

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>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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 Smith³, Djin-Ye Oh⁴, Matthias Budt⁴, Christin Mache⁴, Claudia Schillings¹, Stephan Fuchs⁵, Ralf Dürrwald⁴, Thorsten Wolff⁴, Martin Hölzer⁵, Sofia Paraskevopoulou⁵, and Max von Kleist^{1,3,+}

¹Department of Mathematics & Computer Science, Freie Universität Berlin, Germany

²International Max-Planck Research School “Biology and Computation”

(IMPRS-BAC), Max-Planck Institute for Molecular Genetics, Berlin, Germany

³Project groups, Robert-Koch Institute, Berlin, Germany

⁴Department 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

¹ 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 a 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.

²¹

22 Introduction

23 In 2023, most non-pharmaceutical measures intended to curb the spread of SARS-CoV-2
24 have been lifted as the WHO de-escalated the COVID-19 pandemic from a ‘Public Health
25 Emergency of International Concern’ (PHEIC) to a ‘pandemic that denotes an established
26 and ongoing health issue’ [1]. While the occurrence of explosive surges of COVID-19 has
27 most likely come to an end, SARS-CoV-2 still spreads at alarming rates, with 1.5–5% of the
28 population infected at any given time [2] and thousands of ICU cases and reported fatalities
29 monthly [3], notwithstanding long-term effects [4]. As of August 2023, nearly four years after
30 the emergence of SARS-CoV-2, approximately 770 million cases have been reported [5], with
31 true numbers magnitudes larger, such that it is reasonable to assume that almost the entire
32 world population has been exposed to SARS-CoV-2 antigen to date.
33 Since its first introduction in the human population, SARS-CoV-2 has evolved significantly [6]
34 with most changes allocated to its surface protein (spike), which is both responsible for bind-
35 ing host cells and the target site for neutralizing antibodies elicited by vaccines or infection [7].
36 While many early evolutionary changes of the virus are believed to have been driven by the
37 adaptation of the virus to the human host [8, 9, 10, 11], spike mutations associated with es-
38 cape from neutralizing antibodies were observed already during the first year of the pandemic,
39 and before the roll-out of vaccines [12, 13, 14]. After the broad vaccine roll-out and the global
40 Delta wave, SARS-CoV-2 has circulated across an immunologically more experienced popula-
41 tion, being transmitted more and more through hosts who had been vaccinated or previously
42 infected. Subsequently, the early Omicron variants BA.1 and BA.2 emerged, which exhibited
43 a substantial transmission advantage over Delta [6]. This transmission advantage was associ-
44 ated with a pronounced evasion from humoral immunity [15, 16, 17, 18]. While inequalities
45 in vaccine distribution¹ and differences in the implementation of non-pharmaceutical mea-
46 sures may have created inter-country differences early-on in the pandemic [19], those were
47 relatively minor compared to the vast geographical differences observed in the distribution
48 and spread of emerging Omicron sub-lineages over the last year. This geographical variation
49 may reflect an increasing complexity of the global immunological landscape, where the course
50 of waves with new (sub-) variants in a particular region could be substantially influenced by
51 the infection history of that region, i.e. which variants dominated the preceding waves and
52 at what time.
53 Consequently, it has become increasingly difficult to predict whether a new variant will spread
54 successfully in a given population at a given time. This circumstance also poses challenges
55 in optimizing the design of adapted mRNA vaccines to protect vulnerable groups or heavily
56 exposed individuals. Despite the appreciation of the problem [20] and the abundance of rich
57 data sources for SARS-CoV-2, until now, there has been limited progress in integrating the
58 available data to inform variant risk assessment and vaccine design, as acknowledged by a
59 “limitation of evidence” by the WHO technical advisory group on COVID-19 vaccine com-
60 position (TAG-CO-VAC) [21].
61 Almost four years after the emergence of SARS-CoV-2, the scientific community has collected
62 an unprecedented wealth of data. Amongst others, the Bloom Lab developed high-throughput
63 assays that can precisely define the epitopes of neutralizing antibodies and quantify the phe-
64 notypic effect of virtually any mutation on the spike protein, even those not yet observed in
65 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.

84

85 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.

105 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’ $0 \leq FR \leq \infty$, 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

¹⁵⁴ superimposed onto the Spike RBD structure in **Fig. 2C**. Interestingly, the fold resistances be-
¹⁵⁵ tween the Omicron ‘Spike-pseudo-groups’ in **Fig. 2B** highlight a consecutively increasing fold
¹⁵⁶ resistance of succeeding variants relative to their temporal predecessors (compare **Fig. 1B**).

¹⁵⁷ Immune waning and infection risk reduction

¹⁵⁸ While cross-neutralization potency can qualitatively describe the antigenic overlap between
¹⁵⁹ variants, the ability to neutralize the virus (and prevent infection) will ultimately depend on
¹⁶⁰ the **neutralizing antibody concentrations at the time of viral re-exposure** [30].

¹⁶¹ Neutralizing antibody levels rise within 1 to 2 weeks after initial antigen exposure and
¹⁶² slowly decay afterwards. We parameterized a simple **pharmacokinetic model** to capture the
¹⁶³ **antibody concentration-time profile after initial antigen exposure as shown in Fig. 3A** (details
¹⁶⁴ in *Methods*). Then, based on DMS data², we computed a relative weighing of antibodies
¹⁶⁵ belonging to different epitope classes as shown in **Fig. 3B**. This data-derived weighing may
¹⁶⁶ reflect the accessibility of the distinct epitopes, the strength of binding or the capability
¹⁶⁷ to neutralize the virus, when an antibody is bound. Our analysis showed that antibodies of
¹⁶⁸ epitope classes A, B, D2, and F3 were more potently neutralizing the virus, whereas antibodies
¹⁶⁹ of classes E3 and F1 were less potent.

¹⁷⁰ Based on the previously computed cross-neutralization potency, antibody pharmacokinetics
¹⁷¹ and **relative antibody potencies**, we subsequently estimated the **probability of neutralizing**
¹⁷² the Delta variant following exposure to the Wuhan-variant antigen (genetic profiles in **Suppl.**
¹⁷³ **Table ST5**) after antigen exposure with the Wuhan-variant antigen (**Fig. 3C-D**; data sources
¹⁷⁴ in **Suppl. Table ST6**). For this task, we estimated the only remaining free parameter (the
¹⁷⁵ average normalized antibody potency $\overline{IC_{50}}$) and **assumed that neutralization probability**
¹⁷⁶ **approximates infection risk reduction**.

¹⁷⁷ Notably, utilized clinical data varies considerably with regards to the level of reported
¹⁷⁸ risk reduction due to statistical limitations (few observed ‘infection events’), as well as het-
¹⁷⁹ erogeneity in the study groups, which is a well-known phenomenon for prevention trials [31].
¹⁸⁰ Regarding immune waning dynamics **after exposure to the Wuhan-variant antigen, we see**
¹⁸¹ **that, after antigen exposure, Delta can initially be almost completely neutralized. However,**
¹⁸² **neutralization probabilities decrease to 50%, depending on individual antibody pharmacoki-**
¹⁸³ **netics within 100 to 250 days after exposure to the antigen (Fig. 3C)**.

¹⁸⁴ We then applied the previously estimated parameters in our model to predict Omicron
¹⁸⁵ neutralization after Wuhan-variant antigen exposure. Our predictions align well with vaccine
¹⁸⁶ efficacy data (summarized in **Suppl. Table ST7**). Our model predicts that Omicron (genetic
¹⁸⁷ profile in **Suppl. Table ST5**) is initially neutralized with 45–85% probability approximately
¹⁸⁸ two weeks after exposure to the Wuhan-variant antigen, with neutralizing immunity decaying
¹⁸⁹ rapidly and reaching about 10% infection risk reduction between 80 and 350 days post antigen
¹⁹⁰ exposure.

¹⁹¹ Lineage-resolved population immunity predicts variant dynamics

¹⁹² Notably, with all parameters of our underlying model being derived at this stage, we may
¹⁹³ now **predict the individual infection risk reduction with regards to an arbitrary variant y for a**
¹⁹⁴ **person who was recently exposed to the Spike protein of an arbitrary variant x** , e.g., through

²https://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 population³. 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 $\gamma_{BA.2.86} = 0.095$ (CI: 0.064,

³<https://impfdashboard.de/>

328 0.127) due to population immunity. Consequently, the critical threshold of the relative intrinsic
329 growth rate would be $\alpha_{BA.2.86} / \sum_x \pi_x \cdot \alpha_x = (\gamma_{BA.2.86} + 1)^{-1} = 0.913$ (CI: 0.887, 0.939)
330 for this variant to spread successfully through the population.
331 In summary, we present a novel and innovative mathematical approach that can accurately
332 predict variant dynamics from a model-inferred dynamic immune landscape, utilizing DMS,
333 incidence, and molecular surveillance data. An implementation is freely available [57] and can
334 be adapted to different settings to allow contextualizing the risk of occurrence and spread of
335 variants at the country level, which is a major advancement of current warning systems. In
336 the future, our approach could serve as a basis to identify epitopes most likely to be under re-
337 centive selective pressure and therefore provide cues for designing vaccine candidates that avoid
338 targeting those epitopes, thus maximizing neutralization breadth against emerging variants
339 in a forthcoming season. Furthermore, our conceptional ideas may be transferred to model
340 the evolution of other respiratory viruses that are subject to molecular surveillance.

