Cell Host & Microbe, Volume *29*

Supplemental information

Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2

Haibo Wu, Na Xing, Kaiwen Meng, Beibei Fu, Weiwei Xue, Pan Dong, Wanyan Tang, Yang Xiao, Gexin Liu, Haitao Luo, Wenzhuang Zhu, Xiaoyuan Lin, Geng Meng, and Zhenglin Zhu

Supplementary Figures

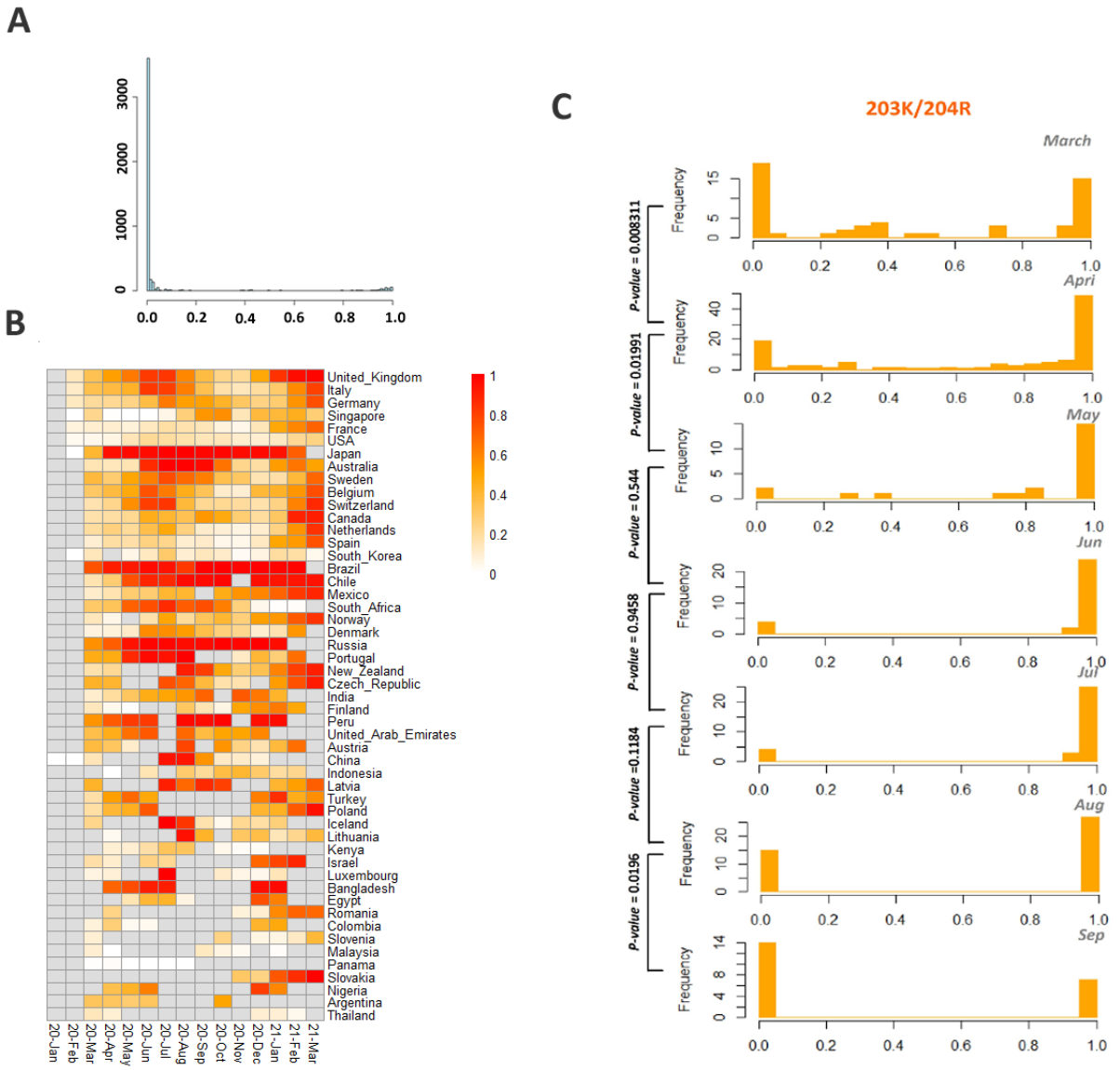

Figure S1. Statistics of SARS-CoV-2 mutations, related to Figure 2.

