

1 **Evidence for adaptive evolution in the receptor-binding domain of**
2 **seasonal coronaviruses**

3

4 Kathryn E. Kistler^{*1,2}, Trevor Bedford^{1,2}

5

6 *¹Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle,*
7 *WA USA*

8 *²Molecular and Cellular Biology Program, University of Washington, Seattle, WA USA*

9 * To whom correspondence should be addressed.

10

11 **Abstract**

12 Seasonal coronaviruses (OC43, 229E, NL63 and HKU1) are endemic to the human population,
13 regularly infecting and reinfecting humans while typically causing asymptomatic to mild
14 respiratory infections. It is not known to what extent reinfection by these viruses is due to
15 waning immune memory or antigenic drift of the viruses. Here, we address the influence of
16 antigenic drift on immune evasion of seasonal coronaviruses. We provide evidence that at least
17 two of these viruses, OC43 and 229E, are undergoing adaptive evolution in regions of the viral
18 spike protein that are exposed to human humoral immunity. This suggests that reinfection may
19 be due, in part, to positively-selected genetic changes in these viruses that enable them to
20 escape recognition by the immune system. It is possible that, as with seasonal influenza, these
21 adaptive changes in antigenic regions of the virus would necessitate continual reformulation of a
22 vaccine made against them.

23

24 **Introduction**

25 Coronaviruses were first identified in the 1960s and, in the decades that followed, human
26 coronaviruses (HCoVs) received a considerable amount of attention in the field of infectious
27 disease research. At this time, two species of HCoV, OC43 and 229E were identified as the
28 causative agents of roughly 15% of common colds (McIntosh 1974; Heikkinen and Järvinen
29 2003). Infections with these viruses were shown to exhibit seasonal patterns, peaking in
30 January-March in the Northern Hemisphere, as well as yearly variation, with the greatest
31 incidence occurring every 2-4 years (Monto and Lim 1974; Hamre and Beem 1972).
32 Subsequently, two additional seasonal HCoVs, HKU1 and NL63, have entered the human
33 population. These 4 HCoVs endemic to the human population usually cause mild respiratory
34 infections, but occasionally result in more severe disease in immunocompromised patients or
35 the elderly (Liu, Liang, and Fung 2020). In the past 20 years, three additional HCoVs
36 (SARS-CoV-1, MERS-CoV and SARS-CoV-2) have emerged, which cause more severe
37 respiratory illness. At the writing of this paper, amidst the SARS-CoV-2 pandemic, no vaccine for
38 any HCoV is currently available, though many candidate SARS-CoV-2 vaccines are in
39 production and clinical trials (Krammer 2020).

40

41 Coronaviruses are named for the ray-like projections of spike protein that decorate their surface.
42 Inside these virions is a positive-sense RNA genome of roughly 30kB (Li 2016). This large
43 genome size can accommodate more genetic variation than a smaller genome (Woo et al.
44 2009). Genome flexibility, coupled with a RNA virus error-prone polymerase (Drake 1993) and a
45 high rate of homologous recombination (Pasternak, Spaan, and Snijder 2006), creates genetic
46 diversity that is acted upon by evolutionary pressures that select for viral replication. This
47 spawns much of the diversity within and between coronaviruses species (Woo et al. 2009; Hon
48 et al. 2008), and can contribute to the virus' ability to jump species-barriers, allowing a
49 previously zoonotic CoV to infect and replicate in humans.

50

51 The battle between virus and host results in selective pressure for mutations that alter viral
52 antigens in a way that evades immune recognition. Antigenic evolution, or antigenic drift, leaves
53 a characteristic mark of positively selected epitopes within the viral proteins most exposed to the
54 host immune system (Smith et al. 2004). For CoVs, this is the spike protein, exposed on the
55 surface of the virion to human humoral immunity. Some human respiratory illnesses caused by
56 RNA viruses, like seasonal influenza (Smith et al. 2004), evolve antigenically while others, like
57 measles, do not (Fulton et al. 2015). Because of this, seasonal influenza vaccines must be
58 reformulated on a nearly annual basis, while measles vaccines typically provide lifelong
59 protection. Whether HCoVs undergo antigenic drift is relevant not only to understanding HCoV
60 evolution and natural immunity against HCoVs, but also to predicting the duration of a vaccine's
61 effectiveness.

62

63 Early evidence that closely-related HCoVs are antigenically diverse comes from a 1980s human
64 challenge study in which subjects were infected and then reinfected with a variety of
65 229E-related strains (Reed 1984). All subjects developed symptoms and shed virus upon initial
66 virus inoculation. After about a year, subjects who were re-inoculated with the same strain did
67 not show symptoms or shed virus. However, the majority of subjects who were re-inoculated
68 with a heterologous strain developed symptoms and shed virus. This suggests that immunity
69 mounted against 229E viruses provides protection against some, but not all, other 229E strains.
70 This is a result that would be expected of an antigenically evolving virus.

71

72 More recent studies have identified 8 OC43 genotypes and, in East Asian populations, certain
73 genotypes were shown to temporally replace other genotypes (Lau et al. 2011; Zhang et al.
74 2015; Zhu et al. 2018). Whether certain genotypes predominate due to antigenic differences
75 that confer a fitness advantage is not known. However, evidence for selection in the spike
76 protein of one of these dominant OC43 genotypes has been provided by *dN/dS*, a standard
77 computational method for detecting positive selection (Ren et al. 2015). This method has also
78 been used to suggest positive selection in the spike protein of 229E (Chibo and Birch 2006).
79 Additionally, two genetically distinct groupings (each of which include multiple of the
80 aforementioned 8 genotypes) of OC43 viruses have been shown to alternate in prevalence
81 within a Japanese community, meaning that the majority of OC43 infections are caused by one
82 group for about 2-4 years at which point the other group begins to account for the bulk of

83 infections. It has been suggested that antigenic differences between these groups contribute to
84 this epidemic switching (Komabayashi et al. 2020).

85

86 Here, we use a variety of computational approaches to detect adaptive evolution in spike and
87 comparator proteins in HCoVs. These methods were designed as improvements to dN/dS with
88 the intention of identifying positive selection within a serially-sampled RNA virus population. We
89 focus on the seasonal HCoVs that have been continually circulating in humans: OC43, 229E,
90 HKU1 and NL63. Our analyses of nonsynonymous divergence, rate of adaptive substitutions,
91 and Time to Most Recent Ancestor (TMRCA) provide evidence that the spike protein of OC43
92 and 229E is under positive selection. Though we conduct these analyses on HKU1 and NL63,
93 we do not observe evidence for adaptive evolution in the spike protein of these viruses. For
94 HKU1, there is not enough longitudinal sequencing data available for us to confidently make
95 conclusions as to whether or not this lack of evidence reflects an actual lack of adaptive
96 evolution.

97

98 **Results**

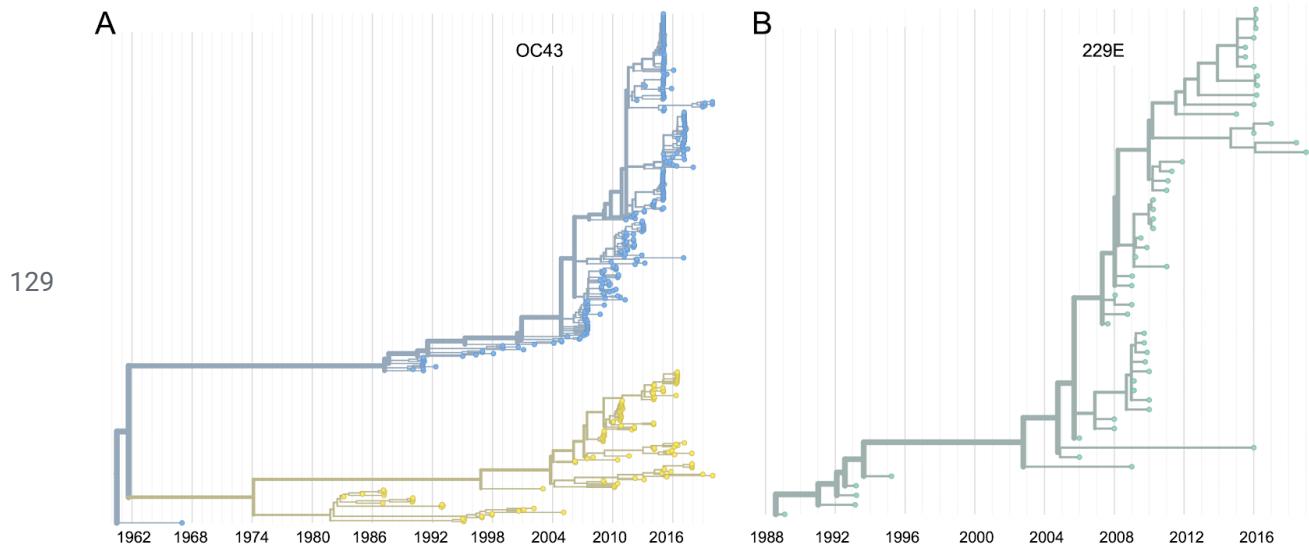
99 We constructed time-resolved phylogenies of the OC43 and 229E using publicly accessible
100 sequenced isolates. A cursory look at these trees confirms previous reports that substantial
101 diversity exists within each viral species (Zhang et al. 2015; Komabayashi et al. 2020; Lau et al.
102 2011). The phylogeny of OC43 bifurcates immediately from the root (Fig. 1), indicating that
103 OC43 consists of multiple, co-evolving lineages. Because of the distinct evolutionary histories, it
104 is appropriate to conduct phylogenetic analyses separately for each lineage. We have arbitrarily
105 labeled these lineages 'A' and 'B' (Fig. 1).

106

107 Because recombination is common amongst coronaviruses (Pasternak, Spaan, and Snijder
108 2006; Hon et al. 2008; Lau et al. 2011), we built separate phylogenies for each viral gene. In the
109 absence of recombination, each tree should show the same evolutionary relationships between
110 viral isolates. A dramatic difference in a given isolate's position on one tree versus another is
111 strongly indicative of recombination (Kosakovsky Pond et al. 2006). Comparing the
112 RNA-dependent RNA polymerase (RdRp) and spike trees reveals this pattern of recombination
113 in some isolates (Fig. 1 Supplement 1A). A comparison of the trees of the S1 and S2
114 sub-domains of spike shows more limited evidence for intragenic recombination (Fig. 1
115 Supplement 1B), which is consistent with the fact that the distance between two genetic loci is
116 inversely-related to the chance that these loci remain linked during a recombination event.
117 Though intragenic recombination likely does occur occasionally, analyzing genes, rather than
118 isolates, greatly reduces the contribution of recombination to genetic variation in our analyses.
119 Because of this, we designate the lineage of each gene separately, based on that gene's
120 phylogeny. Though most isolates contain all genes from the same lineage, some isolates have,
121 say, a lineage A spike gene and a lineage B RdRp gene. This allows us to consider the
122 evolution of each gene separately, and interrogate the selective pressures acting on them.
123 Because of its essential role in viral replication and lack of antibody exposure, we expect RdRp
124 to be under purifying selection to maintain its structure and function. If HCoVs evolve
125 antigenically, we expect to see adaptive evolution in spike, and particularly in the S1 domain of

126 Spike (Hofmann et al. 2006; Hulswit et al. 2019), due to its exposed location at the virion's
127 surface and interaction with the host receptor.