341 Figures

341

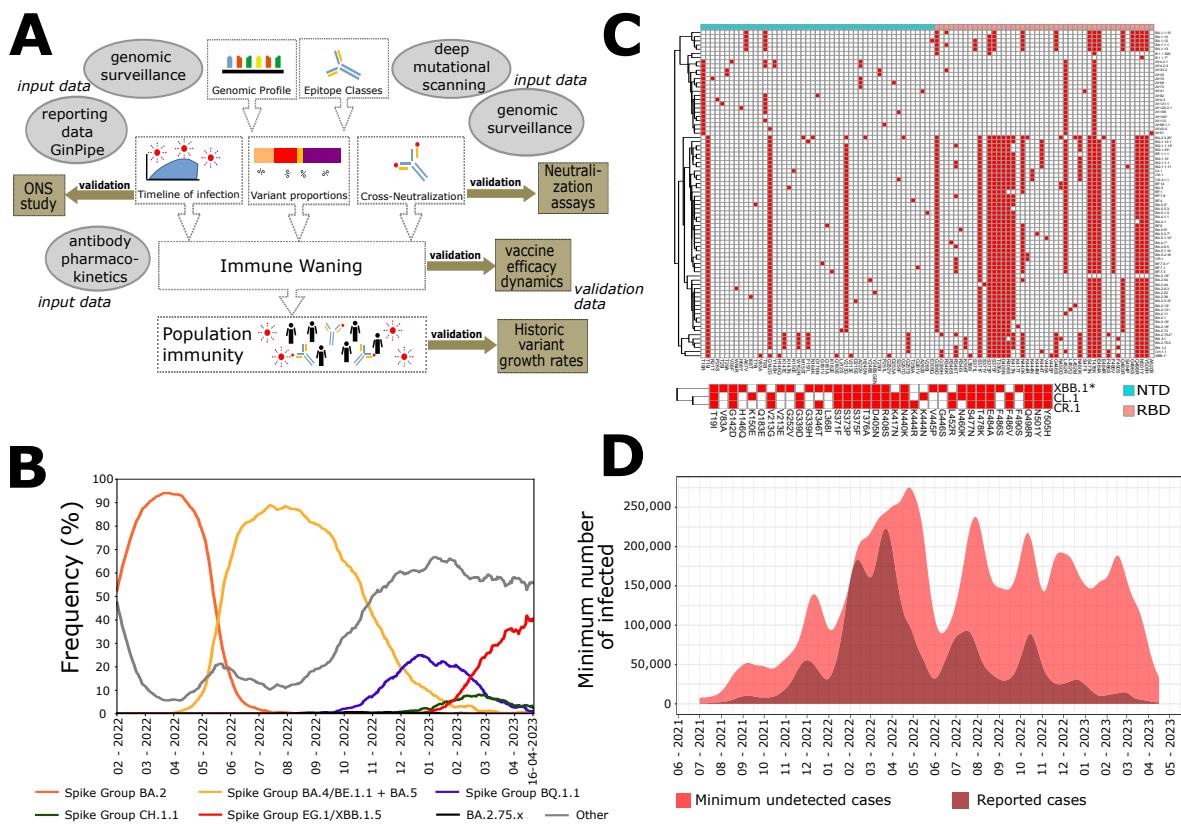


Figure 1: Overview of data sources, data integration, and validation steps for modeling variant-specific immunity. A. Overview. B. Variant dynamics of BA.2*, BA.4/5*, BQ.1.1*, CH.1.1*, EG.1/XBB.1.5*, and BA.2.75* ‘spike-pseudo-groups’. C. Mutation profiles of the 78 most abundant ‘spike-pseudo-groups’ 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 $I_{\text{rep}}(t)$; light red: undetected cases.

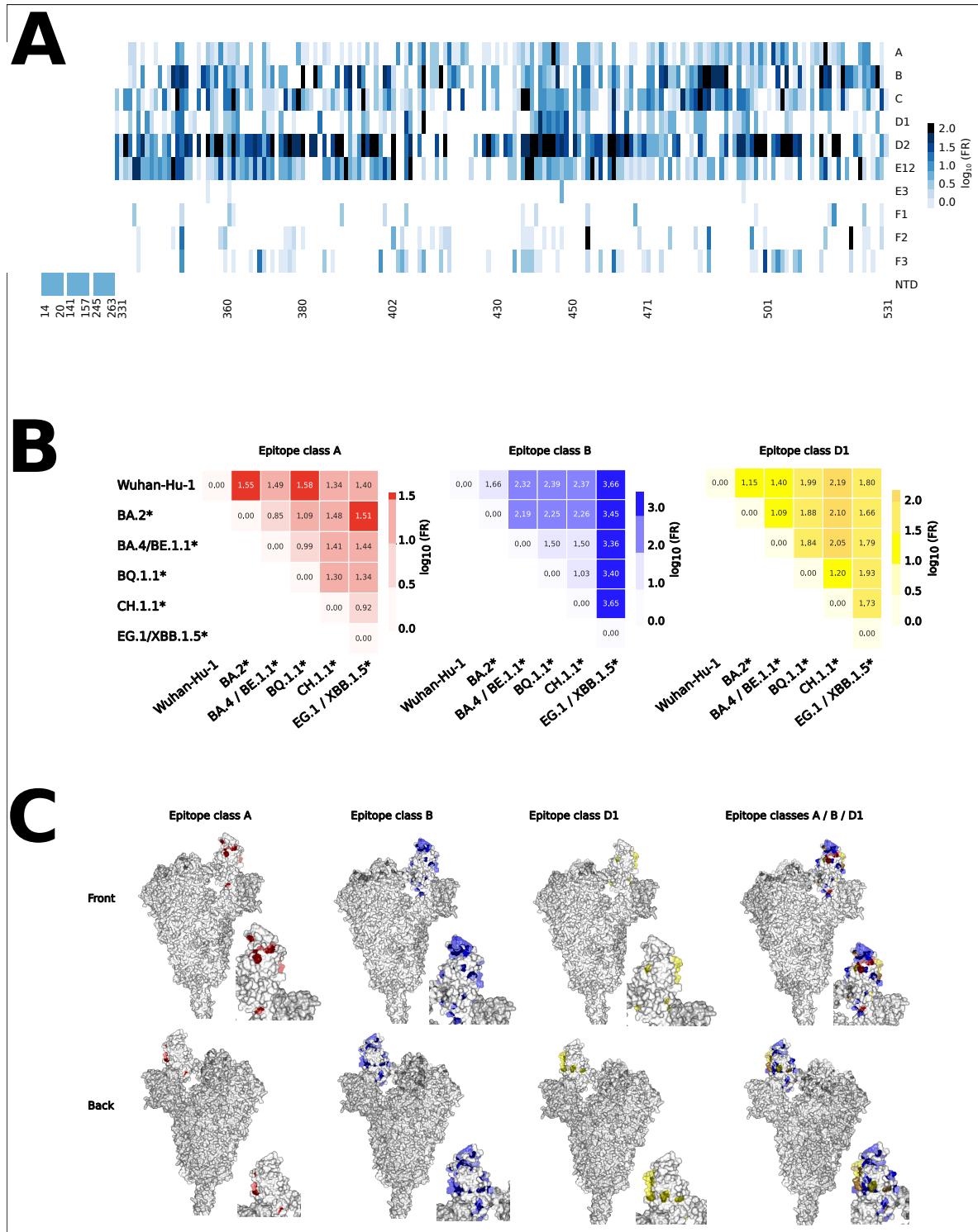


Figure 2: Epitope classes and variant cross-neutralization. A. Fold resistance against antibodies of epitope classes $\vartheta = \{A, B, C, D1, D2, E12, E3, F1, F2, F3, NTD\}$ induced by mutational changes at indicated sites. B. Fold resistance to neutralization against immunity-inducing variants (on 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.

