

Advancing precision vaccinology by molecular and genomic surveillance of SARS-CoV-2 in Germany, 2021

Oh, Hölzer, et al.

Supplementary Information

1. Supplementary methods

2. Supplementary tables and figures

- Suppl. Table S1: Variants of Concern (VOCs) detected in the IMS-SC2
- Suppl. Table S2: Age Group Distribution of IMS-SC2 Random Sample
- Suppl. Table S3: Selected variations (mutations of concern) within the Spike Protein of SARS-CoV-2
- Suppl. Figure S1: Stream graph representing lineages identified in the representative sampling subset of IMS-SC2 genome sequences.
- Suppl. Figure S2: Phylogenetic tree based on the representative sampling subset of IMS-SC2 genome sequences
- Suppl. Figure S3: Graphic depiction of selected variations (mutations of concern)

SUPPLEMENTARY METHODS

Sample selection: For representative sampling, IMS-SC2 network laboratories randomly select SARS-CoV-2 specimens each week which fulfill the following criteria: (i) non-identical zip codes in order to sample cases that are geographically distinct and to minimize the chance of specimens being sampled from the same cluster, (ii) Ct values below 23, indicating higher viral loads which are per experience associated with more accurate whole genome sequencing results. In addition, targeted sampling is performed on an as-needed-basis on samples of particular clinical or epidemiological interest. These may originate from outbreaks, vaccine breakthrough infections, re-infections, or travelers returning from countries with high VOC or new variant prevalence. Specimens are sent to RKI at room temperature using postal or courier services. To ensure seamless sample tracking, sample labels and shipping material are provided upfront (prior to sending) by RKI to network labs.

RNA extraction: Total RNA was extracted from upper respiratory tract specimens (nasal, nasopharyngeal, oropharyngeal, or combined nasal/oropharyngeal swabs) using the MagNA Pure 96 DNA and Viral NA Small Volume Kit (Roche Life Science, Mannheim, Germany) and the MagNA Pure 96 instrument (Roche Life Science) according to the manufacturer's instructions. For about 5% of sequenced samples, RNA extractions were performed on-site at the network labs.

Sequencing and genome reconstruction: Sequencing was performed in-house using either Illumina or Nanopore tiled amplicons, For Illumina, SC2 amplicon-libraries were generated employing the CleanPlex SARS-Cov-2 amplicon panel (Paragon Genomics) and sequencing was performed on an iSeq 100 system (Illumina) with 150 bp paired-end reads, yielding between 100-200k reads per sample. Libraries for nanopore sequencing were generated using NEBNext ARTIC SARS-CoV-2 Companion Kit (Oxford Nanopore Technologies) with the ARTIC V3, V4, and V4.1 primer sets (<https://github.com/artic-network/artic-ncov2019>) and sequencing was performed on MinION and GridION instruments (Oxford Nanopore), resulting in an average of 116k reads per sample and that could be attributed to SARS-CoV-2. Consensus genomes were then reconstructed via covPipe [1] and poreCov [2], respectively. After filtering out samples that failed the reconstruction and genomes with more than 5% N bases, these sequencing and reconstruction approaches resulted in high-quality consensus sequences with, on average, an N content of 1.6% (Nanopore-derived sequences) and 1.8% (Illumina) per genome.

Virus strains: Virus isolates used for neutralization experiments were lineage B (SARS-CoV-2/human/DEU/BavPat2-ChVir984-ChVir1017/2020, Genbank accession: MT270112), lineage B.1.1.7, VOC alpha (SARS-CoV-2 (hCoV-19/Germany/NW-RKI-I-0026/2020, GISAID accession: EPI_ISL_751799), lineage B.1.351, VOC beta (hCoV-19/Germany/NW-RKI-I-0029/2020, GISAID accession: EPI_ISL_803957), lineage P.2,

VOI zeta (hCov-19/Germany/BE-RKI-N-250/2021, ENA Sequence ID: IMSSC2-4-2021-01819) or lineage A.27 (ENA Sequence ID: IMSSC2-4-2021-03311), respectively.

Plaque reduction neutralisation test (PRNT₅₀): Briefly, 1.6E5 VeroE6 were plated in 24 well plates the day before. Sera obtained from BNT162b2-vaccinated healthcare workers [3] were 2-fold serially diluted and incubated with 50 PFU of SARS-CoV-2 isolates in a total volume of 200 µl for 1 h at 37°C. The mixture was then used to infect the cells for 1 h at 37°C. After aspiration of the inoculate, cells were grown for three days in avicel plaque medium and stained with crystal violet. The PRNT₅₀ titre represents the reciprocal value of the highest serum dilution that reduces plaque number by at least 50% compared to untreated infection. Statistical analysis was performed with the GraphPad Prism software version 9. Significance was determined by 2-way ANOVA.

IMS-SC2 data set preparation: We re-run poreCov v0.11.7 [2] in FASTA-mode and with default parameters on an initial data set (n=3714 sequences, date of analyses: 10th January, 2022) to obtain a harmonized and quality-controlled data set and up-to-date lineage annotations. For quality control, poreCov compares sequences against NC_045512.2 (Wuhan reference sequence) using a sequence similarity threshold of 90% and allowing up to 5% N bases per default. We removed 81 sequences with a percentage of ambiguous N bases above 5% ending up with 3633 sequences for lineage assignment. Lineages were assigned using Pangolin v3.1.17[4] with PangoLEARN release 2021-12-06. We then used covSonar v1.1.0 (<https://gitlab.com/s.fuchs/covsonar>) to generate a database with all sequence and metadata information for easy filtering and access to mutation profiles. Finally, we restricted the sampling period to 1st December, 2020 through December 31st, 2021, obtaining a final data set of 3623 IMS-SC2 sequences. This data set included random-selection (n=3282), targeted-selection (n=194) and unspecified-selection (n=147) samples. All IMS-SC2 results presented in Fig. 1 are based on the full data set (n=3623), whereas the results shown in Fig. 2 are based on the random selection (n=3282). Where not otherwise stated, downstream analyses were based on the random-selection data set.

