SARS-CoV-2 Evolution on a Dynamic Immune Landscape

Max von Kleist (KleistM@rki.de)
Robert Koch Institute https://orcid.org/0000-0001-6587-6394
N. Alexia Raharinirina
Freie Universität Berlin https://orcid.org/0000-0003-1253-8927
Nils Gubela
Freie Universität Berlin
Daniela Börnigen
Robert-Koch Institute
Maureen Smith
Robert-Koch Institute
Djin-Ye Oh
Robert-Koch Institute
Matthias Budt
Robert Koch Institut
Christin Mache
Robert-Koch Institute
Claudia Schillings
Freie Universität Berlin
Stephan Fuchs
Robert-Koch Institute
Ralf Dürrwald
Robert-Koch Institute
Thorsten Wolff
Robert Koch Institute https://orcid.org/0000-0001-7688-236X
Martin Hoelzer
Robert-Koch Institute
Sofia Paraskevopoulou
Robert Koch Institute https://orcid.org/0000-0003-2608-2596

Biological Sciences - Article

Keywords:

Posted Date: October 12th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3366919/v1

Additional Declarations: There is **NO** Competing Interest.

SARS-CoV-2 Evolution on a Dynamic Immune Landscape

N. Alexia Raharinirina[1,*], Nils Gubela[1,2,*], Daniela Börnigen[3,*], Maureen Rebecca Smith[3], Djin-Ye Oh[4], Matthias Budt[4], Christin Mache[4], Claudia Schillings[1], Stephan Fuchs[5], Ralf Dürrwald[4], Thorsten Wolff[4], Martin Hölzer[5], Sofia Paraskevopoulou[5], and Max von Kleist[1,3,+]

[1]Department of Mathematics & Computer Science, Freie Universität Berlin, Germany
[2]International Max-Planck Research School "Biology and Computation" (IMPRS-BAC), Max-Planck Institute for Molecular Genetics, Berlin, Germany
[3]Project groups, Robert-Koch Institute, Berlin, Germany
[4]Department 1, Robert-Koch Institute, Berlin, Germany
[5] Department MFI, Robert-Koch Institute, Berlin, Germany
[*]these authors contributed equally
[+]max.kleist@fu-berlin.de

September 18, 2023

Abstract

Since the onset of the pandemic, many SARS-CoV-2 variants have emerged, exhibiting substantial evolution in the virus' spike protein, the main target of neutralizing antibodies. A plausible hypothesis proposes that the virus evolves to evade antibody-mediated neutralization (vaccine- or infection-induced) to maximize its ability to infect an immunologically experienced population. While virus infection induces neutralizing antibodies, viral evolution may thus navigate on an dynamic immune landscape that resulted from the infection history in different regions. Global inequalities in vaccine distribution and differences in infection-prevention measures have shaped this global immunological landscape, resulting in uneven geographic distributions of SARS-CoV-2 variants. Consequently, predicting which variant will spread within particular regions has become increasingly challenging. To tackle this challenge, we developed a comprehensive mechanistic model of the dynamic immunological landscape of SARS-CoV-2. We utilized deep-mutational scanning data and antibody pharmacokinetics to compute time-dependent cross-neutralization between arbitrary variants. Combined with infection history and molecular surveillance data, we could predict the variant-specific *relative* number of susceptibles over time, exemplified for Germany. This quantity precisely matched historical variant dynamics, predicted future variant dynamics, and could explain global differences in variant dynamics. Our work strongly supports the hypothesis that SARS-CoV-2 evolution is driven by escape from humoral immunity, allows contextualizing risk assessment of variants, and provides important clues for vaccine design.

Introduction

In 2023, most non-pharmaceutical measures intended to curb the spread of SARS-CoV-2 have been lifted as the WHO de-escalated the COVID-19 pandemic from a 'Public Health Emergency of International Concern' (PHEIC) to a 'pandemic that denotes an established and ongoing health issue' [1]. While the occurrence of explosive surges of COVID-19 has most likely come to an end, SARS-CoV-2 still spreads at alarming rates, with 1.5–5% of the population infected at any given time [2] and thousands of ICU cases and reported fatalities monthly [3], not-withstanding long-term effects [4]. As of August 2023, nearly four years after the emergence of SARS-CoV-2, approximately 770 million cases have been *reported* [5], with true numbers magnitudes larger, such that it is reasonable to assume that almost the entire world population has been exposed to SARS-CoV-2 antigen to date.

Since its first introduction in the human population, SARS-CoV-2 has evolved significantly [6] with most changes allocated to its surface protein (spike), which is both responsible for binding host cells and the target site for neutralizing antibodies elicited by vaccines or infection [7]. While many early evolutionary changes of the virus are believed to have been driven by the adaptation of the virus to the human host [8, 9, 10, 11], spike mutations associated with escape from neutralizing antibodies were observed already during the first year of the pandemic, and before the roll-out of vaccines [12, 13, 14]. After the broad vaccine roll-out and the global Delta wave, SARS-CoV-2 has circulated across an immunologically more experienced population, being transmitted more and more through hosts who had been vaccinated or previously infected. Subsequently, the early Omicron variants BA.1 and BA.2 emerged, which exhibited a substantial transmission advantage over Delta [6]. This transmission advantage was associated with a pronounced evasion from humoral immunity [15, 16, 17, 18]. While inequalities in vaccine distribution[1] and differences in the implementation of non-pharmaceutical measures may have created inter-country differences early-on in the pandemic [19], those were relatively minor compared to the vast geographical differences observed in the distribution and spread of emerging Omicron sub-lineages over the last year. This geographical variation may reflect an increasing complexity of the global immunological landscape, where the course of waves with new (sub-) variants in a particular region could be substantially influenced by the infection history of that region, i.e. which variants dominated the preceding waves and at what time.

Consequently, it has become increasingly difficult to predict whether a new variant will spread successfully in a given population at a given time. This circumstance also poses challenges in optimizing the design of adapted mRNA vaccines to protect vulnerable groups or heavily exposed individuals. Despite the appreciation of the problem [20] and the abundance of rich data sources for SARS-CoV-2, until now, there has been limited progress in integrating the available data to inform variant risk assessment and vaccine design, as acknowledged by a "limitation of evidence" by the WHO technical advisory group on COVID-19 vaccine composition (TAG-CO-VAC) [21].

Almost four years after the emergence of SARS-CoV-2, the scientific community has collected an unprecedented wealth of data. Amongst others, the Bloom Lab developed high-throughput assays that can precisely define the epitopes of neutralizing antibodies and quantify the phenotypic effect of virtually any mutation on the spike protein, even those not yet observed in circulating variants [22, 23, 24]. On the other hand, national molecular surveillance programs

[1] https://data.undp.org/vaccine-equity/

allow for rapid detection of emerging variants and monitor variant dynamics [25]. While diagnostic surveillance of SARS-CoV-2 has decreased substantially over the last year, innovative methods, including genome-based incidence estimation [26] and virus load quantification in wastewater [27], may nevertheless enable close monitoring of SARS-CoV-2 spread. While each of these achievements provide valuable information individually, it is crucial to consider their intricate interrelation when defining the immunological landscape in a country. This comprehensive understanding would facilitate contextualizing information about emerging variants, including their potential to spread nationally and internationally. Also, it can help identifying antigens that maximize vaccine efficacy both seasonally and regionally.

Here, we present a mathematical approach to rigorously define the adaptive immunological landscape of SARS-CoV-2 in a national context by combining molecular data on antibody-epitope interaction, national molecular surveillance data, and infection history. Based on this data for the German COVID-19 pandemic, we calculate an immunity-driven *relative* transmission rate that precisely predicts variant dynamics and explains why certain globally successful Omicron variants failed to spread in Germany, strongly supporting the hypothesis that SARS-CoV-2 evolution is driven by variant-specific immunity. The proposed methods can be adapted to national surveillance data, substantially extending upon current alert systems and may help to design vaccines that complement infection-induced immunity.

Results

This study aimed to estimate population immunity towards all circulating variants and to assess whether variant-specific immunity is the driving force behind the ongoing evolution of SARS-CoV-2. We integrated a vast set of available data to achieve this goal, summarized in **Fig. 1A**. Sequencing data from national genomic surveillance programs provided information on the mutation profile and the frequency of circulating SARS-CoV-2 variants over time, whereas a recently developed tool (GInPipe [26]) enabled the correction of case reporting data for time-varying under-reporting. These two pieces of information allowed us to recover a complete variant-resolved infection timeline. Next, we used deep mutational scanning data [22, 24] to identify the epitopes of neutralizing (Spike-targeting) antibodies, enabling us to group variants based on their unique antigenic profiles and to compute cross-neutralization probabilities for all variants with distinct antigenic profiles, i.e., how much and how long will recovery from a recent infection with a variant x protect against another variant y? Lastly, we integrated the timeline of infections, cross-neutralization, and immune-waning dynamics to estimate the expected number of susceptible individuals for each variant. If SARS-CoV-2 evolution was decisively driven by population immunity, the variant-specific number of susceptibles should be directly proportional to the variant-specific transmission rate, and hence, allowing to estimate a competitive growth (dis-)advantage of a given variants. While the outlined method is universally applicable, we will demonstrate our analytical approach using data from the German SARS-CoV-2 epidemic.

Variant genomic profiles and infection timeline

We used German SARS-CoV-2 genomes that were randomly collected between July 01, 2021 to April 16, 2023 ($\approx$ 600,000 sequences) to predict immune dynamics between March 01, 2022 to April 16, 2023 (daily variant frequencies, see Suppl. **Fig. S1**). The time from July

01, 2021 to March 01, 2022 served as a 'burn-in' phase to converge to the initial immunological landscape. The viral genomes belonged to 1,205 Pangolin lineages and harbored 280 and 314 mutations in the Receptor-Binding Domain (RBD) and N-Terminal Domain (NTD) regions, respectively (**Suppl. Table ST1**). Of these initially 1,205 lineages, we identified 651 unique Spike profiles (called 'Spike-pseudo-groups' henceforth), **Fig. 1C** (**Suppl. Table ST2)**. Next, we reconstructed the timeline of all 'Spike-pseudo-group' frequencies, as illustrated for the most abundant groups in **Fig. 1B** (full dataset in **Suppl. Data D1**). Lastly, we used genome-based incidence reconstruction (GInPipe [26]) to estimate the timeline of infections, **Fig. 1D**, which allowed us to turn 'Spike-pseudo-group' frequencies into infection counts. Genome-based incidence reconstruction with GInPipe illustrated a considerable increase of undetected cases from April 2022 onwards, when complimentary antigen testing was discontinued in Germany (**Suppl. Fig. S2**). SARS-CoV-2 under-reporting became even more pronounced from summer 2022 onwards, with substantial levels of ongoing SARS-CoV-2 infections between February 2022 and April 2023. To validate our incidence reconstruction tool, we compared infection numbers for the UK derived with GInPipe with data from the COVID-19 Infection Survey of the Office for National Statistics (ONS) [2], **Suppl. Fig. S3**, yielding coherent results. Our analysis demonstrated that reported cases may substantially under-estimate true SARS-CoV-2 incidences and that under-reporting rates may change over time. If reported cases are not corrected, this may create biases in estimating population exposure- and population immunity to SARS-CoV-2. Here, we showed that GInPipe can be utilized to correct for time-dependent under-reporting of SARS-CoV-2.

Deep mutational scanning data and cross-neutralization potency

Next, we estimated cross-neutralization profiles for each possible pair of 'antibody inducing' x and cross-neutralized 'Spike-pseudo-group' y. In the original deep mutational scanning (DMS) data 'escape fractions' were assigned to all sites in the RBD region for a large panel of 836 antibodies [22, 24, 23], which were aggregated into ten epitope classes, **Fig. 2A** [28] (**Suppl. Tables ST3–ST4** for details on each antibody and class). We added an 'NTD class' to these ten epitope classes based on the antigenic super-sites in NTD [29]. Original DMS 'escape fractions' measure the proportion of antibodies not binding to a mutated Spike protein, relative to the immune-inducing variant x, ranging from 0 (no escape) to 1 (total escape). Unfortunately, however, this measure depends on the antibody concentration used, as well as the potency of the antibody against its inducing variant (schematically illustrated in **Suppl. Fig. S4**). However, since utilized antibody concentrations were reported in the original studies [23, 28], we could convert 'escape fractions' to 'fold resistances' $0 \leq \text{FR} \leq \infty$, due to mutational differences in antibody inducing x and antigen-presenting 'Spike-pseudo-groups' y (**Fig. 2A** and *Methods*). Intuitively, fold resistance FR values denote: 'how much more antibody is needed to neutralize the same proportion of mutant virus y, relative to the antibody-inducing strain x?'. A comparison between 'fold resistances' obtained through our method from DMS data with 'fold resistances' obtained through neutralization assays yields overall reasonable agreement (considering the methodological differences in the two experimental methods), **Suppl. Fig. S5**. Next, we computed a fold resistance FR for each of the 10 + 1 epitope classes and for each pair of 'antibody inducing' x and cross-neutralized 'Spike-pseudo-group' y, based on their pairwise mutational differences in sites relevant to each epitope class. Computed fold resistances for three representative epitope classes and a set of relevant omicron 'Spike-pseudo-groups' are shown in **Fig. 2B**. Corresponding epitopes are superimposed onto the Spike RBD structure in **Fig. 2C**. Interestingly, the fold resistances between the Omicron 'Spike-pseudo-groups' in **Fig. 2B** highlight a consecutively increasing fold resistance of succeeding variants relative to their temporal predecessors (compare **Fig. 1B**).

