

9 **Abstract**

10 Prolonged infections of immunocompromised individuals have been proposed as a crucial source of
11 new variants of SARS-CoV-2 during the COVID-19 pandemic. In principle, sustained within-host
12 antigenic evolution in immunocompromised hosts could allow novel immune escape variants to
13 emerge more rapidly, but little is known about how and when immunocompromised hosts play a
14 critical role in pathogen evolution. Here, we use a simple mathematical model to understand the
15 effects of immunocompromised hosts on the emergence of immune escape variants in the presence
16 and absence of epistasis. We show that when the pathogen does not have to cross a fitness valley for
17 immune escape to occur (no epistasis), immunocompromised individuals have no qualitative effect
18 on antigenic evolution (although they may accelerate immune escape if within-host evolutionary
19 dynamics are faster in immunocompromised individuals). But if a fitness valley exists between
20 immune escape variants at the between-host level (epistasis), then persistent infections of
21 immunocompromised individuals allow mutations to accumulate, therefore facilitating rather than
22 simply speeding up antigenic evolution. Our results suggest that better genomic surveillance of
23 infected immunocompromised individuals and better global health equality, including improving
24 access to vaccines and treatments for individuals who are immunocompromised (especially in lower-
25 and middle-income countries), may be crucial to preventing the emergence of future immune escape
26 variants of SARS-CoV-2.

27 **Lay Summary**

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

32 Introduction

33 Understanding how and when variants of SARS-CoV-2, the causative agent of COVID-19, are likely to
34 evolve is key to managing the future of the pandemic. Multiple variants of concern have evolved since
35 the start of the pandemic, with higher transmissibility evolving on at least two occasions, by the Alpha
36 (B.1.1.7) variant (relative to the wildtype) [1], and by the Delta (B.1.617.2) variant (relative to Alpha)
37 [2,3], with the latter becoming the globally dominant strain in 2021 [4]. Other variants such as Beta
38 (B.1.351) and Omicron (B.1.1.529) have additionally shown evidence of immune escape, indicating
39 antigenic evolution [5–7] (Omicron has also been linked with an increase in transmission [8,9]). With
40 increasing numbers of people acquiring immunity to SARS-CoV-2, either through infection or
41 vaccination, we should expect a shift towards antigenic evolution rather than higher intrinsic
42 transmissibility or greater virulence as the primary driver of new variants of concern [10]. The extent
43 to which SARS-CoV-2 may evolve antigenically in future, thereby allowing it to evade host immunity
44 partially or fully, is currently unknown. However, the emergence and rapid spread of Omicron towards
45 the end of 2021 has demonstrated that antigenic evolution is both possible and under strong
46 selection. The unusual nature of Omicron (possessing a large number of mutations in the spike protein
47 but only distantly related to the dominant variant at the time, Delta [11]) has led to speculation that
48 it underwent long-term within-host evolution in an immunocompromised individual who was unable
49 to clear the infection [12]. We explore this hypothesis using a simple mathematical model to
50 understand the potential importance of immunocompromised individuals for the antigenic evolution
51 of SARS-CoV-2.

52

53 A fundamental tenet of evolutionary epidemiology is that the rate of antigenic evolution depends on
54 a balance between immune pressure and mutation supply [13–15]. The greater the proportion of the
55 population that is immune, the greater the strength of selection for immune escape but mutation
56 supply is constrained as few hosts can be infected. Conversely, if many hosts are susceptible to

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

61

62 Rapid deployment of vaccinations against SARS-CoV-2 combined with the relaxation of non-
63 pharmaceutical interventions in many countries led to both strong immune pressure and high
64 numbers of infections in the latter half of 2021. For example, by the end of November 2021 the UK
65 had fully vaccinated 68% of the population while still experiencing over 620 confirmed cases per
66 million (approximately 70% of the previous peak in January 2021) [16]. At the time, the Delta variant
67 was dominant globally and accounted for over 99% of infections in the UK [16]. Yet, despite apparently
68 favourable evolutionary conditions for immune escape there were no indications of the Delta variant
69 exhibiting antigenic evolution in the UK or elsewhere. This may indicate that closely related immune
70 escape variants were suppressed, perhaps by transiently boosted innate, B, and T cell responses, or
71 due to epistasis (e.g. less transmissible). Instead, the initial BA.1 sublineage of the Omicron variant,
72 first detected in South Africa and reported to the World Health Organization on November 24, 2021
73 [11], was able to substantially escape host immunity and evolved from a distant clade. This variant
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, it exhibits substantially
76 lower vaccine effectiveness [17] and is estimated to be over five times as likely to lead to reinfection
77 [6]. The BA.1 sublineage of Omicron became the dominant variant in the UK within a month and
78 replaced Delta in many countries in early 2022 [16], with the BA.2 sublineage later replacing BA.1 [18].

79

80 The BA.1 sublineage of Omicron confirms that substantial immune escape is not only possible for
81 SARS-CoV-2 but also that selection for immune escape towards the end of 2021 was very strong.

82 According to the conceptual model of antigenic evolution as a balance between immune pressure and
83 mutation supply [13], this suggests that the lack of adaptation to evade host immunity by the Delta
84 lineage was simply due to insufficient mutation supply. However, this is difficult to reconcile with the
85 high 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 Several hypotheses have been proposed for the sudden emergence of the Omicron variant from a
90 distant clade. One possibility is that omicron evolved in an animal host following infection by a human,
91 and then jumped back into the human population. Alternatively, it could have evolved in a remote
92 population without being detected until it began to spread more widely in late 2021. However, neither
93 of these explanations are especially convincing. Evolution in an animal host would have not only
94 required two jumps across the human-animal species barrier, but also selection in the animal host
95 would have had to correspond to increased fitness in the human population through immune escape.
96 It is more plausible that Omicron was able to substantially escape immunity in humans because it had
97 experienced selection for immune escape in humans. Similarly, evolution in a remote population does
98 not appear to be plausible as it fails to explain why a similar array of mutations were not seen in
99 regions where mutation supply was significantly higher (due to more infections) and immune pressure
100 was strong due to vaccine and naturally-acquired immunity.

