether SARS-CoV-2 transmission is associated with sociodemographic factors, we examined correlations between our estimated outcomes and the CDC social vulnerability index (SVI) metric [20]. Comparing the state SVI (**Fig. 4A**) to the attack rates for each variant, weighted by the state population sizes (**Fig. 4B**), we found that the Omicron BA.1* (correlation coefficient $R = 0.56$). BA.4* ($R = 0.3$), and BA.5* ($R = 0.31$) attack rates positively correlate with the SVI (**Fig. 4B**). The BA.2* and XBB* emergences occurred immediately following the two largest Omicron waves, BA.1* and BA.5*, respectively. We, therefore, hypothesized that while individuals living in states with higher SVIs have higher exposure rates, they are less susceptible to infection during variant emergence immediately following exposure to a previous novel variant wave.

To test the hypothesis that states with higher SVI had higher exposure rates, we compared the Omicron BA.1* attack rates to those for BA.2* (peaked ~4 months after BA.1) and BA.5* (peaked ~6 months after BA.1). We calculated a negative correlation between the BA.1* and BA.2* attack rates ($R = -0.31$, 95% CI [-0.54, -0.04]; **Fig. 4C**) and a positive correlation between BA.1* and BA.5* ($R = 0.39$, 95% CI [0.13, 0.6]; **Fig. 4D**). States like Kentucky, Louisiana, and Alabama, which are on the higher end of the SVI scale, had attack rates that were relatively low for BA.1*, low for BA.2* attack rates, and high for BA.5*. Four states that did not fit the negative BA.1*-BA.2* correlations were South Dakota, Iowa, Idaho, and

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Nebraska, all of which had low SVI values and relatively low attack rates for both variants. Thus our analysis supports our hypothesis that variant waves are driven by opposing forces of social factors that govern exposure rates and population susceptibility following previous outbreaks.

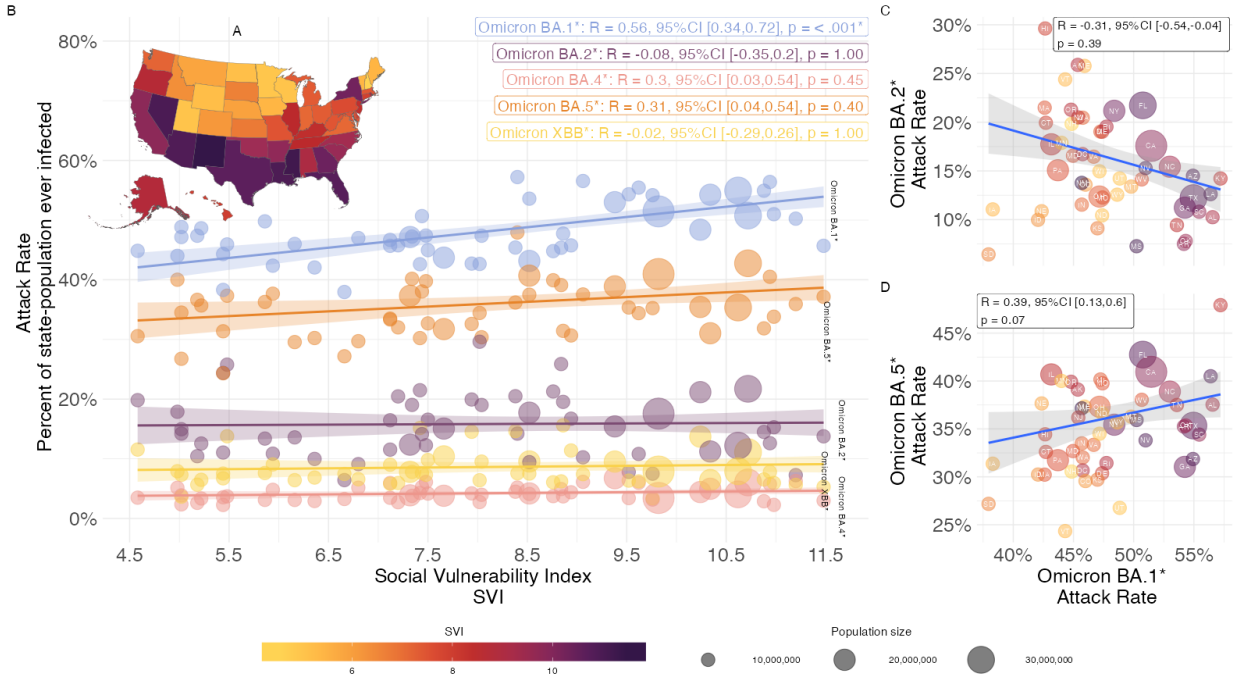


Fig. 4. Correlation between variant attack rates and the social vulnerability index.
A) Map of the SVI for all states, colors correspond to the SVI scores.
B) Scatterplot between attack rates by variant category and the SVI. Sizes are equivalent to the size of the state population and colors correspond to the variant categories as in the Panel B of **Fig. 1**.
C) Scatterplot between the attack rate of Omicron BA.1* and Omicron BA.2*, colors correspond to the SVI quartile, and size is proportional to the state population size. Correlation between the attack rates.
D) Scatterplot between the attack rate of Omicron BA.1* and Omicron BA.5*, colors correspond to the SVI quartile, and size is proportional to the state population size. Correlation between the attack rates.

Discussion:

We investigated the Omicron variant-specific infection dynamics across all US states, estimating daily infections, attack rates, and effective reproduction numbers. By combining sequencing data with infection estimates, we aimed to disambiguate infection dynamics during periods of strain replacement and when variants were co-circulating, revealing features of the epidemic that could not be inferred from the reported epidemiological data alone.

We found that Omicron variants were responsible for approximately 376 million infections across the US, including approximately 161 million during the BA.1* wave. The transmission dynamics of variants differed markedly: BA.1* emerged as a genetically distinct[3], [21]–[23]

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

variant which caused large rapid epidemics, especially in states with a higher degree of social vulnerability. Subsequent Omicron variants, while able to both co-circulate and eventually outcompete extant strains, spread at lower levels and often for longer durations, exhibiting much weaker association with social vulnerability measures than the BA.1* variant. These findings reveal the complex interplay between viral evolution, population susceptibility (driven by previous infections and population-level immunity), and social factors that affect the risk of exposure and infection.

The validity of our estimated variant-specific infections, attack rates, and effective reproduction numbers depends on several assumptions [24]–[27]. Our state-level estimates of total infections (i.e. not stratified by variant) were derived from a Bayesian nowcasting model which used publicly available time series of COVID-19 case notifications, hospitalizations, and deaths, accounting for effective population immunity. These estimates are calibrated to hospitalization and death data, accounting for delays associated with disease progression and estimates of infection hospitalization and infection fatality ratios [17]. We then used publically available SARS-CoV-2 sequencing data from GISAID to estimate variant frequencies at the state level to disaggregate the total number of infections into variant-specific incidence. As such, we assumed the sequencing was done at random within states. We also note that our analysis of the association between state-level attack rates and state-level SVI has the potential for ecological fallacy and should thus be interpreted with caution.

Our findings align with data from blood donors [28] and another modeling study in China [29]. The prevalence of anti-spike and anti-nucleocapsid antibodies (infection-induced and hybrid-induced) in the blood donor sample rose from 20.9% in April - June 2021 (Pre-Omicron), to 54.6% in January - March 2022, and then to 70.3% in July - September 2022. The latter two periods align with our estimates of the Omicron BA.1* and BA.5* waves. After the Omicron BA.5* wave, we estimate a cumulative attack rate of 83.4% of the US population. The China study estimates after the BA.5* introduction in a naive population, that 97% of the population had been infected. The overall attack rate we estimate is larger than the US population, which is explained by reinfections over the Omicron variant waves.