German GISAID reference data set preparation: The GISAID database was filtered for sequences with sampling location “Germany” and with sampling dates for the time frame between 1st December, 2020 and 31st December, 2021 (as above). The resulting dataset contained n=330,357 sequences. Lineage assignments were used as provided by GISAID in the metadata download (obtained 11th January, 2022).

Geographical distribution analysis: We benchmarked the geographical distribution by creating a visualization utilizing a custom *Python* script using the *geopandas* package [<https://geopandas.org/en/stable/about/citing.html>] to create a map of 3-digit zip code areas. These areas, shown in Fig. 1A, were colored by the number of samples submitted to the IMS-SC2.

Phylogenetic tree inference: Genomes that belong to the randomly sampled IMS-SC2 dataset (n=3282) were aligned with MAFFT v7.490 using default parameters [5]. Phylogenetic inference was performed with IQ-TREE v2.1.4-beta [6] under the GTR+F+R2 evolutionary model, using 1000 ultra-fast bootstrap replicates [7] and the resulting tree visualized and colored via Iroki [8]. Phylogenetic analysis revealed two long branch attractions (LBA#1 and LBA#2) based on a tree calculation with the full set of randomly sampled IMS-SC2 sequences. Subsequent investigation of the sequences causing LBAs revealed one sequence annotated as Alpha, but missing the characteristic deletions del69/70 and del144 in the spike protein, also showing two non-synonymous substitutions S:W152R and S:A1078S which are not commonly encountered in this VOC. Therefore, this sequence was excluded from the dataset (see above for details). In the absence of additional Gamma sequences, the single Gamma sequence caused another LBA, branching together with six other Alpha and one B.1.1.524 sequences (LBA#2). We therefore removed these nine sequences, leading to two long branch attractions (LBA#1 and LBA#2) from the tree visualization shown in Fig. 2B, via InkScape (<https://inkscape.org/>). LBA#1 includes: IMSSC2-206-2021-00149 (Gamma), IMSSC2-91-2021-00084 (Alpha), IMSSC2-123-2021-00119 (Alpha), IMSSC2-501-2021-00150 (Alpha), IMSSC2-123-2021-00120 (Alpha), IMSSC2-304-2021-0008 (Alpha), IMSSC2-206-2021-00092 (Alpha), IMSSC2-63-2021-00009 (B.1.1.524), and LBA#2 includes: IMSSC2-100-2021-00118 (Alpha). **Supplementary Figure S2** shows the full tree, including these LBAs.

Data Visualization: The Sankey plot was produced based on counts of the full GISAID (n=330,357) and IMS-SC2 (n=3623) data sets in the selected time period using a custom script (<https://github.com/hoelzer/sankey>) inspired by the Pavian package [9]. As standard Venn and Euler diagrams are an inadequate solution for quantitative visualization of multiple (n > 4) set intersections, we used the UpSetR package, a scalable alternative for visualizing intersecting mutation of concern (MOC) sets and their properties [10]. As input for UpSetR, we selected all IMS-SC2 sequences from the random data set (n=3282) that harbor a specific MOC at the selected site in the Spike. All figures were finalized with InkScape (<https://inkscape.org/>) for publication.

Genome-based incidence estimation and case ascertainment: We used the newly established genome-based incidence estimation pipeline GlnPipe [11] to predict the number of SARS-CoV-2 infections using an IMS-SC2 sequence set [3,282 sequences, provided in ENA (project accession number: PRJEB50616)], as well as all German sequences covering the time frame of the IMS-SC2 data (starting February 2021), available in GISAID (226,316 sequences) using default settings [PID: <https://doi.org/10.5281/zenodo.5519610>]. The timeframe 2021-02-01 to 2021-10-15 was used for visualization and relative case detection calculations. GlnPipe provides estimates

of the incidence correlate Φ , which are proportional to the ‘true’ incidences. For visualization purposes, we scaled the GInPipe’s estimate using a constant factor, such that $\Phi(t) \geq \text{incidences}(t)$ in **Figures 3 and 4**. For the Delta sub-analysis in **Figure 4**, we included all sequences assigned to lineages B.1.617.2 and AY (132,610 sequences for GISAID and 1,497 sequences for the Delta IMS).

Relative case detection analysis was performed as described in GInPipe [11]. Weekly test statistics and new reported cases were obtained from the RKI [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Testzahl.html, https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Fallzahlen_Gesamtuebersicht.html]. The daily positive test proportion was obtained by dividing the number of reported cases by the number of tests performed.

Vaccine efficacy and effect of vaccination campaign on Delta cases

Daily vaccinations: Daily data on vaccination timeline was derived from the German Ministry of Health (<https://impfdashboard.de/daten>) and is depicted in **Fig.4B, main manuscript** (blue round dots and dotted lines). We applied a linear smoothing filter with a window size of seven days to derive a smoothed line (solid red line in **Fig.4B, main manuscript**).

