

Supplementary information

SARS-CoV-2 variants: biology, pathogenicity, immunity and control

**In the format provided by
the authors and unedited**

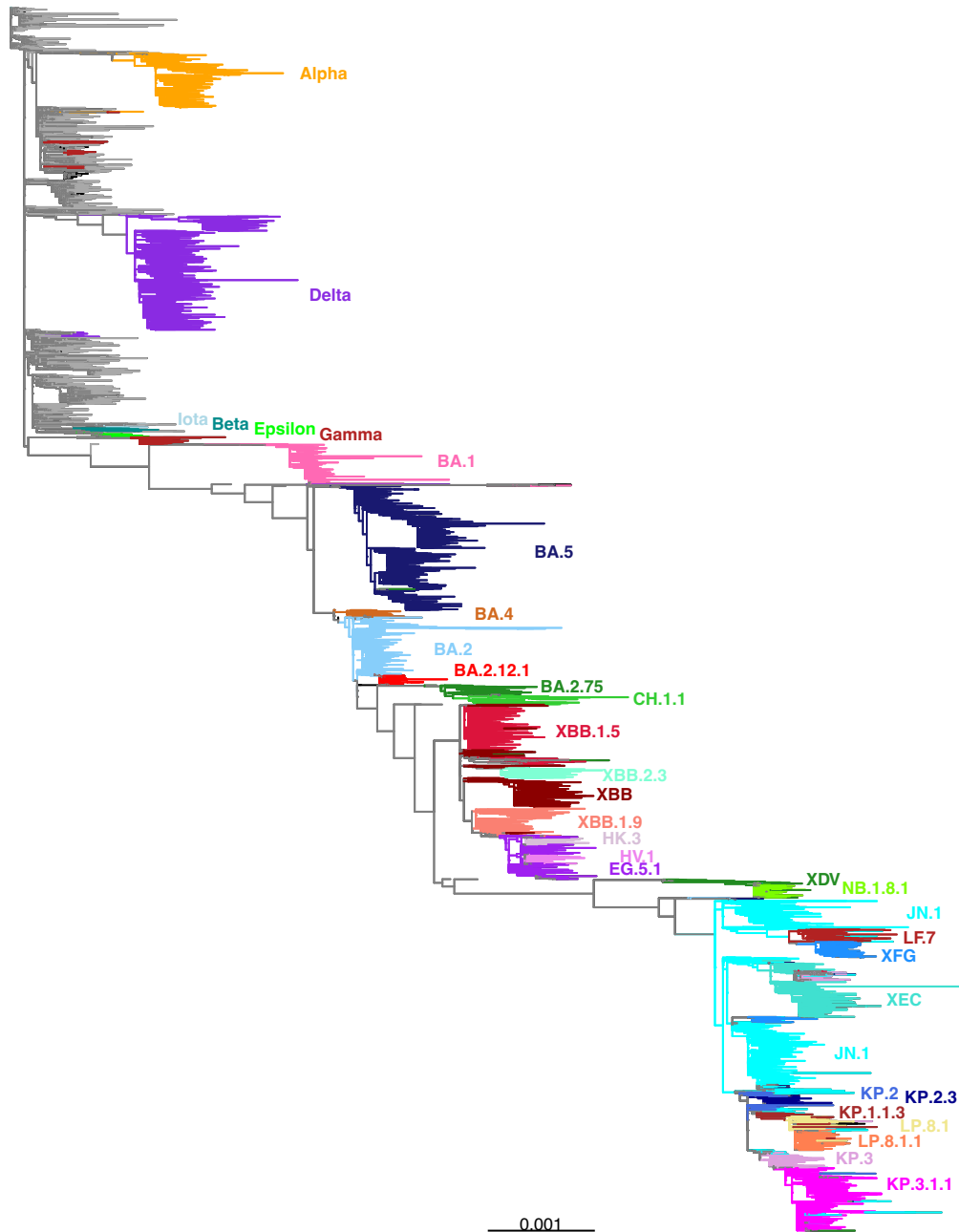


Fig. S1. Phylogenetic tree illustrating the relationships between Spike protein sequences in major lineages sampled during the pandemic over time. To construct the data set, 25 sequences per week were randomly sampled from GISAID^{1,2} starting in Dec 2019 and continuing through July 26, 2025, for a total of 7,249 sequences. The colors match those used in Figure 1, except that the early -pre-alpha variants are all labeled in grey. Note the particularly long ancestral branch lengths leading into the BA.1, BA.2, and JN.1 lineages. XBB was a recombinant lineage that arose from two distinct BA.2 sub-lineages; note, recombined lineages have unreliable placement within a phylogenetic tree³. The lineage that is currently increasing the fastest (as of July, 2025) is XFG, a recombinant of sequences of the LF.7 and LP.8.1.2 parental lineages. The tree was based on a DNA

alignment of the full-length genome built using FastTree^{4,5}. The tree was inferred using FastTree [Version 2.1.11 Double precision (No SSE3), OpenMP (192 threads)] with arguments "-nt -gtr -fastest" (nucleotide data, Generalized Time-Reversible ML model, rate categories using the CAT approximation with 20 rate categories). Inferred ML frequencies and rates were as follows: GTR Frequencies, 0.2992 0.1827 0.1958 0.3223; and GTR rates(ac ag at cg ct gt), 0.4554 1.5520 0.2486 0.3287 4.4822 1.0000

WHO	Pango Lineage	Signal	Pept	NTD										RBD										class- age																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
				Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ

