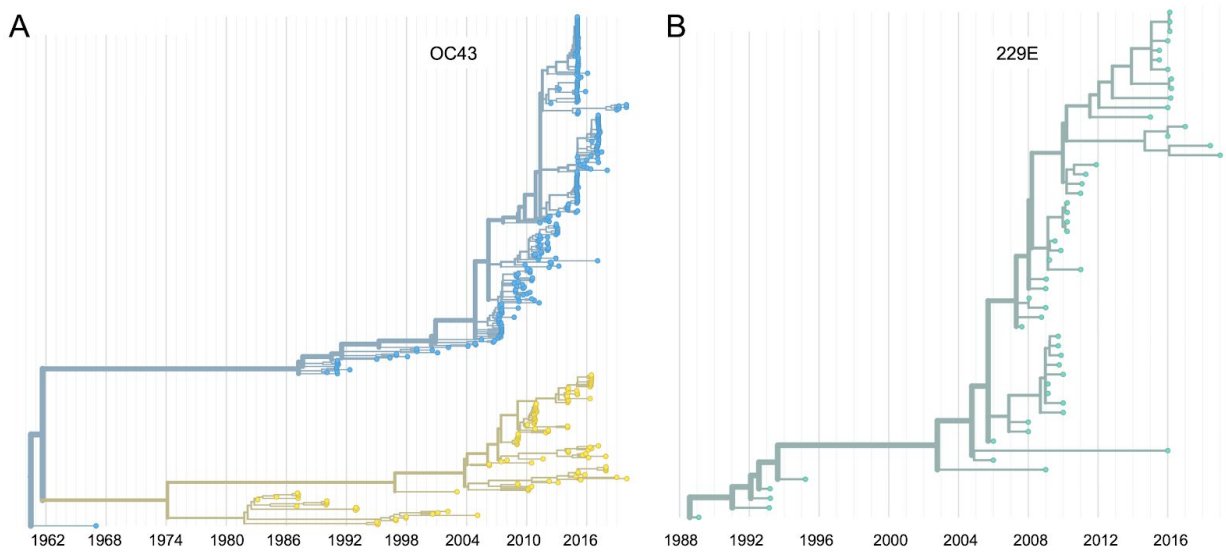

Figure 1. Phylogenetic trees for spike gene of seasonal HCoVs OC43 and 229E. Phylogenies built from A: OC43 spike sequences from 389 isolates over 53 years, and B: 229E spike sequences from 54 isolates over 31 years. HCoVs that bifurcate immediately after the root are split into blue and yellow lineages. 229E and contains just one lineage (teal). For the analyses in this paper, the evolution of each gene (or genomic region) is considered separately, so phylogenies are built for each viral gene and those phylogenies are used to split isolates into lineages for each gene. These are temporally resolved phylogenies with year shown on the x-axis. The clock rate estimate is 5×10^{-4} for OC43 and 6×10^{-4} for 229E.

