Title: Escalating combinations of enhanced infectivity and immune escape define SARS-CoV-2 Omicron lineage replacement

Authors: Nicholas F.G. Chen[1], Kien Pham[1], Chrispin Chaguza[1,2], Rafael Lopes[1,2], Fayette Klaassen[3], Daniel M. Weinberger[1,2], Virginia E. Pitzer[1,2], Joshua L. Warren[2,4], Nathan D. Grubaugh[1,2,5*], Anne M. Hahn[1*,#]

[1] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, United States,

[2] Public Health Modeling Unit, Yale School of Public Health, New Haven, United States,

[3] Department of Global Health & Population, Harvard T. H. Chan School of Public Health, Boston, United States,

[4] Department of Biostatistics, Yale School of Public Health, New Haven, United States

[5] Department of Ecology and Evolutionary Biology, Yale University, New Haven, United States

* These authors contributed equally to this work.

[#] corresponding author: anne.hahn@yale.edu

Abstract: In 2022, consecutive sweeps of the highly transmissible SARS-CoV-2 Omicron-family maintained high viral transmission levels despite extensive antigen exposure on the population level resulting from both vaccinations and infections. To better understand variant fitness in the context of the highly dynamic immunity landscape of 2022, we aimed to dissect the interplay between immunity and fitness advantages of emerging SARS-CoV-2 Omicron lineages on the population-level. We evaluated the relative contribution of higher intrinsic transmissibility or immune escape on the fitness of emerging lineages by analyzing data collected through our local genomic surveillance program from Connecticut, USA. We compared growth rates, estimated infections, effective reproductive rates, average viral copy numbers, and likelihood for causing vaccine break-through infections.

Using these population-level data, we find that newly emerging Omicron lineages reach dominance through a specific combination of enhanced intrinsic transmissibility and immune escape that varies over time depending on the state of the host-population. Using similar frameworks that integrate whole genome sequencing together with clinical, laboratory, and epidemiological data can advance our knowledge on host-pathogen dynamics in the post-emergence phase that can be applied to other communicable diseases beyond SARS-CoV-2.

Main Text

Introduction

Capacities for genomic surveillance saw an unprecedented rise during the COVID-19 pandemic (Brito et al. 2022). The ability to closely monitor SARS-CoV-2 populations in real-time was crucial to discover and track SARS-CoV-2 variants of concern (VOC). The first VOC, Alpha (B.1.1.7) (Hill et al. 2022; Rambaut et al. 2020), emerged towards the end of 2020 and was followed in 2021 by Beta (B.1.351) (Tegally et al. 2021), Gamma (P.1) (Faria et al. 2021), and Delta (B.1.617) (WHO 2021). In late 2021, the highly divergent Omicron (B.1.1.529) lineage (Viana et al. 2022) rapidly displaced Delta globally.

In 2022, SARS-CoV-2 transmission remained, somewhat counterintuitively, high all year, which stands in contrast to the summer-lows observed during 2020 and 2021. This was characterized by consecutive sweeps of multiple Omicron lineages. BA.1 (B.1.1.529.1), the first member of the Omicron family to emerge, shared a common ancestor with the D614G background in late 2020. This ancestor diverged into at least five main sublineages (BA.1-BA.5) (Tegally et al. 2022; Viana et al. 2022). The exact origin and nature of the emergence of these Omicron lineages remains unclear; however, it is likely that these lineages emerged from a common source, likely undergoing selective pressure for an extended amount of time (Markov et al. 2023). A subset of the original Omicron viral population further spread in one or multiple transmission events to a larger community and eventually worldwide (Tegally et al. 2023).

Thus, 2022 saw a unique combination of this highly divergent virus variant facing a host population that, through intensive vaccination campaigns and vaccine roll-out in many parts of the world, had built up significant amounts of immunity towards the original Wuhan-Hu-1 Spike antigen. In this study, we aimed to reconstruct key epidemiological and evolutionary features that led to the particular lineage replacement dynamics of Omicron lineages as exemplified in a well-monitored study setting in Connecticut, United States (US). For this, we drew upon sample data collected from the Yale New

Haven Hospital (YNHH) through our Yale Genomic Surveillance Initiative from January 2022 to January 2023. Whole genome sequencing data together with relevant laboratory and patient metadata were available for approximately 5-10% of total reported cases across the state and the entire length of the study period. To characterize competitive growth advantages of variants during their emergence periods and evaluate their fitness compared to their immediate predecessors, we calculated the growth rates of each incoming Omicron lineage, compared their virus copy number from nasal swabs, and calculated the risk for vaccine breakthrough infections as a metric for immune escape. Doing so, we reveal a high capacity of Omicron lineages to explore different fitness niches through a combination of enhanced transmissibility and/or immune escape depending on the history of antigenic exposure of the host-population.

The synthesis of high-resolution epidemiological, genomic, and immunological data allows us to assess and compare transmissibility and immune escape of each Omicron lineage. Exploring lineage replacement dynamics on a population level allows us to evaluate predictions based on *in-vitro* data and thus evaluate key parameters defining future variant emergence. Doing so, we reveal a high capacity of Omicron lineages to explore different fitness niches through a combination of enhanced transmissibility and/or immune escape depending on the history of antigenic exposure of the host-population. Our study highlights the ability to derive key aspects of pathogen lineage fitness by analyzing data that can be drawn from genomic surveillance efforts with relatively simple technical requirements. Such frameworks will be particularly relevant for further monitoring SARS-CoV-2 in the post-pandemic phase and are easily transferable to other pathogens.

Results

Continuous Omicron lineages replacement causes high levels of community transmission throughout 2022

Close monitoring of the SARS-CoV-2 population composition added valuable insights on how emerging variants were impacting COVID-19 case dynamics. The first VOCs detected in Connecticut were Alpha in late 2020 and Delta around mid-2021 (**Fig 1A**, based on data from GISAID for Connecticut). Looking at the epidemic curve in Connecticut (**Fig 1B**), the introduction of both the Alpha and Delta VOCs were followed by a surge in reported SARS-CoV-2 cases in December 2020 and Spring 2021, respectively. After a summer of relatively low transmission in 2021, cases started to rise again towards the end of the year with a second Delta-dominated peak in December 2021.

Globally, the first Omicron sample was identified in November 2021 and shortly after also detected in the US and Connecticut (**Table 1**). BA.1 is an outgroup to the BA.2/BA.4/BA.5 cluster, and XBB a recombinant of two second generation BA.2 lineages (BA.2.10 & BA.2.75) that likely arose in Southeast-Asia around late summer 2022 (Goh et al. 2023) (**Supplementary Fig 1**). Notably, whereas BA.1-BA.5 most likely originated from a common source (Tegally et al. 2021, 2022; Viana et al. 2022), XBB derived from lineages that emerged through sustained transmission chains in 2022. After BA.1's emergence in Connecticut in November 2021, the rapid replacement of Delta was accompanied by a massive increase in reported cases. Then, BA.2 swiftly replaced BA.1 in March 2022. BA.4 and BA.5 were both introduced to Connecticut in May 2022 and co-circulated for eight months until December 2022 (**Table 1**). Lastly, XBB-based variants outcompeted BA.5 lineages towards the end of 2022. Interestingly, the New England region including our study site was the first where XBB.1.5 was widely circulated within the US (Ma et al. 2023).

Table 1 Major Omicron lineages in Connecticut January 2022 - January 2023, info taken from GISAID

Lineage	**First detected globally**	**First detected in CT**	**Major sub-lineages in Connecticut (+key amino acid changes in Spike)**
BA.1	October 2021	November 2021	BA.1.1 (+R346K)
BA.2	November 2021	January 2022	BA.2.12.1 (+L452Q,S704L)
BA.4	November 2021	May 2022	BA.4.6 (+R346T, N658S)
BA.5	November 2021	May 2022	BQ.1.1 (+R346T, K444T,N460K) BF.7 (+R346T)
XBB (BA.2.10*x BA.2.75*)	September 2022	October 2022	XBB.1.5 (+F486P)

To understand the factors that influenced the consecutive sweeps of Omicron lineages, we first characterized the infection dynamics broken up for each lineage with respect to case counts across our one-year study period. As the northern hemisphere entered into the summer months of 2022, most pandemic response mechanisms and non-pharmaceutical measures were discontinued. Consequently, case reporting is expected to be less reliable in accurately mirroring true infection dynamics compared to earlier in the pandemic. Therefore, we retrieved data from covidestim, a Bayesian nowcasting model incorporating publicly available time series of COVID-19 case notifications, hospitalizations, and deaths, accounting for effective population immunity to estimate infections (Chitwood et al. 2022). Together with the frequencies of the major Omicron lineages derived from state-wide genomic surveillance data, we partitioned the overall estimated infections to calculate variant-specific infection estimates cumulatively and over time.