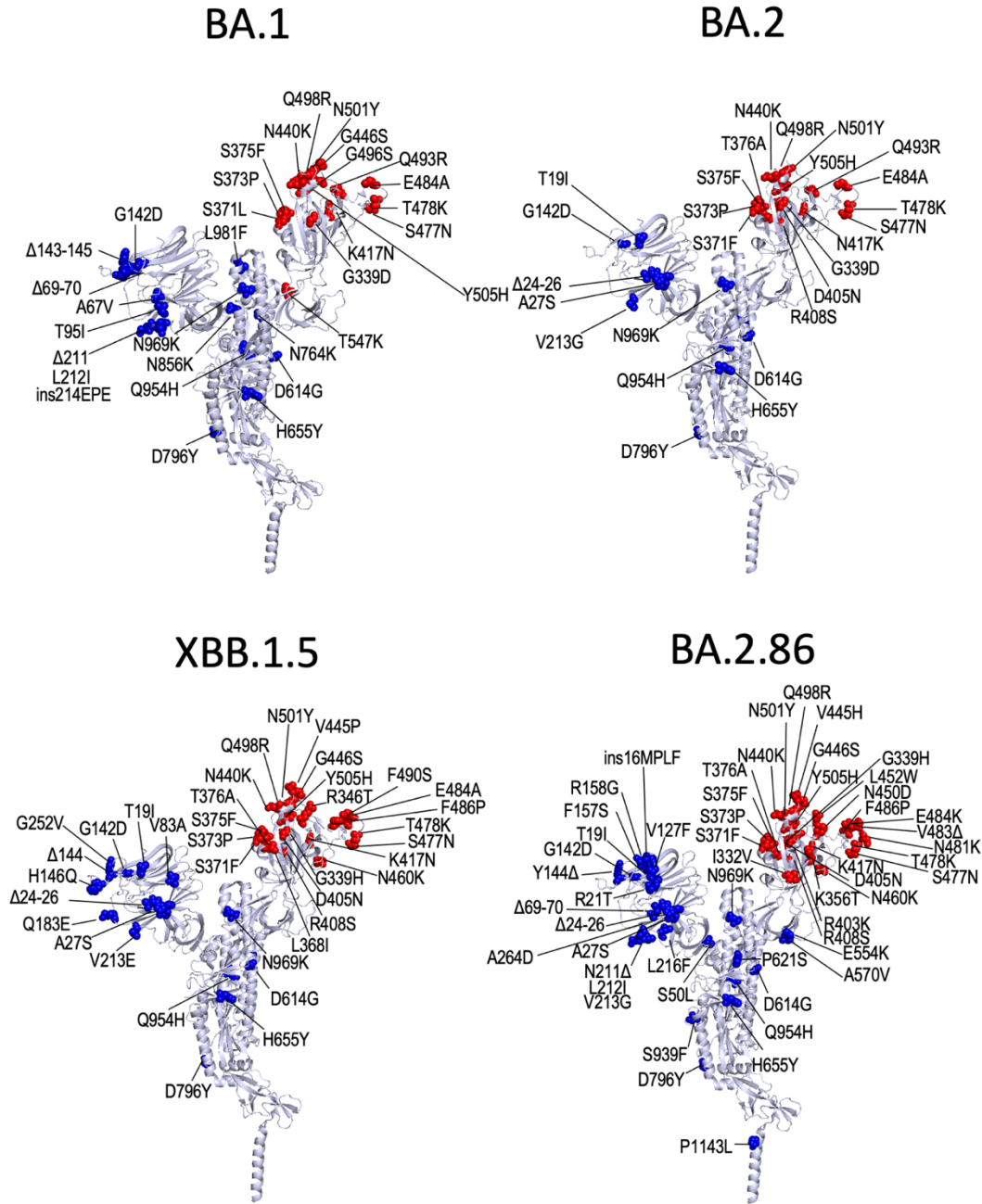
B

Fig. S2. Expanded version of Fig. 2AB. (A) All of the Spike mutations of the major variants are shown relative to the original ancestral variant from Wuhan. The coloring and use of symbols is the same as in Fig.2B. **(B)** Structural mapping of the lineage defining Spike mutations. The red and blue mutations are in the RBD and non-RBD regions, respectively.