Immune waning and infection risk reduction

While cross-neutralization potency can qualitatively describe the antigenic overlap between variants, the ability to neutralize the virus (and prevent infection) will ultimately depend on the neutralizing antibody concentrations at the time of viral re-exposure [30].

Neutralizing antibody levels rise within 1 to 2 weeks after initial antigen exposure and slowly decay afterwards. We parameterized a simple pharmacokinetic model to capture the antibody concentration-time profile after initial antigen exposure as shown in **Fig. 3A** (details in *Methods*). Then, based on DMS data[2], we computed a relative weighing of antibodies belonging to different epitope classes as shown in **Fig. 3B**. This data-derived weighing may reflect the accessibility of the distinct epitopes, the strength of binding or the capability to neutralize the virus, when an antibody is bound. Our analysis showed that antibodies of epitope classes A, B, D2, and F3 were more potently neutralizing the virus, whereas antibodies of classes E3 and F1 were less potent.

Based on the previously computed cross-neutralization potency, antibody pharmacokinetics and relative antibody potencies, we subsequently estimated the probability of neutralizing the Delta variant following exposure to the Wuhan-variant antigen (genetic profiles in **Suppl. Table ST5**) after antigen exposure with the Wuhan-variant antigen (**Fig. 3C-D**; data sources in **Suppl. Table ST6**). For this task, we estimated the only remaining free parameter (the average normalized antibody potency $\widehat{\mathrm{IC50_x}}$) and assumed that neutralization probability approximates infection risk reduction.

Notably, utilized clinical data varies considerably with regards to the level of reported risk reduction due to statistical limitations (few observed 'infection events'), as well as heterogeneity in the study groups, which is a well-known phenomenon for prevention trials [31]. Regarding immune waning dynamics after exposure to the Wuhan-variant antigen, we see that, after antigen exposure, Delta can initially be almost completely neutralized. However, neutralization probabilities decrease to 50%, depending on individual antibody pharmacokinetics within 100 to 250 days after exposure to the antigen (**Fig. 3C**).

We then applied the previously estimated parameters in our model to predict Omicron neutralization after Wuhan-variant antigen exposure. Our predictions align well with vaccine efficacy data (summarized in **Suppl. Table ST7**). Our model predicts that Omicron (genetic profile in **Suppl. Table ST5**) is initially neutralized with 45–85% probability approximately two weeks after exposure to the Wuhan-variant antigen, with neutralizing immunity decaying rapidly and reaching about 10% infection risk reduction between 80 and 350 days post antigen exposure.

Lineage-resolved population immunity predicts variant dynamics

Notably, with all parameters of our underlying model being derived at this stage, we may now predict the individual infection risk reduction with regards to an arbitrary variant y for a person who was recently exposed to the Spike protein of an arbitrary variant x, e.g., through

[2] `https://media.githubusercontent.com/media/jbloomlab/SARS2_RBD_Ab_escape_maps/main/processed_data/escape_data.csv`

infection (**Fig. 4A**). Having a tool to estimate variant-resolved individual infection risk reduction over time, we can now ask: Will a novel SARS-CoV-2 variant spread? Whether or not the variant will spread is determined by its *relative fitness*, which, if SARS-CoV-2 evolution is driven by population immunity, is determined by the expected number of individuals susceptible to the variant. To test this hypothesis, we computed the expected number of susceptibles for each variant by integrating infection history in Germany, cross-neutralization, and immune waning (**Fig. 1A**, **Fig. 4A**, *Methods*). We then calculated the relative fitness of a variant as the number of susceptibles for the variant in relation to the average number of susceptibles across all currently circulating variants. In other words, based on immunity, can the variant infect more (or fewer) individuals than the current viral population? We calculated this immunity-driven relative fitness from our model and compared the predictions with historical changes in variant frequencies, **Fig. 4B**. We found an astonishing match between our predictions and the real-world data: Our model predicts the BA.2 infection point in April–May 2022, for the BA.4/BA.5 pseudo-group between July–October 2022, and for the more recent variants BQ.1.1, CH.1.1, and XBB.1.5 in January, February and April 2023, respectively (compare also **Fig. 1B**). Moreover, the data-derived change in variant frequency and our model-predicted immunity-driven relative fitness correspond in magnitude (see *Methods* for a theoretical justification). Also, our simulations highlight that any of the analyzed variants would have had a fitness advantage weeks before they were actually detected in substantial numbers in Germany.

We next evaluated whether our model could forecast variant dynamics: We used our data set, which ends on April 16, 2023, and evaluated the relative growth advantage of currently circulating and closely monitored variants XBB.1.9, XBB.1.16, and EG.5.1 due to variant-specific immunity. Next to these predictions, we plotted the actual variant dynamics observed after April 16, 2023 in Germany, **Fig. 4C**. Our predictions indicate that XBB.1.9 has a slight growth disadvantage, XBB.1.16 a slight growth advantage, and EG.5.1 a substantial growth advantage by mid-April 2023. Compellingly, these forecasts precisely match the actual variant growth dynamics in Germany during April–July 2023: The proportion of XBB.1.9 declined from 20 to 5%, XBB.1.16 slightly increased from 2 to 10%, and EG.5.1 substantially increased from $< 1\%$ to $> 30\%$.

Regional infection history determines success of a lineage

Next, we wanted to investigate why particular lineages, which dominated in other regions of the world, did not spread in Germany. For example, BA.2.12.1 dominated in the US, with variant proportions $> 50\%$ between the beginning of May 2022 to mid-June 2022, and BE.10 reached $> 20\%$ between the beginning of November 2022 to the beginning of January 2023 (orange lines in **Fig. 5A–B**). None of these lineages achieved comparable dominance in Germany (red lines in **Fig. 5A–B**). To test our hypothesis that infection history and variant-specific immunity may have limited the success of these lineages in Germany, we used our model to calculate their immunity-driven relative fitness, which equals the relative number of susceptibles (green areas in **Fig. 5A–B**). For both lineages we observed that their relative fitness was already declining when they were imported in Germany (mid-May 2022 for BA.2.12.1 and mid-Oct. 2022 for BE.10). Shortly after importation, their relative fitness reached the threshold of zero, which predicts that the invading variants would be less fit than the average viral population in Germany, and subsequently decline. Again, our model predictions for the dynamics of BA.2.12.1 and BE.10 in Germany precisely match their data-derived variant dynamics and suggest that the two lineages did not spread in Germany because they entered the country 'too late': For BA.2.12.1, by the time it entered Germany, the BA.2 wave (March–June 2022; compare **Fig. 1B;D**) had created substantial cross-neutralizing immunity, whereas, for BE.10, a sub-lineage of BA.5.3.1, the preceding BA.5 wave had created cross-neutralizing immunity (July–Oct 2022; compare **Fig. 1B;D**).

Population immunity predicts start and end of an infection wave

Finally, we wanted to assess whether our model could be used to estimate the start and end of infection waves. Intuitively, the more susceptibles present in a population, the larger an infection wave may become. Conversely, if too few individuals are susceptible to a virus variant, its reproductive number may decrease below the critical threshold and infection numbers decrease too. To test this hypothesis, we computed the expected number of susceptibles averaged over all currently circulating variants (see *Methods*). We found that inflection points of infection waves corresponded to a change in sign of the second derivative of the variant-averaged number of susceptibles, **Fig. 5C**. In other words: When the increase or decrease in the number of susceptibles decelerates, we observe an inflection point in the infection wave. However, we also observe that with increasing diversification of the viral population (compare **Fig. 1B**), our predictions become more uncertain and hence it becomes more difficult to determine the corresponding change points.

In summary, our analysis highlights that (i) the evolution of SARS-CoV-2 is predominantly driven by escape from neutralizing immunity, (ii) that the proposed model may predict neutralizing immunity in the population, and (iii) that our model could be used as a variant-alert system that can be tailored to the specific infection history of a given region.

Discussion

SARS-CoV-2 continues to present a serious public health concern, although case fatalities have markedly declined since the beginning of the pandemic. On the one hand, it is believed that compared to Omicron lineages, early SARS-CoV-2 variants displayed a broader cell tropism, allowing them to infect cells of the lower respiratory tract and cause severe illness [32, 33]. On the other hand, previous exposure to SARS-CoV-2 antigens, through infection or vaccination substantially lowers the risk for severe disease [34, 35]. Along these lines, it was estimated that the first year of vaccination saved approximately 20 million lives [36]. Exposure to SARS-CoV-2 triggers adaptive immune responses, which, in simplified terms, denote cellular immune response (CD4+ and CD8+ T-cell responses) and humoral immune response (B-cell-associated antibody production) [37, 38]. Cellular immune responses are believed to help resolve infection and prevent severe disease [39, 40]. With at least 2,000 unique CD4 and CD8 T-cell epitopes on the spike (S), nucleocapsid (N), matrix (M), envelope (E), and non-structural (NSP) proteins [41], the genetic barrier to the development of viral resistance against cellular immune responses may be insurmountable. On the other hand, antibodies target viruses or virus components outside cells. Among the produced antibodies, only certain spike-targeting antibodies are considered virus-neutralizing [42]. In our work, we focused on the group of virus-neutralizing antibodies targeting the S protein, given that the ongoing evolution of SARS-CoV-2 is predominantly occurring in epitopes of neutralizing antibodies (RBD and NTD), which suggests that neutralizing antibodies targeting RBD and NTD are the major component limiting SARS-CoV-2 transmission to date. However, on a technical note, a limitation of our study is that DMS data is only available for RBD-targeting antibodies and not for the NTD [22, 24]. To overcome this limitation, we included an additional class of NTD-targeting antibodies targeting three antigenic super-sites [29], as outlined in the *Methods* section.

Overall, our presumption that neutralizing antibodies against RBD and NTD are the dominating factor that differentiate SARS-CoV-2 variants allowed us to model infection- and variant dynamics based on infection history, and cross-neutralization based on RBD and NTD epitopes. Based on infection history, this would ultimately allow us to determine mutations that evade neutralization and may increase transmission rates. Notably, a related approach [43] estimated the fitness effects of mutations through their deviation from neutral mutation rates. A key advantage of our approach is that we explicitly consider the dynamic immune landscape; thus, the immunological fitness effect of a mutation may change over time, depending on the specific infection history of a region of interest. While this allows us to precisely estimate variant dynamics, a limitation of our study is that, disregarding cellular immune responses, we cannot predict case severity or the number of hospitalizations. Moreover, our model is not suitable for estimating an absolute protective antibody titer against SARS-CoV-2 infection because our model operates with unitless antibody concentrations (compare **Fig. 3A**).

Our analysis did not include seasonality because it is irrelevant for determining the *relative* variant-specific growth rate γ_y (it cancels out in the computation). However, if seasonality was a main driver of SARS-CoV-2 incidence, it would affect the size of different epidemic waves as depicted in **Fig. 5C**. Interestingly, there is still no scientific consensus regarding the seasonality of SARS-CoV-2 incidence [44]: Some studies have reported associations between climate and SARS-CoV-2 incidence [45, 46, 47, 48, 49], while others failed to show a role of environmental factors [50, 51, 52]. Importantly, some previous analyses may be obscured by case reporting being lowest in the summer month [26]. Interestingly, an early study of influenza and endemic coronaviruses suggested that the availability of susceptibles may be a dominant driver for infection incidence [53], which would support our findings presented in **Fig. 5C**.

Furthermore, we did not include the vaccination timeline because (i) Wuhan-variant vaccines have little cross-neutralization ability towards the predominant Omicron lineages that emerged during the prediction horizon (**Fig. 3D**), and (ii) infections far out-weighed the number of vaccinations throughout the time horizon in our study. Updated (second-)booster-vaccines (Wuhan + BA.4/5) became only available during autumn/winter 2022 in Germany (after the BA.5 wave June–Oct. 2022; **Fig. 1B**), with limited uptake in the population[3]. While the vaccination campaign may have prevented many severe cases, we speculate that the impact on the total number of infections may have been limited. Interestingly, the predicted BA.4/5 spike pseudo-group dynamics have been accurately predicted without considering the vaccination timeline, which is explained by the negligible amount of vaccinations with the updated vaccine (<1 million doses) in comparison to ≈22.3 million BA.4 and BA.5 infections.