Our findings provide strong evidence that the dynamic evolution of SARS-CoV-2 variants is a result of the interplay between exposure and immunity to the virus [3], [18], [30], [31]. The pandemic's history has been marked by the initial emergence of highly transmissible variants [3], [7], [27], [32] and the Omicron era is marked with immune escape characteristics [3], [30], [33], necessitating ongoing adaptations in public health responses. By quantifying infection rates, attack rates, and effective reproduction numbers for different variants across all states, we provide valuable insights that can guide preparedness and resource allocation.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Materials and Methods:

First, we describe the processing of lineage information and how the lineages were summarized into categories. Second, we describe how the variant-specific infection estimates are produced by joining the infection estimate time series and variant frequency time series. Third, we describe the use of a modified version of *EpiEstim* tools to estimate the R_t for each of the variants. Lastly, we describe the joint analysis of the attack rate estimates and social vulnerability index scores.

Data Sources

The GISAID database contains more than 16 million genomes, of which approximately one-third come from US genomic surveillance efforts [11]. We processed the metadata and generated counts and frequencies of each variant lineage. Frequencies of variants have been used as a surveillance tool by the Centers for Disease Control and Prevention (CDC) and can give information on new invading variants. The GISAID metadata contains the Pango lineage nomenclature system classification of the genome. We can further distribute lineage information into variant categories by aggregating the major parental lineages and their sublineages into the same category. We categorized those lineages into major lineages categories (which we refer to as "Omicron variants"), such as Omicron BA.1* to incorporate Omicron BA.1 and its sublineages, Omicron BA.2* to Omicron BA.2 and its sublineages, and so on (**Table S1**).

The *covidestim* model estimates the time-series of SARS-CoV-2 infections in the United States by state and county based on cases, hospitalizations, and deaths data. The model output is a median of the infection estimates and its credible interval.

Lineages collapsing into major lineages categories

We pre-processed the metadata downloaded from GISAID and categorized the Pango lineages into 8 major categories: 'Omicron BA.1*', 'Omicron BA.2*', 'Omicron BA.3*', 'Omicron BA.4*', 'Omicron BA.5*', 'Omicron XBB*', 'Other Recombinant' and 'Other', see **Table 1** for details on each lineage and its sublineage alias.

As for the categories such as 'Omicron BA.3*', 'Other Recombinant', and 'Other', had less than 2% in frequency and we suppressed them from the main analysis.

We have collapsed all the sub-lineages into major categories, the following table summarizes our categorization. From the categorization, we count and calculate the frequency of each of these categories in every state and week. See **Fig. S1** in the supplementary material for the counts of genomic sequences for the whole US during the studied period, Dec 2021 to May 2023, into the 8 previously mentioned categories.

Variant-specific estimates by joining genomic frequencies and infection estimate

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We summarized the genomic sequence data to align with the *Covidestim* weekly infection estimates. From weekly counts, we calculate the frequency of each of the major variant categories described above. This process guarantees compatibility between the dates of metadata and infection estimates. We filter out frequencies below 2% on a week. By multiplying the frequencies of each category at each state by the number of total infections estimated for each state weekly, we produce estimates of the infections per variant in each state per week. We round the number of infections estimated to an integer number of infections.

In a formula, we have

$$I_{v,s}(t) = I_s(t) f_{v,s}(t),$$

Where the infection estimate time series for the variant *v* at state *s*, $I_{v,s}(t)$, is given by the infection estimate time series of total infection at state *s*, $I_s(t)$, times the frequencies time series of each variant *s* within the state *s*, $f_{v,s}(t)$. With $f_{v,s}(t) > 0.02$ every week.

We interpolate the weekly time series using a b-spline function to produce a daily time series of infections. We repeated the same procedure to the 2.5th and 97.5th quantiles of the infection estimates generated by *Covidestim*. We report the 2.5th and 97.5th quantiles of the posterior distribution trajectories as the lower and upper bound, respectively, for the credible interval (CrI) of the infection estimates. To compare the incidence estimates by each variant, we calculated the cumulative incidence over the epidemic of each variant for all states. The incidence is given as the percent of the population ever infected with the variant in the state.

Effective reproduction number estimates

The daily time series of each variant in each state was then given to the 'estimate_R()' function from the R package *EpiEstim*. To avoid non-converging problems with the model employed by *EpiEstim*, we only parse time series with more than ten days of continuous infection estimates. The model is parametrized using an uncertain Serial Interval (SI) setting, estimating the serial intervals of SARS-CoV-2 (Omicron variant specific) by drawing from two (truncated) normal distributions for the mean and standard deviation of the SI. The truncated normal distribution of the SI is then parametrized with a mean of 3.5 (1–6 days).

R_t ratios per variant for each state

We calculated the R_t ratio for pairs of variants to compare the R_t values between each variant across states (**Fig. S6**). We created pairs based on temporal succession; for time points with two or more variants co-circulating, we divided the succeeding variant time series by the preceding variant time series (**Fig. S6A**). In all pairs of succession, we found the average R_t ratio was greater than 1, meaning all the succeeding variants were more transmissible and capable of invasion (**Fig. S6B** and **Fig. S5**).

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Social Vulnerability Index aggregation to the state level

The social vulnerability index (SVI) is a metric compiled by the CDC summarizing the social conditions that may affect the outcome in the face of disasters, such as infectious disease outbreaks[18](**Fig. 4**). The SVI is a summary metric, incorporating 4 main themes: socioeconomic status; household characteristics; racial and ethnic minority status; and housing type and transportation. States that are high on the SVI scale tend to have larger populations and are primarily concentrated in the southern half of the US (**Fig. 4A**). Originally the index was compiled at the census tract and county level; we have aggregated them by state to be able to use it with the state-level estimates of infections by variant.

Data availability:

The findings of this study are based on metadata associated with 3,103,250 sequences available on GISAID from September 1st, 2021 up to April 22, 2023, and accessible at https://doi.org/10.55876/gis8.231023hd (GISAID Identifier: EPI_SET_231023hd).
All genome sequences and associated metadata in this dataset are published in GISAID's EpiCoV database. To view the contributors of each sequence with details such as accession number, Virus name, Collection date, Originating Lab and Submitting Lab, and the list of Authors, visit https//doi.org/10.55876/gis8.231023hd

Data Snapshot

EPI_SET_231023hd is composed of 3,103,250 individual genome sequences with collection dates ranging from 2021-09-01 to 2023-04-22. Data were collected in 2 countries and territories. All sequences in this dataset are compared relative to hCoV-19/Wuhan/WIV04/2019 (WIV04), the official reference sequence employed by GISAID (EPI_ISL_402124). Learn more at https://gisaid.org/WIV04.

Code availability

The pipeline used to calculate the variant-specific infections, attack rates, R_t, R_t ratio, and SVI comparison is available on the following GitHub repository:
https://github.com/rafalopespx/Variant_infections_rate

Acknowledgments:

We gratefully acknowledge T. Thornhill for the discussion and helpful insights with the Social Vulnerability Index and P. Jack and S. Taylor for technical support, and the authors from the originating laboratories responsible for obtaining the specimens, as well as the submitting laboratories where the genomic data were generated and shared via GISAID, on which this research is based. This project is supported by Cooperative Agreement NU38OT000297 from the

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Centers for Disease Control and Prevention (CDC) and the Council of State and Territorial Epidemiologists (CSTE), SHEPheRD Contract 200-2016-91779 from the CDC, and the CDC Broad Agency Announcement Contract 75D30122C14697. This work does not necessarily represent the views of the CDC or CSTE. All plots use color palettes from the 'MetBrewer' R package https://github.com/BlakeRMills/MetBrewer.