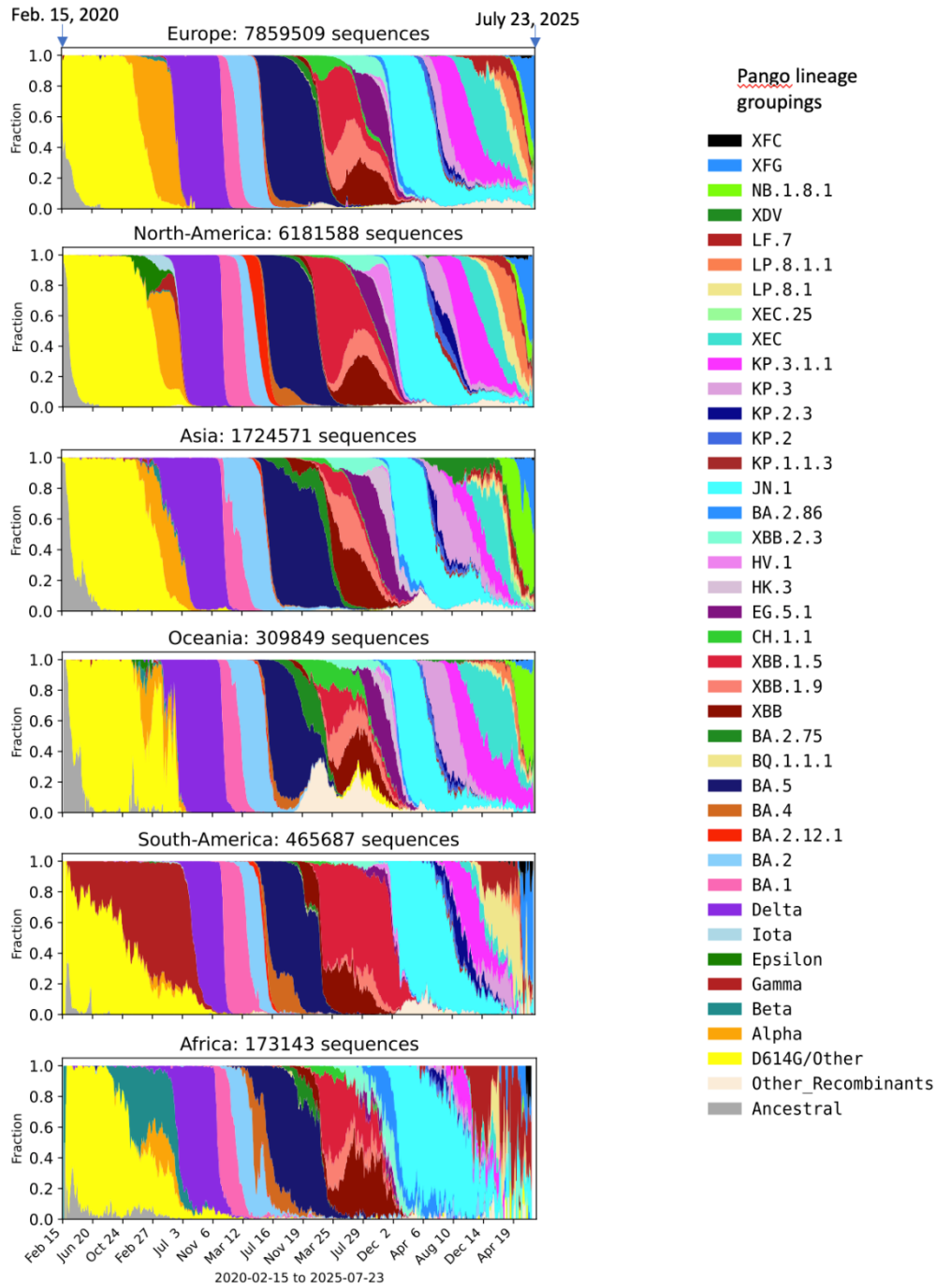


Fig. S3. Variant transitions over time by continent, following the global variant transitions shown in Fig. 1A. The graphs are based on all available GISAID^{1,2} entries and visualized using the “Embers” tool at <https://cov.lanl.gov>. Variants sometimes had very distinctive patterns in different geographic regions, for example, Gamma’s high prevalence in South America (brick red) and Beta’s in Africa (teal) early in the pandemic. Recombinants that became major global lineages are indicated, XBB (and its sub-lineages HV.1 and HK.3) became globally dominant prior to the JN.1 expansion and originally arose from two BA.2 variants. JN.1 major recombinant lineages include: XEC; XDV, which initially expanded in China and eventually gave rise to NB.1.8.1, which then became globally prominent in the spring of 2025; and XFG, is the most rapidly expanding lineage in

the summer of 2025. Other recombinants are grouped together and labeled in beige, and some of these became regionally prominent without spreading globally.

