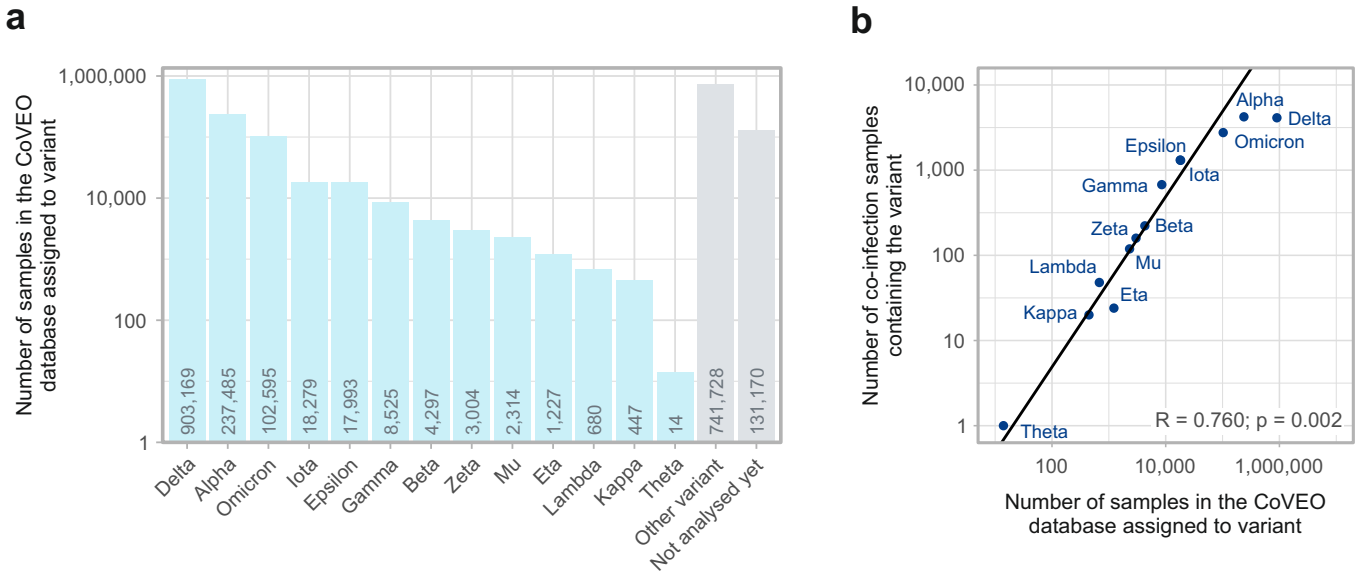
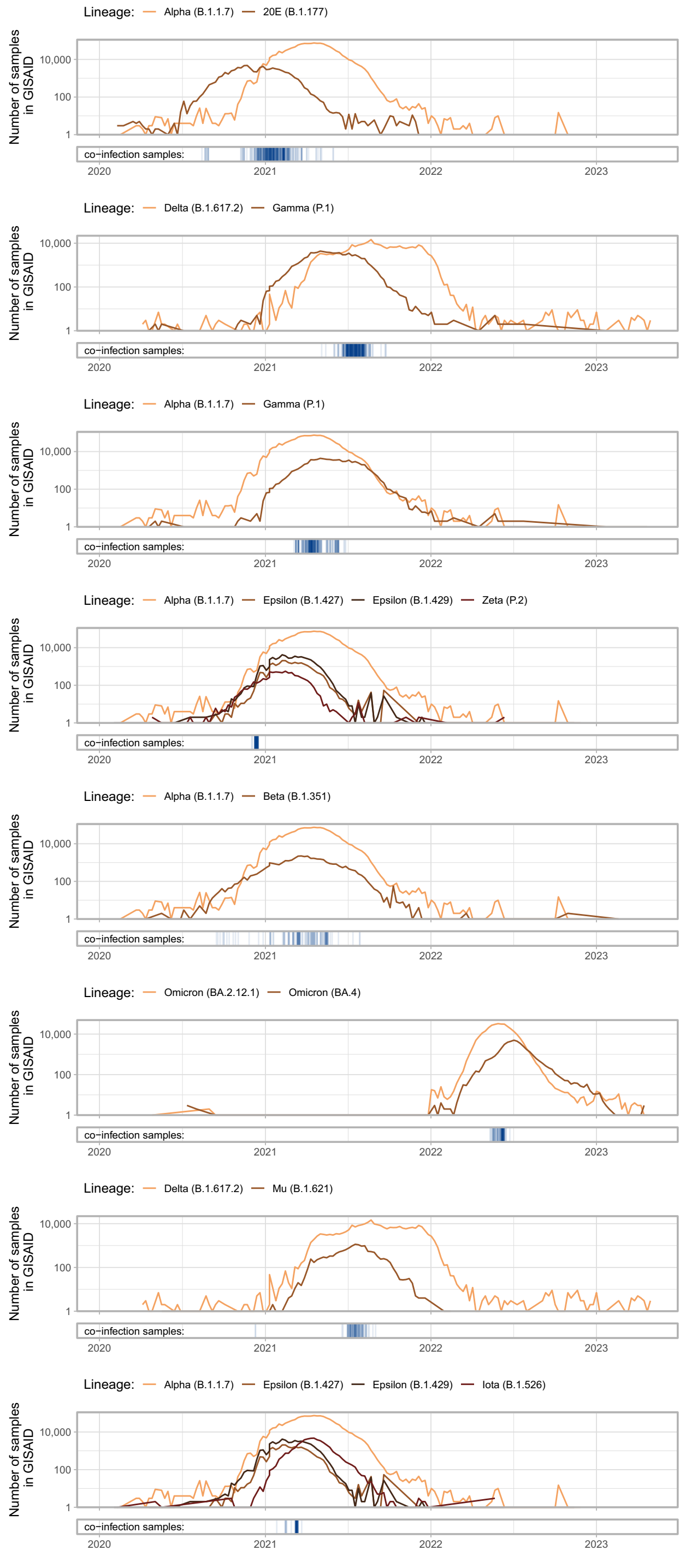
Supplementary Figure 1. The number of samples in the CoVEO database. **a.** The number of good-quality SARS-CoV-2 samples with a human host assigned to different variants in the CoVEO database. **b.** The relationship between the total number of samples assigned to a specific variant and the number of co-infection samples containing the variant. The straight black line represents a linear dependence with a slope of 1 on a log-log graph. R represents the Pearson-correlation coefficient, and the corresponding p-value is derived from a two-sided t-test (n=13 variants).