Lastly, we did not include any intrinsic, variant-specific replication advantages conferred by receptor binding [54], cell tropism [32], or escape from innate immune response [55, 56]. While these parameters are unknown for most SARS-CoV-2 variants, lineage-specific replication parameters could be regarded when computing the *relative* variant growth rate γ_y. For example, the growth advantage at the end of our forecast horizon for variant BA.2.86 'Pirola', recently defined as a variant under monitoring by the WHO, would be $\gamma_{\mathrm{BA.2.86}} = 0.095$ (CI: 0.064,

[3] `https://impfdashboard.de/`

0.127) due to population immunity. Consequently, the critical threshold of the relative intrinsic growth rate would be $\alpha_{\mathrm{BA.2.86}}/\sum_x \pi_x \cdot \alpha_x = (\gamma_{\mathrm{BA.2.86}} + 1)^{-1} = 0.913$ (CI: 0.887, 0.939) for this variant to spread successfully through the population.

In summary, we present a novel and innovative mathematical approach that can accurately predict variant dynamics from a model-inferred dynamic immune landscape, utilizing DMS, incidence, and molecular surveillance data. An implementation is freely available [57] and can be adapted to different settings to allow contextualizing the risk of occurrence and spread of variants at the country level, which is a major advancement of current warning systems. In the future, our approach could serve as a basis to identify epitopes most likely to be under recent selective pressure and therefore provide cues for designing vaccine candidates that avoid targeting those epitopes, thus maximizing neutralization breadth against emerging variants in a forthcoming season. Furthermore, our conceptional ideas may be transferred to model the evolution of other respiratory viruses that are subject to molecular surveillance.

Figures

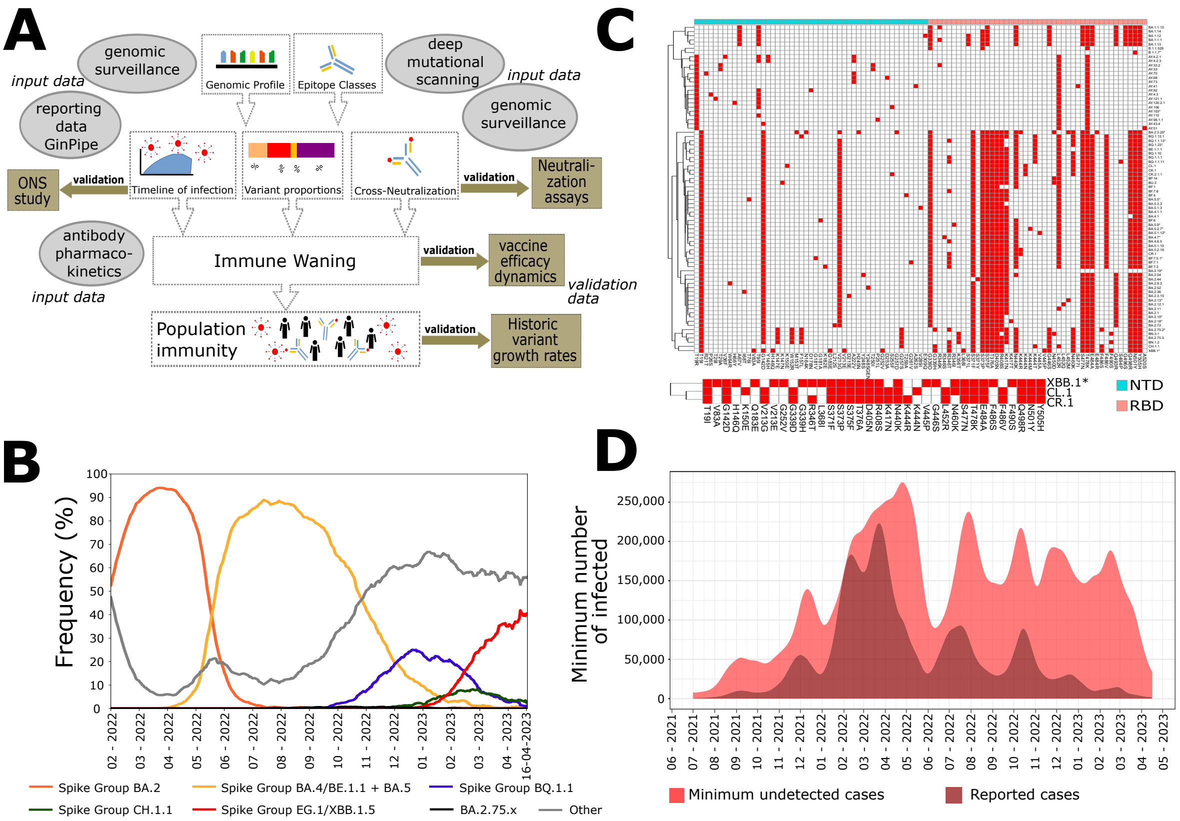

Figure 1: **Overview of data sources, data integration, and validation steps for modeling variant-specific immunity.** A. Overview. B. Variant dynamics of BA.2*, BA.4/5*, BQ.1.1.*, CH.1.1*, EG.1/XBB.1.5*, and BA.2.75* 'spike-pseudo-groups'. C. Mutation profiles of the 78 most abundant 'spike-pseudo-groups' in the NTD and RBD region of the spike protein. D. Reconstructed infection timeline in Germany using GInPipe, i.e., the minimum number of actual cases $I_{\min}(t)$. Dark red: reported cases $I_{\mathrm{rep}}(t)$; light red: undetected cases.

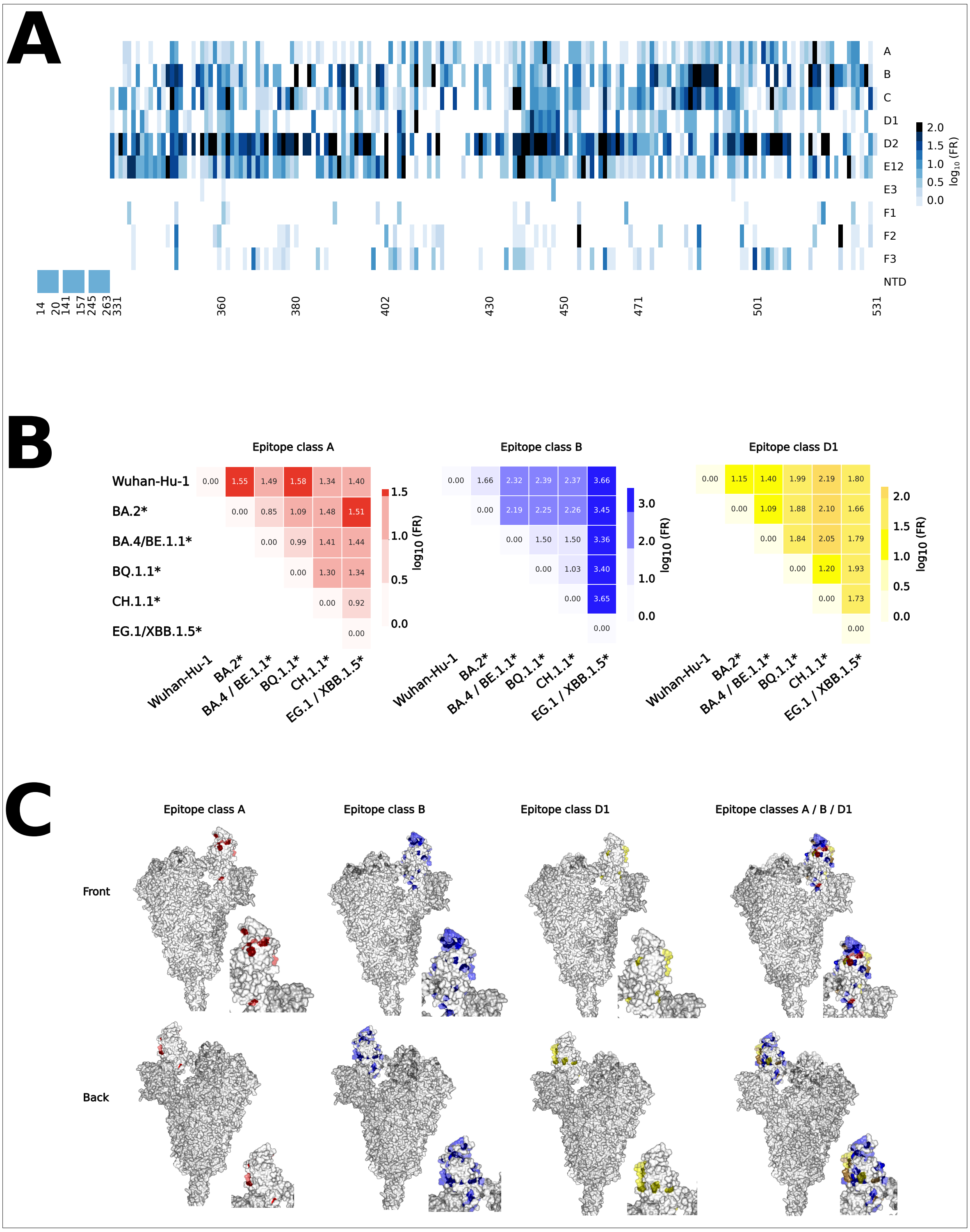

Figure 2: **Epitope classes and variant cross-neutralization.** A. Fold resistance against antibodies of epitope classes $\vartheta = \{A, B, C, D1, D2, E12, E3, F1, F2, F3, NTD\}$ induced by mutational changes at indicated sites. B. Fold resistance to neutralization against immunity-inducing variants (on y-axis) for antibodies of epitope classes A, B, and D1. C. Site-specific fold resistances superimposed on spike RBD structure for antibodies of epitope classes A, B, and D1.

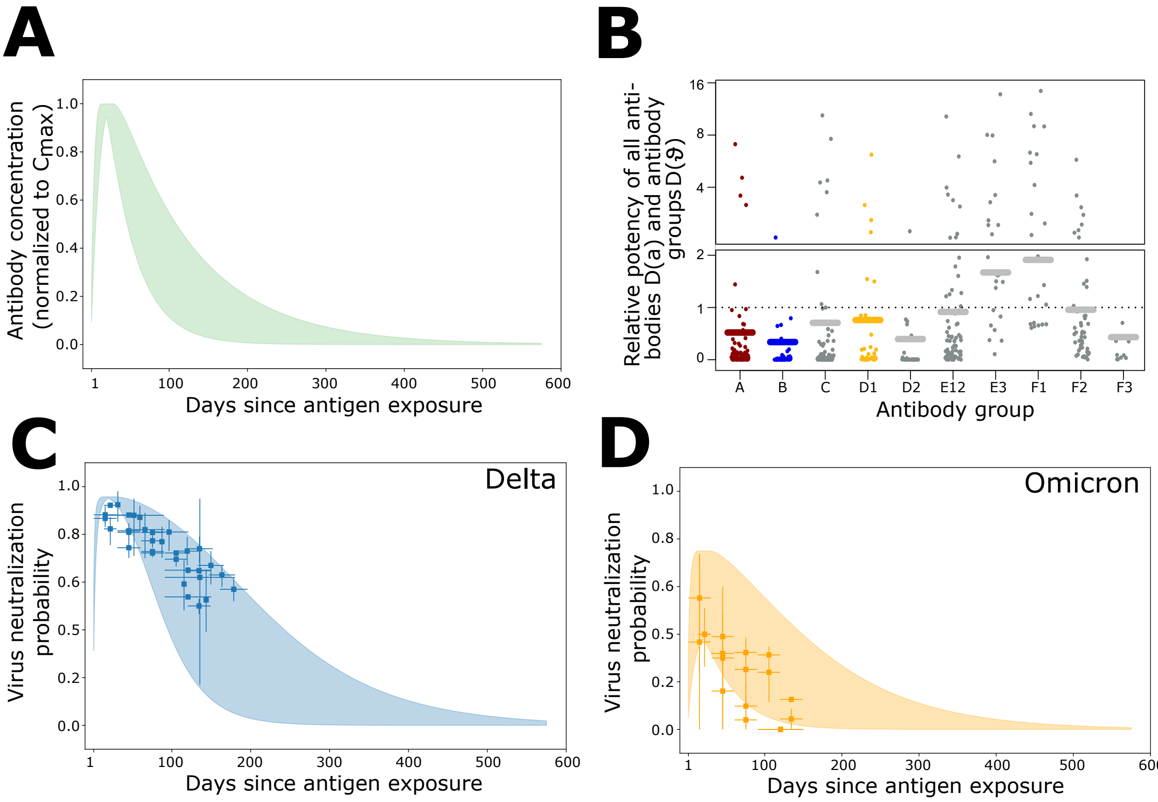

Figure 3: **Immune waning dynamics.** A. Normalized antibody pharmacokinetics after antigen exposure. B. Relative potency $\mathrm{IC50_{DMS}}$ of antibodies of epitope classes $\vartheta = \{A, B, C, D1, D2, E12, E3, F1, F2, F3\}$ in DMS data. The solid horizontal line shows the average $\mathrm{IC50_{DMS}}$. C. Predicted (blue range) neutralization probability of the Delta variant after exposure to the Wuhan antigen $P_{\mathrm{Neut}}(t, \mathrm{Wuhan}, \mathrm{Delta})$ and corresponding clinical vaccine efficacy (blue markers; original data in **Suppl. Table ST6**). D. Predicted (orange range) neutralization probability of the Omicron variant after exposure to the Wuhan antigen $P_{\mathrm{Neut}}(t, \mathrm{Wuhan}, \mathrm{Delta})$ and corresponding clinical vaccine efficacy (orange markers; original data in **Suppl. Table ST7**).