101

102 A more promising hypothesis is that the Omicron variant arose due to long-term within-host evolution
103 in an immunocompromised individual, who was most likely infected between March and August 2021
104 [11]. While an immunocompetent individual would be expected to clear infection after a relatively
105 short period, an immunocompromised person may fail to fully clear the infection, allowing the virus
106 to coevolve with the immune system [19]. Indeed, longitudinal sequencing from an

107 immunocompromised patient who was infected for over 150 days with SARS-CoV-2 revealed rapid
108 accumulation of mutations [20]. These mutations appeared to be adaptive at the within-host level
109 due to their concentration in the spike protein, with several common to other variants of concern.
110 Similar results have been observed in other patients with long-term infections of SARS-CoV-2 [21,22],
111 including those who have been treated with convalescent plasma, indicating antigenic evolution
112 within the host [23] (although some immunocompromised individuals show little to no within-host
113 evolution of SARS-CoV-2 [24]). A study of infection in immunocompromised individuals has found that
114 mutations accumulate in the spike gene receptor binding domain and N-terminal domains, associated
115 with immune escape and viral packaging [25]. Furthermore, a recent study has concluded that the
116 large number of mutations which were associated with the Alpha variant likely occurred in an
117 immunocompromised individual [26].

118

119 It is currently unclear how important immunocompromised individuals are for the antigenic evolution
120 of SARS-CoV-2, or for pathogen evolution more generally. Do infections of immunocompromised
121 individuals simply accelerate antigenic evolution or do they play a key role in facilitating immune
122 escape? In the first case, infection of immunocompromised individuals speeds up antigenic evolution
123 due to a faster rate of adaptation within these hosts, leading to the emergence of new immune escape
124 variants on shorter timescales than would be possible in an immunocompetent population. Such a
125 scenario would suggest that although immunocompromised individuals might speed up antigenic
126 evolution, they are not essential for it to occur. In the second case, long-term infections of
127 immunocompromised individuals allow the virus to accumulate mutations that are advantageous (or
128 neutral) at the within-host level but may be disadvantageous (or neutral) at the between-host level.
129 If there is epistasis between mutations at the between host level (i.e. fitness depends on the context
130 of which other mutations are present), as indicated by recent experiments [27], then sustained
131 adaptation within immunocompromised individuals may allow the virus to traverse valleys in the

132 fitness landscape, which would otherwise be very difficult to cross, to reach another peak. The second
133 scenario would therefore suggest that long-term infections in immunocompromised individuals play
134 a disproportionate role in the antigenic evolution of pathogens such as SARS-CoV-2.