128

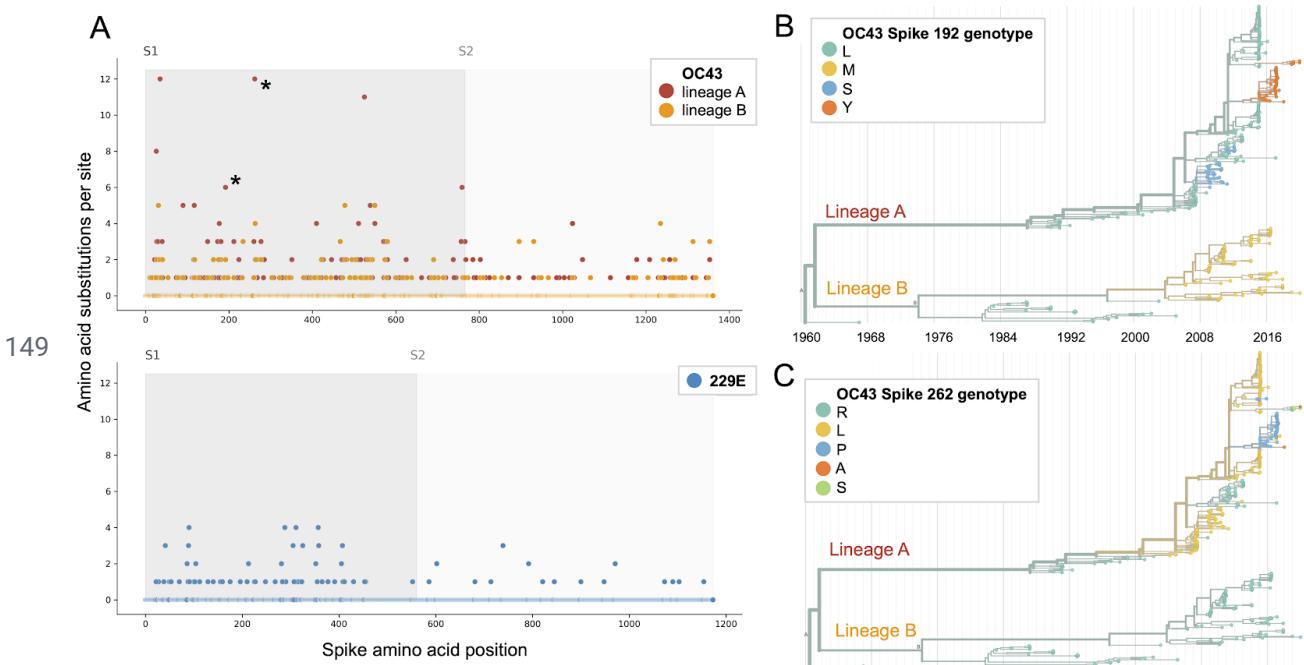


130 **Figure 1. Phylogenetic trees for spike gene of seasonal HCoVs OC43 and 229E.** Phylogenies built
131 from A: OC43 spike sequences from 389 isolates over 53 years, and B: 229E spike sequences from 54
132 isolates over 31 years. HCoVs that bifurcate immediately after the root are split into blue and yellow
133 lineages. 229E and contains just one lineage (teal). For the analyses in this paper, the evolution of each
134 gene (or genomic region) is considered separately, so phylogenies are built for each viral gene and those
135 phylogenies are used to split isolates into lineages for each gene. These are temporally resolved
136 phylogenies with year shown on the x-axis. The clock rate estimate is 5×10^{-4} for OC43 and 6×10^{-4} for
137 229E.

138

139 Using phylogenies constructed from the spike gene, we tallied the number of independent
140 amino acid substitutions at each position within spike. The average number of mutations per site
141 is higher in S1 than S2 for HCoV lineages in OC43 and 229E (Fig. 2A). A greater occurrence of
142 repeated mutations is expected if some mutations within S1 confer immune avoidance. Not only
143 should S1 contain more repeated mutations, but we would also expect these mutations to
144 spread widely after they occur due to their selective advantage. Additionally, we expect sites
145 within S1 to experience diversifying selection due to the ongoing arms race between virus and
146 host immune system. This is visible in the distribution of genotypes at the most
147 repeatedly-mutated sites in OC43 lineage A (Fig. 2B and 2C).

148

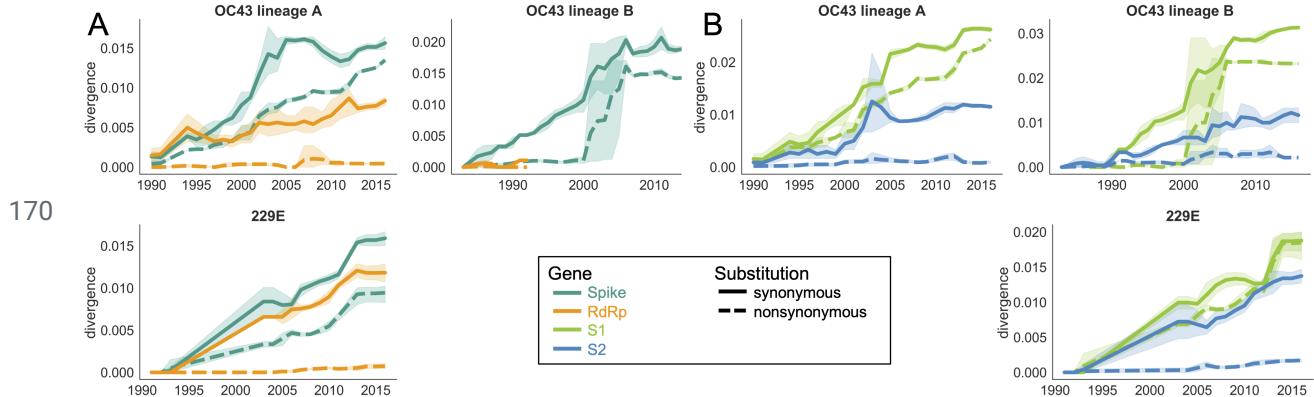


150 **Figure 2. More sites mutate repeatedly within spike S1 versus S2.** A: Number of mutations observed
 151 at each position in the spike gene. S1 (darker gray) and S2 (light gray) are indicated by shading and the
 152 average number of mutations per site is indicated by a dot and color-coded by HCoV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

156

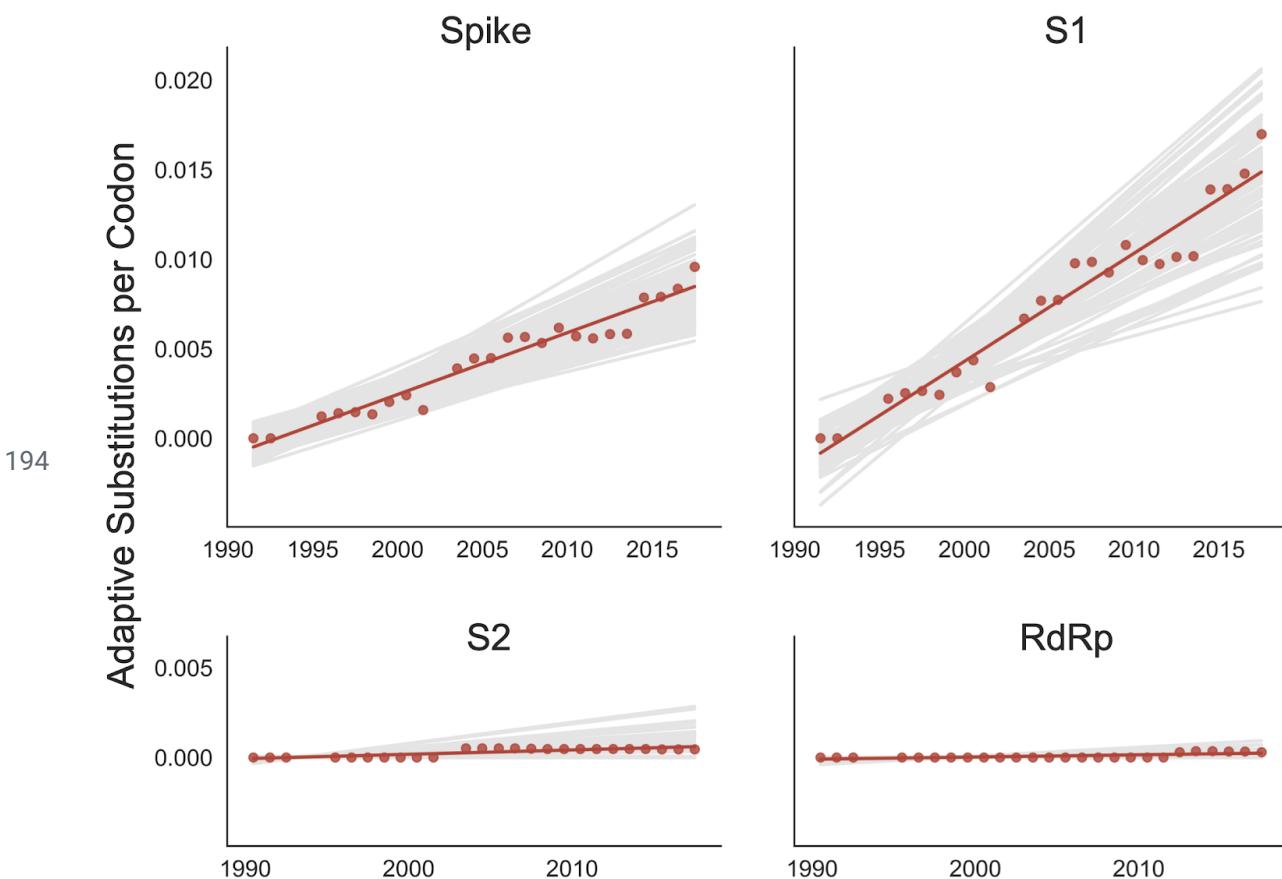
157 An adaptively evolving gene, or region of the genome, should exhibit a high rate of
 158 nonsynonymous substitutions. For each seasonal HCoV lineage, we calculated nonsynonymous
 159 and synonymous divergence as the average Hamming distance from that lineage's common
 160 ancestor (Zanini et al. 2015). The rate of nonsynonymous divergence is markedly higher within
 161 spike versus RdRp of 229E and OC43 lineage A (Fig. 3A). While nonsynonymous divergence
 162 increases steadily over time in spike, it remains roughly constant at 0.0 in RdRp, while rates of
 163 synonymous evolution are similar between spike and RdRp. These results suggest that there is
 164 predominantly positive selection on OC43 and 229E spike, but predominantly purifying selection
 165 on RdRp. Separating spike into the S1 (receptor-binding) and S2 (membrane-fusion) domains
 166 reveals that the majority of nonsynonymous divergence in spike occurs within S1 (Fig. 3B). In
 167 fact, the rates of nonsynonymous divergence in S2 are similar to those seen in RdRp,
 168 suggesting S2 evolves under purifying selection while S1 evolves adaptively.