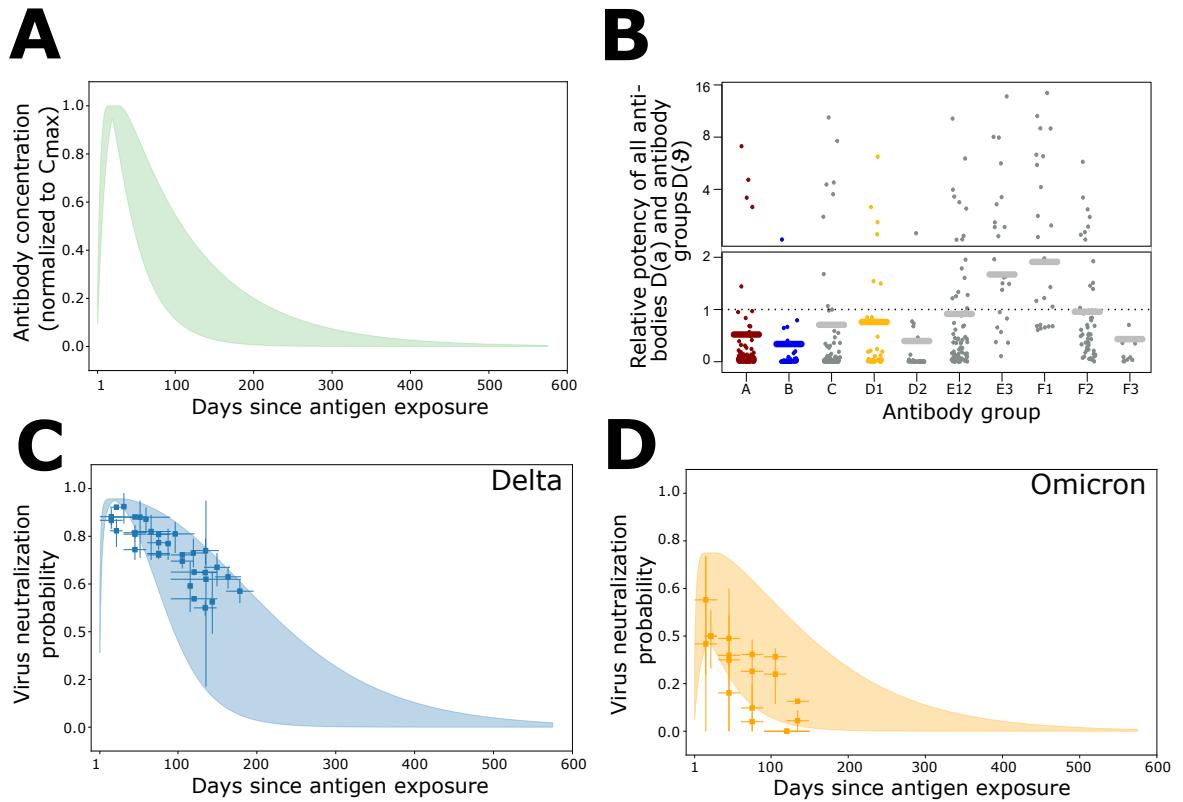


Figure 3: Immune waning dynamics. A. Normalized antibody pharmacokinetics after antigen exposure. B. Relative potency IC50_{DMS} of antibodies of epitope classes $\vartheta = \{A, B, C, D1, D2, E12, E3, F1, F2, F3\}$ in DMS data. The solid horizontal line shows the average IC50_{DMS} . C. Predicted (blue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

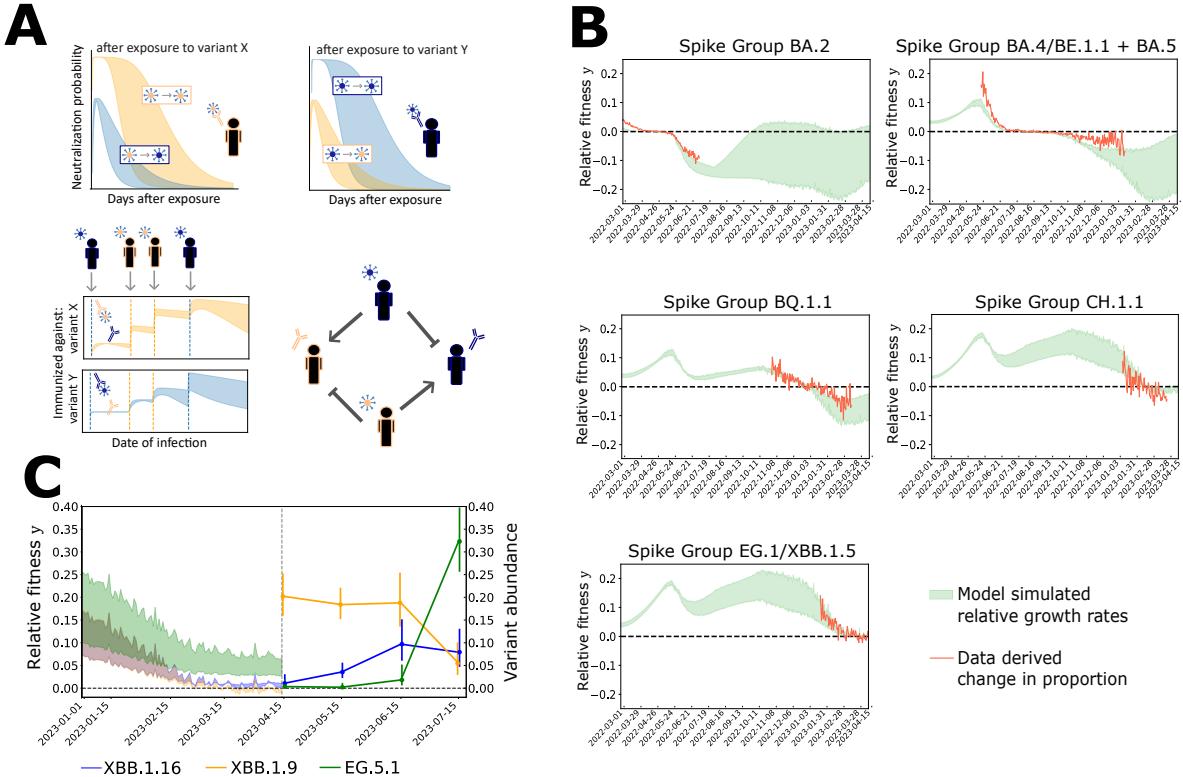


Figure 4: Population immunity and variant dynamics. A. Schematic for calculating population immunity: For each individual, we can compute the probability to cross-neutralize a variant y after exposure to variant x over time. Taking the infection history into account, we can calculate the expected number 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 $(\frac{\pi_y(t+1)}{\pi_y(t)} - 1)$ (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.

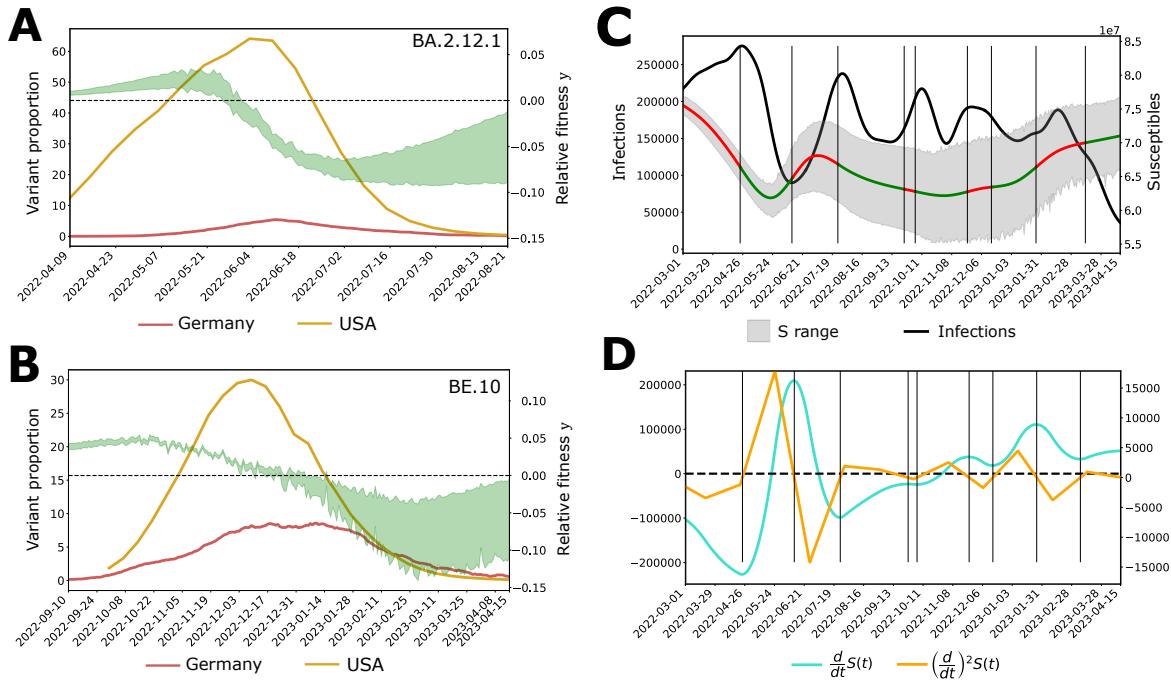


Figure 5: Impact of regional infection history on variant dynamics and prediction of infection waves. A. Dynamics of BA.2.12.1 in the US (orange line) vs. Germany (red line) and model-predicted number of BA.2.12.1 susceptible individuals in Germany (green area). B. Dynamics of BE.10 in the US (orange 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 $I_{\min}(t)$ (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.

