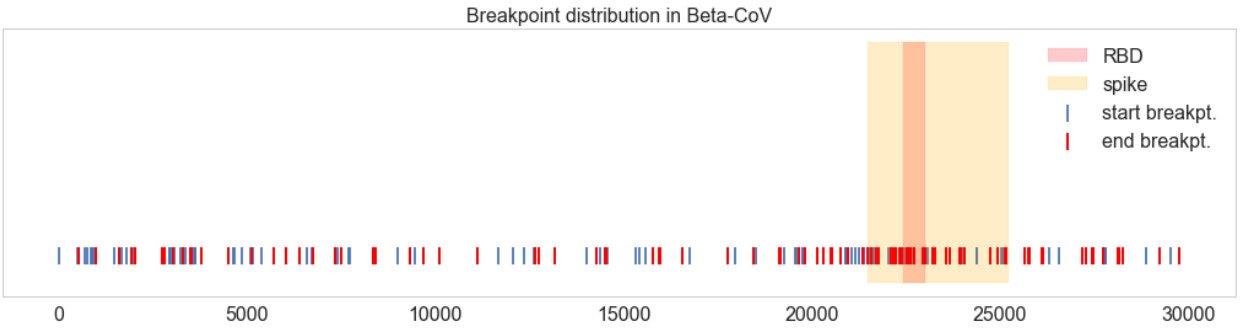
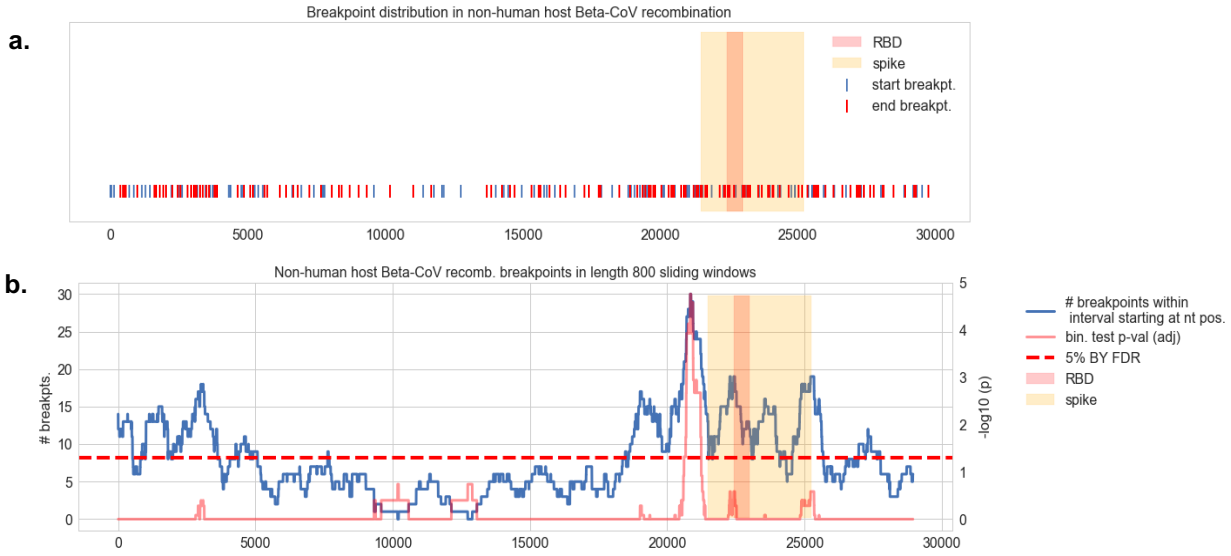
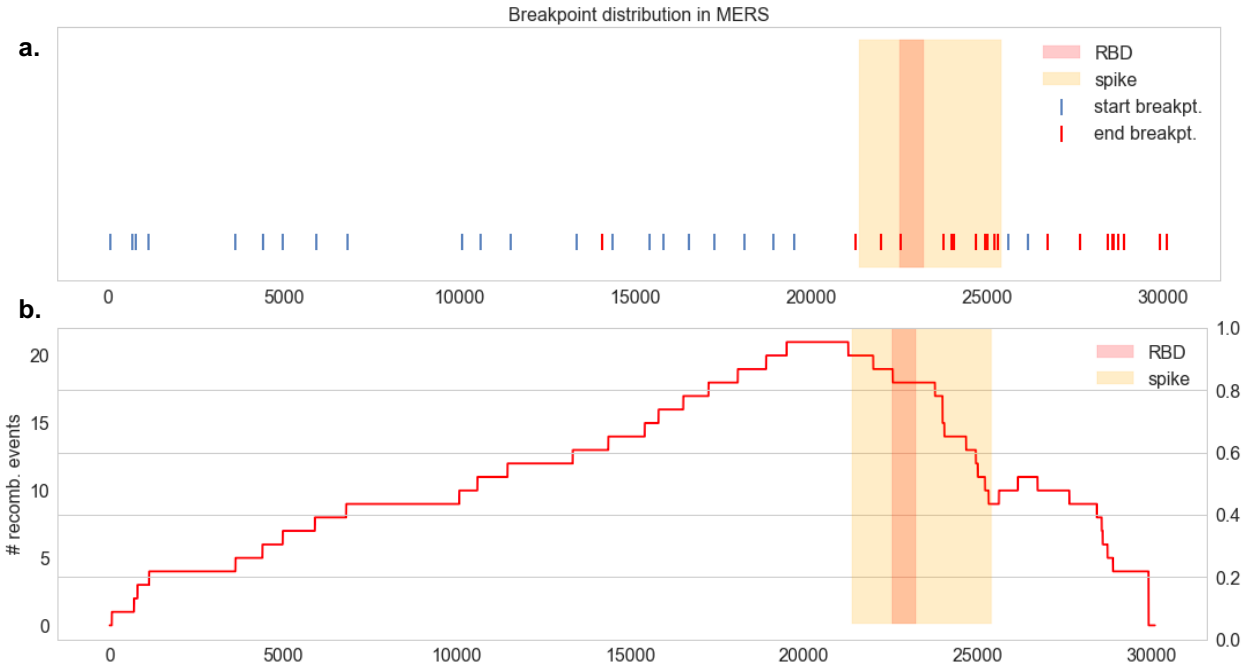


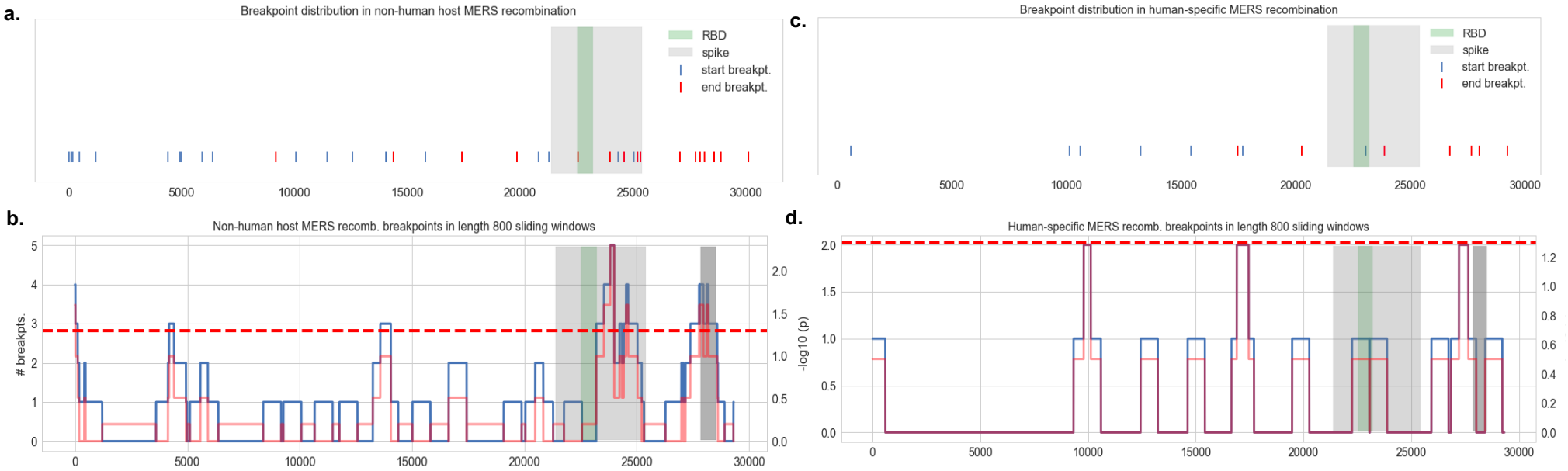
Supplementary Figure 1 | Recombination breakpoints in betacoronaviruses. Start (blue) and end (red) breakpoints from all of the 103 detected recombination events in betacoronaviruses are shown, highlighting the Spike protein and the RBD.