By analyzing the variant-specific estimated infections generated through covidestim, we are further able to characterize dynamics of each lineage that may not be reflected in reported cases **(Fig 1C).** Using this framework, we estimated that the BA.1 lineage caused approximately 1.5 million infections in

Connecticut, USA (**Fig 1D**), representing around 44% (Credible Interval (CrI) 27%-65%) of the state's population. Interestingly, the BA.2, BA.4, and BA.5 lineages caused over 2.1 million combined infections (CrI 1.2-3.5 million) within only nine months after this surge (**Fig 1D**). Another surge driven by the XBB lineage and its descendants caused an estimated 0.4 million infections (CrI 0.2-0.7 million) up until January 2023.

Taken together, we reconstructed in detail the SARS-CoV-2 infection dynamics caused by Omicron sublineages in 2022 by drawing upon data from genomic surveillance and epidemiological modeling. We showed the timing and scale of each of the genetically closely related Omicron lineages differed substantially and we thus aimed to further characterize each of the emergence periods.

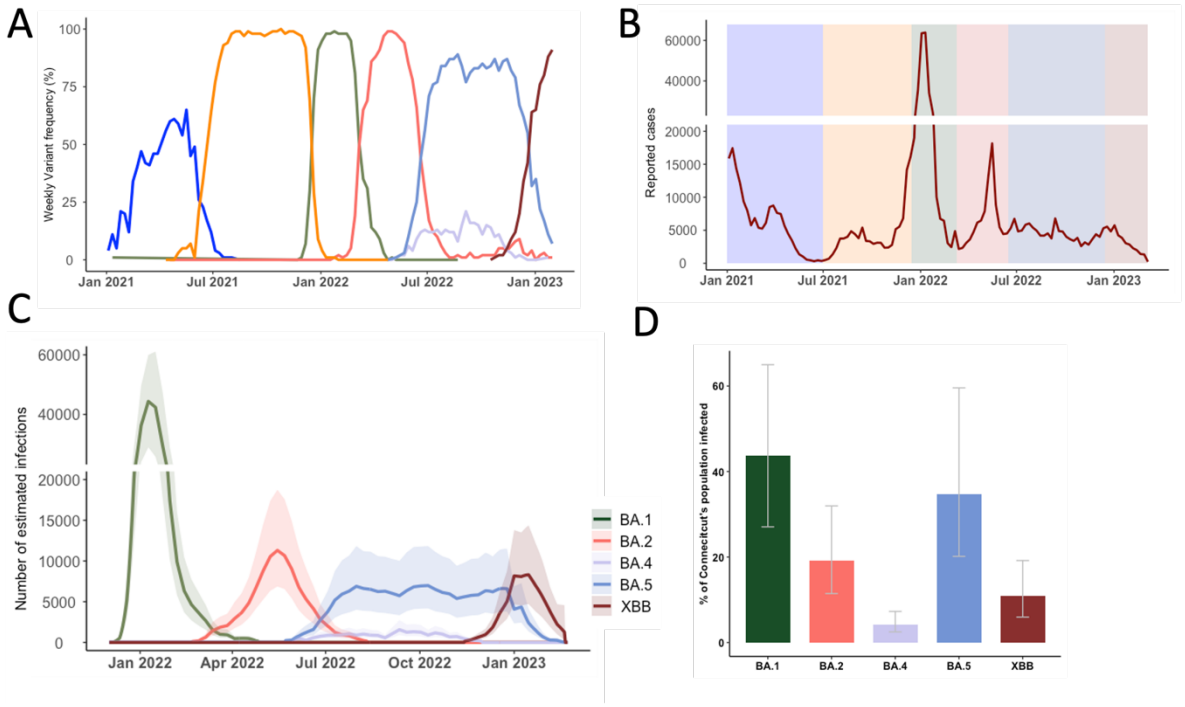


Fig 1. Genomic Epidemiology of SARS-CoV-2 describes the COVID-19 epidemic in Connecticut, US. (A) Frequencies of SARS-CoV-2 variants of concern from January 2021 to January 2023, based on sequences deposited on GISAID. **(B)** Reported COVID-19 infections from January 2021 to 2023 with highlights of the dominant periods of different Omicron lineages, data from the Connecticut Department of Public Health **(C)** Number of estimated infections based on the covidestim model for 2022 where testing and reporting widely changed after the first wave of Omicron BA.1 **(D)** Cumulative estimated cases for each Omicron variant up until January 2023.

Growth advantage of emerging Omicron lineages shrinks towards the end of 2022

Having established that each of these variants exhibits distinct epidemic dynamics in the population, we aimed to further explore the specific features of each of the variants and how these contributed to this emergence pattern. For this, we first determined the rate of lineage replacement by calculating the time it took for each lineage to increase from 5% to 50% of all reported daily cases in Connecticut (Fig 2A). This range of 5% to 50% was chosen to capture the period of time it took for a variant to become reliably established in the population, while avoiding some of the inherent stochasticity associated with initial cases, to when it accounted for the majority of cases and is thus dominant in the population. We then fitted a logistic regression model to the observed frequencies and compared their slope coefficients **(Fig 2B)**.

Doing so, we show that BA.1 had the fastest growth rate, reaching a frequency of 50% in only 10 days from being detected at 5% **(Fig 2A)**. For each of the subsequent lineages, lineage replacement slowed down, ranging from 20 to 59 days for BA.2 and XBB, respectively.

We further compared the emergence of each of these lineages using variant-specific estimated infections from the covidestim model. For this, we partitioned the covidestim weekly estimated COVID-19 infections in Connecticut based on state-wide variant frequency to calculate variant-specific reproduction numbers (R_t) over the course of 2022 (**Fig 2C**). We show that although BA.1 had the highest initial R_t value (2.4, CrI 1.6-3.5), it quickly fell below 1 in early January. BA.2 emerged in February 2022 with an initial R_t of 1.7 (CrI 1.3-2.2), followed by a notable decrease and a plateau at around 1.2 for several weeks before falling below 1 around April 2022. The initial R_t of BA.5 (1.5 CrI 1.2-1.86) and BA.4 (1.2 CrI 1.07-1.36) were overall lower than for BA.1 and BA.2. However, BA.4 and BA.5 both hovered at $R_t \sim 1$ for several months during summer 2022. BA.4 was finally outcompeted by late-stage BA.5 lineages. The R_t for BA.5 fell below 1 once XBB was introduced with an R_t of 1.4 (CrI 1.1-1.7). The R_t of XBB fell below 1 around the beginning of 2023.

Next, we sought to understand the relative transmission advantages during the emergence periods by comparing R_t values of co-circulating lineages (**Fig 2D**). We first identified 14-day windows in which one lineage replaced the previously dominating lineage (**Supplementary Table 1**). Using a method previously developed (Earnest et al. 2022; Petrone et al. 2022), we identified six, 14-day periods of variant overlap and divided the R_t value of the emerging lineage by the R_t value of the previously dominating lineage for each day of the emergence period, yielding 14 R_t ratios. The median of these values results in a comparative fitness measure (**Fig 2D**). We show that BA.1 and BA.2 were 78.4% and 76% more transmissible than Delta and BA.1, respectively (**Fig 2D**). The advantage of both BA.4 and BA.5 over BA.2 was significantly lower at 27.9% and 48.5%, respectively. While BA.5 had a slight fitness advantage of 10% over BA.4 during the emergence of both lineages, this advantage was not sufficient to fully outcompete BA.4 for several months. However, late-stage BA.5 sub-lineages such as BQ.1.1 and BF.7 (**Table 1**) were able to push out BA.4 before XBB emerged by late November. XBB replaced these late-stage BA.5 sequences with a fitness advantage of 28.5%.

By characterizing the growth rates during the emergence windows for the major Omicron lineages, we have shown high variance in lineage replacement dynamics. The lack of a trend in our R_t ratio analysis suggests the fitness advantages of incoming lineages are likely being influenced by several non-uniform factors that vary for different emergence windows. We next sought to further dissect these observations through exploring possible underlying drivers of the incoming lineages' fitness advantages.