³⁴² **Methods**

³⁴³ **SARS-CoV-2 genomic data**

³⁴⁴ We collected SARS-CoV-2 genomes from Germany published via the Robert Koch Institute⁴
³⁴⁵ and extracted their mutation profiles using covSonar⁵, a database-driven system for handling
³⁴⁶ genomic sequences of SARS-CoV-2 and screening genomic profiles. To ensure representative
³⁴⁷ sampling, we only included genomes from the ‘random sampling’ strategy, which denotes the
³⁴⁸ majority of all sequence data. In total, we collected \approx 600,000 SARS-CoV-2 genomes for
³⁴⁹ the time period of July 1, 2021 to April 16, 2023, which were subsequently used to infer
³⁵⁰ variant-specific population from March 01, 2022 to April 16, 2023.

³⁵¹ **Variant proportions and ‘Spike-pseudo-groups’**

³⁵² SARS-CoV-2 genomes were assigned to 1,205 pangolin lineages using established methods
³⁵³ [58, 59]. For each lineage, we collected all ‘characteristic’ spike amino acid (aa) changes for
³⁵⁴ subsequent analyses; in our work ‘characteristic’ implied that an aa change was present in at
³⁵⁵ least 75% of all sequences from that lineage⁶ (**Suppl. Table ST1**). We then determined the
³⁵⁶ genomic profile in the spike protein for all 1,205 SARS-CoV-2 pangolin lineages and found 887
³⁵⁷ mutations in Spike, with 280 distinct mutations in the RBD region (aa pos 318–541) and 314 in
³⁵⁸ the NTD region (aa pos 13–305). The ‘antigenic profile’ for each lineage was then determined
³⁵⁹ based on the set of unique mutations within the NTD and RBD regions. Differences between
³⁶⁰ lineages were defined as the set difference between mutation profiles. Clustering lineages with
³⁶¹ identical ‘antigenic profiles’ yielded 651 ‘Spike-pseudo-groups’ with distinct genomic profiles
³⁶² in the RBD and NTD region of the spike protein (**Suppl. Table ST2**). Based on the
³⁶³ genomic profiles and their clustering into epitope-pseudo-groups, we computed pseudo-group
³⁶⁴ frequencies $\pi_x(t)$ for all Spike-pseudo-groups $x \in \mathcal{X}$ in the entire observation horizon.

³⁶⁵ **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-

⁴https://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 $\tilde{P}_{\text{rep}}(t)$ at their maximum to be able to estimate the *minimum number of infections* $I_{\min}(t)$.

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

The minimum number of infections is then calculated as

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

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

369 The ONS survey is based on large-scale random sampling strategy of households, based
370 on which SARS-CoV-2 test-positivity and the number of currently infected individuals is
371 estimated, overcoming under-reporting issues resulting from changing testing policies or -
372 behaviour. In brief, data for the UK was available for the time frame of May 2021 to April
373 2022 from the officially reported cases⁸ and sequences from GISAID, respectively.

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

380 Deep mutational scanning (DMS) data

381 To assess the phenotype (escape from neutralizing antibodies) of each ‘Spike-pseudo-group’,
382 we used all DMS data [22, 24] available on Feb 13, 2023⁹. DMS measures, in a yeasts-display
383 assay, how much a mutated site s in the RBD affects the binding of antibodies elicited by a
384 variant that is not mutated at site s . We utilized data from 836 antibodies that were classified
385 into 12 distinct epitope classes [22] [28] (see below) and aggregated all values on site-level by
386 their mean, yielding ‘escape fractions’ to antibody a for each mutated site $\text{ef}_{s,a}$ (values were
387 bounded at 0.99). Notably, ‘escape fractions’ denote a proxy for quantifying the probability

⁷https://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>

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

388 that an antibody does not bind an RBD that contains a mutation at site s . As such, the
389 numerical value depends on the antibody potency and hence we aimed to convert these values
390 to ‘fold resistances’ (fold change in antibody potency), see also **Suppl. Fig. S4**.

391 Assuming that mutational effects of sites s on the binding affinity are independent, the binding
392 probability of an antibody elicited by a variant x to a variant y can be expressed as

$$b_a(x, y) = \prod_{s \in \Omega(x, y)} (1 - \text{ef}_{s,a}), \quad (1)$$

393 where $\text{ef}_{s,a}$ is the normalized escape fraction of mutated site s with respect to antibody a and
394 $\Omega(x, y)$ denotes the set of RBD sites that distinguish variants x and y [22].

395 Using classic pharmacodynamic approaches, we then model the binding probability as

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

where $\text{FR}_{x,y}(a)$ denotes the fold resistance of variant y to an antibody a elicited by variant x .
The parameter $\text{IC50}_{\text{DMS}}(a)$ 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,
 $c_a = 400 \mu\text{g/mL}$ was the antibody concentration at which the DMS experiment was
conducted [23, 28]. Combining eqs. (1)–(2) yields:

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

396 Notably, as already evident from **Suppl. Fig. S4**, DMS estimates of $\text{ef}_{s,a}$ as well as corre-
397 sponding $\text{FR}_{x,y}(a)$ can become unreliable depending on antibody concentrations and antibody
398 potency $\text{IC50}_x(a)$, falsely predicting hyper-susceptibility ($\text{FR}_{x,y}(a) < 1$). We enforced that
399 $\text{FR}_{x,y}(a) \geq 1$ to avoid such artifacts.

400 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:

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

401 for all epitope classes ϑ in \mathcal{A} (**Fig. 2A**).

402 As a proof of concept, we compared DMS-derived fold resistances $\text{FR}_{x,y}(\vartheta)$ using the calcula-
403 tions above with fold resistance values obtained from virus neutralization assays (reported in

⁴⁰⁴ Cao et al. [61]) for antibodies targeting all epitope classes \mathcal{A} defined above. As can be seen
⁴⁰⁵ in **Suppl. Fig. S5**, we observed a strong and significant positive correlation between the
⁴⁰⁶ DMS-derived fold resistances $\text{FR}_{x,y}(\vartheta)$ 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:

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

⁴⁰⁷ where $|\Omega(x,y)|$ denotes the number of mutational differences between variants x and y in the
⁴⁰⁸ antigenic super-site of the NTD.

⁴⁰⁹ 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:

$$P_{\text{Neut}}(t, x, y) = 1 - \prod_{\vartheta \in \mathcal{A}_x \setminus y} (1 - b_\vartheta(t, x, y)),$$

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

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

⁴¹⁰ where $c_\vartheta(t)$ is the antibody's concentration in an individual at time t , $\text{IC50}_x(\vartheta)$ is the half-
⁴¹¹ maximal inhibitory antibody concentration against the variant that elicited the antibody.
⁴¹² $\text{FR}_{x,y}(\vartheta)$ is the fold resistance of variant y to binding of antibodies of epitope class ϑ , elicited
⁴¹³ by variant x .

⁴¹⁴ Antibody potency

Next, we quantified $\text{IC50}_x(\vartheta)$ 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 data¹⁰:

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

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

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

$$\text{IC50}_x(\vartheta) = D(\vartheta) \cdot \widehat{\text{IC50}}_x,$$

where $\widehat{\text{IC50}}_x$ denotes the IC50_x averaged over all epitope classes. NTD-targeting antibodies were not included in the DMS data set and hence we set

$$\text{IC50}_x(\text{NTD}) = \widehat{\text{IC50}}_x.$$

Notably, $\widehat{\text{IC50}}_x$ 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 $\widehat{\text{IC50}}_x(t_{\max}, t_{\text{half}})$ for each $t_{\max}, t_{\text{half}}$ 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

$$\underset{\widehat{\text{IC50}}_x(t_{\max}, t_{\text{half}})}{\operatorname{argmin}} \sum_t \left\| P_{\text{Neut}}(t, \text{Wuhan}, \text{Delta}, \widehat{\text{IC50}}_x(t_{\max}, t_{\text{half}})) - VE(t, \text{Wuhan}, \text{Delta}) \right\|^2.$$

where $VE(t, \text{Wuhan}, \text{Delta})$ 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 $\approx P_{\text{Neut}}$.

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

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

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.:

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

433 In our simulations, we considered identical PK for antibodies of the different epitope classes.
 434 Utilized parameters ($t_{\max}, t_{\text{half}}$) were extracted from literature: t_{\max} varied between 14 and
 435 28 days after antigen exposure [63, 38, 64, 65], while t_{half} ranged between 25 and 69 days
 436 [66, 67, 64, 30, 68, 69, 70, 71]. For simulation, we used different combinations of ($t_{\max}, t_{\text{half}}$)
 437 in a 5×15 grid within a range of t_{\max} within [14, 28] and t_{half} within [25, 69] and plotted the
 438 range of predictions (min, max).

