Supporting materials for

Mapping SARS-CoV-2 antigenic relationships and serological responses

Samuel H. Wilks[1,*], Barbara Mühlemann[2,3,*], Xiaoying Shen[4,5,*], Sina Türeli[1], Eric B. LeGresley[1], Antonia Netzl[1], Miguela A. Caniza[6], Jesus N. Chacaltana-Huarcaya[7], Xiaoju Daniell[4], Michael B. Datto[8], Thomas N. Denny[5], Christian Drosten[2,3], Ron A. M. Fouchier[9], Patricia J. Garcia[10], Peter J. Halfmann[11], Agatha Jassem[12], Terry C. Jones[1,2,3], Yoshihiro Kawaoka[11,13,14], Florian Krammer[15,16], Charlene McDanal[4], Rolando Pajon[17], Viviana Simon[15,16,18,19], Melissa Stockwell[20], Haili Tang[4], Harm van Bakel[21], Richard Webby[22], David C. Montefiori[4,5,#], Derek J. Smith[1,#]

Affiliations:
1: Center for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK
2: Institute of Virology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, 10117 Berlin, Germany
3: German Centre for Infection Research (DZIF), partner site Charité, 10117 Berlin, Germany
4: Department of Surgery, Duke University School of Medicine, Durham, NC, USA
5: Duke Human Vaccine Institute, Duke University School of Medicine, Durham, NC, USA
6: Department of Global Pediatric Medicine, Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, TN, USA
7: Hospital Nacional Daniel A Carrión, Callao, Bellavista, Peru
8: Department of Pathology, Duke University School of Medicine, Durham, NC, USA
9: Erasmus Medical Center, Rotterdam, Netherlands
10: School of Public Health, Universidad Peruana Cayetano Heredia, Lima, Peru
11: Department of Pathobiological Science, School of Veterinary Medicine University of Wisconsin-Madison, Madison, WI, USA
12: BC Centre for Disease Control, Vancouver, British Columbia, Canada
13: Division of Virology, Institute of Medical Science, University of Tokyo, Tokyo, Japan
14: The Research Center for Global Viral Diseases, National Center for Global Health and Medicine Research Institute, Tokyo, Japan
15: Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
16: Department of Pathology, Cellular and Molecular Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
17: Moderna, Inc., Cambridge, MA, USA
18: Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
19: The Global Health and Emerging Pathogen Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
20: Division of Child and Adolescent Health, Department of Pediatrics, Columbia University Vagelos College of Physicians and Surgeons, and Department of Population and Family Health, Mailman School of Public Health, New York, NY, USA
21: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
22: Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, TN, USA

* Contributed equally
# Corresponding

This PDF file includes

Materials and Methods

Specimens and study cohorts

Serum samples from individuals that received the Moderna mRNA-1273 vaccine or were infected with the prototype D614G or other SARS-CoV-2 variants are used in this study (n=126). All samples, including those from vaccine recipients were collected in clinical studies with information listed in Table S1. Convalescent sera are assumed but not confirmed first infections. Institutional review board (IRB) approvals were obtained for each study site. All participants were given and signed the informed consent form. The use of the specimens for the research conducted in this manuscript is covered under IRB protocols Pro00093087 and Pro00105358.

Pseudotype virus assay

Neutralization was measured with lentiviral particles pseudotyped with SARS-CoV-2 spike and containing a firefly luciferase (Luc) reporter gene for quantitative measurements of infection by relative luminescence units (RLU). The assay was performed in 293T/ACE2-MF provided by Drs. Michael Farzan and Huihui Mu. Pseudoviruses were prepared, titrated, and used for measurements of neutralizing antibodies essentially as described previously (*45*). Briefly, an expression plasmid encoding codon-optimized full-length spike of the Wuhan-1 ancestral sequence (VRC7480) was provided by Drs. Barney Graham and Kizzmekia Corbett at the Vaccine Research Center, National Institutes of Health (USA). Mutations were introduced into VRC7480 either by site-directed mutagenesis using the QuikChange Lightning Site-Directed Mutagenesis Kit from Agilent Technologies (Catalog # 210518), or were created by a spike gene synthesized by GenScript using the spike sequence in VRC7480 as template. All mutations were confirmed by full-length spike gene sequencing by Sanger Sequencing, using Sequencher and SnapGene for sequence analyses. Pseudovirions were produced in HEK293T/17 cells (ATCC cat. no. CRL-11268) by transfection using Fugene 6 (Promega Cat#E2692) and a combination of spike plasmid, lentiviral backbone plasmid (pCMV ΔR8.2) and firefly Luc reporter gene plasmid (pHR' CMV Luc)36 in a 1:17:17 ratio in Opti-MEM (Life Technologies). Transfection mixtures were added to pre-seeded HEK 293T/17 cells in T-75 flasks containing 12 ml of growth medium and incubated for 16-20 h at 37°C. Medium was removed and 15 ml of fresh growth medium added. Pseudovirus-containing culture medium was collected after an additional 2 days of incubation and clarified of cells by low-speed centrifugation and 0.45 μm micron filtration. TCID50 assays were performed as described previously (*45*). For measurements of neutralization, a pre-titrated dose of virus was incubated with 8 serial 5-fold dilutions of serum samples (1:10 or 1:20 starting dilution) in duplicate in a total volume of 150 μl for 1 h at 37°C in 96-well flat-bottom culture plates. 293T/ACE2-MF cells were detached from T75 culture flasks using TrypLE Select Enzyme solution, suspended in growth medium (100,000 cells/ml) and immediately added to all wells (10,000 cells in 100 μL of growth medium per well). One set of 8 wells received cells and virus (virus control) and another set of 8 wells received cells only (background control). After 66-72 h of incubation, medium was removed by gentle aspiration and 30 μl of Promega 1X lysis buffer was added to all wells. After a 10-minute incubation at room temperature, 100 μl of Bright-Glo luciferase reagent was added to all wells. After 1-2 minutes, 110 μl of the cell lysate was transferred to a black/white plate. Luminescence was measured using a GloMax Navigator luminometer (Promega). Serum samples were heat-inactivated for 30 minutes at 56ºC prior to assay. Neutralization titers are the inhibitory dilution (ID) of serum samples at which RLUs were reduced by either 50% (ID50) or 80% (ID80) compared to virus control wells after subtraction of background RLUs. The spike mutations of all variants used in this study are shown in Table S2.

Note that one titration, a spuriously high value of 3429 for D614G post-vaccine sera against B.1.1.526, was discounted as a likely titration error in all our analyses.

Titer analyses

Calculation of Geometric Mean Titers (GMT) and fold-changes for non-detectable titers

Calculation of the mean titers becomes non-trivial when experiments contain non-detectable titers or, in data science terminology, censored data. Although computation of means of data with censored values has been around

for over 60 years (*46–48*), it does not seem to be commonly employed when analyzing data from neutralization assays. Correct estimation of the GMT of titers becomes especially important if one wants to quantify the amount of escape by a variant like Omicron with many low and non-detectable titers. Setting non-detectable titers to half the value of an assay's limit of detection is a common method employed in virology, however, it may overestimate the GMTs for such variants. For completeness we give a brief overview of the mean estimation method. The main step in estimation of the titers is the construction of the correct likelihood function that describes the probability of observing both the non-detectable and detectable titers. In the dataset used for this study we deal with only left-censored data. We assume that $\log_2$ titer values of a particular serum group are modeled by a normal density function $f(x;\mu,\sigma)$ with unknown mean μ and standard deviation σ and we let y_i (i=1,...,M) denote detectable titers and $<T_i$ (i=1,..,M) denote the non-detectable titers. Then the log likelihood function is given by

$$LL(\mu,\sigma) = \sum_{i=1}^{N} log \int_{-\infty}^{T_i} f(x;\mu,\sigma) + \sum_{i=1}^{M} log\, f(y_i;\mu,\sigma)$$

Note, here the part for the non-detectable observations is simply making the statement that the probability of observing a non-detectable titer, given $f(x;\mu,\sigma)$, is simply the probability that $x < T_i$ under this distribution (up to a normalization factor). This likelihood function then can either be fed into a maximum likelihood estimation pipeline or, if priors are chosen for μ and σ, into a Bayesian posterior calculation pipeline from which the value of μ can be estimated in various ways. For the calculations in this paper, μ was estimated using the “optim” function in R (*49*), with method “L-BFGS-B”, while confidence intervals were calculated using a quadratic approximation based on numerical solutions of the hessian matrix. The package we use for mean titer computation can be found in (*50*). The following parameters were used for the computation; “method”: bayesian, “level”: 0.95, “dilution_stepsize”: 0, “prior_mean_mean”: 0, “prior_mean_sd”: 100, “prior_sd_shape”: 3, “prior_sd_scale”: 0.25.

When calculating the fold-changes in titers, the same approach to non-detectable titers was taken, with for example a change from 40 to <20 equating to a change of ≤-1 on the $\log_2$ scale, for which the likelihood function would be the integral of the normal density function from -∞ to -1.

Titers after accounting for estimated individual effects

Accounting for individual variation in serum reactivity

The data used in this paper contains sera collected from different individuals infected with the same variants. Even when these titers are grouped according to the infecting variant, one can see quite large variations in titer magnitudes whereas the qualitative appearance of the titer curve is similar (Fig. 1). Some of this is due to individual variation in the overall reactivity of sera collected from different individuals, for example related to the magnitude of the serological response. We model this behavior in the following way: Assume that t_{ij} is the $\log_2$ titer measurement of antigen i against serum j from the serum group J. Then

$$t_{ij} = s_{iJ} + r_j + \varepsilon_{ij} \quad (\mathrm{i} = 1, \ldots . \mathrm{n} \text{ and } \mathrm{j} = 1, \ldots . \mathrm{m})$$

where s_{ij} is the average $\log_2$ titer of antigen i against the serum group J, r_j is the serum reactivity bias of the serum j and ε_{ij} is the independently and normally distributed log titer noise. For each serum group, parameters s_{ij}, r_j and standard deviation of log titer noise are then found to maximize the sum of the likelihood of each measured log titer given the observed log titer t_{ij}. For the fitting procedure we used the R package “cmdstanr” (*51*) interface to Stan (*52*) optimization routines.

Note that antigenic cartography has its own method of dealing with serum reactivity bias. The formula for the target distance (section ‘Antigenic cartography’, below) indicates that for any serum, the antigen with the highest titer (usually the homologous antigen, but not always) is automatically adjusted to have 0 distance to that serum. This is

true regardless of the value of the maximum titer and this also takes care of the individual serum reactivity in a way that is compatible with antigenic map making.

Effect of outlier removal on geometric mean titer

We investigated the effect of outliers on the GMT by removing either titrations identified as outliers or entire serum samples containing outlying titrations and re-calculating the GMT of the reduced dataset. We defined outliers as measurements outside the interval

$$I = [q_{0.25} - 1.5x\ IQR;\ q_{0.75} + 1.5x\ IQR]$$

where $q_{0.25}$ corresponds to the first quartile, $q_{0.75}$ to the third quartile and IQR to the interquartile range. The effect of titration and sample removal on the GMT of different serum groups is shown in fig. S3. A total of 16 samples contained titrations that were identified as outliers. The majority of measurements fell within the defined interval, titrations against the B.1.617.2 variant were the most frequent above this interval. Only these titrations were extreme or frequent enough to have a substantial impact on the GMT in B.1.1.7, P.1, and D614G sera. In cases where these titrations originated from samples with generally high titers, removing the entire sample reduced GMTs to 0.5 - 0.7 of the full GMT (B.1.1.7, P.1, B.1.351 sera). Only in vaccinated sera there were titrations below the lower bound, resulting in higher GMTs when removing these titrations or samples (mRNA-1273: B.1.1.7+E484K, B.1.617.2(AY.3)+E484K).

Higher-than-homologous titers

Most serum groups had highest GMTs against variants homologous to the variant associated with their infection (Fig. 1, figs. S1-S3). However, individuals infected with B.1.526+E484K and C.37 had highest GMTs against P.1 and B.1.617.2+K417N, respectively, possibly indicating that those individuals might have been infected more than once. Individuals in the B.1.351 serum group showed higher titers against the P.1 variant than the homologous B.1.351 antigen. Rather than a particular immunological or antigenically related phenomenon, we find it more likely due to assay-related biases inflating titers for the particular P.1 variant, as it has been observed in other assays using lentiviral pseudotypes (*10*).

Calculating fold-change estimates for B.1.1.529

The B.1.1.529 variant was titrated against only a subset of post-vaccination, B.1.351, B.1.617.2, and B.1.526+E484K sera. Therefore, a direct comparison of the GMTs between variants shown in Fig. 1 may not be a reliable indicator of the extent of B.1.1.529 antigenic escape. We therefore also compared estimates of individually-paired titer fold-change of B.1.1.529 and the other variants relative to the homologous strain for each serum group, to account for any differences in individual response magnitude (fig. S5), and compared GMT differences between variants for only the subset of sera against which B.1.1.529 had been titrated (fig. S6).

Antigenic cartography

Antigenic cartography is a multidimensional scaling method (MDS) (*53*) for the reduction of dimensions in large titer datasets in order to not only visualize large amounts of titer data in a lower-dimensional space but to also draw conclusions about the antigenic evolution of pathogens by studying the positions occupied by these viruses on the antigenic map. First developed in (*14*, *54*) and applied to influenza A/H3N2 to quantify its antigenic evolution from 1968 to 2003, it has since been used to study the evolution of influenza A H1 (*55*), H3 (*14*), H5 (*56*), and dengue viruses (*21*) and has been a part of the global influenza vaccine strain selection strategy.

In order to construct an antigenic map, one starts with a titer table where columns are sera and rows are antigens. Denoting the values of such a table by t_{ij} (where i denotes the antigen index and j the sera index), titer values represent a degree of dissimilarity of the i^{th} antigen to the j^{th} sera. We turn this dissimilarity score into a similarity score by the following transformation:

$$D_{ij} = \max_i(\log_2(\ t_{ij}\ )) - \log_2(\ t_{ij}\ )$$

The D_{ij} values are called the target distances. Antigenic cartography strives to find either a two or three dimensional placement of antigens and sera such that the Euclidean distance between antigens and sera conform as well as possible to the target distances. This is achieved by an optimization of the antigen and serum coordinates that minimize the stress function

$$\sum_{i,j}(D_{ij} - d_{ij})^2$$

where d_{ij} represents the Euclidean distance between the i^{th} antigen and j^{th} serum points in the antigenic map. The minimization is achieved using the L-BFGS algorithm (*57*) to find the best cost-minimizing coordinates starting from a set of randomly determined initial conditions. Note unlike most common MDS problems, the similarity matrix is not square and might contain masked or censored values and therefore requires a slightly custom approach. In particular if t_{ij} is a non-detectable titer, then the stress is computed as

$$\varphi(D_{ij} - d_{ij} + s)(D_{ij} - d_{ij} + s)^2$$

where s is a step size factor which is 1 for discrete data (and their means) such as those obtained from Haemagglutinin Inhibition assays, or 0 when the output data is continuous and measured from titer curves, and

$$\varphi(x) = \frac{1}{1+10^{-x}}$$

is a damping function whose role is to give more freedom to non-detectable titers to be optimized for a value less than the limit of detection of the assay.

We used the Racmacs package to compute the antigenic maps presented in this manuscript (*58*). The map was constructed using 1000 optimisations, dilution step size of 0, with the minimum column basis parameter set to “none”.

Assessing model fit

We performed different analyses to ascertain that the two-dimensional antigenic map is a good representation of the titer data, and to ensure that the map is robust to noise from measurement errors.

First, to get an understanding of the measurement error of individual titrations, we analyzed a set of repeat titrations of 60 sera against 8 of the variants used in the assay, and found that the mean difference in $\log_2$ titers between repeated titrations was 0.209, with a standard deviation of 0.84, with measurement error being approximately normally distributed (fig. S7). Accounting for the systematic bias of titrations between repeats, and assuming a mean of 0 without it, resulted in a total standard deviation of 0.87 on the $\log_2$ scale. Since measurement error would be present for both the first and second repeat titration this would be consistent with noise per measurement with a standard deviation of $\sqrt{0.87^2/2} = 0.62$.

Subsequently, we investigated how well antigenic maps of different dimensions represent the measured patterns of serological reactivity. In our previous experience with other antigenically variable pathogens, the dimensionality of relationships can be lower (for example two dimensions in seasonal influenza and dengue virus (*14*, *21*)) or higher (for example three dimensions for A/H5N1 and swine influenza(*18*)), or may simply not be well represented in Euclidean space. Under cross-validation, a map of three dimensions represented the data best (RMSE = 1.2), but performed only marginally better compared to four (RMSE = 1.204), five (RMSE = 1.212) and two dimensions (RMSE = 1.273), with the one-dimensional map performing worst (RMSE = 1.766) (fig. S8A). The overall arrangement of the variants and sera in three dimensions is similar to two dimensions (fig. S8B-C).