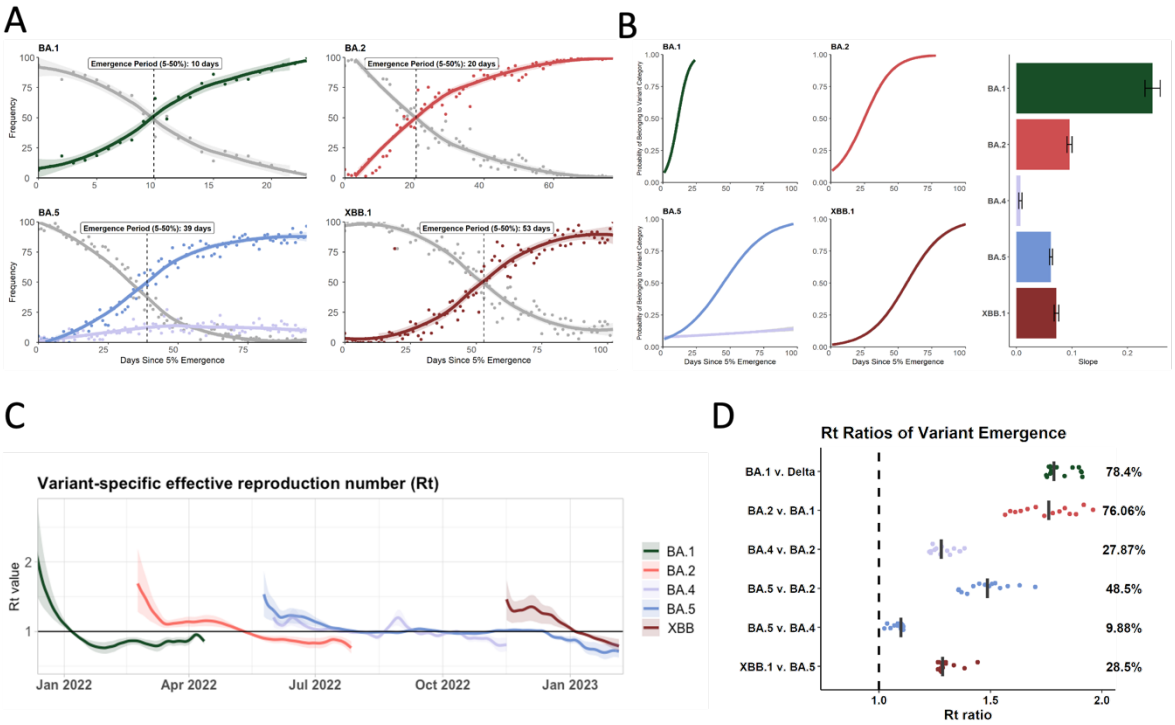


Fig 2. Comparison of key epidemiological parameters for each Omicron lineage in 2022. (A) Frequencies of the incoming Omicron lineage during its emergence period highlight the time interval between being detected at 5% and 50% of all sequences (based on GISAID data for Connecticut). **(B)** Comparison of growth rates between the different Omicron lineages and the average slope of the growth curve during emergence periods **(C)** Variant-specific R_t numbers for each of the Omicron lineage over time derived from the modeled overall R_t from the covidestim model **(D)** Based on (C), the R_t ratios for each of the emergence periods were calculated to estimate the average advantage of each incoming lineage compared to the previously dominating one.

Average inter- and intra-lineage viral copy numbers vary over time and only partially explain lineage replacement patterns

To explain the specific mechanisms by which variants gain advantages over another, we evaluated the fitness advantage of each variant as a combination of intrinsic transmissibility and immune escape. For the purpose of this study, we used proxies derived from data collected through our genomic surveillance program for each of these factors.

As a proxy for intrinsic transmissibility, we measured virus copies in nasopharyngeal swab material using RT-qPCR. We used the same assay (Vogels et al. 2021) for each sample processed for sequencing which enables us to compare average cycle threshold (Ct) values per variant across a total of 11,111 samples (**Fig 3A**). We also converted the RT-qPCR values to genome equivalents (**Supplementary figure 3**).

Using cross-sectional data has caveats and is sensitive to several epidemiological and viral factors (Fryer et al. 2023; Hay et al. 2021). Changes in tissue tropism, symptom severity, and test-seeking behavior are expected to differ for Omicron infections compared to infections with pre-Omicron lineages. In the following, we are thus focusing on Omicron-lineage samples only.

For this, we analyzed 6,856 samples identified as Omicron-lineages (**Fig 3B**) collected from a wide range of individuals and disease statuses, including asymptomatic individuals detected through baseline surveillance of outpatients as well as inpatients and emergency department visits. We were particularly interested in Ct-value trends over time and plotted the raw data with temporal resolution (**Fig 3C**). We then tested the data for the assumptions of linearity and fit linear regressions **(Fig 3D)** to derive coefficients as a measure of Ct-value change over time for each lineage (**Table 2**).

Table 2 Trends in median Ct values of the major Omicron lineages over time

Lineage	Median Ct early samples	Median Ct late samples	**Regression Coefficient (Ct / day) (p-value)**
BA.1	27.297 (IQR: 24.179 - 30.918)	27.799 (IQR: 26.139 - 28.760)	**+0.016 (0.045)**
BA.1.1	25.637 (IQR: 22.8- 28.15)	28.399 (IQR: 25.563, 30-184)	**+0.016 (0.011)**
BA.2	26.681, IQR: 22.557 - 29.128	22.544 (IQR: 20.304, 25.718)	**-0.015 (0.002)**
BA.2.12.1	22.972 (IQR: 22.928 - 22.540)	23.335 (IQR: 22.928 - 22.540)	**-0.021 (<0.001)**
BA.4	21.2 (IQR: 18.944 - 23.75)	19.379 (IQR: 19.088 - 21.132)	**+0.022 (0.227)**
BA.5	23.077 (IQR: 19.61 - 25.606)	25.131 (IQR: 22.750 - 28.133)	**+0.015 (<0.001)**
BQ.1.1	23.231 (IQR: 21.580 - 24.561)	25.828 (IQR: 23.446 - 29.863)	**+0.005 (0.758)**
XBB	21.794 (IQR: 20.573 - 26.42)	24.753 (IQR: 19.846 - 27.566)	**+0.03 (0.024)**

Looking at trends over time, we noticed that the average Ct value for BA.1, BA.4, BA.5, and XBB samples increased over time (decreasing viral copy numbers) towards the end of each wave with a rate around 0.015-0.03 Ct values/day (**Table 2**). However, for the BA.2 and BA.2.12.1 samples, we observed a decreasing trend towards the end of the BA.2 wave with a decrease of 0.015 and 0.021 Ct values/day.

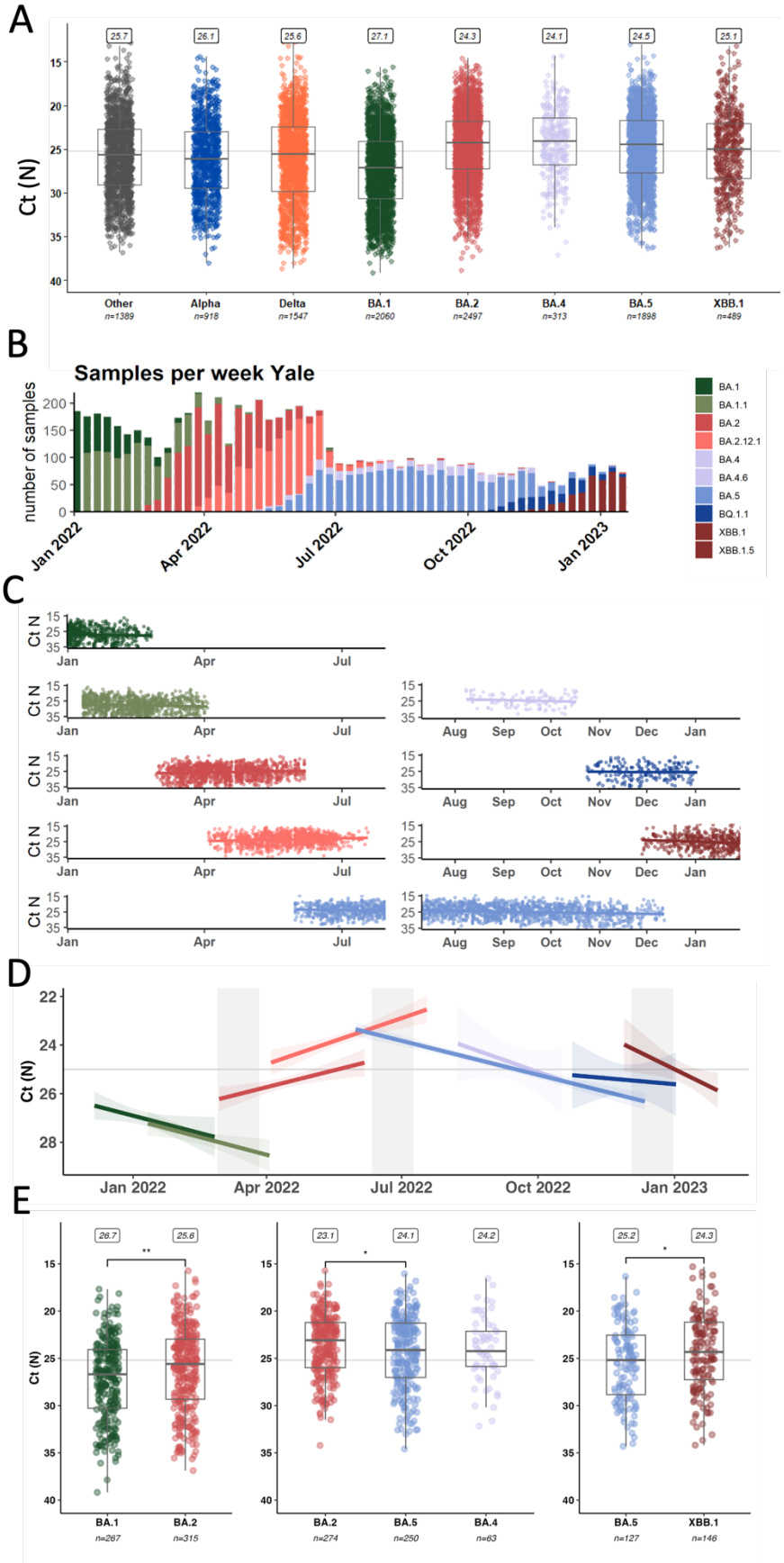
Next, we compared only samples that were collected specifically during the emergence period of an incoming lineage, defined as 14 days on either side of the date where the previously dominant and the incoming variant each account for 50% of the samples (**Supplement Table 2**).