169



171 **Figure 3. Nonsynonymous divergence is higher in OC43 and 229E Spike S1 versus S2 or RdRp.** A:
 172 Nonsynonymous (dashed lines) and synonymous divergence (solid lines) of the spike (teal) and RdRp
 173 (orange) genes of all 229E and OC43 lineages over time. Divergence is the average Hamming distance
 174 from the ancestral sequence, computed in sliding 3-year windows which contain at least 2 sequenced
 175 isolates. Shaded region shows 95% confidence intervals. B: Nonsynonymous and synonymous
 176 divergence within the S1 (light green) and S2 (blue) domains of spike. Year is shown on the x-axis. Note
 177 that x- and y-axis scales are not shared between plots.
 178

179 As a complement to the divergence analysis, we implemented an alternative to the dN/dS
 180 method that was specifically designed to detect positive selection within RNA virus populations
 181 (Bhatt, Holmes, and Pybus 2011). Compared with traditional dN/dS methods, the Bhatt method
 182 has the advantages of: 1) measuring the strength of positive selection within a population given
 183 sequences collected over time, 2) higher sensitivity to identifying mutations that occur only once
 184 and sweep through the population, and 3) correcting for deleterious mutations (Bhatt,
 185 Katzourakis, and Pybus 2010; Bhatt, Holmes, and Pybus 2011). We adapted this method to
 186 detect adaptive substitutions in seasonal HCoVs and compare these rates to H3N2, the
 187 canonical example of antigenic evolution (Rambaut et al. 2008; Yang 2000). As shown in Figure
 188 4, OC43 lineage A has continuously amassed adaptive substitutions in spike over the past >30
 189 years while RdRp has accrued few, if any, adaptive substitutions. These adaptive substitutions
 190 are located within the S1, and not the S2, domain of spike (Fig. 4). We observe a largely linear
 191 accumulation of adaptive substitutions in spike and S1 through time, although the method does
 192 not dictate a linear increase.
 193

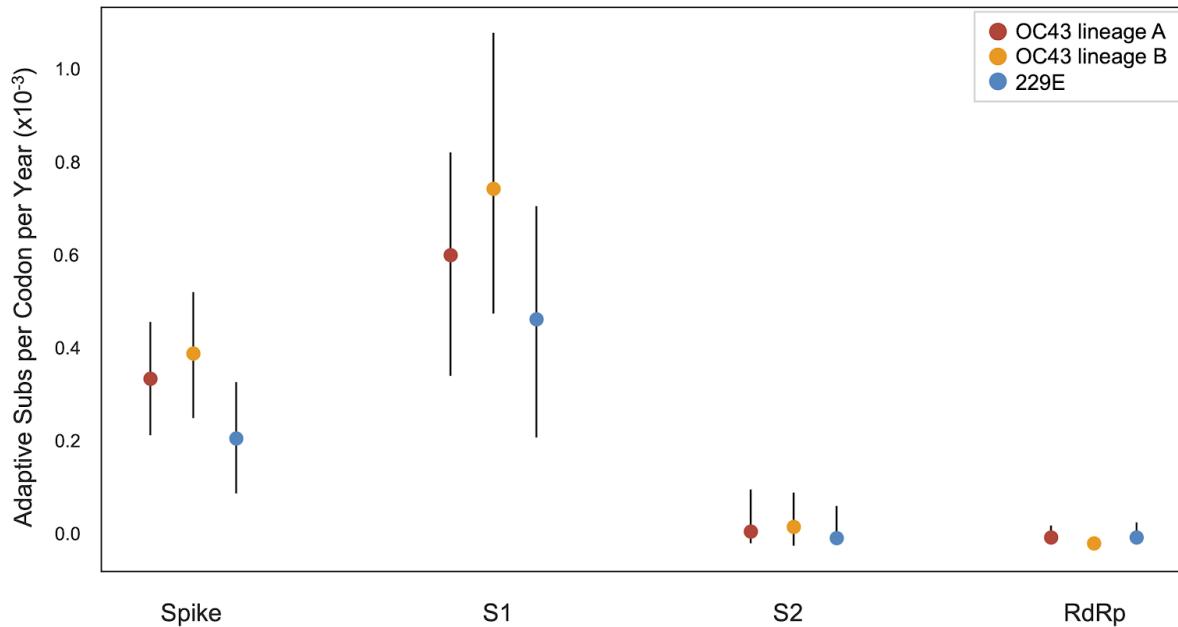


195 **Figure 4. Adaptive substitutions accumulate over time in OC43 lineage A spike S1.** Adaptive
 196 substitutions per codon within OC43 lineage A spike, S1, S2 and RdRp as calculated by our
 197 implementation of the Bhatt method. Adaptive substitutions are computed in sliding 3-year windows, and
 198 only for timepoints that contain 3 or more sequenced isolates. Red dots display estimated values
 199 calculated from the empirical data and red lines show linear regression fit to these points. Grey lines show
 200 the distribution of regressions fit to the computed number of adaptive substitutions from 100 bootstrapped
 201 datasets. Year is shown on the x-axis.
 202
 203 We estimate that OC43 lineage A accumulates roughly 0.6×10^{-3} adaptive substitutions per
 204 codon per year (or 0.45 adaptive substitutions each year) in the S1 domain of spike while the
 205 rate of adaptation in OC43 lineage B is slightly higher and is estimated to result in an average
 206 0.56 adaptive substitutions in S1 per year (Fig. 5). The S1 domain of 229E is estimated to
 207 accrue 0.26 adaptive substitutions per year. A benefit of the Bhatt method is the ability to
 208 calculate the strength of selection, which allows us to compare these seasonal HCoVs to other
 209 viruses. We used our implementation of the Bhatt method to calculate the rate of adaptation for
 210 influenza H3N2, which is known to undergo rapid antigenic evolution, and measles, which does
 211 not. We estimate that the receptor-binding domain of influenza H3N2 accumulates adaptive
 212 substitutions about 3 times faster than the HCoVs OC43 and 229E (Fig. 6). We detect no
 213 adaptive substitutions in the measles receptor-binding protein. These results put the evolution of
 214 the S1 domain of OC43 and 229E in context, indicating that the S1 domain is under positive

215 selection, and that this positive selection generates new variants in the putative antigenic
216 regions of these HCoVs at about a third of the rate of the canonical example of antigenic
217 evolution, the HA1 domain of influenza H3N2.