(A) The distribution of ρ^2 for pairs of mutations. (B) A heatmap showing the IF changes in different countries over months. The data corresponds to Figure 2C. We did not show the countries without an IF > 0 identified in any month, mostly due to an insufficient number of samples. The empty IFs are coloured grey. (C) The changes in the distribution of IFs over time for 203K/204R. The iSNV analyses are based on SARS-CoV-2 raw-sequence data collected in the United Kingdom. The X-axis represents the IF, whose value ranges from 0 to 1 (fixed), and the Y-axis refers to the counts of samples with different IF values. The results of the Wilcoxon test for the difference between two distributions are shown at the left of each figure.

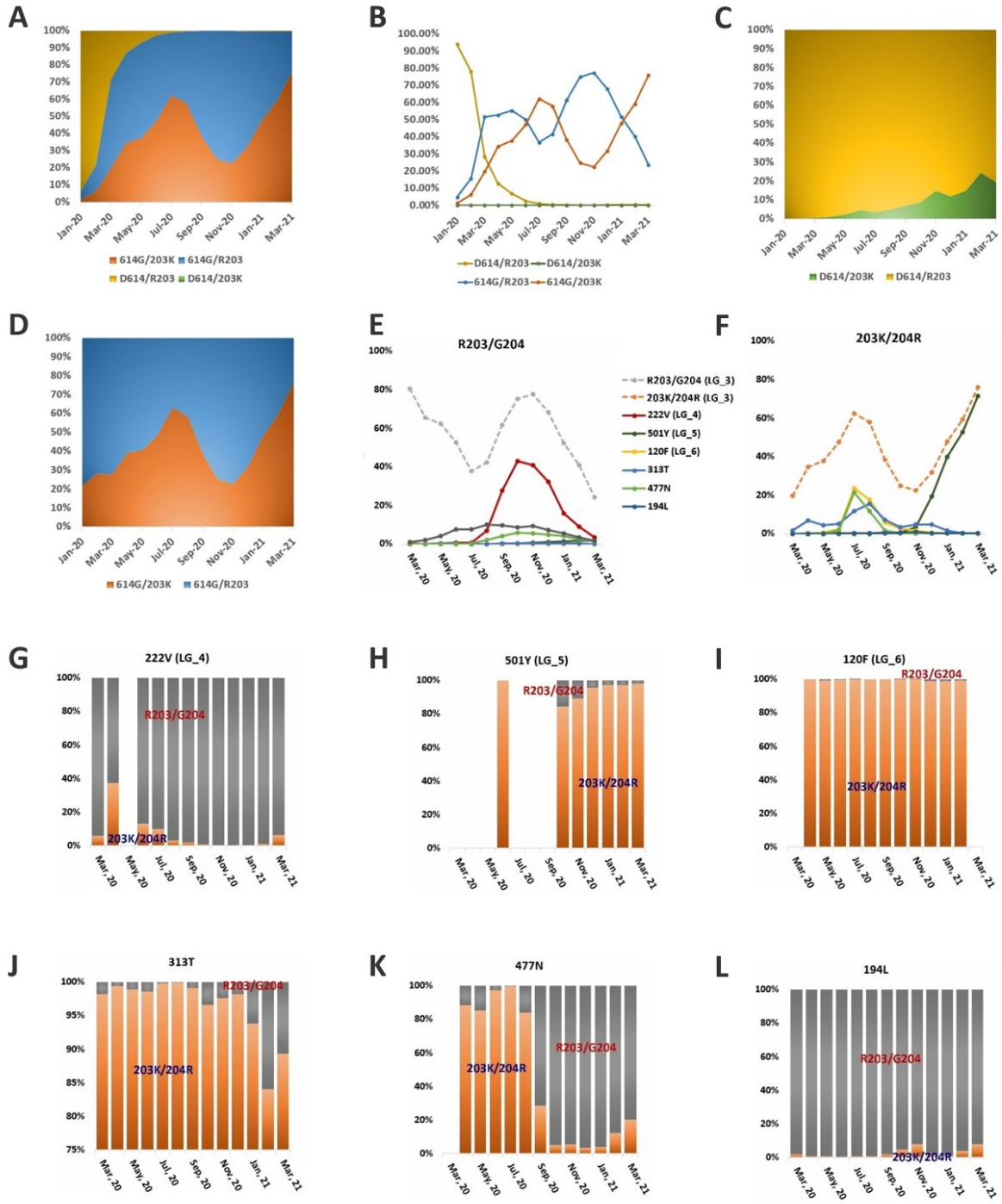

Figure S2. Competition and cooperation of 203K/204R and other mutants, related to Figure 3.