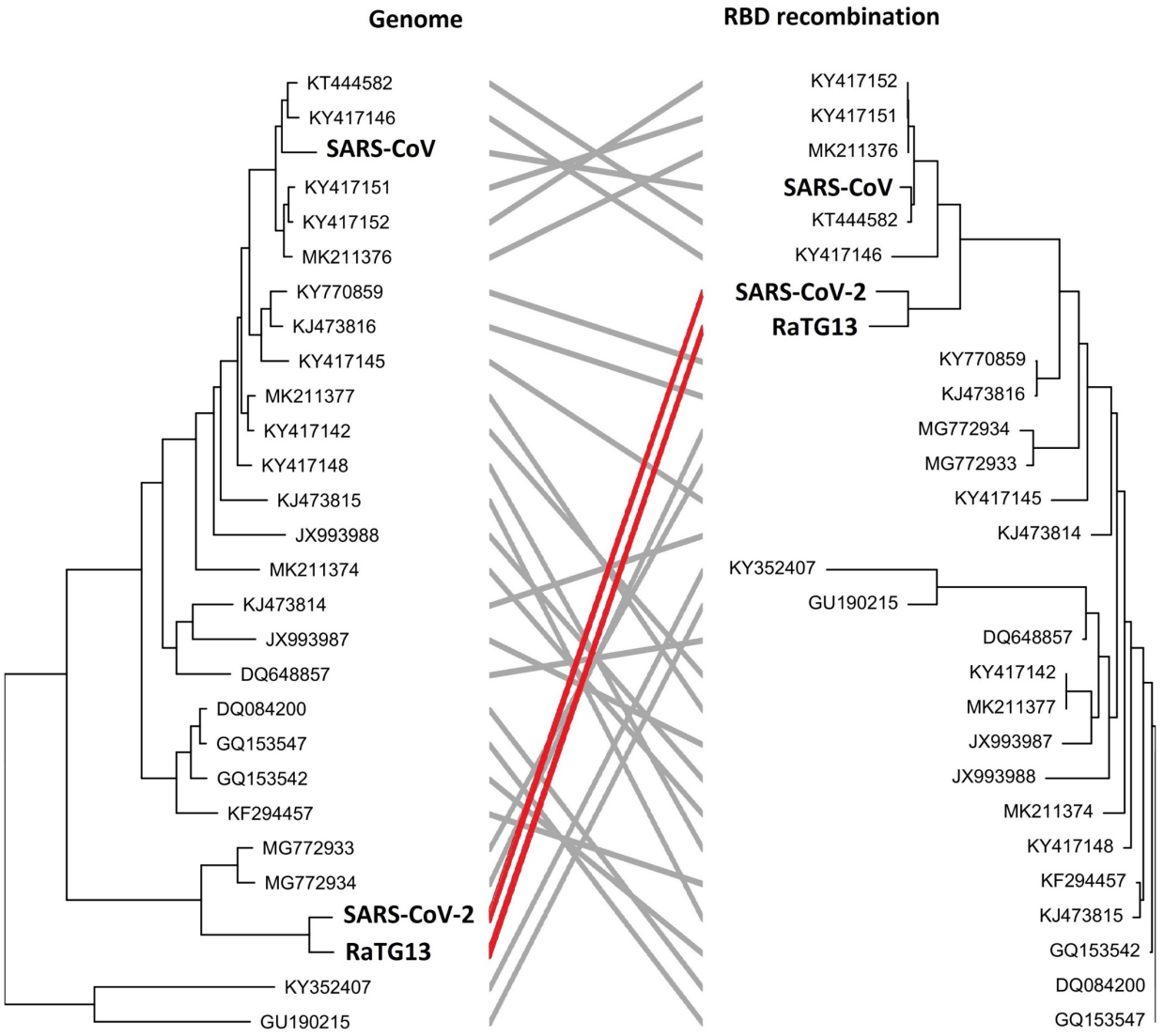
Supplementary Figure 2 | Recombination breakpoints in betacoronaviruses (non-human host only). a. Distribution of start and end breakpoints recombination events detected in betacoronaviruses infecting non-human hosts. The Spike protein and the RBD are highlighted. **b.** Sliding window analysis shows the distribution of recombination breakpoints (either start or end) in 800 nucleotide (nt) length windows upstream (namely, in the 5' to 3' direction) of every nt position along the viral genome (blue curve). BY-adjusted *p*-values are shown in red.