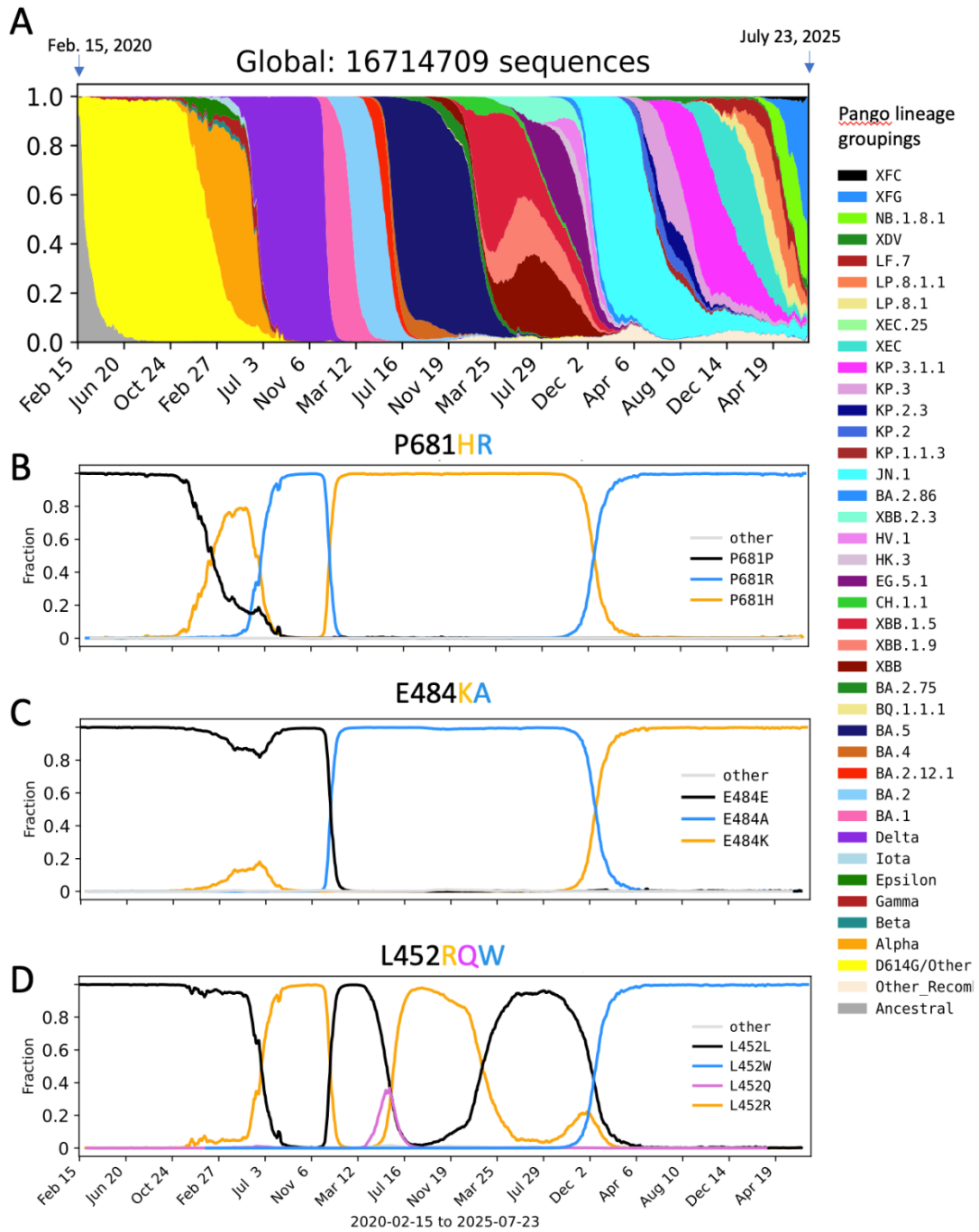


Fig. S4. Three examples of sites showing patterns of recurrent changes associated with major lineage transitions during the pandemic. (A) The top plot shows global transitions in major SARS-CoV-2 lineages by plotting the changes in frequency of the lineages, on the y-axis, throughout the pandemic, similar to Fig. 1A. **(B)** Changes to positively charged P681H/R are associated with increased S1/S1 cleavage and viral fusogenicity⁶ and were found in the early VOCs Alpha and Delta; by the summer of 2021, ancestral P681 was almost never sampled. Site 681 has transitioned several times, with Omicron carrying Alpha's P681H form, and JN.1 re-introducing Delta's 681R. **(C)** E484K was a highly recurrent mutation among pre-Omicron lineages; for example it arose

independently in WHO VOCs Gamma, Delta and Iota (see **Fig. 2A**) and conferred neutralization resistance^{7,8}. Omicron introduced a new substitution, E484A, which stayed as the dominant form for nearly 2 years, until JN.1 carried E484K back into dominance. **(D)** Position 452 mutations can also confer neutralization resistance⁹, and seem to be under strong selective pressure, historically oscillating between ancestral L452 and several other forms (i.e., L452R/Q/W) in different major lineages. The substitution L452R was highly recurrent both pre-Omicron and within Omicron (Fig. 2A), but the L452W change was relatively rare until JN.1 brought it into global dominance, and L452Q was unusual outside of BA.2.12.1. While these events may have been stochastic, the rarer mutations W and Q may have been favored over L452R given epistatic effects in the context of particular Spike proteins, or selective pressure on site may have been impacted by the contemporary immunological context of the human host population at the time it was in circulation. The base mutations that give rise to these amino acid changes are: L (CTG)-to-R (CGG), or L-to-W (TTG), or L-to-Q (CAG), so the recurrent L-to-R substitution is the result of a T-to-G transversion, while the more unusual L-to-Q is the result of a transition. Thus, base mutation rates seem an unlikely explanation for patterns of recurrence in this case. In Parts B–D: black lines indicate the frequency of the ancestral amino acid in a given position over time; blue, orange and magenta lines track the frequency of common amino acid substitutions, as indicated in each legend; and light grey indicates the frequency of all other non-specified amino acids.