(A-D) The changes in the IFs of four mutation combinations, D614/R203, D614/203K, 614G/R203 and 614G/203K. (A, C and D) are the percentage accumulated area maps. (A) is for four mutation combinations. (C and D) are for D614/R203 vs D614/203K and 614G/R203 vs 614G/203K, respectively. (B) is the line chart of the IFs according to (A). (E-L) are the comparison of IF tracks for the six mutants with a significant correlation with LG_3. (E and F) are the IFs of the six mutants (real lines) in the R203/G204 and 203K/204R variants, respectively. In (E) and (F), dotted lines denote the changes in the IFs of the R203/G204 and 203K/204R variants, respectively. (G-L) are the ratios of R203/G204 and 203K/204R in the six mutants in different months.

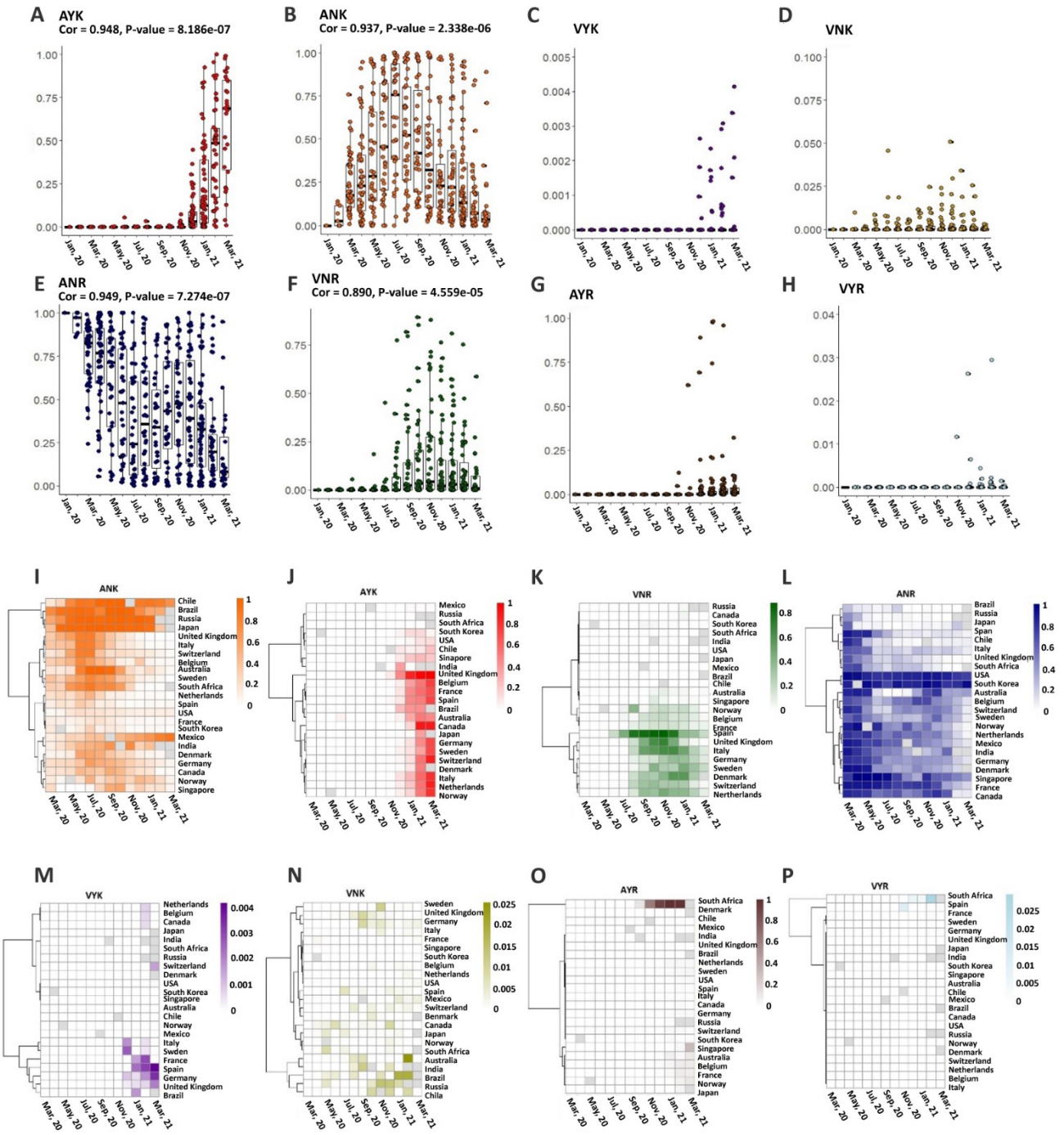

Figure S3. IFs of lineages in different countries, related to Figure 3.