Assessing goodness of fit using fitted and measured titers

Next, we assessed how well the two-dimensional map fits the measured titers. We first investigated whether the measured $\log_2$ titers correlated with the titers fitted in the antigenic map. Overall, the fitted and measured titers corresponded well (fig. S9). When comparing the mean differences of the fitted versus measured titers (mean: -0.01, standard deviation: 1.04, shown as a histogram in fig. S10; when assuming a mean of 0, the standard deviation is 1.03), they are comparable to the estimated size of measurement error (fig. S7). This suggests that the majority of the residual error observed in the antigenic map fit of the data is consistent with measurement error, but that not all variation in the titers is captured.
We investigated whether any variant and serum group combinations exhibited consistently larger differences between fitted and measured titers (fig. S11). We find some variants showing differences in fitted versus measured titers in some serum groups, such as the B.1.617.2 variant having larger measured than fitted titers against the B.1.617.2 sera. Such differences do not necessarily result from antigenic dissimilarity but may also be related to changes in avidity of antibody binding. (*40*). Some of these systematic residual errors in the map fit will also be reflective of the fact that not all patterns of cross-reactivity between the serum groups can be captured in the constraints of the two-dimensional Euclidean space of the map. Overall, we judge that the two-dimensional map is a reasonable representation of the broader scale underlying antigenic relationships among variants, providing a useful complement to raw titer analysis, and in some cases may compensate for systematic titer biases like that of P.1 that are otherwise spurious.

Assessing the robustness of variant and serum positions to measurement error and presence/absence of variants and serum groups

We performed three experiments to determine the uncertainty of the positions of variants and sera in the antigenic map. First, we estimated how titration measurement error and variability in antigen reactivity would affect conclusions about antigenic map positions (fig. S12). We performed a “smooth” bootstrap with normally distributed noise added to the titer measurements and / or to antigen reactivity. Noise added to the titer measurement had a standard deviation of 0.62, in keeping with the standard deviation of repeat titrations shown in fig. S7, and random noise for antigen reactivity had a standard deviation of 0.4, indicating that systematic biases in titers like that suspected for P.1 may also be present to some degree for other strains in the assay. We performed 1000 repeats of adding random noise, and summarized the variant positions by showing the area in which 68% (one standard deviation) of the positional variation of a serum or antigen is captured. In general, we found the antigen and serum positions are robust to both the random and systematic noise added, without affecting the main conclusions about the antigenic relationships between the variants included.

Next, we assessed whether the positions of the variants are robust to the exclusion of each variant in turn. We find that positions of variants are robust to the exclusion of variants (fig. S13), with the greatest movement observed with the exclusion of B.1.1.7. We then investigated whether positions of variants are robust to the exclusion of each serum group in turn (fig. S14). Overall, the map was robust to the exclusion of single serum groups, with the greatest movement observed when excluding the B.1.617.2 convalescent serum group. We also removed the outlier titrations and sera identified in fig. S3, and found little difference in map topology (fig. S15).

We also evaluated the robustness of map positions on missing titrations by performing 1000 resample bootstrap repeats, where in each repeat, a random sample of titers was selected with replacement. We performed bootstrap runs on both variants and sera combined, as well as antigens and sera separately (fig. S16). Map positions were largely robust to sampling random subsets of titers, with the biggest variability observed for the B.1.1.529 variant, likely due to that variant having only a small number of titrations.

We also show that the error is approximately equally distributed between all variants and sera (fig. S17A). Positions of variants and sera are also robust when allowing the variant and sera to move in the map while not increasing the overall map RMSE between map and table distance by more than one unit (fig. S17B).

Assessing predictive power of the antigenic map (cross-validation)

To investigate whether the antigenic map is able to predict missing titers, we performed 1000 bootstrap replicates, where in each replicate we removed 10% of all titrations, and predicted the missing titers from the resulting map.

The mean difference between predicted and measured titer was -0.38 with a standard deviation of 1.3 (fig. S18) when assuming a mean of 0. This is comparable to the difference observed between fitted and measured titers (fig. S10), and the difference between titer repeat measurements (fig. S7). We also considered the difference between predicted and measured titers split by variant and serum group (fig. S19) and found that no variant is consistently predicted as having higher or lower titers by all serum groups.

Construction of the Antibody Landscapes

The construction of the antibody landscapes differs from the methodology used in previous publications (*40*, *59*), since each serum represents a likely first infection and was used to construct the map itself with explicit assumptions about how reactivity should be distributed. Given a serum with coordinates $(x_m,y_{m,})$ let t_m be its maximum $\log_2$ titer/10. Then (x_m,y_m,t_m) represents the highest point of the antibody landscape for this serum. Given any other point p=(x,y) on the two-dimensional base space, the third coordinate t of the surface at this point is given by

$$t = t_m - d(p, p_m)$$

where $d(p,p_m)$ is the Euclidean distance between p and p_m. In particular, the serum point and maximum titer for a given serum describe how reactivity would be expected to vary in a cone-like fashion across antigenic space, with its apex at the serum point at a height equal to the maximum measured $\log_2$ titer/10 and diminishing at a rate of a two-fold reduction in titer with each antigenic unit on the map. The serum group averages shown in Fig. 3 therefore simply represent an average of all these individual slope = 1 cone-like responses for the relevant sera, displaying the GMT of all the individual titer predictions for the group.

Supplementary figures

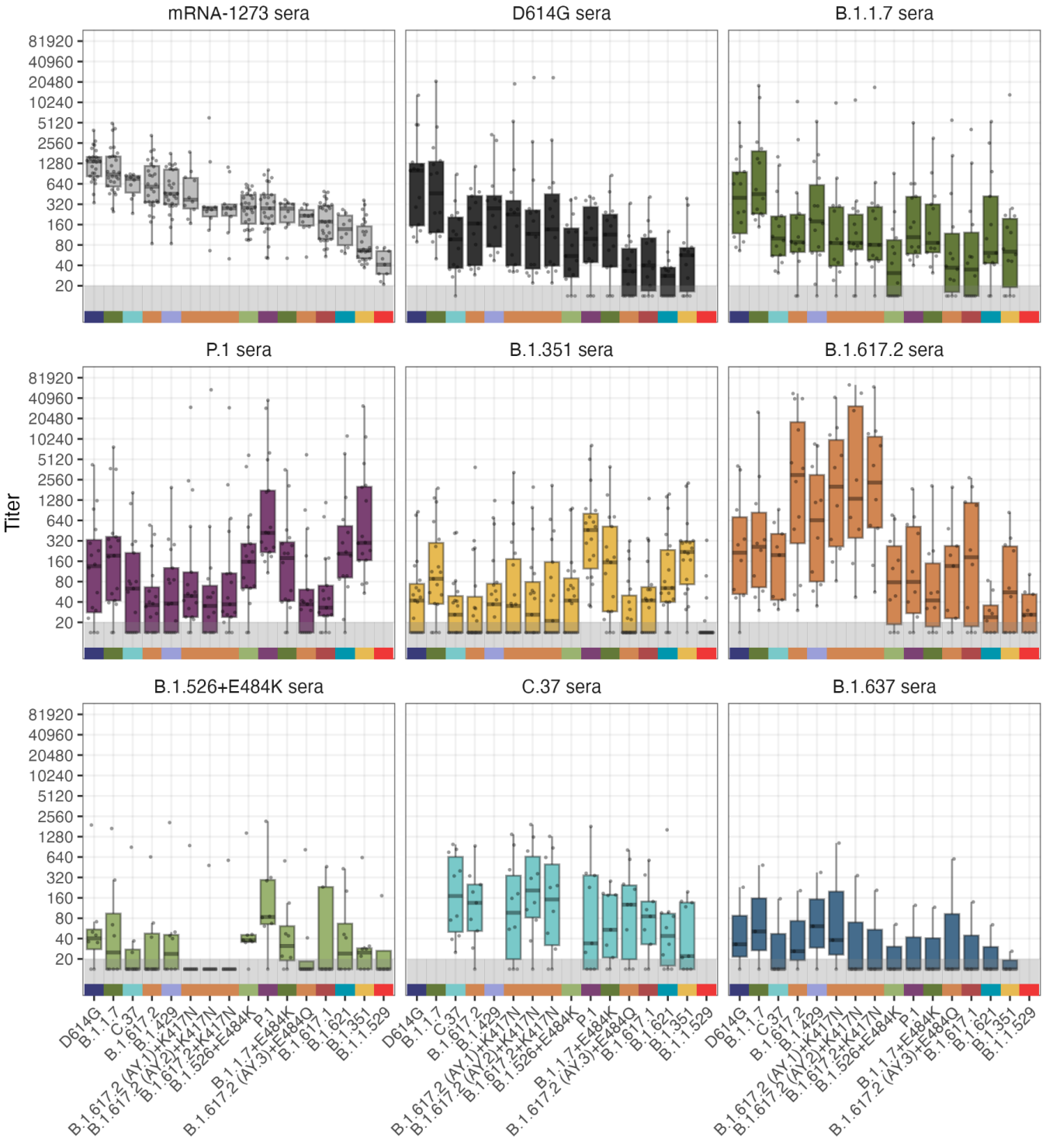

Figure S1: Neutralization of lentivirus pseudotypes encoding different SARS-CoV-2 spike proteins by sera from patients vaccinated or infected with different SARS-CoV-2 variants. The box plot indicates the median and 25th and 75th percentile. Variants are ordered according to GMT in mRNA-1273 sera (top-left), which allows for direct comparison in a column. The gray region designates titers that fell below the limit of detection of 20. Titers shown correspond to the line plot in the main manuscript Fig. 1.

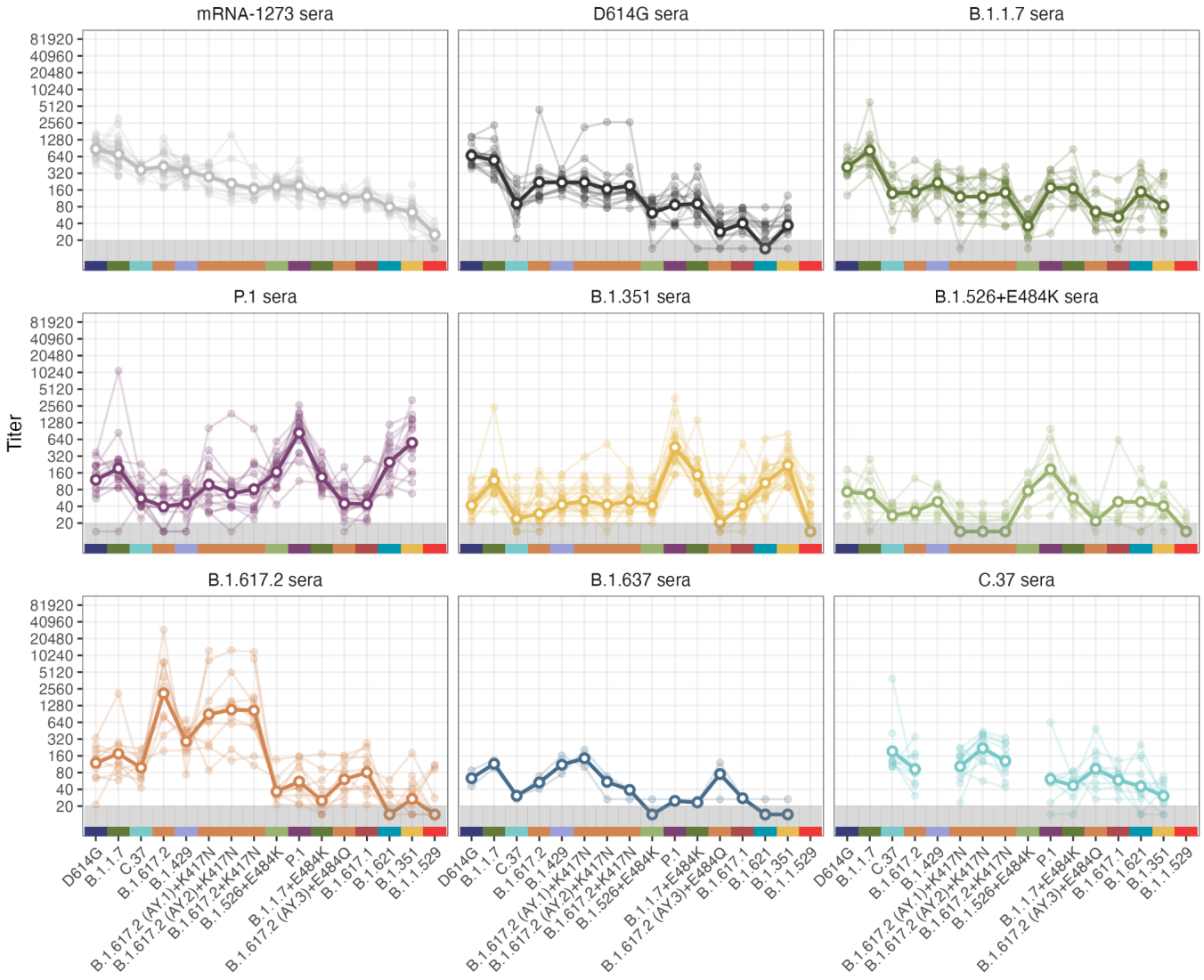

Figure S2: Neutralization of lentivirus pseudotypes encoding different SARS-CoV-2 spike proteins by sera from patients vaccinated or infected with different SARS-CoV-2 variants after accounting for individual effects. Individual effects were accounted for as described in the methods section 'Titers after accounting for estimated individual effects'. Variants are ordered according to GMT in mRNA-1273 sera (top-left), which allows for direct comparison in a column. White outlined points and the dense line shows the GMT, fainter individual lines and solid points show individual titer traces. Points in the gray region at the bottom of the plots show titers and GMTs that fell below the detection threshold of 20. GMT for data including non-detectable titers was calculated as described in the methods, 'Titer Analyses' section.

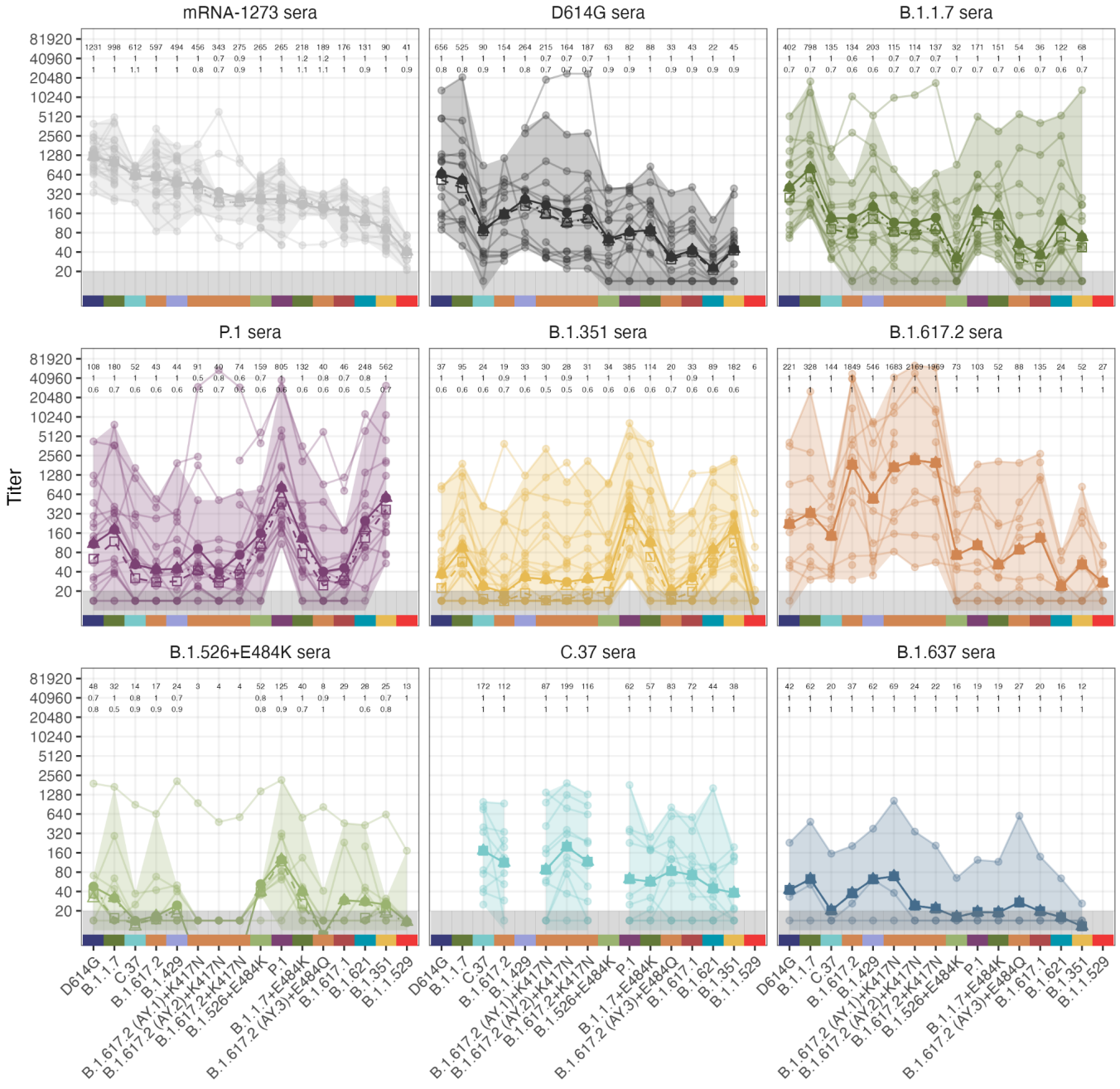

Figure S3: Effect of outlier removal on geometric mean titer against lentiviral pseudotypes encoding different SARS-CoV-2 spike proteins by patients vaccinated or infected with different variants. Shown are the GMT without outlier removal (solid line full circles), the GMT after removing outlying titrations (dotted line open triangles) and the GMT after removing serum samples with outlying titrations (dashed line open squares). Transparent lines represent individual samples. Outliers were defined as titrations outside the interval [$q_{0.25}$ - 1.5x IQR; $q_{0.75}$ + 1.5x IQR] designated by the shaded area (IQR: interquartile range; $q_{0.25/0.75}$: 1st/3rd quartile). Numbers in the top row show the GMT without outlier removal, middle row the GMT without outlying titrations relative to the overall GMT, bottom row the GMT without outlying samples relative to the overall GMT. Points in the gray region at the bottom of the plots show titers and GMTs that fell below the detection threshold of 20. GMT for data including non-detectable titers was calculated as described in the methods, 'Titer Analyses' section.