Author contribution:

Conceptualization: RL, TC, NDG

Methodology: RL, KP, FK, TC, NDG

Investigation: RL, SR, AH

Visualization: RL

Funding acquisition: TC, NDG

Supervision: JAS, NAM, TC, NDG

Writing – original draft: RL, TC, NDG

Writing – review & editing: all authors

Competing Interests:

NDG is a paid consultant for BioNTech.

References:

[1] "WHO Coronavirus (COVID-19) Dashboard." Accessed: Oct. 22, 2023. [Online]. Available: https://covid19.who.int

[2] N. G. Davies *et al.*, "Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England," *Science*, vol. 372, no. 6538, p. eabg3055, Apr. 2021, doi: 10.1126/science.abg3055.

[3] C. Roemer *et al.*, "SARS-CoV-2 evolution in the Omicron era," *Nat. Microbiol.*, pp. 1–8, Oct. 2023, doi: 10.1038/s41564-023-01504-w.

[4] D. P. Martin *et al.*, "The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages," *Cell*, vol. 184, no. 20, pp. 5189-5200.e7, Sep. 2021, doi: 10.1016/j.cell.2021.09.003.

[5] H. Tegally *et al.*, "Detection of a SARS-CoV-2 variant of concern in South Africa," *Nature*, vol. 592, no. 7854, Art. no. 7854, Apr. 2021, doi: 10.1038/s41586-021-03402-9.

[6] N. R. Faria *et al.*, "Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil," *Science*, vol. 372, no. 6544, pp. 815–821, May 2021, doi: 10.1126/science.abh2644.

[7] R. Earnest *et al.*, "Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA," *Cell Rep. Med.*, vol. 3, no. 4, Apr. 2022, doi: 10.1016/j.xcrm.2022.100583.

[8] R. Viana *et al.*, "Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa," *Nature*, vol. 603, no. 7902, Art. no. 7902, Mar. 2022, doi: 10.1038/s41586-022-04411-y.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[9] H. Tegally *et al.*, "Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa," *Nat. Med.*, vol. 28, no. 9, Art. no. 9, Sep. 2022, doi: 10.1038/s41591-022-01911-2.

[10] E. Volz *et al.*, "Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England," *Nature*, vol. 593, no. 7858, Art. no. 7858, May 2021, doi: 10.1038/s41586-021-03470-x.

[11] A. F. Brito *et al.*, "Global disparities in SARS-CoV-2 genomic surveillance," *Nat. Commun.*, vol. 13, no. 1, Art. no. 1, Nov. 2022, doi: 10.1038/s41467-022-33713-y.

[12] K. L. Bajema *et al.*, "Estimated SARS-CoV-2 Seroprevalence in the US as of September 2020," *JAMA Intern. Med.*, vol. 181, no. 4, pp. 450–460, Apr. 2021, doi: 10.1001/jamainternmed.2020.7976.

[13] D. P. Oran and E. J. Topol, "Prevalence of Asymptomatic SARS-CoV-2 Infection," *Ann. Intern. Med.*, vol. 173, no. 5, pp. 362–367, Sep. 2020, doi: 10.7326/M20-3012.

[14] M. D. T. Hitchings *et al.*, "The Usefulness of the Test-Positive Proportion of Severe Acute Respiratory Syndrome Coronavirus 2 as a Surveillance Tool," *Am. J. Epidemiol.*, vol. 190, no. 7, pp. 1396–1405, Jul. 2021, doi: 10.1093/aje/kwab023.

[15] B. Rader, "Use of At-Home COVID-19 Tests — United States, August 23, 2021–March 12, 2022," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 71, 2022, doi: 10.15585/mmwr.mm7113e1.

[16] F. Klaassen *et al.*, "Population Immunity to Pre-Omicron and Omicron Severe Acute Respiratory Syndrome Coronavirus 2 Variants in US States and Counties Through 1 December 2021," *Clin. Infect. Dis.*, vol. 76, no. 3, pp. e350–e359, Feb. 2023, doi: 10.1093/cid/ciac438.

[17] F. Klaassen *et al.*, "Changes in Population Immunity Against Infection and Severe Disease From Severe Acute Respiratory Syndrome Coronavirus 2 Omicron Variants in the United States Between December 2021 and November 2022," *Clin. Infect. Dis.*, vol. 77, no. 3, pp. 355–361, Aug. 2023, doi: 10.1093/cid/ciad210.

[18] P. O. Ankomah, M. J. Siedner, and R. P. Bhattacharyya, "Pre-Existing Population Immunity and severe acute respiratory syndrome coronavirus 2 Variant Establishment and Dominance Dynamics in the United States: An Ecological Study," *Open Forum Infect. Dis.*, vol. 9, no. 12, p. ofac621, Dec. 2022, doi: 10.1093/ofid/ofac621.

[19] M. H. Chitwood *et al.*, "Reconstructing the course of the COVID-19 epidemic over 2020 for US states and counties: Results of a Bayesian evidence synthesis model," *PLOS Comput. Biol.*, vol. 18, no. 8, p. e1010465, Aug. 2022, doi: 10.1371/journal.pcbi.1010465.

[20] "CDC/ATSDR SVI Data and Documentation Download | Place and Health | ATSDR." Accessed: Oct. 06, 2023. [Online]. Available: https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html

[21] M. Kandeel, M. E. M. Mohamed, H. M. Abd El-Lateef, K. N. Venugopala, and H. S. El-Beltagi, "Omicron variant genome evolution and phylogenetics," *J. Med. Virol.*, vol. 94, no. 4, pp. 1627–1632, 2022, doi: 10.1002/jmv.27515.

[22] A. Lentini, A. Pereira, O. Winqvist, and B. Reinius, "Monitoring of the SARS-CoV-2 Omicron BA.1/BA.2 lineage transition in the Swedish population reveals increased viral RNA levels in BA.2 cases," *Med*, vol. 3, no. 9, pp. 636-643.e4, Sep. 2022, doi: 10.1016/j.medj.2022.07.007.

[23] C. van Dorp, E. Goldberg, R. Ke, N. Hengartner, and E. Romero-Severson, "Global estimates of the fitness advantage of SARS-CoV-2 variant Omicron," *Virus Evol.*, vol. 8, no. 2, p. veac089, Jul. 2022, doi: 10.1093/ve/veac089.

[24] P. Nouvellet *et al.*, "A simple approach to measure transmissibility and forecast incidence," *Epidemics*, vol. 22, pp. 29–35, Mar. 2018, doi: 10.1016/j.epidem.2017.02.012.

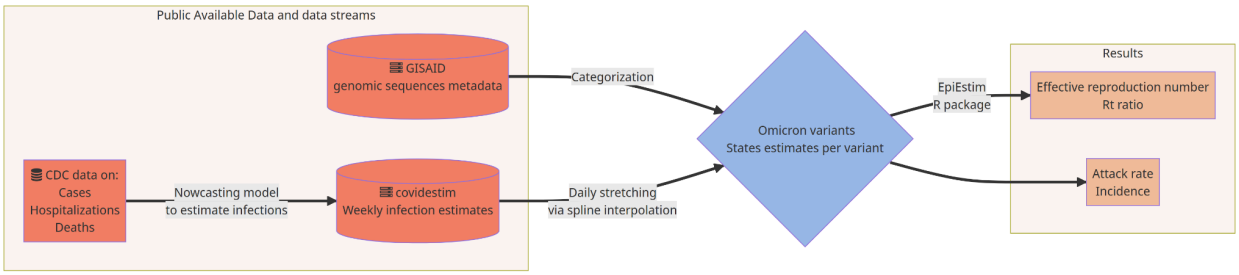
[25] R. K. Nash, P. Nouvellet, and A. Cori, "Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges," *PLOS Digit. Health*, vol. 1, no. 6, p. e0000052, Jun. 2022, doi: 10.1371/journal.pdig.0000052.