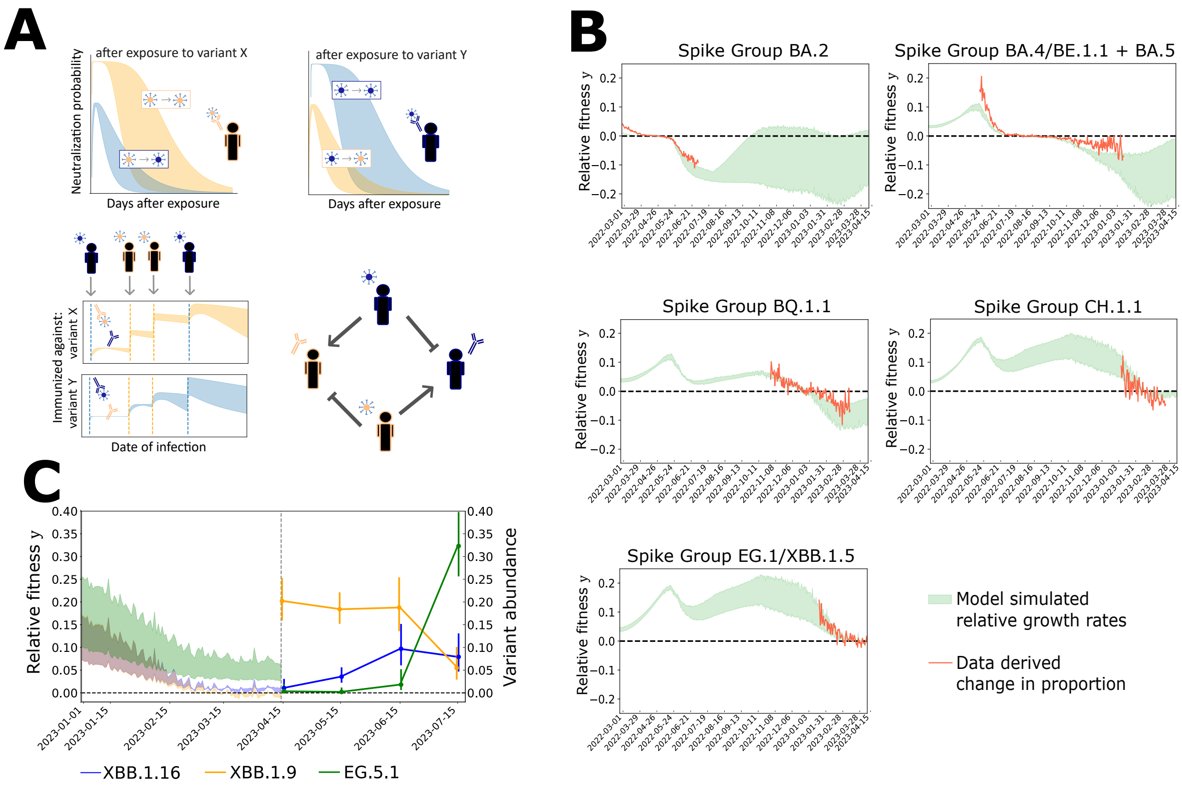

Figure 4: **Population immunity and variant dynamics.** A. Schematic for calculating population immunity: For each individual, we can compute the probability to cross-neutralize a variant y after exposure to variant x over time. Taking the infection history into account, we can calculate the expected number of individuals immune to any variant y. B. Historical variant dynamics. Model-predicted relative growth advantage $\gamma_y(t)$ of 'spike-pseudo-groups' BA.2*, BA.4/5*, BQ.1.1.*, CH.1.1*, and EG.1/XBB.1.5* (green range), and data-derived change in frequency $\left(\frac{\pi_y(t+1)}{\pi_y(t)} - 1\right)$ (red line). C. Prospective variant dynamics. Model-predicted relative growth advantage γ_y of emerging 'spike-pseudo-groups' XBB.1.9, XBB.1.16, and EG.5.1 (yellow, blue, and green ranges) until April 2023. Data-derived frequencies π_y of 'spike-pseudo-groups' XBB.1.9, XBB.1.16, and EG.5.1 from April-July 2023. Confidence intervals (95%) were calculated using Wilson's method.

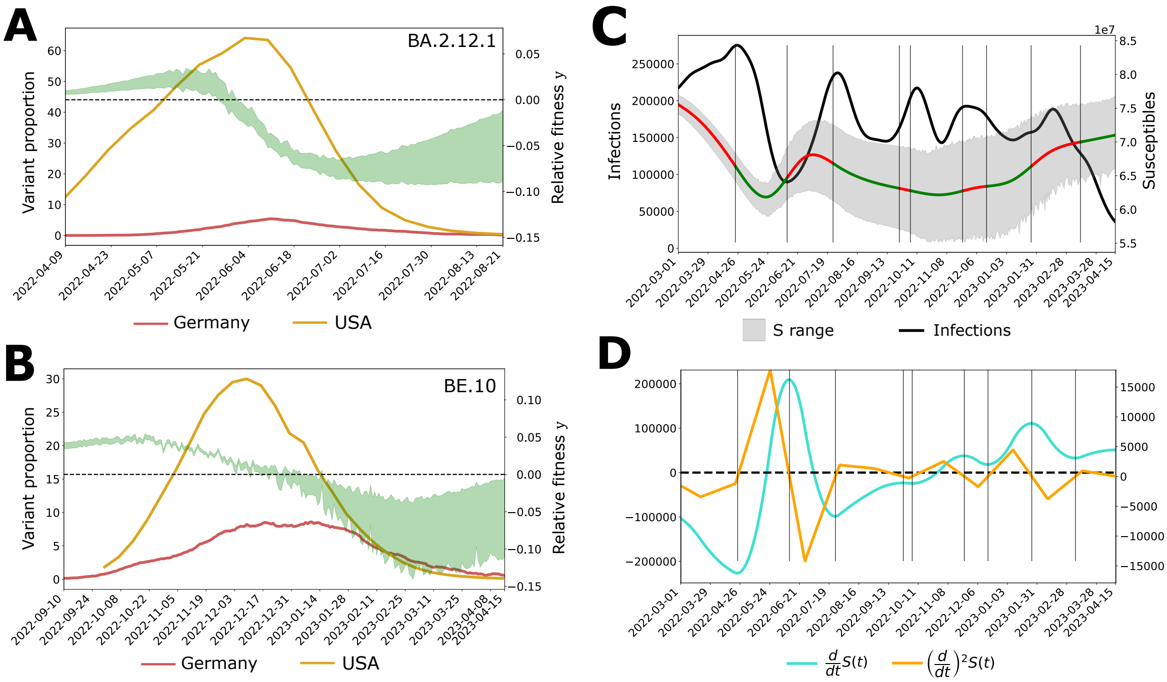

Figure 5: **Impact of regional infection history on variant dynamics and prediction of infection waves.** A. Dynamics of BA.2.12.1 in the US (orange line) vs. Germany (red line) and model-predicted number of BA.2.12.1 susceptible individuals in Germany (green area). B. Dynamics of BE.10 in the US (orange line) vs. Germany (red line) and model-predicted number of BE.10 susceptible individuals in Germany (green area). C. Prediction of infection waves. When superimposing the model-predicted number of susceptibles (grey ranges) and the minimum number of infections per day $I_{\text{min}}(t)$(black line), we realized that the peak and trough of the infection dynamics coincided with the second derivative of the model-predicted number of susceptibles. I.e., whenever the absolute velocity of the number of susceptibles decreases, we observe an infection peak or trough.

Methods

SARS-CoV-2 genomic data

We collected SARS-CoV-2 genomes from Germany published via the Robert Koch Institute[4] and extracted their mutation profiles using covSonar[5], a database-driven system for handling genomic sequences of SARS-CoV-2 and screening genomic profiles. To ensure representative sampling, we only included genomes from the 'random sampling' strategy, which denotes the majority of all sequence data. In total, we collected ≈600,000 SARS-CoV-2 genomes for the time period of July 1, 2021 to April 16, 2023, which were subsequently used to infer variant-specific population from March 01, 2022 to April 16, 2023.

Variant proportions and 'Spike-pseudo-groups'

SARS-CoV-2 genomes were assigned to 1,205 pangolin lineages using established methods [58, 59]. For each lineage, we collected all 'characteristic' spike amino acid (aa) changes for subsequent analyses; in our work 'characteristic' implied that an aa change was present in at least 75% of all sequences from that lineage[6] (**Suppl. Table ST1**). We then determined the genomic profile in the spike protein for all 1,205 SARS-CoV-2 pangolin lineages and found 887 mutations in Spike, with 280 distinct mutations in the RBD region (aa pos 318–541) and 314 in the NTD region (aa pos 13–305). The 'antigenic profile' for each lineage was then determined based on the set of unique mutations within the NTD and RBD regions. Differences between lineages were defined as the set difference between mutation profiles. Clustering lineages with identical 'antigenic profiles' yielded 651 'Spike-pseudo-groups' with distinct genomic profiles in the RBD and NTD region of the spike protein (**Suppl. Table ST2**). Based on the genomic profiles and their clustering into epitope-pseudo-groups, we computed pseudo-group frequencies $\pi_x(t)$ for all Spike-pseudo-groups $x \in \mathcal{X}$ in the entire observation horizon.

Genome-based reconstruction of infection timeline

To correct for under-reporting of infected individuals, we reconstructed the actual SARS-CoV-2 infection timeline in Germany for the entire prediction time horizon using the Genome-based Incidence reconstruction Pipeline (GInPipe) [26].

The sequences were pooled into consecutive temporal bins according to their collection date, such that bins b had either the same size $n(b)$ or spanned the same number of days $\Delta d(b)$. As parameters, we chose time spans of $\Delta d(b) = 7, 14$, and 21 days and bin sizes of $n(b) =$ 3,000, 5,000, 10,000, and 15,000 sequences.

For each of the bins $\phi_t(b) \approx c \cdot I(t)$ was calculated, which is a correlate of the actual number of infections $I(t)$ [26]. Here, we filtered the bin-wise incidence correlates $\phi_t(b)$ if the time span of a bin was smaller than 7 days, or if a bin comprised less than 1,000 sequences. Bin-wise $\phi_t(b)$ estimates were smoothed using kernel smoothing with a bandwidth of 21.

For approximating the actual case numbers, we inferred the time-dependent case ascertain-

[4] `https://github.com/robert-koch-institut/SARS-CoV-2-Sequenzdaten_aus_Deutschland`
[5] `https://github.com/rki-mf1/covsonar`
[6] `https://outbreak.info/situation-reports/methods#characteristic`

ment probabilities $P_{\text{rep}}(t) \leq 1$, i.e., the probability of an infection being reported:

$$P_{\text{rep}}(t) = \frac{I_{\text{rep}}(t)}{I(t)} \approx \frac{I_{\text{rep}}(t)}{\phi_t \cdot c}$$
$$\Leftrightarrow P_{\text{rep}}(t) \cdot c \approx \frac{I_{\text{rep}}(t)}{\phi_t},$$

where $I_{\text{rep}}(t) \leq I(t)$ denotes the daily reported infections (weekly cases/7)[7]. We smoothed the reported cases $I_{\text{rep}}(t)$ with a bandwidth of 14.
Lastly, we normalized the case ascertainment probabilities $\widetilde{P}_{\text{rep}}(t)$ at their maximum to be able to estimate the *minimum number of infections* $I_{\min}(t)$.

$$P_{\text{rep}}(t) \leq \widetilde{P}_{\text{rep}}(t) = \frac{P_{\text{rep}}(t)}{\max_t\left(P_{\text{rep}}(t)\right)}.$$

The minimum number of infections is then calculated as

$$I(t) \geq I_{\min}(t) = \frac{I_{\text{rep}}(t)}{\widetilde{P}_{\text{rep}}(t)} = \phi_t \max_t\left(\frac{I_{\text{rep}}(t)}{\phi_t}\right) \geq I_{\text{rep}}(t).$$

To confirm the validity of GInPipe-estimated incidences, we compared our genome-based incidence predictions with a UK data set from the representative COVID-19 Infection Survey of the Office for National Statistics (ONS) [2]

The ONS survey is based on large-scale random sampling strategy of households, based on which SARS-CoV-2 test-positivity and the number of currently infected individuals is estimated, overcoming under-reporting issues resulting from changing testing policies or -behaviour. In brief, data for the UK was available for the time frame of May 2021 to April 2022 from the officially reported cases[8] and sequences from GISAID, respectively.