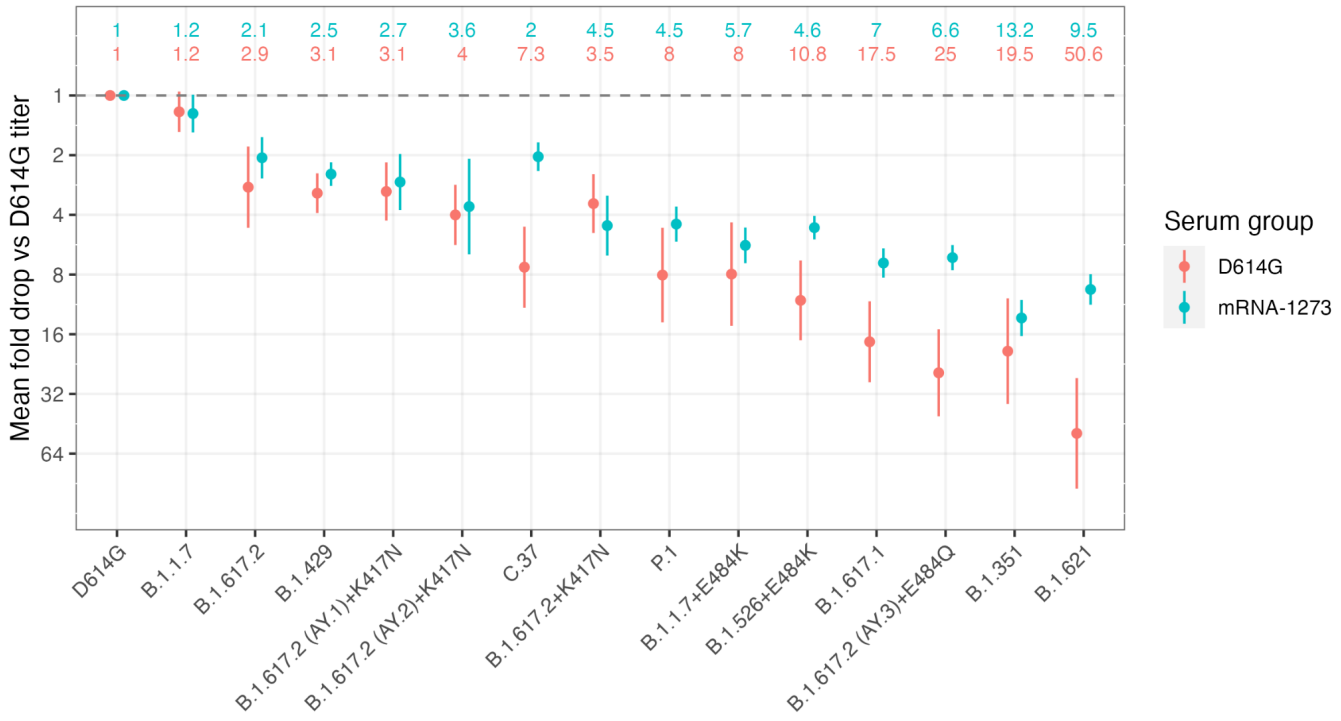

Figure S4: Comparison mean fold drop relative to the D614G variant between mRNA-1273 and D614G sera. Points show the estimate for the mean fold drop compared to the titer for D614G, while lines represent the 95% confidence interval for this estimate. The point D614G to the left of the plot represents the homologous virus against which fold-change for other strains was compared, and is therefore fixed at 1. Numbers at the top of each plot show point estimates for the fold-drop against each variant for that serum group. Fold drops and confidence intervals were calculated as described in methods 'Calculation of Geometric Mean Titers (GMT) and fold-changes for non-detectable titers'.

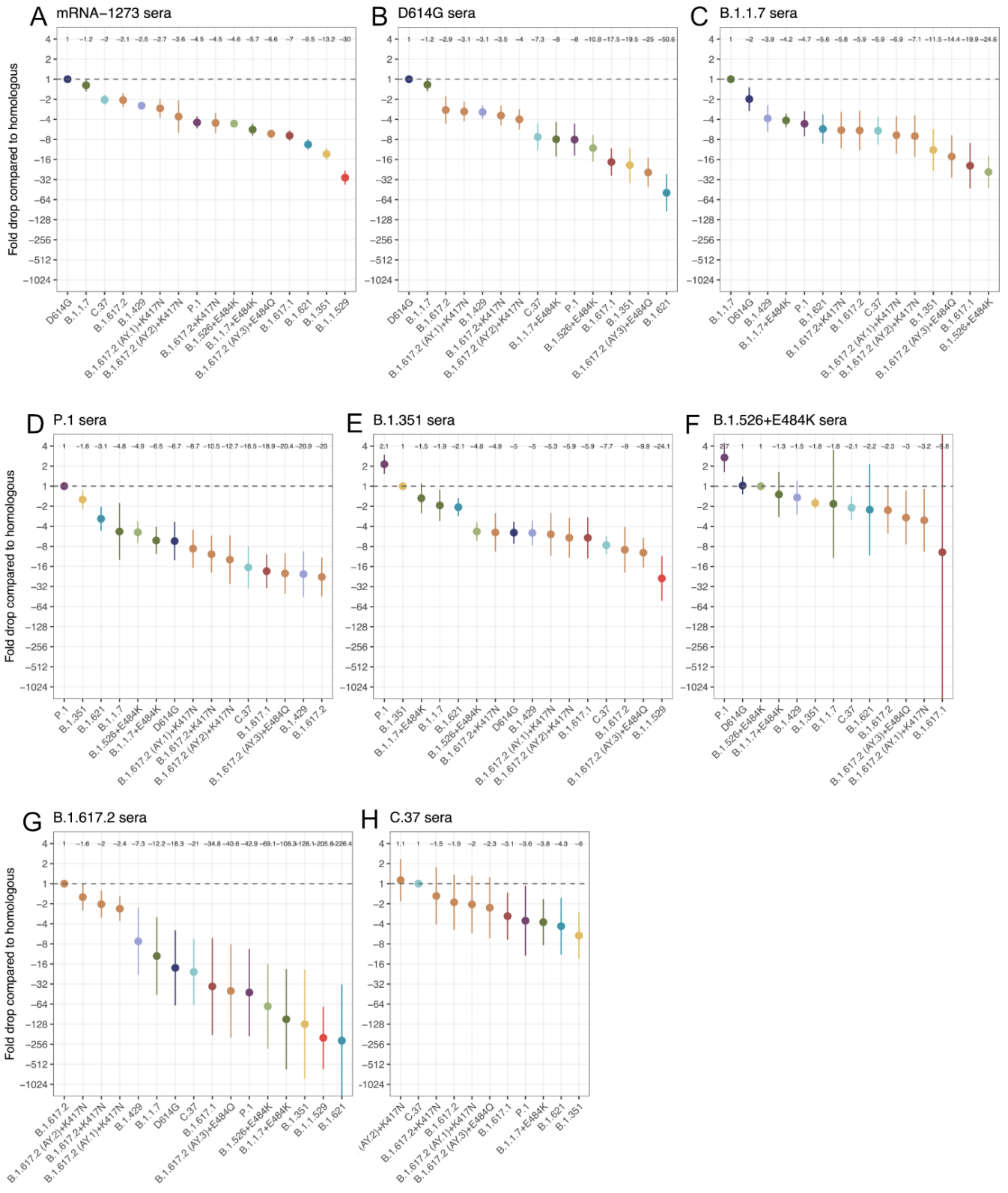

Figure S5: Estimates of fold-change compared to the homologous variant. As with fig. S4, mean fold differences for each serum group were compared to the homologous variant but here all serum groups and variants are shown. Points show the point estimate for the fold-change difference while lines show the 95% confidence interval for the estimate. Numbers at the top of each plot show point estimates for the fold-change against each variant for that serum group. Fold drops and confidence intervals were calculated as described in methods 'Calculation of Geometric Mean Titers (GMT) and fold-changes for non-detectable titers'.

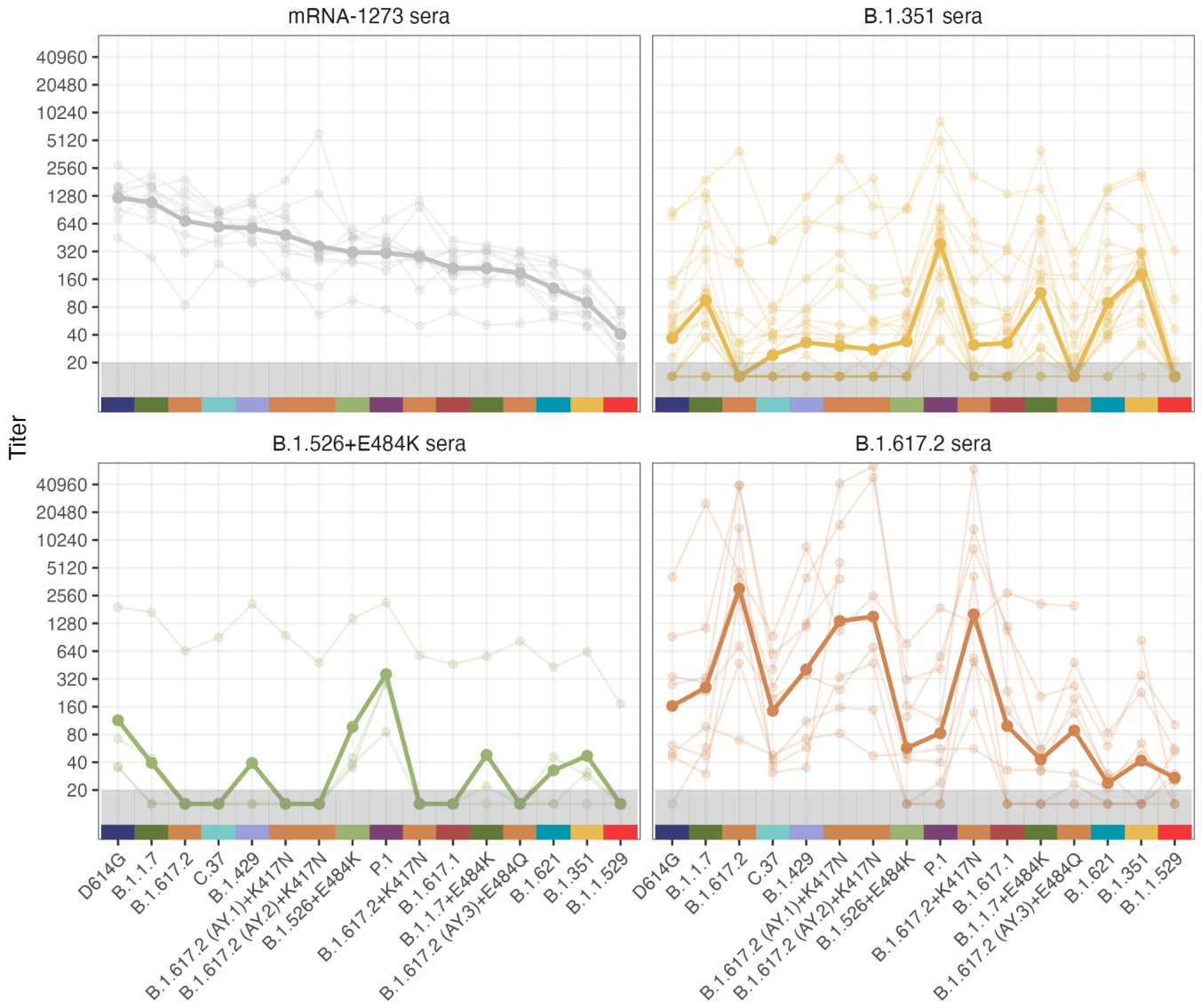

Figure S6: Neutralization of lentivirus pseudotypes encoding different SARS-CoV-2 spike proteins, subset to only those sera that B.1.1.529 was titrated against. This figure is analogous to Fig. 1, but subsetted to those sera that B.1.1.529 was titrated against. The faint lines indicate individual sera, the thicker line indicates the GMT across all individual sera. Variants are ordered according to reactivity against mRNA-1273 sera. Points in the gray region at the bottom of the plots show titers and GMTs that fell below the detection threshold of 20. GMTs for data including non-detectable titers was calculated as described in the Methods, 'Titer Analyses' section.

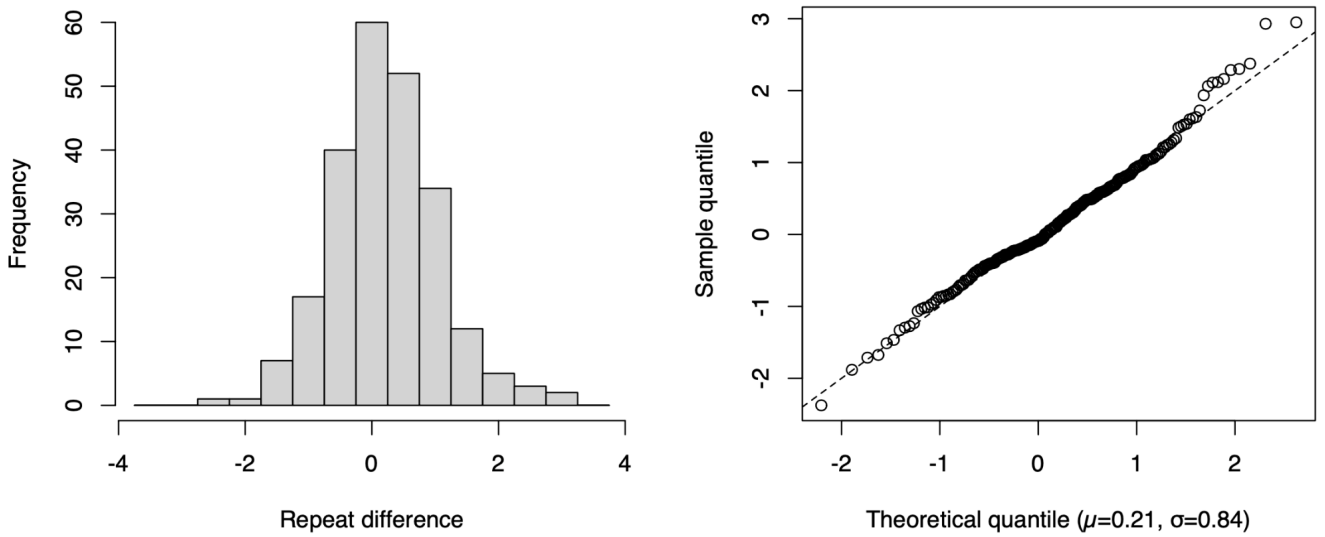

Figure S7: RMSE from repeat titrations. The left figure shows the histogram of differences between repeat titrations, while the quantile-quantile plot on the right indicates that measurement error is approximately normally distributed. The mean difference between repeated titrations was 0.209 [0.1, 0.317], with a standard deviation of 0.84. Assuming instead a mean of 0 to account for the systematic bias of titrations between repeats, the standard deviation for measurement error is 0.87 on the $\log_2$ scale. Accounting for the presence of measurement error in both the first and second titration the standard deviation of noise per measurement is $\sqrt{0.87^2/2} = 0.62$.

