

Reviewer #1 (General assessment and major comments (Required)):

The authors use dN/dS-based statistics to show that the S1 domain has been adapting in coronaviruses OC43 and 229E, but not in NL63. For at least OC43 and 229E, this agrees qualitatively with earlier studies. The authors should make a detailed, quantitative comparison with these earlier results.

Comments/questions:

1. I had a hard time understanding exactly what was going on in the simulations, and whether, e.g., they produced phylogenies that looked like the real ones. But perhaps the simulations should be cut or de-emphasized. From my understanding of the Bhatt et al papers and the authors' code, their selection-detection method is entirely site-based and doesn't use the tree at all, and so should be insensitive to recombination a priori. If this is correct, the simulations are largely unnecessary (but see the next point).

**We agree that Bhatt method should be robust to recombination since it is a site-based method. However, we think the simulated data in Figure 7 provides a useful confirmation of this in addition to providing a framework to test how limited longitudinal sequencing data affects the estimated rate of adaptation (Figure 7-figure supplement). In order to provide clarity, we have added Figure 7- figure supplement 2, which shows representative phylogenies for each of the recombination/selection. When presenting early iterations of this work to colleagues, we received repeated questions about potential bias from recombination and so although the Bhatt method is indeed robust to recombination (in theory and in practice), we feel it's still useful to show this robustness explicitly.**

2. It looks like there is substantial geographic population structure, as well as time-varying geographic biases in sampling. This seems like it could generate false signals of substitutions, where standing spatial variation appears to be temporal variation. Can the authors bound how large this contribution could be? If not, perhaps they could still say that they've found signatures of either global positive selection or local adaptation, both of which are interesting.

**Examining the phylogenies with isolates colored by country**

**(<https://nextstrain.org/community/blab/seasonal-cov-adaptive-evolution/oc43/spike?c=country>), we can see that the tree forms a ladder-like topology where the tips are organized by date, rather than by country. This is especially visible when examining sequences from the 2010's, where isolates from several countries were sequenced. In these years, isolates frequently appear interspersed within the same temporal cluster rather than forming separate geographic clusters. This suggests that HCoV are rapidly transmitted globally and this geographic mixing justifies analyzing all isolates together as one population, rather conducting separate analyses on geographically-partitioned data. We have added an explanation of this to the Results section, which reads as follows:**

**“Additionally, the trees form ladder-like topologies with isolate tips arranged into temporal clusters rather than geographic clusters, indicating a single global population rather than geographically-isolated populations of virus”.**

3. Why is synonymous divergence also much higher in OC43 S1 than it is in RDRP or S2?

**We are not sure why synonymous divergence differs between genes. To address this concern, we have calculated the ratio of nonsynonymous divergence in spike to nonsynonymous divergence in RdRp and compare this to the equivalent ratio of synonymous divergence. We also compute the same ratios for S1 versus S2. We find that the nonsynonymous divergence ratio is higher than the synonymous divergence ratio. This indicates that, while spike synonymous divergence is often higher than RdRp synonymous divergence, this difference is much smaller than the difference in nonsynonymous divergences. We take this to be a convincing result showing that nonsynonymous divergence in spike is indeed higher than nonsynonymous divergence in RdRp (and the same for the comparison of S1 to S2). We have added Figure 3- figure supplement 2 showing these results, and have inserted the following text in the Results section to acknowledge this observation:**

**“Though we would expect synonymous divergence to be equivalent in all areas of the genome, this is not born out in our results. It is unclear whether the difference in synonymous divergence between genes reflects an actual biological difference. However, the ratio of nonsynonymous divergence in spike to nonsynonymous divergence in RdRp is consistently higher than the equivalent ratio of synonymous divergence (Figure 3- figure supplement 2). Thus, despite differences in synonymous divergence, spike is accumulating more relatively more nonsynonymous divergence than RdRp.”**

4. As I mention in the general assessment above, the authors should compare their results to previous studies. This includes both overall substitution rates as well as specific sites found to have repeated substitutions. For NL63, how does the null result compare to what's known? I believe Kiyuka et al, J of Infectious Diseases 2018 found widespread reinfection by similar NL63 genotypes, which may be relevant.