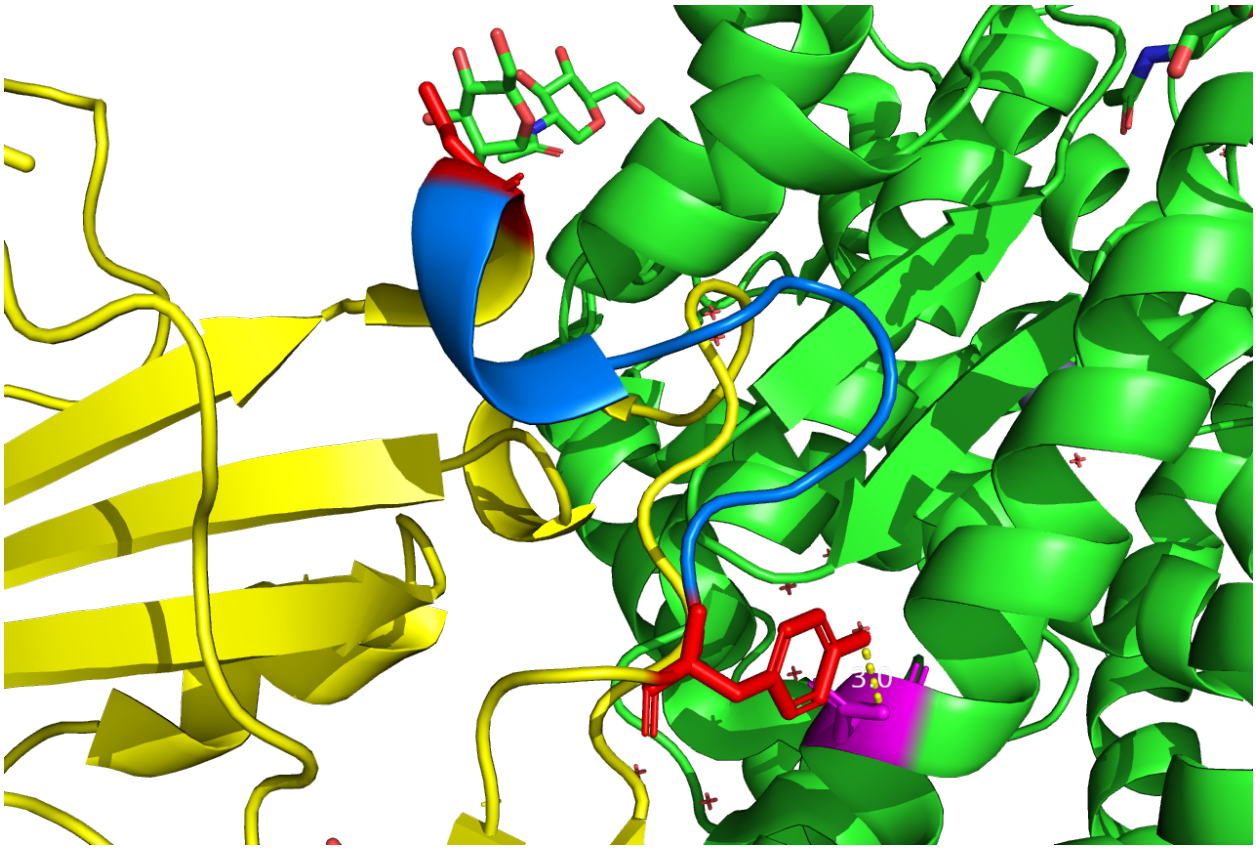
Supplementary Figure 3 | Recombination breakpoints in MERS-CoV. a. Distribution of start and end breakpoints from 24 recombination events detected from MERS-CoV isolates. **b.** Number of recombination events covering each nt position of the MERS-CoV genome. The Spike gene, and the n-terminus in particular, overlaps with the majority of the detected recombination segments.