[26] T. Britton and G. Scalia Tomba, "Estimation in emerging epidemics: biases and remedies," *J. R. Soc. Interface*, vol. 16, no. 150, p. 20180670, Jan. 2019, doi: 10.1098/rsif.2018.0670.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[27] E. Volz, “Fitness, growth and transmissibility of SARS-CoV-2 genetic variants,” *Nat. Rev. Genet.*, pp. 1–11, Jun. 2023, doi: 10.1038/s41576-023-00610-z.
[28] J. M. Jones, “Estimates of SARS-CoV-2 Seroprevalence and Incidence of Primary SARS-CoV-2 Infections Among Blood Donors, by COVID-19 Vaccination Status — United States, April 2021–September 2022,” *MMWR Morb. Mortal. Wkly. Rep.*, vol. 72, 2023, doi: 10.15585/mmwr.mm7222a3.
[29] E. E. Goldberg, Q. Lin, E. O. Romero-Severson, and R. Ke, “Swift and extensive Omicron outbreak in China after sudden exit from ‘zero-COVID’ policy,” *Nat. Commun.*, vol. 14, no. 1, Art. no. 1, Jul. 2023, doi: 10.1038/s41467-023-39638-4.
[30] X. Zhang *et al.*, “SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance,” *Signal Transduct. Target. Ther.*, vol. 6, no. 1, Art. no. 1, Dec. 2021, doi: 10.1038/s41392-021-00852-5.
[31] S. M. Hirabara *et al.*, “SARS-COV-2 Variants: Differences and Potential of Immune Evasion,” *Front. Cell. Infect. Microbiol.*, vol. 11, 2022, Accessed: Oct. 16, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcimb.2021.781429
[32] B. A. Petros *et al.*, “Early Introduction and Rise of the Omicron Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Variant in Highly Vaccinated University Populations,” *Clin. Infect. Dis.*, vol. 76, no. 3, pp. e400–e408, Feb. 2023, doi: 10.1093/cid/ciac413.
[33] A. M. Carabelli *et al.*, “SARS-CoV-2 variant biology: immune escape, transmission and fitness,” *Nat. Rev. Microbiol.*, vol. 21, no. 3, Art. no. 3, Mar. 2023, doi: 10.1038/s41579-022-00841-7.

Supplementary material:



medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Fig. S1. Flowchart of the process of joining genomic and epidemiological data streams. From the genomic sequences metadata GISAID and infection estimates from *covidestim*, we produced infection estimates by multiplying the frequencies of each Omicron variant by the infection estimates. Those estimates are then imputed to *EpiEstim* functions to produce variant-specific effective reproduction numbers, $R_{t,}$ and state attack rates.

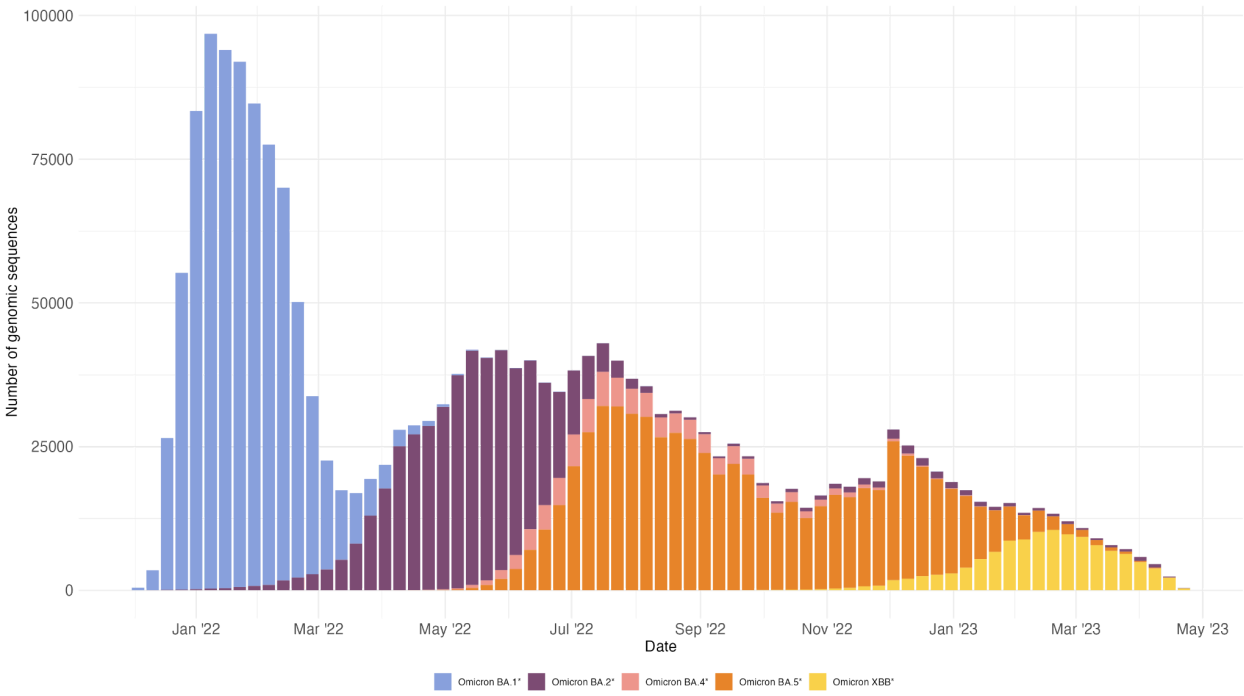


Fig. S2. Number of genomic sequences per variant category per week during the period of December 1st, 2021 to May 1st, 2023, to the whole country. From the GISAID metadata, we calculate the amount of sequences deposited to the database per week, during the analyzed period. Each bar is a week of the period and the filling of the bar is the frequency of each variant during that week. It is possible to see the pattern of succession of variants over the year.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

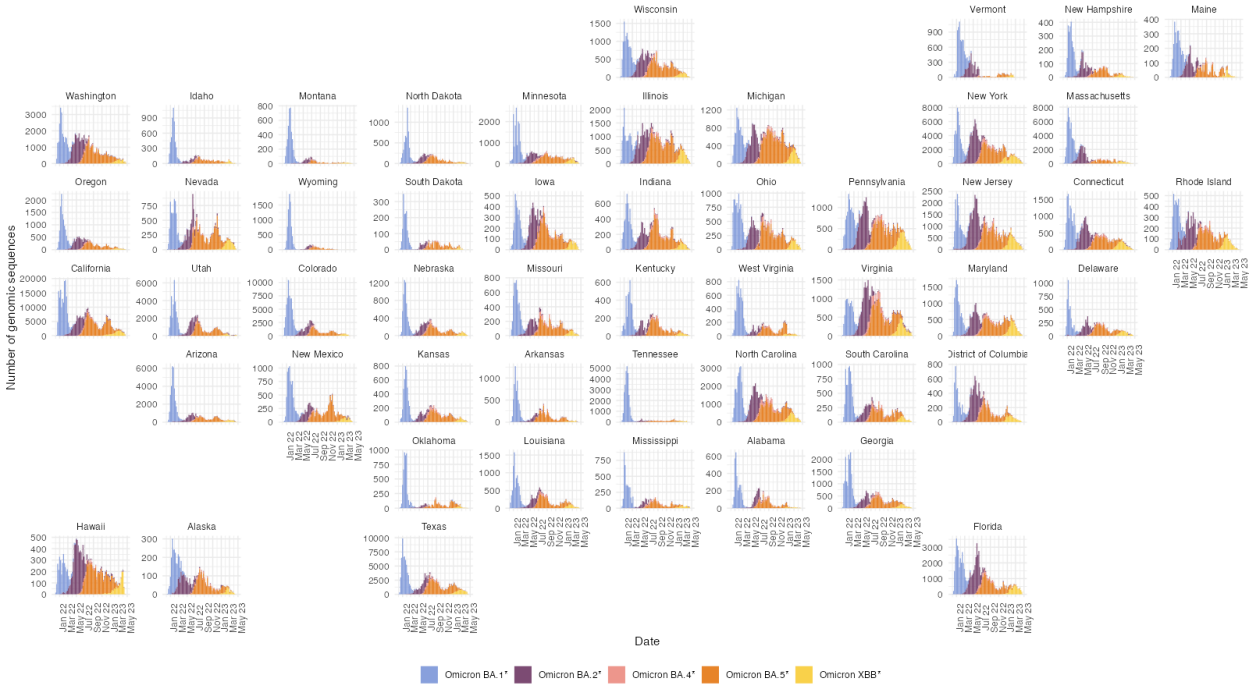


Fig. S3. Number of genomic sequences per variant category per week during the period of December 1st, 2021 to May 1st, 2023, to all individual states. From the GISAID metadata, we calculate the amount of sequences deposited to the database per week per state, during the analyzed period. Each bar is a week of the period and the filling of the bar is the frequency of each variant during that week. It is possible to see the pattern of succession of variants over the year.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