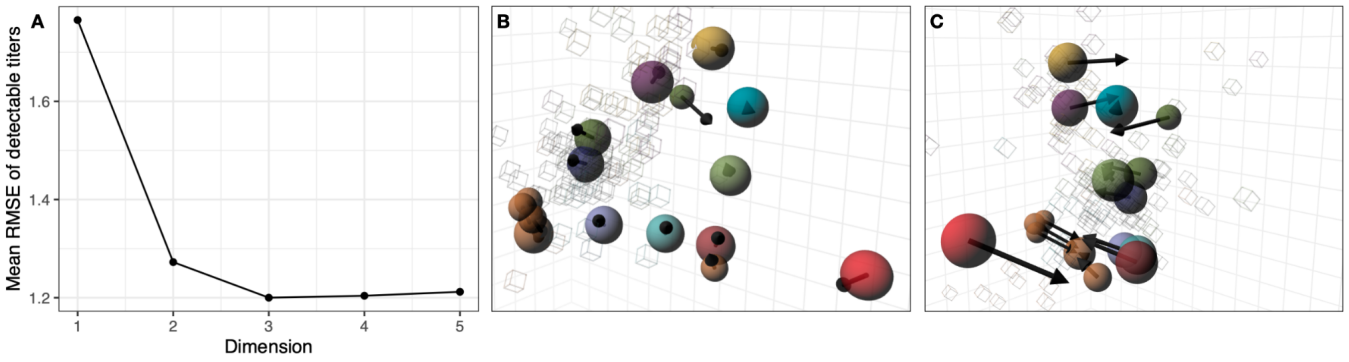

Figure S8: Dimensionality tests and 3D map. A) RMSE of detectable titers in 1 to 5 dimensions compared to known titers. Per dimension, 1000 repeats were performed wherein a map was constructed from 90% of all titers. For each run, the RMSE is calculated by comparing the titers predicted from the map against the known titers on the $\log_2$ scale. No large reduction in error is achieved when using more than two dimensions. B, C) The map optimized in 3 dimensions, with arrows pointing to where each variant is placed in two dimensions. B) 'Side' view. C) 'Front' view.

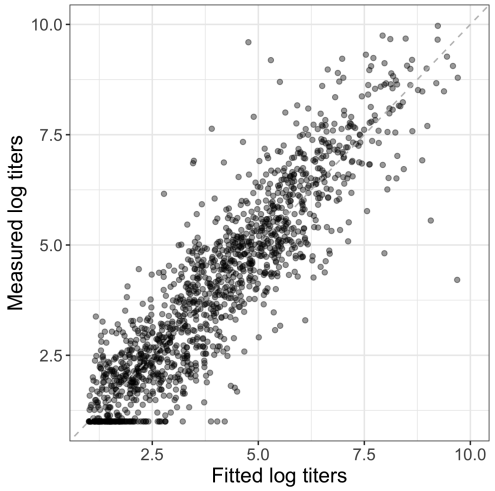

Figure S9: Scatter plot of fitted titers against measured titers. Titers are plotted on the $\log_2$ scale. Fitted titers are determined from the distances in the antigenic map. The dashed gray line is the line of best fit. Non-detectable measured titers were set to 1.

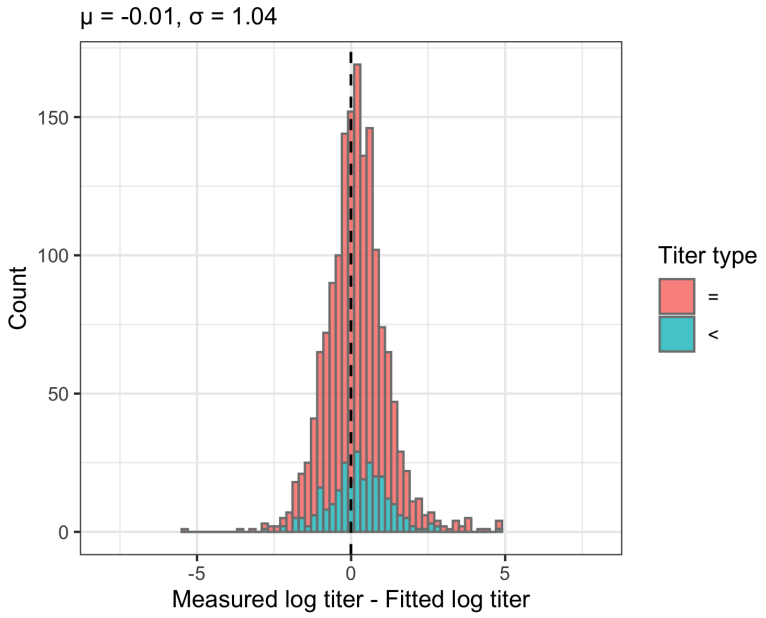

Figure S10: Histogram of differences between fitted and measured $\log_2$ titers. For non-detectable measured titers differences between fitted and measured titers were imputed as described in the methods and are indicated in blue. Differences for detectable measured titers are indicated in red. Measured titers are on average 0.01 lower than fitted titers. Assuming a mean of 0 to account for the systematic bias of titrations between repeats, the standard deviation is 1.04 on the $\log_2$ scale. The black dashed line indicates the mean.

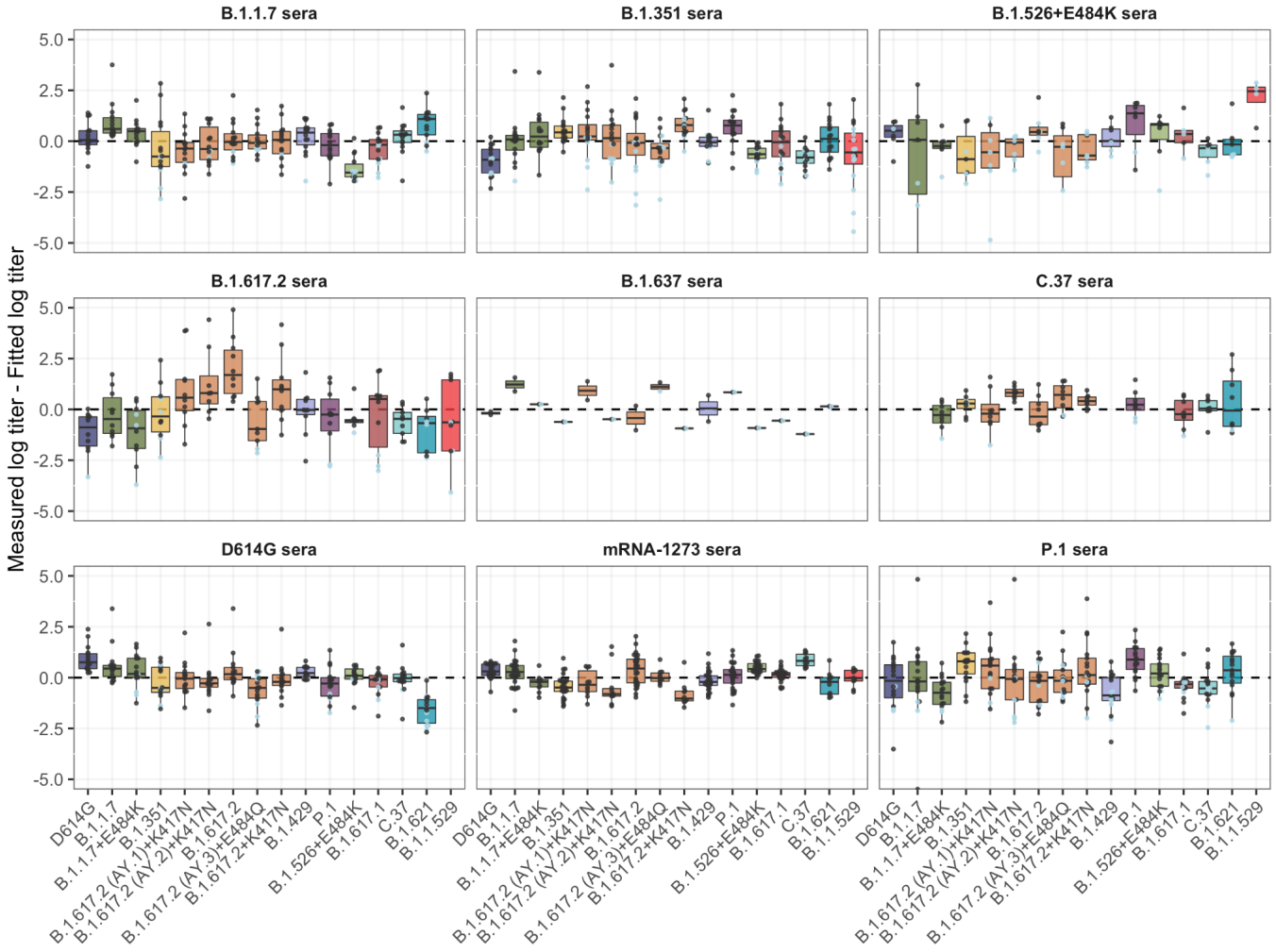

Figure S11: Measured minus fitted $\log_2$ titers split by serum group and variant. The difference between fitted and measured $\log_2$ titers is shown split by variant and serum group. Residuals for where the measured titers are non-detectable are shown in blue, residuals where the measured titers were detectable in black. The box plot indicates the median and 25th and 75th percentile.

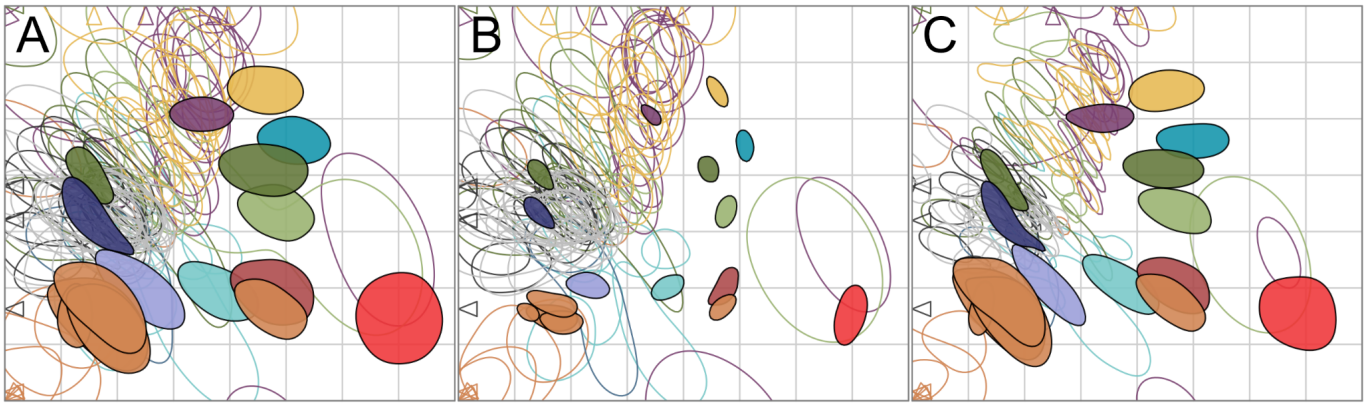

Figure S12: Assessing the effect of the uncertainty in titers and variant reactivity for the antigenic map using "smooth" bootstrap. 1000 bootstrap repeats were performed with 100 optimisations per repeat. A) Bootstrap with noise added to titers and antigen reactivity. B) Bootstrap with noise added to titers. C) Bootstrap with noise added to antigen reactivity. The noise added to the titers was normally distributed and had a standard deviation of 0.62, the noise added to the antigen reactivity a standard deviation of 0.4. Each colored region indicates the area in which 68% (one standard deviation) of the positional variation of a serum or variant position is captured. Triangles indicate the positions of variants and sera that are outside the area of the plot. For color correspondence of variants and sera, refer to Fig. 2.

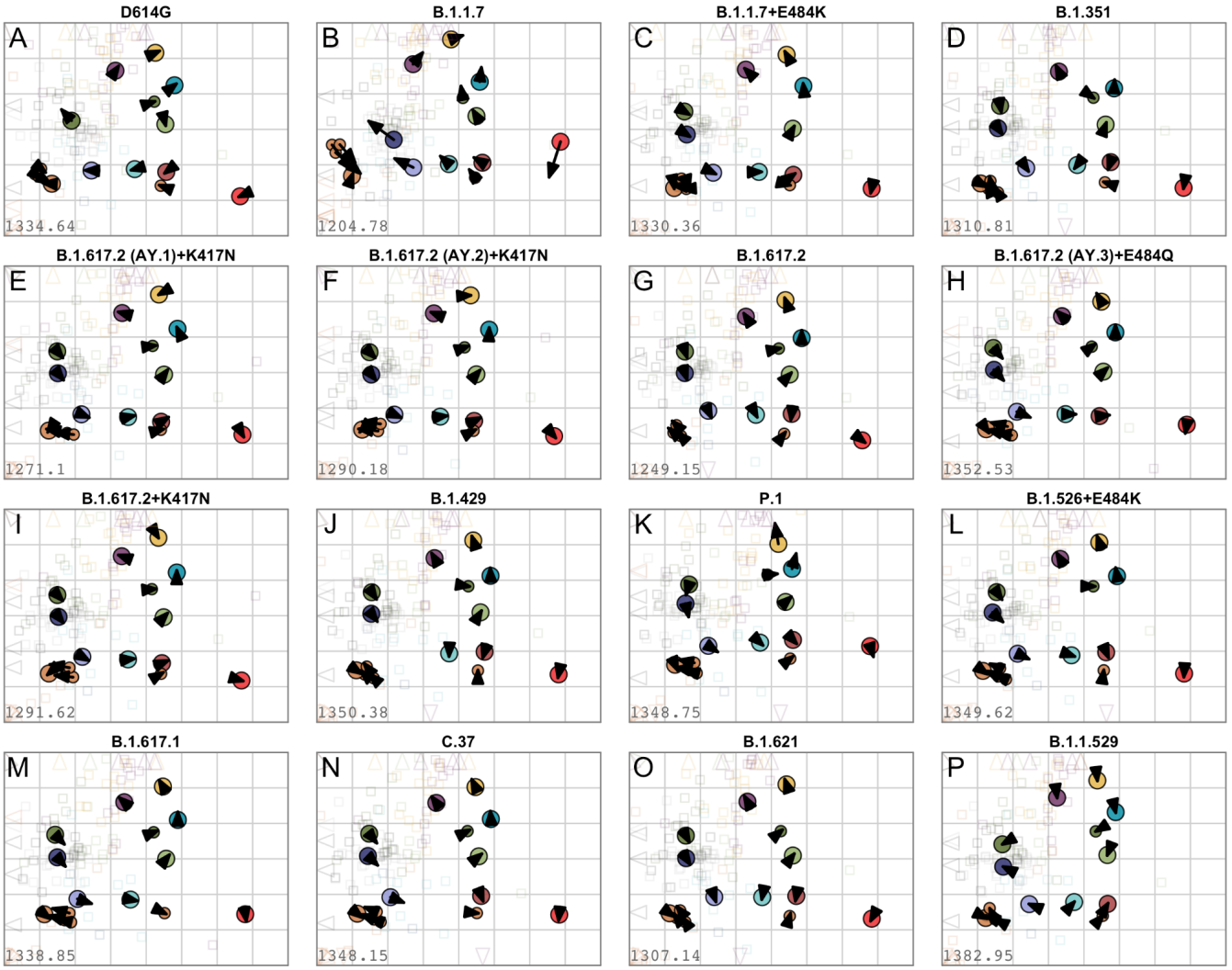

Figure S13: Assessing the effect of removing variants on map topology. Each variant was removed in turn and the map re-optimized. Arrows point to the positions of the variants in the map with all variants (shown in Fig. 2). The short length of the arrows in each case gives a visual confirmation of the robustness of the antigenic map. Triangles indicate the positions of sera that are outside the area of the plot. Removed variants: A) D614G, B) B.1.1.7, C) B.1.1.7+E484K, D) B.1.351, E) B.1.617.2 (AY.1)+K417N, F) B.1.617.2 (AY.2)+K417N, G) B.1.617.2, H) B.1.617.2 (AY.3)+E484Q, I) B.1.617.2+K417N, J) B.1.429, K) P.1, L) B.1.526+E484K, M) B.1.617.1, N) C.37, O) B.1.621, P) B.1.1.529. For color correspondence of variants and sera, refer to Fig. 2.