439 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

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

440 where \mathcal{X} denotes the set of all variants present within the time horizon of interest, $P_{\text{Neut}(t-s,x,y)}$
 441 denotes the probability that an infection with strain x , that occurred $t - s$ days ago cross-
 442 neutralizes a variant y . In the equation above, $\pi_x(s)$ denotes the proportion of variant x at
 443 day $s < t$, derived from the molecular surveillance of SARS-CoV-2 (**Suppl. Data D1**), while
 444 $I(s)$ denotes the number of infected individuals at some previous time point s . As our best
 445 case for under-reporting-corrected infection numbers, we used the GInPipe-derived estimate
 446 $I_{\min}(s)$, as derived above.
 447 The expected number of susceptible individuals to a variant y at time t is then calculated
 448 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
 449 Germany.

450 Variant dynamics

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

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

451 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
 452 at time t and where $\alpha_x > 0$ denotes a variants' intrinsic (antibody-independent) transmission
 453 fitness, which we assumed to be near identical for all currently circulating variants $\alpha_x \approx \alpha$,
 454 implying that variant dynamics are dominated by infection history and immune dynamics.
 455 We ignored low abundance variants with $\pi_x(t) < 1\%$ and re-normalized accordingly.

456 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 $\mathbf{F}(t) = \text{diag}([\dots, f_y(t)/\bar{f}(t), \dots])$. Consequently, for $p_y(t) > 0$, we get

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

457 if fitness is determined by population immunity.

458 **Population susceptibility**

The total number of susceptibles was calculated as

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

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

460 **Variant proportions in the US**

461 The proportions of SARS-CoV-2 variants in the USA were obtained from the Centers for
462 Disease Control and Prevention's website¹¹ on July 18, 2023. This report provides weekly
463 estimations of the predominant SARS-CoV-2 variants, presented as proportions of sequences
464 under surveillance. The analysis covers a time span from July 31, 2021, as the earliest period,
465 and extends to May 6, 2023, marking the latest period in the report. Throughout this
466 duration, a total of 44 distinct SARS-CoV-2 variants were identified.

467 **Code availability**

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

473 **Author Contributions**

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

478 **Acknowledgements**

479 We are very grateful to the German Electronic Sequence Data Hub (DESH) and all of its data
480 contributors, i.e., the authors from the originating laboratories responsible for obtaining the
481 specimens and the submitting laboratories where genetic sequence data were generated and
482 shared. We further thank the Sequencing Core Facility of the Genome Competence Center
483 (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>

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

488 **Funding**

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

496 **Conflicts of interest**

497 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 circle 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 $400 \mu\text{g/mL}$. 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 $\text{FR}_{x,y}$ and fold resistances derived from neutralization assays.** Distinct markers show the epitope groups and distinct colors the spike-pseudo groups.

⁴⁹⁹ **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]

500 **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-mfl/covsonar>).

501 **References**

- 502 [1] World Health Organization (WHO). Statement on the fifteenth meeting of the ihr (2005)
503 emergency committee on the covid-19 pandemic. <https://www.who.int>, 2023. Accessed:
504 2023-09-08.
- 505 [2] Office for National Statistics (ONS). Coronavirus (COVID-19) infection survey, uk sta-
506 tistical bulletins, 2023. Accesed: 2023-05-15.
- 507 [3] World Health Organization (WHO). Weekly epidemiological update on covid-19 - 1 june
508 2023. <https://www.who.int>, 2023. Accessed: 2023-09-08.
- 509 [4] Hannah E Davis, Lisa McCorkell, Julia Moore Vogel, and Eric J Topol. Long covid: major
510 findings, mechanisms and recommendations. *Nature Reviews Microbiology*, 21(3):133–
511 146, 2023.
- 512 [5] World Health Organization (WHO). Coronavirus disease (covid-19) pandemic.
513 <https://www.who.int>, 2023. Accessed: 2023-09-08.
- 514 [6] Peter V Markov, Mahan Ghafari, Martin Beer, Katrina Lythgoe, Peter Simmonds, Niko-
515 laos I Stilianakis, and Aris Katzourakis. The evolution of sars-cov-2. *Nature Reviews*
516 *Microbiology*, 21(6):361–379, 2023.
- 517 [7] Yanjia Chen, Xiaoyu Zhao, Hao Zhou, Huanzhang Zhu, Shibo Jiang, and Pengfei Wang.
518 Broadly neutralizing antibodies to sars-cov-2 and other human coronaviruses. *Nature*
519 *reviews Immunology*, 23(3):189–199, 2023.
- 520 [8] Yixuan J Hou, Shiho Chiba, Peter Halfmann, Camille Ehre, Makoto Kuroda, Kenneth H
521 Dinnon III, Sarah R Leist, Alexandra Schäfer, Noriko Nakajima, Kenta Takahashi, et al.
522 Sars-cov-2 d614g variant exhibits efficient replication ex vivo and transmission in vivo.
523 *Science*, 370(6523):1464–1468, 2020.
- 524 [9] Bette Korber, Will M Fischer, Sandrasegaram Gnanakaran, Hyejin Yoon, James Theiler,
525 Werner Abfalterer, Nick Hengartner, Elena E Giorgi, Tanmoy Bhattacharya, Brian Foley,
526 et al. Tracking changes in sars-cov-2 spike: evidence that d614g increases infectivity of
527 the covid-19 virus. *Cell*, 182(4):812–827, 2020.
- 528 [10] Jessica A Plante, Yang Liu, Jianying Liu, Hongjie Xia, Bryan A Johnson, Kumari G
529 Lokugamage, Xianwen Zhang, Antonio E Muruato, Jing Zou, Camila R Fontes-Garfias,
530 et al. Spike mutation d614g alters sars-cov-2 fitness. *Nature*, 592(7852):116–121, 2021.
- 531 [11] Ewen Callaway et al. The mutation that helps delta spread like wildfire. *Nature*,
532 596(7873):472–473, 2021.
- 533 [12] Emanuele Andreano, Giulia Piccini, Danilo Licastro, Lorenzo Casalino, Nicole V John-
534 son, Ida Paciello, Simeone Dal Monego, Elisa Pantano, Noemi Manganaro, Alessandro
535 Manenti, et al. Sars-cov-2 escape from a highly neutralizing covid-19 convalescent plasma.
536 *Proceedings of the National Academy of Sciences*, 118(36):e2103154118, 2021.
- 537 [13] Pengfei Wang, Manoj S Nair, Lihong Liu, Sho Iketani, Yang Luo, Yicheng Guo, Maple
538 Wang, Jian Yu, Baoshan Zhang, Peter D Kwong, et al. Antibody resistance of sars-cov-2
539 variants b. 1.351 and b. 1.1. 7. *Nature*, 593(7857):130–135, 2021.

- 540 [14] Pengfei Wang, Ryan G Casner, Manoj S Nair, Maple Wang, Jian Yu, Gabriele Cerutti,
541 Lihong Liu, Peter D Kwong, Yaoxing Huang, Lawrence Shapiro, et al. Increased re-
542 sistance of sars-cov-2 variant p. 1 to antibody neutralization. *Cell host & microbe*,
543 29(5):747–751, 2021.
- 544 [15] Wanwisa Dejnirattisai, Jiandong Huo, Daming Zhou, Jiří Zahradník, Piyada Supasa,
545 Chang Liu, Helen ME Duyvesteyn, Helen M Ginn, Alexander J Mentzer, Aekkachai
546 Tuekprakhon, et al. Sars-cov-2 omicron-b. 1.1. 529 leads to widespread escape from
547 neutralizing antibody responses. *Cell*, 185(3):467–484, 2022.
- 548 [16] Matthew McCallum, Nadine Czudnochowski, Laura E Rosen, Samantha K Zepeda,
549 John E Bowen, Alexandra C Walls, Kevin Hauser, Anshu Joshi, Cameron Stewart,
550 Josh R Dillen, et al. Structural basis of sars-cov-2 omicron immune evasion and re-
551 ceptor engagement. *Science*, 375(6583):864–868, 2022.
- 552 [17] Rungtiwa Nutalai, Daming Zhou, Aekkachai Tuekprakhon, Helen M Ginn, Piyada Su-
553 pasa, Chang Liu, Jiandong Huo, Alexander J Mentzer, Helen ME Duyvesteyn, Aiste
554 Dijokaitė-Guraliuc, et al. Potent cross-reactive antibodies following omicron break-
555 through in vaccinees. *Cell*, 185(12):2116–2131, 2022.
- 556 [18] Frederik Plesner Lyngse, Carsten Thure Kirkeby, Matthew Denwood, Lasse Engbo Chris-
557 tiansen, Kåre Mølbak, Camilla Holten Møller, Robert Leo Skov, Tyra Grove Krause,
558 Morten Rasmussen, Raphael Niklaus Sieber, et al. Household transmission of sars-
559 cov-2 omicron variant of concern subvariants ba. 1 and ba. 2 in denmark. *Nature*
560 *Communications*, 13(1):5760, 2022.
- 561 [19] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby
562 Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar,
563 et al. A global panel database of pandemic policies (oxford covid-19 government response
564 tracker). *Nature human behaviour*, 5(4):529–538, 2021.
- 565 [20] Nathalie Worp, Lorenzo Subissi, Mark D Perkins, Maria D Van Kerkhove, Anurag
566 Agrawal, Meera Chand, Janko van Beek, Bas B Oude Munnink, and Marion PG Koop-
567 mans. Towards the development of a sars-cov-2 variant risk assessment tool: expert
568 consultation on the assessment of scientific evidence on emerging variants. *The Lancet*
569 *Microbe*, 2023.
- 570 [21] World Health Organization (WHO). Statement on the antigen composition of covid-19
571 vaccines. <https://www.who.int>, 2023. Accessed: 2023-09-08.
- 572 [22] Allison J Greaney, Tyler N Starr, and Jesse D Bloom. An antibody-escape estimator for
573 mutations to the SARS-CoV-2 receptor-binding domain. *Virus Evolution*, 8(1):veac021,
574 2022.
- 575 [23] Allison J Greaney, Tyler N Starr, Pavlo Gilchuk, Seth J Zost, Elad Binshtein, Andrea N
576 Loes, Sarah K Hilton, John Huddleston, Rachel Eguia, Katharine H D Crawford, Adam S
577 Dingens, Rachel S Nargi, Rachel E Sutton, Naveenchandra Suryadevara, Paul W Roth-
578 lauf, Zhuoming Liu, Sean P J Whelan, Robert H Carnahan, James E Crowe, Jr, and
579 Jesse D Bloom. Complete mapping of mutations to the SARS-CoV-2 spike receptor-
580 binding domain that escape antibody recognition. *Cell Host Microbe*, 29(1):44–57.e9,
581 2021.

