



## 9 **Abstract**

10 The apparent lack of antigenic evolution by the Delta variant (B.1.617.2) of SARS-CoV-2 during the  
11 COVID-19 pandemic is puzzling. The combination of increasing immune pressure due to the rollout  
12 of vaccines and a relatively high number of infections following the relaxation of non-pharmaceutical  
13 interventions should have created perfect conditions for immune escape variants to evolve from the  
14 Delta lineage. Instead, the Omicron variant (B.1.1.529), which is hypothesised to have evolved in an  
15 immunocompromised individual, is the first major variant to exhibit significant immune escape  
16 following vaccination programmes and is set to become globally dominant in 2022. Here, we use a  
17 simple mathematical model to explore possible reasons why the Delta lineage did not exhibit  
18 antigenic evolution and to understand how and when immunocompromised individuals affect the  
19 emergence of immune escape variants. We show that when the pathogen does not have to cross a  
20 fitness valley for immune escape to occur, immunocompromised individuals have no qualitative  
21 effect on antigenic evolution (although they may accelerate immune escape if within-host  
22 evolutionary dynamics are faster in immunocompromised individuals). But if a fitness valley exists  
23 between immune escape variants at the between-host level, then persistent infections of  
24 immunocompromised individuals allow mutations to accumulate, therefore facilitating rather than  
25 simply speeding up antigenic evolution. Our results suggest that better global health equality,  
26 including improving access to vaccines and treatments for individuals who are immunocompromised  
27 (especially in lower- and middle-income countries), may be crucial to preventing the emergence of  
28 future immune escape variants of SARS-CoV-2.

## 29 **Lay Summary**

30 We study the role that immunocompromised individuals may play in the evolution of novel variants  
31 of the coronavirus responsible for the COVID-19 pandemic. We show that immunocompromised  
32 hosts can be crucial for the evolution of immune escape variants. Targeted treatment and  
33 surveillance may therefore prevent the emergence of new variants.

## 34 Introduction

35 Understanding how and when variants of SARS-CoV-2, the causative agent of COVID-19, are likely to  
36 evolve is key to managing the future of the pandemic. Multiple variants of concern have evolved  
37 since the start of the pandemic, with higher transmissibility evolving on at least two occasions, by  
38 the Alpha (B.1.1.7) variant (relative to the wildtype) [1], and by the Delta (B.1.617.2) variant (relative  
39 to Alpha) [2,3], with the latter becoming the globally dominant strain in 2021 [4]. Other variants  
40 such as Beta (B.1.351) and Omicron (B.1.1.529) have instead shown evidence of immune escape,  
41 indicating antigenic evolution [5–7]. With increasing numbers of people acquiring immunity to SARS-  
42 CoV-2, either through infection or vaccination, we should expect a shift towards antigenic evolution  
43 rather than higher transmissibility or greater virulence as the primary driver of new variants of  
44 concern [8]. The extent to which SARS-CoV-2 may evolve antigenically in future, thereby allowing it  
45 to evade host immunity partially or fully, is currently unknown. However, the emergence and rapid  
46 spread of Omicron towards the end of 2021 has demonstrated that antigenic evolution is both  
47 possible and under strong selection. The unusual nature of Omicron (possessing a large number of  
48 mutations in the spike protein but only distantly related to the dominant variant at the time, Delta  
49 [9]) has led to speculation that it underwent long-term within-host evolution in an  
50 immunocompromised individual who was unable to clear the infection [10]. We explore this  
51 hypothesis using a simple mathematical model to understand the potential importance of  
52 immunocompromised individuals for the antigenic evolution of SARS-CoV-2.

53

54 A fundamental tenet of evolutionary epidemiology is that the rate of antigenic evolution depends on  
55 a balance between immune pressure and mutation supply [11–13]. The greater the proportion of  
56 the population that is immune, the greater the strength of selection for immune escape but  
57 mutation supply is constrained as few hosts can be infected. Conversely, if many hosts are  
58 susceptible to infection, then mutation supply may be plentiful but selection for immune escape is

59 relatively weak. Hence, the rate of antigenic evolution should be maximised at an intermediate level  
60 of immune pressure, whereby moderate pathogen prevalence leads to a plentiful supply of  
61 mutations for selection to act upon, and the strength of selection for immune escape is reasonably  
62 strong.

63

64 Rapid deployment of vaccinations against SARS-CoV-2 combined with the relaxation of non-  
65 pharmaceutical interventions in many countries led to both strong immune pressure and high  
66 numbers of infections in the latter half of 2021. For example, by the end of November 2021 the UK  
67 had fully vaccinated 68% of the population while still experiencing over 620 confirmed cases per  
68 million (approximately 70% of the previous peak in January 2021) [14]. At the time, the Delta variant  
69 was dominant globally and accounted for over 99% of infections in the UK [14]. Yet, despite  
70 apparently favourable evolutionary conditions for immune escape there were no indications of the  
71 Delta variant exhibiting antigenic evolution in the UK or elsewhere. Instead, the Omicron variant,  
72 first detected in South Africa and reported to the World Health Organization on November 24, 2021  
73 [9], was able to substantially escape host immunity and evolved from a different clade. Omicron  
74 contains 30 mutations to the spike protein (used for binding to host cell receptors) and has been  
75 shown to evade over 85% of neutralizing antibodies [7]. Relative to Delta, Omicron exhibits  
76 substantially lower vaccine effectiveness [15] and is estimated to be over five times as likely to lead  
77 to reinfection [6]. Omicron became the dominant variant in the UK within a month and is on course  
78 to replace Delta in many countries in early 2022 [14].

79

80 The Omicron variant confirms that substantial immune escape is not only possible for SARS-CoV-2  
81 but also that selection for immune escape towards the end of 2021 was very strong. According to  
82 the conceptual model of antigenic evolution as a balance between immune pressure and mutation  
83 supply [11], this suggests that the lack of adaptation to evade host immunity by the Delta lineage

84 was simply due to insufficient mutation supply. However, this is difficult to reconcile with the high  
85 number of cases at the time, implying mutation supply was plentiful. Furthermore, if mutation  
86 supply was the key constraint, how did an immune escape variant appear from an obscure clade that  
87 was responsible for few infections?

88

89 A promising hypothesis is that the Omicron variant arose due to long-term within-host evolution in  
90 an immunocompromised individual, who was most likely infected between March and August 2021  
91 [9]. While an immunocompetent individual would be expected to clear infection after a relatively  
92 short period, an immunocompromised person may fail to fully clear the infection, allowing the virus  
93 to coevolve with the immune system [16]. Indeed, longitudinal sequencing from an  
94 immunocompromised patient who was infected for over 150 days with SARS-CoV-2 revealed rapid  
95 accumulation of mutations [17]. These mutations appeared to be adaptive at the within-host level  
96 due to their concentration in the spike protein, with several common to other variants of concern.  
97 Similar results have been observed in other patients with long-term infections of SARS-CoV-2  
98 [18,19], including those who have been treated with convalescent plasma, indicating antigenic  
99 evolution within the host [20].