135

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

145

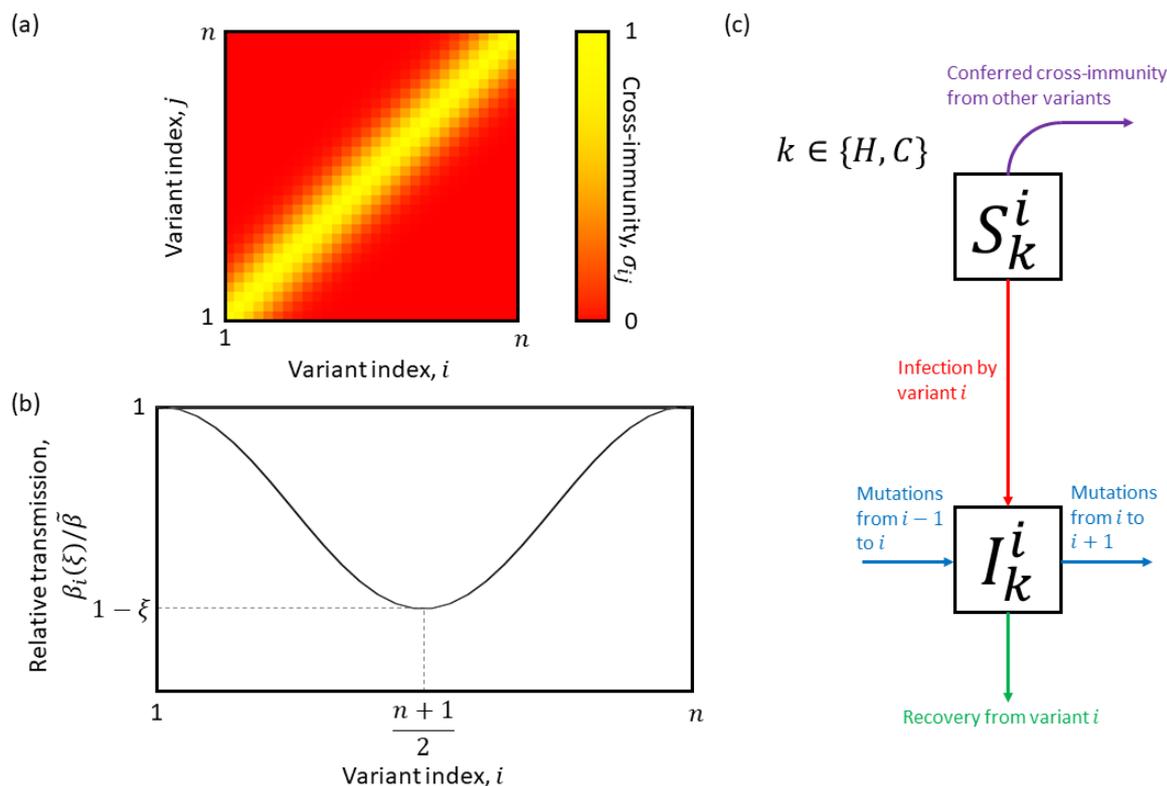
146 **Model description**

147 We adapt the model of antigenic evolution presented by Gog and Grenfell [28] to incorporate
148 immunocompromised individuals and epistasis. The model assumes that there are $n = 30$ variants
149 equally spaced in a line, with adjacent variants differing by a single mutation. Hosts are classed as
150 either entirely susceptible to a variant, or entirely immune to it. Cross-immunity between variants is
151 therefore ‘polarising’, which means that when an individual is infected by variant i , a proportion σ_{ij}
152 of those currently susceptible to variant j become fully immune to it for life (no waning immunity) and
153 a proportion $1 - \sigma_{ij}$ remain fully susceptible to variant j . This assumption greatly reduces the
154 complexity of the model as it means we do not need to track all infection histories, which would

155 require at least $(2 + n)2^n \approx 137$ billion classes with $n = 30$. The strength of cross-immunity between
156 variants i and j is given by

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

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



165

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

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

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

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

177 variants. When $0 < \xi < 1$, epistasis reduces the transmission rate for variants intermediate between
178 1 and n , reaching a minimum of $\beta_i(\xi) = 1 - \xi$, with $\beta_1(\xi) = \beta_n(\xi) = \tilde{\beta}$ for all ξ (Fig. 1b).

179

180 Healthy and immunocompromised individuals are identical in our model except for their infectious
181 periods and rates of within-host antigenic evolution. The infectious period for immunocompromised
182 individuals, $1/\gamma_C = 140$ days, is assumed to be 20 times longer than that for healthy individuals,
183 $1/\gamma_H = 7$ days. These values are chosen to be illustrative and reasonable parameter variation does
184 not qualitatively affect our results. The rate of within-host antigenic evolution (i.e. the per-capita
185 transition rate between adjacent variants in the antigenic space) is governed by parameters μ_H and
186 μ_C in healthy and immunocompromised individuals, respectively. We assume that the virus mutates
187 at a constant rate, leading to a constant rate of antigenic evolution for a given host type. While our
188 primary model implicitly captures a simplified version of within-host evolutionary dynamics by
189 assuming a constant rate of antigenic evolution per host type, we justify this assumption by exploring
190 a separate within-host only model in the Appendix, which demonstrates that a constant rate of
191 antigenic evolution is a reasonable approximation. In our primary model, we either set $\mu_H = \mu_C$ so
192 that the rate of within-host antigenic evolution is the same in healthy and immunocompromised hosts,
193 or set $\mu_H < \mu_C$ to investigate the impact of a faster rate of within-host antigenic evolution in
194 immunocompromised individuals. The rate of antigenic evolution may differ between host types due
195 to differences in the viral population size within a host or the strength of selection for immune escape.

196

197 To allow for random mutations, we simulate our model using the stochastic τ -leaping method [29] for
198 the underlying ordinary differential equations (ODEs)

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

$$\frac{dI_k^i}{dt} = \beta_i(\xi)S_k^i(I_H^i + I_C^i) - (\gamma_k + (1 - \delta_{i,n})\mu_k)I_k^i + (1 - \delta_{i,1})\mu_k I_k^{i-1}, \quad (4)$$

199 where $\delta_{i,j}$ is the Kronecker delta, which takes the value 1 if $i = j$ and is 0 otherwise. A schematic for
200 this system can be found in Fig. 1c.