**Thanks for pointing us towards the Kiyuka et al paper – we have included a brief discussion of this paper in the Introduction and Discussion sections. Kiyuka et al designate multiple NL63 genotypes and show that each genotype circulates within Kilifi Kenya for a relatively long period of time. Additionally, people become reinfected by the same genotype and Kiyuka et al observe that NL63 reinfections are often enhanced by a previous infection. Kiyuka et al hypothesize that enhanced reinfection signals that NL63 is exploiting an immune response and that epitopes would actually be under purifying selection because of this. Whether or not this hypothesis is correct, their observations of persistent circulation of NL63 genotypes and reinfection by the same genotype fit well with our results showing a lack of adaptive evolution in NL63 S1.**

**“Our finding fits with a study of NL63 in Kenya, which identified multiple genotypes of NL63 and show that people regularly become reinfected with the same genotype of NL63 (Kiyuka et al. 2018). Additionally, Kiyuka et al found that these genotypes circulate locally for a long period of time, suggesting a decent amount of viral diversity and a potential lack of evolution due to immune**

**selection. Though our results cannot explain why OC43 and 229E likely evolve antigenically while NL63 does not, Kiyuka et al observe that NL63 reinfections are sometimes enhanced by a previous infection and hypothesize that NL63 is actually under purifying selection at epitope sites (Kiyuka et al. 2018)."**

**We have also added a sentence in the Results comparing our calculation of dN/dS in OC43 spike to the estimates reported in Ren et al 2015. The Chibo and Birch 2006 paper states that "a probability of 1.0 was obtained when testing for positive selection", but does not elaborate on specific dN/dS values.**

**"Our estimate of dN/dS in OC43 spike is similar to the previously reported estimate of roughly 0.3 (Ren et al. 2015)."**

Reviewer #1 (Minor Comments):

1. Fig 2 caption: Maybe clarify that 192 and 262 are just shown as illustrative examples---when I first read this, I started looking through the paper to see what was special about them.

**Figure 2 caption has been changed to clarify that these sites are examples.**

2. Fig 3: Maybe make the axis ranges uniform across subplots?

**We have changed the x- and y-axes on figure 3 so that the range is consistent for all subplots.**

3. Fig 4 seems like it should be a supplement to Fig 5, along with the corresponding plots for OC43 lineage B and 229E.

**We think it is important to include Figure 4 as a main figure, rather than a supplementary figure in order to show the roughly linear increase in adaptive substitutions over time. We believe this point is an important indication of antigenic evolution. We have added text in the paragraph describing this figure in order to emphasize this point.**

**"We observe a largely linear accumulation of adaptive substitutions in spike and S1 through time, although the method does not dictate a linear increase. This observation suggests that spike (and S1 in particular) is evolving in response to a continually changing selective pressure. This is exactly what would be expected if these adaptive substitutions are evidence of antigenic evolution resulting from an evolutionary arms race between spike and the host immune system."**

4. It would be helpful to have a more detailed description of the method for the TMRCA analysis. Not repeating what's in the PACT paper, but just explaining how it's applied in this dataset and how it's affected by recombination.

**We have included more detail on how PACT works in the Methods description of the TMRCA analysis.**

**“Briefly, PACT computes TMRCA values by creating a series of subtrees that include only tips positioned within a temporal slice of the full tree and finding the common ancestor of these tips. The overall mean and 95% confidence interval were calculated from the list of TMRCA values in these time slices.”**

**We have also expanded our discussion of how recombination may affect our TMRCA results in the Results section, including how differences in S1 and S2 may indicate recombination uncoupling these regions and how recombination can increase TMRCA estimates.**

**“The difference in TMRCA between S1 and S2 is indicative not only of differing selective pressures acting on these two spike domains, but also of intra-spike recombination. This is because the immune selection imposed on S1, should also propagate neutral hitch-hiker mutations in closely-linked regions such as S2. The difference in TMRCA suggests that recombination may uncouple these regions. Recombination can also push TMRCA to higher values, though this should not have a larger impact on RdRp than S1. The contributions of the forces of directional selection and recombination are difficult to parse from the TMRCA results.”**

5. I love that the authors made an accompanying nextstrain page and comprehensive github repo. I had some trouble navigating the nextstrain page (trying to change strains between OC43 and 229E using the dropdown menu under Dataset resulted in an error page). The github repo should probably be cleaned up a bit at some point (remove system files, fix error messages, more function descriptions in, e.g., bhatt\_bootstrapping.ipynb).