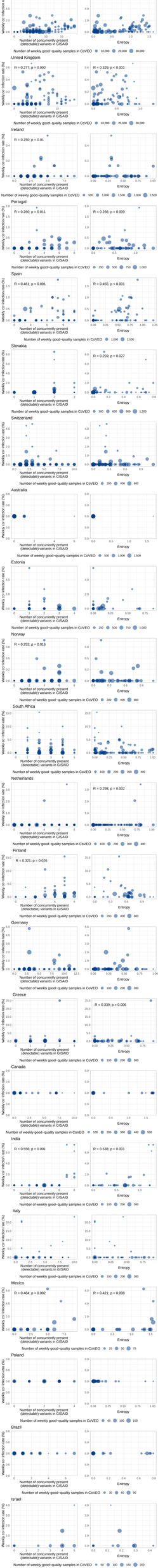


Supplementary Figure 2. Temporal distribution of co-infection samples for variant combinations with more than 50 samples. (The same figures for the top 4 most abundant combinations are shown in Figure 1c of the main manuscript.) Prevalence curves indicate the number of GISAID samples assigned to the respective variants (binned weekly). Blue vertical lines on the bottom panels mark the collection date of co-infection samples of the given variants.

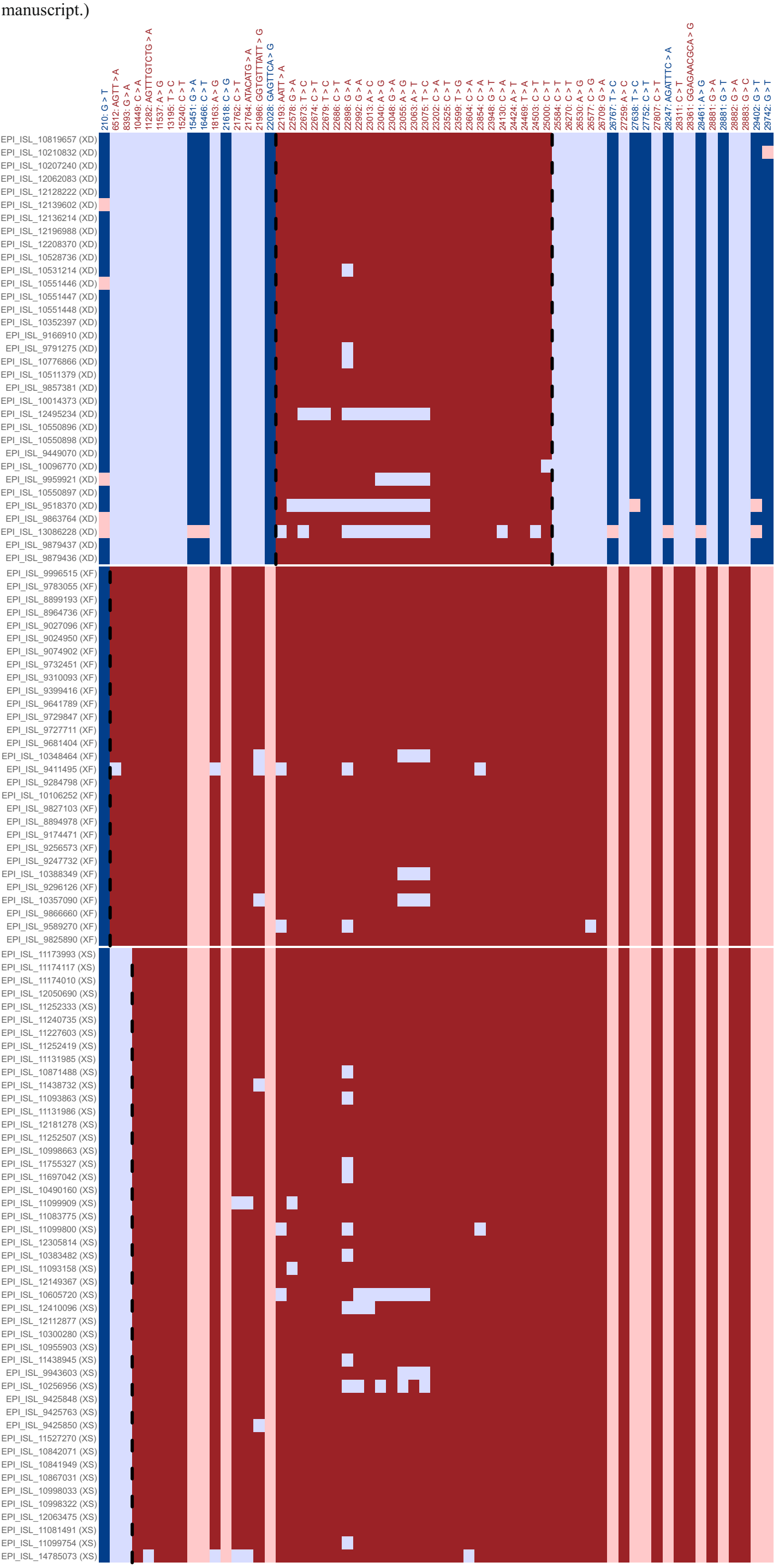


Supplementary Figure 3. Temporal distribution of co-infection samples for countries with at least 1,000 good-quality samples in the database. The top panel on each subfigure depicts the number of samples in the CoVEO database collected each week from the appropriate country. The second panel shows the collection date of co-infection samples identified in the given country. The third panel demonstrates the country-wise weekly prevalence curves of variants in the GISAID database that were included in the study. Coloured curves correspond to the most abundant variants in the whole set of detected co-infection samples (Delta, Omicron (BA.1), Alpha, Iota, Epsilon, 20E), while different shades of grey demonstrate the remaining lineages. For the United States and the United Kingdom, on the fourth panel various measures (the number of concurrently circulating lineages, the cumulative number of lineages and information entropy) of genetic diversity are plotted based on GISAID data from the given country.