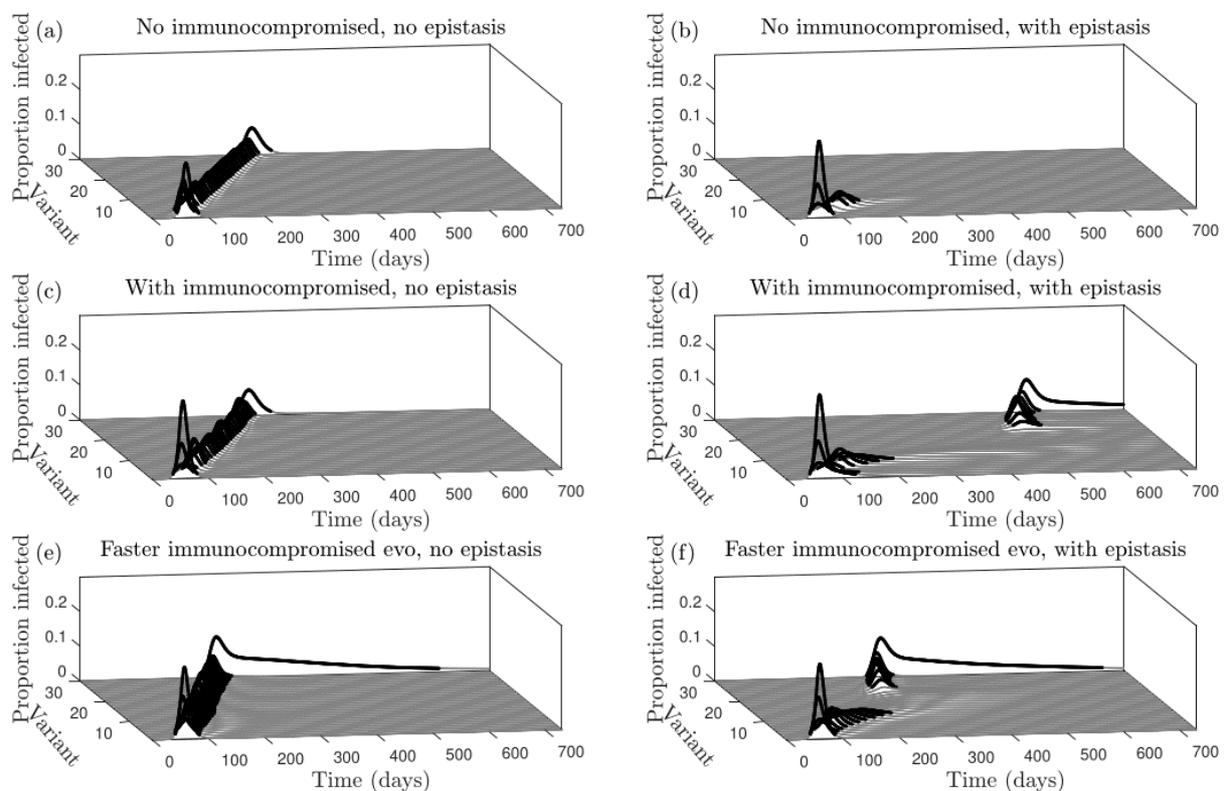
201

202 We run 10 simulations for each parameter combination up to $t_{max} = 1460$ time steps (days), as
203 preliminary analysis revealed that either antigenic evolution reaches the boundary of antigenic space
204 within this timeframe, or the infection is driven extinct. Note however, that this duration is arbitrary
205 and varies inversely with μ_C and μ_H . We say that a variant is ‘observed’ if it exceeds a threshold of
206 0.01. In each simulation, we summarise the dynamics by measuring the total number of variants
207 observed and the maximum distance in antigenic space between observed variants.

208

209 Results

210 We wish to establish when antigenic evolution proceeds as a gradual diffusion through antigenic
211 space, or in large jumps. We therefore focus our analysis on the strength of epistasis on
212 transmissibility ξ , the strength of cross-immunity η , the proportion of the population that is
213 immunocompromised p , the relative rate of adaptation (antigenic evolution) in immunocompromised
214 hosts μ_C/μ_H and the relative infectious period γ_H/γ_C . In the absence of epistasis (or when epistasis is
215 sufficiently weak), the virus diffuses gradually through antigenic space (Figs. 2a and 2c). As the host
216 population accumulates immunity to the current dominant variant, selection favours the next variant
217 in line that can substantially escape immunity, leading to successive epidemic waves at regular
218 intervals. This occurs regardless of whether within-host evolution is assumed to be faster in
219 immunocompromised individuals (Fig. 2e).



220

221 **Figure 2: Antigenic evolution with or without immunocompromised individuals and epistasis.** (a) No epistasis in an entirely
222 immunocompetent population ($p = 0, \xi = 0$). (b) Strong epistasis in an entirely immunocompetent population ($p = 0, \xi =$
223 0.8). (c) No epistasis and a small immunocompromised subpopulation ($p = 0.05, \xi = 0$). (d) Strong epistasis and a small

224 *immunocompromised subpopulation ($p = 0.05, \xi = 0.8$). (e) No epistasis and a small immunocompromised subpopulation*
225 *with faster within-host evolution in immunocompromised individuals ($p = 0.05, \xi = 0.8, \mu_C = 5\mu_H$). (f) Strong epistasis and*
226 *a small immunocompromised subpopulation with faster within-host evolution in immunocompromised individuals ($p =$*
227 *$0.05, \xi = 0.8, \mu_C = 5\mu_H$). All other parameter values given in Table A3. Dynamics are shown for a single simulation.*

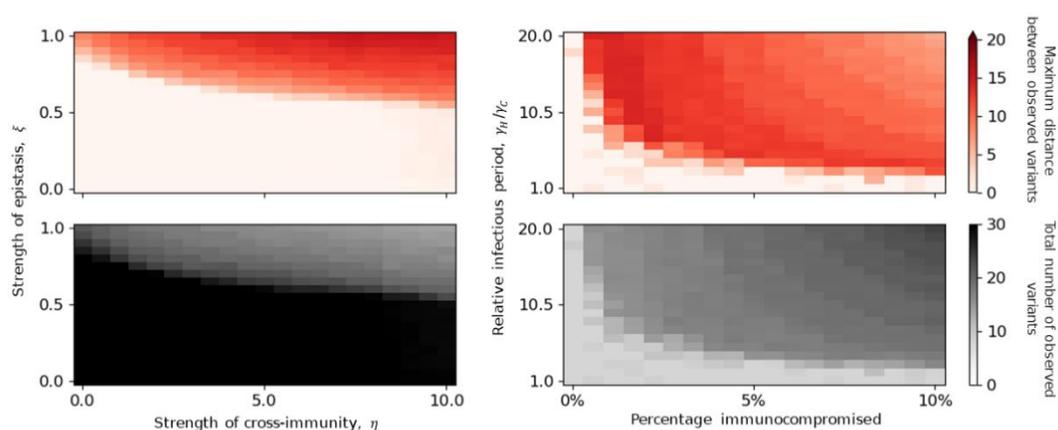
228

229 When epistasis is sufficiently strong, however, the proportion of the population that is
230 immunocompromised plays a crucial role in antigenic evolution (Figs. 2b, d, f). If very few individuals
231 are immunocompromised, the epidemic quickly burns out with little antigenic evolution, as the virus
232 is unable to cross the fitness valley caused by epistasis at the between-host level (Fig. 2b). But if a
233 sufficient proportion of the population is immunocompromised, then the virus can cross this fitness
234 valley due to within-host evolution in this subpopulation (Fig. 2d). Immunocompromised hosts
235 experience longer infections, on average, which allows the virus to accumulate mutations and cross
236 the fitness valley. When the virus has acquired enough mutations in the immunocompromised such
237 that between-host transmissibility is restored to a sufficiently high level, it is able to spread in the rest
238 of the host population. Again, this process is sped up if the within-host evolutionary dynamics are
239 assumed to be faster in immunocompromised individuals, but the qualitative dynamics are unchanged
240 (Fig. 2f).