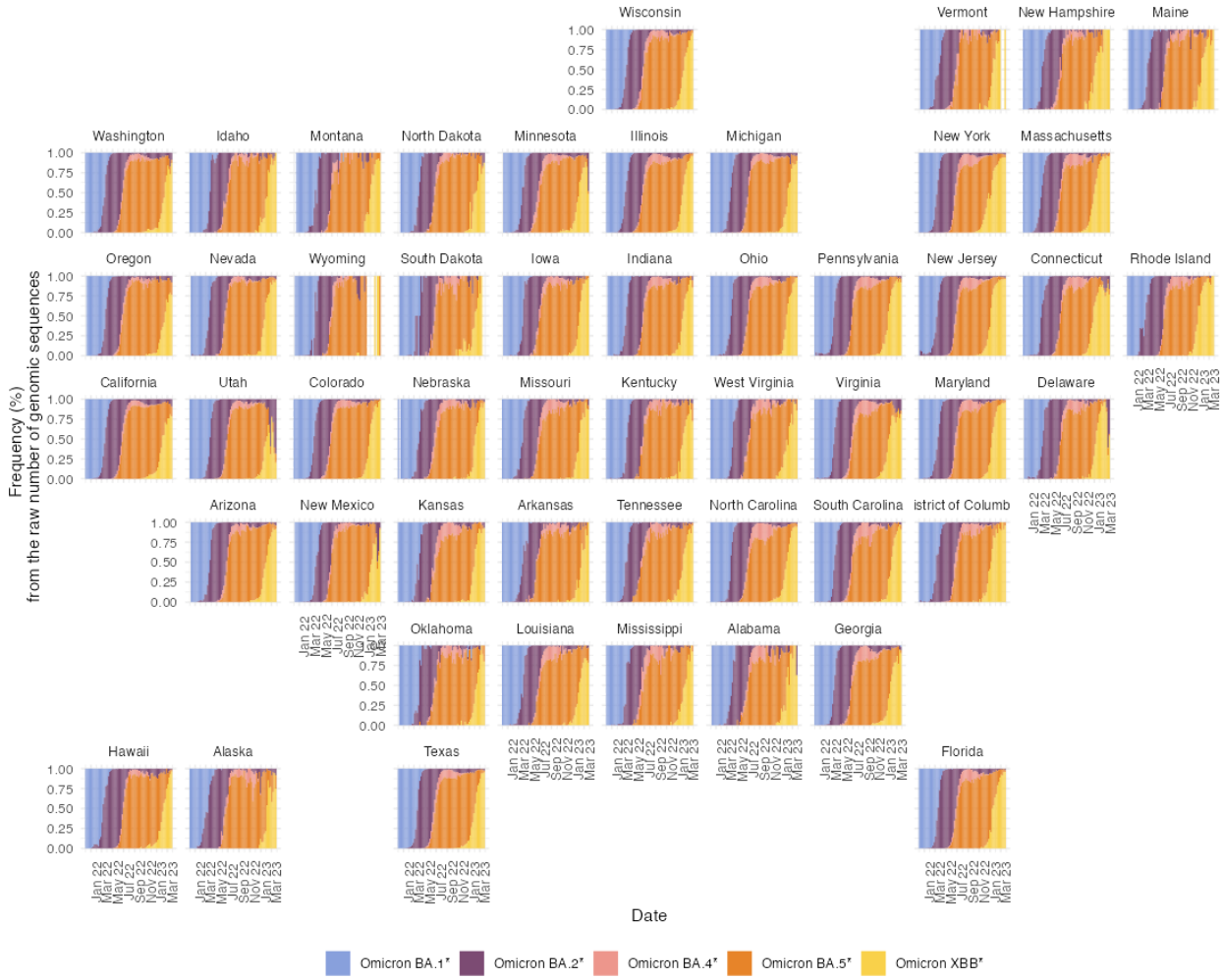


Fig. S4. Frequency from the raw number of genomic sequences per variant category, during the period ranging from December 1st, 2021 to May 1st, 2023, over all the individual states. From the GISAID metadata, we calculate the amount of sequences deposited to the database per week, during the analyzed period. Each bar is a week of the period and the filling of the bar is the frequency of each variant during that week. It is possible to see the pattern of succession of variants over the year.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

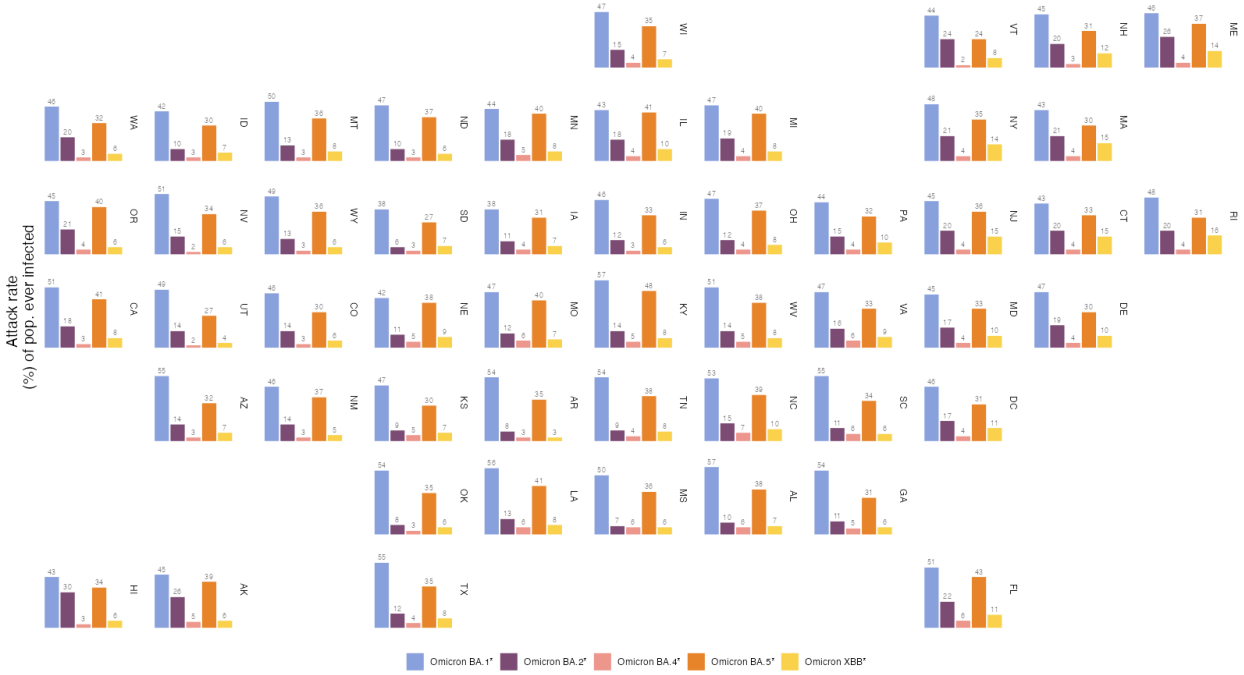


Fig. S5. Attack rate per each variant category for all individual states. Bar chart to the variant-specific attack rates estimates in the layout of the US states. Each chart is the attack rates of the variants with the corresponding color. The double-letter state abbreviation is displayed on the right side of each subchart.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