**The error seen when navigating between viruses using the drop-down menu is a bug that the developers of Nextstrain are aware of and will be addressing (<https://github.com/nextstrain/nextstrain.org/issues/240>) . If this error is encountered, all datasets can still be viewed by clicking on the desired dataset from <https://nextstrain.org/community/blab/seasonal-cov-adaptive-evolution>. We have cleaned up the GitHub repo to remove all system files that we saw. We have added more function descriptions in bhatt\_bootstrapping.ipynb.**

Reviewer #2 (General assessment and major comments (Required)):

The paper by Kistler and Bedford explores whether adaptive evolution has led to diversification of coronaviruses responsible for the common cold in the human population. The paper is very well written and presents evidence for adaptive evolution in some strains and lack of evidence in others. I enjoyed reading the paper, and all my comments are relatively minor.

Reviewer #2 (Minor Comments):

1. Line 141-2 states that more repeated mutations are expected if they confer immune avoidance. This is indeed correct, but it is also worth mentioning that an alternative hypothesis for repeat mutations is high mutation rate and genetic drift. This has in fact been shown for many recurrent mutations observed in SARS-CoV-2, and is also purportedly associated with higher mutation rate at certain types of sites (see Van Dorp et al. Nature Communications 2020).

**While preferential mutation occurs at particular types of sites, resulting in a greater recurrence of mutations at these sites between isolates, this effect should occur roughly evenly throughout the genome. In other words, this effect should not be specific to spike or S1 and likely does not account for observation of a higher rate of repeated substitutions in S1 than S2. We have added a couple sentences to the manuscript to explain this.**

**“Alternatively, these repeated substitutions could be a result of high mutation rate and random genetic drift as has been shown at particular types of sites in SARS-CoV-2 (van Dorp et al. 2020). However, this latter hypothesis should affect all regions of the genome equally and should not result in a greater number of repeated substitutions in S1 than S2.”**

2. Figure 3 shows higher synonymous rates at OC43 lineages for Spike/S1 and. Could the authors discuss reasons? How does this affect their results and interpretations on non-synonymous divergence?

**We are not sure why synonymous divergence differs between genes. To address this concern, we have calculated the ratio of nonsynonymous divergence in spike to nonsynonymous divergence in RdRp and compare this to the equivalent ratio of synonymous divergence. We also compute the same ratios for S1 versus S2. We find that the nonsynonymous divergence ratio is higher than the synonymous divergence ratio. This indicates that, while spike synonymous divergence is often higher than RdRp synonymous divergence, this difference is much smaller than the difference in nonsynonymous divergences. We take this to be a convincing result showing that nonsynonymous divergence in spike is indeed higher than nonsynonymous divergence in RdRp (and the same for the comparison of S1 to S2). We have added Figure 3- figure supplement 2 showing these results, and have inserted the following text in the Results section to acknowledge this observation:**

**“Though we would expect synonymous divergence to be equivalent in all areas of the genome, this is not born out in our results. It is unclear whether the difference in synonymous divergence between genes reflects an actual biological difference. However, the ratio of nonsynonymous divergence in spike to nonsynonymous divergence in RdRp is consistently higher than the equivalent ratio of synonymous divergence (Figure 3- figure supplement 2). Thus, despite differences**

**in synonymous divergence, spike is accumulating more relatively more nonsynonymous divergence than RdRp.”**

3. The authors employed an alternative to the canonical dN/dS metric for adaptive evolution, which is convincing. It would be nevertheless nice to see results on dN/dS using codeml or some similar model-based phylogenetic approach, which takes into account tree topology and many other factors. Fig. 3 seems to support higher dN/dS in line with the authors hypotheses but it would be useful to see the results.