241

242 Our results are qualitatively robust to variation in key model parameters, although our sensitivity
243 analysis reveals two notable interactions (Fig. 3). When varying the strength of epistasis and the extent
244 of cross-immunity between variants, we find that, intuitively, immunocompromised individuals are
245 especially important for traversing the fitness valley if epistasis is stronger or if cross-immunity is
246 broader (Fig. 3a). This is because stronger epistasis makes the fitness valley deeper and broader cross-
247 immunity reduces the pool of susceptible hosts across a wider range of variants. However, if epistasis
248 is sufficiently strong (around $\xi = 0.8$ in Fig. 3a) a large jump in antigenic space to a distant variant
249 occurs regardless of the strength of cross-immunity. Our sensitivity analysis also reveals that as the

250 proportion of the population that is immunocompromised decreases, a jump in antigenic space
251 becomes less likely and requires a longer relative infectious period in immunocompromised hosts (Fig.
252 3b). This suggests that better treatment of immunocompromised hosts (to reduce the average
253 duration of infection), improved genomic surveillance of these hosts (to identify novel variants of
254 concern), and better prevention and treatment of pre-existing conditions (to reduce the proportion
255 of the population that is immunocompromised) may greatly reduce the likelihood of new variants
256 emerging at distant fitness peaks.



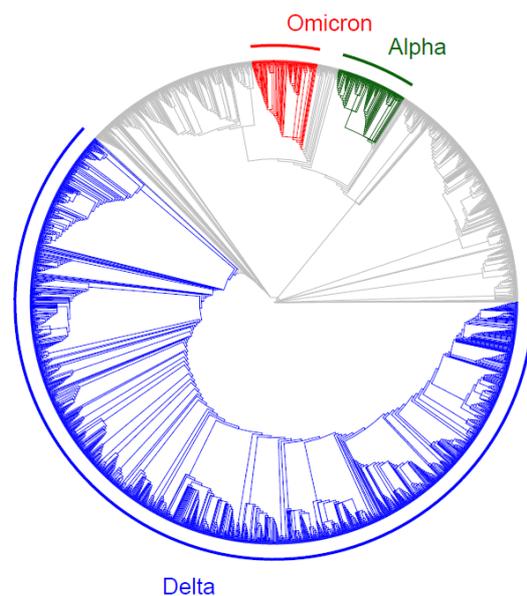
257
258 **Figure 3: Sensitivity analysis.** Top row: maximum distance between observed variants (darker shading indicates larger jumps
259 in antigenic space); bottom row: total number of variants observed. (a) Varying the strength of cross-immunity (η) and
260 epistasis (ξ) when 5% of the population is immunocompromised ($p=0.05$). (b) Varying the percentage of the population that
261 is immunocompromised and the relative recovery periods (with $\eta = 5$ and $\xi = 0.8$). All other parameters as in Table A3. All
262 datapoints are averaged over 10 simulations.

263

264 Discussion

265 The presence of immunocompromised individuals has been suggested as an important driver behind
266 not only the emergence of the Omicron variant of SARS-CoV-2, but also other variants of concern,
267 including Alpha and Delta [20]. Using a simple model of antigenic evolution, we have shown that
268 prolonged infections of immunocompromised individuals allow pathogens to accumulate sufficient
269 mutations to overcome epistasis at the between-host level, facilitating the emergence of novel

270 immune escape variants. Our model was motivated by the sudden emergence of the Omicron variant
271 from a distant clade to the dominant variant at the time, coupled with longitudinal sequencing from
272 immunocompromised patients that indicate rapid within-host evolution [18–20]. Given relatively high
273 levels of infection (and hence mutation supply; Fig. 4) combined with rapidly increasing immune
274 pressure in mid- to late-2021, conditions for the Delta lineage to exhibit antigenic evolution seemed
275 to be favourable. Mutation supply or lack of immune pressure therefore do not appear to have been
276 the fundamental constraint for the lack of antigenic evolution by the Delta variant, which suggests
277 that either epistasis or transiently-boosted innate immunity constrained immune escape. Indeed, our
278 model suggests that novel immune escape variants readily evolve when epistasis is relatively weak.
279 When epistasis is stronger, reducing transmissibility for variants between fitness peaks, we find that
280 immunocompromised individuals may play a key role in antigenic evolution, effectively allowing the
281 pathogen to traverse a fitness valley to reach a new peak. Note that while faster within-host
282 adaptation in immunocompromised individuals speeds up the rate of antigenic evolution, unlike
283 epistasis it does not qualitatively affect the outcome. Crucially, our results also suggest that improving
284 treatment for those who are immunocompromised can greatly reduce the likelihood of new variants
285 emerging.



286

287 **Figure 4: Phylogenetic tree for SARS-CoV-2 variants shortly after the emergence of Omicron.** Three variants of concern

288 (Alpha, Delta and Omicron) are highlighted to illustrate that there had been high mutation supply for the Delta variant.