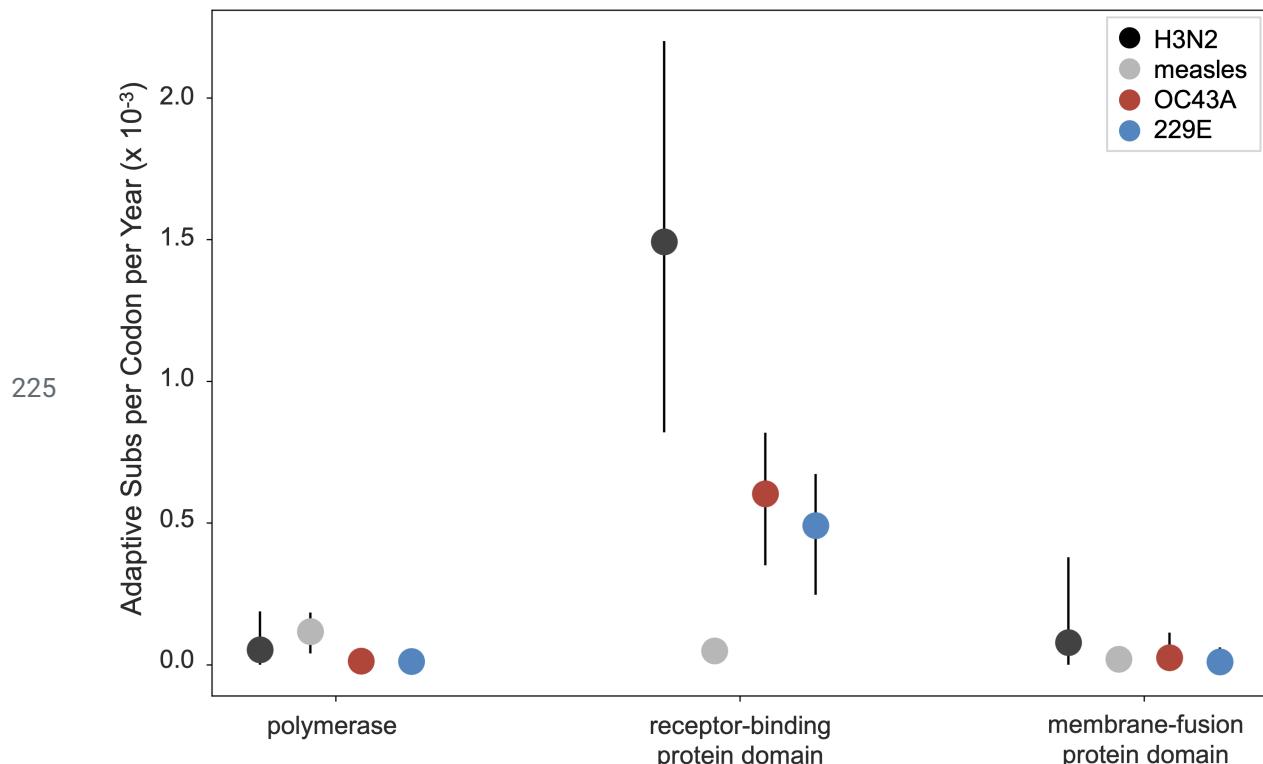
218

219



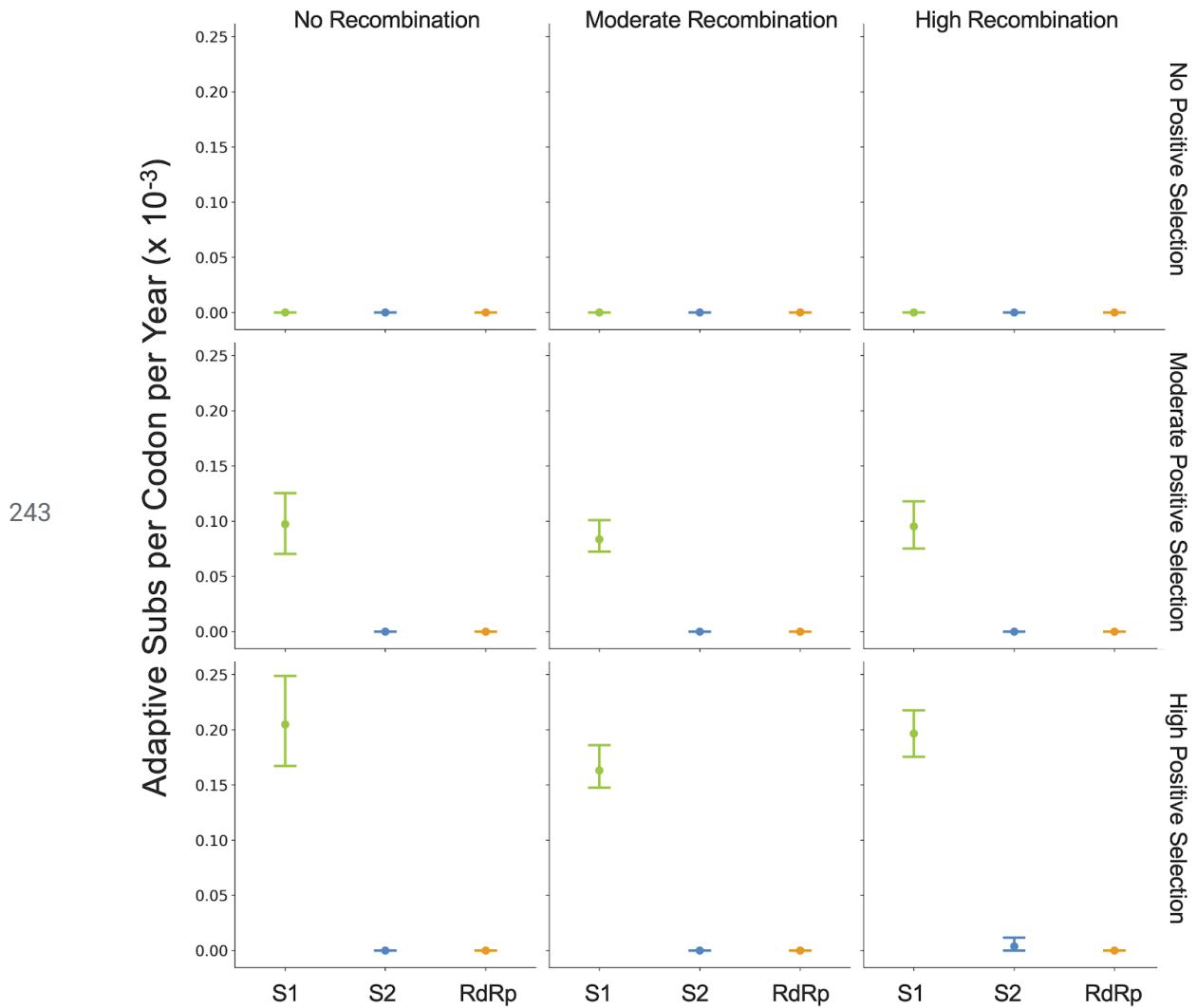
220 **Figure 5. The rate of adaptive substitution is highest in spike S1.** Adaptive substitutions per codon
221 per year as calculated by our implementation of the Bhatt method. Rates are calculated within Spike, S1,
222 S2 and RdRp for 229E and OC43 lineages. Error bars show 95% bootstrap percentiles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

224



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

233
234 Because coronaviruses are known to recombine, and recombination has the potential to impact
235 evolutionary analyses of selection, we sought to verify that our results are not swayed by the
236 presence of recombination. To do this, we simulated the evolution of OC43 lineage A spike and
237 RdRp genes under varying levels of recombination and positive selection and used our
238 implementation of the Bhatt method to identify adaptive substitutions. As the strength of positive
239 selection increases, we detect more adaptive substitutions, regardless of the level of
240 recombination (Fig. 7). This demonstrates that our estimates of adaptive evolution are not
241 biased by recombination events.
242



244 **Figure 7. Detection of positive selection is not biased by recombination.** OC43 lineage A sequences
 245 were simulated with varying levels of recombination and positive selection. The Bhatt method was used to
 246 calculate the rate of adaptive substitutions per codon per year for S1 (light green), S2 (blue) and RdRp
 247 (orange). The mean and 95% confidence interval of 5 independent simulations is plotted.
 248

249 Finally, we know that strong directional selection skews the shape of phylogenies (Volz, Koelle,
 250 and Bedford 2013). In influenza H3N2, immune selection causes the genealogy to adopt a
 251 ladder-like shape where the rungs are formed by viral diversification and each step is created by
 252 the appearance of new, antigenically-superior variants that replace previous variants. This
 253 ladder-like shape can also be seen in the phylogenies of the OC43 and 229E (Fig. 1). In this
 254 case, selection can be quantified by the timescale of population turnover as measured by the
 255 Time to Most Recent Common Ancestor (TMRCA), with the expectation that stronger selection
 256 will result in more frequent steps and therefore a smaller TMRCA measure (Bedford, Cobey,

257 and Pascual 2011). We computed average TMRCA values from phylogenies built on Spike, S1,
258 S2 or RdRp sequences of OC43 and 229E (Table 1). We observe that, for both OC43 lineage A
259 and 229E, the average TMRCA is lower in spike than RdRp and lower in S1 versus S2. These
260 results suggest strong directional selection in S1, likely driven by pressures to evade the
261 humoral immune system. The difference in TMRCA between S1 and S2 is indicative not only of
262 differing selective pressures acting on these two spike domains, but also of intra-spike
263 recombination, which emphasizes the importance of using methods that are robust to
264 recombination to detect adaptive evolution.

265

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

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

268

269 Because HKU1 was identified in the early 2000's, there are fewer longitudinally-sequenced
270 isolates available for this HCoV compared to 229E and OC43 (Fig. 1 Supplement 2).
271 Consequently, the phylogenetic reconstructions and divergence analysis of HKU1 have a higher
272 level of uncertainty. To begin with, it is less clear from the phylogenies whether HKU1
273 represents a single HCoV lineage like 229E or, instead, should be split into multiple lineages like
274 OC43 (Fig. 1). Because of this, we completed all antigenic analyses for HKU1 twice: once
275 considering all isolates to be members of a single lineage, and again after splitting isolates into
276 2 separate lineages. These lineages are arbitrarily labeled 'A' and 'B' as was done for OC43.
277 When HKU1 is considered to consist of just one lineage, there is no signal of antigenic evolution
278 by divergence analysis (Fig. 3 Supplement 1B) or by the Bhatt method of estimating adaptive
279 evolution (Fig. 5 Supplement 1A). However, when HKU1 is assumed to consist of 2
280 co-circulating lineages, HKU1 lineage A has a markedly higher rate of adaptive substitutions in
281 S1 than in S2 or RdRp (Fig. 5 Supplement 1B).

282

283 To demonstrate the importance of having a well-sampled longitudinal series of sequenced
284 isolates for our antigenic analyses, we returned to our simulated OC43 S1 datasets. We
285 mimicked shorter longitudinal series by truncating the dataset to only 24, 14, 10, or 7 years of
286 samples and ran the Bhatt analysis on these sequentially shorter time series (Fig. 7
287 Supplement). This simulated data reveals a general trend that less longitudinal data reduces the
288 ability to detect adaptive evolution and increases the uncertainty of the analysis. Given the
289 dearth of longitudinal data for HKU1, we do not feel that it is appropriate to make strong
290 conclusions about adaptive evolution, or lack thereof, in this HCoV.

291

292 Despite being identified at roughly the same time as HKU1, substantially more NL63 isolates
293 have been sequenced (Fig. 1 Supplement 2) making the phylogenetic reconstruction and
294 evolutionary analyses of this virus correspondingly more reliable. We do not observe evidence

295 for adaptive evolution in NL63 (Fig. 3 Supplement 1A and Fig. 5 Supplement 1A) and this lack
296 of support for adaptive evolution in the NL63 spike gene is more likely to reflect an actual lack of
297 adaptive evolution in this virus.

298

299 Discussion

300 Using several corroborating methods, we provide evidence that the seasonal HCoVs OC43 and
301 229E undergo adaptive evolution in S1, the region of the spike protein exposed to human
302 humoral immunity (Figs. 3, 4 and 5). We additionally confirm that RdRp and S2 do not show
303 signals of adaptive evolution. We observe that S1 accumulates between 0.3 (229E) and 0.5
304 (OC43) adaptive substitutions per year. We infer that these viruses accumulate adaptive
305 substitutions at roughly a third of the rate of influenza H3N2 (Fig. 6). The most parsimonious
306 explanation for the observation of substantial adaptive evolution in S1 is that antigenic drift is
307 occurring in which mutations that escape from human population immunity are selectively
308 favored in the viral population leading to repeated adaptive changes. However, it is formally
309 possible that the adaptive evolution we detect is a result of selective pressures other than
310 evasion of the adaptive immune system. Showing that this is truly antigenic evolution could
311 involve a serological comparison of isolates that differ at S1 residues under positive selection.
312 We do not observe evidence of antigenic evolution in NL63 or HKU1 (Figs. 3 and 5
313 Supplements). For NL63, this likely represents a true lack of marked adaptive evolution in S1.
314 There is much less longitudinal sequencing data available for HKU1 and it is possible that a
315 more completely sampled time series of genome sequences could alter the result for this virus
316 (Fig. 7 Supplement 1).

317

318 Our conclusions of adaptive evolution in S1, arrived at through computational analyses of
319 sequencing data, agree with studies that observe reinfection of subjects by heterologous
320 isolates of 229E (Reed 1984), sequential dominance of specific genotypes of OC43 (Lau et al.
321 2011; Zhang et al. 2015), and common reinfection by seasonal HCoVs from longitudinal
322 serological data (Edridge et al. 2020). In this latter study, HCoV infections were identified from
323 longitudinal serum samples by assaying for increases in antibodies against the nucleocapsid (N)
324 protein of representative OC43, 229E, HKU1, and NL63 viruses. This study concluded that the
325 average time between infections was 1.5–2.5 years, depending on the HCoV (Edridge et al.
326 2020). In comparison, influenza H3N2 reinfects people roughly every 5 years (Kucharski et al.
327 2015). Thus, frequent reinfection by seasonal HCoVs is likely due to a combination of factors
328 and suggests waning immune memory, and/or incomplete immunity against reinfection, in
329 addition to antigenic drift.

330

331 Human coronaviruses are a diverse grouping split, phylogenetically, into two genera: NL63 and
332 229E are alphacoronaviruses, while OC43, HKU1, MERS, SARS, and SARS-CoV-2 are
333 betacoronaviruses. Transmissibility and pathology do not seem to correlate with genus, nor
334 does the method of cell-entry. Coronaviruses bind to a remarkable range of host-cell receptors
335 including peptidases, cell adhesion molecules and sugars. Amongst the seasonal HCoVs, OC43
336 and HKU1 both bind 9-O-acetylsialic acid (Hulswit et al. 2019) while 229E binds human
337 aminopeptidase N (hAPN) and NL63 binds angiotensin-converting enzyme 2 (ACE2) (Liu,

338 Liang, and Fung 2020). Despite a relatively large phylogenetic distance and divergent S1
339 structures, NL63 and SARS-CoV-1 and SARS-CoV-2 bind to the same host receptor using the
340 same virus-binding motifs (VBMs) (Li 2016). This VBM is located in the C-terminal domain of S1
341 (S1-CTD), which fits within the trend of S1-CTD receptor-binding in CoVs that bind protein
342 receptors (Hofmann et al. 2006; Li 2016). This is opposed to the trend amongst CoVs that bind
343 sugar receptors, where receptor-binding is located within the S1 N-terminal domain (S1-NTD)
344 (Li 2016). This localization roughly aligns with our observations that the majority of the
345 repeatedly-mutated sites occur toward the C-terminal end of 229E S1 and the N-terminal end of
346 OC43 S1 (Fig. 2).