We then used GInPipe with the UK sequencing and reporting data to derive a minimum number of infections in the UK. Applying a right-sided rolling sum over 10 days over the minimum incidences approximates PCR-positivity from these numbers [60]. For comparison, we depict GInPipe's estimates side by side with the number of currently infected individuals from the ONS study in **Suppl. Fig. S3**, indicating highly congruent predictions between the two methods.

Deep mutational scanning (DMS) data

To assess the phenotype (escape from neutralizing antibodies) of each 'Spike-pseudo-group', we used all DMS data [22, 24] available on Feb 13, 2023[9]. DMS measures, in a yeasts-display assay, how much a mutated site s in the RBD affects the binding of antibodies elicited by a variant that is not mutated at site s. We utilized data from 836 antibodies that were classified into 12 distinct epitope classes [22] [28] (see below) and aggregated all values on site-level by their mean, yielding 'escape fractions' to antibody a for each mutated site $\text{ef}_{s,a}$ (values were bounded at 0.99). Notably, 'escape fractions' denote a proxy for quantifying the probability

[7] https://github.com/robert-koch-institut/COVID-19_7-Tage-Inzidenz_in_Deutschland/blob/main/COVID-19-Faelle_7-Tage-Inzidenz_Deutschland.csv

[8] https://coronavirus.data.gov.uk/details/cases

[9] https://github.com/jbloomlab/SARS2_RBD_Ab_escape_maps/blob/main/processed_data/escape_data.csv

that an antibody does not bind an RBD that contains a mutation at site s. As such, the numerical value depends on the antibody potency and hence we aimed to convert these values to 'fold resistances' (fold change in antibody potency), see also **Suppl. Fig. S4**.
Assuming that mutational effects of sites s on the binding affinity are independent, the binding probability of an antibody elicited by a variant x to a variant y can be expressed as

$$b_a(x,y) = \prod_{s\in\Omega(x,y)} \left(1 - \mathrm{ef}_{s,a}\right), \tag{1}$$

where $\mathrm{ef}_{s,a}$ is the normalized escape fraction of mutated site s with respect to antibody a and $\Omega(x,y)$ denotes the set of RBD sites that distinguish variants x and y [22].
Using classic pharmacodynamic approaches, we then model the binding probability as

$$b_a(x,y) = \frac{c_a}{\mathrm{FR}_{x,y}(a)\cdot \mathrm{IC50}_{\mathrm{DMS}}(a) + c_a}, \tag{2}$$

where $\mathrm{FR}_{x,y}(a)$ denotes the fold resistance of variant y to an antibody a elicited by variant x. The parameter $\mathrm{IC50}_{\mathrm{DMS}}(a)$ corresponds to the potency of the antibody against the antibody-eliciting variant (we used wild-type data), which was extracted from the DMS dataset. Notably, c_a =400 µg/mL was the antibody concentration at which the DMS experiment was conducted [23, 28]. Combining eqs. (1)–(2) yields:

$$\mathrm{FR}_{x,y}(a) = \frac{c_a}{\mathrm{IC50}_{\mathrm{DMS}}(a)}\left(\frac{1}{b_a(x,y)} - 1\right).$$

Notably, as already evident from **Suppl. Fig. S4**, DMS estimates of $\mathrm{ef}_{s,a}$ as well as corresponding $\mathrm{FR}_{x,y}(a)$ can become unreliable depending on antibody concentrations and antibody potency $\mathrm{IC50}_{\mathrm{x}}(a)$, falsely predicting hyper-susceptibility ($\mathrm{FR}_{x,y}(a) < 1$). We enforced that $\mathrm{FR}_{x,y}(a) \geq 1$ to avoid such artifacts.

Epitope classes

Based on the similarity of antibody profiles in DMS data, antibodies were previously classified into 12 epitope classes {A, B, C, D1, D2, E1, E2.1, E2.2, E3, F1, F2, F3} [28] (**Suppl. Tables ST3–ST4**). As we encountered more than 30% missing values for epitope classes E2.1 and E2.2, we merged them with E1 into a new class (E12), as they bind to similar regions including aa site R346 [61]. Finally, we retrieved a matrix of 10 epitope classes $\mathcal{A}$ = {A, B, C, D1, D2, E12, E3, F1, F2, F3} for 179 RBD sites. This classification indicates that antibodies belonging to the same class bind to overlapping epitopes and there is little overlap between epitope classes (**Fig. 2A**). Consequently, we assumed that antibodies within the same epitope class would be similarly affected by RBD mutations, whereas phenotypic changes between epitope classes may vary considerably.
We then quantified the fold resistance associated with each epitope class as the average fold resistance of all antibodies belonging to the class:

$$\mathrm{FR}_{x,y}(\vartheta) = \mathrm{mean}\{\mathrm{FR}_{x,y}(a) : a \in \vartheta\}$$

for all epitope classes ϑ in $\mathcal{A}$ (**Fig. 2A**).
As a proof of concept, we compared DMS-derived fold resistances $\mathrm{FR}_{x,y}(\vartheta)$ using the calculations above with fold resistance values obtained from virus neutralization assays (reported in

Cao et al. [61]) for antibodies targeting all epitope classes $\mathcal{A}$ defined above. As can be seen in **Suppl. Fig. S5**, we observed a strong and significant positive correlation between the DMS-derived fold resistances $\mathrm{FR}_{x,y}(\vartheta)$ and those obtained by neutralization assays.

Since the DMS experiments were generated for RBD-targeting antibodies only, no escape data was available to quantify the fold resistance of NTD-targeting antibodies. To overcome this limitation, we included an additional class of NTD-targeting antibodies targeting three antigenic super-sites [29]: Spike aa positions 14–20, 140–158, and 245–264. Consequently, we assigned mutations in the antigenic super-sites fold resistance (FR) values of 10, which is in range with corresponding ELISA experiments [29]. However, the model can be updated if comprehensive DMS data for the NTD domain becomes available [62].
Assuming independence between mutational effects, the total fold resistance of a variant y to binding of an NTD targeting antibody elicited by a variant x was computed as:

$$\mathrm{FR}_{x,y}(\mathrm{NTD}) = 10^{|\Omega(x,y)|},$$

where $|\Omega(x,y)|$ denotes the number of mutational differences between variants x and y in the antigenic super-site of the NTD.

Variant cross-neutralization probability

We assumed that a virus is neutralized if *at least* one antibody is bound to its surface (either at the RBD or NTD of the spike protein). Here, we collectively regard all antibodies from the same epitope class as they compete for the same binding site. By assuming binding independence *between* epitope classes, the neutralization probability can be computed as:

$$P_{\mathrm{Neut}}(t,x,y) = 1 - \prod_{\vartheta \in \mathcal{A}_{x \setminus y}} \left(1 - b_\vartheta(t,x,y)\right),$$

with $b_\vartheta(t,x,y)$ denoting the probability that an antibody of epitope class ϑ in $\mathcal{A} \cup \{\mathrm{NTD}\}$ binds to the virus with

$$b_\vartheta(t,x,y) = \frac{c_\vartheta(t)}{\mathrm{FR}_{x,y}(\vartheta) \cdot \mathrm{IC50}_{\mathrm{x}}(\vartheta) + c_\vartheta(t)},$$

where $c_\vartheta(t)$ is the antibody's concentration in an individual at time t, $\mathrm{IC50}_{\mathrm{x}}(\vartheta)$ is the half-maximal inhibitory antibody concentration against the variant that elicited the antibody. $\mathrm{FR}_{x,y}(\vartheta)$ is the fold resistance of variant y to binding of antibodies of epitope class ϑ, elicited by variant x.

Antibody potency

Next, we quantified $\mathrm{IC50}_{\mathrm{x}}(\vartheta)$ for each epitope class. Since the DMS data was derived from yeast-display RBD mutant libraries, *absolute* antibody potencies may not directly translate to a clinical setting. However, the ranking of antibody potencies may be preserved. Consequently, we estimated the *relative* potency $D(\vartheta)$ from the DMS data[10]:

$$D(\vartheta) = \frac{\widehat{\mathrm{IC50}}_{\mathrm{DMS}}(\vartheta)}{\frac{1}{|\mathcal{A}|}\sum_{\varsigma \in \mathcal{A}} \widehat{\mathrm{IC50}}_{\mathrm{DMS}}(\varsigma)},$$

[10] https://media.githubusercontent.com/media/jbloomlab/SARS2_RBD_Ab_escape_maps/main/processed_data/escape_data.csv

where $\vartheta \in \mathcal{A}$ and $\widehat{\mathrm{IC50}_{\mathrm{DMS}}}(\vartheta)$ denotes the average potency of all antibodies belonging to epitope class ϑ. Epitope-class specific clinical antibody potency $\mathrm{IC50_x}(\vartheta)$ were then inferred using the following relation

$$\mathrm{IC50_x}(\vartheta) = D(\vartheta) \cdot \widehat{\mathrm{IC50_x}},$$

where $\widehat{\mathrm{IC50_x}}$ denotes the $\mathrm{IC50_x}$ averaged over all epitope classes. NTD-targeting antibodies were not included in the DMS data set and hence we set

$$\mathrm{IC50_x}(\mathrm{NTD}) = \widehat{\mathrm{IC50_x}}.$$

Notably, $\widehat{\mathrm{IC50_x}}$ was the only free parameter in the model, which we estimated by fitting our model to (Wuhan-strain) vaccine efficacy (VE) data against the Delta lineage (B.1.617.2) present between the period of July 4, 2021 till December 31, 2021 (Fig. 3C; genomic profile in Suppl. Table ST5).

We considered inter-individual differences in antibody pharmacokinetics (see below), implemented as combinations of parameters t_{max} (time of maximal antibody concentrations) and t_{half} (antibody half-life). For parameter estimation, we first estimated optimal drug potencies $\widehat{\mathrm{IC50_x}}(t_{\mathrm{max}}, t_{\mathrm{half}})$ for each $t_{\mathrm{max}}, t_{\mathrm{half}}$ combination in a 5×15 grid (ranges below) and then averaged over these 75 estimates. Parameter estimation was done using scipy.optimize.root, applying the Levenberg-Marquardt (lm) method to solve the ordinary least-square problem

$$\underset{\widehat{\mathrm{IC50_x}}(t_{\mathrm{max}},t_{\mathrm{half}})}{\operatorname{argmin}} \sum_t \left\| P_{\mathrm{Neut}}(t, \mathrm{Wuhan}, \mathrm{Delta}, \widehat{\mathrm{IC50_x}}(t_{\mathrm{max}}, t_{\mathrm{half}})) - VE(t, \mathrm{Wuhan}, \mathrm{Delta}) \right\|^2.$$

where $VE(t, \mathrm{Wuhan}, \mathrm{Delta})$ denotes the vaccine efficacy against the Delta strain t days after antigen exposure with the Wuhan strain. Here, we assumed that VE = infection risk reduction $\approx P_{\mathrm{Neut}}$.

As a proof of concept, we then tested our predictions with Wuhan-strain VE data against Omicron infection, Fig. 3D (genomic profile in **Suppl. Table ST5**). Utilized VE data include all studies where Wuhan-strain vaccines were tested and which were computed based on hazard ratios or rate of confirmed infection, in **Suppl. Tables ST6–ST7**.

Antibody pharmacokinetics

To determine the duration of sterilizing immunity against any variant y we accounted for antibody pharmacokinetics (PK) after antigen exposure to strain x (by means of infection or vaccination). Pharmacokinetics were considered using a classical, descriptive linear model with analytical solution

$$c(t) = \frac{e^{-k_e t} - e^{-k_a t}}{e^{-k_e t_{\mathrm{max}}} - e^{-k_a t_{\mathrm{max}}}},$$

where t denotes the time after antigen exposure and $c(t)$ denotes the normalized (fraction of maximum) concentrations of the antibody. Parameters k_e and k_a were related to known quantities through established PK relations, i.e.:

$$k_e = \frac{\ln(2)}{t_{\mathrm{half}}}, \qquad t_{\mathrm{max}} = \frac{\ln\left(\frac{k_a}{k_e}\right)}{k_a - k_e}.$$

In our simulations, we considered identical PK for antibodies of the different epitope classes. Utilized parameters ($t_{\max}, t_{\text{half}}$) were extracted from literature: $t_{\max}$ varied between 14 and 28 days after antigen exposure [63, 38, 64, 65], while t_{half} ranged between 25 and 69 days [66, 67, 64, 30, 68, 69, 70, 71]. For simulation, we used different combinations of ($t_{\max}, t_{\text{half}}$) in a 5×15 grid within a range of $t_{\max}$ within $[14, 28]$ and t_{half} within $[25, 69]$ and plotted the range of predictions (min, max).

Expected sterilizing immunity against variant y.