(A-H) The IFs in different countries for the 8 lineages shown in Figure 3. For the four dominant lineages shown in Figure 3A, we calculated the correlations between the median IFs in different countries and the worldwide IFs shown in Figure 3A. The results are displayed at the top. (I-P) The IF changes in different countries for the eight lineages shown in Figure 3. We did not show the countries without an IF > 0 identified in any month, mostly due to an insufficient number of samples.

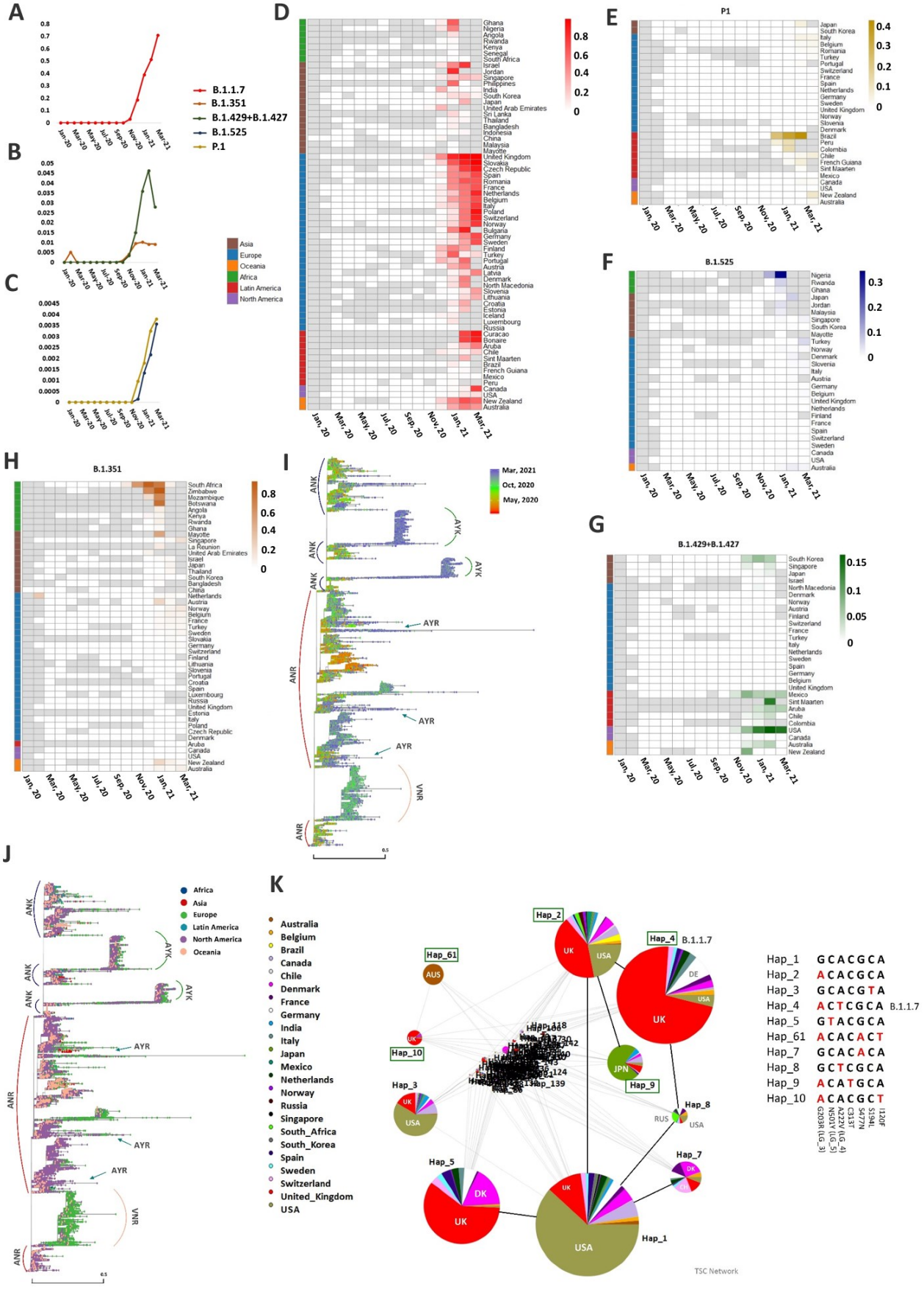

Figure S4. Analyses of R203K/G204R related lineages, related to Figure 3.