- 582 [24] Yunlong Cao, Fanchong Jian, Jing Wang, Yuanling Yu, Weiliang Song, Ayijiang Yisi-
583 mayi, Jing Wang, Ran An, Xiaosu Chen, Na Zhang, Yao Wang, Peng Wang, Lijuan
584 Zhao, Haiyan Sun, Lingling Yu, Sijie Yang, Xiao Niu, Tianhe Xiao, Qingqing Gu, Fei
585 Shao, Xiaohua Hao, Yanli Xu, Ronghua Jin, Zhongyang Shen, Youchun Wang, and
586 Xiaoliang Sunney Xie. Imprinted SARS-CoV-2 humoral immunity induces convergent
587 omicron RBD evolution. *Nature*, 614(7948):521–529, 2023.
- 588 [25] Djin Ye Oh, Martin Hölzer, Sofia Paraskevopoulou, Maria Trofimova, Felix Hartkopf,
589 Matthias Budt, Marianne Wedde, Hugues Richard, Berit Haldemann, Teresa Do-
590 maszewska, et al. Advancing precision vaccinology by molecular and genomic surveillance
591 of severe acute respiratory syndrome coronavirus 2 in germany, 2021. *Clinical infectious*
592 *diseases*, 75(Supplement_1):S110–S120, 2022.
- 593 [26] Maureen Rebecca Smith, Maria Trofimova, Ariane Weber, Yannick Duport, Denise
594 Kühnert, and Max von Kleist. Rapid incidence estimation from sars-cov-2 genomes re-
595 veals decreased case detection in europe during summer 2020. *Nature Communications*,
596 12(1):6009, 2021.
- 597 [27] Katharina Jahn, David Dreifuss, Ivan Topolsky, Anina Kull, Pravin Ganesanandamoor-
598 thy, Xavier Fernandez-Cassi, Carola Bänziger, Alexander J Devaux, Elyse Stachler, Lea
599 Caduff, et al. Early detection and surveillance of sars-cov-2 genomic variants in wastew-
600 ater using cojac. *Nature Microbiology*, 7(8):1151–1160, 2022.
- 601 [28] Yunlong Cao, Jing Wang, Fanchong Jian, Tianhe Xiao, Weiliang Song, Ayijiang Yisi-
602 mayi, Weijin Huang, Qianqian Li, Peng Wang, Ran An, Jing Wang, Yao Wang, Xiao
603 Niu, Sijie Yang, Hui Liang, Haiyan Sun, Tao Li, Yuanling Yu, Qianqian Cui, Shuo Liu,
604 Xiaodong Yang, Shuo Du, Zhiying Zhang, Xiaohua Hao, Fei Shao, Ronghua Jin, Xiangxi
605 Wang, Junyu Xiao, Youchun Wang, and Xiaoliang Sunney Xie. Omicron escapes the
606 majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*, 602(7898):657–663,
607 2022.
- 608 [29] Matthew McCallum, Anna De Marco, Florian A Lempp, M Alejandra Tortorici, Dora
609 Pinto, Alexandra C Walls, Martina Beltramello, Alex Chen, Zhuoming Liu, Fabrizia
610 Zatta, Samantha Zepeda, Julia di Julio, John E Bowen, Martin Montiel-Ruiz, Jiayi Zhou,
611 Laura E Rosen, Siro Bianchi, Barbara Guarino, Chiara Silacci Fregni, Rana Abdelnabi,
612 Shi-Yan Caroline Foo, Paul W Rothlauf, Louis-Marie Bloyet, Fabio Benigni, Elisabetta
613 Cameroni, Johan Neyts, Agostino Riva, Gyorgy Snell, Amalio Telenti, Sean P J Whe-
614 lan, Herbert W Virgin, Davide Corti, Matteo Samuele Pizzuto, and David Veesler. N-
615 terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell*,
616 184(9):2332–2347.e16, 2021.
- 617 [30] David S Khoury, Deborah Cromer, Arnold Reynaldi, Timothy E Schlub, Adam K Wheat-
618 ley, Jennifer A Juno, Kanta Subbarao, Stephen J Kent, James A Triccas, and Miles P
619 Davenport. Neutralizing antibody levels are highly predictive of immune protection from
620 symptomatic sars-cov-2 infection. *Nature medicine*, 27(7):1205–1211, 2021.
- 621 [31] Richard J. Kryscio, Marta S. Mendiondo, Frederick A. Schmitt, and William R. Markes-
622 bery. Designing a large prevention trial: statistical issues. *Statistics in Medicine*,
623 23(2):285–296, 2004.

- 624 [32] Bo Meng, Adam Abdullahi, Isabella ATM Ferreira, Niluka Goonawardane, Akatsuki
625 Saito, Izumi Kimura, Daichi Yamasoba, Pehuén Pereyra Gerber, Saman Fatihi, Surabhi
626 Rathore, et al. Altered tmprss2 usage by sars-cov-2 omicron impacts infectivity and
627 fusogenicity. *Nature*, 603(7902):706–714, 2022.
- 628 [33] Anupriya Aggarwal, Anouschka Akerman, Vanessa Milogiannakis, Mariana Ruiz Silva,
629 Gregory Walker, Alberto Ospina Stella, Andrea Kindinger, Thomas Angelovich, Emily
630 Waring, Supavadee Amatayakul-Chantler, et al. Sars-cov-2 omicron ba. 5: Evolving
631 tropism and evasion of potent humoral responses and resistance to clinical immunother-
632 apeutics relative to viral variants of concern. *EBioMedicine*, 84, 2022.
- 633 [34] Niklas Bobrovitz, Harriet Ware, Xiaomeng Ma, Zihan Li, Reza Hosseini, Christian Cao,
634 Anabel Seleton, Mairead Whelan, Zahra Premji, Hanane Issa, et al. Protective effective-
635 ness of previous sars-cov-2 infection and hybrid immunity against the omicron variant
636 and severe disease: a systematic review and meta-regression. *The Lancet Infectious
637 Diseases*, 23(5):556–567, 2023.
- 638 [35] Daniel R Feikin, Melissa M Higdon, Laith J Abu-Raddad, Nick Andrews, Rafael Araos,
639 Yair Goldberg, Michelle J Groome, Amit Huppert, Katherine L O'Brien, Peter G Smith,
640 Annelies Wilder-Smith, Scott Zeger, Maria Deloria Knoll, and Minal K Patel. Duration
641 of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results
642 of a systematic review and meta-regression. *Lancet*, 399(10328):924–944, 2022.
- 643 [36] Oliver J Watson, Gregory Barnsley, Jaspreet Toor, Alexandra B Hogan, Peter Winskill,
644 and Azra C Ghani. Global impact of the first year of covid-19 vaccination: a mathemat-
645 ical modelling study. *The Lancet Infectious Diseases*, 22(9):1293–1302, 2022.
- 646 [37] Alessandro Sette and Shane Crotty. Adaptive immunity to sars-cov-2 and covid-19. *Cell*,
647 184(4):861–880, 2021.
- 648 [38] Baolin Liao, Zhao Chen, Peiyan Zheng, Linghua Li, Jianfen Zhuo, Fang Li, Suxiang
649 Li, Dingbin Chen, Chunyan Wen, Weiping Cai, Shanhui Wu, Yanhong Tang, Linwei
650 Duan, Peilan Wei, Fangli Chen, Jinwei Yuan, Jinghong Yang, Jiaxin Feng, Jingxian
651 Zhao, Jincun Zhao, Baoqing Sun, Airu Zhu, Yimin Li, and Xiaoping Tang. Detection
652 of anti-sars-cov-2-s2 igg is more sensitive than anti-rbd igg in identifying asymptomatic
653 covid-19 patients. *Frontiers in Immunology*, 12:724763, 2021.
- 654 [39] Carolyn Rydznski Moderbacher, Sydney I Ramirez, Jennifer M Dan, Alba Grifoni,
655 Kathryn M Hastie, Daniela Weiskopf, Simon Belanger, Robert K Abbott, Christina
656 Kim, Jinyong Choi, et al. Antigen-specific adaptive immunity to sars-cov-2 in acute
657 covid-19 and associations with age and disease severity. *Cell*, 183(4):996–1012, 2020.
- 658 [40] Anthony T Tan, Martin Linster, Chee Wah Tan, Nina Le Bert, Wan Ni Chia, Kamini
659 Kunasegaran, Yan Zhuang, Christine YL Tham, Adeline Chia, Gavin JD Smith, et al.
660 Early induction of functional sars-cov-2-specific t cells associates with rapid viral clear-
661 ance and mild disease in covid-19 patients. *Cell reports*, 34(6):108728, 2021.
- 662 [41] Alessandro Sette, John Sidney, and Shane Crotty. T cell responses to sars-cov-2. *Annual
663 Review of Immunology*, 41(1):343–373, 2023.