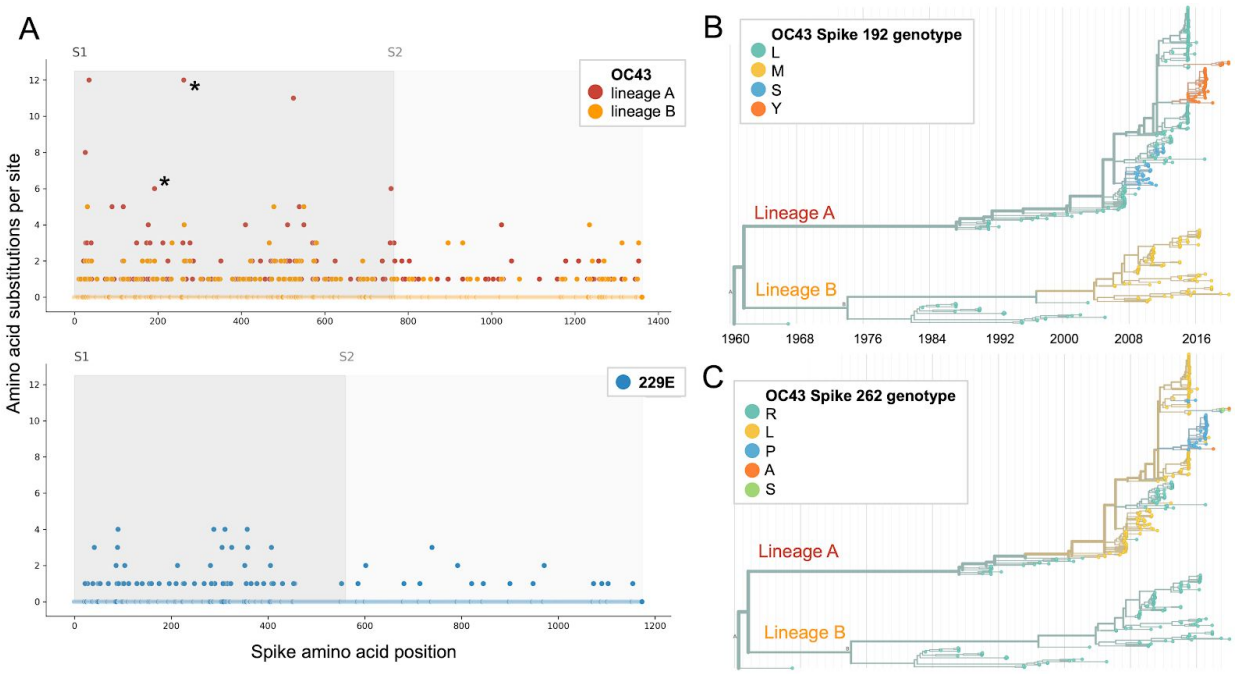

Figure 2. More sites mutate repeatedly within spike S1 versus S2. A: Number of mutations observed at each position in the spike gene. S1 (darker gray) and S2 (light gray) are indicated by shading and the average number of mutations per site is indicated by a dot and color-coded by HCoV lineage. Asterisks indicate positions 192 and 262, which mutate repeatedly throughout the OC43 lineage A phylogeny. The OC43 phylogeny built from spike sequences and color-coded by genotype at position 192 and 262 is shown in B) and C), respectively.