289 Data downloaded from Nextstrain (nextstrain.org) on 08/02/2022 [18,30] and provided by the Global Initiative for Sharing

290 All Influenza Data (GISAID, gisaid.org) [31–33]. Data plotted using the *ggtree* software package in R [34–36].

291

292 In real populations, individuals cannot simply be classed as either healthy or immunocompromised;

293 they vary in the extent to which they are able to mount an immune response due to age,

294 comorbidities, genetic or other environmental factors. Fitness landscapes for the pathogen may also

295 differ between individuals and populations and real antigenic space is likely to be far more complex

296 than our simplified one-dimensional space (note that multi-dimensional antigenic evolution can lead

297 to more varied behaviour including branching and coalescence of phylogenetic pathways; see for

298 example [37] and [38]). While our model does not capture the full complexity of antigenic evolution

299 in real populations, it has important implications for our understanding of future immune escape

300 variants of SARS-Cov-2, and for pathogen evolution more generally. Crucially, our model suggests that

301 the lack of antigenic evolution by Delta followed by the emergence of Omicron is consistent with

302 epistasis constraining immune escape in Delta, but this epistasis may have been overcome if

303 immunocompromised individuals were infected for sufficiently long periods. Hence, rather than

304 simply accelerating antigenic evolution, prolonged infections of immunocompromised individuals may
305 have been critical for the evolution of Omicron. In the preprint of this manuscript, which was written
306 shortly after the emergence of Omicron, we tentatively speculated that the lack of antigenic evolution
307 by Delta suggested it may be difficult for SARS-CoV-2 to escape immunity through incremental
308 mutations, and future variants may require multiple (epistatic) mutations to substantially escape host
309 immunity. However, the subsequent emergence of immune evasion sub-variants of Omicron such as
310 BA.4 and BA.5 suggests that either Delta was unusual in its limited scope for antigenic evolution and
311 that other variants do not experience similar constraints, or that Delta may have eventually exhibited
312 antigenic evolution if Omicron had not emerged. Regardless, our model suggests that
313 immunocompromised individuals may remain a source of new variants that can substantially escape
314 immunity. While not a focus of the current study, in principle prolonged infections of
315 immunocompromised individuals could also facilitate the emergence and coexistence of multiple
316 variants lacking cross-immunity, allowing the pathogen population to occupy different niches in a
317 multidimensional antigenic space [39].

318

319 Our results agree with previous models which suggest that immunocompromised individuals are more
320 likely to facilitate or accelerate within-host pathogen evolution, for example due to a longer average
321 duration of infection or higher viral load [40,41]. However, while we find immunocompromised hosts
322 to play a crucial role in pathogen evolution at the population level, other studies have concluded the
323 opposite as these individuals only make up a small proportion of infections [40,41]. The reason for this
324 discrepancy is likely due to contrasting assumptions regarding within-host fitness, immunity, and traits
325 under selection. For example, van Egeren et al. [41] assume that a fitness valley exists at the within-
326 host level with two or three mutations required to cross, whereas our model assumes that the fitness
327 valley only exists at the between-host level (transmission) but may require many more mutations to
328 traverse. If a fitness valley exists at the within-host level, then intuitively the importance of

329 immunocompromised individuals for pathogen evolution will be lower. In general, there is no reason
330 why the fitness landscape should have the same shape at the within- and between-host levels, as
331 these are two very different environments with potentially contrasting selection pressures (e.g. UV
332 exposure, temperature, interactions with the immune system). In general, one should not expect that
333 a beneficial mutation in one context will necessarily be beneficial in another (“antagonistic
334 pleiotropy”). Some mutations may therefore confer a fitness advantage at the within-host level, while
335 being neutral or detrimental for transmission. For example, a mutation that increases the growth rate
336 in the lower lung may be advantageous at the within-host level, but may lead to fewer transmission
337 stages being produced, resulting in lower fitness at the between-host level. In SARS-CoV-2, the
338 mutation D796H protects against neutralising antibodies but reduces infectivity, unless the mutation
339 Δ H69/ Δ V70 is also present [23]. It is therefore reasonable to expect differences in the fitness
340 landscape at the within and between-host levels. In addition to the different assumptions about the
341 fitness landscape, the model by van Egeren et al. [41] also focused on a static measure of relative
342 fitness and did not consider antigenic evolution explicitly, whereas in our model the fitness of a
343 particular variant depends on the level of immunity in the population, and so will vary over the course
344 of the epidemic. Nevertheless, both models concur that longer duration infections, especially those
345 of immunocompromised individuals, can play a disproportionate role in the evolution of novel
346 variants, and are of particular concern for SARS-CoV-2 evolution.

347

348 We assumed that the rate of antigenic evolution during an infection was constant (but may vary by
349 host type), which was motivated by the separate within-host model discussed in the Appendix. For
350 immunocompetent hosts, who typically clear infection within two weeks [42], this means that there
351 is relatively little time for new variants to emerge for onwards transmission, which slows down
352 adaptation and can prevent epistatic mutations accumulating. But for immunocompromised hosts,
353 who may experience much longer infections (upwards of 150 days [20]), the coevolutionary dynamics

354 between the virus and the host immune system could allow many (potentially epistatic) mutations to
355 accumulate. Interestingly, this hypothesis is consistent with previous theoretical [43] and
356 experimental [44–46] studies showing that coevolution can both accelerate adaptation and allow a
357 pathogen to cross fitness valleys caused by epistasis. For simplicity, our model assumed that the shape
358 of the fitness landscape was similar in immunocompromised and immunocompetent hosts,
359 preventing the pathogen from specialising on one type of host. Specialisation on
360 immunocompromised individuals has been observed for other pathogens (e.g. *Pseudomonas*
361 *aeruginosa* infections in cystic fibrosis patients [47]) and longitudinal studies of prolonged SARS-CoV-
362 2 infections reveal the rapid accumulation of many mutations [20–24], but it is unclear if and when
363 this leads to specialisation with a reduction in fitness on immunocompetent hosts.