(A-H) The changes in the IFs of fast-spreading lineages. (A-C) The changes in the IFs of five lineages around the world. (D-H) The changes in the IFs of lineages in different countries. Continents are marked in the annotation row at the left. (I and J) The distributions of the collection months and continents in the SARS-CoV-2 phylogenetic tree. In (I), the colours of the tip nodes denote the collection months. In (J), the colours of tip nodes denote the collection continents. Lineages are marked. (K) A network of haplotypes constructed from 7 LGs/mutations. The geological positions are differentiated by different colours. The sequences of the haplotypes are shown on the right, with mutants marked in red.

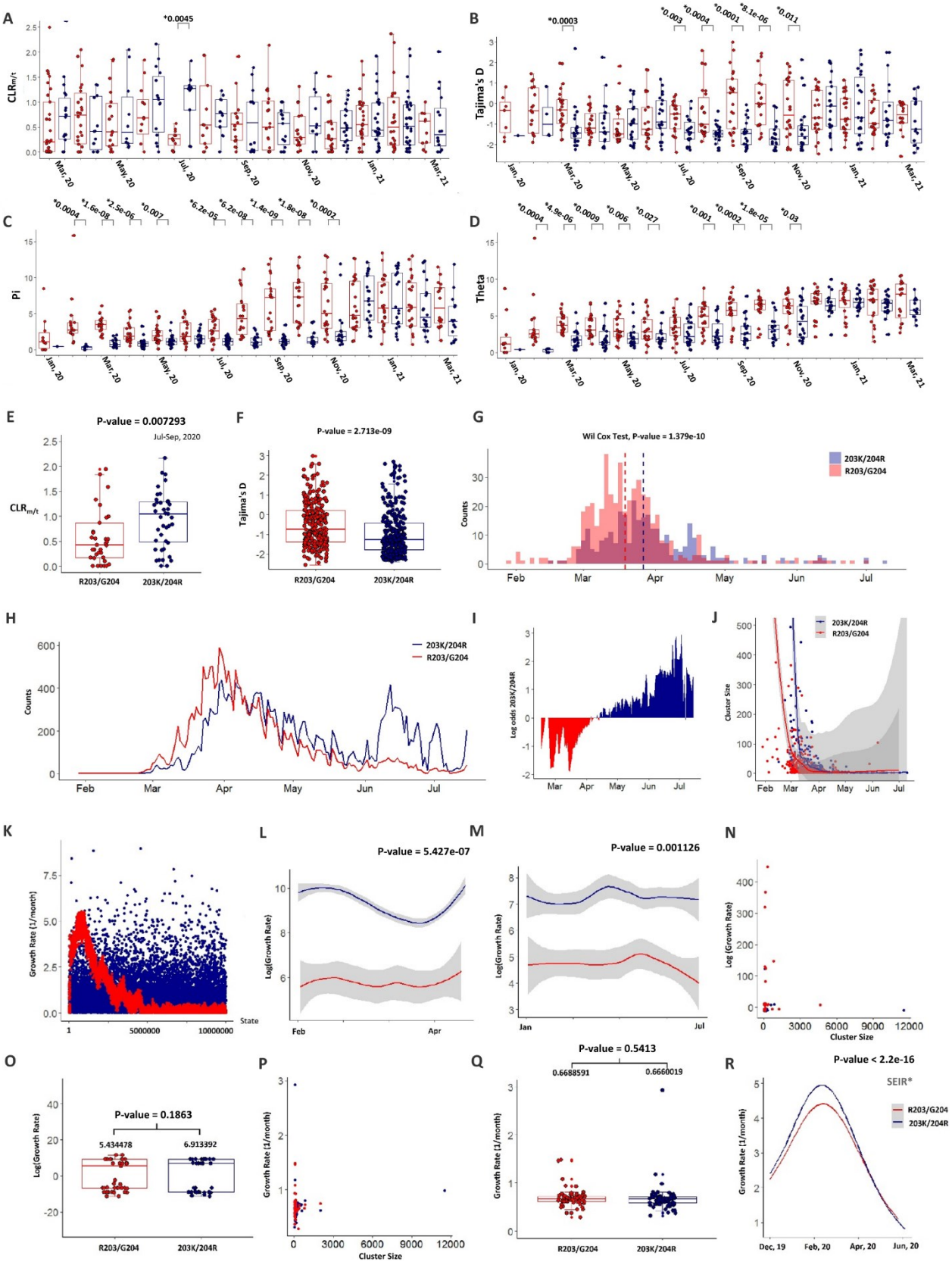

Figure S5. Selection signatures for R203K/G204R, related to Figure 4.