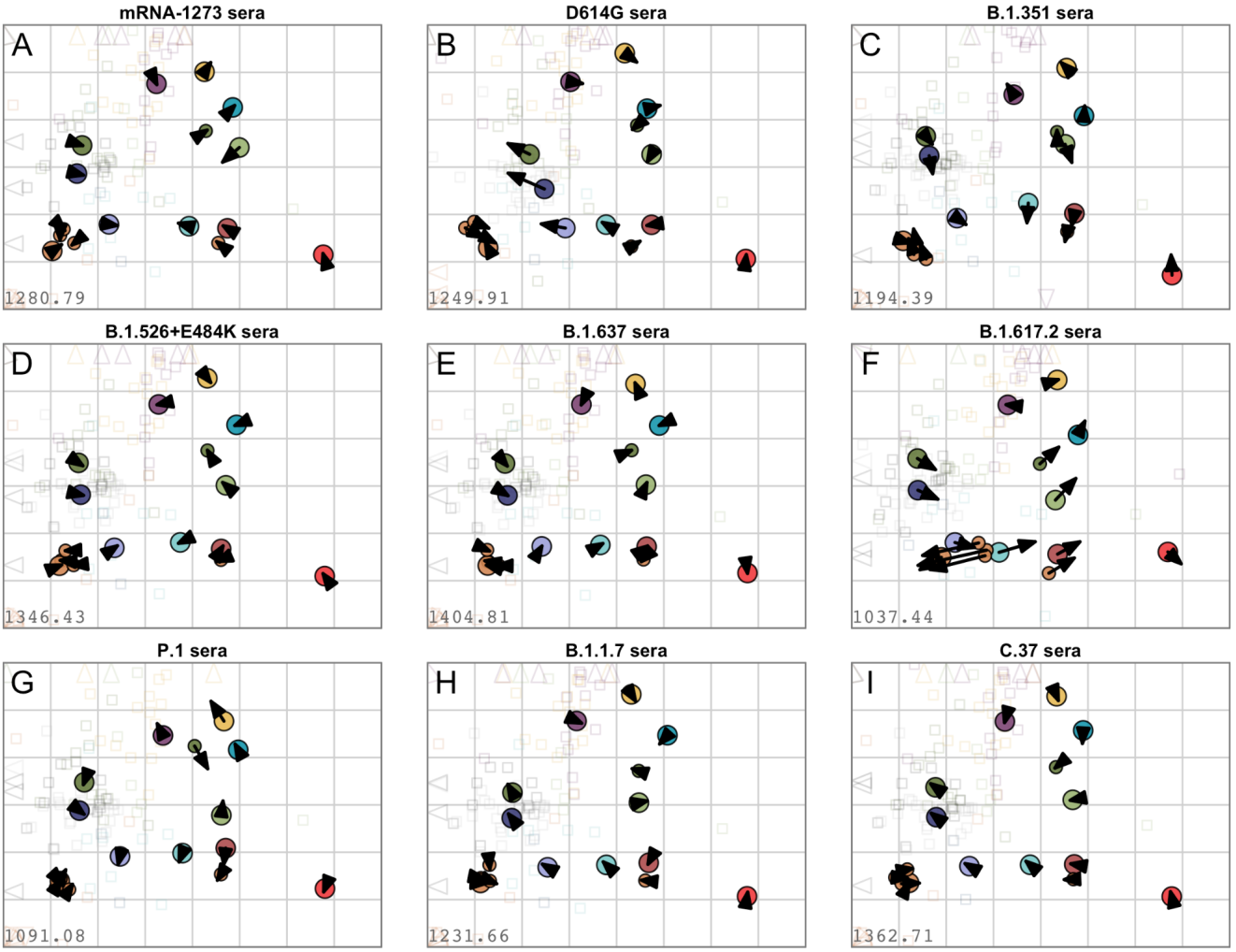

Figure S14: Assessing the effect of removing different serum groups on map topology. Each serum group was removed in turn and the map re-optimized. Arrows point to the positions of the variants in the map with all variants (shown in Fig. 2), a short arrow thus indicates little change in position of a variant. Triangles indicate the positions of sera that are outside the area of the plot. Removed serum group: A) mRNA-1273 sera, B) D614G sera, C) B.1.351 sera, D) B.1.526+E484K sera, E) B.1.637 sera, F) B.1.617.2 sera, G) P.1 sera, H) B.1.1.7 sera, I) C.37 sera. For color correspondence of variants and sera, refer to Fig. 2.

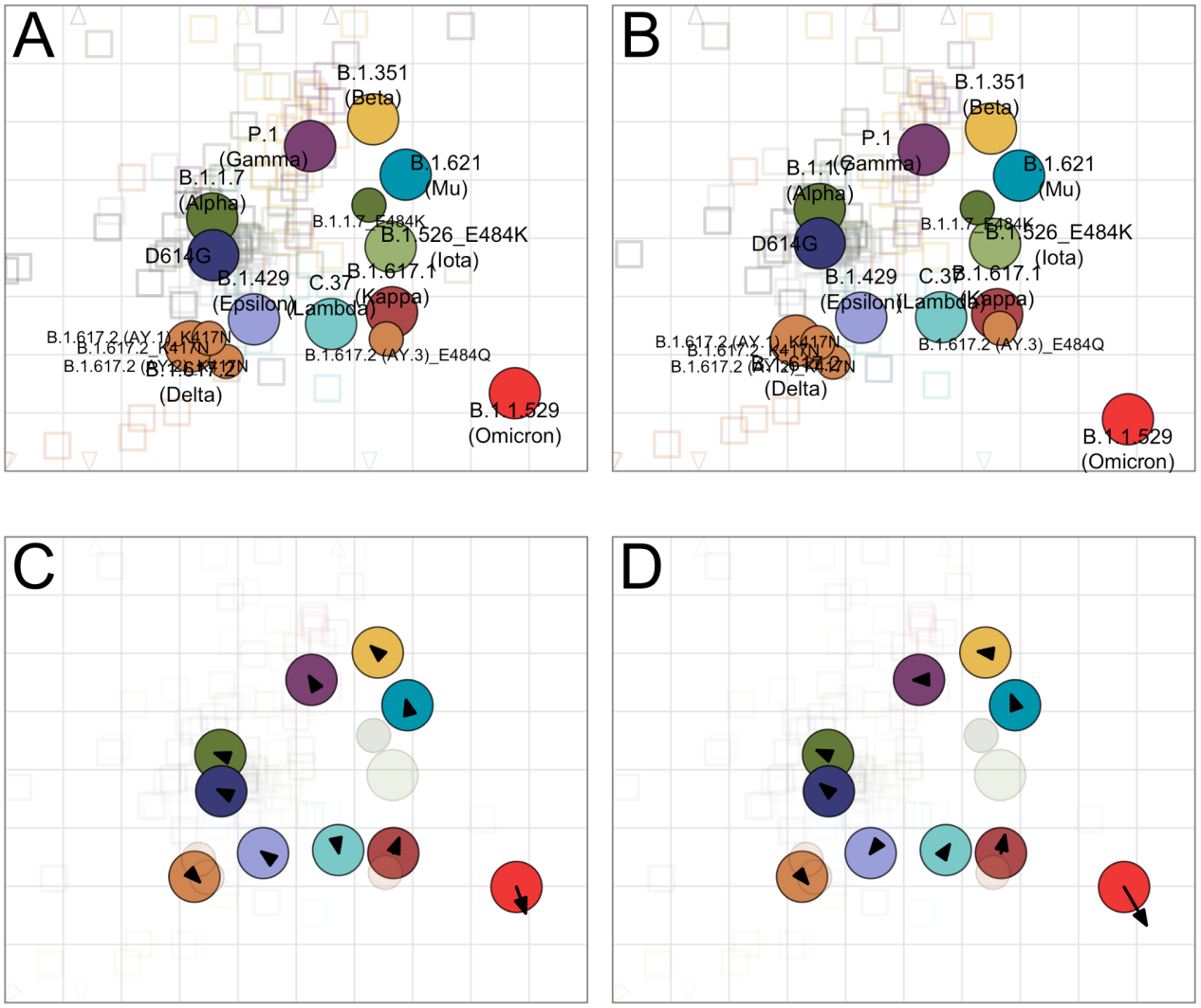

Figure S15: Assessing the effect of outlier removal on map topology. A-B) Reoptimized maps after removing outlying titrations (A) or serum samples with outlying titrations (n=16 samples) (B). C) Procrustes of the map in A to the full map with all titrations. D) Procrustes of the map in B to the full map with all sera. Outliers were identified as described in section 'Effect of outlier removal on geometric mean titer' and shown in fig. S3.

Figure S16: Assessing the robustness of the antigenic map to the inclusion of particular titer measurements. 1000 bootstrap repeats were performed where a random sample of titers were taken with replacement. A) Bootstrapping is applied to variants and sera. B) Bootstrapping is applied to variants only. C) Bootstrapping is applied to sera only. Each colored region indicates the area in which 68% (one standard deviation) of the positional variation of a serum or variant position is captured. Triangles indicate the positions of variants and sera that are outside the area of the plot. For color correspondence of variants and sera, refer to Fig. 2.

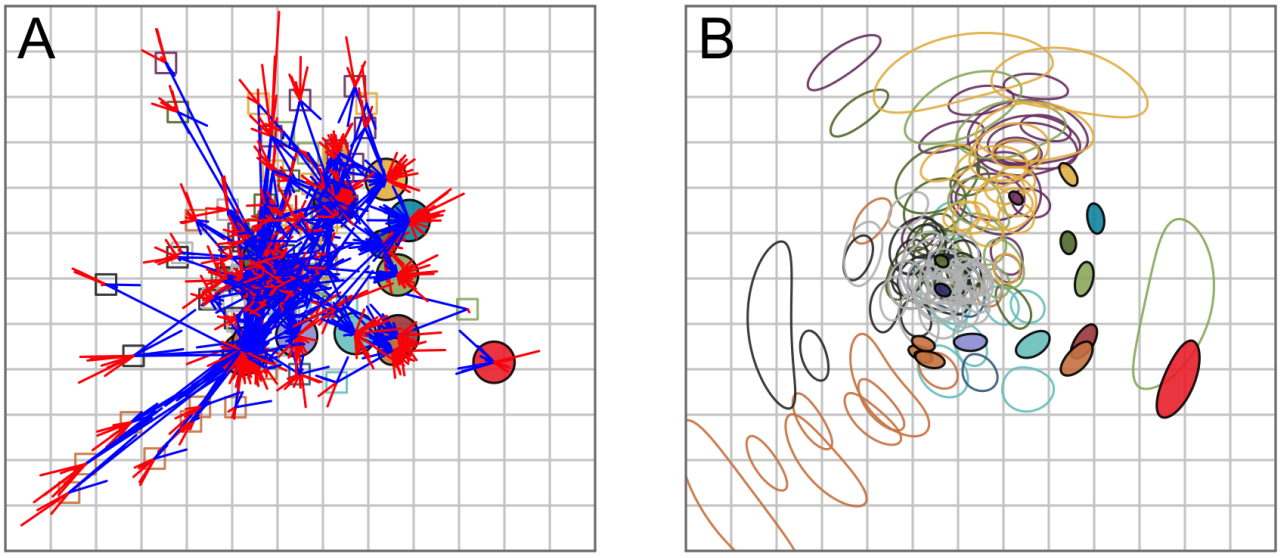

Figure S17: Maps showing the uncertainty of the placements of variants and sera. A) Antigenic map with error lines. For each titration, error lines are drawn between the variants and the serum that show the target distance between the variant and serum. Blue lines indicate a target distance shorter than the distance between the variant and serum in the map, red lines indicate that the distance on the map is shorter than the distance in the table. B) Triangulation blobs. The colored region for each serum and variant point indicates the area which the variant or serum can occupy without the stress of the map increasing by more than one unit. For color correspondence of variants and sera, refer to Fig. 2.

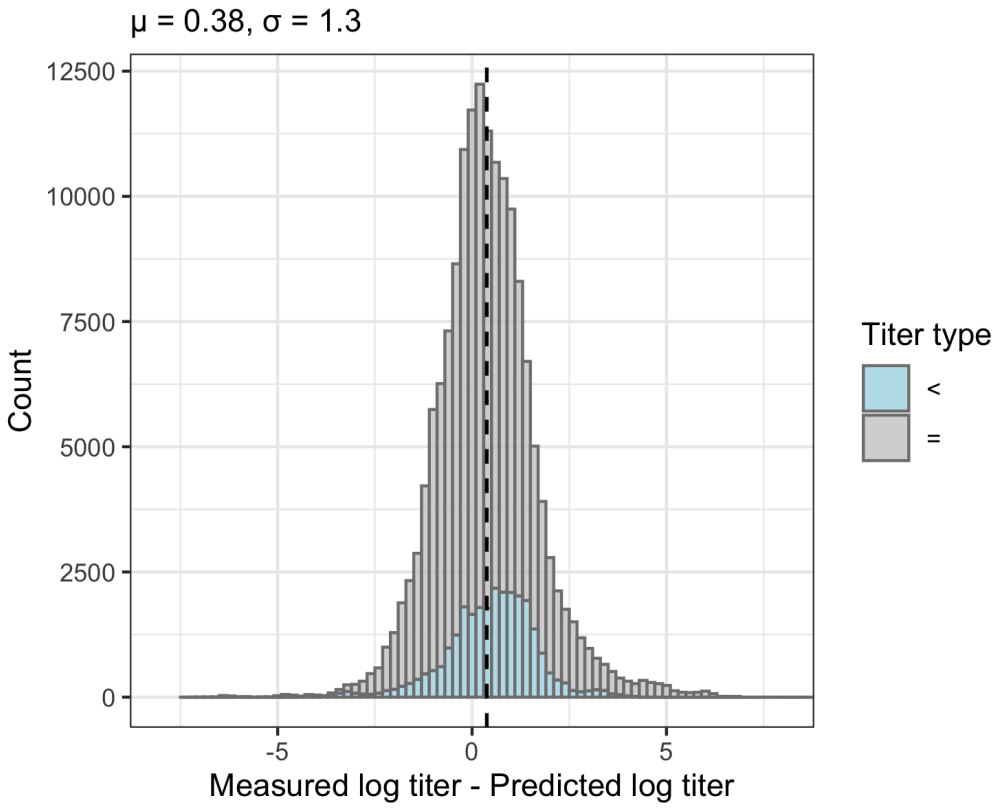

Figure S18: Histogram of differences between measured and predicted $\log_2$ titers. For non-detectable measured titers, differences between fitted and measured titers were imputed as described in the methods section, and are indicated in gray. Differences for detectable measured titers are indicated in blue. Predicted titers are on average 0.38 lower than measured titers. Assuming a mean of 0 to account for the systematic bias of titrations between repeats, the standard deviation is 1.3 on the $\log_2$ scale. The black dashed line indicates the mean.

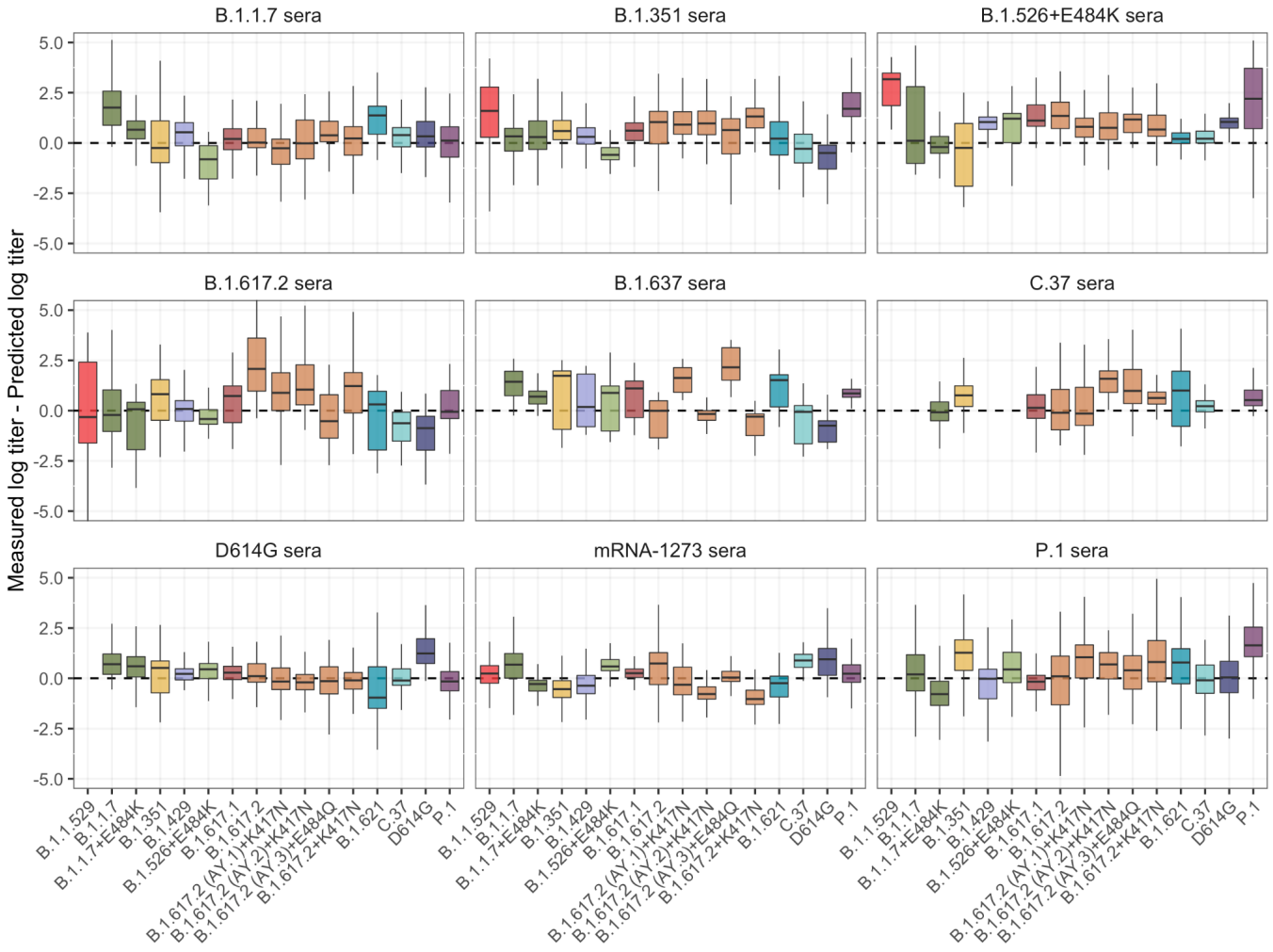

Figure S19: Measured minus predicted log_2 titers split by serum group and variant. The difference between predicted and measured log_2 titers is shown split by variant and serum group. The box plot indicates the median and 25th and 75th percentile.