Vaccine efficacy against Delta infections: The vast majority of vaccinations in Germany in 2021 were BNT162b2 (<https://impfdashboard.de/daten>). For this reason, we used published data on BNT162b2 as a proxy to estimate vaccine efficacy against Delta. Data on time-dependent vaccine efficacy after BNT162b2 vaccination was derived from the following sources: Chemaitelly et al. [12] report time-dependent data after the first- and second shot of BNT162b2 (**Suppl. Fig. S4**, black triangles), whereas Tartof et al. [13] and Eyre et al. [14] report waning vaccine efficacy after the second dose of BNT162b2 (magenta diamonds and red circles in **Suppl. Fig. S4**).

Derivation of a continuous-time model of vaccine efficacy against Delta infections: We then set-up a semi-mechanistic model to estimate a continuous function that reflects the time-course of vaccine efficacy after the first- and second dose of BNT162b2. To this end, we minimized the sum of squared deviations between the data deduced- and model-predicted vaccine efficacy to fit model parameters $IC_{50}, t_{1/2}, F_{1stdose}$, i.e. the 50% inhibitory antibody concentration, their half-life and the fraction of antibody levels induced by the first does only:

$$\min_{IC_{50}, t_{1/2}, F_{1stdose}} \sum (VE_{\text{data}}(t_i) - VE_{\text{estim}}(t_i))^2,$$

Where we used a classical Emax equation for estimating vaccine efficacy

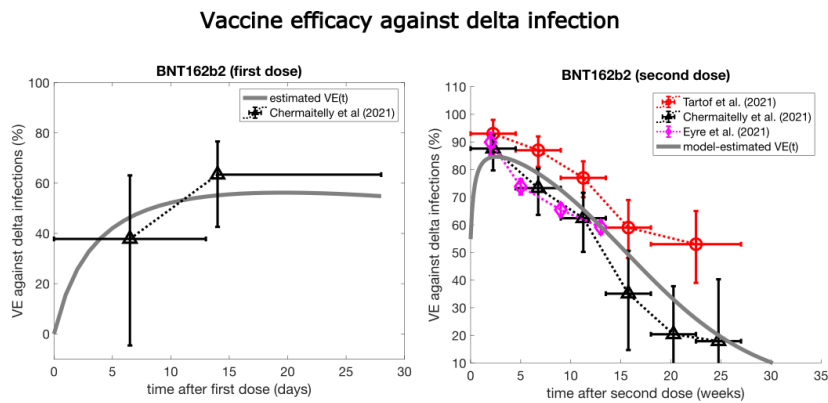
$$VE_{\text{estim}}(t_i) = \frac{X(t_i)}{X(t_i) + IC_{50}},$$

and a classical one compartment pharmacokinetic model to estimate a correlate $X(t_i)$ of neutralizing antibody levels:

$$X(t_i) = F_{(\cdot)} \frac{k_a}{k_a - k_e} (e^{-k_e \cdot t_i} - e^{-k_a \cdot t_i}) + X(t_0) \cdot e^{-k_e \cdot t_i},$$

where we set $F_{(2nddose)} = 1$ and estimated $F_{1stdose}$ as outlined above. The absorption rate was fixed to $k_a = 0.1$ (day⁻¹) and $k_e = \frac{\ln(2)}{t_{1/2}}$ (day⁻¹) following established pharmacokinetic principles. In the equation above, $X(t_0)$ denotes the concentrations of the antibody correlate before the second dose. The later was estimated by first solving the equation for the first dose with $X(t_0) = 0$ and $t = 28$ days (the typical time between first and second dose) and setting $X_{2nd\ dose}(t_0) = X_{1st\ dose}(t = 28)$.

Using this modelling, we deduced the following parameters for our semi-mechanistic model: $IC_{50} = 0.14$, $t_{1/2} = 32.3$ (days) and $F_{1stdose} = 0.27$. The resulting time-dependent vaccine efficacy estimate is shown in **Suppl. Fig. S4** below (solid grey line) and serves as a reasonable input for estimating the time-dependent vaccine efficacy at any arbitrary time instance after vaccination.



Supplementary Figure S4: Onset of vaccine efficacy against Delta after the first- (left panel) and second shot of BNT162b2 (right panel). The graphics show published data [12-14] on the onset- and waning of vaccine efficacy against infections (black triangles, magenta diamonds and red circles), as well as model predicted vaccine efficacy (solid grey line).

Expected vaccination effect on the population level: Lastly, we combined the data on the vaccination dates **Fig.4B (main manuscript)** with the model-predicted vaccine efficacy after BNT162b2 vaccination **Supplementary Figure S4** (grey line) to estimate the expected reduction of SARS-CoV-2 cases in the German population T days after the start of the vaccination campaign (on Dec 27th 2020 the first person received their second BNT162b2 shot). Note that vaccine

efficacy is defined as the (relative) reduction of infections, i.e. $VE = \frac{P(\text{inf}|\emptyset) - P(\text{inf}|\text{vacc.})}{P(\text{inf}|\emptyset)}$, with $P(\text{inf}|\emptyset)$, $P(\text{inf}|\text{vacc.})$ Denoting the probability (or incidence) of infections in a placebo vs. a vaccination cohort. Hence, the expected reduction of infections is given by

$$\mathbb{E}_{\text{pop}}[VE(t)] = \sum \pi(\tau_j) \cdot VE_{\text{estim}}(\tau_j),$$

where $\pi(\tau_j)$ denotes the fraction of the population that received their second shot $\tau_j = 0, \dots, t$ days ago and $VE_{\text{estim}}(\tau_j)$ denotes the corresponding, model-predicted vaccine efficacy in individuals that received their second BNT162b2 shot τ_j days ago. The above equation was solved recursively and the expected reduction of Delta cases resulting from the vaccination campaign is depicted in [Fig. 4C](#) (main manuscript).