(A) Comparison of CLR $_{m/t}$ between R203/G204 (red) and 203K/204R variants (dark blue) in different months. (B-D)

Comparisons of Tajima's D (B), Pi (C) and Theta (D) in the whole genome of R203/G204 (coloured in red) and 203K/204R strains (coloured in dark blue). Comparisons with statistical significance are marked. (E) Comparison of CLR $_{m/t}$ in countries and in the months from July to September between R203/G204 (red) and 203K/204R variants (dark blue). (F) Comparison of Tajima's D calculated by month and country between R203/G204 (red) and 203K/204R viruses (dark blue). (G-J) The temporal distribution of R203/G204 and 203K/204R phylogenetic clusters in the United Kingdom. (G) Counts of UK R203/G204 and 203K/204R clusters when first detected and over time. (H) Numbers of collected R203/G204 or 203K/204R samples over time. (I) Log odds of the frequency of R203K/G204R over time. (J) Relationship between cluster size (Y-axis) and the date when the first sample was collected within a cluster (X-axis). (K-R) Comparison of the phylodynamic growth rates between R203/G204 and 203K/204R variants. (K) The growth rates of cocirculating clusters across states in a logistic growth model. (L and M) are the growth rates (logged in the plot) over time simulated in the skygrowth coalescent model for cocirculating clusters and all clusters, respectively. (N) The median growth rates of clusters (logged in the plot, Y-axis) versus cluster sizes (X-axis). (O) Comparison of the median growth rates of clusters between R203/G204 and 203K/204R. The legends in (P) follow (N) but without log transformation. Those in (Q) follow (O). (N and O) are the simulation results of the skygrowth coalescent model. (P and Q) are the simulation results of the logistic growth model. (R) is the fitted curves of the growth rates from the phylodynamic SEIR model. Both genetic and phylogenetic data are used (SEIR*).