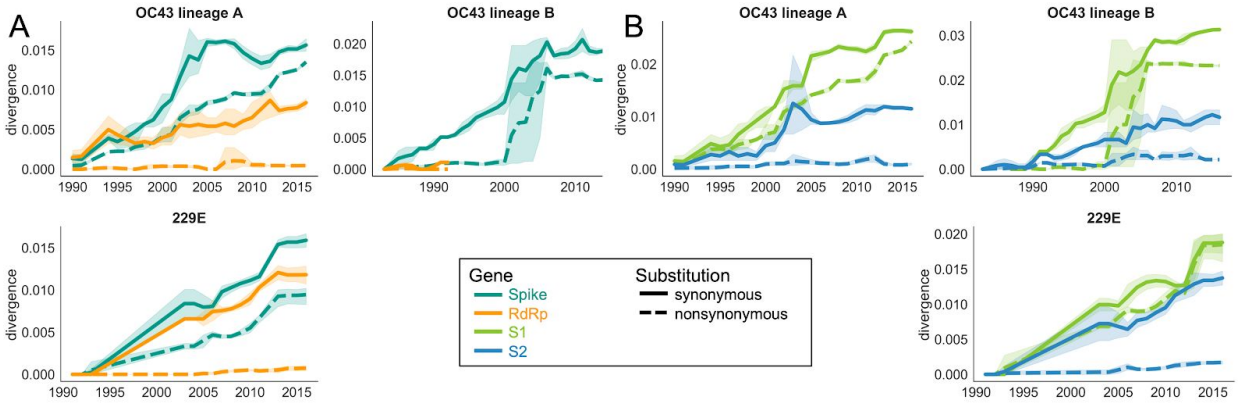

Figure 3. Nonsynonymous divergence is higher in OC43 and 229E Spike S1 versus S2 or RdRp. A: Nonsynonymous (dashed lines) and synonymous divergence (solid lines) of the spike (teal) and RdRp (orange) genes of all 229E and OC43 lineages over time. Divergence is the average Hamming distance from the ancestral sequence, computed in sliding 3-year windows which contain at least 2 sequenced isolates. Shaded region shows 95% confidence intervals. B: Nonsynonymous and synonymous divergence within the S1 (light green) and S2 (blue) domains of spike. Year is shown on the x-axis. Note that x- and y-axis scales are not shared between plots.