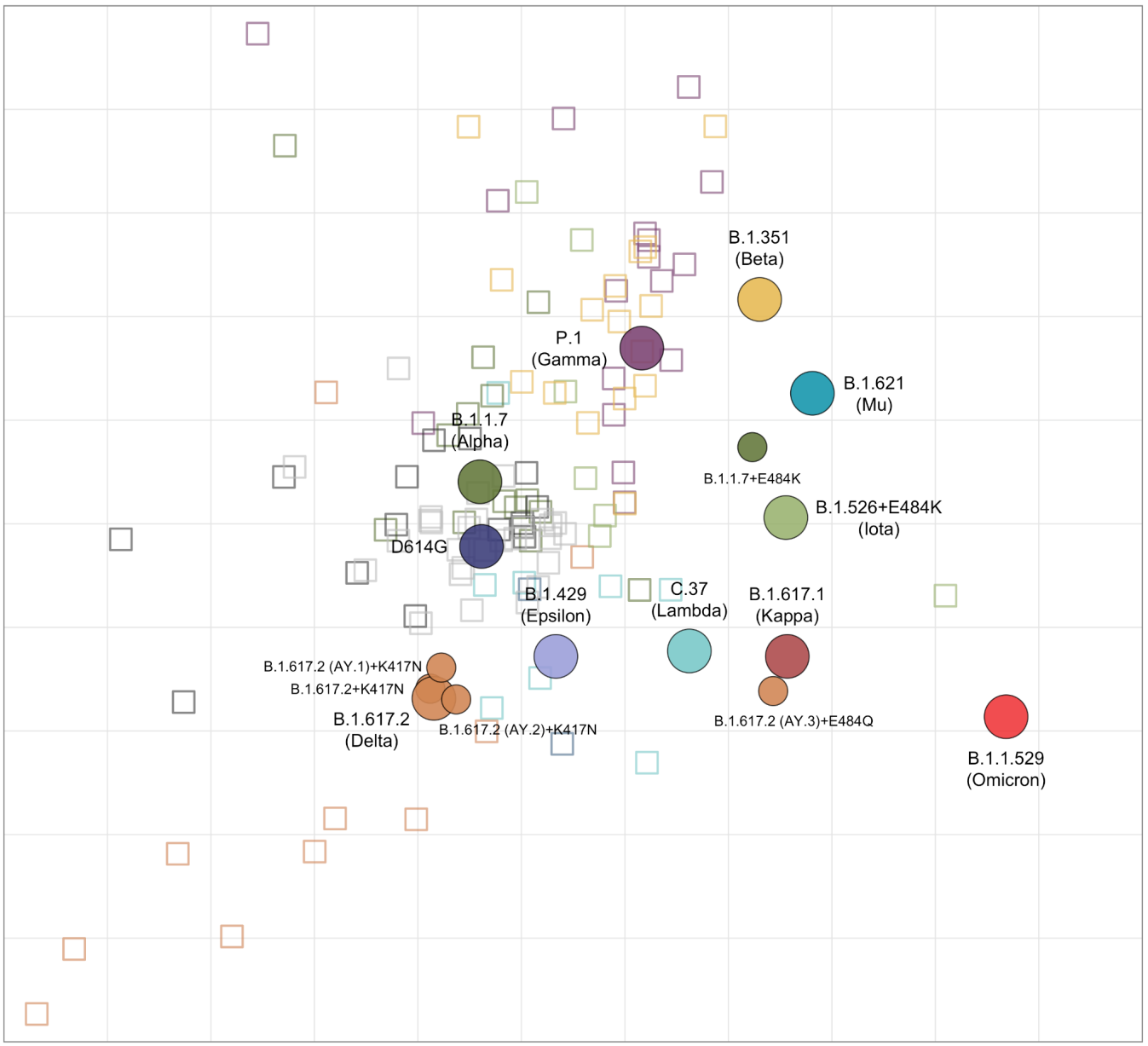

Figure S20: Zoomed out version of the antigenic map shown in Fig. 2. Variants are shown as circles, sera as squares, sera colors correspond to the color of the eliciting variant. Variants with additional substitutions from a root variant are denoted by smaller circles, in the color of their root variant. The x and y-axes both represent antigenic distance, with one grid square corresponding to one two-fold serum dilution in the neutralization assay.

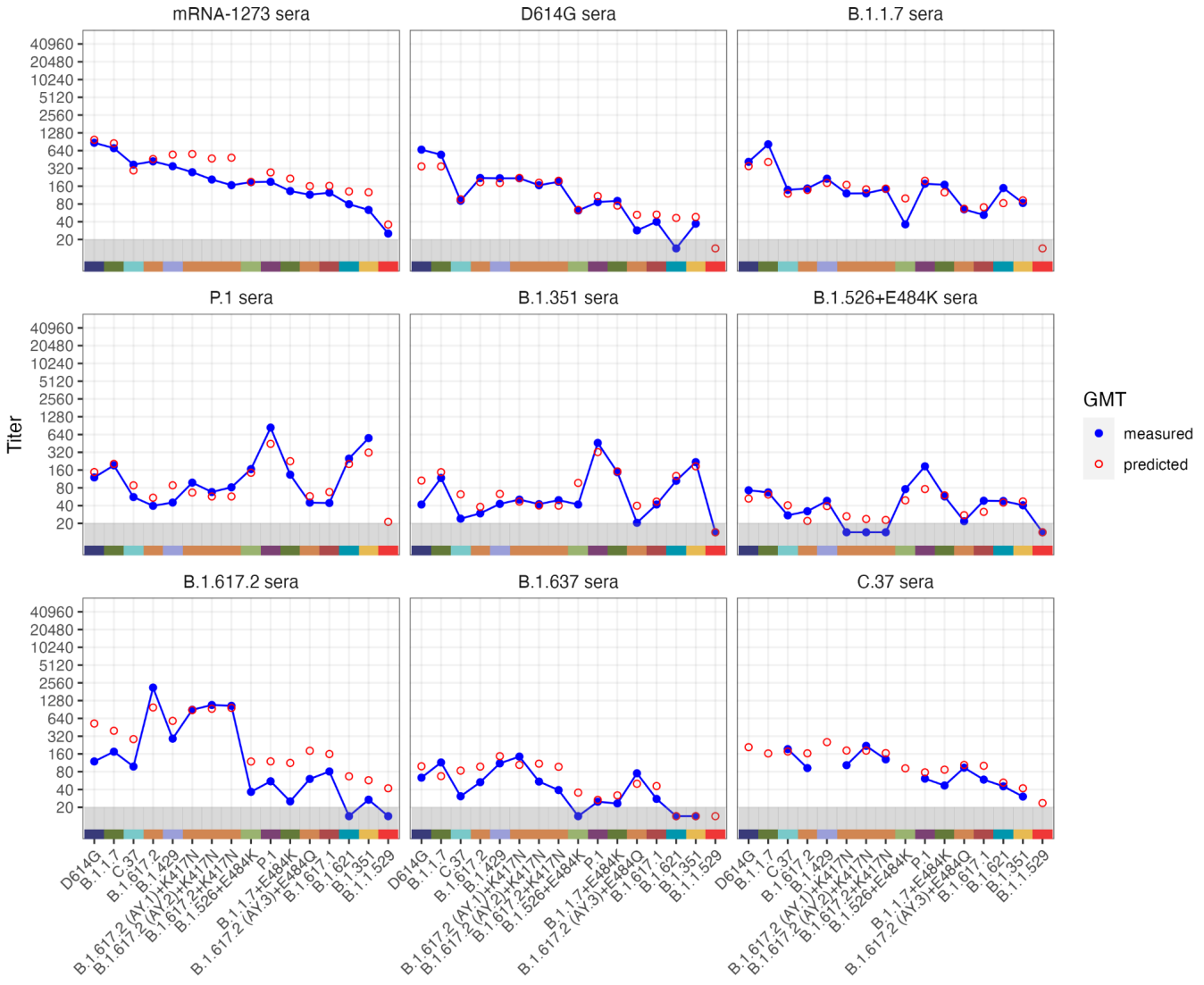

Figure S21: Measured vs antibody landscapes-predicted GMTs for each serum group. Blue points and the line show the measured GMT for each variant, after accounting for individual effects, as described above and plotted also in fig. S2, which would otherwise bias the GMT for variants not titrated against all sera. Red outlined circles show GMT predictions according to the map and antibody landscape fits as described in the section ‘Construction of the antibody landscapes’. Points in the gray region at the bottom of the plot show where the measured or predicted GMT has fallen below the assay detection threshold of 20.

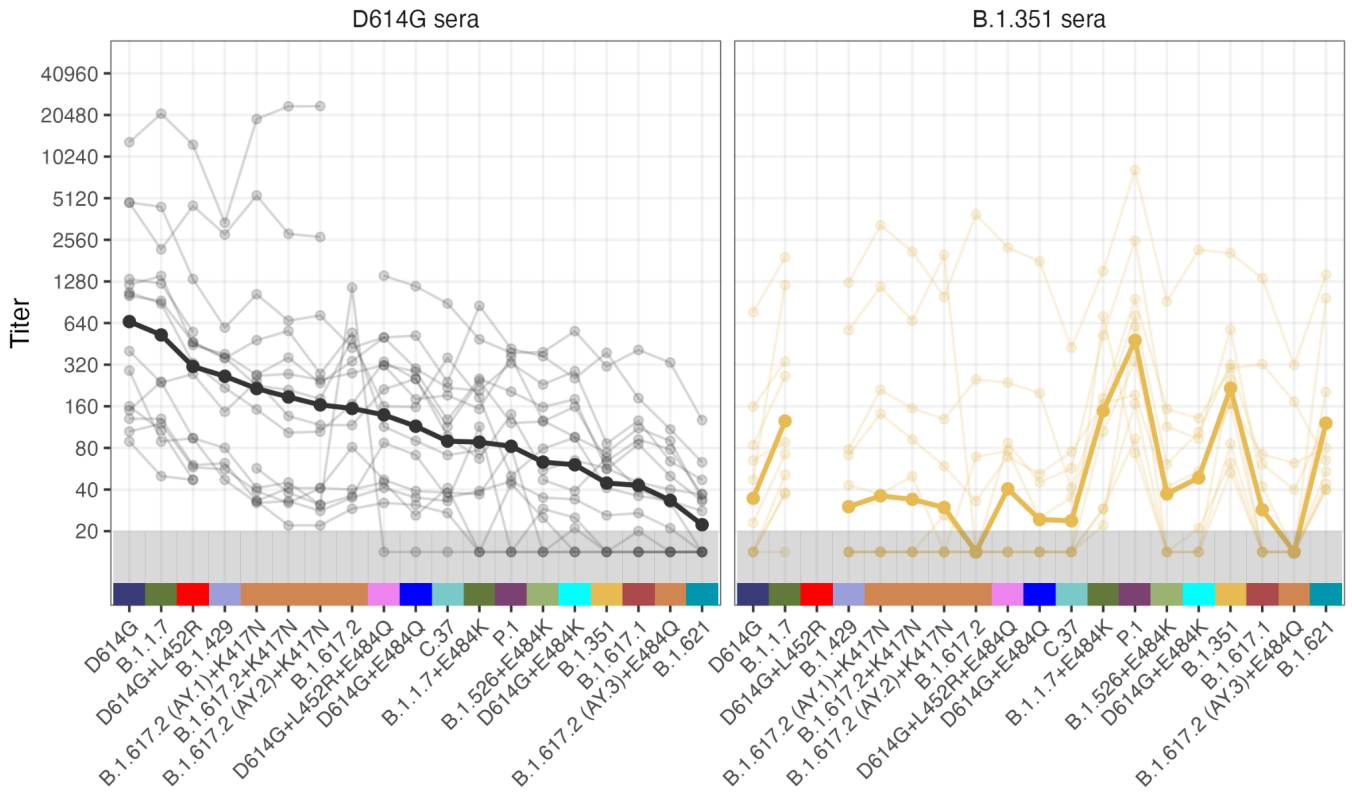

Figure S22: Titer plots including the mutants D614G+E484K, D614G+E484Q, D614G+L452R, and D614G+L452R+E484Q. The mutants were only titrated against D614G and B.1.351 serum groups. Faint lines correspond to individual sera, and the bold line indicates the GMT across all sera. Points in the gray region at the bottom of the plots show titers and GMTs that fell below the detection threshold of 20. GMT for data including non-detectable titers was calculated as described in the Methods, 'Titer Analyses' section.

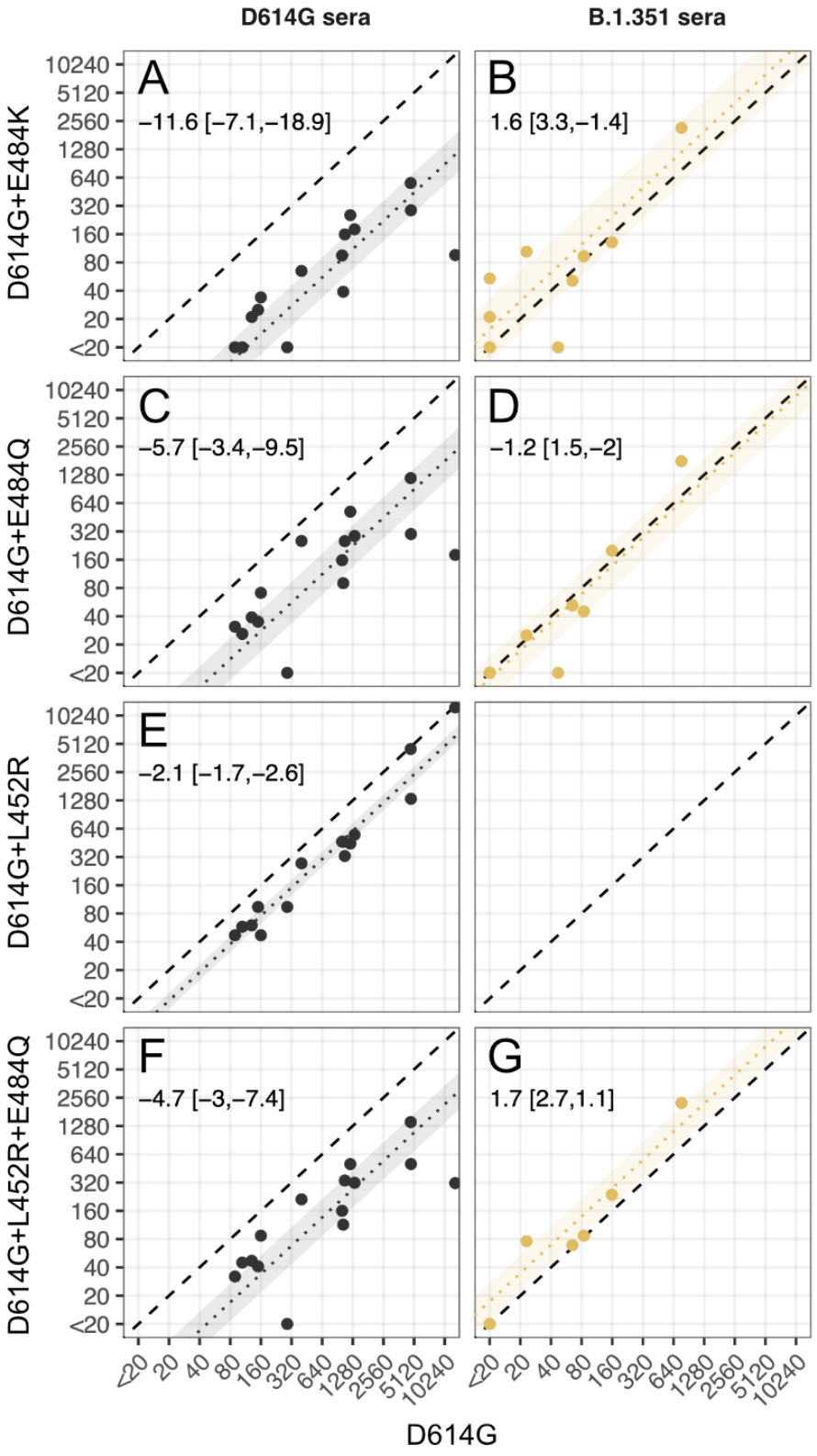

Figure S23: Scatterplots showing the comparison of titers of the single and double mutants D614G+E484K, D614G+E484Q, D614G+L452R, and D614G+L452R+E484Q to titers of the root virus D614G for the D614G and B.1.351 sera. The dotted line shows a line with slope 1 and intercept equal to the fold-change point estimate of difference between the variants on the x and y-axes; shaded regions represent the 95% confidence interval for these estimates. The dashed black line shows the line of equality. Numbers in the top left corner show the associated point estimates and confidence intervals expressed as fold-changes. A, C, E, F) D614G sera, B, D, G) B.1.351 sera.