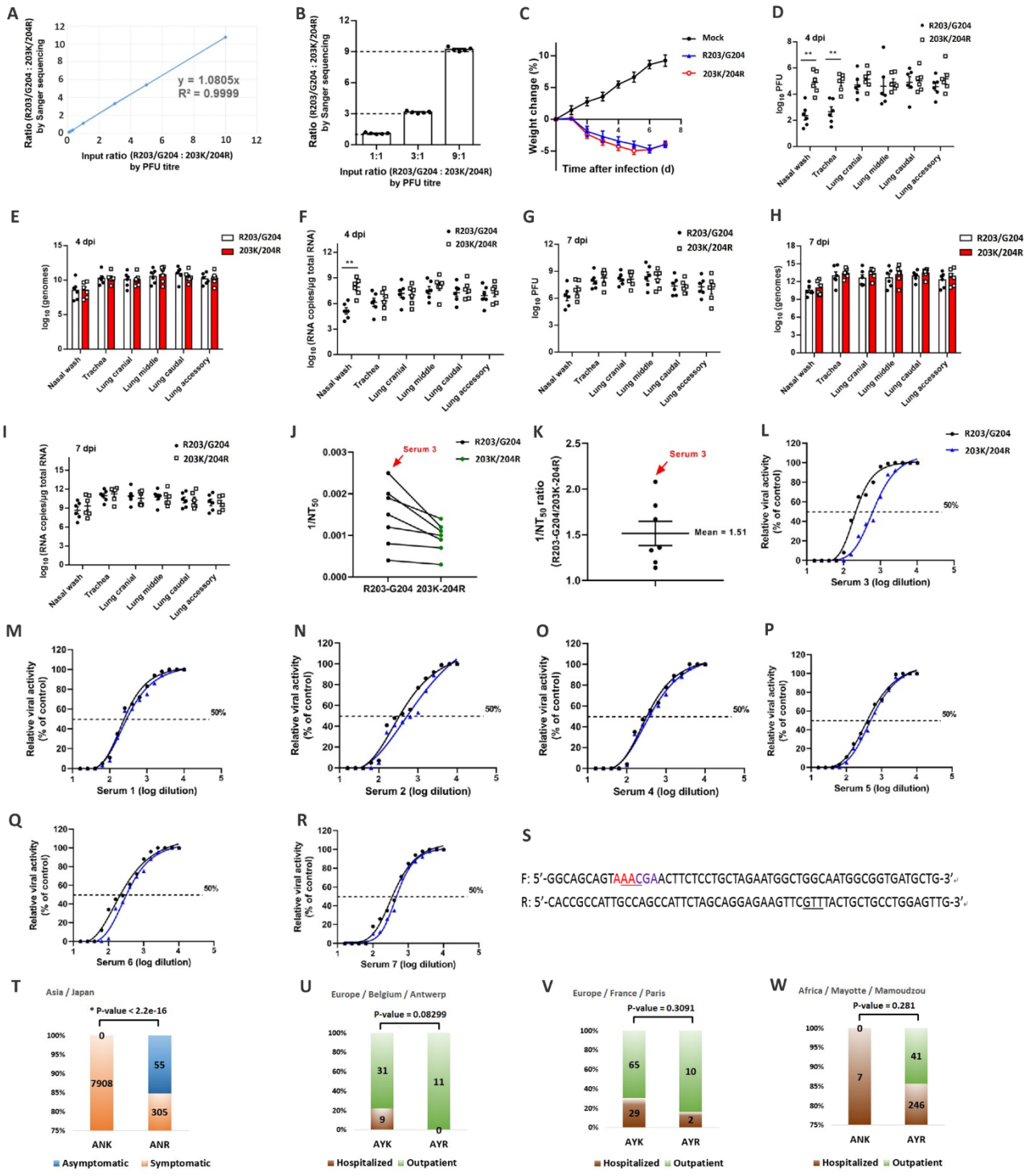

Figure S6. Experiments and analysis results focusing on the fitness, infectivity and virulence of R203K/G204R, related to Figures 4-6.

(A and B) show the validation of competition assay. (A) The correlation between input PFU ratios and output RT–PCR amplicon ratios. The relative amounts of R203/G204 and 203K/204R RNAs was quantified by RT–PCR and Sanger sequencing. R203/G204 and 203K/204R viruses were mixed at PFU ratios of 1:1, 1:3, 1:5, 1:10, 3:1, 5:1 and 10:1. To quantify R203G204/203K204R ratios, RT-PCR products were amplified from extracted RNA and submitted to Sanger sequencing. Data were analyzed by linear regression with coefficient of determination (R2). (B)

Verification of the actual ratios of R203/G204: 203K/204R achieved upon viral mixing. The mixing procedures were repeated independently for five times. The samples were collected immediately after mixing. The results showed that the R203/G204: 203K/204R ratios of intended 1:1, 3:1 and 9:1 were actually 1.06±0.04: 1, 3.21±0.13: 1 and 9.48±0.36: 1, respectively. (C-F) show an independent repeat of hamster experiments infected with R203/G204 and 203K/204R viruses. (C-F) In each independent experiment, 12 hamsters received wild-type R203/G204 virus, 12 received mutant 203K/204R virus, 12 received a 1:1 mixture of R203/G204 and 203K/204R virus, and 12 received PBS (Mock). Weight loss (C) was monitored for 7 dpi. Data are presented as mean ± s.e.m.; n = 12 (all cohorts) at days 0–4; n = 6 (all cohorts) at days 5–7. Weight loss was analysed by two-factor analysis of variance (ANOVA) with Tukey's post hoc test. Infectious titres (D) and amounts of viral genomes (E) were quantified in nasal wash, trachea and lung samples on the 4th dpi. Dots represent individual hamsters (n = 6). The E sgRNA loads (F) at 4 dpi were calculated as a measurement of infectivity. Dots represent individual hamsters (n = 6). Data are presented as the mean ± s.e.m.. **, $p<0.01$. (G-I) are the results of hamster experiments infected with R203/G204 and 203K/204R viruses at 7 dpi. Infectious titres (G) and amounts of viral genomes (H) were quantified in nasal wash, trachea and lung samples on the 7 dpi. Dots represent individual hamsters (n = 6). The E sgRNA loads (I) at 7 dpi were calculated as a measurement of infectivity. Dots represent individual hamsters (n = 6). Data are presented as the mean ± s.e.m.. (J-R) show the neutralizing activities of hamster sera against R203/G204 and 203K/204R mNeonGreen SARS-CoV-2 viruses. (J) Neutralizing activities of hamster sera against R203/G204 and 203K/204R viruses with a mNeonGreen reporter. $1/NT_{50}$ values were plotted. Symbols represent sera from individual hamsters. (K) Ratio of $1/NT_{50}$ between R203/G204 and 203K/204R viruses. Symbols represent sera from individual hamsters. (L-R) Neutralization curves of serum from individual hamsters. The solid line represents the fitted curve, and the dotted line indicates 50% viral inhibition. Data in (K) are represented as the mean ± s.e.m.. (S) The primers used for overlap-extension PCR in the generation of the R203K/G204R mutant virus. Underlined nucleotides are nucleotide substitutions. Colored nucleotides denote the reading frame (R203K/G204R mutation). (T-W) are the prediction of the clinical outcomes of ANK, ANR, AYK and AYR infections based on clinical data collected on a small scale. Legends follow Figure 6.