Early BA.2 samples had significantly lower Ct values (median 25.6, Interquartile Range (IQR) 23.0-29.3) compared to late BA.1 samples (median 26.7, IQR 24.1-30.3) (p-value: 0.003) (**Fig 3E**). This supports previous reports from two European surveillance programs that also showed higher viral copies

of BA.2 over BA.1 as measured by qPCR (Lentini et al. 2022; Musalkova et al. 2023). When looking at the emergence period of BA.4 and BA.5 compared to late-stage BA.2 samples, samples infected with the incoming variant had significantly higher Ct values than BA.2 samples (BA.4 median 24.2, IQR 22.1-25,9; BA.5 median 24.1, IQR 21.3-27.0, BA.2 median 23.1, IQR 21.2-26.0) from that same period (p-value: 0.034). There was no statistically significant difference in the Ct values between BA.4 and BA.5 samples (p-value: 0.829). Finally, early XBB samples again had lower Ct values (median 24.3, IQR 21.2-27.3) compared to late-stage BA.5 samples (including BQ.1.1) (median 25.2, IQR 22.5-28.8) (p-value: 0.033).

We also sought to investigate whether there was any association of patient metadata with Ct value in these first XBB cases by stratifying the samples according to age, sex, vaccination status, and patient class (**Supplementary Figure 4**). There was no significant difference in Ct values based on vaccination status. However, we found that people admitted to the emergency department had, on average, lower Ct values compared to those tested as inpatients or outpatients.

In summary, using an RT-qPCR assay for assessing viral copy number for each sample we submitted to whole-genome sequencing, we revealed striking temporal patterns. A higher viral copy number was not necessarily decisive for a lineage's fitness advantage as, for example, incoming BA.5 samples had lower viral copy numbers than late BA.2 samples. Together, this hints at a varying advantage conferred by enhanced intrinsic transmissibility, as measured by viral copy numbers, that changed over the course of the year and seems mostly linked to BA.2-derived lineages.



201

Figure 3 Comparing average qPCR Ct values as a proxy for variant intrinsic transmissibility. (A) Summary of all samples processed in our genomic surveillance programme from 2021 to January 2023 **(B)** Overview about the numbers and distribution of Omicron samples **(C)** Ct values from Omicron samples plotted together with modeled average over time **(D)** Based on (C), summary values for each of the Omicron lineages with emergence periods highlighted **(E)** Statistical analysis of Ct values retrieved from samples collected during the emergence periods.

Likelihood of incoming lineages to cause breakthrough infections in recent vaccinees varies across different Omicron lineages

As BA.4 and BA.5 replaced BA.2 without clear advantage in their viral copy numbers, we next sought to assess the immune escape capacity to circumvent vaccine-induced immunity as another potential feature to increase a variant's fitness.

A blood donor seroprevalence survey showed that in December 2021, 95.5% of donors (95% Confidence Interval (CI) 93.5-96.9%) had antibodies against the Spike antigen, compared to 17.8% (CI 15.3-20.5%) with infection-induced antibodies (Busch et al. 2022; Jones et al. 2022). The US CDC reports high vaccination coverage in Connecticut (70% coverage with 2 doses in September 2021) as well as uptake of a third dose booster towards the end of 2021 (55% in January 2022) (**Fig 4A**). Further, bi-valent booster shots (Wuhan-Hu-1+BA.5) were administered to 23% of the population (**Fig 4A**). During 2022, the estimated proportion of individuals with at least one previous SARS-CoV-2 infection passed 75% in July (Klaassen et al. 2022, 2023) (**Fig 4B**). Thus, the pool of antigen-naïve hosts was greatly depleted towards mid-2022. Consequently, the fitness advantage of variants being able to escape from this high population immunity were expected to increase over time (**Fig 4B**). Accordingly, in experimental assays with serum collected from curated cohorts with defined antigen exposures, Omicron lineages BA.1, BA.2, BA.5 and XBB showed an astounding sequential drop in sensitivity to neutralization by vaccine- and infection induced antibodies (Cao et al. 2023; Mykytyn et al. 2023a,b; Rössler et al. 2022).

To examine how immune escape contributed to the fitness advantages of incoming Omicron lineages, we compared the likelihood of the incoming lineage to cause more vaccine breakthrough infections (BTI) than the dominant lineage during a five-week window of variant emergence (**Supplement Table 3**). We fitted mixed-effect multivariable logistic regression models adjusted for several potential confounding variables, including vaccination status, sex, age, location, and calendar time.

We show an increased odds of being infected with BA.1 compared to Delta amongst those vaccinated with at least 2 doses and who were more than 5 months from their most recent vaccination (OR: 1.92, 95% CI: 1.03 - 3.56) **(Fig 4C)**. Thus, enhanced immune escape played a role in the emergence of BA.1 while not being a significant driver of fitness advantage in the emergence of Delta in 2021. This result is in line with our previous study based on a similar population that found enhanced odds of being infected with Omicron BA.1 in triple-vaccinated individuals (Chaguza et al. 2022).

For the emergence of BA.2, we did not find overrepresentation in causing breakthrough infections during its emergence period, likely suggesting little to no effect of immune escape in our model **(Figure 4C)**. During the emergence of BA.5, we found a tendency for an increased risk of having been infected with BA.5 vs. BA.2 when having been vaccinated in the last 5 months prior to infection (OR: 2.441, 95% CI: 0.995 - 5.988), though this did not reach statistical significance **(Figure 4C)**. Finally, for the emergence of XBB, we found a tendency for a decreased likelihood of being infected with XBB amongst those who are more than 5 months from their most recent vaccination (OR: 0.586, 95% CI: 0.339-1.013) albeit this did not reach statistical significance in our analysis **(Figure 4C)**.

While these models are limited in understanding vaccine breakthrough infections from a population-level rather than mechanistic perspective, various in-vitro studies lend support to and provide context for our findings. Notably, the seeming lack of vaccine effects in the emergence of BA.2 can be explained by the low antigenic distance between BA.2 and BA.1, especially in people having received 3 monovalent doses (Arora et al. 2022b; Hoffmann et al. 2022; Kurhade et al. 2022; Tan et al. 2022). Additionally, several vaccine cohort studies have shown that BA.5 is more immune-evasive than BA.2 (Arora et al. 2022a; Cao et al. 2022b; Hoffmann et al. 2023b). The lack of statistical significance seen

for the emergence of BA.5 in our analysis may be attributable to the comparatively small sample size for these strata in our dataset **(Supplement Table 10)**.

Our analysis shows that the advantage an emerging lineage gained through immune escape varies over time and that in-vitro findings may not fully translate to a heterologous immunity landscape influenced by individual combinations of infections, boosters, shifts in test-seeking behavior, and other population-level factors.

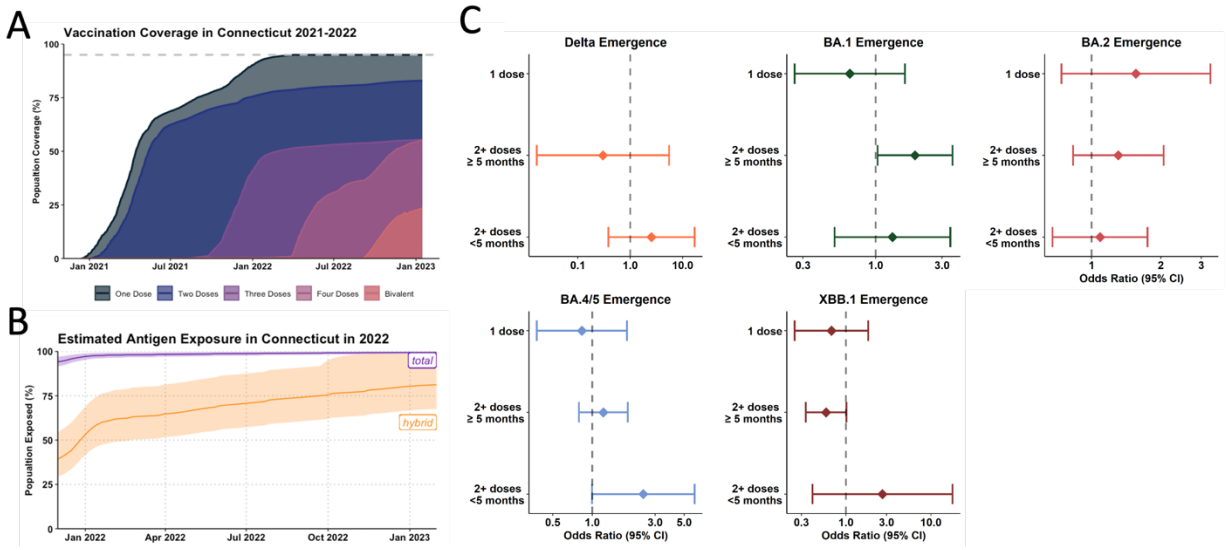


Figure 4 Influence of vaccine uptake and community-immunity levels on Omicron lineage emergence. (A) Vaccination coverage of Connecticut as reported by the US CDC for 2021 and 2022 **(B)** Estimated antigen exposure based on estimates from covidestim model according to either total (vaccination and/or infection) (purple) and hybrid exposure (infection and/or vaccination) (orange) **(C)** Logistic regression model to compare the ratios of infections caused by the incoming or previously dominant variant based on vaccination status during the emergence windows. BA.4 and BA.5 were analyzed together due to the similarity in the Spike region.