364

365

366 While our model is informative, it does not capture the true complexity of antigenic space, the impact
367 of vaccinations and non-pharmaceutical interventions, variation in disease outcomes, and the
368 evolution of other disease characteristics such as transmissibility and virulence. This is by design so
369 that our model requires as few assumptions as possible and so that the model can be adapted for
370 other pathogens in future. We did not attempt to capture these effects, as our results are intended to
371 be illustrative of the key roles that epistasis and immunocompromised individuals may play in the
372 antigenic evolution of SARS-CoV-2 (and other pathogens). Modelling of immunocompromised
373 individuals during the COVID-19 pandemic has largely focused on their increased risk of mortality,
374 rather than their potential importance for pathogen evolution [48–50]. Our study emphasises the
375 need to consider both aspects. We also did not explicitly model within-host dynamics in the main text,
376 instead approximating these dynamics following analysis of a separate within-host model in the
377 appendix. The within-host model in the appendix demonstrated that immune pressure leads to
378 diffusion through the antigenic space at a constant rate, therefore justifying our assumption of a

379 constant rate of antigenic evolution in our primary model. This allowed us to assume infected
380 individuals would substitute variant i with variant $i + 1$ at a constant rate, which mimicked typical
381 within-host dynamics without the need for a fully nested model, which would be much more complex
382 but would likely provide no additional insights to our simpler model.

383

384 We stress that while our results suggest that infected immunocompromised individuals may play a
385 significant role in the antigenic evolution of SARS-CoV-2, we urge caution in how this message is
386 interpreted and communicated. We urge particular caution with regards to the implications of our
387 results for people who are immunocompromised. People may be immunocompromised for a variety
388 of reasons, including uncontrolled HIV, undergoing treatment for cancer, or as a transplant recipient,
389 and some conditions still wrongly attract stigma. Although Omicron was first detected in South Africa,
390 which is estimated to have the highest HIV prevalence in the world (7.7 million people, with many
391 infections uncontrolled [51]), this variant may have evolved in an individual without HIV and may have
392 evolved elsewhere. Rather than stigmatising people who are immunocompromised, our results
393 emphasise the need for global health equality and for better genomic surveillance, especially for
394 immunocompromised people infected with SARS-CoV-2. Improving access to vaccines and
395 treatments, especially in lower- and middle-income countries, and facilitating wider surveillance for
396 new variants is crucial for limiting the emergence of new variants in the COVID-19 pandemic.

397

398 **Appendix**

399 ***Within-host model***

400 The model in the main text focuses on population-level dynamics and implicitly models within-host
401 dynamics by assuming that: (1) immunocompetent and immunocompromised hosts differ in terms of
402 their average infectious period; and (2) antigenic evolution occurs at a constant rate. Here, we

403 consider the dynamics of a simple within-host model to justify the implicit within-host dynamics in
404 our population-level model.

405

406 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
407 strength of the corresponding immune response. The virus grows exponentially with rate r in the
408 absence of an immune response and decreases through the immune response at rate $\kappa \sum_{j=1}^n \tilde{\sigma}_{ij} R_j$,
409 where κ is the per-capita rate of virus removal by the host immune system and $\tilde{\sigma}_{ij}$ is the probability
410 that an immune response for variant j causes cross-immunity to variant i such that

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

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

417

418 As with the between-host model, we use the stochastic τ -leaping method [29] to simulate the within-
419 host dynamics, corresponding to the following set of ODEs

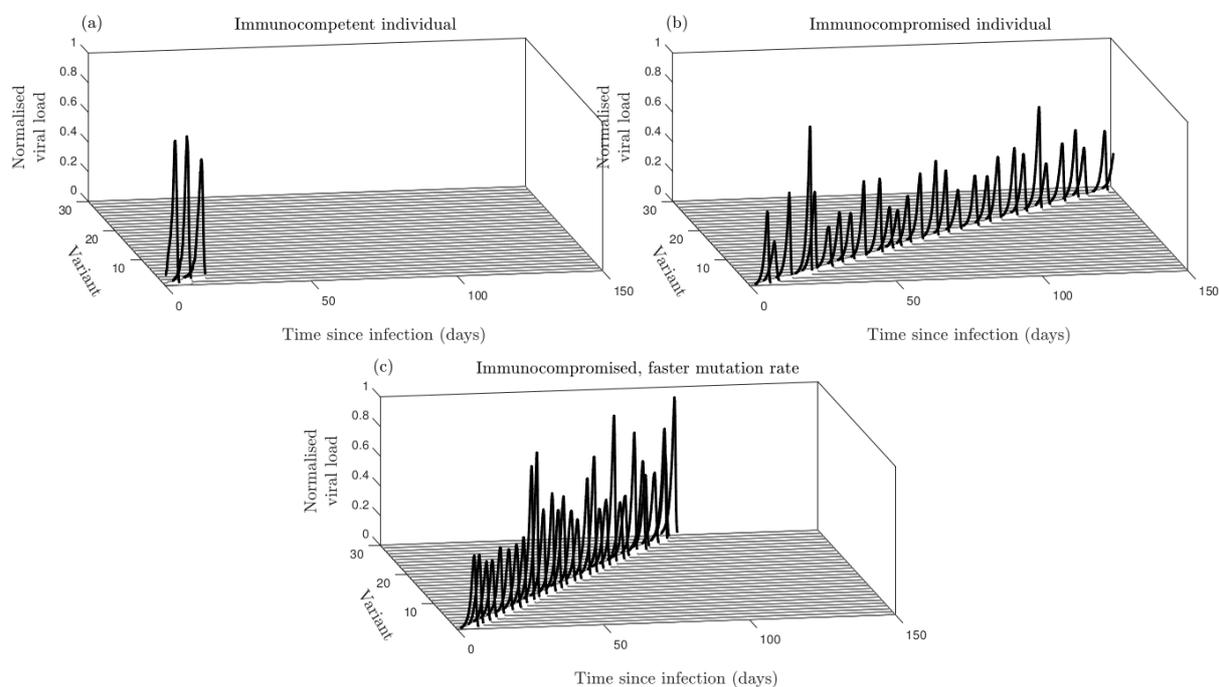
$$\frac{dV_i}{dt} = rV_i - \kappa V_i \sum_{j=1}^n \tilde{\sigma}_{ij} R_j - \frac{\tilde{\mu}V_i}{1 + \delta_i} + \frac{1}{2}\tilde{\mu}\tilde{M}_i \quad (\text{A2})$$

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

420 where \tilde{M}_i is the set of variants adjacent to i in the one-dimensional antigenic space ($\tilde{M}_i = V_{i-1} + V_{i+1}$
 421 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 $i \in \{1, n\}$ and
 422 is 0 otherwise to control the mutation rate at the boundaries.