- 664 [42] Gregory A Poland, Inna G Ovsyannikova, and Richard B Kennedy. Sars-cov-2 immunity:
665 review and applications to phase 3 vaccine candidates. *The Lancet*, 396(10262):1595–
666 1606, 2020.
- 667 [43] Jesse D. Bloom and Richard A. Neher. Fitness effects of mutations to sars-cov-2 proteins.
668 *bioRxiv*, 2023.
- 669 [44] Hannah McClymont and Wenbiao Hu. Weather variability and covid-19 transmission:
670 a review of recent research. *International journal of environmental research and public
671 health*, 18(2):396, 2021.
- 672 [45] Yiqun Ma, Sen Pei, Jeffrey Shaman, Robert Dubrow, and Kai Chen. Role of meteorological
673 factors in the transmission of sars-cov-2 in the united states. *Nature Communications*,
674 12(1):3602, 2021.
- 675 [46] Jordi Landier, Juliette Paireau, Stanislas Rebaudet, Eva Legendre, Laurent Lehut, Ar-
676 naud Fontanet, Simon Cauchemez, and Jean Gaudart. Cold and dry winter conditions
677 are associated with greater sars-cov-2 transmission at regional level in western countries
678 during the first epidemic wave. *Scientific Reports*, 11(1):12756, 2021.
- 679 [47] Mohammad M Sajadi, Parham Habibzadeh, Augustin Vintzileos, Shervin Shokouhi, Fer-
680 nando Miralles-Wilhelm, and Anthony Amoroso. Temperature, humidity, and latitude
681 analysis to estimate potential spread and seasonality of coronavirus disease 2019 (covid-
682 19). *JAMA network open*, 3(6):e2011834, 2020.
- 683 [48] Simiao Chen, Klaus Prettner, Michael Kuhn, Pascal Geldsetzer, Chen Wang, Till
684 Bärnighausen, and David E Bloom. Climate and the spread of covid-19. *Scientific
685 Reports*, 11(1):9042, 2021.
- 686 [49] Tomáš Gavenčíak, Joshua Teperowski Monrad, Gavin Leech, Mrinank Sharma, Sören
687 Mindermann, Samir Bhatt, Jan Brauner, and Jan Kulveit. Seasonal variation in sars-
688 cov-2 transmission in temperate climates: A bayesian modelling study in 143 european
689 regions. *PLoS Computational Biology*, 18(8):e1010435, 2022.
- 690 [50] Canelle Poirier, Wei Luo, Maimuna S Majumder, Dianbo Liu, Kenneth D Mandl, Todd A
691 Mooring, and Mauricio Santillana. The role of environmental factors on transmission
692 rates of the covid-19 outbreak: an initial assessment in two spatial scales. *Scientific
693 reports*, 10(1):17002, 2020.
- 694 [51] Colin J Carlson, Ana CR Gomez, Shweta Bansal, and Sadie J Ryan. Misconcep-
695 tions about weather and seasonality must not misguide covid-19 response. *Nature
696 communications*, 11(1):4312, 2020.
- 697 [52] Riccardo Cappi, Luca Casini, Davide Tosi, and Marco Roccetti. Questioning the season-
698 ality of sars-cov-2: A fourier spectral analysis. *BMJ open*, 12(4):e061602, 2022.
- 699 [53] Rachel E Baker, Wenchang Yang, Gabriel A Vecchi, C Jessica E Metcalf, and Bryan T
700 Grenfell. Susceptible supply limits the role of climate in the early sars-cov-2 pandemic.
701 *Science*, 369(6501):315–319, 2020.

- 702 [54] Gavin Y. Oudit, Kaiming Wang, Anissa Viveiros, Max J. Kellner, and Josef M. Pen-
703 nninger. Angiotensin-converting enzyme 2—at the heart of the covid-19 pandemic. *Cell*,
704 186(5):906–922, 2023.
- 705 [55] Manuel Hayn, Maximilian Hirschenberger, Lennart Koepke, Rayhane Nchioua, Jan Hen-
706 drik Straub, Susanne Klute, Victoria Hunszinger, Fabian Zech, Caterina Prelli Bozzo,
707 Wasim Aftab, Maria Hønholt Christensen, Carina Conzelmann, Janis Alexander Müller,
708 Smitha Srinivasachar Badarinarayanan, Christina Martina Stürzel, Ignasi Forne, Steffen
709 Stenger, Karl-Klaus Conzelmann, Jan Münch, Florian Ingo Schmidt, Daniel Sauter, Axel
710 Imhof, Frank Kirchhoff, and Konstantin Maria Johannes Sparrer. Systematic functional
711 analysis of sars-cov-2 proteins uncovers viral innate immune antagonists and remaining
712 vulnerabilities. *Cell Reports*, 35(7):109126, 2021.
- 713 [56] Judith M Minkoff and Benjamin tenOever. Innate immune evasion strategies of sars-cov-
714 2. *Nature Reviews Microbiology*, 21(3):178–194, 2023.
- 715 [57] N. Alexia Raharinirina, Nils Gubela, Daniela Börnigen, and et al. Sars-cov-2 evolution on
716 a dynamic immune landscape. *Zenodo*, 2023, <https://doi.org/10.5281/zenodo.8349296>.
- 717 [58] Andrew Rambaut, Edward C Holmes, Áine O’Toole, Verity Hill, John T McCrone,
718 Christopher Ruis, Louis du Plessis, and Oliver G Pybus. A dynamic nomenclature
719 proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.*,
720 5(11):1403–1407, 2020.
- 721 [59] Áine O’Toole, Emily Scher, Anthony Underwood, Ben Jackson, Verity Hill, John T Mc-
722 Crone, Rachel Colquhoun, Chris Ruis, Khalil Abu-Dahab, Ben Taylor, Corin Yeats, Louis
723 du Plessis, Daniel Maloney, Nathan Medd, Stephen W Attwood, David M Aanensen, Ed-
724 ward C Holmes, Oliver G Pybus, and Andrew Rambaut. Assignment of epidemiological
725 lineages in an emerging pandemic using the pangolin tool. *Virus Evol.*, 7(2):veab064,
726 2021.
- 727 [60] Wiep van der Toorn, Djin-Ye Oh, Daniel Bourquin, Janine Michel, Eva Krause, Andreas
728 Nitsche, and Max von Kleist. An intra-host sars-cov-2 dynamics model to assess testing
729 and quarantine strategies for incoming travelers, contact management, and de-isolation.
730 *Patterns*, 2(6):100262, 2021.
- 731 [61] Yunlong Cao, Ayijiang Yisimayi, Fanchong Jian, Weiliang Song, Tianhe Xiao, Lei Wang,
732 Shuo Du, Jing Wang, Qianqian Li, Xiaosu Chen, Yuanling Yu, Peng Wang, Zhiying
733 Zhang, Pulan Liu, Ran An, Xiaohua Hao, Yao Wang, Jing Wang, Rui Feng, Haiyan Sun,
734 Lijuan Zhao, Wen Zhang, Dong Zhao, Jiang Zheng, Lingling Yu, Can Li, Na Zhang, Rui
735 Wang, Xiao Niu, Sijie Yang, Xuetao Song, Yangyang Chai, Ye Hu, Yansong Shi, Linlin
736 Zheng, Zhiqiang Li, Qingqing Gu, Fei Shao, Weijin Huang, Ronghua Jin, Zhongyang
737 Shen, Youchun Wang, Xiangxi Wang, Junyu Xiao, and Xiaoliang Sunney Xie. BA.2.12.1,
738 BA.4 and BA.5 escape antibodies elicited by omicron infection. *Nature*, 608(7923):593–
739 602, 2022.
- 740 [62] Bernadeta Dadonaite, Katharine H D Crawford, Caelan E Radford, Ariana G Farrell,
741 Timothy C Yu, William W Hannon, Panpan Zhou, Raees Andrabi, Dennis R Burton,
742 Lihong Liu, David D Ho, Helen Y Chu, Richard A Neher, and Jesse D Bloom. A