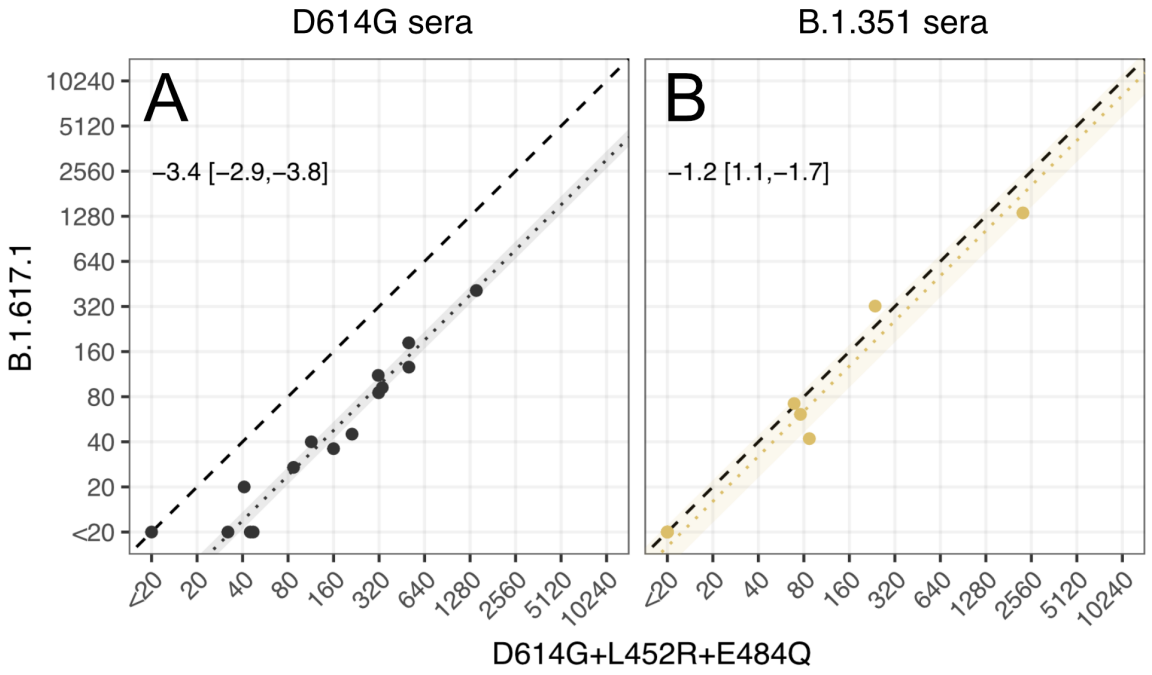

Figure S24: Scatterplots showing the comparison of titers between the D614G+L452R+E484Q mutant and the B.1.617.1 variant. A) Titers against D614G sera. B) Titers against B.1.351 sera. The dotted line shows a line with slope 1 and intercept equal to the fold-change point estimate of difference between the variants on the x and y-axes; shaded regions represent the 95% confidence interval for these estimates. The dashed black line shows the line of equality. Numbers in the top left corner show the associated point estimates and confidence intervals expressed as fold-changes.

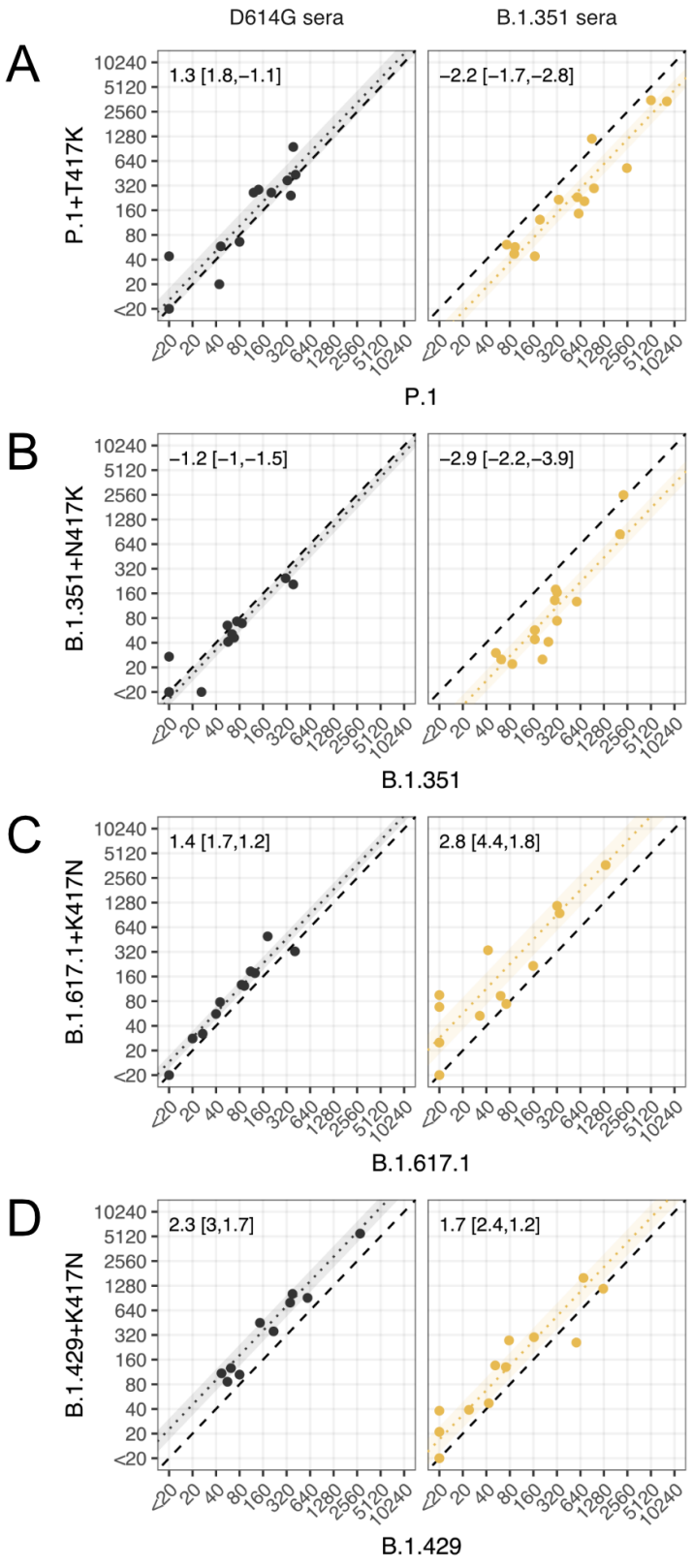

Figure S25: Scatterplots showing the effect of introducing substitutions at 417 into different variant backgrounds. The dotted line shows a line with slope 1 and intercept equal to the fold-change point estimate of difference between the variants on the x and y-axes; shaded regions represent the 95% confidence interval for these estimates. The dashed black line shows the line of equality. Numbers in the top left corner show the associated point estimates and confidence intervals expressed as fold-changes.

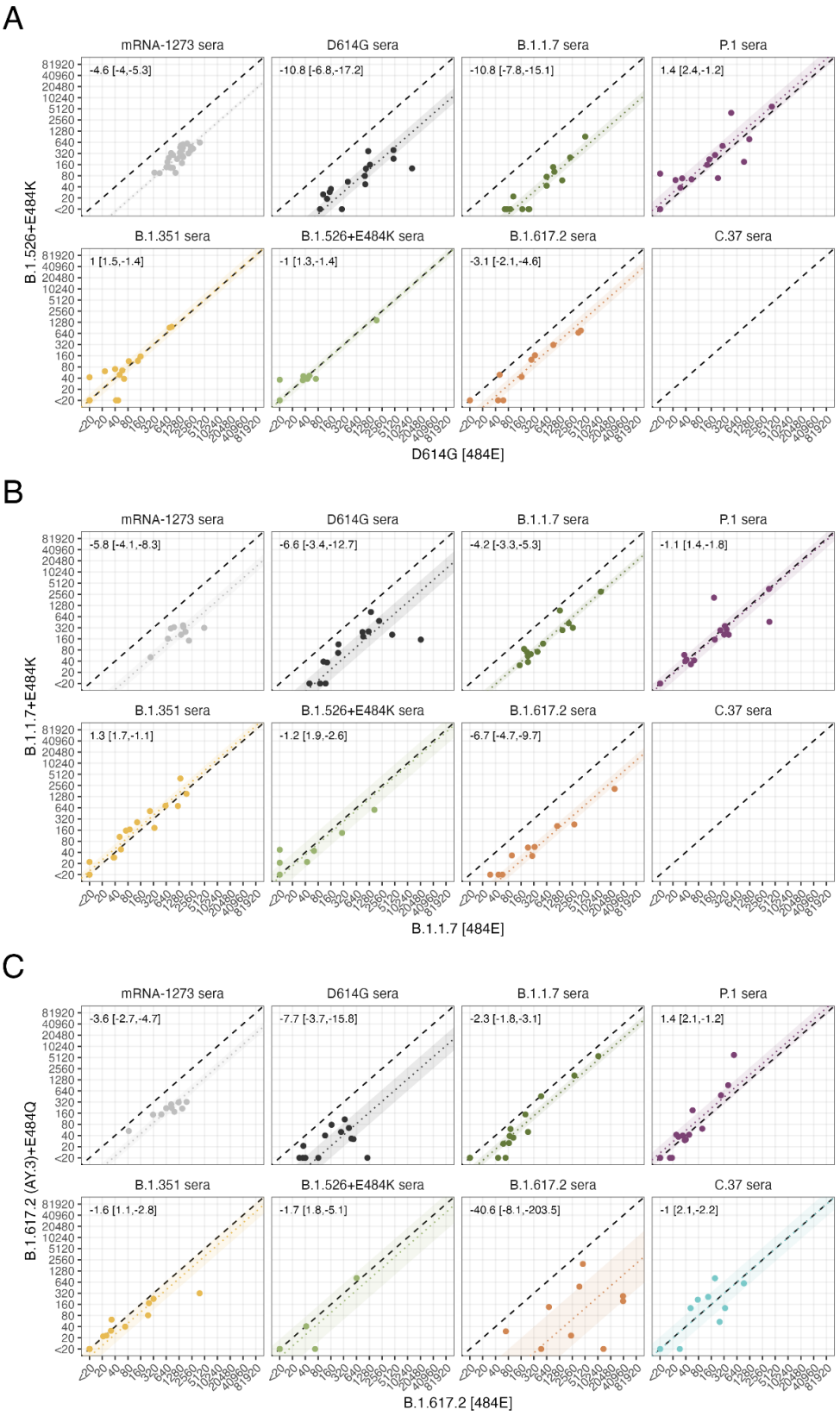

Figure S26: Effect of a pairwise difference at position 484. A) Titers against D614G vs B.1.526+E484K variants. B) Titers against B.1.1.7 vs B.1.1.7+E484K variants. C) Titers against B.1.617.2 vs B.1.617.2 (AY.3)+E484Q variants. The dotted line shows a line with slope 1 and intercept equal to the fold-change point estimate of difference between the variants on the x and y-axes; shaded regions represent the 95% confidence interval for these estimates. The dashed black line shows the line of equality. Numbers in the top left corner show the associated point estimates and confidence intervals expressed as fold-changes.

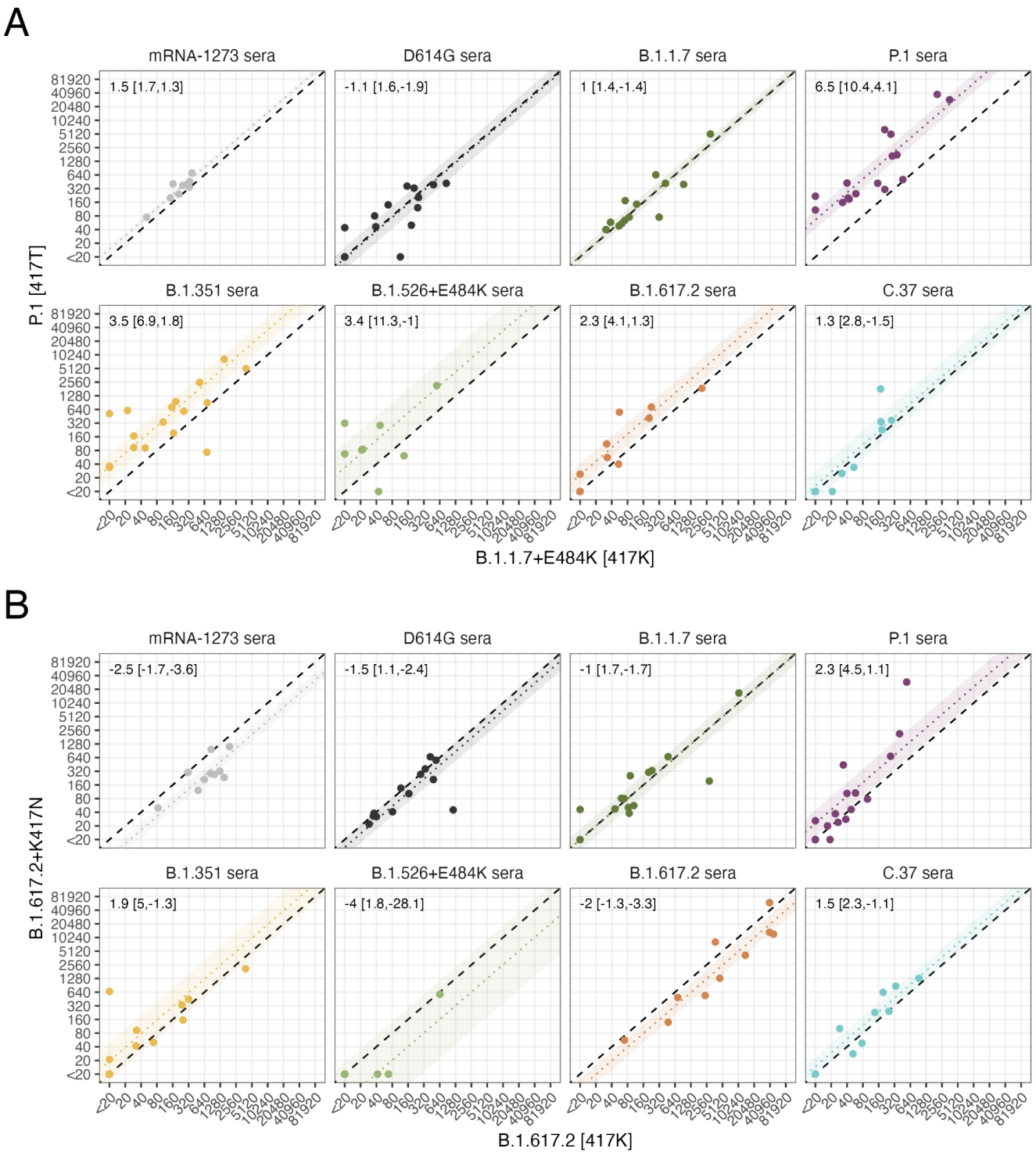

Figure S27: Effect of a pairwise difference at position 417. The dotted line shows a line with slope 1 and intercept equal to the fold-change point estimate of difference between the variants on the x and y-axes; shaded regions represent the 95% confidence interval for these estimates. The dashed black line shows the line of equality. Numbers in the top left corner show the associated point estimates and confidence intervals expressed as fold-changes. A) B.1.1.7 vs P.1 variants. B) B.1.6172 vs B.1.617.2+K417N variants.