The estimation of cross-neutralization probability P_{Neut} enabled us to estimate the expected number of individuals being immune against infection with a variant y for a given time point t by taking the infection history prior to time t into account. The expected number of individuals immune to infection with strain y is given by

$$\mathbb{E}[\text{Immune}_y(t)] = \sum_{x \in \mathcal{X}} \int_{s<t} \pi_x(s) \cdot I(s) \cdot P_{\text{Neut}}(t-s, x, y)\, ds$$

where $\mathcal{X}$ denotes the set of all variants present within the time horizon of interest, $P_{\text{Neut}(t-s,x,y)}$ denotes the probability that an infection with strain x, that occurred $t - s$ days ago cross-neutralizes a variant y. In the equation above, $\pi_x(s)$ denotes the proportion of variant x at day $s < t$, derived from the molecular surveillance of SARS-CoV-2 (**Suppl. Data D1**), while $I(s)$ denotes the number of infected individuals at some previous time point s. As our best case for under-reporting-corrected infection numbers, we used the GInPipe-derived estimate $I_{\min}(s)$, as derived above.
The expected number of susceptible individuals to a variant y at time t is then calculated as $\mathbb{E}[S_y(t)] = Pop - \mathbb{E}[\text{Immune}_y(t)]$, where $Pop = 84.3 \cdot 10^6$ denotes the population size of Germany.

Variant dynamics

To estimate whether an emerging variant may successfully out-compete existing variants, we estimated the *relative* growth advantage of a variant $\gamma_y(t)$:

$$\gamma_y(t) = \frac{\alpha_y \cdot \mathbb{E}[S_y(t)] - \sum_{x \in \mathcal{X}} \pi_x(t) \cdot \alpha_x \cdot \mathbb{E}[S_x(t)]}{\sum_{x \in \mathcal{X}} \pi_x(t) \cdot \alpha_x \cdot \mathbb{E}[S_x(t)]}$$

where $\sum_{x \in \mathcal{X}} \pi_x(t) \cdot \alpha_x \cdot \mathbb{E}[S_x(t)]$ denotes the *average* growth rate across all variants existing at time t and where $\alpha_x > 0$ denotes a variants' intrinsic (antibody-independent) transmission fitness, which we assumed to be near identical for all currently circulating variants $\alpha_x \approx \alpha$, implying that variant dynamics are dominated by infection history and immune dynamics. We ignored low abundance variants with $\pi_x(t) < 1\%$ and re-normalized accordingly.

Relation to variant dynamics

In a discrete-time quasi-species model [72], the theoretical variant frequencies $\mathbf{p} \in [0, 1]^{|\mathcal{X}|}$ are given by

$$\mathbf{p}(t + 1) = \mathbf{Q} \cdot \mathbf{F}(t) \cdot \mathbf{p}(t)$$

where $\mathbf{Q} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ denotes a transition matrix between different variants, which we set to the identity matrix $\mathbf{Q} = \text{Id}$, ignoring any mutational transitions from one- to another variant.

Fitness values of any variant y, relative to the population average $\bar{f}(t) = \sum_x p_x(t) \cdot f_x(t)$ are contained in matrix $\mathbf{F}(t) = \text{diag}\left([\cdots, f_y(t)/\bar{f}(t), \cdots]\right)$. Consequently, for $p_y(t) > 0$, we get

$$\frac{p_y(t+1)}{p_y(t)} - 1 = \gamma_y(t),$$

if fitness is determined by population immunity.

Population susceptibility

The total number of susceptibles was calculated as

$$\mathbb{E}[S(t)] = \sum_{x \in \mathcal{X}} \pi_x(t) \cdot \mathbb{E}[S_x(t)].$$

Derivatives were calculated from cubic spline-interpolated $\mathbb{E}[S(t)]$ estimates.

Variant proportions in the US

The proportions of SARS-CoV-2 variants in the USA were obtained from the Centers for Disease Control and Prevention's website[11] on July 18, 2023. This report provides weekly estimations of the predominant SARS-CoV-2 variants, presented as proportions of sequences under surveillance. The analysis covers a time span from July 31, 2021, as the earliest period, and extends to May 6, 2023, marking the latest period in the report. Throughout this duration, a total of 44 distinct SARS-CoV-2 variants were identified.

Code availability

Codes were written in Python 3.11.3 and R version 4.2.3 (2023-03-15). Simulations were performed on the high-performance compute (HPC) cluster at ZEDAT, Freie Universität Berlin [73]. The pipeline for the genome-based incidence estimation (GInPipe) is available at `https://github.com/KleistLab/GInPipe`, version 3.0.0. All custom codes are available at `https://github.com/KleistLab/VASIL`, version 1.1 [57].

Author Contributions

Conceptualization, D-Y.O., M.v.K. ; Methodology, N.A.R., N.G., D.B., M.R.S., C.S., M.v.K.; Investigation, N.A.R., N.G., D.B., M.R.S., S.P.; Writing Original Draft, N.A.R., N.G., D.B., M.R.S., M.B., M.H., S.P., M.v.K.; Funding Acquisition, S.F., T.W., M.v.K.; Supervision, R.D., T.W., M.H., M.v.K.;

Acknowledgements

We are very grateful to the German Electronic Sequence Data Hub (DESH) and all of its data contributors, i.e., the authors from the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared. We further thank the Sequencing Core Facility of the Genome Competence Center (MF1) at Robert Koch Institute for providing excellent sequencing services, and we thank all

[11] `https://data.cdc.gov/Laboratory-Surveillance/SARS-CoV-2-Variant-Proportions/jr58-6ysp`

labs contributing to the German SARS-CoV-2 surveillance. We thank all international labs for contributing SARS-CoV-2 sequences to the GISAID EpiCoV database. We acknowledge high-performance computing services provided through ZEDAT at the FU Berlin and the Robert Koch Institute.

Funding

This work was supported by the German Ministry for Science and Education (BMBF) and the Ministry for Economic Affairs and Climate action (BMWK), grant numbers 01KI2016 and 01MK21009H and the German Science Foundation (DFG) through the excellence center MATH+ to M.v.K., as well as the European Center for Disease Prevention and Control, grant number 2021/008 ECD.122222 to S.F. and T.W and grant D82015 by the German Ministry of Health (BMG) to T.W.. The funders had no role in designing the research nor the decision to publish.

Conflicts of interest

The authors declare that no conflicts of interest exist.

498 # Supplementary Figures

Supplementary Figure S1: **Spike pseudo-group dynamics over the investigated time horizon**. For readability we only plot pseudo-groups that appear at a frequency of $> 1\%$.

Supplementary Figure S2: **Testing statistics for Germany over time.** The upper purple line shows the number of conducted PCR tests per day. The lower green line depicts the fraction of positive tests. (The weekly numbers of tests are divided by 7 to get the daily tests and taking the 7th date of the week)

Supplementary Figure S3: **Prevalence estimation for the UK.** Black-dashed line: Reported cases in the UK. Red line: Prevalence estimation using GInPipe. A right-sided rolling sum of 10 days is applied to the newly reported cases and minimum incidence estimates to approximate the prevalence (PCR positivity). Blue line: Prevalence estimation from the representative COVID-19 Infection Survey of the Office for National Statistics (ONS). Shaded areas denote inter-quartile ranges. Incidences were calculated with GInPipe in temporal and lineage-wise chunks due to the high number of sequences (>3 million) and were added up afterwards.

Supplementary Figure S4: **Schematic depicting measures of escape from antibody biding. A.** Theoretical binding curve of an antibody a to a wild-type epitope of SARS-CoV-2 (black line) and corresponding binding curve to an epitope that is mutated at site s (red line). The black dot and red cricle mark the concentrations (x-axis) where the binding is half maximal (y-axis) for the wild type $\text{IC50}_{\text{DMS}}(a)$ and mutant virus $\text{IC50}_{\text{DMS}}(a, s)$ respectively. The 'fold resistance' (red-black dashed line) denotes the shift in the IC50, such that $\text{IC50}_{\text{DMS}}(a, s) = \text{FR}(a, s) \cdot \text{IC50}_{\text{DMS}}(a)$. The red square marks the DMS-measured unbound fraction (escape fraction $\text{ef}(s, a)$) and the upward-pointing grey arrow the concentration at which the DMS experiment was conducted. All in all, $\text{IC50}_{\text{DMS}}(a)$, $\text{ef}(s, a)$ were measured and the DMS experiment was conducted with an antibody concentrations of 400 μg/mL. The same DMS experiment, performed with a more- (panel **B.**) or less (panel **C.**) potent antibody would yield a smaller, respectively bigger escape fraction, while the phenotypic effect $\text{FR}(a, s)$ of the mutation s is quantitatively identical.

Supplementary Figure S5: **Comparison between DMS-derived fold resistances $\mathbf{FR}_{x,y}$ and fold resistances derived from neutralization assays.** Distinct markers show the epitope groups and distinct colors the spike-pseudo groups.

Supplementary Tables

Supplementary Table ST1: **Pangolin lineages in the study.** List of 1,205 pangolin lineages in the sequence data set with mutations in the RBD and NTD region of the spike protein.

Supplementary Table ST2: **Spike-pseudo groups and their genomic profiles.** Assignment of pangolin lineages to 651 Spike-pseudo groups. Groups of lineages that show the same genomic profile with regard to mutations in the RBD and NTD region of the spike protein were grouped into the same Spike-pseudo-group.

Supplementary Table ST3: **Epitope classes.** Number of antibodies assigned per epitope class A-F

Supplementary Table ST4: **Epitope classes and antibodies.** Assignment of antibodies to epitope classes A-F

Supplementary Table ST5: **Omicron and Delta spike profiles for vaccine efficacy simulations.** Mutations in the spike protein of Omicron and Delta lineages as observed in the German sequences between November 20, 2021 until January 31, 2022 (for studies evaluating Wuhan-strain vaccine efficacy against Omicron) and July 4, 2021 till December 31, 2021 (for studies evaluating Wuhan-strain vaccine efficacy against Delta) with a minimum frequency of $> 75\%$.

Supplementary Table ST6: **Vaccine efficacy against Delta.** Vaccine efficacy data against Delta as extracted from the literature [74] [75] [76] [35]

Supplementary Table ST7: **Vaccine efficacy against Omicron.** Vaccine efficacy data against Omicron as extracted from the literature [74] [75]

Supplementary Data Files

Supplementary Data File D1: **Lineage Frequencies.** Lineage Frequencies per day as collected between Mar 2022 and Dec 2022 via covSonar (https://github.com/rki-mf1/covsonar).

References

[1] World Health Organization (WHO). Statement on the fifteenth meeting of the ihr (2005) emergency committee on the covid-19 pandemic. https://www.who.int, 2023. Accessed: 2023-09-08.

[2] Office for National Statistics (ONS). Coronavirus (COVID-19) infection survey, uk statistical bulletins, 2023. Accesed: 2023-05-15.

[3] World Health Organization (WHO). Weekly epidemiological update on covid-19 - 1 june 2023. https://www.who.int, 2023. Accessed: 2023-09-08.

[4] Hannah E Davis, Lisa McCorkell, Julia Moore Vogel, and Eric J Topol. Long covid: major findings, mechanisms and recommendations. Nature Reviews Microbiology, 21(3):133–146, 2023.

[5] World Health Organization (WHO). Coronavirus disease (covid-19) pandemic. https://www.who.int, 2023. Accessed: 2023-09-08.

[6] Peter V Markov, Mahan Ghafari, Martin Beer, Katrina Lythgoe, Peter Simmonds, Nikolaos I Stilianakis, and Aris Katzourakis. The evolution of sars-cov-2. Nature Reviews Microbiology, 21(6):361–379, 2023.

[7] Yanjia Chen, Xiaoyu Zhao, Hao Zhou, Huanzhang Zhu, Shibo Jiang, and Pengfei Wang. Broadly neutralizing antibodies to sars-cov-2 and other human coronaviruses. Nature reviews Immunology, 23(3):189–199, 2023.

[8] Yixuan J Hou, Shiho Chiba, Peter Halfmann, Camille Ehre, Makoto Kuroda, Kenneth H Dinnon III, Sarah R Leist, Alexandra Schäfer, Noriko Nakajima, Kenta Takahashi, et al. Sars-cov-2 d614g variant exhibits efficient replication ex vivo and transmission in vivo. Science, 370(6523):1464–1468, 2020.

[9] Bette Korber, Will M Fischer, Sandrasegaram Gnanakaran, Hyejin Yoon, James Theiler, Werner Abfalterer, Nick Hengartner, Elena E Giorgi, Tanmoy Bhattacharya, Brian Foley, et al. Tracking changes in sars-cov-2 spike: evidence that d614g increases infectivity of the covid-19 virus. Cell, 182(4):812–827, 2020.

[10] Jessica A Plante, Yang Liu, Jianying Liu, Hongjie Xia, Bryan A Johnson, Kumari G Lokugamage, Xianwen Zhang, Antonio E Muruato, Jing Zou, Camila R Fontes-Garfias, et al. Spike mutation d614g alters sars-cov-2 fitness. Nature, 592(7852):116–121, 2021.

[11] Ewen Callaway et al. The mutation that helps delta spread like wildfire. Nature, 596(7873):472–473, 2021.

[12] Emanuele Andreano, Giulia Piccini, Danilo Licastro, Lorenzo Casalino, Nicole V Johnson, Ida Paciello, Simeone Dal Monego, Elisa Pantano, Noemi Manganaro, Alessandro Manenti, et al. Sars-cov-2 escape from a highly neutralizing covid-19 convalescent plasma. Proceedings of the National Academy of Sciences, 118(36):e2103154118, 2021.