- 743 pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike.
744 *Cell*, 186(6):1263–1278.e20, 2023.
- 745 [63] Sheila F Lumley, Jia Wei, Denise O'Donnell, Nicole E Stoesser, Philippa C Matthews,
746 Alison Howarth, Stephanie B Hatch, Brian D Marsden, Stuart Cox, Tim James, Liam J
747 Peck, Thomas G Ritter, Zoe de Toledo, Richard J Cornall, E Yvonne Jones, David I Stu-
748 art, Gavin Screaton, Daniel Ebner, Sarah Hoosdally, Derrick W Crook, Christopher P
749 Conlon, Koen B Pouwels, A Sarah Walker, Tim E A Peto, Timothy M Walker, Katie
750 Jeffery, David W Eyre, and Oxford University Hospitals Staff Testing Group. The Dura-
751 tion, Dynamics, and Determinants of Severe Acute Respiratory Syndrome Coronavirus 2
752 (SARS-CoV-2) Antibody Responses in Individual Healthcare Workers. *Clinical Infectious*
753 *Diseases*, 73(3):e699–e709, 2021.
- 754 [64] Jian-Hang Xue, Yong-Jing Wang, Wei Li, Qiu-Ling Li, Qiu-Yan Xu, Jian-Jun Niu, and
755 Li-Li Liu. Anti-receptor-binding domain immunoglobulin g antibody as a predictor
756 of seropositivity for anti-sars-cov-2 neutralizing antibody. *Archives of Pathology &*
757 *Laboratory Medicine*, 146(7):814–821, 2022.
- 758 [65] Irina Arkhipova-Jenkins, Mark Helfand, Charlotte Armstrong, Emily Gean, Joanna An-
759 derson, Robin A Paynter, and Katherine Mackey. Antibody response after sars-cov-2 in-
760 fection and implications for immunity: a rapid living review. *Annals of internal medicine*,
761 174(6):811–821, 2021.
- 762 [66] Jean-Louis Bayart, Jonathan Douxfils, Constant Gillot, Clara David, François Mullier,
763 Marc Elsen, Christine Eucher, Sandrine Van Eckhoudt, Tatiana Roy, Vincent Gerin,
764 Grégoire Wieers, Christine Laurent, Mélanie Closset, Jean-Michel Dogné, and Julien
765 Favresse. Waning of igg, total and neutralizing antibodies 6 months post-vaccination
766 with bnt162b2 in healthcare workers. *Vaccines*, 9(10):1092, 2021.
- 767 [67] Thomas W. Barnes, Johannes Schulte-Pelkum, Laura Steller, Daniel Filchtinski, Robin
768 Jenness, Michelle R. Williams, Christina Kober, Sandro Manni, Thomas Hauser, Aaron
769 Hahn, Uwe Kalina, Toby L. Simon, Patrick Schuetz, and Nathan J. Roth. Determination
770 of neutralising anti-sars-cov-2 antibody half-life in covid-19 convalescent donors. *Clinical*
771 *Immunology*, 232:108871, 2021.
- 772 [68] Alicia T. Widge, Nadine G. Roush, Lisa A. Jackson, Evan J. Anderson, Paul C.
773 Roberts, Mamodikoe Makhene, James D. Chappell, Mark R. Denison, Laura J. Stevens,
774 Andrea J. Pruijssers, Adrian B. McDermott, Britta Flach, Bob C. Lin, Nicole A. Doria-
775 Rose, Siyy O'Dell, Stephen D. Schmidt, Kathleen M. Neuzil, Hamilton Bennett, Brett
776 Leav, Mat Makowski, Jim Albert, Kaitlyn Cross, Venkata-Viswanadh Edara, Katharine
777 Floyd, Mehul S. Suthar, Wendy Buchanan, Catherine J. Luke, Julie E. Ledgerwood,
778 John R. Mascola, Barney S. Graham, and John H. Beigel. Durability of responses after
779 sars-cov-2 mrna-1273 vaccination. *New England Journal of Medicine*, 384(1):80–82, 2021.
- 780 [69] Jennifer M. Dan, Jose Mateus, Yu Kato, Kathryn M. Hastie, Esther Dawen Yu, Cate-
781 rina E. Faliti, Alba Grifoni, Sydney I. Ramirez, Sonya Haupt, April Frazier, Catherine
782 Nakao, Vamseedhar Rayaprolu, Stephen A. Rawlings, Bjoern Peters, Florian Krammer,
783 Viviana Simon, Erica Ollmann Saphire, Davey M. Smith, Daniela Weiskopf, Alessan-
784 dro Sette, and Shane Crotty. Immunological memory to sars-cov-2 assessed for up to 8
785 months after infection. *Science*, 371(6529):eabf4063, 2021.

- 786 [70] Adam K Wheatley, Jennifer A Juno, Jing J Wang, Kevin J Selva, Arnold Reynaldi,
787 Hyon-Xhi Tan, Wen Shi Lee, Kathleen M Wragg, Hannah G Kelly, Robyn Esterbauer,
788 et al. Evolution of immune responses to sars-cov-2 in mild-moderate covid-19. *Nature*
789 *communications*, 12(1):1162, 2021.
- 790 [71] Jan Van Elslande, Lien Gruwier, Lode Godderis, and Pieter Vermeersch. Estimated
791 Half-Life of SARS-CoV-2 Anti-Spike Antibodies More Than Double the Half-Life of
792 Anti-nucleocapsid Antibodies in Healthcare Workers. *Clinical Infectious Diseases*,
793 73(12):2366–2368, 2021.
- 794 [72] Alexander S Bratus, Artem S Novozhilov, and Yuri S Semenov. *Advanced Mathematical*
795 *Methods in Biosciences and Applications*, chapter Rigorous mathematical analysis of the
796 quasispecies model: From Manfred Eigen to the recent developments. Springer, 2019.
- 797 [73] Loris Bennett, Bernd Melchers, and Boris Proppe. Curta: A general-
798 purpose high-performance computer at zedat, freie universität berlin.
799 <http://dx.doi.org/10.17169/refubium-26754>, 2020.
- 800 [74] Mie Agermose Gram, Hanne-Dorthe Emborg, Astrid Blicher Schelde, Nikolaj Ulrik Friis,
801 Katrine Finderup Nielsen, Ida Rask Moustsen-Helms, Rebecca Legarth, Janni Uyen Hoa
802 Lam, Manon Chaine, Aisha Zahoor Malik, et al. Vaccine effectiveness against sars-cov-2
803 infection or covid-19 hospitalization with the alpha, delta, or omicron sars-cov-2 variant:
804 A nationwide danish cohort study. *PLoS medicine*, 19(9):e1003992, 2022.
- 805 [75] Mie Agermose Gram, Hanne-Dorthe Emborg, Astrid Blicher Schelde, Nikolaj Ulrik
806 Friis, Katrine Finderup Nielsen, Ida Rask Moustsen-Helms, Rebecca Legarth, Janni
807 Uyen Hoa Lam, Manon Chaine, Aisha Zahoor Malik, Morten Rasmussen, Jannik Fon-
808 ager, Raphael Niklaus Sieber, Marc Stegger, Steen Ethelberg, Palle Valentiner-Branth,
809 and Christian Holm Hansen. Vaccine effectiveness against SARS-CoV-2 infection or
810 COVID-19 hospitalization with the alpha, delta, or omicron SARS-CoV-2 variant: A
811 nationwide danish cohort study. *PLoS Med*, 19(9):e1003992, 2022.
- 812 [76] Sara Y Tartof, Jeff M Slezak, Heidi Fischer, Vennis Hong, Bradley K Ackerson, Omesh N
813 Ranasinghe, Timothy B Frankland, Oluwaseye A Ogun, Joann M Zamparo, Sharon Gray,
814 Sriniwas R Valluri, Kaije Pan, Frederick J Angulo, Luis Jodar, and John M McLaughlin.
815 Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large inte-
816 grated health system in the USA: a retrospective cohort study. *Lancet*, 398(10309):1407–
817 1416, 2021.

Supplementary Files

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

- [SupplementaryDataFile.csv](#)
- [SupplementaryFiguresTables.pdf](#)