347

348 Here, we have provided support that at least 2 of the 4 seasonal HCoVs evolve adaptively in the
349 region of spike that is known to interact with the humoral immune system. These two viruses
350 span both genera of HCoVs, though due to the complexity of HCoV receptor-binding and
351 pathology mentioned above, it is not clear whether or not this suggests that other HCoVs, such
352 as SARS-CoV-2, will also evolve adaptively in S1. This is important because, at the time of
353 writing of this manuscript, many SARS-CoV-2 vaccines are in production and most of these
354 exclusively include spike (Krammer 2020). If SARS-CoV-2 evolves adaptively in S1 as the
355 closely-related HCoV OC43 does, it is possible that the SARS-CoV-2 vaccine would need to be
356 frequently reformulated to match the circulating strains, as is done for seasonal influenza
357 vaccines.

358

359 Materials and methods

360 All data, source code and analyses can be found at
361 <https://github.com/blab/seasonal-cov-adaptive-evolution>. All phylogenetic trees constructed and
362 analyzed in this manuscript can be viewed interactively at
363 <https://nextstrain.org/community/blab/seasonal-cov-adaptive-evolution>.

364

365 Sequence data

366 All viral sequences are publicly accessible and were downloaded from ViPR (www.viprbrc.org)
367 under the “Coronaviridae” with host “human” (Pickett et al. 2012). Sequences labeled as
368 “OC43”, “229E”, “HKU1” and “NL63” were pulled out of the downloaded FASTA file into 4
369 separate data files. Additionally, a phylogeny of all downloaded human coronaviruses was made
370 and unlabeled isolates that clustered within clades formed by labeled OC43, 229E, HKU1 or
371 NL63 isolates were marked as belonging to that HCoV type and added to our data files. Code
372 for these data-parsing steps is located in

373 `data-wrangling/postdownload_formatting_for_rerun.ipynb`.

374

375 Phylogenetic inference

376 For each of the 4 HCoV datasets, full-length sequences were aligned to a reference genome
377 using the augur align command (Hadfield et al. 2018) and MAFFT (Katoh et al. 2002). Individual
378 gene sequences were then extracted from these alignments if sequencing covered 50% or more
379 of the gene using the code in

380 `data-wrangling/postdownload_formatting_for_rerun.ipynb`. Sequence files for

381 each gene are located in the `data/` directory within each HCoV parent directory (ex:
382 `oc43/data/oc43_spike.fasta`). A Snakemake file (Köster and Rahmann 2012) within each
383 HCoV directory was then used to aligned each gene to a reference strain and a time-resolved
384 phylogeny was built with IG-Tree (Nguyen et al. 2015) and TimeTree (Sagulenko, Puller, and
385 Neher 2018). Phylogenies were viewed to identify the distribution of genotypes throughout the
386 tree, different lineages, and signals of recombination using the nextstrain view command
387 (Hadfield et al. 2018). The clock rate of the phylogeny based on spike sequences for each
388 isolate (as shown in Fig. 1 and Fig. 1 Supplement 2) was 0.0005 for OC43, 0.0006 for 229E,
389 0.0007 for NL63, and 0.0062 for HKU1. All NL63 and HKU1 trees were rooted on an outgroup
390 sequence. For NL63, the outgroup was 229e/AF304460/229e_ref/Germany/2000 and for HKU1
391 the outgroup was mhv/NC_048217_1/mhv/2006. Clock rates for the phylogenies built on each
392 individual gene can be found within the `results/` directory within each HCoV parent directory
393 (ex: `oc43/results/branch_lengths_oc43_spike.json`).
394

395 **Mutation counting**

396 Amino acid substitutions at each position in spike were tallied from the phylogeny using code in
397 `antigenic_evolution/site_mutation_rank.ipynb`.
398

399 **Divergence analysis**

400 For each HCoV lineage and each gene, synonymous and nonsynonymous divergence was
401 calculated at all timepoints as the average Hamming distance between each sequenced isolate
402 and the consensus sequence at the first timepoint (founder sequence). The total number of
403 observed differences between the isolate and founder nucleotide sequences that result in
404 nonsynonymous (or synonymous) substitutions is divided by the number of possible nucleotide
405 mutations that result in nonsynonymous (or synonymous) substitutions, weighted by kappa, to
406 yield an estimate of divergence. Kappa is the ratio of rates of transitions:transversions, and was
407 calculated by averaging values from spike and RdRp trees built by BEAST 2.6.3 (Bouckaert et
408 al. 2019) using the HKY+gamma4 model with 2 partitions and “coalescent constant population”.
409 All BEAST results are found in `.log` files in gene- and HCoV-specific subdirectories within
410 `beast/`. Divergence is calculated from nucleotide alignments. Sliding 3-year windows were
411 used and only timepoints that contained at least 2 sequences were considered. The concept for
412 this analysis is from (Zanini et al. 2015) and code for our adaptation is in
413 `antigenic_evolution/divergence_weighted.ipynb`.
414

415 **Implementation of the Bhatt method**

416 The rate of adaptive evolution was computed using an adaptation of the Bhatt method (Bhatt,
417 Holmes, and Pybus 2011; Bhatt, Katzourakis, and Pybus 2010). Briefly, this method defines a
418 class of neutrally-evolving nucleotide sites, then identifies other classes with higher rates of
419 nonsynonymous nucleotide fixations and high-frequency polymorphisms. This method
420 compares nucleotide sequences at each timepoint (the ingroup) to the consensus nucleotide
421 sequence at the first time point (the outgroup) and yields an estimate of the number of adaptive
422 substitutions within a given genomic region at each of these timepoints. Eight estimators (silent
423 fixed, replacement fixed, silent high frequency, replacement high frequency, silent

424 mid-frequency, replacement mid-frequency, silent low frequency and replacement
425 low-frequency) are then calculated by the site-counting method (Bhatt, Katzourakis, and Pybus
426 2010). In the site-counting method, each estimator is the product of the fixation or polymorphism
427 score times the silent or replacement score, summed for each site in that frequency class.
428 Fixation and polymorphism scores depend on the number of different nucleotides observed at
429 the site and whether the outgroup base is present in the ingroup. Selectively neutral sites are
430 assumed to contain the classes of silent polymorphisms and replacement polymorphisms
431 occurring at a frequency between 0.15 and 0.75. A class of nonneutral, adaptive sites is then
432 identified as having an excess of replacement fixations or polymorphisms (Bhatt, Holmes, and
433 Pybus 2011). Sliding 3-year windows were used and only timepoints that contained at least 2
434 sequences were considered. For each lineage and gene, 100 bootstrap alignments and
435 ancestral sequences were generated and run through the Bhatt method to assess the statistical
436 uncertainty of our estimates of rates of adaptation (Bhatt, Holmes, and Pybus 2011). The rate of
437 adaptation (per codon per year) shown in Fig. 5 is calculated by linear regression of the time
438 series values of adaptive substitutions per codon (Fig. 4). Our code for implementing the Bhatt
439 method is at `antigenic_evolution/bhatt_bootstrapping.ipynb`.
440

441 **Estimation of rates of adaptation of H3N2 and measles**

442 Influenza H3N2 and measles sequencing data was downloaded from
443 <https://github.com/nextstrain/seasonal-flu> and <https://github.com/nextstrain/measles>,
444 respectively. The rates of adaptation of different genes was calculated using our implementation
445 of the Bhatt method described above. The receptor-binding domain used for H3N2 was HA1, for
446 measles was the H protein, and for the HCoVs was S1. The membrane fusion protein used for
447 H3N2 was HA2, for measles was the F protein, and for the HCoVs was S2. The polymerase for
448 H3N2 was PB1, for measles was the P protein, and for the HCoVs was RbRd (nsp12). Our code
449 for this analysis is at `antigenic_evolution/measles_h3n2_bhatt.ipynb`.
450

451 **Simulation of evolving OC43 sequences**

452 The evolution of OC43 lineage A Spike and RdRp genes was simulated using SANTA-SIM
453 (Jariani et al. 2019). The OC43 lineage A root sequence was used as a starting point and the
454 simulation was run for 500 generations and 10 simulated sequences were sampled every 50
455 generations. Evolution was simulated in the absence of recombination and with moderate and
456 high levels of recombination during replication. Under each of these recombination paradigms,
457 we simulated evolution in the absence of positive selection within spike and with moderate and
458 high levels of positive selection. Positive selection was simulated at a subset of Spike sites
459 proportional to the number of epitope sites in H3N2 HA (Luksza and Lässig 2014). All
460 simulations were run with a nucleotide mutation rate of 1x10-4 (Vijgen et al. 2005). Config files,
461 results and source code for these simulations can be at `santa-sim_oc43a/`.
462

463 **Estimation of TMRCA**

464 Mean TMRCA values were estimated for each gene and each HCoV using PACT (Bedford,
465 Cobey, and Pascual 2011). The PACT config files and results for each run are in the directory

466 antigenic_evolution/pact/. The TMRCA estimations and subsequent analyses are
467 executed by code in antigenic_evolution/tmrca_pact.ipynb.

468

469 Acknowledgments

470 We thank Jesse Bloom and members of the Bedford lab for useful feedback. KEK was
471 supported by the National Science Foundation Graduate Research Fellowship Program under
472 Grant No. DGE-1762114. TB is a Pew Biomedical Scholar and is supported by NIH R35
473 GM119774-01.