SIR model of the SARS-CoV-2 delta pandemic: Next, we wanted to predict the hypothetical trajectory of the Delta pandemic in case the vaccination campaign had not been rolled out in Germany. To this end, we first wanted to estimate the time-dependent force of infection,

$$k_{\text{vacc}}(t) = k_{\emptyset}(t) \cdot (1 - \mathbb{E}_{\text{pop}}[VE(t)])$$

that gave rise to the reported cases in Germany. In the equation above, $k_{\emptyset}(t)$ denotes the time-dependent ‘force of infection’ if no vaccine campaign had been rolled out, whereas $(1 - \mathbb{E}_{\text{pop}}[VE(t)])$ denotes the (time-dependent) reduction of the ‘force of infection’ that is solely contributed to the vaccination campaign. To do this, we set up a simple SIR model with:

$$\frac{dS}{dt} = -S(t) \cdot I(t) \cdot k_{\text{vacc}}(t) + \delta_2 \cdot R(t)$$

$$\frac{dI}{dt} = S(t) \cdot I(t) \cdot k_{\text{vacc}}(t) - \delta_1 \cdot I(t)$$

$$\frac{dR}{dt} = \delta_1 \cdot I(t) - \delta_2 \cdot R(t)$$

$$\frac{dI^+}{dt} = S(t) \cdot I(t) \cdot k_{\text{vacc}}(t),$$

where S, I and R denotes the expected number of susceptible, infected (and infectious) and recovered individuals respectively. New infections occur with rate $S(t) \cdot I(t) \cdot k_{\text{vacc}}(t)$, infected individuals become non-infectious with rate $\delta_1 \cdot I(t)$, entering the recovered compartment. Recovered individuals are insusceptible to Delta infections until they become susceptible again with rate $\delta_2 \cdot R(t)$ because of waning immunity. The variable I^+ denotes the number of infections that occur (\propto incidence of SARS-CoV-2).

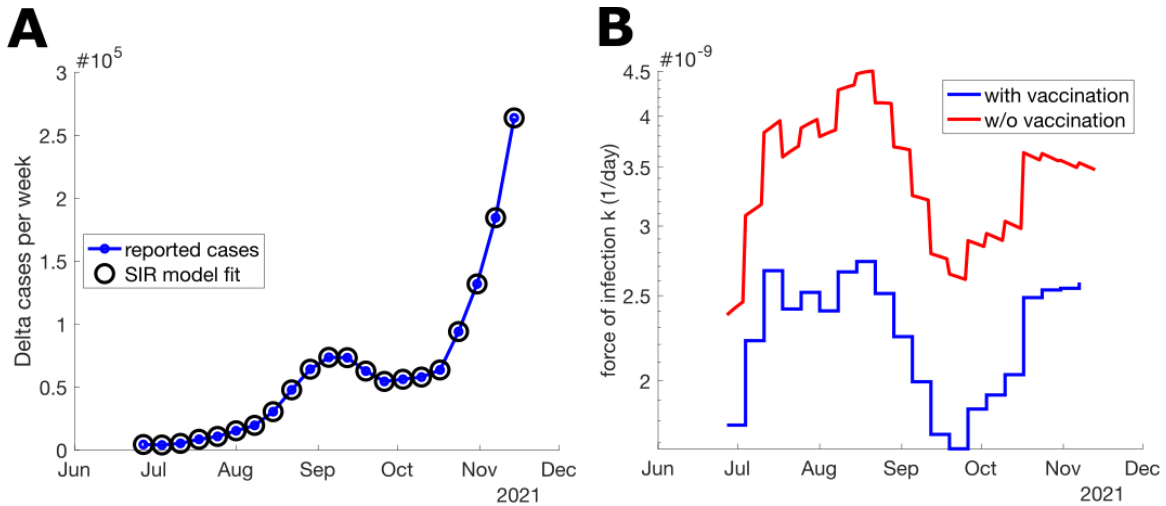
For the ODE system above, we used the following initial conditions $S(t_0) = 88 \cdot 10^6$ (the German population) and $I(t_0) = R(t_0) = I^+(t_0) = 0$. We set $\delta_1 = 1/6$ (day⁻¹) [15] and $\delta_2 = 1/90$ (day⁻¹), based on [Supplementary Figure S4](#), i.e. to the inverse of the mean duration of the infectious period and inverse of the average duration of immunity after infection.

We then estimated the time-varying ‘force of infection’ $k_{vacc}(t)$ by fitting the number of infections $I^+(t_i)$ to the reported cases during the onset of the delta wave. During estimation, we assumed that the ‘force of infection’ is piece-wise constant, i.e. $k_{vacc}(t) = k_{vacc}(t_i)$ if $t \in [t_i, t_{i+1})$.

$$\min_{k_{vacc}(t_0), \dots, k_{vacc}(t_n)} \sum \left(I_{\text{data}}^+(t_i) - I_{\text{estim}}^+(t_i, k_{vacc}(t_0), \dots, k_{vacc}(t_i)) \right)^2,$$

where $I_{\text{data}}^+(t_i)$ denotes the reported number of cases for week t_i and $I_{\text{estim}}^+(t_i, k_{vacc}(t_0), \dots, k_{vacc}(t_i))$ denotes the estimated number of new infections from the SIR model (above) with step-wise constant ‘force of infection’ parameters $k_{vacc}(t_0), \dots, k_{vacc}(t_i)$.

The resulting model fit to the reported cases is shown in [Supplementary Figure S5A](#) (filled blue dots = reported cases, empty circles = corresponding model predictions), indicating an excellent approximation of the pandemic dynamics using the fitted SIR model. The estimated, time-dependent ‘force of infection’ parameters $k_{vacc}(t_i)$ are depicted in [Supplementary Figure S5B](#) (blue line).