100

101 It is currently unclear how important immunocompromised individuals are for the antigenic  
102 evolution of SARS-CoV-2. Do infections of immunocompromised individuals simply accelerate  
103 antigenic evolution or do they play a key role in facilitating immune escape? In the first case,  
104 infection of immunocompromised individuals speeds up antigenic evolution due to faster within-  
105 host dynamics, leading to the emergence of new immune escape variants on shorter timescales than  
106 would be possible in an immunocompetent population. Such a scenario would suggest that although  
107 immunocompromised individuals might speed up the process, they are not essential for antigenic  
108 evolution to occur. In the second case, long-term infections of immunocompromised individuals

109 allow the virus to accumulate mutations that are advantageous at the within-host level but may be  
110 disadvantageous at the between-host level. If there is epistasis between mutations at the between  
111 host level (i.e. fitness depends on the context of which other mutations are present), then sustained  
112 adaptation within immunocompromised individuals may allow the virus to traverse valleys in the  
113 fitness landscape, which would otherwise be very difficult to cross, to reach another peak (Fig. 1b).  
114 The second scenario would therefore suggest that long-term infections in immunocompromised  
115 individuals play a crucial role in the antigenic evolution of SARS-CoV-2.

116

117 Here, we analyse a simple phenomenological model to explore the potential importance of  
118 immunocompromised hosts for the antigenic evolution of SARS-CoV-2. We show that in the absence  
119 of epistasis, antigenic evolution readily occurs regardless of the frequency of immunocompromised  
120 individuals in the population. If epistasis is present, however, such that the virus must traverse a  
121 fitness valley at the between-host level to escape host immunity, then immunocompromised hosts  
122 are crucial for antigenic evolution to occur. These patterns are robust irrespective of whether within-  
123 host evolutionary dynamics are faster in immunocompromised individuals and for a wide range of  
124 parameters affecting cross-immunity, the strength of epistasis, the proportion of the population that  
125 is immunocompromised and their duration of infection relative to immunocompetent hosts.

126

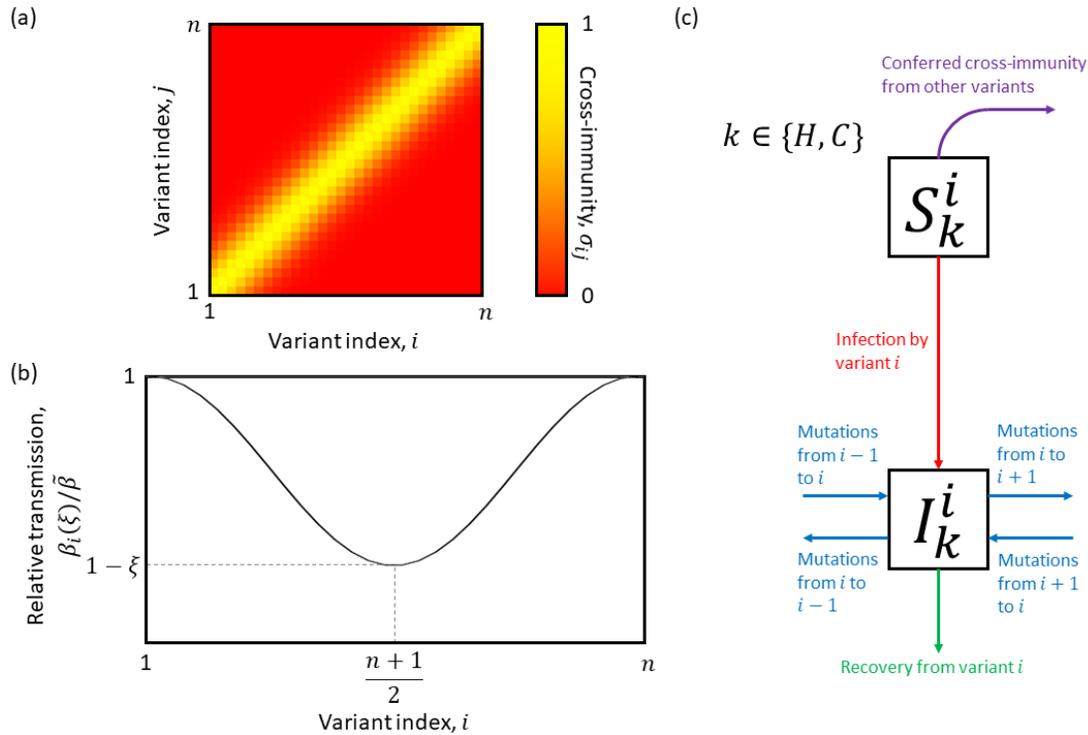
## 127 **Model description**

128 We adapt the model of antigenic evolution presented by Gog and Grenfell [21] to incorporate  
129 immunocompromised individuals and epistasis. The model assumes that there are  $n = 30$  variants  
130 equally spaced in a line, with adjacent variants differing by a single mutation. Hosts are classed as  
131 either entirely susceptible to a variant, or entirely immune to it. Cross-immunity between variants is  
132 therefore ‘polarising’, which means that when an individual is infected by variant  $i$ , a proportion  $\sigma_{ij}$

133 of those currently susceptible to variant  $j$  become fully immune to it for life (no waning immunity)  
134 and a proportion  $1 - \sigma_{ij}$  remain fully susceptible to variant  $j$ . This assumption greatly reduces the  
135 complexity of the model as it means we do not need to track all infection histories, which would  
136 require at least  $(2 + n)2^n \approx 137$  billion classes with  $n = 30$ . The strength of cross-immunity  
137 between variants  $i$  and  $j$  is given by

$$\sigma_{ij} = \exp\left\{-\frac{(i-j)^2}{2\eta}\right\}, \quad (1)$$

138 where  $\eta > 0$  controls the breadth of cross-immunity (large values of  $\eta$  give broad cross-immunity  
139 between distant variants, whereas small values of  $\eta$  limit cross-immunity to closely related variants;  
140 Fig. 1a). We assume that the population is large, well-mixed, and of constant size ( $N = 10^7$ ), with a  
141 proportion  $p$  of individuals who are immunocompromised (only able to produce a weak immune  
142 response; subscript  $C$ ) and a proportion  $1 - p$  who are immunocompetent (able to produce a  
143 normal or “healthy” immune response; subscript  $H$ ). For simplicity, we ignore host demographics  
144 (births and deaths) and mortality from infection, as we are only interested in the antigenic evolution  
145 of the virus over a relatively short timescale.