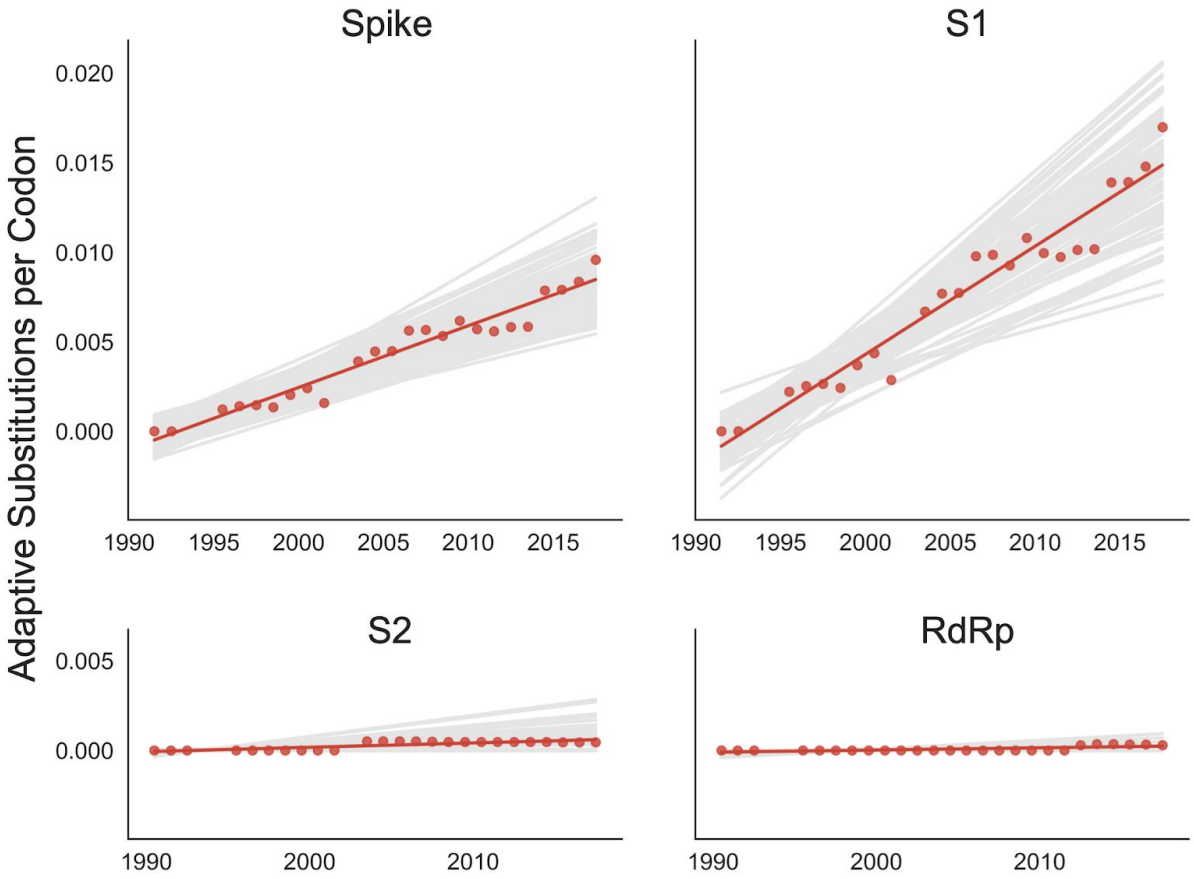

Figure 4. Adaptive substitutions accumulate over time in OC43 lineage A spike S1. Adaptive substitutions per codon within OC43 lineage A spike, S1, S2 and RdRp as calculated by our implementation of the Bhatt method. Adaptive substitutions are computed in sliding 3-year windows, and only for timepoints that contain 3 or more sequenced isolates. Red dots display estimated values calculated from the empirical data and red lines show linear regression fit to these points. Grey lines show the distribution of regressions fit to the computed number of adaptive substitutions from 100 bootstrapped datasets. Year is shown on the x-axis.

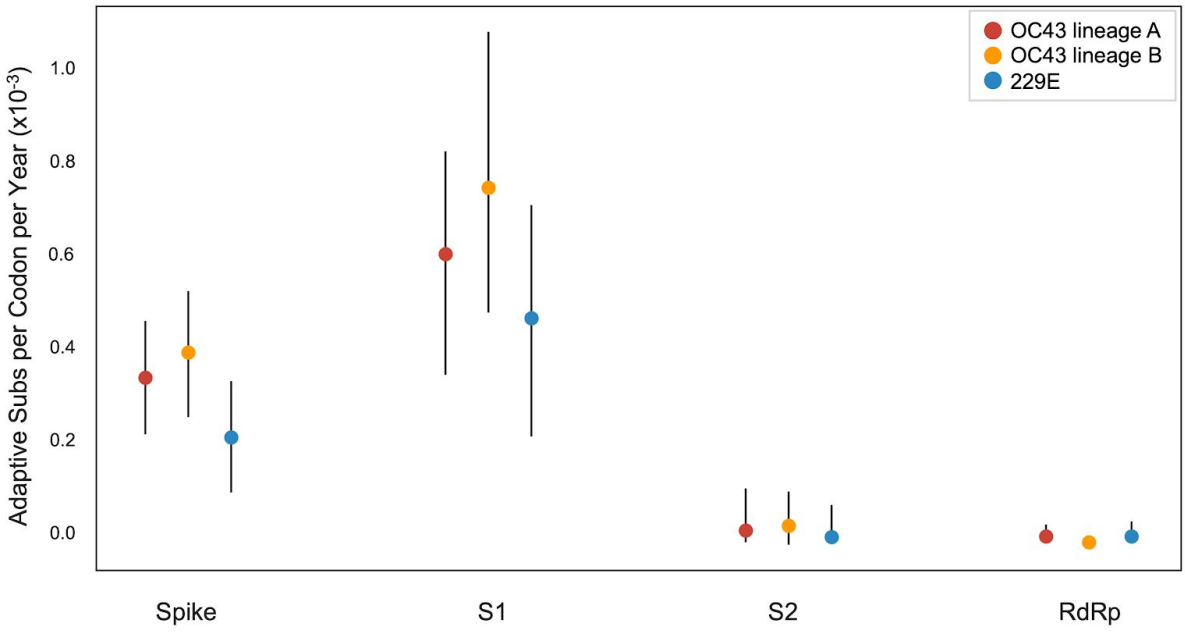

Figure 5. The rate of adaptive substitution is highest in spike S1. Adaptive substitutions per codon per year as calculated by our implementation of the Bhatt method. Rates are calculated within Spike, S1, S2 and RdRp for 229E and OC43 lineages. Error bars show 95% bootstrap percentiles from 100 bootstrapped datasets.