Supplementary Figure S5: A. Weekly reported SARS-CoV-2 cases in Germany from 27th June to 14th November 2021 (blue dots) and model predicted number of cases (empty circles). B. Blue line: fitted, time-dependent ‘force of infection’ $k_{vacc}(t_i)$. Red line = ‘force of infection’ without the effect of vaccination $k_{\emptyset}(t) = k_{vacc}(t)/(1 - \mathbb{E}_{pop}[VE(t)])$.

Simulated Delta incidences in Germany 2021 without vaccination: After having derived the time-dependent ‘force of infection’ parameter $k_{vacc}(t_i)$, it is possible to calculate out the effect of vaccination on the pandemic and to subsequently simulate the pandemic trajectory if the vaccination campaign never happened. For this, we first compute the ‘force of infection’ in the absence of vaccination (red line in [Supplementary Figure S5B](#)),

$$k_{\emptyset}(t) = \frac{k_{vacc}(t)}{(1 - \mathbb{E}_{pop}[VE(t)])}$$

where the expected vaccination effect $\mathbb{E}_{pop}[VE(t)]$ is shown in [Fig. 4C](#) (main manuscript). We then solved the SIR model as before with initial conditions $S(t_0) = 88 \cdot 10^6$, $I(t_0) = R(t_0) = I^+(t_0) = 0$ and parameters $\delta_1 = 1/6$ (day⁻¹) and $\delta_2 = 1/90$ (day⁻¹) as before, but by replacing $k_{vacc}(t)$ with $k_{\emptyset}(t)$. The resulting incidence estimates, as well as the total number of cases averted are depicted in [Figure 4D-E](#).

References

1. Hufsky F, Lamkiewicz K, Almeida A, et al. Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. *Brief Bioinform* **2021**; 22(2): 642-63.
2. Brandt C, Krautwurst S, Spott R, et al. poreCov-An Easy to Use, Fast, and Robust Workflow for SARS-CoV-2 Genome Reconstruction via Nanopore Sequencing. *Front Genet* **2021**; 12: 711437.
3. Hein S, Herrlein ML, Mhedhbi I, et al. Analysis of BNT162b2- and CVnCoV-elicited sera and of convalescent sera toward SARS-CoV-2 viruses. *Allergy* **2021**.
4. O'Toole A, Scher E, Underwood A, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* **2021**; 7(2): veab064.
5. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**; 30(4): 772-80.
6. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **2020**; 37(5): 1530-4.
7. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* **2018**; 35(2): 518-22.
8. Moore RM, Harrison AO, McAllister SM, Polson SW, Wommack KE. Iroki: automatic customization and visualization of phylogenetic trees. *PeerJ* **2020**; 8: e8584.
9. Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* **2020**; 36(4): 1303-4.
10. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **2017**; 33(18): 2938-40.

11. Smith MR, Trofimova M, Weber A, Duport Y, Kuhnert D, von Kleist M. Rapid incidence estimation from SARS-CoV-2 genomes reveals decreased case detection in Europe during summer 2020. *Nat Commun* **2021**; 12(1): 6009.
12. Chemaitelly H, Tang P, Hasan MR, et al. Waning of BNT162b2 Vaccine Protection against SARS-CoV-2 Infection in Qatar. *N Engl J Med* **2021**; 385(24): e83.
13. Tartof SY, Slezak JM, Fischer H, et al. 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* **2021**; 398(10309): 1407-16.
14. Eyre DW, Taylor D, Purver M, et al. The impact of SARS-CoV-2 vaccination on Alpha and Delta variant transmission. *medRxiv* **2021**: 2021.09.28.21264260.
15. van der Toorn W, Oh DY, Bourquain D, et al. An intra-host SARS-CoV-2 dynamics model to assess testing and quarantine strategies for incoming travelers, contact management, and de-isolation. *Patterns (N Y)* **2021**; 2(6): 100262.

Supplementary Table S1: Variants of Concern (VOCs) detected in the IMS-SC2

VOC Name (WHO)	Pangolin Designation	Spike AA substitutions or deletions ¹	Sub-lineages ²	Declared VOC by WHO	Dominant Variant in Germany ³	First Detected in IMS-SC2
Alpha	B.1.1.7	del69/70, del144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Q.1-Q.8	18-Dec-2020	week 09/2021 (March 2021) ⁴	7-Dec-2020
Beta	B.1.351	L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V	B.1.351.1- B.1.351.3, B.1.351.5	18-Dec-2020	-	9-Dec-2020
Gamma	P.1	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F	P.1.1-P.1.17	11-Jan-2021	-	9-Jul-2021
Delta	B.1.617.2	T19R, del157/158, L452R, T478K, D614G, P681R, D950N	AY.1-AY.133	11-May-2021	week 25/2021 (June 2021) ⁵	15-Apr-2021
Omikron	B.1.1.529	A67V, del69/70, T95I, G142D/del143/145, del211/L212I, R214REPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	BA.1-BA.3	26-Nov-2021	week 01/2022 (January 2022) ⁶	13-Dec-2021

¹ Parental VOC lineage

² Including all descendent lineages

³ Time at which variant accounted for >50% of SARS-CoV-2 circulating in Germany per RKI, based on representative genotyping PCR or sequencing results

⁴ https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/DESH/Bericht_VOC_2021-04-21.pdf?__blob=publicationFile

⁵ https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Wochenbericht/Wochenbericht_2021-07-29.pdf?__blob=publicationFile

⁶ https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/Wochenbericht/Wochenbericht_2022-01-13.pdf?__blob=publicationFile

**Supplementary Table S2:
Age Group Distribution of IMS-SC2 Random Sample**

Age Group	Percentage
0-4	2,7%
05-14	11,8%
15-34	28,9%
35-59	34,1%
60-79	14,7%
80+	7,2%
unknown	0,5%

Supplementary Table S3.