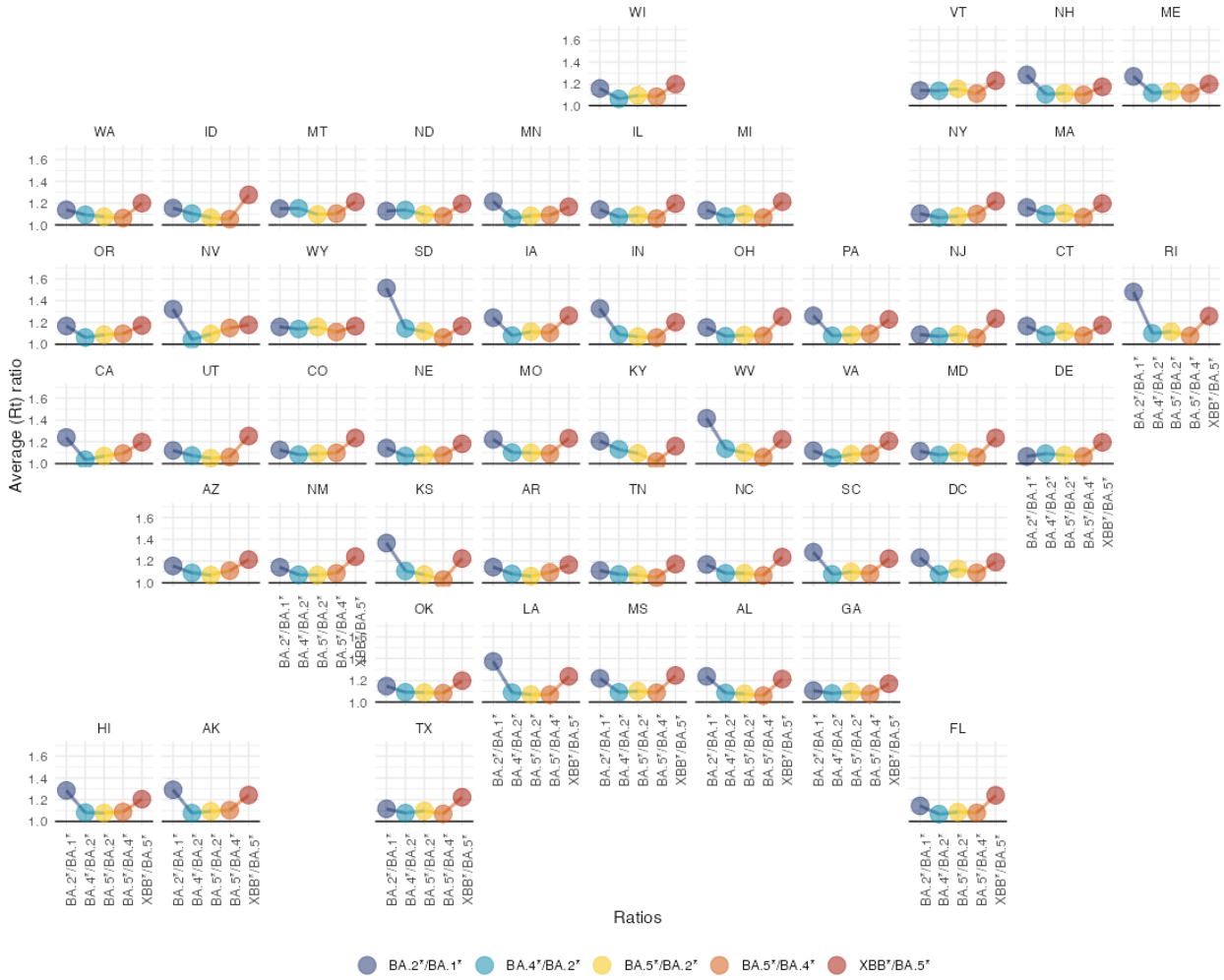


Fig. S6. Effective reproduction number (R_t) ratios to each pair of succeeding variants by each state overall. The R_t ratio is calculated by dividing the average R_t of the predecessor variant by the successor variant. When the slope rises it means the entering variant has a larger value of average R_t over the predecessor variant.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Replacement of variants is marked with a higher R_t ratio, and the co-circulation of variants is marked with a smaller R_t ratio

We estimate the advantage of one variant over another by taking the R_t ratios during their period of coexistence. From the R_t ratios, we can classify two different periods to the succession of Omicron variants. Periods of complete clearance of the previous variants are marked with higher R_t ratios, as for the R_t between BA.2*/BA.1* and XBB/BA.5* (**Fig. 3.**). Conversely, we see periods of coexistence of more than one variant have smaller R_t ratios, e.g., the ratios between BA.4*/BA.2*, BA.5*/BA.2* and BA.5*/BA.4*. The median R_t values of BA.2*, across the US, were almost 20% higher than the R_t values of BA.1*, and to XBB* distribution of R_t values it was more than 20% bigger than the R_t values of BA.5*. In summary, variants with comparable higher R_t (BA.2* and XBB*) values to their predecessor, can completely invade the dominant variants. See **Fig. S6.** for the R_t ratios for all states.

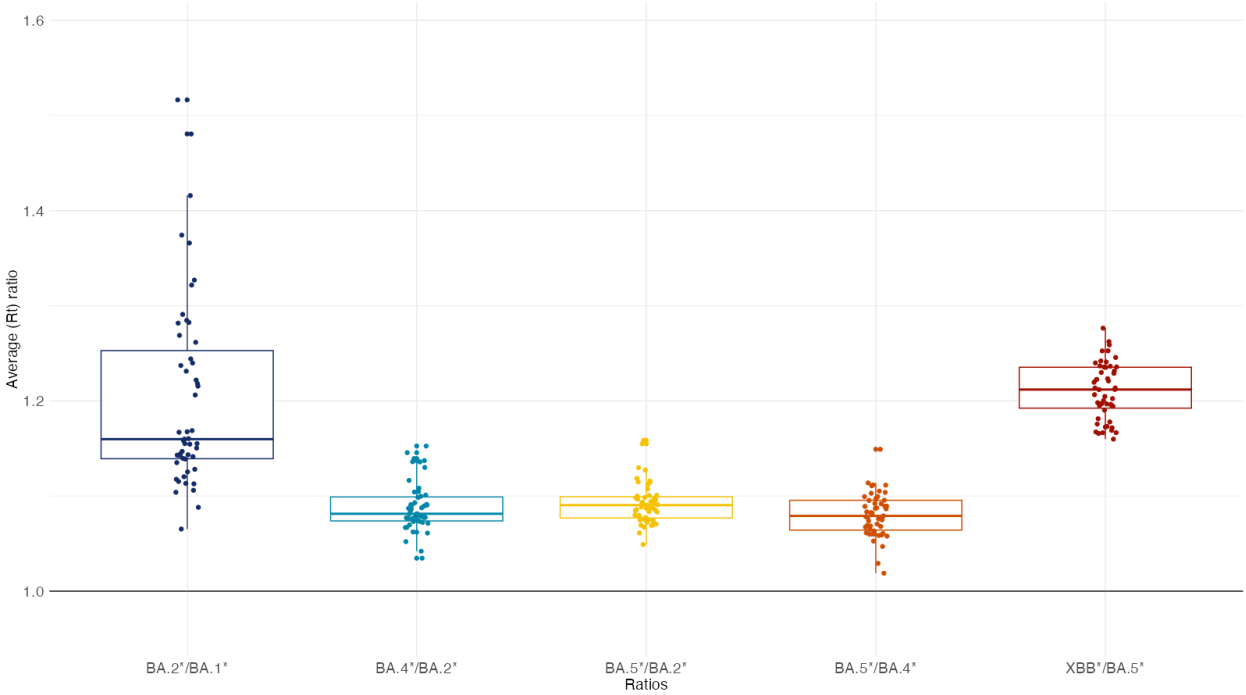


Fig. S7. The ratio between variant-specific R_t boxplot and dots to the state-specific ratios. Dots are the state-level R_t ratio and the boxplot is the distribution over all the states. The pairs of variants are chosen as the succeeding history of variants throughout 2022. After the BA.1* wave and before the XBB* Introduction to the US, the R_t ratios are pretty similar, which was a period of coexistence of variants. The BA.2*/BA.1* and BA.5*/XBB* are significantly higher, and mark the complete clearance of the previous variant, respectively BA.1* and XBB*.