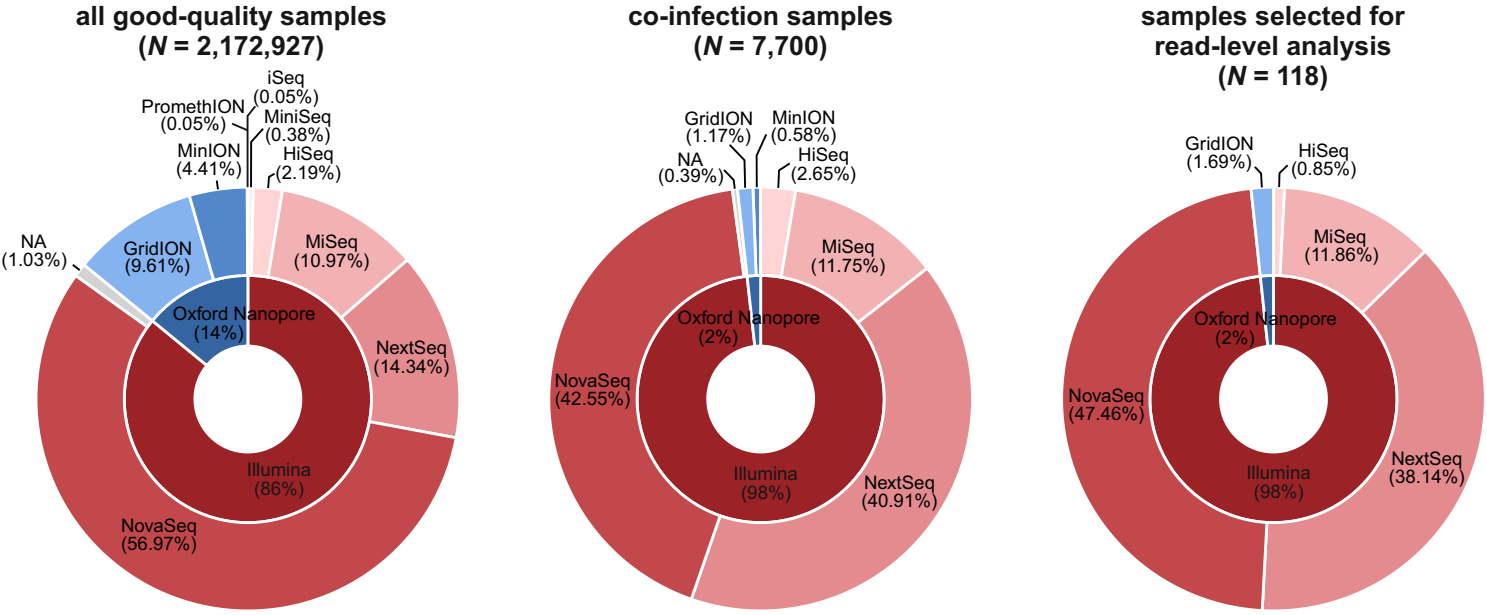
Supplementary Figure 4. Weekly co-infection rate in the function of genetic diversity. Weekly co-infection rate was calculated as the percentage of co-infection samples out of all good-quality samples in the given country, in the given week. Genetic diversity was either defined as the number of lineages concurrently present in the given country in the given week in GISAID data (left panels) or the information entropy (right panels, see Methods). Only lineages investigated by the present study were used. Weeks for which the number of good-quality samples was smaller in the CoVEO database did not reach 10 were discarded. Each marker represents the data for a single week. Marker size corresponds to the number of good-quality samples available. "R" and "p" indicate Pearson-correlation coefficients and respective two-sided t-test p-values (non-significant ($p \geq 0.05$) results are not displayed).



Supplementary Figure 5. Mutually exclusive defining mutations in GISAID samples assigned to the XD, XF and XS 'Deltacron' (recombinant Delta and Omicron) lineages. Samples (rows) are listed by their GISAID ID, with their Pango lineage indicated in brackets. Defining mutations (columns) are colored by their respective variant (blue for Delta and red for Omicron). Heatmap colors show whether the given sample contains the given mutation or not (dark blue: present Delta-mutation; light blue: missing Omicron-mutation; dark red: present Omicron-mutation; light red: missing Delta-mutation). Dashed vertical lines mark the recombination breakpoint range(s) for each lineage. (The same genomic regions are indicated with red shaded areas in Figure 4 of the main manuscript.)



Supplementary Figure 6. Distribution of sequencing platforms and instruments within all good-quality samples included in the study, for co-infection samples and for samples selected for downstream read-level analysis.



Supplementary Figure 7. Schematic diagram of the workflow used to produce the data analysed in the study. Only those steps of the VEO variant calling pipeline (github.com/enasequence/covid-sequence-analysis-workflow) and features of the COVID-19 Data Portal (https://www.covid19dataportal.org/) are shown that are relevant to the current analysis. More details on both can be found in Rahman et al. (2023). For the CoVEO database, only those tables and fields are displayed that were queried during data processing. Queries, codes, data files and vizualizations are uploaded to the csabaiBio/SARSCoV2-coinf github repository. Postgres, PostgreSQL and the Slonik Logo are trademarks or registered trademarks of the PostgreSQL Community Association of Canada and used with their permission.