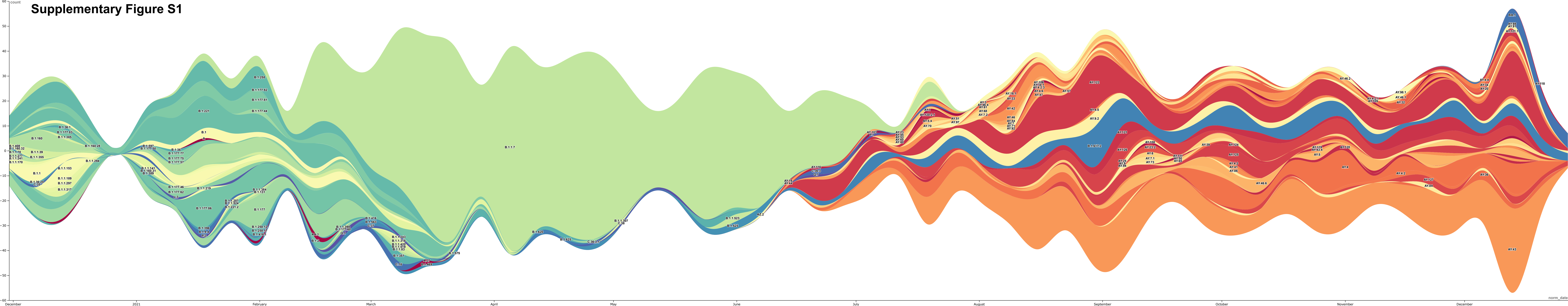
Selected variations within the Spike protein of SARS-CoV-2 (compared to Wuhan-1 strain) and known associations related to antibody and immune escape. Positions within the receptor binding domain (RBD) of the protein are highlighted in **bold**.

Position	Reference allele	Alternate alleles	Associations	References
18	L	F, V	Escape from different monoclonal antibodies	[1]
69-70		Deletion	Increased infectivity by conformational change	[2–4]
144-145		Deletion	Escape from different monoclonal and polyclonal antibodies	[1,5,6]
240-244		Deletion	Escape from monoclonal antibody 4A8; described as deletion of residues 241-243 (https://outbreak.info/compare-lineages) but in our data we see a deletion from residues 240-244	[5]
417	K	N, T	Escape from different monoclonal antibodies	[7–9]
477	S	I, N, R	Escape from different monoclonal and polyclonal antibodies; increased binding affinity to ACE2	[10–13]
478	T	I, K	Predicted to affect the Spike/ACE2 interaction	[14]
452	L	M, Q, R	Reduced antibody neutralization	[14–16]
484	E	A, K, Q	Escape from different monoclonal antibodies; increased binding affinity to ACE2; mediates resistance to therapeutic monoclonal antibodies	[12,17–22]
490	F	L, S, Y	Resistance to monoclonal antibodies and convalescent sera	[2,15,19,23]
501	N	Y, T	Potential association with host specificity; escape from different monoclonal and polyclonal antibodies; increased binding affinity to ACE2	[7,19,24–26]
655	H	P, Y	Adaption to cats and hamsters; antibody escape; identified in the sequence of super spreaders	[27–29]
681	P	H, R, Y	Altered immune and antibody (class 3) recognition	[6]

Suppl. Table S3 References

1. McCallum M, De Marco A, Lempp FA, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **2021**; 184.
2. Harvey WT, Carabelli AM, Jackson B, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 2021; 19.
3. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **2021**; 592.
4. Meng B, Kemp SA, Papa G, et al. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep* **2021**; 35.
5. McCarthy KR, Rennell LJ, Nambulli S, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* (80-) **2021**; 371.
6. Haynes WA, Kamath K, Lucas C, Shon J, Iwasaki A. Impact of B.1.1.7 variant mutations on antibody recognition of linear SARS-CoV-2 epitopes. *medRxiv* **2021**;
7. Tian F, Tong B, Sun L, et al. N501y mutation of spike protein in sars-cov-2 strengthens its binding to receptor ace2. *Elife* **2021**; 10.
8. Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Reports Med* **2021**; 2.
9. Thomson EC, Rosen LE, Shepherd JG, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **2021**; 184.
10. Chen J, Wang R, Wang M, Wei GW. Mutations Strengthened SARS-CoV-2 Infectivity. *J Mol Biol* **2020**; 432.
11. Gaebler C, Wang Z, Lorenzi JCC, et al. Evolution of antibody immunity to SARS-CoV-2. *Nature* **2021**; 591.
12. Zahradník J, Marciano S, Shemesh M, et al. SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. *Nat Microbiol* **2021**; 6.
13. Liu Z, VanBlargan LA, Rothlauf PW, et al. Landscape Analysis of Escape Variants Identifies SARS-CoV-2 Spike Mutations that Attenuate Monoclonal and Serum Antibody Neutralization. *SSRN Electron J* **2020**;
14. Di Giacomo S, Mercatelli D, Rakhimov A, Giorgi FM. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. *J Med Virol* **2021**; 93.
15. Li Q, Wu J, Nie J, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **2020**; 182.
16. Deng X, Garcia-Knight MA, Khalid MM, et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell* **2021**; 184.
17. Barton MI, Macgowan S, Kutuzov M, Dushek O, Barton GJ, Anton Van Der Merwe P. Effects of common mutations in the sars-cov-2 spike rbd and its ligand the human ace2 receptor on binding affinity and kinetics. *Elife* **2021**; 10.
18. Jangra S, Ye C, Rathnasinghe R, et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *The Lancet Microbe.* 2021; 2.
19. Wang Z, Schmidt F, Weisblum Y, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **2021**; 592.
20. Wang P, Nair MS, Liu L, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **2021**; 593.
21. Andreano E, Piccini G, Licastro D, et al. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. *Proc Natl Acad Sci U S A* **2021**; 118.
22. Greaney AJ, Loes AN, Crawford KHD, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **2021**; 29.
23. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies 1 by SARS-CoV-2 spike protein variants. *Elife* **2020**; 9.
24. Richard M, Kok A, de Meulder D, et al. SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nat Commun* **2020**; 11.
25. Gu H, Chen Q, Yang G, et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* (80-) **2020**; 369.
26. Teruel N, Mailhot O, Najmanovich RJ. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. *PLoS Comput Biol* **2021**; 17.
27. Dieterle ME, Haslwanter D, Bortz RH, et al. A Replication-Competent Vesicular Stomatitis Virus for Studies of SARS-CoV-2 Spike-Mediated Cell Entry and Its Inhibition. *Cell Host Microbe* **2020**; 28.
28. Braun KM, Moreno GK, Halfmann PJ, et al. Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck. *PLoS Pathog* **2021**; 17.
29. Yang X, Dong N, Chan EWC, Chen S. Identification of super-transmitters of SARS-CoV-2. *medRxiv* **2020**;