146

147 **Figure 1: Population-level model.** (a) Cross-immunity,  $\sigma_{ij}$ , for variants  $i$  and  $j$ , with lighter colours corresponding to greater  
 148 cross-immunity. (b) Illustration of the normalised transmission rate for each variant, showing a fitness valley. (c) Model  
 149 schematic.

150 Let  $S_H^i$  (respectively  $S_C^i$ ) be the proportion of the population that is immunocompetent (resp.  
 151 immunocompromised) and susceptible to variant  $i \in \{1, \dots, n\}$ , and  $I_H^i$  (resp.  $I_C^i$ ) be the proportion of  
 152 the population that is immunocompetent (resp. immunocompromised) and infected with variant  $i$ .  
 153 To incorporate a fitness valley at the between-host level, we assume that the transmission rate of  
 154 variant  $i$  is given by

$$\beta_i(\xi) = \tilde{\beta} \left( 1 + \frac{\xi}{2} \left( \cos \left( \frac{2\pi(i-1)}{n-1} \right) - 1 \right) \right), \quad (2)$$

155 where  $\tilde{\beta}$  is the maximum transmission rate and  $\xi$  controls the strength of epistasis (Fig. 1b).  
 156 Preliminary analysis revealed that other functional forms with qualitatively similar properties  
 157 produce results consistent with those presented below. When  $\xi = 0$ , there is no epistasis as  
 158  $\beta_i(\xi) = \tilde{\beta}$  for all variants. When  $0 < \xi < 1$ , epistasis reduces the transmission rate for variants

159 intermediate between 1 and  $n$ , reaching a minimum of  $\beta_i(\xi) = 1 - \xi$ , with  $\beta_1(\xi) = \beta_n(\xi) = \tilde{\beta}$  for  
160 all  $\xi$  (Fig. 1b).

161

162 To allow for random mutations, we simulate our model using the stochastic  $\tau$ -leaping method [22]  
163 for the underlying ordinary differential equations (ODEs)

$$\frac{dS_k^i}{dt} = - \sum_{j=1}^n \beta_j(\xi) S_k^i \sigma_{ij} (I_H^j + I_C^j), \quad (3)$$

$$\frac{dI_k^i}{dt} = \beta_i(\xi) S_k^i (I_H^i + I_C^i) - \left( \gamma_k + \frac{\mu_k}{1 + \delta_i} \right) I_k^i + \frac{1}{2} \mu_k M_k^i, \quad (4)$$

164 where  $M_k^i$  is the set of variants adjacent to  $i$  in the one-dimensional antigenic space ( $M_k^i = I_k^{i-1} +$   
165  $I_k^{i+1}$  for  $i \in \{2, \dots, n-1\}$ , with boundary conditions  $M_k^1 = I_k^2$  and  $M_k^n = I_k^{n-1}$ ), and  $\delta_i = 1$  if  
166  $i \in \{1, n\}$  and is 0 otherwise to control the rate of antigenic evolution at the boundaries. A  
167 schematic for this system can be found in Fig. 1c.

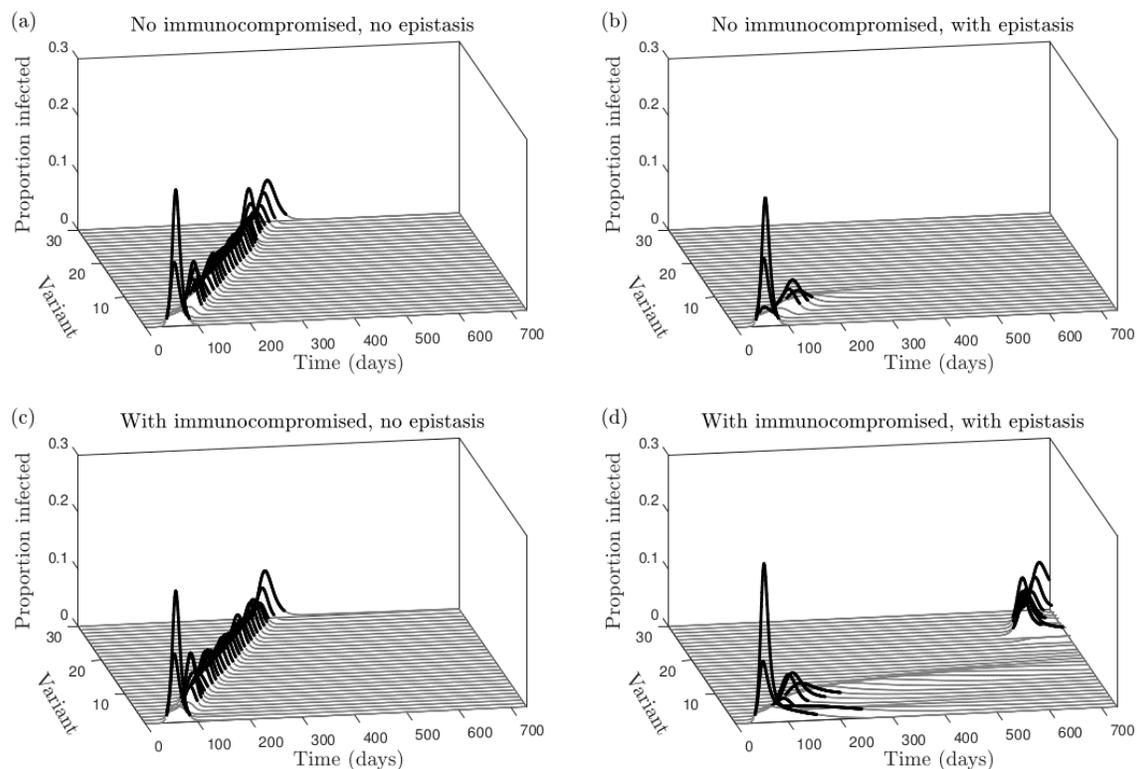
168

169 We run 10 simulations for each parameter combination up to  $t_{max} = 1460$  time steps (days), as  
170 preliminary analysis revealed that either antigenic evolution reaches the boundary of antigenic  
171 space within this timeframe, or the infection is driven extinct. Note however, that this duration is  
172 arbitrary and varies inversely with  $\mu_C$  and  $\mu_H$ . We say that a variant is 'observed' if it exceeds a  
173 threshold of 0.01. In each simulation, we summarise the dynamics by measuring the total number of  
174 variants observed and the maximum distance in antigenic space between observed variants.