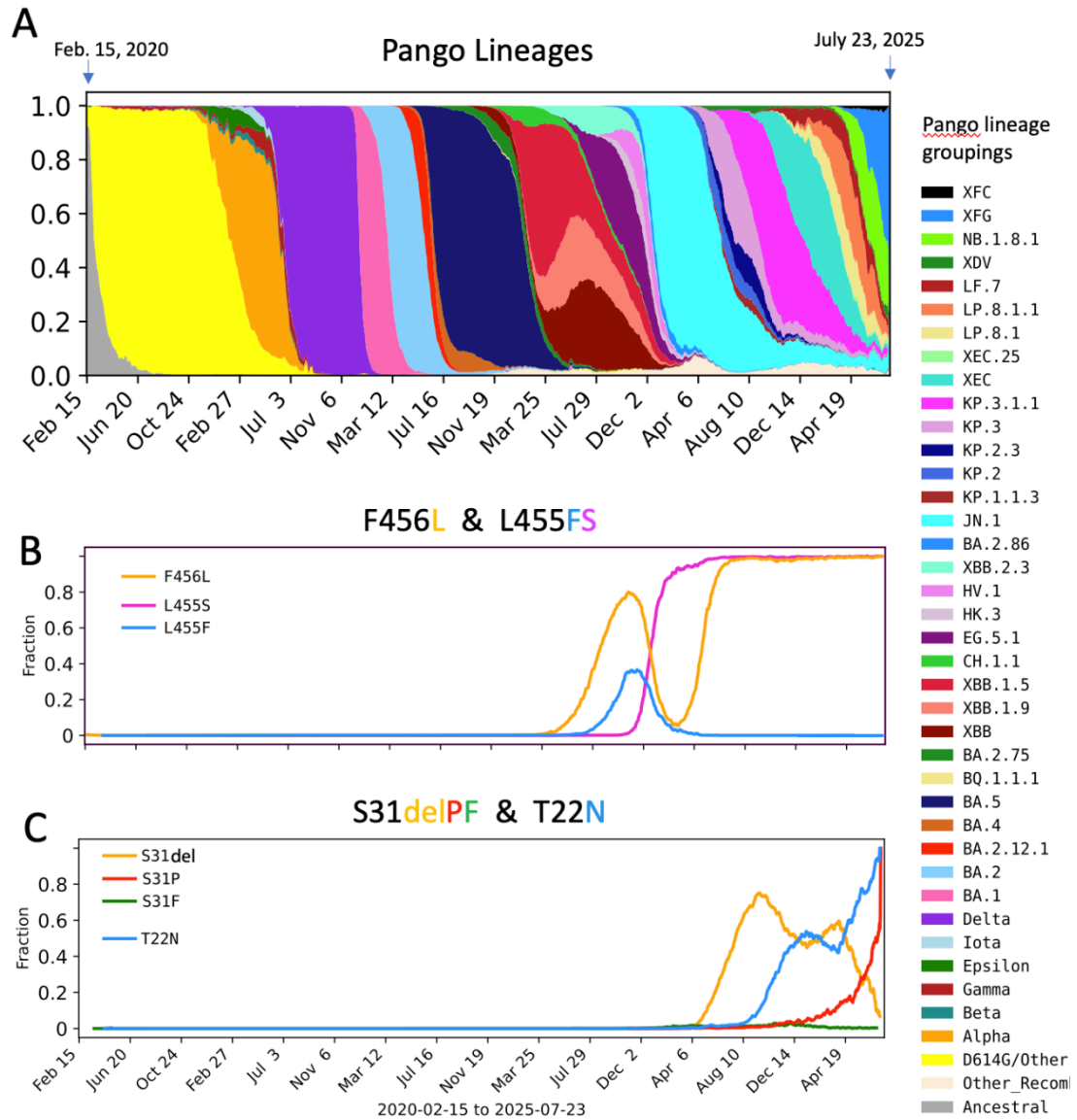


Fig. S5. Two examples of pairs of Spike sites showing patterns of coincident changes that may reflect co-dependence and site interactions. (A) As in Fig. S4, the top plot shows global transitions in major SARS-CoV-2 lineages by plotting the changes in frequency of the lineages on the y-axis, throughout the pandemic. **(B)** Sites 455 and 456 were rarely mutated until XBB lineages first started acquiring F456L, followed by L455F and the FLip pairing, which had strongly coupled impacts on both ACE2 binding and neutralizing resistance¹⁰. BA.2.86 reset both positions to the ancestral form, but JN.1 acquired L455S and led the rapid global expansion of BA.2.86 lineages. The most successful JN.1 sub-lineages then began recurrently introducing F456L (see Fig. 2A). **(C)** The deletion S31del and substitution T22N both create a potential N-linked glycosylation site in the NTD and are associated with immunological resistance in the RBD¹¹. They both were extremely rare prior to the JN.1 lineage expansion, and then both became highly recurrent and under positive

selection, first S31del followed by T22N. Despite each being highly recurrent, at least so far, they are only rarely simultaneously present on the same Spike, suggesting that while one glycan may be beneficial, having both simultaneously may be disfavored. Interestingly, by the end of the period studied (July 23, 2025), variants with the T22N change have become globally dominant accompanied by a S31P substitution. The fact that the sets of mutations in both B and C were very rare early in the pandemic, then suddenly highly recurrent and under apparent selective pressure may reflect epistasis and mutations that are only advantageous in certain Spike contexts, or alternatively these sites may have been impacted by the contemporary immunological context of the human host population at the time they were coming into circulation.