**We have added a new Table 1 which shows dN/dS estimates for each HCoV and gene. For 229E, these results agree well with our general expectation that dN/dS should be higher in spike versus RdRp and S1 versus S2. In OC43 lineages, dN/dS results follow this same trend but the differences in S1 versus S2 are much smaller. Though we believe dN/dS is not the ideal way to detect adaptive evolution in HCoV spike (for reasons laid out in the paper), the dN/dS estimates are at least in line with what we would expect.**

4. The authors detect recombination in the data, how does this affect the results shown in Figure 3?

**Though it is usually assumed that immune selection acts on SNPs, it is possible that recombination is an alternative source of genetic variation. Because co-circulating HCoV isolates are closely-related (as indicated by the ladder-like shape of the phylogenies), it is possible that recombination events contribute some genetic variation without introducing larger changes that would disrupt viral functions. We have added text to the first subsection of the Results to explain this. Divergence as measured by Figure 3 is measured by directly comparing sequences rather than walking branches on a phylogenetic tree. Divergence can accrue due to either SNPs or recombination events.**

5. All in all, figure 3 is a key figure, but it is very small (much smaller than similar supplementary figures). Why the lack of data for RdRp in OC43 lineage B?

**We have reformatted Figure 3 in an effort to make it larger and more clear to the reader. We have changed the colors so that colors are used consistently throughout the paper to denote certain genes (or genomic regions) or certain HCoV types. We have also added a sentence in the Figure 3 legend to explain why the OC43 lineage B RdRp line covers only a few years. While there are 115 sequences of OC43 lineage B spike available, there are only 17 RdRp sequences. This means that, at most timepoints, there are fewer than 2 sequenced isolates and we, therefore, do not compute divergence at these timepoints.**

6. Line 192 specifies a linear accumulation of adaptive substitutions. What is the interpretation ?

**A roughly linear accumulation of adaptive substitutions supports our hypothesis that adaptive substitutions in S1 are evidence of antigenic evolution. This is because there is an evolutionary arms race between an antigenically-evolving virus and the immune system. In other words, the observation that spike (and S1 in particular), is consistently accumulating adaptive substitutions is an indication that these adaptive substitutions are a result of a continuous selective pressure (i.e. the humoral immune system). We have added this text to the manuscript.**

**“This observation suggests that spike (and S1 in particular) is evolving in response to a continual selective pressure. This is exactly what would be expected if these adaptive substitutions are evidence of antigenic evolution resulting from an evolutionary arms race between spike and the host immune system.”**

7. It would be useful to keep the colors consistent across different figures, for example, Fig. 1A and 2A with regards to lineages A and B.

**The colors used in the figures have been changed so that each virus and each gene (or genomic region) is consistently shown in the same color throughout the figures and supplements.**

8. I am a little concerned that divergence was calculated relative to the first sequence ("founder"), since this does not necessarily represent the true ancestral sequence. Due to the ladder-like tree structure it probably is a good approximation, but could this be validated by ancestral sequence reconstruction?

**We have run the same divergence analysis using the inferred root sequence of the phylogeny rather than the consensus at the first time point, so we have opted to keep Figure 3 as the divergence from the founder sequence. The results are fairly equivalent. In these screenshots, the top panel is divergence computed from the root sequence, and the bottom panel is divergence computed from consensus sequence at the first time point.**

9. Table 1 show lineage A results only, what about lineage B?

**We did not perform this analysis on OC43 lineage B because the limited number of RdRp sequences available for this lineage could artificially skew the TMRCA estimates. We have added a sentence to the Results section stating this.**

**“We did not compute TMRCA for OC43 lineage B because the limited number of available RdRp sequences for this lineage mean that TMRCA can only be calculated for about 4 years, which could artificially skew the TMRCA estimates.”**

10. Line 262: differences in TMRCA indicate selective pressure and/or recombination, but not necessarily both.

**We agree that it is difficult to parse the relative contributions of selective pressures and/or recombination to the differences in TMRCA we see between the S1 and S2 subunits. We have added a few sentences to the portion of the Results section that focuses on TMRCA to clarify this.**