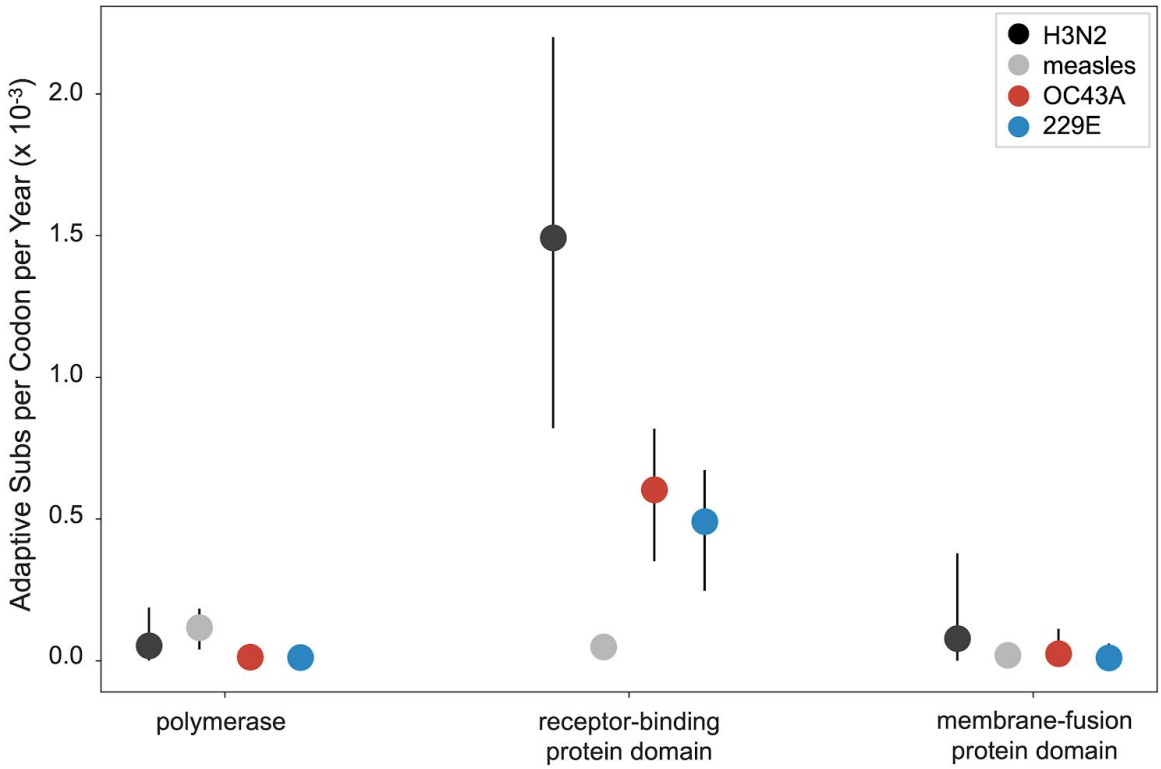

Figure 6. OC43 and 229E spike S1 accumulates adaptive substitutions faster than measles but slower than influenza H3N2. Comparison of adaptive substitutions per codon per year between influenza H3N2 (black), measles (gray), OC43 lineage A (red), and 229E (orange). The polymerase, receptor binding domain and membrane fusion domain for H3N2 are PB1, HA1 and HA2. For both HCoVs, they are RdRp, S1 and S2, respectively. For measles, the polymerase is the P gene, the receptor-binding protein is the H gene and the fusion protein is the F gene. Error bars show 95% bootstrap percentiles from 100 bootstrapped datasets.