[13] Pengfei Wang, Manoj S Nair, Lihong Liu, Sho Iketani, Yang Luo, Yicheng Guo, Maple Wang, Jian Yu, Baoshan Zhang, Peter D Kwong, et al. Antibody resistance of sars-cov-2 variants b. 1.351 and b. 1.1. 7. Nature, 593(7857):130–135, 2021.

[14] Pengfei Wang, Ryan G Casner, Manoj S Nair, Maple Wang, Jian Yu, Gabriele Cerutti, Lihong Liu, Peter D Kwong, Yaoxing Huang, Lawrence Shapiro, et al. Increased resistance of sars-cov-2 variant p. 1 to antibody neutralization. Cell host & microbe, 29(5):747–751, 2021.

[15] Wanwisa Dejnirattisai, Jiandong Huo, Daming Zhou, Jiří Zahradník, Piyada Supasa, Chang Liu, Helen ME Duyvesteyn, Helen M Ginn, Alexander J Mentzer, Aekkachai Tuekprakhon, et al. Sars-cov-2 omicron-b. 1.1. 529 leads to widespread escape from neutralizing antibody responses. Cell, 185(3):467–484, 2022.

[16] Matthew McCallum, Nadine Czudnochowski, Laura E Rosen, Samantha K Zepeda, John E Bowen, Alexandra C Walls, Kevin Hauser, Anshu Joshi, Cameron Stewart, Josh R Dillen, et al. Structural basis of sars-cov-2 omicron immune evasion and receptor engagement. Science, 375(6583):864–868, 2022.

[17] Rungtiwa Nutalai, Daming Zhou, Aekkachai Tuekprakhon, Helen M Ginn, Piyada Supasa, Chang Liu, Jiandong Huo, Alexander J Mentzer, Helen ME Duyvesteyn, Aiste Dijokaite-Guraliuc, et al. Potent cross-reactive antibodies following omicron breakthrough in vaccinees. Cell, 185(12):2116–2131, 2022.

[18] Frederik Plesner Lyngse, Carsten Thure Kirkeby, Matthew Denwood, Lasse Engbo Christiansen, Kåre Mølbak, Camilla Holten Møller, Robert Leo Skov, Tyra Grove Krause, Morten Rasmussen, Raphael Niklaus Sieber, et al. Household transmission of sars-cov-2 omicron variant of concern subvariants ba. 1 and ba. 2 in denmark. Nature Communications, 13(1):5760, 2022.

[19] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, et al. A global panel database of pandemic policies (oxford covid-19 government response tracker). Nature human behaviour, 5(4):529–538, 2021.

[20] Nathalie Worp, Lorenzo Subissi, Mark D Perkins, Maria D Van Kerkhove, Anurag Agrawal, Meera Chand, Janko van Beek, Bas B Oude Munnink, and Marion PG Koopmans. Towards the development of a sars-cov-2 variant risk assessment tool: expert consultation on the assessment of scientific evidence on emerging variants. The Lancet Microbe, 2023.

[21] World Health Organization (WHO). Statement on the antigen composition of covid-19 vaccines. https://www.who.int, 2023. Accessed: 2023-09-08.

[22] Allison J Greaney, Tyler N Starr, and Jesse D Bloom. An antibody-escape estimator for mutations to the SARS-CoV-2 receptor-binding domain. Virus Evolution, 8(1):veac021, 2022.

[23] Allison J Greaney, Tyler N Starr, Pavlo Gilchuk, Seth J Zost, Elad Binshtein, Andrea N Loes, Sarah K Hilton, John Huddleston, Rachel Eguia, Katharine H D Crawford, Adam S Dingens, Rachel S Nargi, Rachel E Sutton, Naveenchandra Suryadevara, Paul W Rothlauf, Zhuoming Liu, Sean P J Whelan, Robert H Carnahan, James E Crowe, Jr, and Jesse D Bloom. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host Microbe, 29(1):44–57.e9, 2021.

[24] Yunlong Cao, Fanchong Jian, Jing Wang, Yuanling Yu, Weiliang Song, Ayijiang Yisimayi, Jing Wang, Ran An, Xiaosu Chen, Na Zhang, Yao Wang, Peng Wang, Lijuan Zhao, Haiyan Sun, Lingling Yu, Sijie Yang, Xiao Niu, Tianhe Xiao, Qingqing Gu, Fei Shao, Xiaohua Hao, Yanli Xu, Ronghua Jin, Zhongyang Shen, Youchun Wang, and Xiaoliang Sunney Xie. Imprinted SARS-CoV-2 humoral immunity induces convergent omicron RBD evolution. Nature, 614(7948):521–529, 2023.

[25] Djin Ye Oh, Martin Hölzer, Sofia Paraskevopoulou, Maria Trofimova, Felix Hartkopf, Matthias Budt, Marianne Wedde, Hugues Richard, Berit Haldemann, Teresa Domaszewska, et al. Advancing precision vaccinology by molecular and genomic surveillance of severe acute respiratory syndrome coronavirus 2 in germany, 2021. Clinical infectious diseases, 75(Supplement_1):S110–S120, 2022.

[26] Maureen Rebecca Smith, Maria Trofimova, Ariane Weber, Yannick Duport, Denise Kühnert, and Max von Kleist. Rapid incidence estimation from sars-cov-2 genomes reveals decreased case detection in europe during summer 2020. Nature Communications, 12(1):6009, 2021.

[27] Katharina Jahn, David Dreifuss, Ivan Topolsky, Anina Kull, Pravin Ganesanandamoorthy, Xavier Fernandez-Cassi, Carola Bänziger, Alexander J Devaux, Elyse Stachler, Lea Caduff, et al. Early detection and surveillance of sars-cov-2 genomic variants in wastewater using cojac. Nature Microbiology, 7(8):1151–1160, 2022.

[28] Yunlong Cao, Jing Wang, Fanchong Jian, Tianhe Xiao, Weiliang Song, Ayijiang Yisimayi, Weijin Huang, Qianqian Li, Peng Wang, Ran An, Jing Wang, Yao Wang, Xiao Niu, Sijie Yang, Hui Liang, Haiyan Sun, Tao Li, Yuanling Yu, Qianqian Cui, Shuo Liu, Xiaodong Yang, Shuo Du, Zhiying Zhang, Xiaohua Hao, Fei Shao, Ronghua Jin, Xiangxi Wang, Junyu Xiao, Youchun Wang, and Xiaoliang Sunney Xie. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature, 602(7898):657–663, 2022.

[29] Matthew McCallum, Anna De Marco, Florian A Lempp, M Alejandra Tortorici, Dora Pinto, Alexandra C Walls, Martina Beltramello, Alex Chen, Zhuoming Liu, Fabrizia Zatta, Samantha Zepeda, Julia di Iulio, John E Bowen, Martin Montiel-Ruiz, Jiayi Zhou, Laura E Rosen, Siro Bianchi, Barbara Guarino, Chiara Silacci Fregni, Rana Abdelnabi, Shi-Yan Caroline Foo, Paul W Rothlauf, Louis-Marie Bloyet, Fabio Benigni, Elisabetta Cameroni, Johan Neyts, Agostino Riva, Gyorgy Snell, Amalio Telenti, Sean P J Whelan, Herbert W Virgin, Davide Corti, Matteo Samuele Pizzuto, and David Veesler. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell, 184(9):2332–2347.e16, 2021.

[30] David S Khoury, Deborah Cromer, Arnold Reynaldi, Timothy E Schlub, Adam K Wheatley, Jennifer A Juno, Kanta Subbarao, Stephen J Kent, James A Triccas, and Miles P Davenport. Neutralizing antibody levels are highly predictive of immune protection from symptomatic sars-cov-2 infection. Nature medicine, 27(7):1205–1211, 2021.

[31] Richard J. Kryscio, Marta S. Mendiondo, Frederick A. Schmitt, and William R. Markesbery. Designing a large prevention trial: statistical issues. Statistics in Medicine, 23(2):285–296, 2004.

[32] Bo Meng, Adam Abdullahi, Isabella ATM Ferreira, Niluka Goonawardane, Akatsuki Saito, Izumi Kimura, Daichi Yamasoba, Pehuén Pereyra Gerber, Saman Fatihi, Surabhi Rathore, et al. Altered tmprss2 usage by sars-cov-2 omicron impacts infectivity and fusogenicity. Nature, 603(7902):706–714, 2022.

[33] Anupriya Aggarwal, Anouschka Akerman, Vanessa Milogiannakis, Mariana Ruiz Silva, Gregory Walker, Alberto Ospina Stella, Andrea Kindinger, Thomas Angelovich, Emily Waring, Supavadee Amatayakul-Chantler, et al. Sars-cov-2 omicron ba. 5: Evolving tropism and evasion of potent humoral responses and resistance to clinical immunotherapeutics relative to viral variants of concern. EBioMedicine, 84, 2022.

[34] Niklas Bobrovitz, Harriet Ware, Xiaomeng Ma, Zihan Li, Reza Hosseini, Christian Cao, Anabel Selemon, Mairead Whelan, Zahra Premji, Hanane Issa, et al. Protective effectiveness of previous sars-cov-2 infection and hybrid immunity against the omicron variant and severe disease: a systematic review and meta-regression. The Lancet Infectious Diseases, 23(5):556–567, 2023.

[35] Daniel R Feikin, Melissa M Higdon, Laith J Abu-Raddad, Nick Andrews, Rafael Araos, Yair Goldberg, Michelle J Groome, Amit Huppert, Katherine L O'Brien, Peter G Smith, Annelies Wilder-Smith, Scott Zeger, Maria Deloria Knoll, and Minal K Patel. Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression. Lancet, 399(10328):924–944, 2022.

[36] Oliver J Watson, Gregory Barnsley, Jaspreet Toor, Alexandra B Hogan, Peter Winskill, and Azra C Ghani. Global impact of the first year of covid-19 vaccination: a mathematical modelling study. The Lancet Infectious Diseases, 22(9):1293–1302, 2022.

[37] Alessandro Sette and Shane Crotty. Adaptive immunity to sars-cov-2 and covid-19. Cell, 184(4):861–880, 2021.

[38] Baolin Liao, Zhao Chen, Peiyan Zheng, Linghua Li, Jianfen Zhuo, Fang Li, Suxiang Li, Dingbin Chen, Chunyan Wen, Weiping Cai, Shanhui Wu, Yanhong Tang, Linwei Duan, Peilan Wei, Fangli Chen, Jinwei Yuan, Jinghong Yang, Jiaxin Feng, Jingxian Zhao, Jincun Zhao, Baoqing Sun, Airu Zhu, Yimin Li, and Xiaoping Tang. Detection of anti-sars-cov-2-s2 igg is more sensitive than anti-rbd igg in identifying asymptomatic covid-19 patients. Frontiers in Immunology, 12:724763, 2021.

[39] Carolyn Rydyznski Moderbacher, Sydney I Ramirez, Jennifer M Dan, Alba Grifoni, Kathryn M Hastie, Daniela Weiskopf, Simon Belanger, Robert K Abbott, Christina Kim, Jinyong Choi, et al. Antigen-specific adaptive immunity to sars-cov-2 in acute covid-19 and associations with age and disease severity. Cell, 183(4):996–1012, 2020.

[40] Anthony T Tan, Martin Linster, Chee Wah Tan, Nina Le Bert, Wan Ni Chia, Kamini Kunasegaran, Yan Zhuang, Christine YL Tham, Adeline Chia, Gavin JD Smith, et al. Early induction of functional sars-cov-2-specific t cells associates with rapid viral clearance and mild disease in covid-19 patients. Cell reports, 34(6):108728, 2021.

[41] Alessandro Sette, John Sidney, and Shane Crotty. T cell responses to sars-cov-2. Annual Review of Immunology, 41(1):343–373, 2023.

[42] Gregory A Poland, Inna G Ovsyannikova, and Richard B Kennedy. Sars-cov-2 immunity: review and applications to phase 3 vaccine candidates. The Lancet, 396(10262):1595–1606, 2020.

[43] Jesse D. Bloom and Richard A. Neher. Fitness effects of mutations to sars-cov-2 proteins. bioRxiv, 2023.

[44] Hannah McClymont and Wenbiao Hu. Weather variability and covid-19 transmission: a review of recent research. International journal of environmental research and public health, 18(2):396, 2021.

[45] Yiqun Ma, Sen Pei, Jeffrey Shaman, Robert Dubrow, and Kai Chen. Role of meteorological factors in the transmission of sars-cov-2 in the united states. Nature Communications, 12(1):3602, 2021.

[46] Jordi Landier, Juliette Paireau, Stanislas Rebaudet, Eva Legendre, Laurent Lehot, Arnaud Fontanet, Simon Cauchemez, and Jean Gaudart. Cold and dry winter conditions are associated with greater sars-cov-2 transmission at regional level in western countries during the first epidemic wave. Scientific Reports, 11(1):12756, 2021.