**“This is because the immune selection imposed on S1, should also propagate neutral hitch-hiker mutations in closely-linked regions such as S2. The difference in TMRCA suggests that recombination may uncouple these regions. Recombination can also push TMRCA to higher values, though this should not have a larger impact on RdRp than S1. The contributions of the forces of directional selection and recombination are difficult to parse from the TMRCA results.”**

11. It is intriguing that NL63 does not display evidence for adaptive evolution and this merits more discussion why the differences among different strains that all infect in a similar manner and presumably should be under similar selective pressures.

**We agree that it is interesting that NL63 seems not to be evolving antigenically while 229E and OC43 are. In their study, Kiyuka et al examine NL63 circulating in Kilifi Kenya and find that several genotypes persist for long periods of time, and that people are commonly reinfected by the same genotype. This is suggestive of a lack of evolution driven by immune selection and is very much in line with our findings. We do not think**



our results can explain *why* this is, though Kiyuka et al 2018 have hypothesized that NL63 is actually under purifying selection at epitope sites because their data indicate that NL63 infection is enhanced by previous infections. We have updated the Introduction and Discussion sections to include this information.

**“Our finding fits with a study of NL63 in Kenya, which identified multiple genotypes of NL63 and show that people regularly become reinfected with the same genotype of NL63 (Kiyuka et al. 2018). Additionally, Kiyuka et al found that these genotypes circulate locally for a long period of time, suggesting a decent amount of viral diversity and a potential lack of evolution due to immune selection. Though our results cannot explain why OC43 and 229E likely evolve antigenically while NL63 does not, Kiyuka et al observe that NL63 reinfections are sometimes enhanced by a previous infection and hypothesize that NL63 is actually under purifying selection at epitope sites (Kiyuka et al. 2018).”**

12. The title reads as if all seasonal coronaviruses experience adaptive evolution; whereas the authors do not find this. I would thus suggest changing to "Evidence for adaptive evolution ... of some seasonal coronaviruses".

**Good point. We have changed the title to be more specific. It now reads: “Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229E”**

Reviewer #3 (Evaluation Summary (Required)):

This paper analyzing potential adaptive evolution in seasonal coronaviruses ("common colds") is highly relevant with well-supported conclusions. Its results will be widely applicable to myriad fields, including evolutionary biology, epidemiology and public health, vaccinology, immunology, and virology.

Reviewer #3 (Public Review (required)):

Kistler and Bedford present a timely and highly relevant analysis of adaptive evolution in seasonal "common cold" coronaviruses. Overall, I find the research compelling, well-performed, and mostly well-presented. The research contains sufficient statistical rigor with commendable computational reproducibility to be considered highly reliable. The authors conclude that at least two of the four known common cold coronaviruses have been undergoing adaptive evolution in their human hosts. These results may shed light on the emerging long-term evolutionary dynamics of SARS-CoV-2, the causative agent of COVID-19, as it continues circulating in humans, as well as informing ongoing vaccine design.

Reviewer #3 (Recommendations for the authors (required)):

Kistler and Bedford present a timely and highly relevant analysis of adaptive evolution in seasonal "common cold" coronaviruses. Overall, I find the research compelling, well-performed, and mostly well-presented (though some minor organizational changes are needed). As always, the commitment to open code and data from the Bedford lab is admirable and successfully performed/communicated. In my comments below, I offer advice to improve the clarity and presentation of the paper, with a few small requested analyses or need for further explanations.

1. On Page 3 line 105: "We have arbitrarily labeled these lineages 'A' and 'B' (Fig. 1)." The figure then shows A/B panels for different hCoV strains, which is a little confusing at first until you orient to the figure presentation. I recommend labeling the phylogeny in panel A with "A" and "B" rather than just using colors, and including that lineage information in the caption.