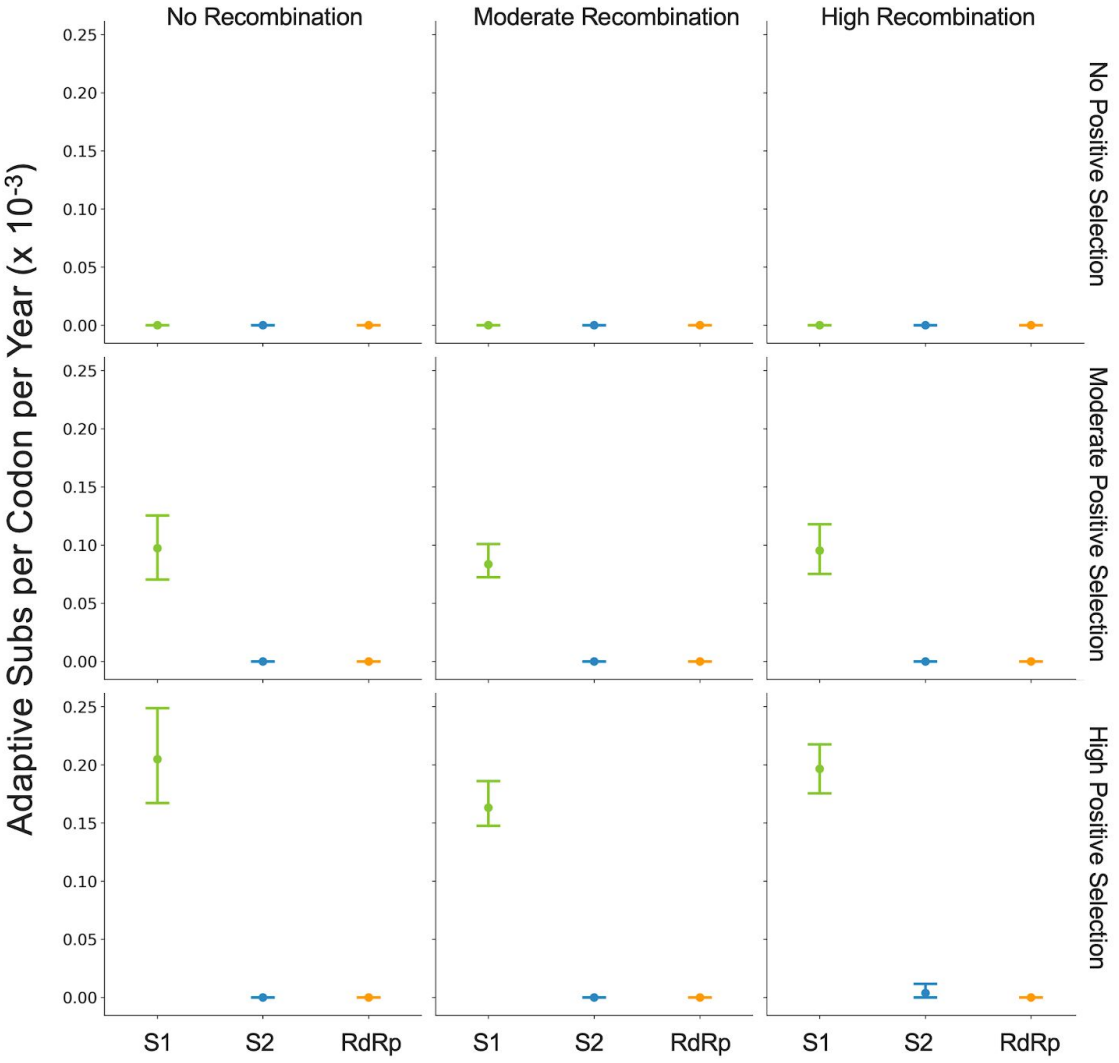

Figure 7. Detection of positive selection is not biased by recombination. OC43 lineage A sequences were simulated with varying levels of recombination and positive selection. The Bhatt method was used to calculate the rate of adaptive substitutions per codon per year for S1 (light green), S2 (blue) and RdRp (orange). The mean and 95% confidence interval of 5 independent simulations is plotted.

	Spike	S1	S2	RdRp
OC43A	4.67	3.45	13.05	17.39
229E	4.19	2.23	5.08	4.86

Table 1. Mean TMRCA is lower in S1 than RdRp or S2. Average TMRCA values (in years) for OC43 lineage A and 229E.