A conceptual model describing Omicron lineage fitness in a highly dynamic fitness landscape

Lastly, we explored how the infection dynamics of each variant influenced the emergence of the subsequent replacement variant. For this, we analyzed population-level immune exposure by calculating the number of infections 60-, 90-, or 120-days prior to the time point where an incoming lineage reached 5% of sequence frequency (**Fig 5A**). Emerging BA.2 lineages in March 2022 faced a population that experienced around 1.5 million BA.1 infections within the last 90 days, the time frame where antibody-mediated protection against re-infections is expected to be highest. For the emergence of BA.5 and XBB, this figure was around 0.7 million. This extensive infection dynamic led to a gradual rise in the proportion of Connecticut's population that was protected against reinfection as estimated through the covidestim model (**Fig 5B)** (Klaassen et al. 2022, 2023). Based on these estimates, we schematically plotted the relative importance of either intrinsic transmissibility or antigenic distance for an emerging variant's fitness advantage over the course of 2022 (**Fig 5C).** While immune escape increases gradually, following the number of people with Omicron antigen exposure, the advantage conferred through higher intrinsic transmissibility mainly played a role after the BA.1 and BA.5 waves, which infected up to 50% of the population each.

Summarizing our findings presented here, together with evidence from the literature, we developed a conceptual model based on a classical susceptible, infected, and recovered (SIR) compartmental dynamic transmission model highlighting fitness determinants of Omicron lineages (**Fig 5D**). Since Omicron BA.1 encompassed a vastly divergent Spike protein from pre-Omicron viruses, the previously observed high efficacy of vaccines, especially mRNA-platform based formulations, was greatly diminished (Andrews et al. 2022; Cao et al. 2022; Cele et al. 2022; Garcia-Beltran et al. 2022). In the model, this is reflected by opening a novel antigenic space (also defined as serotype) compared to previous pre-Omicron lineages (Simon-Loriere & Schwartz 2022). We include three different levels of susceptibility depending on vaccine status (S_N, S_V, S_{V3+}), as these strata have shown to harbor different quantities and qualities in antibody levels and thus protection against a first Omicron infection

(Muecksch et al. 2022; Park et al. 2022; Schaefer-Babajew et al. 2022). The exploration of this new antigenic space confers a fitness advantage in itself, moving 40-50% of Connecticut's population from the 'susceptible' (S_N = Omicron-naive, i.e. no previous Omicron antigen exposure) compartment to 'recovered' (R_I for unvaccinated or R_H for vaccinated).

The fraction of the population that just recovered from BA.1 (**Fig 5A**) was likely fairly well protected against a swift reinfection with either BA.1 or the antigenically similar BA.2 (Zhou et al. 2023). Based on epidemiological data, the transition from the R_I/R_H bin back to the 'susceptible' bin within the same antigenic space takes around 6 months (Andeweg et al. 2023; Malato et al. 2023; 2023). BA.2 was competing with BA.1 to infect the remaining 'susceptible' population (i.e. Omicron-naive S_N) fraction of the population. Based on our analysis, we conclude BA.2 outcompeted BA.1 in reaching this susceptible population faster through its higher intrinsic transmissibility (yellow arrow) (**Figs 3E and 5D**).

After the BA.1 and BA.2 waves, approximately one-third of the population remained in the Omicron-naive compartment (**Fig 1D**). The fitness advantage of variants in the same antigenic space, even with higher transmissibility (like BA.2.12.1), was restricted to this remaining pool of susceptibles. BA.4 and BA.5 successfully escaped from immunity conferred through pre-Omicron antigen contacts as well as BA.1/BA.2 infections and thus opened a second antigenic space (**Fig 5D**). Vaccinated or previously infected individuals fall back from being fairly protected ('recovered' bin) in the previous antigenic space to the 'susceptible' (S_I, S_H) state in the second antigenic space (**Fig 5D**). Although BA.4 and BA.5 displayed lower intrinsic transmissibility compared to BA.2 (**Fig 3D**), these variants outcompeted BA.2 as they could spread in a larger fraction of people (the R_I & R_H bins) (red arrow) (**Fig 5D**). The fairly high and steady transmission levels of BA.4 and 5 (**Fig 1C**) were likely defined by the rate of individuals moving from the recovered bin of the first antigen space to the S compartment in the second antigenic space (**Fig 5D**).

315 At the time XBB emerged, over 80% of the population had at least one Omicron antigen exposure (**Fig**
316 **4B**). For the purpose of our model, we assume that XBB covers a similar antigenic niche as BA.5.
317 Although XBB displays immune evasion from BA.5-induced immunity, some cross-neutralization,
318 especially in individuals with previous Omicron antigen contact, was reported (Kurhade et al. 2023;
319 Springer et al. 2023). In order to outcompete the ongoing BA.5 transmission chains (**Fig 2D**), any
320 incoming lineage had to maintain at least similar or higher levels of immune escape while also being
321 more transmissible (orange arrow) (**Fig 5D**). Thus, XBB outcompeted late-stage BA.5 samples through
322 its extensive immune escape profile (Hoffmann et al. 2023) paired with higher intrinsic transmissibility
323 (**Fig 3E**) and thus spreading faster within the susceptibles in the second antigenic space (**Fig 5D**).

324 In summary, we provide a framework to assess critical parameters that define variant fitness of Omicron
325 lineages in 2022. Ongoing SARS-CoV-2 evolution led to the emergence of lineages with greater
326 flexibility in exploring different fitness niches in a highly dynamic host population through a differential
327 combination of intrinsic transmissibility and immune escape.

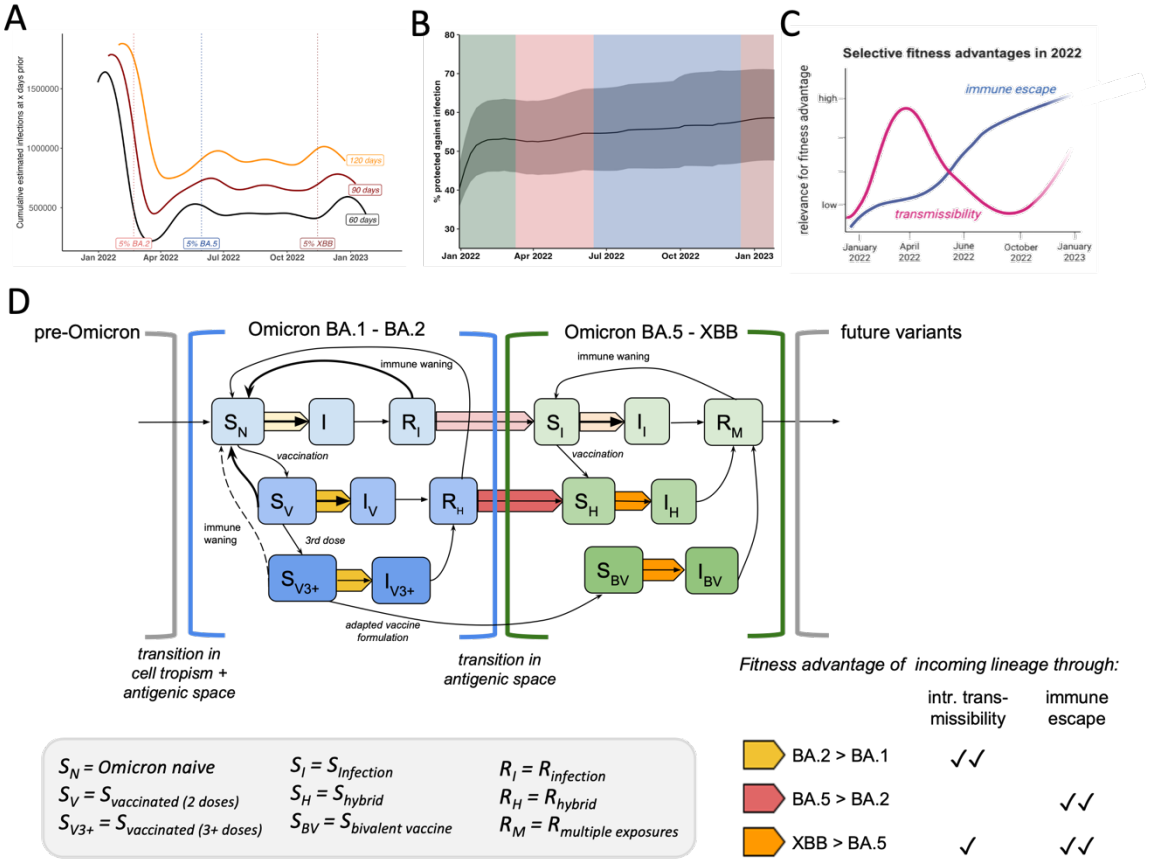


Figure 5 Summary of selective fitness advantages of Omicron lineage emergence in a dynamic host immunity landscape. (A) Sliding-window analysis of cumulative estimated infections 60, 90 or 120 days prior to each date on the plot with highlights when emerging Omicron lineages reached 5% of total samples **(B)** Frequency of estimated % of the population thought to be protected against infection based on vaccine and or infection status based on the covidestim model **(C)** Conceptual visualization of selective advantages of incoming Omicron lineages over the year **(D)** SIR-based conceptual transmission model highlighting different fitness advantages of incoming Omicron lineages depending on the status of the host population. Shades of yellow, orange and red represent the different combinations of advantageous fitness traits as outlined in the table in the right corner.