**A legend showing the color for each lineage has been added to Figure 1 and the associated figure supplements.**

2. The caption for Figure 1 needs \_units\_ for the clock rate. Presumably this is the codon sub rate per year, which is used elsewhere? Or is this nucleotide? Similarly, units are also needed... -->Page 7 line 204 for the parenthetical "(or 0.45 adaptive substitutions each year)". Are the authors converting to nucleotide?

--> Page 14 lines 388-389 page 14

**The clock rate is given in substitutions per nucleotide site per year. These units have been added to the text. The parenthetical on line 204 was altered for clarification and now reads: "(or 0.45 adaptive amino acid substitutions in S1 each year)".**

3. In general, I find the references to supplementary figures in the text confusing. For example, I first read the phrase "Fig 1 Supplement 1A" to mean both Fig 1 and Supplement 1A. Writing this as, "Supplement 1A, Fig 1" will make it more clear that Fig 1 of the main text is not being referenced.

**References to figures and supplements have been changed according to eLife convention, so that figures are referred to by (Figure 1) in the text and supplements are referred to by (Figure 1- figure supplement 1).**

4. Page 3 lines 119-122 sentences are not well ordered. Should be in order: 1) Though..., 2) Because..., 3) This....

**We have restructured this paragraph to, hopefully, make the explanation of how we assigned lineages and completed gene-specific analyses more clear.**

5. The last two sentences in the paragraph spanning page 3-4 seem tacked on and not immediately relevant to recombination. Please move these sentences or include a paragraph break.

**We have introduced a paragraph break as well as changing the text in previous sentences to, hopefully, make this section more clear to the reader.**

6. Related to the previous comment, the Results section as a whole will benefit from improved organization, specifically by creating subsections. I highly recommend adding these to improve readability.

**We have added subsections to the Results section.**

7. Comments for Figure 2:

--> Unless it becomes too busy, small indicators (or at least in the caption to avoid figure noise) might be added to emphasize the RBD domain within S1 specifically.

**We have added the 229E RBD and the putative OC43 RBD to Figure 2. We have also added a sentence to the Results that explains that we conduct analyses on all of S1 rather than just the RBD within it because neutralizing antibodies have been identified which bind to the N-Terminal Domain (NTD) as well as the RBD.**

--> I recommend changing the color of the asterisks in panel A to match the lineage A (red) color.

**We have changed the layout of this figure to put OC43 lineage A and lineage B in separate panels, which should reduce the confusion due to over-layed data. We have decided to keep the asterisks black so that is clear that they are asterisks rather than data points and hope that separating the 2 lineages into different panels will obviate the need for coloring the asterisks.**

8. Page 5 line 160 authors write, "from that lineage's common ancestor." It would be more precise to say the "from that lineage's MOST RECENT common ancestor."

**We agree and have changed this sentence to read "most recent common ancestor".**

9. Figure 3, specifically in panel A for RdRp, leaves some ambiguous interpretations: Is the line missing for OC43 lineage B because there is no RdRp data after the early 1990s (seems unlikely?), or because there were no adaptive substitutions, in which case the orange lines should remain steady at 0? This aspect of the figure should be clarified or fixed.

**We have added a sentence in the Figure 3 legend to clarify this.**

--> There is a similar situation for panel C, HKU1 lineage B, in fig s3. In addition, the "C" for that panel is cut off at the bottom, so this figure needs to be slightly reformatted.

**We have reformatted the supplement to avoid the panels being cut off.**

10. Page 6 line 185, the authors introduce the H3N2 analysis, but the actual analysis is not really presented or described for another 2 pages. The H3N2 comparison is definitely not part of figure 4, which this paragraph is introducing. I think this sentence is likely misplaced? Again, this comment shows that adding subsections to the Results section will be helpful for overall organization.

**We have removed the mention of H3N2 in the introduction of the Bhatt analysis. We now mention calculating rates of adaptation in H3N2, measles and flu B several paragraphs later, when we actually compare these rates to seasonal HCoVs.**

11. Page 6 line 179, I would like to see more details about what constitutes an \_adaptive substitution\_ in the Bhatt method, which is not as widely used as dN/dS, within the main text

itself. A couple additional sentences briefly and "birds eye view" explaining what constitutes adaptive will help orient readers. The easiest way to this - just move lines 418-422 from page 14 ("Briefly,.....each of these timepoints.") to the results section.