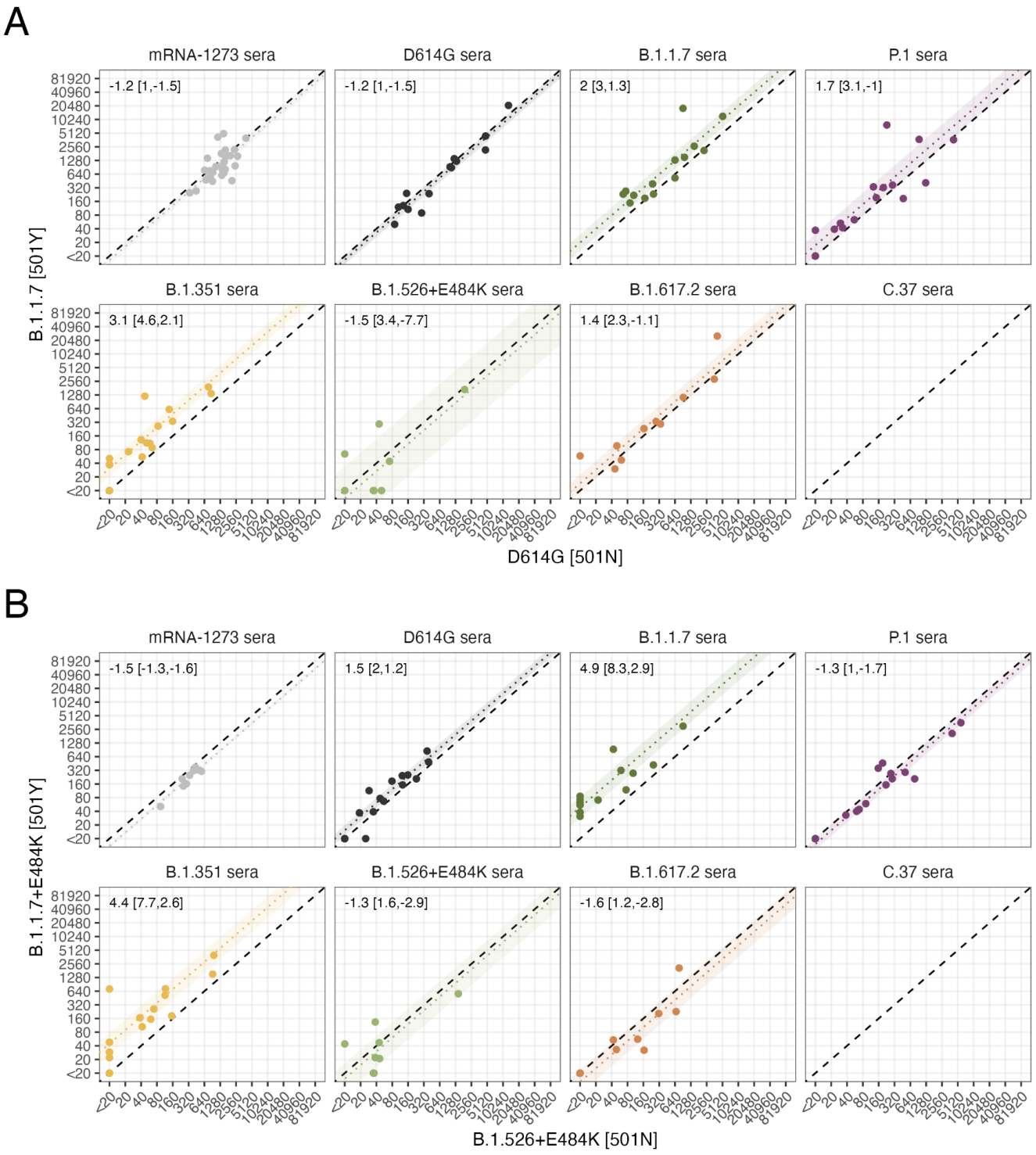

Figure S28: Effect of a pairwise difference at position 501. The dotted line shows a line with slope 1 and intercept equal to the fold-change point estimate of difference between the variants on the x and y-axes; shaded regions represent the 95% confidence interval for these estimates. The dashed black line shows the line of equality. Numbers in the top left corner show the associated point estimates and confidence intervals expressed as fold-changes. A) D614G vs B.1.1.7 variants. B) B.1.526+E484K vs B.1.1.7+E484K variants.

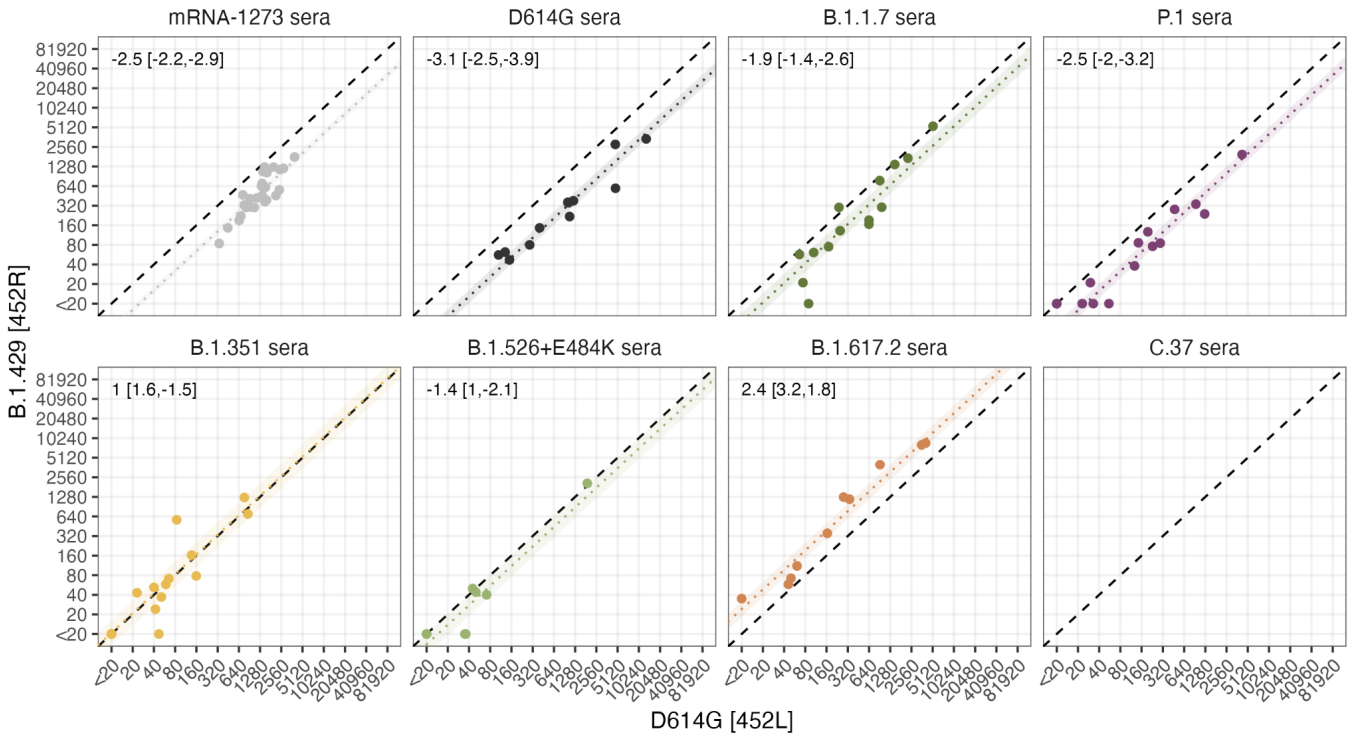

Figure S29: Effect of a pairwise difference at position 452. The dotted line shows a line with slope 1 and intercept equal to the fold-change point estimate of difference between the variants on the x (D614G) and y-axes (B.1.429); shaded regions represent the 95% confidence interval for these estimates. The dashed black line shows the line of equality. Numbers in the top left corner show the associated point estimates and confidence intervals expressed as fold-changes.

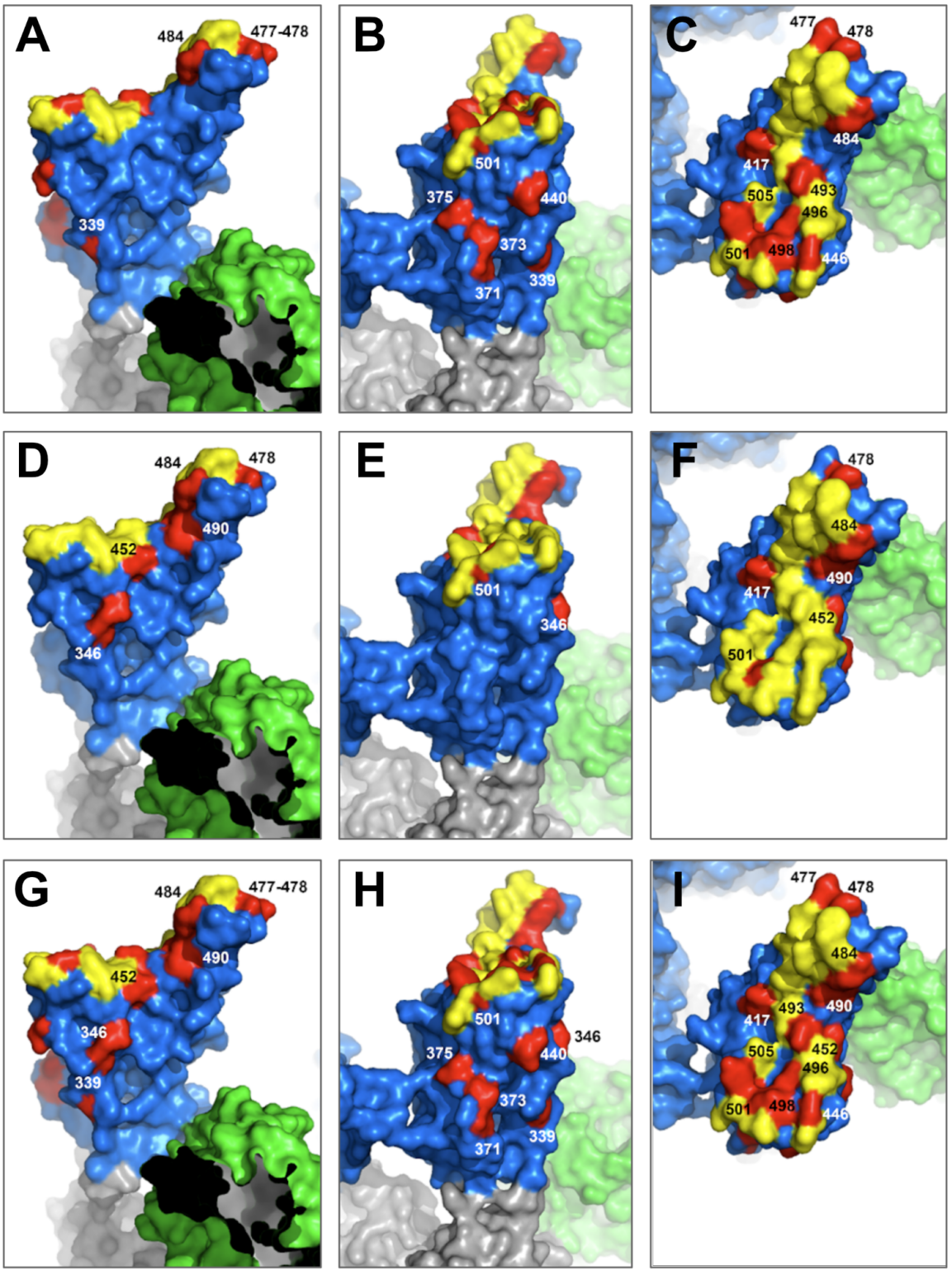
A
484
477-478
339
B
501
375
440
373
339
371
C
477
478
484
417
505
493
496
501
498
446
D
484
478
490
452
346
E
501
346
F
478
484
490
417
452
501
G
484
477-478
490
452
346
339
H
501
346
375
440
373
339
371
I
477
478
484
493
490
417
505
452
496
501
498
446

Figure S30: Visualization of RBD substitutions present in variants in this study on the SARS-CoV-2 spike structure. RBD in blue, RBD-ACE2 interface residues in yellow, NTD in green, variant substitutions in red. Substitutions labeled only on the "up" conformation of the RBD. A-C) Positions with substitutions in the B.1.1.529 variant. D-F) Substitutions of the positions with substitutions in non-B.1.1.529 variants from this study. G-I) Positions with substitutions in all variants in this study. A, D, G) Solvent-accessible front of the RBD. B, E, H) Side of the RBD (at the RBD-RBD interface when the RBD is in the “closed” conformation). C, F, I) RBD-ACE2 interaction interface of the RBD. The structure was produced using Modeller, by combining data from three structures with protein data bank accessions 7KNB, 7C2L, 7L2C.

Figure S31: Comparison of putative antigenic substitutions in SARS-CoV-2 and Influenza A/H3N2. See figure 2 in Koel et al., 2013 (*36*).

Supplementary Tables

Sample Group	Agency /Cohort[#]	Lead Contacts	Collection Location	Infection Stage at Collection[$]
mRNA-1273 (N=28)	Moderna	Pajon	United States	Non infected
D614G sera (N=15)	CoVPN	Hural	United States	Convalescent
B.1.1.7 sera (N=14)	BC CDC	Jassem	Canada	Convalescent
B.1.351 sera (N=19)	CoVPN	Hural	Africa	Convalescent
P.1 sera (N=17)	BC CDC	Jassem	Canada	Convalescent
	UW-Madison	Kawaoka	Japan	Active infection
B.1.617.2 sera (N=12)	Duke University	Datto & Denny	United States	Active infection
	BC CDC	Jassem	Canada	Convalescent
	UW-Madison	Kawaoka	Japan	Active infection
B.1.525+E484K sera (N=8)	CDC/C-HEaRT	Veguilla	United States	Convalescent
	Icahn School of Medicine at Mount Sinai	Krammer & Simon	United States	Convalescent/Longevity
C.37 sera (N=10)	St. Jude Children's Research Hospital	Webby	Peru	Active infection /Convalescent
B.1.637 sera (N=3)	CDC/C-HEaRT	Veguilla	United States	Convalescent

Table S1: Description of sera used in this study. [#]BC CDC-BC Centre for Disease Control; CoVPN- COVID-19 Prevention Network; UW-Madison- University of Wisconsin, Madison; CDC-Centers for Disease Control and Prevention; C-HEaRT- Coronavirus Household Evaluation and Respiratory Testing Cohort. [$]Defined by days since symptom onset or first positive diagnosis. Active infection: <20 days; Convalescent: ≥28 days and <60 days; Longevity: ≥60 days.

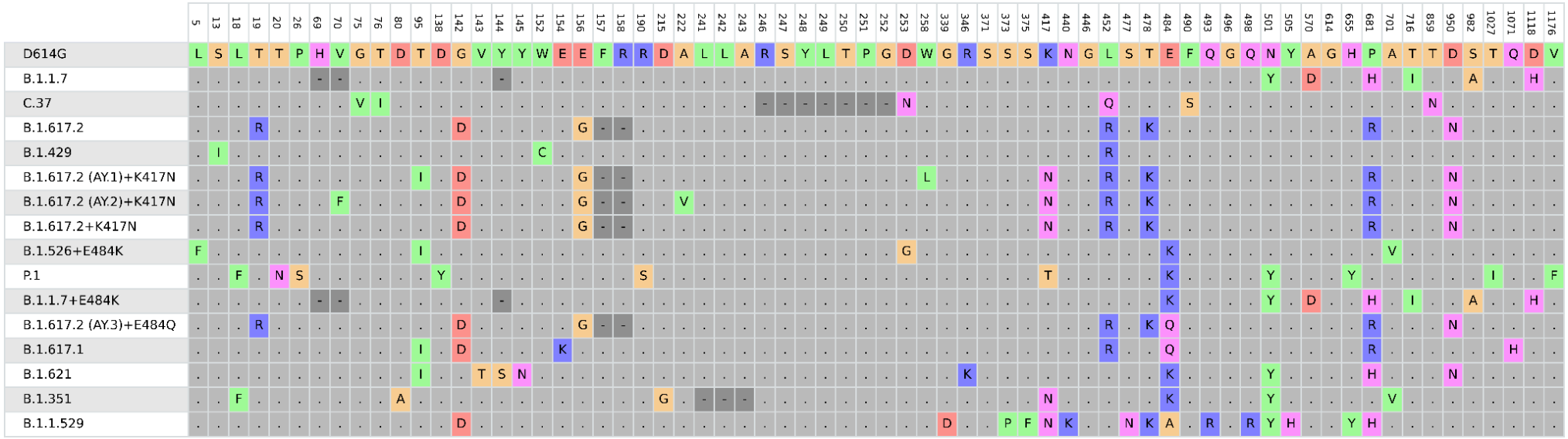

Table S2: A comparison of sequence differences of variants used in this study. This table displays the amino acids for each of the variants for positions in the entire spike protein. The coloring is according to the Lesk scheme which categorizes amino acids based on physico-chemical similarity (*60*) as follows; orange: small nonpolar, green: hydrophobic, magenta: polar, red: negatively charged, blue: positively charged.