Discussion

We evaluated the underlying drivers of fitness advantages that led to the sequential sweeps of Omicron lineages in Connecticut during 2022. By determining selective traits that best explain the observed lineage dynamics, our analysis revealed the ability of SARS-CoV-2 lineages to thrive in a variety of highly dynamic fitness landscapes. A key result from the data presented here is that SARS-CoV-2 fitness was not dictated by a single factor, but rather a fine-tuned combination of intrinsic transmissibility and immune escape. This intrinsic volatility of the fitness landscape is, by definition, coupled with the antigen exposure history of the host population, making fitness advantages moving targets.

While novel virus lineages are designated according to their phylogenetic identity, inferring epidemiological relevance and antigenic space from genetic data alone remains challenging. Thus, in addition to continuing efforts of genomic surveillance, it remains crucial to collect relevant clinical and epidemiological patient data, ideally coupled with broad-scale and representative sero-prevalence surveys. Being able to link genetic information with additional laboratory and epidemiological data of locally matched host-pathogen populations and their epidemic outcomes is crucial for evaluating the potential of novel variants (Meijers et al. 2023).

2022 was unique in terms of the spread of the Omicron family, followed by highly synchronized antigen exposures resulting in fairly homogenous population-level immunity. Going forward, the immunity landscape will likely greatly diversify depending on individuals' antigen exposure histories and on the synchronicity of waning rates between individuals. This has implications for future SARS-CoV-2 transmission as a significant portion of the population might become susceptible to newly circulating lineages entering the northern hemisphere winter season of 2023/2024. We showed that in 2022, high levels of population immunity towards the end of the year slowed down the growth rates of BA.5 and XBB. Thus, variant-turnovers will take longer in future and might be more subtle than the rapid lineage

displacements in 2022. Additionally, the high number of co-circulating XBB-variants might further limit the growth rates of newly emerging variants (Beesley et al. 2023).

We demonstrate the utility of an integrated approach to genomic surveillance data collection for identifying key characteristics of SARS-CoV-2 fitness dynamics. Data linkage to key patient metadata was crucial to evaluate existing data from laboratory studies and curated vaccinee cohorts in a real-world, population-level setting and should be considered when designing surveillance systems. Such frameworks provide critical infrastructure to evaluate emerging lineages of known and unknown pathogens and can be further incorporated into tailoring public health responses and informing vaccine formulations. Lessons learned from SARS-CoV-2 can be utilized and further adapted to other pathogens to maximize the usefulness of such surveillance programmes for community health.

This retrospective analysis re-visits the lineage replacement dynamics of Omicron lineages that dominated Connecticut in 2022 and sheds light on the underlying drivers of their sequential fitness advantages. This study exemplifies a use-case for combining sequencing data with epidemiological and laboratory metadata to better understand pathogen emergence. Based on the high-level data used in this study, we showcase that linking vaccination status and viral copy numbers to genomic surveillance data may provide crucial tools to evaluate variant fitness for SARS-CoV-2 and beyond.

Limitations

One limitation of this study might be the validity of these findings to other populations and settings. As outlined above, variant transmission advantages are intrinsically coupled to the local host population and thus our results are most applicable to other places with similar population characteristics, vaccination uptake rates, and public health policies. Whereas the specific dynamics of the different Omicron waves may vary locally, the data presented here is valuable for examining overall trends that are broadly applicable to other settings. Overall fitness advantages of each new Omicron lineage over their respective predecessor were fairly consistent globally as recently shown when analyzing different settings globally (Meijers et al. 2023).

A further potential limitation is that our dataset for Ct value comparison is based on a pre-selection from the overall pool of available SARS-CoV-2 positive samples based on a Ct value cut-off in the diagnostic test that were expected to yield high sequencing coverage. We think this reflects the proportion of samples that would possess the viral burden that is relevant for onward transmission and thus suitable for the purpose of this study. Further, it could be argued that Ct values do not necessarily scale with the amounts of infectious viral particles and that the kinetics of RNA clearance may not be directly related to clearance of infectious virus. For our specific assay, we used a standard curve to convert Ct values to absolute viral genome copies and we are usually able to isolate infectious virus from samples with Ct values below 30 (based on internal data). Further, symptom severity is thought to have remained similar over all Omicron lineages, reducing the potential for bias due to differences in test seeking behavior. Lastly, for the purposes of this study, we were interested in comparing trends between lineages in contrast to absolute values.

Materials & Methods

Ethics statement

The Institutional Review Board from the Yale University Human Research Protection Program determined that the RT-qPCR testing and sequencing of de-identified remnant COVID-19 clinical samples obtained from clinical partners conducted in this study is not research involving human subjects (IRB Protocol ID: 2000028599).

Data sources

Clinical sample collection, RT-qPCR, and sequencing

SARS-CoV-2 positive samples (nasal swabs in viral transport media) were collected through the Yale New Haven Hospital (YNHH) System as a part of routine inpatient and outpatient testing and sent to the Yale SARS-CoV-2 Genomic Surveillance Initiative. Using the MagMAX viral/pathogen nucleic acid isolation kit, nucleic acid was extracted from 300µl of each clinical sample and eluted into 75µl of elution buffer. Extracted nucleic acid was then tested using a "research use only" (RUO) RT-qPCR assay (Vogels et al. 2021) for SARS-CoV-2 RNA. Libraries were prepared for sequencing using the Illumina COVIDSeq Test (RUO version) and quantified using the Qubit High Sensitivity dsDNA kit. Negative controls were included for RNA extraction, cDNA synthesis, and amplicon generation.

Prepared libraries were sequenced at the Yale Center for Genomic Analysis on the Illumina NovaSeq with a 2x150 approach and at least 1 million reads per sample.

Reads were then aligned to the Wuhan-Hu-1 reference genome (GenBank MN908937.3) using BWA-MEM v.0.7.15 (Li 2013). Adaptor sequences were then trimmed, primer sequences masked, and consensus genomes called (simple majority >60% frequency) using iVar v1.3.133 (Grubaugh et al. 2019) and SAMtools(Danecek et al. 2021). When <20 reads were present at a site an ambiguous "N" was used, with negative controls consisting of ≥99% Ns. The Pangolin lineage assignment tool (O'Toole et al. 2021) was used for assigning viral lineages.

Clinical metadata

We obtained patient metadata and vaccination records from the YNHH system and the Center for Outcomes Research and Evaluation (CORE) and matched these records to sequencing data through unique sample identifiers. Duplicate patient records or those with missing or inconsistent vaccination data were removed. We determined vaccination status at time of infection by comparing the sample collection date to the patient's vaccination record dates.

We then categorized vaccine breakthrough statuses with respect to both the number of vaccine doses received more than 14 days prior to the collection date and the timing of the most recent vaccination relative to a 5 month period of time. Patient vaccination statuses at time of infection were thus categorized as: non-vaccine breakthrough, one dose vaccine breakthrough, two or more dose vaccine breakthrough greater than 5 months since the most recent vaccination, or two or more dose vaccine breakthrough within 5 months since the most recent vaccination.

Population vaccination trends

We obtained data on vaccination trends in Connecticut from the Centers for Disease Control and Prevention (CDC) (CDC 2021).

Population variant trends

We obtained variant trend data for Connecticut from the Global Initiative on Sharing All Influenza Data (GISAID).

Variant R_t and immunity estimates

We obtained variant cumulative estimated case counts from covidestim, a Bayesian nowcasting approach that incorporates reported cases, hospitalizations, immunity (Klaassen et al. 2022, 2023), exposures, and vaccination data to generate state and county level estimates of variant specific infections (Chitwood et al. 2022). To obtain variant-specific Rt estimates, we used the variant-specific infection estimates from covidestim with the EpiEstim R package (Nouvellet et al. 2018).