**We agree that it would be useful to orient the reader to the Bhatt method by providing a brief overview of the method when it is introduced in the Results section. We have included this explanation in the Results section as suggested and rearranged the Methods description of the Bhatt method accordingly.**

**"Briefly, this method defines a class of neutrally-evolving nucleotide sites as those with synonymous mutations or where nonsynonymous polymorphisms occur at medium frequency. Then, the number of fixed and high-frequency nonsynonymous sites that exceed the neutral expectation are calculated. This method compares nucleotide sequences at each timepoint (the ingroup) to the consensus nucleotide sequence at the first time point (the outgroup) and yields an estimate of the number of adaptive substitutions within a given genomic region at each of these timepoints."**

12. Jumping off of the last comment - why wasn't dN/dS done? Given that dN/dS is more commonly applied, I think a comparison of these results to standard dN/dS analysis is merited. In fact, including a dN/dS analysis may bolster the authors' overall conclusions and/or contribute to justifying using the Bhatt method, especially if dN/dS is not sufficient sensitive for this data.

**We originally only included an analysis of the data using the Bhatt method because this method (and implementations of the McDonald-Kreitman method in general) is better at distinguishing relaxed selection from positive selection, whereas dN/dS can get confused in these situations. However, while we think dN/dS is less appropriate for detecting selection in RNA virus populations, we agree that including this standard method for comparison is useful. We have included dN/dS estimates for each HCoV and gene in Table 1. As expected, here we see higher dN/dS values for spike than RdRp and higher dN/dS values for S1 than S2.**

13. In general, the authors should clarify their use of the terms "positive selection" and "adaptive substitutions." The former is traditionally associated with interpreting dN/dS, which isn't calculated in the manuscript, and the latter term is more mechanistically-oriented regarding effects of mutations. Therefore, what is meant by "positive selection" in the simulations, and how does this definition/implementation compare to the authors' measurements of "adaptive substitutions"?

**While the terms "positive selection" and "adaptive evolution" are not synonymous, they are related in that adaptive evolution is a result of positive selection acting on genetic variation. We have added a sentence to the Results section to explain this and aim to use these terms in throughout our paper as they are used in the literature (i.e. "positive selection" to refer to dN/dS results and "adaptive substitutions" to refer to the Bhatt method).**

**“ Mutations that escape from population immunity are beneficial to the virus and so are driven to fixation by positive selection. This results in adaptive evolution of the virus population.”**

**For the simulations, we specify a fitness function that imitates positive selection at epitope sites, then we use the Bhatt method to detect adaptive evolution, the result of this positive selection. We have added more detail to the Methods section describing how these simulations are done.**

**“Positive selection was simulated through exposure-dependent selection at a subset of spike S1 sites proportional to the number of epitope sites in H3N2 HA (Luksza and Lässig 2014). The simulated selection allows mutations in these “epitope” sites to rise in frequency while also encouraging “epitopes” to change over time (to mimic antigenic novelty).”**

14. Figure 7 and its associated analysis raised some minor concerns for me. It seems like only simulation 5 replicates were performed for each condition. Is there a reason so few simulations were performed (eg too computationally expensive?). Further, mean and CI bars for only 5 replicates in Figure 7 gives the impression that there are more than 5 replicates. A strip plot would be more forthcoming about the analyses conducted here, and some additional explanation about why only 5 replicates were performed per condition would help.

**We have run additional simulations so Figure 7 now displays mean and 95% CI for 10 replicates instead of 5.**

15. Table 1 and its associated analysis:

--> CI's or some measure of statistical bounds should be included in Table 1.

**We have added 95% confidence intervals to the table of TMRCA values (now Table 2).**

--> Where is OC43B in the table? Was the analysis not performed on this lineage, and if so why not?