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table S1. Categorization of Pango lineages and sublineages alias

Category	'Omicron BA.1*'	'Omicron BA.2*'	'Omicron BA.3*'	'Omicron BA.4*'	'Omicron BA.5*'	'Omicron XBB*'	'Recombinant'	'Other'
Pango lineage	BA.1 or B.1.1.529.1	BA.2 or B.1.1.529.2	BA.3 or B.1.1529.3	BA.4 or B.1.1.529.4	BA.5 or B.1.1.529.5	XBB	X, excluding XBB	Any other
Sublineages alias	BD.1	B["G" "H" "J" "L" "M" "N" "R" "S" "Y"] or C["A" "B" "H" "J" "M" "V"] or D["D" "S" "V"] or E["J" "P"] or F["J"]	No alias for sublineages	C["S"] or D["C"]	B["E" "F" "K" "Q" "T" "U" "V" "W" "Z"] or C["C" "D" "E" "F" "G" "K" "L" "N" "P" "Q" "R" "T" "U" "W" "Y" "Z"] or D["A" "B" "E" "F" "G" "H" "J" "K" "L" "M" "N" "P" "Q" "R" "T" "U" "W" "Y" "Z"] or E["A" "C" "D" "E" "F" "H" "N" "Q" "R" "S" "T" "V" "W" "Y" "Z"] or F["A" "B" "C" "F"]	No alias for sublineages	No alias for sublineages	No alias for sublineages

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table S2. Peak of infections by variant categories

	Omicron BA.1*	Omicron BA.2*	Omicron BA.4*	Omicron BA.5*	Omicron XBB*
USA	4,204,319	625,656	140,772	799,728	303,653
State					
Alabama	80,699	6,097	4,461	14,819	5,983
Alaska	9,027	1,904	430	2,717	790
Arizona	103,970	13,127	2,523	19,627	5,975
Arkansas	49,773	2,257	1,106	8,238	1,715
California	548,049	87,067	16,727	113,263	35,291
Colorado	75,425	11,865	2,127	12,888	5,004
Connecticut	44,145	10,623	1,420	6,548	6,842
Delaware	13,389	2,883	414	2,136	1,211
District of Columbia	10,903	1,518	323	1,510	1,065
Florida	318,587	67,102	17,295	79,468	26,388
Georgia	159,257	14,898	6,119	28,439	8,480
Hawaii	18,682	6,339	499	3,713	1,600
Idaho	17,138	2,344	952	4,127	1,873
Illinois	140,808	29,216	5,329	31,065	14,913
Indiana	81,697	9,031	2,489	16,823	5,317
Iowa	30,717	4,489	1,356	7,029	3,755
Kansas	37,585	3,332	2,005	5,739	2,845
Kentucky	58,792	8,449	2,282	16,826	4,674
Louisiana	75,108	8,542	3,172	13,962	5,042
Maine	10,069	5,770	526	3,478	3,065
Maryland	83,371	14,839	2,258	12,542	7,509
Massachusetts	82,828	22,513	2,456	12,387	11,933
Michigan	114,189	28,113	2,793	23,520	11,321
Minnesota	62,361	14,423	2,987	12,736	6,290
Mississippi	43,877	2,782	2,699	8,180	1,981
Missouri	78,656	9,588	4,167	13,998	6,102
Montana	12,300	2,195	850	2,451	1,314
Nebraska	19,585	2,689	1,087	5,505	3,251
Nevada	46,073	6,622	1,192	10,153	2,259
New Hampshire	14,377	4,075	600	2,807	1,727

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

New Jersey	128,193	23,330	3,824	19,382	18,473
New Mexico	25,694	4,354	860	5,324	1,984
New York	281,267	56,751	7,882	39,921	30,444
North Carolina	143,645	21,360	6,347	27,913	12,096
North Dakota	10,915	1,168	277	1,896	682
Ohio	143,117	18,910	4,482	27,233	13,252
Oklahoma	64,238	3,899	1,560	10,404	3,910
Oregon	46,662	11,592	2,075	11,071	3,625
Pennsylvania	141,384	25,348	3,860	24,278	16,019
Rhode Island	16,010	3,321	359	2,102	2,315
South Carolina	85,567	7,491	3,939	13,580	4,193
South Dakota	11,348	1,062	392	1,913	1,274
Tennessee	97,156	8,163	3,680	19,420	6,692
Texas	422,616	35,932	14,956	80,743	25,776
Utah	41,564	5,510	1,004	6,323	2,295
Vermont	6,486	2,543	276	1,109	709
Virginia	103,246	18,160	4,652	19,349	10,278
Washington	89,584	18,463	2,486	17,473	6,482
West Virginia	20,050	3,665	1,173	5,200	1,805
Wisconsin	72,726	12,375	2,153	11,674	5,731
Wyoming	7,480	1,174	649	1,823	582

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table S3. Total of infections by variant categories

	Omicron BA.1*	Omicron BA.2*	Omicron BA.4*	Omicron BA.5*	Omicron XBB*	Total
USA	161,762,895	52,112,475	14,179,483	119,878,658	28,635,157	376,568,668
State						
Alabama	2,772,841	502,481	301,070	1,840,577	364,149	5,781,118
Alaska	331,852	189,315	36,086	286,095	45,496	888,844
Arizona	3,997,666	1,055,156	210,600	2,319,194	476,758	8,059,374
Arkansas	1,634,641	226,639	95,847	1,063,683	102,172	3,122,982
California	20,332,095	6,943,130	1,359,685	16,185,013	3,108,330	47,928,253
Colorado	2,648,940	782,758	178,813	1,702,499	369,914	5,682,924
Connecticut	1,522,891	710,982	149,978	1,162,133	518,107	4,064,091
Delaware	461,352	185,235	38,615	295,717	92,789	1,073,708
District of Columbia	322,927	118,000	30,969	216,461	80,392	768,749
Florida	10,901,127	4,667,189	1,296,781	9,182,990	2,386,455	28,434,542
Georgia	5,760,126	1,192,386	528,765	3,295,697	684,121	11,461,095
Hawaii	603,820	419,328	39,099	487,560	90,698	1,640,505
Idaho	751,594	178,263	52,393	540,414	122,580	1,645,244
Illinois	5,468,759	2,247,330	532,089	5,155,806	1,211,730	14,615,714
Indiana	3,075,722	777,050	233,370	2,254,682	391,892	6,732,716
Iowa	1,208,498	348,540	112,319	989,727	212,310	2,871,394
Kansas	1,368,857	265,362	150,426	865,498	191,838	2,841,981
Kentucky	2,555,786	635,456	231,573	2,140,746	338,583	5,902,144
Louisiana	2,621,453	586,838	267,115	1,882,854	362,779	5,721,039
Maine	617,251	346,511	49,827	501,764	185,229	1,700,582
Maryland	2,718,440	1,001,262	252,671	1,976,392	592,713	6,541,478
Massachusetts	2,936,602	1,481,011	294,630	2,083,316	1,044,955	7,840,514
Michigan	4,717,583	1,900,200	378,516	4,010,475	779,358	11,786,132
Minnesota	2,481,968	1,008,431	289,711	2,256,693	446,887	6,483,690
Mississippi	1,495,347	216,336	187,023	1,069,372	167,479	3,135,557
Missouri	2,909,771	753,251	375,551	2,440,057	449,121	6,927,751
Montana	532,094	143,079	33,344	387,768	81,431	1,177,716
Nebraska	819,490	210,089	92,888	728,462	175,299	2,026,228
Nevada	1,570,551	470,852	70,311	1,041,307	183,352	3,336,373
New Hampshire	609,876	269,619	47,485	415,388	157,023	1,499,391
New Jersey	4,032,876	1,819,971	398,646	3,214,700	1,300,745	10,766,938