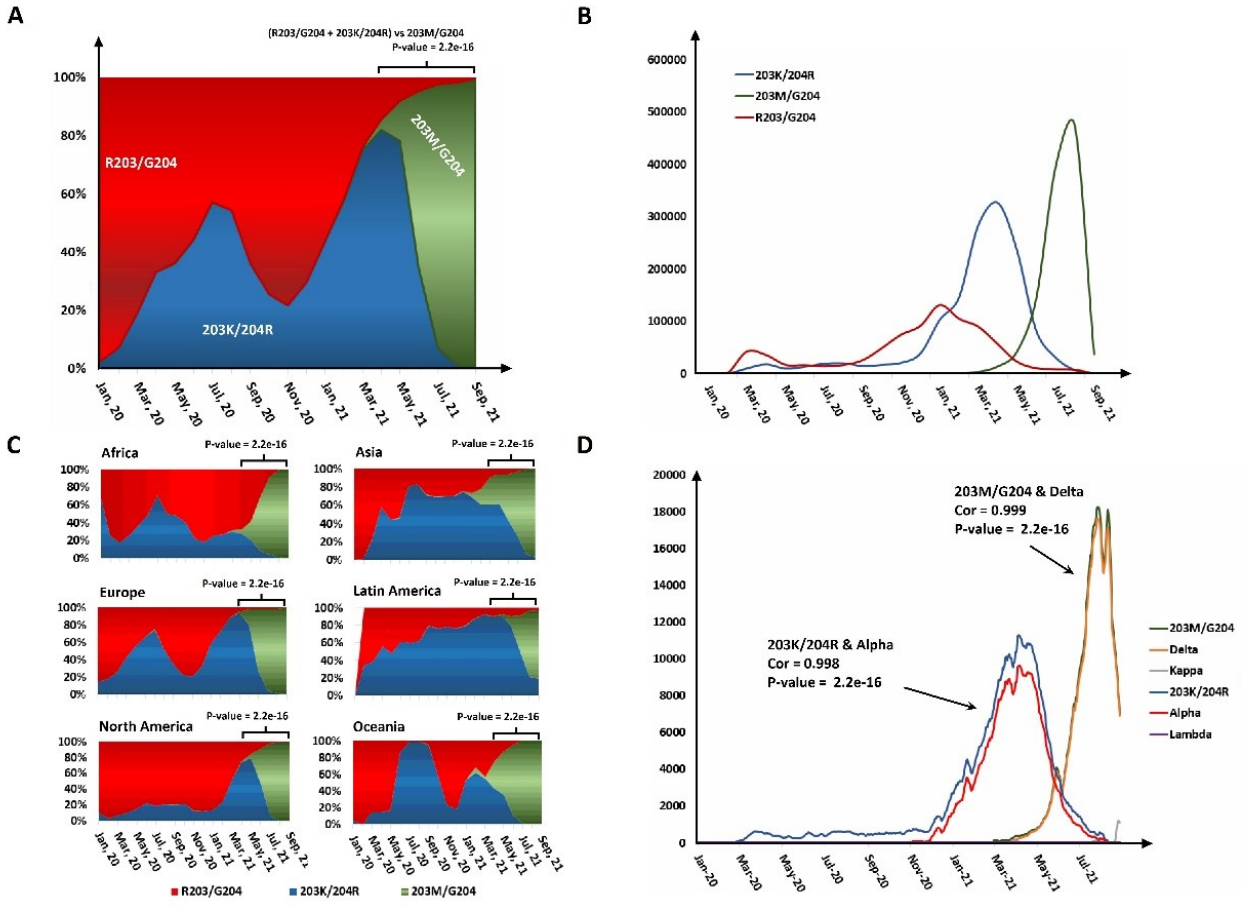

Figure S7. The IF change of mutations in 28881 to 28883, related to Figure 2.

(A) The IF change of R203/G204, 203K/204R and 203M/G204 in the world from January, 2020 to September 2021. (B) The change of the counts of the three alleles in the world up to date. (C) The IF change of the three alleles in different continents. (D) the change of the counts of 203K/204R and 203M/G204 and four main lineages with one of the two mutations. In (A and C), Fisher's exact test was performed on the fraction of R203/G204 + 203K/204R and 203M/G204 in the beginning and the recent months of emergence of R203M.