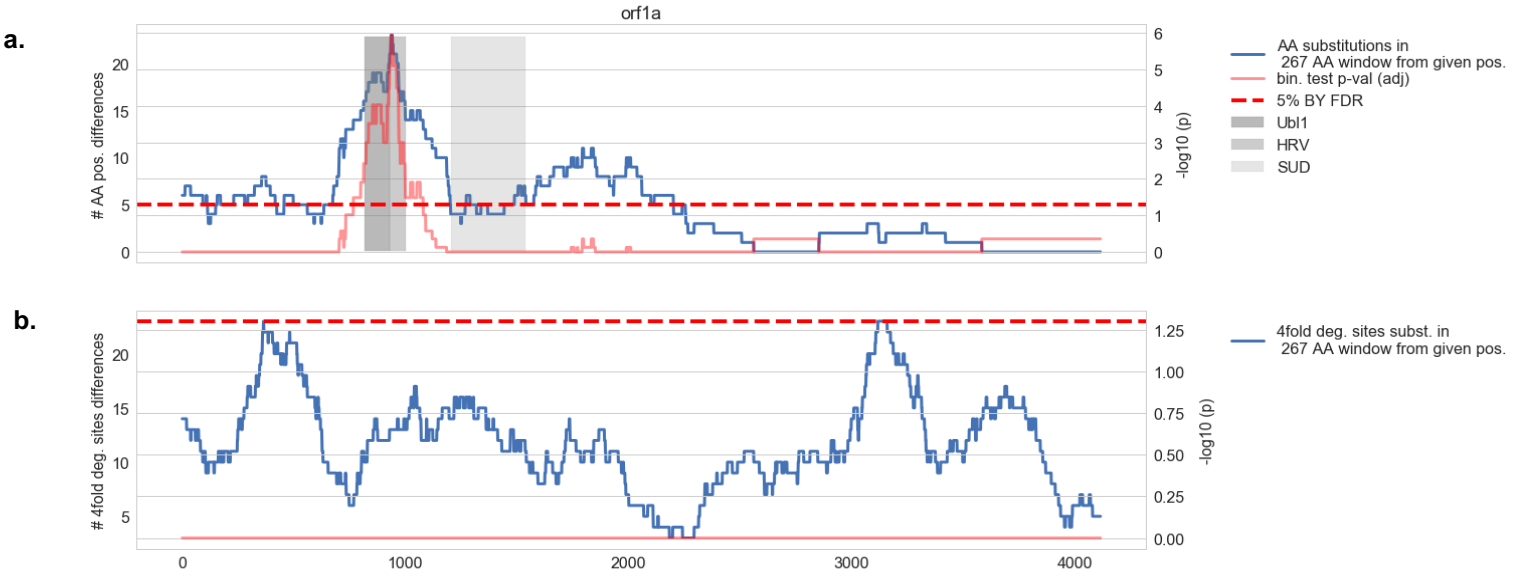
Supplementary Figure 4 | Host-specific recombination breakpoints in MERS-CoV. a, c. Distribution of start and end recombination breakpoints in MERS-CoV isolates from non-human hosts (a) and human hosts (c). **b, d.** Sliding window analysis: 800 nt length windows for non-human (b) and human-specific recombination events (d); recombination breakpoint count in each interval shown in blue; binomial test nominal p-values in red. Spike protein (light gray), RBD of Spike (green) and membrane protein (dark gray) are highlighted.