Analyses

Variant Rt ratios

To compare R_t values between variants, we first selected 14-day periods when a new variant was emerging in the population. For each of these time periods, we then divided the daily R_t value of the emerging variant by the R_t value of the established variant for the same calendar day to calculate daily R_t ratios.

Variant emergence periods and logistic growth rates

To determine the length of time each variant needed to become established in the population, we calculated daily frequencies for each variant across Connecticut using case count data from GISAID. We then defined a variant's emergence period as the date from when a variant first accounted for 5% of all cases to the first date the variant reached its maximum frequency in the population. We then fitted a locally estimated scatterplot smoothing (loess) curve to the data and extracted the fitted value corresponding to a frequency of 50% in the population. With this value we determined the number of days it took each variant to increase from a frequency of 5% in the population to 50%.

Using the same emergence periods, we fitted binomial logistic regression models where the variant lineage was modeled as a function of calendar time. For each model, the variant(s) being displaced in the population served as the reference group. We then extracted the coefficient of the predictor variable of each model to determine the logistic growth rate for each variant.

Variant Ct values over time and in periods of emergence

To understand how Ct values change across time, we subset each variant to the period of time where it was above 10% frequency in the population using RT-qPCR data from the Yale SARS-CoV-2 Genomic Surveillance Initiative. For each of these time periods and variants we tested for heteroscedasticity via a Breusch-Pagan test, checked for non-linearity or outlier values by plotting the residual values against

the fitted values, and tested for normality via a Q-Q plot and frequency histogram of model residuals. We then fitted linear regression models where the Ct value was modeled as a function of calendar time.

To compare Ct values between variants, we identified 3, 4-week periods of variant co-circulation using the same dataset. For the two periods of pairwise comparisons, we performed Wilcoxon rank sum tests. For the comparison between three variants, we performed a Welch's ANOVA test as well as a one-way ANOVA test to test for concordance, followed by post-hoc pairwise tests via Tukey's honest significance test.

Mixed effect multivariable logistic regression models

To determine the impact of vaccinations in periods of variant emergence, we used sequencing data from the Yale SARS-CoV-2 Genomic Surveillance Initiative matched to vaccination data from the YNHH System and CORE to identify five, five-week periods of variant co-circulation. We selected the specific date ranges of these periods so as to balance the number of unvaccinated individuals attributed to each variant in each period.

For these five periods we then fit mixed effect multivariable logistic regression models with a dichotomous outcome of the co-circulating variants found in each time period. To dichotomize the outcome in periods when more than two variants were circulating, we aggregated variants that emerged or were displaced contemporaneously, with the reference set as the variant(s) that was being displaced in the population.

Model covariates were selected via an Akaike information criterion (AIC) selection criteria test and included vaccination status at the time of infection, patient sex (male or female), age (5-17, 18-39, 40-64, 65+), town of residence as a random effect, and calendar time as a linear predictor. Non-vaccine breakthrough, male sex, and the 18-39 age group served as variable reference levels. Due to the inability for under 5 years olds to receive vaccinations for the majority of our study period, we restricted our analysis to individuals 5 years and older.

To test the impact of the date interval lengths, we performed a sensitivity analysis by modifying the period of emergence from 5 weeks to 3,4,6,7, and 8 weeks and found minimal differences in the results.

Factors impacting XBB.1.5 Ct values

To investigate the factors that impact Ct values associated with the XBB.1 variant infections, we subset sequencing data from the Yale SARS-CoV-2 Genomic Surveillance Initiative to only those infections caused by lineages within the XBB.1 parent lineage. We then identified calendar time, sex (female or male), patient class (inpatient, outpatient, or emergency), age (<18, 18-49, 50-69, 70+), and vaccination status at time of infection as variables that could impact Ct values.

To understand the impact of calendar time, we tested for heteroscedasticity via a Breusch-Pagan test, checked for non-linearity or outlier values by plotting the residual values against the fitted values, and tested for normality via a Q-Q plot and frequency histogram of model residuals. We then fit a linear regression model to the data with Ct value as a function of calendar time.

For the remaining variables we assessed the normality of the Ct value distributions via Shapiro-Wilk tests and the variance of the Ct value distributions via Bartlett or F-tests. To test for differences in Ct values by sex, we performed equal variance t-tests. For the remaining variables, we performed Welch's ANOVA tests as well as one-way ANOVA tests to test for concordance, followed by post-hoc pairwise comparisons via Tukey's honest significance test.

Conceptual SIR Model

To synthesize our findings, we constructed a conceptual framework of variant fitness based on the traditional SIR transmission model. As a conceptual framework, this model does not attempt to explicitly simulate transmission through quantitative approaches. Rather, the model provides a visual representation of the proposed mechanisms by which the Omicron lineages may have gained selective advantages based on the results of our other analyses and findings from the literature.

Statistical analysis and data availability

We used the R statistical software (v. 4.2.1) (R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.) for all statistical analysis and figures.

Data and code used in this study are publicly available on GitHub (https://github.com/NickChen10/Omicron_project).

Acknowledgments

We gratefully acknowledge the work of all members of the Yale Genomic Surveillance Initiative especially Center for Outcome Research at Yale New Haven Hospital. We thank all 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. This project is supported by the CDC Broad Agency Announcement Contracts 75D30122C14697 (awarded to NDG) and 75D30121C10273 (supports KF), the Connecticut Department of Public Health (CDPH) contract 21PSX0049 (awarded to the Yale Center for Genome Analysis and supports NDG), and the Council of State and Territorial Epidemiologists contract NU38OT000297 (supports KF). This work does not necessarily represent the views of the CDC or CDPH.

Author contribution:

Conceptualization: NC, NDG, AMH

Methodology: NC, KP, NDG, AMH

Investigation: NC, KP, AMH, DMW, JLW, VP, NDG, RL, FK

Visualization: NC, KP, AMH

Funding acquisition: NDG

Supervision: NDG, AMH

Writing – original draft: NC, AMH

Writing – review & editing: all authors

Competing Interests

NDG is a paid consultant for BioNTech, DMW has received consulting fees from Pfizer, Merck, and GSK, unrelated to this manuscript, and has been PI on research grants from Pfizer and Merck to Yale, unrelated to this manuscript. JLW has received consulting fees from Pfizer and Revelar Biotherapeutics Inc unrelated to this manuscript.

References

Andeweg SP, Gier B de, Vennema H, Walle I van, Maarseveen N van, et al. 2023. Higher risk of SARS-CoV-2 Omicron BA.4/5 infection than of BA.2 infection after previous BA.1 infection, the Netherlands, 2 May to 24 July 2022. Eurosurveillance. 28(7):2200724

Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, et al. 2022. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. N. Engl. J. Med. 386(16):1532–46

Arora P, Zhang L, Krüger N, Rocha C, Sidarovich A, et al. 2022a. SARS-CoV-2 Omicron sublineages show comparable cell entry but differential neutralization by therapeutic antibodies. Cell Host Microbe. 30(8):1103-1111.e6

Arora P, Zhang L, Rocha C, Sidarovich A, Kempf A, et al. 2022b. Comparable neutralisation evasion of SARS-CoV-2 omicron subvariants BA.1, BA.2, and BA.3. Lancet Infect. Dis. 22(6):766–67

Beesley LJ, Moran KR, Wagh K, Castro LA, Theiler J, et al. 2023. SARS-CoV-2 variant transition dynamics are associated with vaccination rates, number of co-circulating variants, and convalescent immunity. eBioMedicine. 91:104534

Brito AF, Semenova E, Dudas G, Hassler GW, Kalinich CC, et al. 2022. Global disparities in SARS-CoV-2 genomic surveillance. Nat. Commun. 13(1):7003

Busch MP, Stramer SL, Stone M, Yu EA, Grebe E, et al. 2022. Population-weighted seroprevalence from SARS-CoV-2 infection, vaccination, and hybrid immunity among U.S. blood donations from January-December 2021. Clin. Infect. Dis.: Off. Publ. Infect. Dis. Soc. Am. 75(Supplement_2):ciac470

Cao Y, Jian F, Wang J, Yu Y, Song W, et al. 2023. Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. Nature. 614(7948):521–29

Cao Y, Wang J, Jian F, Xiao T, Song W, et al. 2022a. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature. 602(7898):657–63

Cao Y, Yisimayi A, Jian F, Song W, Xiao T, et al. 2022b. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. Nature. 608(7923):593–602

CDC. 2021. COVID-19 Vaccination Trends in the United States,National and Jurisdictional | Data | Centers for Disease Control and Prevention

Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, et al. 2022. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. Nature. 602(7898):654–56

Chaguza C, Coppi A, Earnest R, Ferguson D, Kerantzas N, et al. 2022. Rapid emergence of SARS-CoV-2 Omicron variant is associated with an infection advantage over Delta in vaccinated persons. Med (N. York, Ny). 3(5):325-334.e4

Chitwood MH, Russi M, Gunasekera K, Havumaki J, Klaassen F, et al. 2022. 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. 18(8):e1010465