474

475 Competing Interests

476 The authors declare no competing interests.

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

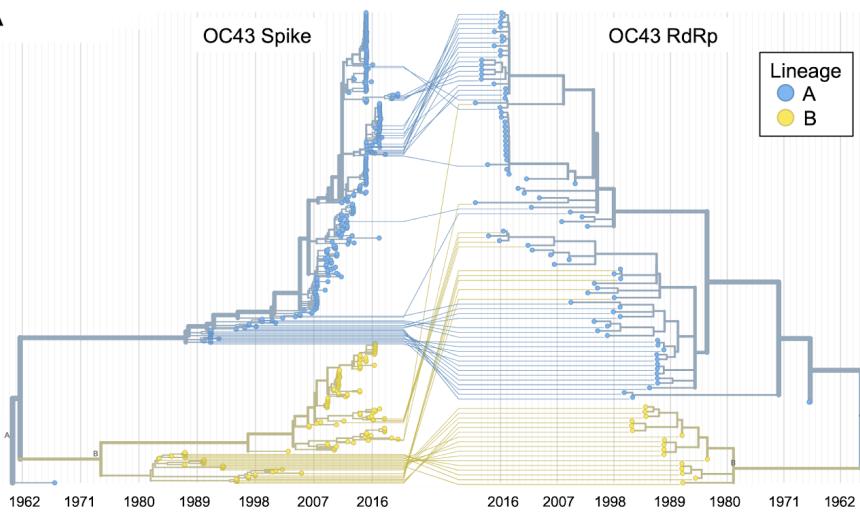
505

506

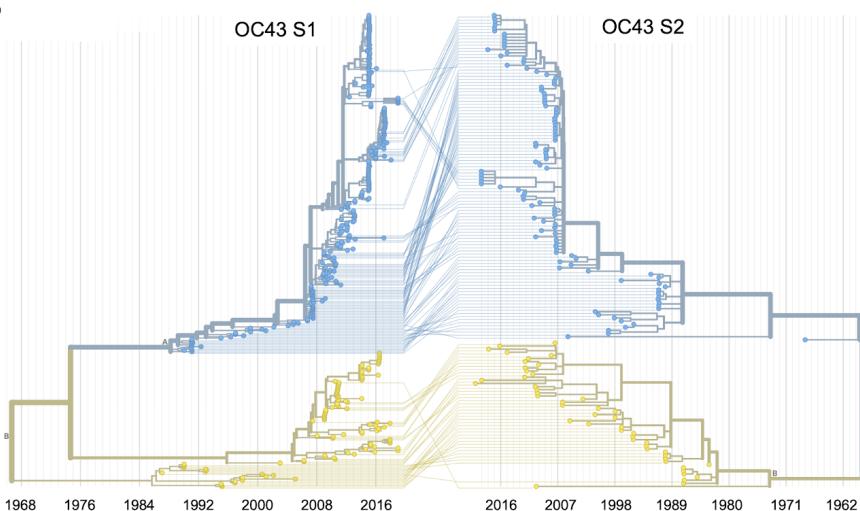
507

508 **Supplemental Figures**

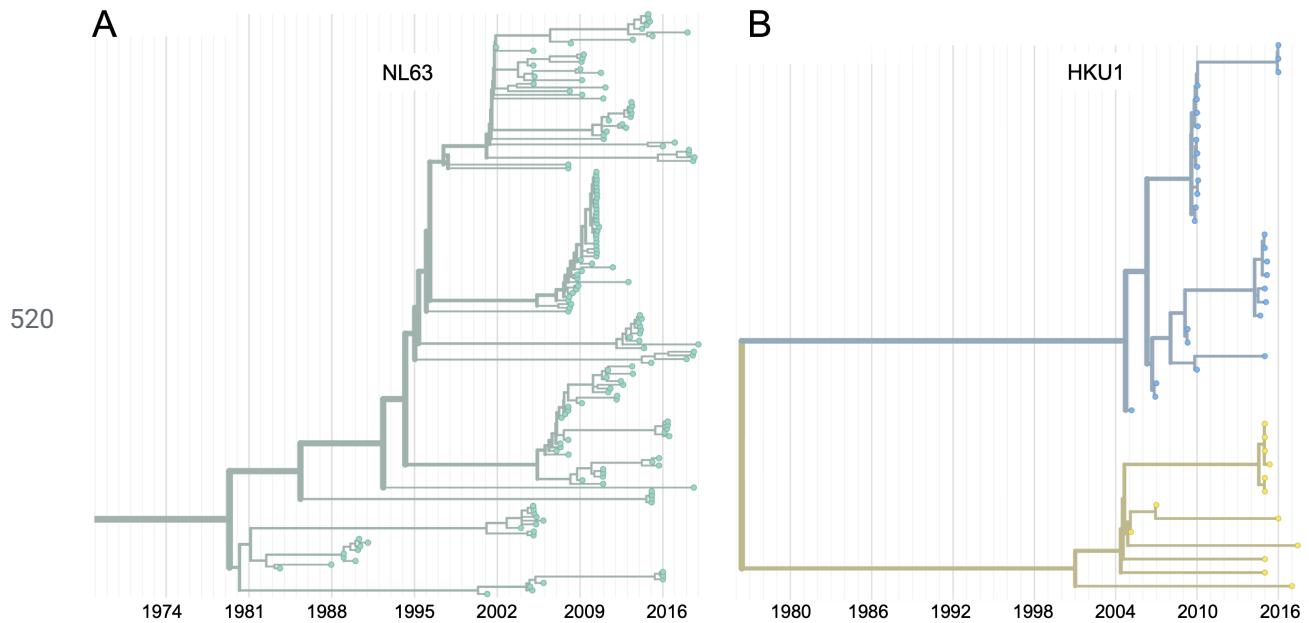
509

A

510

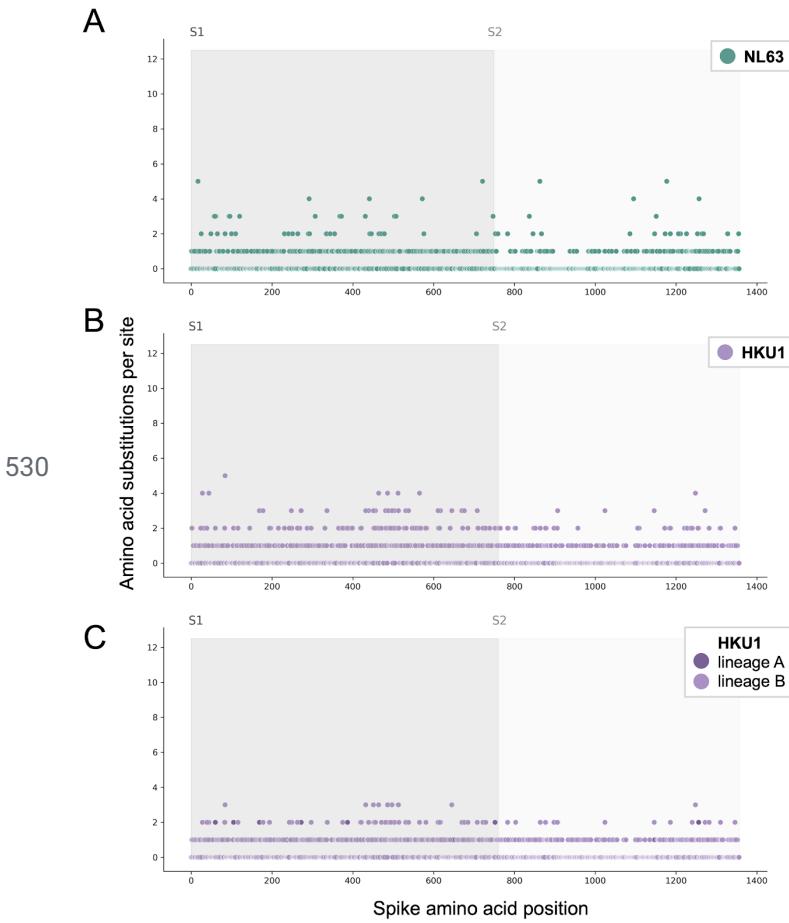
B

511 **Figure 1 Supplement 1. Recombination occurs between HCoV isolates.** A tanglegram draws lines
 512 between an isolate's position on two phylogenies built on different genes (or genomic regions). Dramatic
 513 differences in an isolate's position on one tree versus another is indicative of recombination. A)
 514 Phylogenetic relationships between OC43 isolates based on RdRp sequences versus relationships based
 515 on Spike sequences. Blue lines that connect isolates classified as lineage A based on their RdRp
 516 sequence to isolates classified as lineage B based on their Spike sequence, suggest that recombination
 517 occurred in these isolates or their ancestors. B) Phylogenetic reconstruction of OC43 isolates based on
 518 S1 sequences versus S2 sequences. Year is shown on the x-axis.
 519

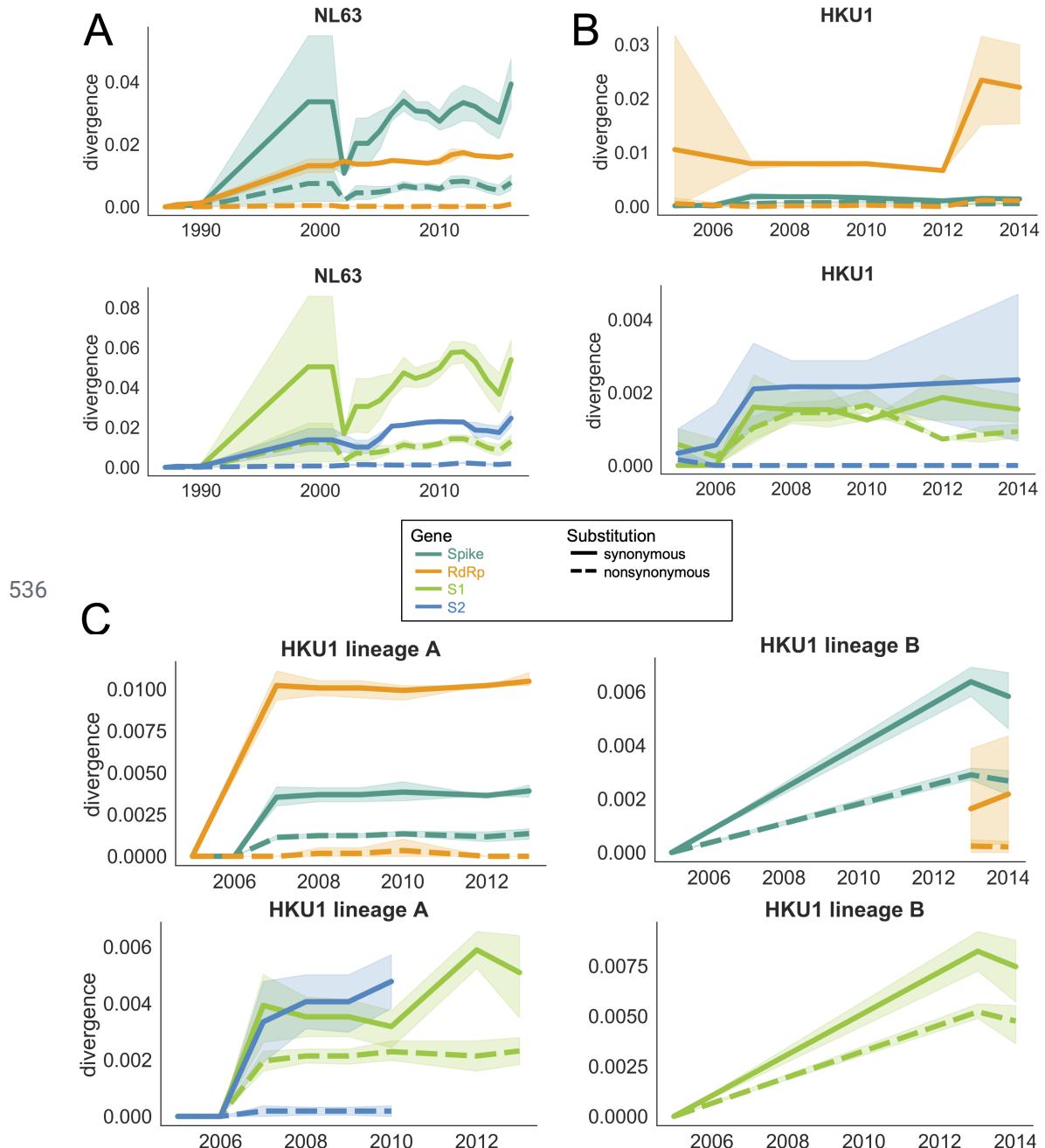


521 **Figure 1 Supplement 2. Phylogenetic trees for seasonal HCoVs NL63 and HKU1.** Phylogenies built
 522 from A: NL63 spike sequences from 159 isolates over 37 years, and B: HKU1 spike sequences from 41
 523 isolates over 13 years. HCoVs that bifurcate immediately after the root are split into blue and yellow
 524 lineages. NL63 contains just one lineage (teal). Both HCoVs are rooted on an outgroup sequence. For the
 525 analyses in this paper, the evolution of each gene (or genomic region) is considered separately, so
 526 phylogenies are built for each viral gene and those phylogenies are used to split isolates into lineages for
 527 each gene. These are temporally resolved phylogenies with year shown on the x-axis. The clock rate of
 528 each HCoV is listed in the Methods “Phylogenetic inference” section.