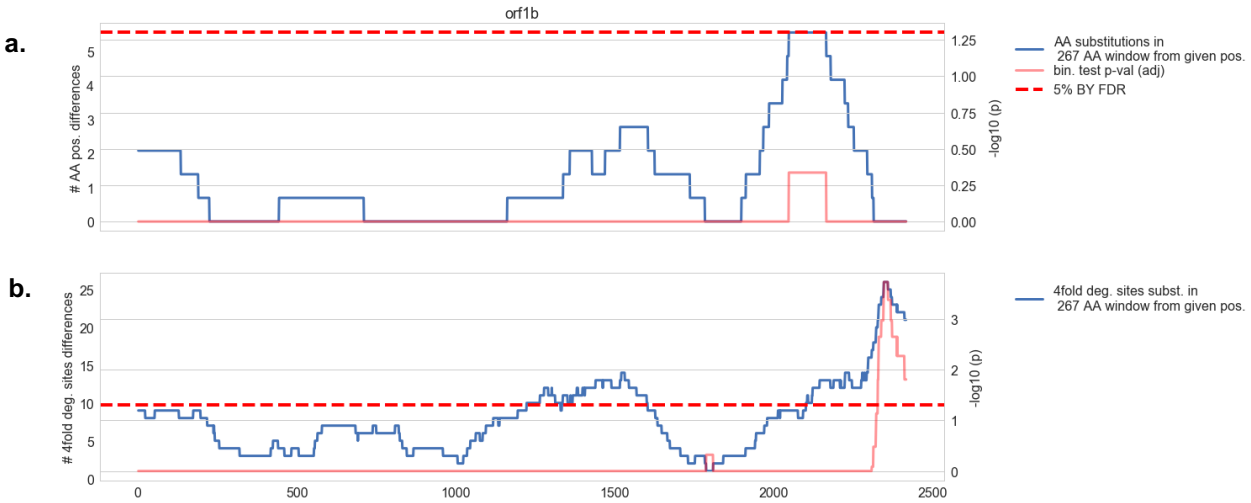
Supplementary Figure 5 | Tanglegram comparing the *Sarbecovirus* tree topologies derived from the full genome and the RBD (3rd codon positions). Diagonal lines connecting both trees illustrate topological incongruences. The lines highlighted in red represent the phylogenetic incongruences involving SARS-CoV-2 and RatG13 as evidence of their recombinant origin.



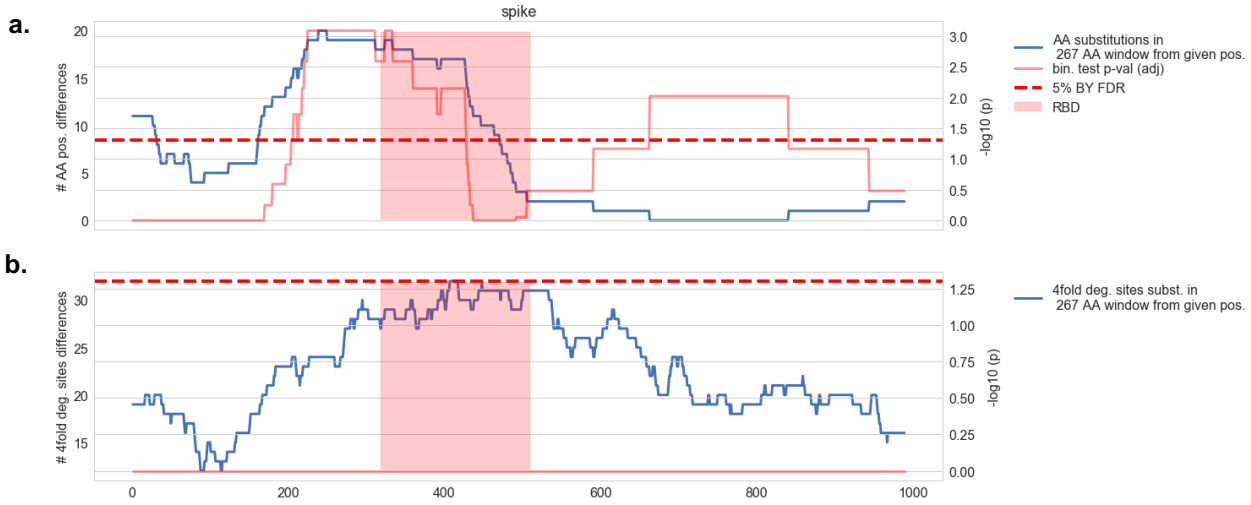
Supplementary Figure 6 | Functional impact of amino acids 427N and 436Y in the SARS-CoV-2 Spike protein. Interaction between the human ACE2 receptor (green) and the spike protein (yellow) based on SARS-CoV (PDB accession code: 2AJF), highlighting the short helix 427-436 (blue) that lies at the interface of the Spike-ACE2 interaction.