Supplementary Figure S1

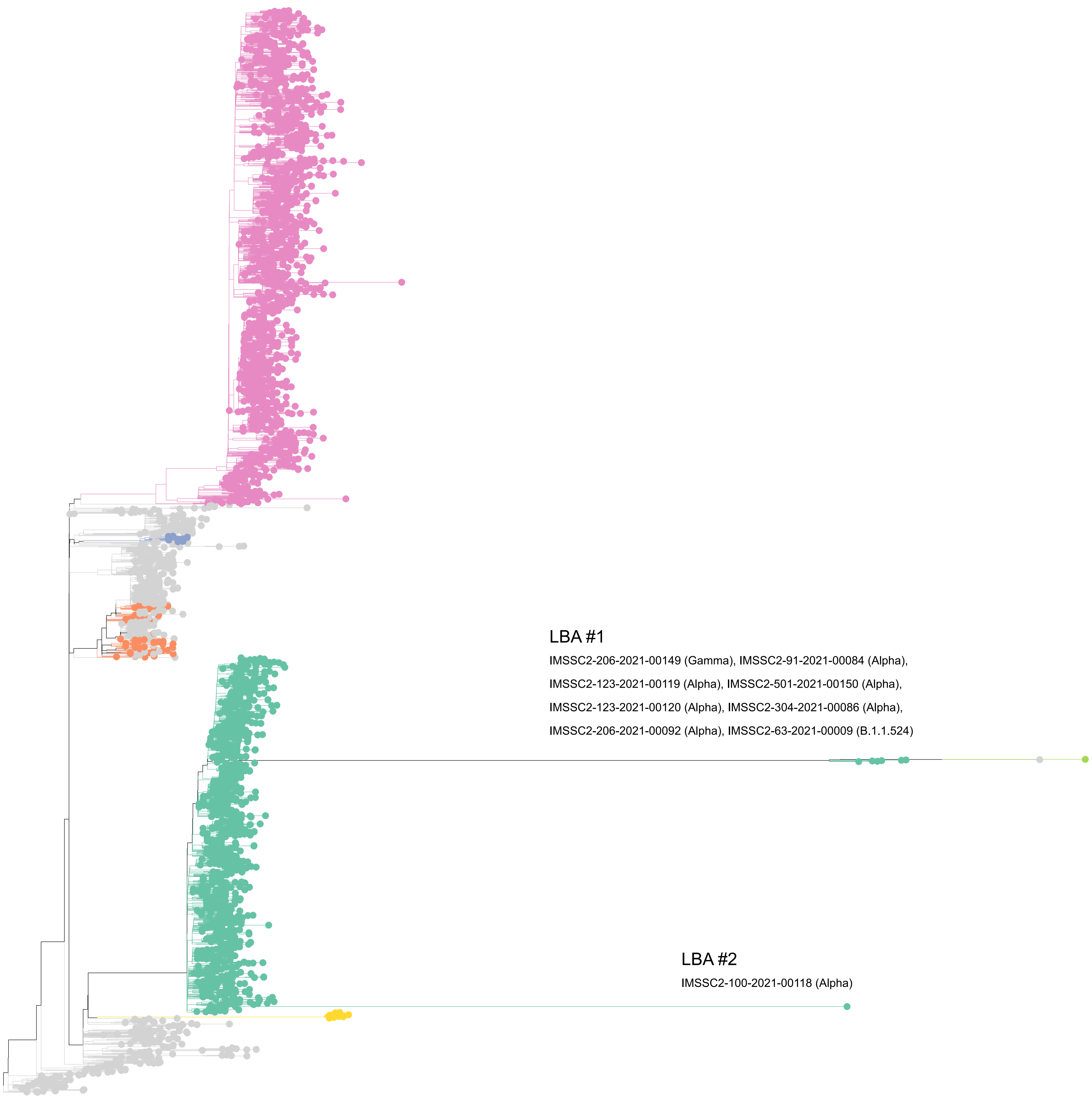


Suppl. Figure S1

Lineages identified in the representative sampling subset of IMS-SC2 genome sequences.