529



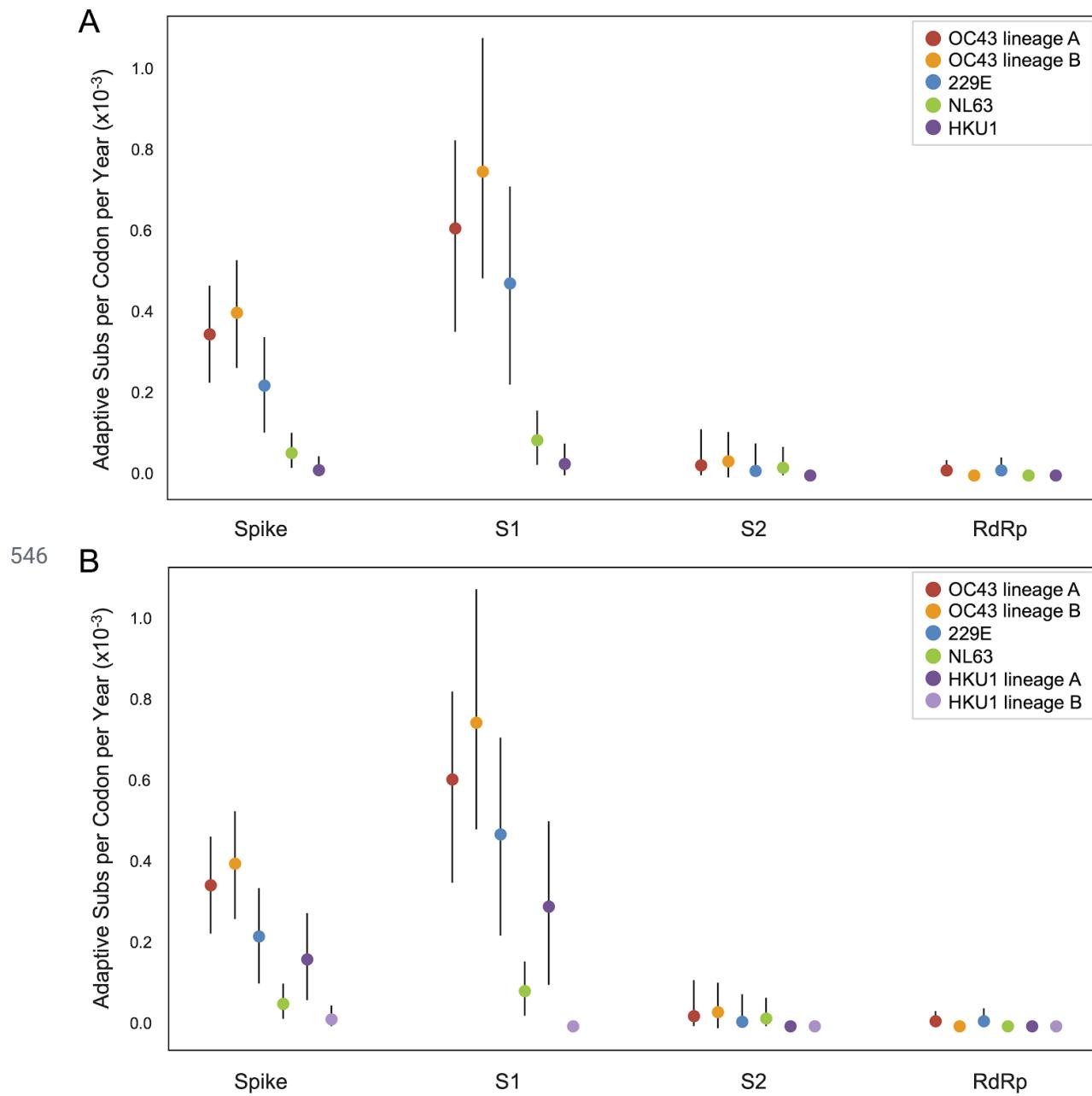
531 **Figure 2 Supplement 1. Mutations per at each position within Spike for NL63 and HKU1.** Number of
 532 mutations observed at each position in the Spike gene. S1 (darker gray) and S2 (light gray) are indicated
 533 by shading and the average number of mutations per site is indicated by a dot and color-coded by HCoV
 534 lineage. A: NL63, B: HKU1 (assuming all HKU1 isolates are a single lineage), C: HKU1 (assuming there
 535 are 2 co-circulating HKU1 lineages).



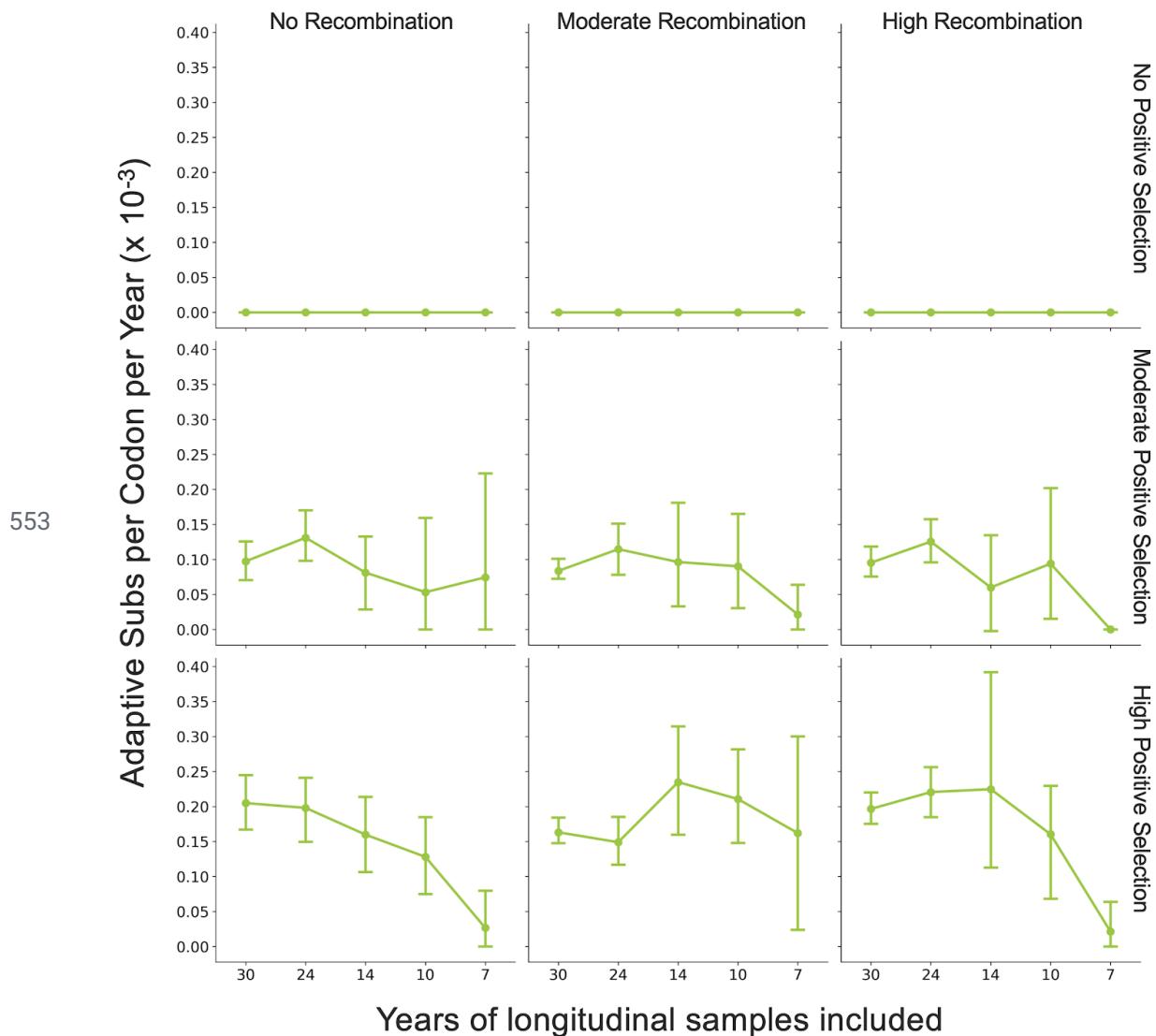
537 **Figure 3 Supplement 1. Nonsynonymous divergence in NL63 and HKU1.** Nonsynonymous (dashed
 538 lines) and synonymous divergence (solid lines) within the Spike (teal) and RdRp (orange) genes and
 539 within S1 (light green) and S2 (blue) over time. Divergence is the average Hamming distance from the
 540 ancestral sequence, computed in sliding 3-year windows which contain at least 2 sequenced isolates.
 541 Shaded region shows 95% confidence intervals. A: NL63, B: HKU1 (assuming all HKU1 isolates belong
 542 to a single lineage), and C: HKU1 (divided into 2 co-circulating lineages). Year is shown on the x-axis.
 543 Note that x- and y-axis scales are not shared between plots.

544

545



547 **Figure 5 Supplement 1. NL63 and HKU1 have low rates of adaptation in Spike.** As in Figure 4,
548 adaptive substitutions per codon per year are calculated by our implementation of the Bhatt method. A:
549 NL63 (teal) and HKU1 (purple) are both considered to consist of a single lineage. B: HKU1 is divided into
550 2 co-circulating lineages (dark and light purple). The calculated rates of adaptive substitution within Spike,
551 S1, S2 and RdRp are plotted alongside 229E and OC43 for comparison. Error bars show 95% bootstrap
552 percentiles from 100 bootstrapped datasets



554 **Figure 7 Supplement 1. Fewer years of longitudinally-sampled isolates reduces ability to detect
555 rate of adaptation.** OC43 lineage A S1 sequences were simulated under conditions of no, moderate and
556 high rates of recombination in combination with no, moderate or high strength of positive selection. The
557 Bhatt method was used to calculate the rate of adaptive evolution under each of these scenarios using all
558 available sequence data (30 years), or only the most recent 24, 14, 10 or 7 years of simulated
559 sequences.

560

561 References

- 562 Bedford, Trevor, Sarah Cobey, and Mercedes Pascual. 2011. "Strength and Tempo of Selection
563 Revealed in Viral Gene Genealogies."
- 564 Bhatt, Samir, Edward C. Holmes, and Oliver G. Pybus. 2011. "The Genomic Rate of Molecular
565 Adaptation of the Human Influenza A Virus." *Molecular Biology and Evolution* 28 (9):
566 2443–51.