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10(2):giab008

Earnest R, Uddin R, Matluk N, Renzette N, Turbett SE, et al. 2022. Comparative transmissibility of SARS-CoV-2 variants Delta and Alpha in New England, USA. Cell Reports Medicine. 3(4):100583

586 Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, et al. 2021. Genomics and epidemiology of
587 the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 372(6544):815–21

588 Fryer HR, Golubchik T, Hall M, Fraser C, Hinch R, et al. 2023. Viral burden is associated with age,
589 vaccination, and viral variant in a population-representative study of SARS-CoV-2 that accounts for time-
590 since-infection-related sampling bias. PLOS Pathog. 19(8):e1011461

591 Garcia-Beltran WF, Denis KJSt, Hoelzemer A, Lam EC, Nitido AD, et al. 2022. mRNA-based COVID-19
592 vaccine boosters induce neutralizing immunity against SARS-CoV-2 Omicron variant. Cell. 185(3):457-
593 466.e4

594 GISAID. https://gisaid.org

595 Goh AXC, Chae S-R, Chiew CJ, Tang N, Pang D, et al. 2023. Characteristics of the omicron XBB subvariant
596 wave in Singapore. Lancet. 401(10384):1261–62

597 Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, Jesus JGD, et al. 2019. An amplicon-based
598 sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome
599 Biol. 20(1):8

600 Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, et al. 2021. Estimating epidemiologic
601 dynamics from cross-sectional viral load distributions. Science. 373(6552):eabh0635

602 Hill V, Plessis LD, Peacock TP, Aggarwal D, Colquhoun R, et al. 2022. The origins and molecular evolution
603 of SARS-CoV-2 lineage B.1.1.7 in the UK. Virus Evol. 8(2):veac080

604 Hoffmann M, Arora P, Nehlmeier I, Kempf A, Cossmann A, et al. 2023. Profound neutralization evasion
605 and augmented host cell entry are hallmarks of the fast-spreading SARS-CoV-2 lineage XBB.1.5. Cell. Mol.
606 Immunol. 20(4):419–22

607 Hoffmann M, Behrens GMN, Arora P, Kempf A, Nehlmeier I, et al. 2022. Effect of hybrid immunity and
608 bivalent booster vaccination on omicron sublineage neutralisation. Lancet Infect. Dis. 23(1):25–28

609 Jones JM, Opsomer JD, Stone M, Benoit T, Ferg RA, et al. 2022. Updated US Infection- and Vaccine-
610 Induced SARS-CoV-2 Seroprevalence Estimates Based on Blood Donations, July 2020-December 2021.
611 JAMA. 328(3):298–301

612 Klaassen F, Chitwood MH, Cohen T, Pitzer VE, Russi M, et al. 2022. Population Immunity to Pre-Omicron
613 and Omicron Severe Acute Respiratory Syndrome Coronavirus 2 Variants in US States and Counties
614 Through 1 December 2021. Clin. Infect. Dis.: Off. Publ. Infect. Dis. Soc. Am. 76(3):e350–59

615 Klaassen F, Chitwood MH, Cohen T, Pitzer VE, Russi M, et al. 2023. Changes in Population Immunity
616 Against Infection and Severe Disease From Severe Acute Respiratory Syndrome Coronavirus 2 Omicron
617 Variants in the United States Between December 2021 and November 2022. Clin. Infect. Dis.: Off. Publ.
618 Infect. Dis. Soc. Am. 77(3):355–61

619 Kurhade C, Zou J, Xia H, Cai H, Yang Q, et al. 2022. Neutralization of Omicron BA.1, BA.2, and BA.3
620 SARS-CoV-2 by 3 doses of BNT162b2 vaccine. Nat. Commun. 13(1):3602

621 Kurhade C, Zou J, Xia H, Liu M, Chang HC, et al. 2023. Low neutralization of SARS-CoV-2 Omicron
622 BA.2.75.2, BQ.1.1 and XBB.1 by parental mRNA vaccine or a BA.5 bivalent booster. Nat. Med. 29(2):344–
623 47

624 Lentini A, Pereira A, Winqvist O, Reinius B. 2022. Monitoring of the SARS-CoV-2 Omicron BA.1/BA.2
625 lineage transition in the Swedish population reveals increased viral RNA levels in BA.2 cases. Med (N.
626 York, Ny). 3(9):636-643.e4

627 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv

628 Ma KC, Shirk P, Lambrou AS, Hassell N, Zheng X, et al. 2023. Genomic Surveillance for SARS-CoV-2
629 Variants: Circulation of Omicron Lineages — United States, January 2022–May 2023. Morb. Mortal. Wkly.
630 Rep. 72(24):651–56

631 Malato J, Ribeiro RM, Fernandes E, Leite PP, Casaca P, et al. 2023. Stability of hybrid versus vaccine
632 immunity against BA.5 infection over 8 months. Lancet Infect. Dis. 23(2):148–50

633 Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, et al. 2023. The evolution of SARS-CoV-2. Nat
634 Rev Microbiol. 1–19

635 Meijers M, Ruchnewitz D, Eberhardt J, Łuksza M, Lässig M. 2023. Population immunity predicts
636 evolutionary trajectories of SARS-CoV-2. Cell

637 Muecksch F, Wang Z, Cho A, Gaebler C, Tanfous TB, et al. 2022. Increased memory B cell potency and
638 breadth after a SARS-CoV-2 mRNA boost. Nature. 1–7

639 Musalkova D, Piherova L, Kwasny O, Dindova Z, Stancik L, et al. 2023. Trends in SARS-CoV-2 cycle
640 threshold values in the Czech Republic from April 2020 to April 2022. Sci. Rep. 13(1):6156

641 Mykytyn AZ, Fouchier RA, Haagmans BL. 2023a. Antigenic evolution of SARS coronavirus 2. Curr. Opin.
642 Virol. 62:101349

643 Mykytyn AZ, Rosu ME, Kok A, Rissmann M, Amerongen G van, et al. 2023b. Antigenic mapping of
644 emerging SARS-CoV-2 omicron variants BM.1.1.1, BQ.1.1, and XBB.1. Lancet Microbe. 4(5):e294–95
645 Nouvellet P, Cori A, Garske T, Blake IM, Dorigatti I, et al. 2018. A simple approach to measure
646 transmissibility and forecast incidence. Epidemics. 22:29–35

647 O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, et al. 2021. Assignment of Epidemiological Lineages
648 in an Emerging Pandemic Using the Pangolin Tool. Virus Evol. 7(2):veab064-

649 Park Y-J, Pinto D, Walls AC, Liu Z, Marco AD, et al. 2022. Imprinted antibody responses against SARS-
650 CoV-2 Omicron sublineages. Science. 378(6620):619–27

651 Petrone ME, Rothman JE, Breban MI, Ott IM, Russell A, et al. 2022. Combining genomic and
652 epidemiological data to compare the transmissibility of SARS-CoV-2 variants Alpha and Iota. Commun
653 Biology. 5(1):439

654 R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical
655 Computing, Vienna. www.R-project.org

656 Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, et al. 2020. Preliminary genomic characterisation of
657 an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. virological.org

658 Rössler A, Netzl A, Knabl L, Schäfer H, Wilks SH, et al. 2022. BA.2 and BA.5 omicron differ
659 immunologically from both BA.1 omicron and pre-omicron variants. Nat. Commun. 13(1):7701

660 Schaefer-Babajew D, Wang Z, Muecksch F, Cho A, Loewe M, et al. 2022. Antibody feedback regulates
661 immune memory after SARS-CoV-2 mRNA vaccination. Nature. 1–3

662 Simon-Loriere E, Schwartz O. 2022. Towards SARS-CoV-2 serotypes? Nat. Rev. Microbiol. 20(4):187–88

663 Springer DN, Medits I, Weseslindtner L, Stiasny K, Aberle JH. 2023. SARS-CoV-2 neutralising antibody
664 response to bivalent booster after omicron infection. Lancet Microbe

Tan CW, Chia WN, Zhu F, Young BE, Chantasrisawad N, et al. 2022. SARS-CoV-2 Omicron variant emerged under immune selection. Nat. Microbiol. 7(11):1756–61

Team C-19 F, Stein C, Nassereldine H, Sorensen RJD, Amlag JO, et al. 2023. Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. Lancet. 401(10379):833–42

Tegally H, Moir M, Everatt J, Giovanetti M, Scheepers C, et al. 2022. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. Nat. Med. 28(9):1785–90

Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, et al. 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature. 1–6

Tegally H, Wilkinson E, Tsui JL-H, Moir M, Martin D, et al. 2023. Dispersal Patterns and Influence of Air Travel During the Global Expansion of SARS-CoV-2 Variants of Concern. Cell

Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, et al. 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature. 1–10

Vogels CBF, Breban MI, Ott IM, Alpert T, Petrone ME, et al. 2021. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. Plos Biol. 19(5):e3001236

WHO. 2021. COVID-19Weekly Epidemiological Update.

Zhou J, Sukhova K, Peacock TP, McKay PF, Brown JC, et al. 2023. Omicron breakthrough infections in vaccinated or previously infected hamsters. Proc. Natl. Acad. Sci. 120(45):e2308655120