References

- 1 Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33-46 (2017).
<https://doi.org/10.1002/gch2.1018>
- 2 Khare, S. *et al.* GISAID's Role in Pandemic Response. *China CDC Wkly* **3**, 1049-1051 (2021).
<https://doi.org/10.46234/ccdcw2021.255>
- 3 Giorgi, E. E. *et al.* Recombination and low-diversity confound homoplasy-based methods to detect the effect of SARS-CoV-2 mutations on viral transmissibility. *bioRxiv*, 2021.2001.2029.428535 (2021). <https://doi.org/10.1101/2021.01.29.428535>
- 4 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641-1650 (2009).
<https://doi.org/10.1093/molbev/msp077>
- 5 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
<https://doi.org/10.1371/journal.pone.0009490>
- 6 Gellenoncourt, S. *et al.* The Spike-Stabilizing D614G Mutation Interacts with S1/S2 Cleavage Site Mutations To Promote the Infectious Potential of SARS-CoV-2 Variants. *J Virol* **96**, e0130122 (2022). <https://doi.org/10.1128/jvi.01301-22>
- 7 Wang, P. *et al.* Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host Microbe* **29**, 747-751 e744 (2021). <https://doi.org/10.1016/j.chom.2021.04.007>
- 8 Wang, P. *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130-135 (2021). <https://doi.org/10.1038/s41586-021-03398-2>
- 9 Zhang, X. *et al.* SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance. *Signal Transduct Target Ther* **6**, 430 (2021).
<https://doi.org/10.1038/s41392-021-00852-5>
- 10 Jian, F. *et al.* Convergent evolution of SARS-CoV-2 XBB lineages on receptor-binding domain 455-456 synergistically enhances antibody evasion and ACE2 binding. *PLoS Pathog* **19**, e1011868 (2023). <https://doi.org/10.1371/journal.ppat.1011868>
- 11 Wang, Q. *et al.* Antibody evasiveness of SARS-CoV-2 subvariants KP.3.1.1 and XEC. *bioRxiv*, 2024.2011.2017.624037 (2024). <https://doi.org/10.1101/2024.11.17.624037>