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

New Mexico	957,983	288,810	62,929	778,920	111,065	2,199,707
New York	9,414,502	4,117,501	861,538	6,893,665	2,667,404	23,954,610
North Carolina	5,558,845	1,618,409	707,725	4,080,326	1,000,189	12,965,494
North Dakota	361,065	79,590	19,952	279,255	43,128	782,990
Ohio	5,512,264	1,445,588	454,099	4,360,627	912,709	12,685,287
Oklahoma	2,149,327	308,975	135,519	1,399,006	251,413	4,244,240
Oregon	1,889,525	897,735	157,995	1,683,766	255,883	4,884,904
Pennsylvania	5,595,975	1,929,311	488,119	4,062,804	1,332,612	13,408,821
Rhode Island	505,780	206,988	39,350	333,036	165,326	1,250,480
South Carolina	2,854,649	556,436	307,723	1,771,568	320,376	5,810,752
South Dakota	335,727	56,862	28,966	240,432	66,014	728,001
Tennessee	3,659,591	645,005	281,411	2,561,121	513,954	7,661,082
Texas	15,937,226	3,541,687	1,172,640	10,249,347	2,293,542	33,194,442
Utah	1,565,741	455,929	75,906	857,997	117,272	3,072,845
Vermont	276,505	152,197	14,149	151,874	46,990	641,715
Virginia	3,984,707	1,404,674	507,862	2,845,993	770,285	9,513,521
Washington	3,485,480	1,557,749	208,230	2,438,627	452,457	8,142,543
West Virginia	908,103	253,293	94,442	681,815	136,347	2,074,000
Wisconsin	2,745,629	871,016	225,331	2,008,904	425,997	6,276,877
Wyoming	281,475	72,670	19,401	206,535	37,509	617,590

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table S4. Attack rates by variant categories

	Omicron BA.1*	Omicron BA.2*	Omicron BA.4*	Omicron BA.5*	Omicron XBB*
USA	47.6%	15.3%	4.2%	35.3%	8.4%
States					
Alabama	56.6%	10.2%	6.1%	37.5%	7.4%
Alaska	45.4%	25.9%	4.9%	39.1%	6.2%
Arizona	54.9%	14.5%	2.9%	31.9%	6.6%
Arkansas	54.2%	7.5%	3.2%	35.2%	3.4%
California	51.5%	17.6%	3.4%	41.0%	7.9%
Colorado	46.0%	13.6%	3.1%	29.6%	6.4%
Connecticut	42.7%	19.9%	4.2%	32.6%	14.5%
Delaware	47.4%	19.0%	4.0%	30.4%	9.5%
District of Columbia	45.8%	16.7%	4.4%	30.7%	11.4%
Florida	50.8%	21.7%	6.0%	42.8%	11.1%
Georgia	54.3%	11.2%	5.0%	31.0%	6.4%
Hawaii	42.6%	29.6%	2.8%	34.4%	6.4%
Idaho	42.1%	10.0%	2.9%	30.2%	6.9%
Illinois	43.2%	17.7%	4.2%	40.7%	9.6%
Indiana	45.7%	11.5%	3.5%	33.5%	5.8%
Iowa	38.3%	11.0%	3.6%	31.4%	6.7%
Kansas	47.0%	9.1%	5.2%	29.7%	6.6%
Kentucky	57.2%	14.2%	5.2%	47.9%	7.6%
Louisiana	56.4%	12.6%	5.7%	40.5%	7.8%
Maine	45.9%	25.8%	3.7%	37.3%	13.8%
Maryland	45.0%	16.6%	4.2%	32.7%	9.8%
Massachusetts	42.6%	21.5%	4.3%	30.2%	15.2%
Michigan	47.2%	19.0%	3.8%	40.2%	7.8%
Minnesota	44.0%	17.9%	5.1%	40.0%	7.9%
Mississippi	50.2%	7.3%	6.3%	35.9%	5.6%
Missouri	47.4%	12.3%	6.1%	39.8%	7.3%
Montana	49.8%	13.4%	3.1%	36.3%	7.6%
Nebraska	42.4%	10.9%	4.8%	37.7%	9.1%
Nevada	51.0%	15.3%	2.3%	33.8%	6.0%
New Hampshire	44.9%	19.8%	3.5%	30.5%	11.5%
New Jersey	45.4%	20.5%	4.5%	36.2%	14.6%
New Mexico	45.7%	13.8%	3.0%	37.1%	5.3%

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

New York	48.4%	21.2%	4.4%	35.4%	13.7%
North Carolina	53.0%	15.4%	6.7%	38.9%	9.5%
North Dakota	47.4%	10.4%	2.6%	36.6%	5.7%
Ohio	47.2%	12.4%	3.9%	37.3%	7.8%
Oklahoma	54.3%	7.8%	3.4%	35.4%	6.4%
Oregon	44.8%	21.3%	3.7%	39.9%	6.1%
Pennsylvania	43.7%	15.1%	3.8%	31.7%	10.4%
Rhode Island	47.7%	19.5%	3.7%	31.4%	15.6%
South Carolina	55.4%	10.8%	6.0%	34.4%	6.2%
South Dakota	37.9%	6.4%	3.3%	27.2%	7.5%
Tennessee	53.6%	9.4%	4.1%	37.5%	7.5%
Texas	55.0%	12.2%	4.0%	35.3%	7.9%
Utah	48.8%	14.2%	2.4%	26.8%	3.7%
Vermont	44.3%	24.4%	2.3%	24.3%	7.5%
Virginia	46.7%	16.5%	5.9%	33.3%	9.0%
Washington	45.8%	20.5%	2.7%	32.0%	5.9%
West Virginia	50.7%	14.1%	5.3%	38.0%	7.6%
Wisconsin	47.2%	15.0%	3.9%	34.5%	7.3%
Wyoming	48.6%	12.6%	3.4%	35.7%	6.5%

medRxiv preprint doi: https://doi.org/10.1101/2023.11.07.23298178; this version posted November 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table S5. Interval values to the R_t estimates by variant categories

	Variant Categories				
	Omicron BA.1*	**Omicron BA.2***	**Omicron BA.4***	**Omicron BA.5***	**Omicron XBB***
median	1.01 (4.23, 0.36)	1.03 (3.94, 0.33)	1.03 (2.59, 0.38)	1.04 (2.94, 0.41)	1.11 (2.35, 0.48)
upper	1.20 (7.74, 0.54)	1.12 (6.20, 0.56)	1.13 (3.97, 0.56)	1.12 (4.79, 0.56)	1.21 (3.68, 0.66)
lower	0.85 (2.13, 0.22)	0.95 (2.00, 0.18)	0.94 (1.65, 0.24)	0.97 (1.86, 0.29)	1.02 (1.56, 0.33)