175

## 176 Results

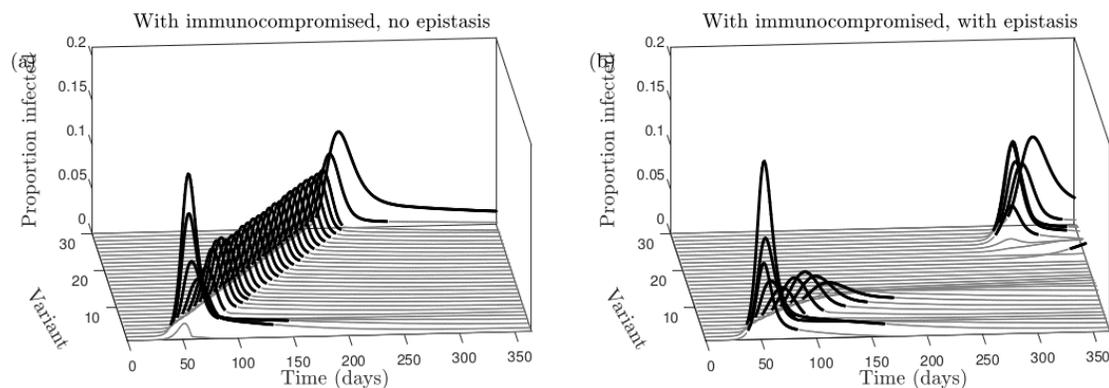
177 We focus our analysis on the strength of epistasis on transmissibility, the strength of cross-  
178 immunity, the proportion of the population that is immunocompromised, the relative rate of  
179 adaptation in immunocompromised hosts. In the absence of epistasis (or when epistasis is  
180 sufficiently weak), the virus diffuses gradually through antigenic space (Figs. 2a and 2c). As the host  
181 population accumulates immunity to the current dominant variant, selection favours the next  
182 variant in line that can substantially escape immunity, leading to successive epidemic waves at  
183 regular intervals. This occurs regardless of whether within-host evolution is assumed to be faster in  
184 immunocompromised individuals (Fig. 3a).



185

186 **Figure 2: Antigenic evolution with or without immunocompromised individuals and epistasis.** (a) No epistasis in an  
187 entirely immunocompetent population. (b) Strong epistasis in an entirely immunocompetent population  
188. (c) No epistasis and a small immunocompromised subpopulation. (d) Strong epistasis  
189 and a small immunocompromised subpopulation. All other parameter values given in Table A3.  
190 Dynamics are shown for a single simulation.

191 When epistasis is sufficiently strong, however, the proportion of the population that is  
192 immunocompromised plays a crucial role in antigenic evolution (Figs. 2b and 2d). If very few  
193 individuals are immunocompromised, the epidemic quickly burns out with little antigenic evolution,  
194 as the virus is unable to cross the fitness valley caused by epistasis at the between-host level (Fig.  
195 2b). But if a sufficient proportion of the population is immunocompromised, then the virus can cross  
196 this fitness valley due to within-host evolution in this subpopulation (Fig. 2d). Immunocompromised  
197 hosts experience longer infections, on average, which allows the virus to gradually accumulate  
198 mutations and cross the fitness valley. When the virus has acquired enough mutations in the  
199 immunocompromised such that between-host fitness is restored, it is able to spread in the rest of  
200 the host population. Again, this process is sped up if the within-host evolutionary dynamics are  
201 assumed to be faster in immunocompromised individuals, but the qualitative dynamics are  
202 unchanged (Fig. 3b).

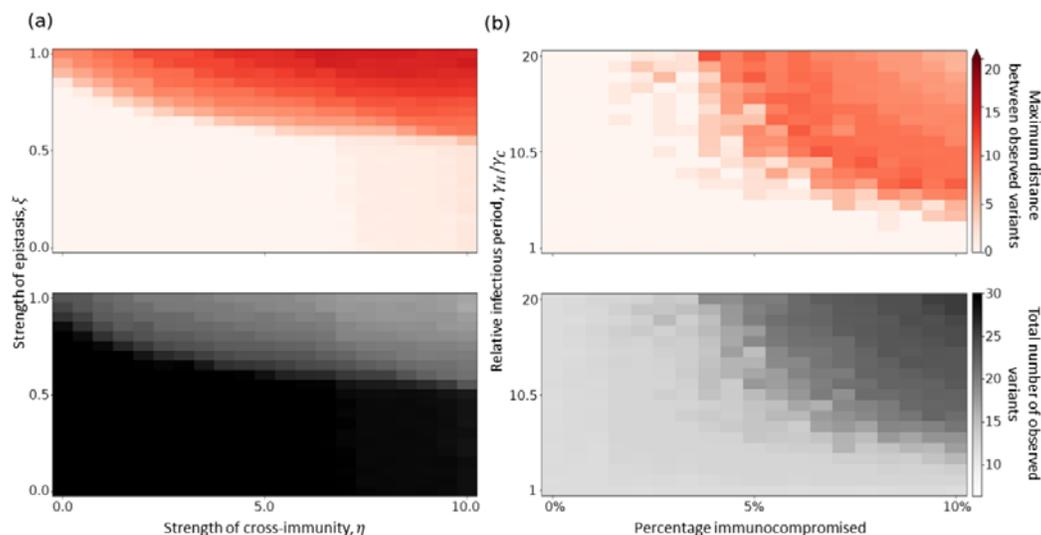


203

204 *Figure 3: Antigenic evolution when the rate of within-host evolution is faster in immunocompromised individuals. In both*  
205 *plots, there is a small immunocompromised proportion and the rate of within-host evolution in the*  
206 *immunocompromised population is 5 times higher than in the immunocompetent population. (a) No epistasis*  
207 *and (b) strong epistasis. All other parameters as in Figure 2. Dynamics are shown for a single simulation.*

208 Our results are robust to variation in key model parameters, although our sensitivity analysis reveals  
209 two notable interactions (**Figure 4**). When varying the strength of epistasis and the extent of cross-  
210 immunity between variants, we find that, intuitively, immunocompromised individuals are especially  
211 important for traversing the fitness valley if epistasis is stronger or if cross-immunity is broader

212 (Figure 4a). This is because stronger epistasis makes the fitness valley deeper and broader cross-  
213 immunity reduces the pool of susceptible hosts across a wider range of variants. However, if  
214 epistasis is sufficiently strong (around  $\xi = 0.9$  in Figure 4a) a large jump in antigenic space to a  
215 distant variant occurs regardless of the strength of cross-immunity. Our sensitivity analysis also  
216 reveals that as the proportion of the population that is immunocompromised decreases, a jump in  
217 antigenic space becomes less likely and requires a longer relative infectious period in  
218 immunocompromised hosts (Figure 4b). This suggests that better treatment of  
219 immunocompromised hosts (to reduce the average duration of infection) and better prevention and  
220 treatment of pre-existing conditions (to reduce the proportion of the population that is  
221 immunocompromised) may greatly reduce the likelihood of new variants emerging at distant fitness  
222 peaks.