**We did not perform this analysis on OC43 lineage B because the limited number of RdRp sequences available for this lineage could artificially skew the TMRCA estimates. We have added a sentence to the Results section stating this.**

--> The authors motivate this analysis by explaining how TMRCA is meaningful for H3N2. Can the authors perform this analysis for H3N2 proteins as well to provide further context for the HCoV values, just as they did for these analyses associated with Fig 5?

**Based on Bedford et al 2011, our estimates of HCoV TMRCA are roughly 2-2.5 times longer than TMRCA of H3N2 hemagglutinin. We have included this information in the Results section.**

16. Page 11 paragraph beginning on line 283. To motivate this analysis, the authors may also wish to cite this paper, co-authored by Trevor, that uses a downsampling strategy from empirical to study 2009 H1N1 dynamics, and shows time dependency in evolutionary metrics.

<https://bedford.io/papers/meyer-time-dependence/>

--> In addition, why did the authors use simulated data here? If we have HCoV sequence data since at least the 1990s, it seems possible to have used real data here. Further explanation/justification is therefore needed.

**While we could use down-sampling strategy of the empirical data to show a decrease in power with a decrease in data, we think that using simulated data to show this relationship is better because it gives a “truth” to compare to. In other words, in the simulated data we know whether or not there is positive selection on residues within S1 and, therefore, whether or not adaptive substitutions should be detected. Using *all* of the simulated data gives us a “true” rate of adaptation that we can compare other rates of adaptation to. Empirical HCoV data is already somewhat sparse and, therefore, we do not think the rates of adaptation we calculate can be regarded as a “truth”, but rather are estimates.**

--> All that said, looking at the CI's (assuming these are CI's - the caption needs to add this info) in Fig s7, the bounds across time points are often overlapping. One might expect that CIs would be wider as sample size decreases, which is not always the case. To my eye, the ONLY panel in this figure that truly shows the authors' conclusion is the "no recombination/high positive selection" panel.

**We have added clarification to the Figure 7 figure Supplement to state that the mean and 95% CI are plotted for each point. Additionally, we ran more simulations (for a total of 10). Our reasoning in including this figure supplement is to show that when the 30-year point is taken as the “true” rate of adaptation, it is evident that a reduced amount of temporal data alters the estimated rate of adaptation. Though the general trend is that less longitudinal sequence data results in a lower estimated rate of adaptation and higher uncertainty, this is not the case at every time point. However, whether a truncated time-series results in a lower or higher rate of adaptation, the estimated rate is still skewed from the “true” rate. We have added a sentence to the Results section to try to clarify this interpretation of Figure 7 figure supplement.**

17. Please add a reference for the sentence page 12 line 333-334 about transmissibility/pathology correlates.

**We have rewritten this sentence to remove this statement.**

18. Line 384 of page 14: IG-Tree typo should be IQ-TREE. In addition, The authors may also wish to confirm on their own end which IQ-TREE version was used - a major version 2 was released in 2020 and has a different citation. Either way, please indicate the IQ-TREE version used and make sure the citation is right for whichever was used.

**The typo in the methods section has been fixed, the version number was added, and the citation was verified.**

19. Figure 2 supplement 1 panel C - please choose different colors. Since some transparency is used for points, it's very hard to distinguish precisely light from dark purple.

**We have changed the colors used to represent each virus and each gene to be more consistent throughout the paper. For the figure in question, we have also expanded the**

**panel into two panels (one for each lineage) so that data points from the two lineages can be more readily distinguished.**

20. Grammar and spelling:

-->Page 1 line 27, there should be a comma after 229E (as in, "...two species of HCoV, OC43 and 229E, were identified...")

**Done.**

-->Page 2 sentence beginning on 55 "Some human respiratory illness..." is a runon sentence. Please add a comma before ",while others,".

**Done.**

-->Caption of figure 1: Sentence in the middle of line 134 is a runon. Please add a comma at "...for each viral gene, and those..."

**Done.**

-->Page 5 line 163 typo, "spikenand" → "spike and"

**Addressed. Thank you.**

-->Page 7 sentence beginning on line 203 is a run-on sentence. Please add a comma before ",while the rate of adaptation..."

**Done.**

-->Page 12 line 335-336, again a comma is needed before "while."

**Fixed.**