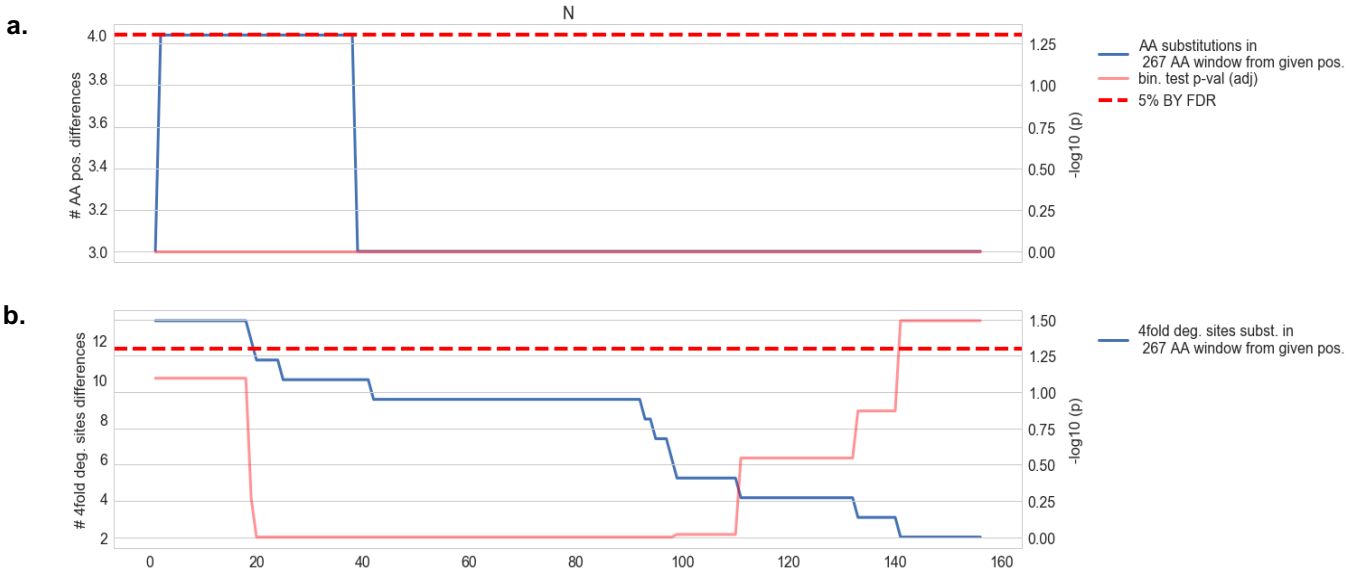
Supplementary Figure 7 | Divergence between SARS-CoV-2 and the RaTG13 bat virus in orf1a. a. Sliding window analysis of 267 aa length windows showing non-synonymous substitutions where the two strains differ (blue). BY-adjusted *p*-values for each window shown in red. Several domains of non-structural protein three are highlighted in gray: ubiquitin-like domain 1 (Ubl1), hypervariable region (HRV) and SARS-unique domain (SUD). **b.** Sliding window analysis (267 aa length windows) containing 4-fold degenerate sites that differ between SARS-CoV-2 and RaTG13. No windows are statistically significant after BY *p*-value adjustment (red line).



Supplementary Figure 8 | Divergence between SARS-CoV-2 and the RaTG13 bat virus in orf1b. a. Sliding window analysis of 267 aa length windows showing non-synonymous substitutions where the two strains differ (blue). BY-adjusted *p*-values for each window shown in red. **b.** Sliding window analysis (267 aa length windows) containing 4-fold degenerate sites that differ between SARS-CoV-2 and RaTG13.



Supplementary Figure 9 | Divergence between SARS-CoV-2 and the RaTG13 bat virus in Spike. a. Sliding window analysis of 267 aa length windows showing non-synonymous substitutions where the two strains differ (blue). BY-adjusted *p*-values for each window shown in red. RBD is highlighted. **b.** Sliding window analysis (267 aa length windows) containing 4-fold degenerate sites that differ between SARS-CoV-2 and RaTG13.



Supplementary Figure 10 | Divergence between SARS-CoV-2 and the RaTG13 bat virus in the Nucleocapsid. a. Sliding window analysis of 267 aa length windows showing non-synonymous substitutions where the two strains differ (blue). BY-adjusted *p*-values for each window shown in red. **b.** Sliding window analysis (267 aa length windows) containing 4-fold degenerate sites that differ between SARS-CoV-2 and RaTG13.