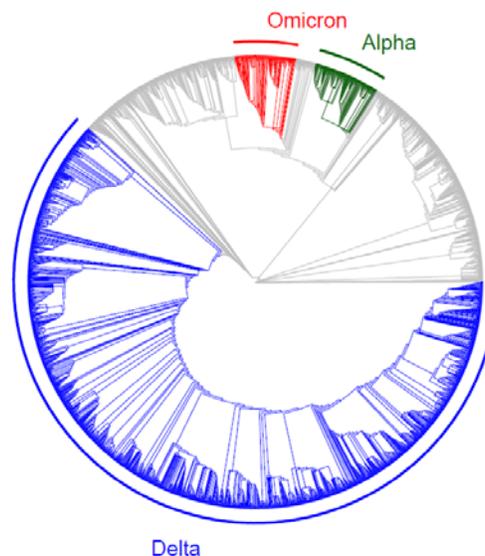
223

224 **Figure 4: Sensitivity analysis.** Top row: maximum distance between observed variants (darker shading indicates larger  
225 jumps in antigenic space); bottom row: total number of variants observed. (a) Varying the strength of cross-immunity ( $\eta$ )  
226 and epistasis ( $\xi$ ) when 5% of the population is immunocompromised ( $p=0.05$ ). (b) Varying the percentage of the population  
227 that is immunocompromised and the relative recovery periods (with  $\eta = 5$  and  $\xi = 0.8$ ). All other parameters as in Table  
228 A3. All datapoints are averaged over 10 simulations.

229

## 230 Discussion

231 The presence of immunocompromised individuals has been suggested as an important driver behind  
232 not only the emergence of the Omicron variant, but also other variants of concern, including Alpha  
233 and Delta [17]. Using a simple model of antigenic evolution, we have shown that prolonged  
234 infections of immunocompromised individuals allow pathogens to accumulate sufficient mutations  
235 to overcome epistasis at the between-host level, facilitating the emergence of novel immune escape  
236 variants. Our model was motivated by the surprising lack of antigenic evolution arising from the  
237 Delta lineage in 2021. Given relatively high levels of infection (and hence mutation supply) combined  
238 with rapidly increasing immune pressure in mid- to late-2021, conditions for the Delta lineage to  
239 exhibit antigenic evolution seemed to be favourable. The local mutational space appears to have  
240 been thoroughly explored by the Delta variant (Fig. 5), which suggests that mutation supply was not  
241 the fundamental constraint for the lack of antigenic evolution. Indeed, our model suggests that  
242 novel immune escape variants readily evolve when epistasis is relatively weak. When epistasis is  
243 stronger, reducing transmissibility for variants between fitness peaks, we find that  
244 immunocompromised individuals play a key role in antigenic evolution, effectively allowing the  
245 pathogen to traverse the fitness valley to reach a new peak. Note that while faster within-host  
246 adaptation in immunocompromised individuals speeds up the rate of antigenic evolution, unlike  
247 epistasis it does not qualitatively affect the outcome. Crucially, we have also shown that improving  
248 treatment for those who are immunocompromised can greatly reduce the likelihood of new variants  
249 emerging.



250

251 *Figure 5: Phylogenetic tree for SARS-CoV-2 variants. Three variants of concern (Alpha, Delta and Omicron) are highlighted*  
252 *to illustrate that there has been high mutation supply for the Delta variant. Data downloaded from Nextstrain*  
253 *(nextstrain.org) [23,24] and provided by the Global Initiative for Sharing All Influenza Data (GISAID, gisaid.org) [25–27].*  
254 *Data plotted using the ggtree software package in R [28–30].*

255 While our model does not capture the full complexity of antigenic evolution, it has important  
256 implications for our understanding of future immune escape variants of SARS-Cov-2. Crucially, our  
257 model suggests that the lack of antigenic evolution by Delta followed by the emergence of Omicron  
258 is consistent with epistasis constraining immune escape, but epistasis may be overcome when  
259 immunocompromised individuals are infected for sufficiently long periods. Hence, rather than  
260 simply accelerating antigenic evolution, prolonged infections of immunocompromised individuals  
261 may be critical for the evolution of immune escape variants. Based on the lack of antigenic evolution  
262 by Delta, we tentatively speculate that it may be difficult for SARS-CoV-2 to evolve antigenically  
263 through incremental mutations, and future variants may require multiple (epistatic) mutations to  
264 substantially escape host immunity. If true, a possible implication is that vaccines against SARS-CoV-  
265 2 may not need to be regularly updated in a similar manner to seasonal influenza vaccines, but if  
266 new variants do emerge, they may substantially escape prior immunity and be harder to predict.

267 However, it is also possible that Delta was unusual in being limited in its scope for antigenic  
268 evolution and that other variants may not experience similar constraints in the fitness landscape.

269

270 Our results agree with previous models which suggest that immunocompromised individuals are  
271 more likely to facilitate or accelerate within-host pathogen evolution, for example due to a longer  
272 average duration of infection or higher viral load [31,32]. However, while we find  
273 immunocompromised hosts to play a crucial role in pathogen evolution at the population level,  
274 other studies have concluded the opposite as these individuals only make up a small proportion of  
275 infections [31,32]. The reason for this discrepancy is likely due to contrasting assumptions regarding  
276 within-host fitness, immunity, and traits under selection. For example, van Egeren et al. [32] assume  
277 that a fitness valley exists at the within-host level with two or three mutations required to cross,  
278 whereas our model assumes that the fitness valley only exists at the between-host level  
279 (transmission) but may require many more mutations to traverse. If a fitness valley exists at the  
280 within-host level, then intuitively the importance of immunocompromised individuals for pathogen  
281 evolution will be lower. Their model also focused on a static measure of relative fitness and did not  
282 consider antigenic evolution explicitly, whereas in our model the fitness of a particular variant  
283 depends on the level of immunity in the population, and so will vary over the course of the  
284 epidemic. Nevertheless, both models concur that longer duration infections, especially those of  
285 immunocompromised individuals, can play a disproportionate role in the evolution of novel variants,  
286 and are of particular concern for SARS-CoV-2 evolution.

287

288 We assumed that the rate of antigenic evolution during an infection was constant (but may vary by  
289 host type), which was motivated by the within-host model discussed in the appendix. For  
290 immunocompetent hosts, who typically clear infection within two weeks [33], this means that there  
291 is relatively little time for new variants to emerge for onwards transmission, which slows down

292 adaptation and can prevent epistatic mutations accumulating. But for immunocompromised hosts,  
293 who may experience much longer infections (upwards of 150 days [17]), the coevolutionary  
294 dynamics between the virus and the host immune system could allow many (potentially epistatic)  
295 mutations to accumulate. Interestingly, this hypothesis is consistent with previous theoretical [34]  
296 and experimental [35–37] studies showing that coevolution can both accelerate adaptation and  
297 allow a pathogen to cross fitness valleys caused by epistasis.