[47] Mohammad M Sajadi, Parham Habibzadeh, Augustin Vintzileos, Shervin Shokouhi, Fernando Miralles-Wilhelm, and Anthony Amoroso. Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of coronavirus disease 2019 (covid-19). JAMA network open, 3(6):e2011834, 2020.

[48] Simiao Chen, Klaus Prettner, Michael Kuhn, Pascal Geldsetzer, Chen Wang, Till Bärnighausen, and David E Bloom. Climate and the spread of covid-19. Scientific Reports, 11(1):9042, 2021.

[49] Tomáš Gavenčiak, Joshua Teperowski Monrad, Gavin Leech, Mrinank Sharma, Sören Mindermann, Samir Bhatt, Jan Brauner, and Jan Kulveit. Seasonal variation in sars-cov-2 transmission in temperate climates: A bayesian modelling study in 143 european regions. PLoS Computational Biology, 18(8):e1010435, 2022.

[50] Canelle Poirier, Wei Luo, Maimuna S Majumder, Dianbo Liu, Kenneth D Mandl, Todd A Mooring, and Mauricio Santillana. The role of environmental factors on transmission rates of the covid-19 outbreak: an initial assessment in two spatial scales. Scientific reports, 10(1):17002, 2020.

[51] Colin J Carlson, Ana CR Gomez, Shweta Bansal, and Sadie J Ryan. Misconceptions about weather and seasonality must not misguide covid-19 response. Nature communications, 11(1):4312, 2020.

[52] Riccardo Cappi, Luca Casini, Davide Tosi, and Marco Roccetti. Questioning the seasonality of sars-cov-2: A fourier spectral analysis. BMJ open, 12(4):e061602, 2022.

[53] Rachel E Baker, Wenchang Yang, Gabriel A Vecchi, C Jessica E Metcalf, and Bryan T Grenfell. Susceptible supply limits the role of climate in the early sars-cov-2 pandemic. Science, 369(6501):315–319, 2020.

[54] Gavin Y. Oudit, Kaiming Wang, Anissa Viveiros, Max J. Kellner, and Josef M. Penninger. Angiotensin-converting enzyme 2—at the heart of the covid-19 pandemic. Cell, 186(5):906–922, 2023.

[55] Manuel Hayn, Maximilian Hirschenberger, Lennart Koepke, Rayhane Nchioua, Jan Hendrik Straub, Susanne Klute, Victoria Hunszinger, Fabian Zech, Caterina Prelli Bozzo, Wasim Aftab, Maria Hønholt Christensen, Carina Conzelmann, Janis Alexander Müller, Smitha Srinivasachar Badarinarayan, Christina Martina Stürzel, Ignasi Forne, Steffen Stenger, Karl-Klaus Conzelmann, Jan Münch, Florian Ingo Schmidt, Daniel Sauter, Axel Imhof, Frank Kirchhoff, and Konstantin Maria Johannes Sparrer. Systematic functional analysis of sars-cov-2 proteins uncovers viral innate immune antagonists and remaining vulnerabilities. Cell Reports, 35(7):109126, 2021.

[56] Judith M Minkoff and Benjamin tenOever. Innate immune evasion strategies of sars-cov-2. Nature Reviews Microbiology, 21(3):178–194, 2023.

[57] N. Alexia Raharinirina, Nils Gubela, Daniela Börnigen, and et al. Sars-cov-2 evolution on a dynamic immune landscape. Zenodo, 2023, https://doi.org/10.5281/zenodo.8349296.

[58] Andrew Rambaut, Edward C Holmes, Áine O'Toole, Verity Hill, John T McCrone, Christopher Ruis, Louis du Plessis, and Oliver G Pybus. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol., 5(11):1403–1407, 2020.

[59] Áine O'Toole, Emily Scher, Anthony Underwood, Ben Jackson, Verity Hill, John T McCrone, Rachel Colquhoun, Chris Ruis, Khalil Abu-Dahab, Ben Taylor, Corin Yeats, Louis du Plessis, Daniel Maloney, Nathan Medd, Stephen W Attwood, David M Aanensen, Edward C Holmes, Oliver G Pybus, and Andrew Rambaut. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol., 7(2):veab064, 2021.

[60] Wiep van der Toorn, Djin-Ye Oh, Daniel Bourquain, Janine Michel, Eva Krause, Andreas Nitsche, and Max von Kleist. An intra-host sars-cov-2 dynamics model to assess testing and quarantine strategies for incoming travelers, contact management, and de-isolation. Patterns, 2(6):100262, 2021.

[61] Yunlong Cao, Ayijiang Yisimayi, Fanchong Jian, Weiliang Song, Tianhe Xiao, Lei Wang, Shuo Du, Jing Wang, Qianqian Li, Xiaosu Chen, Yuanling Yu, Peng Wang, Zhiying Zhang, Pulan Liu, Ran An, Xiaohua Hao, Yao Wang, Jing Wang, Rui Feng, Haiyan Sun, Lijuan Zhao, Wen Zhang, Dong Zhao, Jiang Zheng, Lingling Yu, Can Li, Na Zhang, Rui Wang, Xiao Niu, Sijie Yang, Xuetao Song, Yangyang Chai, Ye Hu, Yansong Shi, Linlin Zheng, Zhiqiang Li, Qingqing Gu, Fei Shao, Weijin Huang, Ronghua Jin, Zhongyang Shen, Youchun Wang, Xiangxi Wang, Junyu Xiao, and Xiaoliang Sunney Xie. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by omicron infection. Nature, 608(7923):593–602, 2022.

[62] Bernadeta Dadonaite, Katharine H D Crawford, Caelan E Radford, Ariana G Farrell, Timothy C Yu, William W Hannon, Panpan Zhou, Raiees Andrabi, Dennis R Burton, Lihong Liu, David D Ho, Helen Y Chu, Richard A Neher, and Jesse D Bloom. A

pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. Cell, 186(6):1263–1278.e20, 2023.

[63] Sheila F Lumley, Jia Wei, Denise O'Donnell, Nicole E Stoesser, Philippa C Matthews, Alison Howarth, Stephanie B Hatch, Brian D Marsden, Stuart Cox, Tim James, Liam J Peck, Thomas G Ritter, Zoe de Toledo, Richard J Cornall, E Yvonne Jones, David I Stuart, Gavin Screaton, Daniel Ebner, Sarah Hoosdally, Derrick W Crook, Christopher P Conlon, Koen B Pouwels, A Sarah Walker, Tim E A Peto, Timothy M Walker, Katie Jeffery, David W Eyre, and Oxford University Hospitals Staff Testing Group. The Duration, Dynamics, and Determinants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antibody Responses in Individual Healthcare Workers. Clinical Infectious Diseases, 73(3):e699–e709, 2021.

[64] Jian-Hang Xue, Yong-Jing Wang, Wei Li, Qiu-Ling Li, Qiu-Yan Xu, Jian-Jun Niu, and Li-Li Liu. Anti–receptor-binding domain immunoglobulin g antibody as a predictor of seropositivity for anti–sars-cov-2 neutralizing antibody. Archives of Pathology & Laboratory Medicine, 146(7):814–821, 2022.

[65] Irina Arkhipova-Jenkins, Mark Helfand, Charlotte Armstrong, Emily Gean, Joanna Anderson, Robin A Paynter, and Katherine Mackey. Antibody response after sars-cov-2 infection and implications for immunity: a rapid living review. Annals of internal medicine, 174(6):811–821, 2021.

[66] Jean-Louis Bayart, Jonathan Douxfils, Constant Gillot, Clara David, François Mullier, Marc Elsen, Christine Eucher, Sandrine Van Eeckhoudt, Tatiana Roy, Vincent Gerin, Grégoire Wieers, Christine Laurent, Mélanie Closset, Jean-Michel Dogné, and Julien Favresse. Waning of igg, total and neutralizing antibodies 6 months post-vaccination with bnt162b2 in healthcare workers. Vaccines, 9(10):1092, 2021.

[67] Thomas W. Barnes, Johannes Schulte-Pelkum, Laura Steller, Daniel Filchtinski, Robin Jenness, Michelle R. Williams, Christina Kober, Sandro Manni, Thomas Hauser, Aaron Hahn, Uwe Kalina, Toby L. Simon, Patrick Schuetz, and Nathan J. Roth. Determination of neutralising anti-sars-cov-2 antibody half-life in covid-19 convalescent donors. Clinical Immunology, 232:108871, 2021.

[68] Alicia T. Widge, Nadine G. Rouphael, Lisa A. Jackson, Evan J. Anderson, Paul C. Roberts, Mamodikoe Makhene, James D. Chappell, Mark R. Denison, Laura J. Stevens, Andrea J. Pruijssers, Adrian B. McDermott, Britta Flach, Bob C. Lin, Nicole A. Doria-Rose, Sijy O'Dell, Stephen D. Schmidt, Kathleen M. Neuzil, Hamilton Bennett, Brett Leav, Mat Makowski, Jim Albert, Kaitlyn Cross, Venkata-Viswanadh Edara, Katharine Floyd, Mehul S. Suthar, Wendy Buchanan, Catherine J. Luke, Julie E. Ledgerwood, John R. Mascola, Barney S. Graham, and John H. Beigel. Durability of responses after sars-cov-2 mrna-1273 vaccination. New England Journal of Medicine, 384(1):80–82, 2021.

[69] Jennifer M. Dan, Jose Mateus, Yu Kato, Kathryn M. Hastie, Esther Dawen Yu, Caterina E. Faliti, Alba Grifoni, Sydney I. Ramirez, Sonya Haupt, April Frazier, Catherine Nakao, Vamseedhar Rayaprolu, Stephen A. Rawlings, Bjoern Peters, Florian Krammer, Viviana Simon, Erica Ollmann Saphire, Davey M. Smith, Daniela Weiskopf, Alessandro Sette, and Shane Crotty. Immunological memory to sars-cov-2 assessed for up to 8 months after infection. Science, 371(6529):eabf4063, 2021.

[70] Adam K Wheatley, Jennifer A Juno, Jing J Wang, Kevin J Selva, Arnold Reynaldi, Hyon-Xhi Tan, Wen Shi Lee, Kathleen M Wragg, Hannah G Kelly, Robyn Esterbauer, et al. Evolution of immune responses to sars-cov-2 in mild-moderate covid-19. Nature communications, 12(1):1162, 2021.

[71] Jan Van Elslande, Lien Gruwier, Lode Godderis, and Pieter Vermeersch. Estimated Half-Life of SARS-CoV-2 Anti-Spike Antibodies More Than Double the Half-Life of Anti-nucleocapsid Antibodies in Healthcare Workers. Clinical Infectious Diseases, 73(12):2366–2368, 2021.

[72] Alexander S Bratus, Artem S Novozhilov, and Yuri S Semenov. Advanced Mathematical Methods in Biosciences and Applications, chapter Rigorous mathematical analysis of the quasispecies model: From Manfred Eigen to the recent developments. Springer, 2019.

[73] Loris Bennett, Bernd Melchers, and Boris Proppe. Curta: A general-purpose high-performance computer at zedat, freie universität berlin. http://dx.doi.org/10.17169/refubium-26754, 2020.

[74] Mie Agermose Gram, Hanne-Dorthe Emborg, Astrid Blicher Schelde, Nikolaj Ulrik Friis, Katrine Finderup Nielsen, Ida Rask Moustsen-Helms, Rebecca Legarth, Janni Uyen Hoa Lam, Manon Chaine, Aisha Zahoor Malik, et al. Vaccine effectiveness against sars-cov-2 infection or covid-19 hospitalization with the alpha, delta, or omicron sars-cov-2 variant: A nationwide danish cohort study. PLoS medicine, 19(9):e1003992, 2022.

[75] Mie Agermose Gram, Hanne-Dorthe Emborg, Astrid Blicher Schelde, Nikolaj Ulrik Friis, Katrine Finderup Nielsen, Ida Rask Moustsen-Helms, Rebecca Legarth, Janni Uyen Hoa Lam, Manon Chaine, Aisha Zahoor Malik, Morten Rasmussen, Jannik Fonager, Raphael Niklaus Sieber, Marc Stegger, Steen Ethelberg, Palle Valentiner-Branth, and Christian Holm Hansen. Vaccine effectiveness against SARS-CoV-2 infection or COVID-19 hospitalization with the alpha, delta, or omicron SARS-CoV-2 variant: A nationwide danish cohort study. PLoS Med, 19(9):e1003992, 2022.

[76] Sara Y Tartof, Jeff M Slezak, Heidi Fischer, Vennis Hong, Bradley K Ackerson, Omesh N Ranasinghe, Timothy B Frankland, Oluwaseye A Ogun, Joann M Zamparo, Sharon Gray, Srinivas R Valluri, Kaije Pan, Frederick J Angulo, Luis Jodar, and John M McLaughlin. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. Lancet, 398(10309):1407–1416, 2021.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryDataFile.csv
- SupplementaryFiguresTables.pdf