To visualize the dynamics in the virus population over time, virus lineages were determined with pangolin based on the randomly sampled genome sequences (n=3282, see *Methods*). Lineage frequencies were aggregated based on the date of sampling relative to calendar weeks. Missing values have been interpolated. Visualization was performed using <https://rawgraphs.io>. This figure corresponds to Fig. 2A in the main manuscript, but shows all Pangolin lineages separately, rather than focusing on VOC- lineages and -sublineages selected variants.

Supplementary Figure S2



LBA #1

IMSSC2-206-2021-00149 (Gamma), IMSSC2-91-2021-00084 (Alpha),
IMSSC2-123-2021-00119 (Alpha), IMSSC2-501-2021-00150 (Alpha),
IMSSC2-123-2021-00120 (Alpha), IMSSC2-304-2021-00086 (Alpha),
IMSSC2-206-2021-00092 (Alpha), IMSSC2-63-2021-00009 (B.1.1.524)

LBA #2

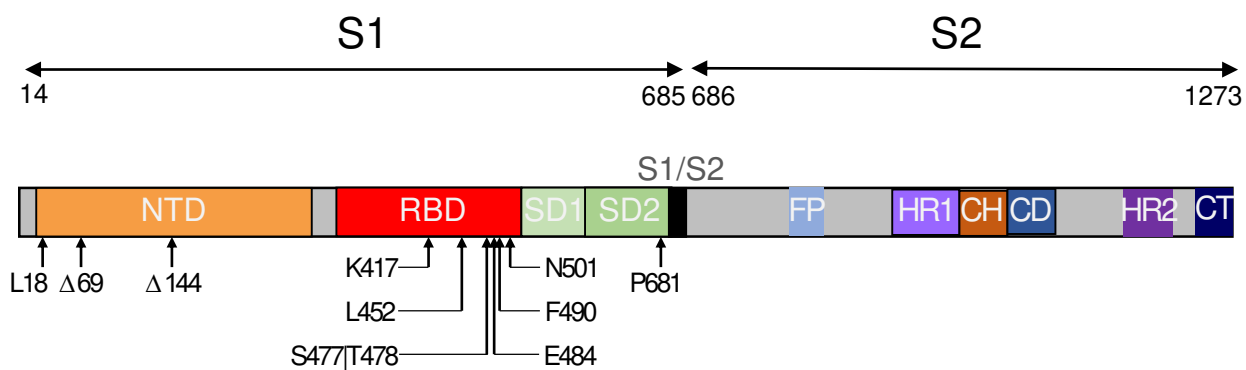
IMSSC2-100-2021-00118 (Alpha)

Suppl. Figure S2

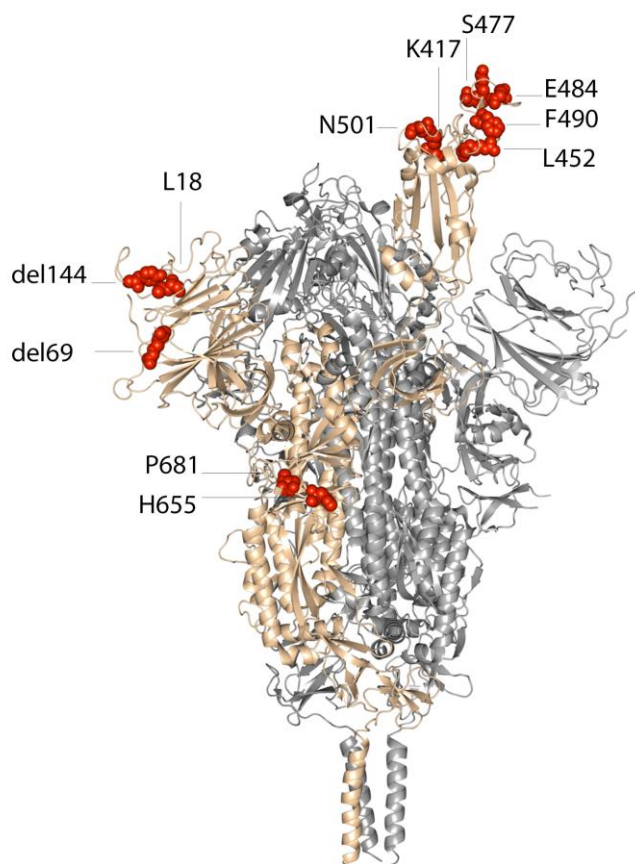
Phylogenetic tree highlighting VOC clades. Sequencing data presented here is based on all randomly selected SARS-CoV-2 positive specimens from the IMS-SC2 laboratory network (n=3282). Lineage B.1.177 is also shown as an early variant that emerged in Europe in early summer 2020 as well as three A.27 samples. This phylogenetic tree corresponds to the phylogenetic tree shown in Fig. 2B, featuring two LBAs as described in *Supplementary Methods*.

Supplementary Figure S3

A



B



Suppl. Figure S3

Graphic depiction of selected variations (mutations of concern).

A. Schematic overview of the SARS-CoV-2 spike gene with arrows indicating positions of selected amino acid substitutions and deletions (i.e. MOCs, see also Suppl. Table S3). NTD, N-terminal domain; RBD, receptor-binding domain; SD1, subdomain 1; SD2, subdomain 2; S1/S2, S1/S2 protease cleavage site; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; CT, cytoplasmic tail.

B. Positions of selected amino acid substitutions and deletions on the SARS-CoV-2 Spike.