298

299 In this study we have focused on evolution in immunocompromised individuals as a source of  
300 immune escape variants such as Omicron, but two alternative explanations must also be considered  
301 [10]. The first is that Omicron evolved early in the pandemic in a remote population (e.g., in  
302 southern Africa) before eventually taking off globally in late 2021. However, the combination of a  
303 large number of mutations in the spike protein and strong selection for immune escape in the wider  
304 population renders this explanation unsatisfactory, as it does not explain why these mutations did  
305 not appear in places where mutation supply and immune pressure were both high, such as the UK in  
306 late 2021. The second is that Omicron evolved in an animal host following transmission from  
307 humans, before crossing the species barrier into humans once again. This is a plausible but requires  
308 crossing the species barrier twice and selection to favour mutations that are beneficial in both the  
309 animal species and humans. We believe it is more plausible that Omicron evolved in an  
310 immunocompromised individual, especially since this is consistent with longitudinal sequencing of  
311 within-host evolution [17,18].

312

313 We stress that while our results suggest that infected immunocompromised individuals may play a  
314 significant role in the antigenic evolution of SARS-CoV-2, we urge caution in how this message is  
315 interpreted and communicated. While our model is informative, it does not capture the true  
316 complexity of antigenic space, the impact of vaccinations and non-pharmaceutical interventions,

317 variation in disease outcomes, and the evolution of other disease characteristics such as  
318 transmissibility and virulence. This is by design so that our model requires as few assumptions as  
319 possible. We did not attempt to capture these effects, as our results are intended to be illustrative of  
320 the key roles that epistasis and immunocompromised individuals may play in the antigenic evolution  
321 of SARS-CoV-2 (and other pathogens). We urge particular caution with regards to the implications of  
322 our results for people who are immunocompromised. People may be immunocompromised for a  
323 variety of reasons, including uncontrolled HIV, undergoing treatment for cancer, or as a transplant  
324 recipient, and some conditions still wrongly attract stigma. Although Omicron was first detected in  
325 South Africa, which is estimated to have the highest HIV prevalence in the world (7.7 million people,  
326 with many infections uncontrolled [38]), this variant may have evolved in an individual without HIV  
327 and may have evolved elsewhere. Rather than stigmatising people who are immunocompromised,  
328 our results emphasise the need for global health equality. Improving access to vaccines and  
329 treatments, especially in lower- and middle-income countries, and facilitating wider surveillance for  
330 new variants is crucial for controlling the COVID-19 pandemic.

331

## 332 **Appendix**

### 333 *Within-host model*

334 The model in the main text focuses on population-level dynamics and implicitly models within-host  
335 dynamics by assuming that: (1) immunocompetent and immunocompromised hosts differ in terms  
336 of their average infectious period; and (2) antigenic evolution occurs at a constant rate. Here, we  
337 consider the dynamics of a simple within-host model to justify the implicit within-host dynamics in  
338 our population-level model.

339

340 Let  $V_i$  be the viral abundance of variant  $i \in \{1, \dots, n\}$  within a single infected host and let  $R_i$  be the  
341 strength of the corresponding immune response. The virus grows exponentially with rate  $r$  in the  
342 absence of an immune response and decreases through the immune response at rate  $\kappa \sum_{j=1}^n \tilde{\sigma}_{ij} R_j$ ,  
343 where  $\kappa$  is the per-capita rate of virus removal by the host immune system and  $\tilde{\sigma}_{ij}$  is the probability  
344 that an immune response for variant  $j$  causes cross-immunity to variant  $i$  such that

$$\tilde{\sigma}_{ij} = \exp\left\{-\frac{(i-j)^2}{2\tilde{\eta}}\right\} \quad (\text{A1})$$

345 where  $\tilde{\eta}$  controls the breadth of cross-immunity between variants (similar to  $\eta$  in the main text). The  
346 virus also mutates to adjacent variants in the antigenic space with rate  $\tilde{\mu}$ . The immune response to  
347 variant  $i$  increases at per-capita rate  $\kappa q V_i$  and decays with rate  $d$ . The parameter  $q$  controls the  
348 strength of host immune system such that larger values indicate an immune system that can  
349 respond well to infection (immunocompetent) and smaller values indicate a weaker immune  
350 response (immunocompromised).

351

352 As with the between-host model, we use the stochastic  $\tau$ -leaping method [22] to simulate the  
353 within-host dynamics, corresponding to the following set of ODEs

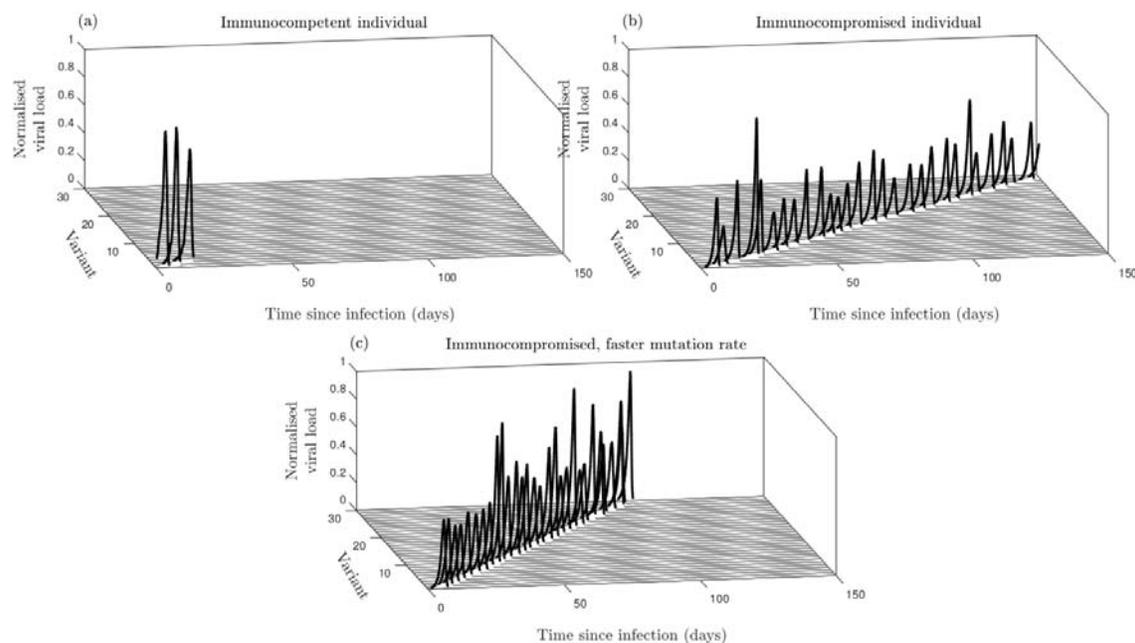
$$\frac{dV_i}{dt} = rV_i - \kappa V_i \sum_{j=1}^n \tilde{\sigma}_{ij} R_j - \frac{\tilde{\mu} V_i}{1 + \delta_i} + \frac{1}{2} \tilde{\mu} \tilde{M}_i \quad (\text{A2})$$