- 567 Bhatt, Samir, Aris Katzourakis, and Oliver G. Pybus. 2010. "Detecting Natural Selection in RNA
568 Virus Populations Using Sequence Summary Statistics." *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 10 (3):
569 421–30.
- 570 Bouckaert, Remco, Timothy G. Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu
571 Fourment, Alexandra Gavryushkina, Joseph Heled, et al. 2019. "BEAST 2.5: An Advanced
572 Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 15 (4):
573 e1006650.
- 574 Chibo, Doris, and Chris Birch. 2006. "Analysis of Human Coronavirus 229E Spike and
575 Nucleoprotein Genes Demonstrates Genetic Drift between Chronologically Distinct Strains." *The Journal of General Virology* 87 (Pt 5): 1203–8.
- 576 Drake, J. W. 1993. "Rates of Spontaneous Mutation among RNA Viruses." *Proceedings of the National Academy of Sciences of the United States of America* 90 (9): 4171–75.
- 577 Edridge, Arthur W. D., Joanna Kaczorowska, Alexis C. R. Hoste, Margreet Bakker, Michelle
578 Klein, Katherine Loens, Maarten F. Jebbink, et al. 2020. "Seasonal Coronavirus Protective
579 Immunity Is Short-Lasting." *Nature Medicine*, September.
580 https://doi.org/10.1038/s41591-020-1083-1.
- 581 Fulton, Benjamin O., David Sachs, Shannon M. Beaty, Sohui T. Won, Benhur Lee, Peter Palese,
582 and Nicholas S. Heaton. 2015. "Mutational Analysis of Measles Virus Suggests Constraints
583 on Antigenic Variation of the Glycoproteins." *Cell Reports* 11 (9): 1331–38.
- 584 Hadfield, James, Colin Megill, Sidney M. Bell, John Huddleston, Barney Potter, Charlton
585 Callender, Pavel Sagulenko, Trevor Bedford, and Richard A. Neher. 2018. "Nextstrain:
586 Real-Time Tracking of Pathogen Evolution." *Bioinformatics* 34 (23): 4121–23.
- 587 Hamre, D., and M. Beem. 1972. "Virologic Studies of Acute Respiratory Disease in Young
588 Adults. V. Coronavirus 229E Infections during Six Years of Surveillance." *American Journal
589 of Epidemiology* 96 (2): 94–106.
- 590 Heikkinen, Terho, and Asko Järvinen. 2003. "The Common Cold." *The Lancet* 361 (9351):
591 51–59.
- 592 Hofmann, Heike, Graham Simmons, Andrew J. Rennekamp, Chawaree Chaipan, Thomas
593 Gramberg, Elke Heck, Martina Geier, et al. 2006. "Highly Conserved Regions within the
594 Spike Proteins of Human Coronaviruses 229E and NL63 Determine Recognition of Their
595 Respective Cellular Receptors." *Journal of Virology* 80 (17): 8639–52.
- 596 Hon, Chung-Chau, Tsan-Yuk Lam, Zheng-Li Shi, Alexei J. Drummond, Chi-Wai Yip, Fanya
597 Zeng, Pui-Yi Lam, and Frederick Chi-Ching Leung. 2008. "Evidence of the Recombinant
598 Origin of a Bat Severe Acute Respiratory Syndrome (SARS)-like Coronavirus and Its
599 Implications on the Direct Ancestor of SARS Coronavirus." *Journal of Virology* 82 (4):
600 1819–26.
- 601 Hulswit, Ruben J. G., Yifei Lang, Mark J. G. Bakkers, Wentao Li, Zeshi Li, Arie Schouten, Bram
602 Ophorst, et al. 2019. "Human Coronaviruses OC43 and HKU1 Bind to 9-O-Acetylated Sialic
603 Acids via a Conserved Receptor-Binding Site in Spike Protein Domain A." *Proceedings of the National Academy of Sciences of the United States of America* 116 (7): 2681–90.
- 604 Jariani, Abbas, Christopher Warth, Koen Deforche, Pieter Libin, Alexei J. Drummond, Andrew
605 Rambaut, Frederick A. Matsen Iv, and Kristof Theys. 2019. "SANTA-SIM: Simulating Viral
606 Sequence Evolution Dynamics under Selection and Recombination." *Virus Evolution* 5 (1):
607 vez003.
- 608 Katoh, Kazutaka, Kazuharu Misawa, Kei-Ichi Kuma, and Takashi Miyata. 2002. "MAFFT: A
609 Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform."
610 *Nucleic Acids Research* 30 (14): 3059–66.

- 615 Komabayashi, Kenichi, Yohei Matoba, Shizuka Tanaka, Junji Seto, Yoko Aoki, Tatsuya Ikeda,
616 Yoshitaka Shimotai, Yoko Matsuzaki, Tsutomu Itagaki, and Katsumi Mizuta. 2020.
617 "Longitudinal Epidemiology of Human Coronavirus OC43 in Yamagata, Japan, 2010–2017:
618 Two Groups Based on Spike Gene Appear One after Another." *Journal of Medical Virology*
619 7 (August): 825.
- 620 Kosakovsky Pond, Sergei L., David Posada, Michael B. Gravenor, Christopher H. Woelk, and
621 Simon D. W. Frost. 2006. "Automated Phylogenetic Detection of Recombination Using a
622 Genetic Algorithm." *Molecular Biology and Evolution* 23 (10): 1891–1901.
- 623 Köster, Johannes, and Sven Rahmann. 2012. "Snakemake—a Scalable Bioinformatics
624 Workflow Engine." *Bioinformatics* 28 (19): 2520–22.
- 625 Krammer, Florian. 2020. "SARS-CoV-2 Vaccines in Development." *Nature*, September.
626 <https://doi.org/10.1038/s41586-020-2798-3>.
- 627 Kucharski, Adam J., Justin Lessler, Jonathan M. Read, Huachen Zhu, Chao Qiang Jiang, Yi
628 Guan, Derek A. T. Cummings, and Steven Riley. 2015. "Estimating the Life Course of
629 Influenza A(H3N2) Antibody Responses from Cross-Sectional Data." *PLOS Biology*.
630 <https://doi.org/10.1371/journal.pbio.1002082>.
- 631 Lau, Susanna K. P., Paul Lee, Alan K. L. Tsang, Cyril C. Y. Yip, Herman Tse, Rodney A. Lee,
632 Lok-Yee So, et al. 2011. "Molecular Epidemiology of Human Coronavirus OC43 Reveals
633 Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype
634 due to Natural Recombination." *Journal of Virology* 85 (21): 11325–37.
- 635 Li, Fang. 2016. "Structure, Function, and Evolution of Coronavirus Spike Proteins." *Annual
636 Review of Virology* 3 (1): 237–61.
- 637 Liu, Ding X., Jia Q. Liang, and To S. Fung. 2020. "Human Coronavirus-229E, -OC43, -NL63,
638 and -HKU1." *Reference Module in Life Sciences*.
639 <https://doi.org/10.1016/b978-0-12-809633-8.21501-x>.
- 640 Luksza, Marta, and Michael Lässig. 2014. "A Predictive Fitness Model for Influenza." *Nature* 507
641 (7490): 57–61.
- 642 McIntosh, Kenneth. 1974. "Coronaviruses: A Comparative Review." In *Current Topics in
643 Microbiology and Immunology / Ergebnisse Der Mikrobiologie Und Immunitätsforschung*,
644 85–129. Springer Berlin Heidelberg.
- 645 Monto, Arnold S., and Sook K. Lim. 1974. "The Tecumseh Study of Respiratory Illness. VI.
646 Frequency of and Relationship between Outbreaks of Coronavims Infection." *The Journal
647 of Infectious Diseases* 129 (3): 271–76.
- 648 Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015.
649 "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood
650 Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74.
- 651 Pasternak, Alexander O., Willy J. M. Spaan, and Eric J. Snijder. 2006. "Nidovirus Transcription:
652 How to Make Sense...?" *The Journal of General Virology* 87 (6): 1403–21.
- 653 Pickett, Brett E., Eva L. Sadat, Yun Zhang, Jyothi M. Noronha, R. Burke Squires, Victoria Hunt,
654 Mengya Liu, et al. 2012. "ViPR: An Open Bioinformatics Database and Analysis Resource
655 for Virology Research." *Nucleic Acids Research* 40 (Database issue): D593–98.
- 656 Rambaut, Andrew, Oliver G. Pybus, Martha I. Nelson, Cecile Viboud, Jeffery K. Taubenberger,
657 and Edward C. Holmes. 2008. "The Genomic and Epidemiological Dynamics of Human
658 Influenza A Virus." *Nature* 453 (7195): 615–19.
- 659 Reed, Sylvia E. 1984. "The Behaviour of Recent Isolates of Human Respiratory Coronavirus in
660 Vitro and in Volunteers: Evidence of Heterogeneity among 229E-Related Strains." *Journal
661 of Medical Virology* 13 (2): 179–92.
- 662 Ren, Lili, Yue Zhang, Jianguo Li, Yan Xiao, Jing Zhang, Ying Wang, Lan Chen, Gláucia

- 663 Paranhos-Baccalà, and Jianwei Wang. 2015. "Genetic Drift of Human Coronavirus OC43
664 Spike Gene during Adaptive Evolution." *Scientific Reports* 5 (June): 11451.
- 665 Sagulenko, Pavel, Vadim Puller, and Richard A. Neher. 2018. "TreeTime: Maximum-Likelihood
666 Phylodynamic Analysis." *Virus Evolution* 4 (1): vex042.
- 667 Smith, Derek J., Alan S. Lapedes, Jan C. de Jong, Theo M. Bestebroer, Guus F. Rimmelzwaan,
668 Albert D. M. E. Osterhaus, and Ron A. M. Fouchier. 2004. "Mapping the Antigenic and
669 Genetic Evolution of Influenza Virus." *Science* 305 (5682): 371–76.
- 670 Vijgen, Leen, Philippe Lemey, Els Keyaerts, and Marc Van Ranst. 2005. "Genetic Variability of
671 Human Respiratory Coronavirus OC43." *Journal of Virology*.
- 672 Volz, Erik M., Katia Koelle, and Trevor Bedford. 2013. "Viral Phylodynamics." *PLoS
673 Computational Biology* 9 (3): e1002947.
- 674 Woo, Patrick C. Y., Susanna K. P. Lau, Yi Huang, and Kwok-Yung Yuen. 2009. "Coronavirus
675 Diversity, Phylogeny and Interspecies Jumping." *Experimental Biology and Medicine* 234
676 (10): 1117–27.
- 677 Yang, Z. 2000. "Maximum Likelihood Estimation on Large Phylogenies and Analysis of Adaptive
678 Evolution in Human Influenza Virus A." *Journal of Molecular Evolution* 51 (5): 423–32.
- 679 Zanini, Fabio, Johanna Brodin, Lina Thebo, Christa Lanz, Göran Bratt, Jan Albert, and Richard
680 A. Neher. 2015. "Population Genomics of Intrapatient HIV-1 Evolution." *eLife* 4 (December).
681 <https://doi.org/10.7554/eLife.11282>.
- 682 Zhang, Yue, Jianguo Li, Yan Xiao, Jing Zhang, Ying Wang, Lan Chen, Gláucia
683 Paranhos-Baccalà, Lili Ren, and Jianwei Wang. 2015. "Genotype Shift in Human
684 Coronavirus OC43 and Emergence of a Novel Genotype by Natural Recombination."
685 *Journal of Infection*. <https://doi.org/10.1016/j.jinf.2014.12.005>.
- 686 Zhu, Yun, Changchong Li, Li Chen, Baoping Xu, Yunlian Zhou, Ling Cao, Yunxiao Shang, et al.
687 2018. "A Novel Human Coronavirus OC43 Genotype Detected in Mainland China."
688 *Emerging Microbes & Infections* 7 (1): 173.
- 689