$$\frac{dR_i}{dt} = (\kappa q V_i - d) R_i \quad (\text{A3})$$

354 where  $\tilde{M}_i$  is the set of variants adjacent to  $i$  in the one-dimensional antigenic space ( $\tilde{M}_i = V_{i-1} +$   
355  $V_{i+1}$  for  $i \in \{2, \dots, n-1\}$ , with boundary conditions  $\tilde{M}_1 = V_2$  and  $\tilde{M}_n = V_{n-1}$ ), and  $\delta_i = 1$  if  
356  $i \in \{1, n\}$  and is 0 otherwise to control the mutation rate at the boundaries.

357

358 When the host is immunocompetent (large  $\beta$ ), the infection is rapidly cleared, with little within-host  
359 evolution (Fig. A1a). But when the host is immunocompromised (small  $\beta$ ), the infection persists over  
360 much longer timescales, with immune pressure leading to successive selective sweeps as the virus  
361 diffuses through the antigenic space at a constant rate (Fig. A1b). If the mutation rate is faster in  
362 immunocompromised hosts (larger  $\mu$ ), the coevolutionary dynamics of the virus and the immune  
363 response are accelerated but the rate of antigenic evolution remains constant (Fig. A1c). These  
364 results justify the simplifying assumptions in our population-level model regarding within-host  
365 dynamics, where we assume that there is a constant rate of antigenic evolution, which may differ  
366 between host types.



367

368 **Figure A1: Within-host dynamics.** (a) Immunocompetent host  $\beta = 0.001$ , (b) Immunocompromised host  $\beta = 0.0001$ , (c)  
369 immunocompromised host with faster mutation  $\mu = 0.001$ . Values of the viral load are normalised by the  
370 maximum value attained. All other parameters as in Table A4. Dynamics are shown for a single simulation.

371

372 **Simulation algorithm**

373 We simulate our within-host and population-level models using the  $\tau$ -leaping method [22], which is  
374 an approximate stochastic simulation algorithm. We define the propensity functions  $\alpha_E^i$  in Table A1,  
375 which give the rates of event type  $E$  for each variant index  $i$ . These propensity functions are then  
376 used to update the system synchronously at a time interval of one day using random numbers from  
377 the Poisson distribution  $P(\alpha_E^i)$ . Source code for the simulations is available in the Supplementary  
378 Material and at:  
  
379 [https://github.com/ecoevotheory/Smith\\_and\\_Ashby\\_2022](https://github.com/ecoevotheory/Smith_and_Ashby_2022).

Event	Rate
Infection of host type $k \in \{H, C\}$ by variant $i$	$\alpha_{\text{Inf}, k}^i = \beta_i S_k^i (I_H^i + I_S^i)$
Gain of full immunity to variant $i$ due to infection by variant $j$ (host type $k \in \{H, C\}$ )	$\alpha_{\text{Imm}, j, k}^i = \beta_j \sigma_{ij} S_k^i (I_H^j + I_S^j)$
Recovery by host type $k \in \{H, C\}$ from variant $i$	$\alpha_{\text{Rec}, k}^i = \gamma_k I_k^i$
Mutation from variant $i$ to variant $i \pm 1$ for host type $k \in \{H, C\}$ .	$\alpha_{\text{Mut}, k}^i = \frac{\mu_k}{1 + \delta_i} I_k^i$

380 *Table A1: The propensity functions for each of the event types in the population-level model.*

Event	Rate
Growth of variant $i$	$\alpha_{\text{Grow}}^i = rV_i$
Removal of variant $i$ by the corresponding immune response	$\alpha_{\text{Rem}}^i = \kappa V_i R_i$
Removal of variant $i$ by immune response with index $j$	$\alpha_{\text{Cross-rem}, j}^i = \kappa \tilde{\sigma}_{ij} V_i R_j$
Mutation from variant $i$ to variant $i \pm 1$	$\alpha_{\text{Mut}}^i = \frac{\tilde{\mu}}{1 + \delta_i} V_i$
Decay of immune response for variant $i$	$\alpha_{\text{Dec}}^i = dR_i$

381 *Table A2: The propensity functions for the various event types for the within-host model.*

Parameter	Description	Value
$p$	Proportion immunocompromised	0.05
$\xi$	Strength of epistasis (proportional reduction in transmission)	0.8
$N$	Population size	$10^7$
$n$	Number of variants	30
$\mu_k$	Mutation rate for $k \in \{H, C\}$	0.01
$R_0$	Basic reproductive number	3.0
$\gamma_H$	Recovery rate (immunocompetent)	1/7
$\gamma_C$	Recovery rate (immunocompromised)	1/140
$t_{max}$	Final time (days)	1460
$\eta$	Strength of cross-immunity	10.0

383 *Table A3: Default parameters for the between-host model.*

384

Parameter	Description	Value
$n$	Number of variants	30
$\tilde{\mu}$	Mutation rate	$5 \times 10^{-3}$
$r$	Viral growth rate	1.0
$\kappa$	Viral clearance rate by immune response	$2 \times 10^{-2}$
$q$	Relative strength of host immune system	0.1
$t_{max}$	Final time (days)	150
$d$	Decay rate of immune response	$10^{-4}$
$\tilde{\eta}$	Strength of cross-immunity	0.1

385 *Table A4: Default parameters for the within-host model.*

## 386 **Acknowledgements**

387 We thank Angus Buckling for helpful discussions. CAS is funded by the Natural Environment  
388 Research Council (NE/V003909/1). BA is funded by the Natural Environment Research Council  
389 (NE/N014979/1 and NE/V003909/1).

390

## 391 **Author contributions**

392 BA conceived the study, CAS carried out the analysis, and both co-authored the manuscript.

393

## 394 **Data accessibility**

395 Source code for the simulations is available in the Supplementary Material and at:

396 [https://github.com/ecoevotheory/Smith\\_and\\_Ashby\\_2022](https://github.com/ecoevotheory/Smith_and_Ashby_2022)

397

## 398 **References**

- 399 1. Davies NG, Abbott S, Barnard RC *et al.* Estimated transmissibility and impact of SARS-CoV-2  
400 lineage B.1.1.7 in England. *Science* 2021;**372**, DOI: 10.1126/science.abg3055.
- 401 2. Hagen A. How Dangerous is the Delta Variant. *American Society for Microbiology* 2021.
- 402 3. SPI-M. *SPI-M-O: Consensus Statement on COVID-19*, 2021.
- 403 4. Nexstrain. Genomic epidemiology of novel coronavirus - Global subsampling. 2021.
- 404 5. Cele S, Gazy I, Jackson L *et al.* Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent  
405 plasma. *Nature* 2021;**593**:142–6.

- 406 6. Ferguson N, Ghani A, Cori A *et al.* *Report 49: Growth, Population Distribution and Immune Escape*  
407 *of Omicron in England.*, 2021.
- 408 7. Cao Y, Wang J, Jian F *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing  
409 antibodies. *Nature* 2021, DOI: 10.1038/d41586-021-03796-6.
- 410 8. Day T, Gandon S, Lion S *et al.* On the evolutionary epidemiology of SARS-CoV-2. *Current Biology*  
411 2020;**30**:R849–57.
- 412 9. Viana R, Moyo S, Amoako DG *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant  
413 in southern Africa. *medRxiv* 2021:2021.12.19.21268028.
- 414 10. Kupferschmidt K. Where did “weird” omicron come from? 2021.
- 415 11. Grenfell BT, Pybus OG, Gog JR *et al.* Unifying the Epidemiological and Evolutionary Dynamics of  
416 Pathogens. *Science* 2004;**303**:327–32.
- 417 12. Saad-Roy CM, Morris SE, Metcalf CJE *et al.* Epidemiological and evolutionary considerations of  
418 SARS-CoV-2 vaccine dosing regimes. *Science* 2021;**372**:363–70.
- 419 13. Ashby B, Thompson RN. Non-pharmaceutical interventions and the emergence of pathogen  
420 variants. *Medrxiv* 2021, DOI: 10.1101/2021.05.27.21257938.
- 421 14. Ritchie H, Mathieu E, Rodés-Guirao L *et al.* Coronavirus Pandemic (COVID-19). *Published online at*  
422 *OurWorldInData.org* 2007.
- 423 15. Andrews N, Stowe J, Kirsebom F *et al.* Effectiveness of COVID-19 vaccines against the Omicron  
424 (B.1.1.529) variant of concern. *medRxiv* 2021:2021.12.14.21267615.
- 425 16. Marchi J, Lässig M, Walczak AM *et al.* Antigenic waves of virus–immune coevolution. *Proceedings*  
426 *of the National Academy of Sciences* 2021;**118**:e2103398118.
- 427 17. Corey L, Beyrer C, Cohen MS *et al.* SARS-CoV-2 Variants in Patients with Immunosuppression.  
428 *New England Journal of Medicine* 2021;**385**:562–6.

- 429 18. Borges V, Isidro J, Cunha M *et al.* Long-Term Evolution of SARS-CoV-2 in an Immunocompromised  
430 Patient with Non-Hodgkin Lymphoma. *mSphere* 2021;**6**, DOI: 10.1128/msphere.00244-21.
- 431 19. Weigang S, Fuchs J, Zimmer G *et al.* Within-host evolution of SARS-CoV-2 in an  
432 immunosuppressed COVID-19 patient as a source of immune escape variants. *Nature*  
433 *Communications* 2021;**12**:6405.
- 434 20. Kemp SA, Collier DA, Datir RP *et al.* SARS-CoV-2 evolution during treatment of chronic infection.  
435 *Nature* 2021;**592**:277–82.
- 436 21. Gog JR, Grenfell BT. *Dynamics and Selection of Many-Strain Pathogens.*, 2002.
- 437 22. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *The*  
438 *Journal of Chemical Physics* 2001;**115**:1716–33.
- 439 23. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus*  
440 *Evolution* 2018;**4**, DOI: 10.1093/ve/vex042.
- 441 24. Hadfield J, Megill C, Bell SM *et al.* Nextstrain: real-time tracking of pathogen evolution.  
442 *Bioinformatics* 2018;**34**:4121–3.
- 443 25. Khare S, Gurry C, Freitas L *et al.* GISAID’s Role in Pandemic Response. *China CDC Weekly*  
444 2021;**3**:1049–51.
- 445 26. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID’s innovative contribution to  
446 global health. *Global Challenges* 2017;**1**:33–46.
- 447 27. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality.  
448 *Eurosurveillance* 2017;**22**, DOI: 10.2807/1560-7917.ES.2017.22.13.30494.
- 449 28. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics*  
450 2020;**69**, DOI: 10.1002/cpbi.96.

- 451 29. Yu G, Lam TT-Y, Zhu H *et al.* Two Methods for Mapping and Visualizing Associated Data on  
452 Phylogeny Using *Ggtree*. *Molecular Biology and Evolution* 2018;**35**:3041–3.
- 453 30. Yu G, Smith DK, Zhu H *et al.* *ggtree*: an R package for visualization and annotation of phylogenetic  
454 trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;**8**:28–  
455 36.
- 456 31. Eden JS, Chisholm RH, Bull RA *et al.* Persistent infections in immunocompromised hosts are  
457 rarely sources of new pathogen variants. *Virus Evolution* 2017;**3**, DOI: 10.1093/ve/vex018.
- 458 32. van Egeren D, Novokhodko A, Stoddard M *et al.* Controlling long-term SARS-CoV-2 infections is  
459 important for slowing viral evolution. 2021, DOI: 10.1101/2021.04.10.21255251.
- 460 33. Kissler SM, Fauver JR, Mack C *et al.* Viral dynamics of acute SARS-CoV-2 infection and  
461 applications to diagnostic and public health strategies. *PLoS Biology* 2021;**19**, DOI:  
462 10.1371/journal.pbio.3001333.
- 463 34. Ashby B, Gupta S, Buckling A. Effects of epistasis on infectivity range during host-parasite  
464 coevolution. *Evolution* 2014;**68**:2972–82.
- 465 35. Paterson S, Vogwill T, Buckling A *et al.* Antagonistic coevolution accelerates molecular evolution.  
466 *Nature* 2010;**464**:275–8.
- 467 36. Hall AR, Scanlan PD, Buckling A. Bacteria-Phage Coevolution and the Emergence of Generalist  
468 Pathogens. *The American Naturalist* 2011;**177**:44–53.
- 469 37. Meyer JR, Dobias DT, Weitz JS *et al.* Repeatability and Contingency in the Evolution of a Key  
470 Innovation in Phage Lambda. *Science* 2012;**335**:428–32.
- 471 38. Mahy M, Marsh K, Sabin K *et al.* HIV estimates through 2018. *AIDS* 2019;**33**:S203–11.
- 472