423

424 When the host is immunocompetent (large q), the infection is rapidly cleared, with little within-host
 425 evolution (Fig. A1a). But when the host is immunocompromised (small q), the infection persists over
 426 much longer timescales, with immune pressure leading to successive selective sweeps as the virus
 427 diffuses through the antigenic space at a constant rate (Fig. A1b). If the mutation rate is faster in
 428 immunocompromised hosts (larger $\tilde{\mu}$), the coevolutionary dynamics of the virus and the immune
 429 response are simply accelerated (Fig. A1c compared with Fig. A1b). These results justify the simplifying
 430 assumptions in our population-level model regarding within-host dynamics, where we assume that
 431 there is a constant rate of antigenic evolution, which may differ between host types.



432

433 **Figure A1: Within-host dynamics.** (a) Immunocompetent host ($q = 1.0$), (b) Immunocompromised host ($q = 0.1$), (c)
 434 immunocompromised host with faster mutation $q = 0.1, \tilde{\mu} = 0.02$). Values of the viral load are normalised by the maximum
 435 value attained. All other parameters as in Table A4. Dynamics are shown for a single simulation.

436 ***Simulation algorithm***

437 We simulate our within-host and population-level models using the τ -leaping method [29], which is
438 an approximate stochastic simulation algorithm. We define the propensity functions α_E^i in Table A1,
439 which give the rates of event type E for each variant index i . These propensity functions are then used
440 to update the system synchronously at a time interval of one day using random numbers from the
441 Poisson distribution $P(\alpha_E^i)$. Source code for the simulations is available in the Supplementary Material
442 and at https://github.com/ecevotheory/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 $k \in \{H, C\}$.	$\alpha_{\text{Mut}, k}^i = \frac{\mu_k}{1 + \delta_i} I_k^i$

443 **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	$\alpha_{\text{Dec}}^i = dR_i$

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

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

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

447

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

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

449 **Acknowledgements**

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

453

454 **Author contributions**

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

456

457 **Data accessibility**

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

459 https://github.com/ecevotheory/Smith_and_Ashby_2022

460

461 **References**

462 1. Davies NG, Abbott S, Barnard RC *et al.* Estimated transmissibility and impact of SARS-CoV-2
463 lineage B.1.1.7 in England. *Science* (1979) 2021;**372**, DOI: 10.1126/science.abg3055.

464 2. Hagen A. How Dangerous is the Delta Variant. *American Society for Microbiology* 2021.

465 3. SPI-M. *SPI-M-O: Consensus Statement on COVID-19.*, 2021.

466 4. Nexstrain. Genomic epidemiology of novel coronavirus - Global subsampling. 2021.

467 5. Cele S, Gazy I, Jackson L *et al.* Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent
468 plasma. *Nature* 2021;**593**:142–6.

469 6. Ferguson N, Ghani A, Cori A *et al.* *Report 49: Growth, Population Distribution and Immune Escape*
470 *of Omicron in England.*, 2021.

471 7. Cao Y, Wang J, Jian F *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing
472 antibodies. *Nature* 2021, DOI: 10.1038/d41586-021-03796-6.

473 8. Lyngse FP, Mortensen LH, Denwood MJ *et al.* SARS-CoV-2 Omicron VOC Transmission in Danish
474 Households. *medRxiv* 2021:2021.12.27.21268278.

- 475 9. Sofonea MT, Roquebert B, Foulongne V *et al.* From Delta to Omicron: analysing the SARS-CoV-2
476 epidemic in France using variant-specific screening tests (September 1 to December 18, 2021).
477 *medRxiv* 2022:2021.12.31.21268583.
- 478 10. Day T, Gandon S, Lion S *et al.* On the evolutionary epidemiology of SARS-CoV-2. *Current Biology*
479 2020;**30**:R849–57.
- 480 11. Viana R, Moyo S, Amoako DG *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant
481 in southern Africa. *medRxiv* 2021:2021.12.19.21268028.
- 482 12. Kupferschmidt K. Where did “weird” omicron come from? 2021.
- 483 13. Grenfell BT, Pybus OG, Gog JR *et al.* Unifying the Epidemiological and Evolutionary Dynamics of
484 Pathogens. *Science (1979)* 2004;**303**:327–32.
- 485 14. Saad-Roy CM, Morris SE, Metcalf CJE *et al.* Epidemiological and evolutionary considerations of
486 SARS-CoV-2 vaccine dosing regimes. *Science (1979)* 2021;**372**:363–70.
- 487 15. Ashby B, Thompson RN. Non-pharmaceutical interventions and the emergence of pathogen
488 variants. *Medrxiv* 2021, DOI: 10.1101/2021.05.27.21257938.
- 489 16. Ritchie H, Mathieu E, Rodés-Guirao L *et al.* Coronavirus Pandemic (COVID-19). *Published online at*
490 *OurWorldInData.org* 2020.
- 491 17. Andrews N, Stowe J, Kirsebom F *et al.* Effectiveness of COVID-19 vaccines against the Omicron
492 (B.1.1.529) variant of concern. *medRxiv* 2021:2021.12.14.21267615.
- 493 18. Hadfield J, Megill C, Bell SM *et al.* Nextstrain: real-time tracking of pathogen evolution.
494 *Bioinformatics* 2018;**34**:4121–3.
- 495 19. Marchi J, Lässig M, Walczak AM *et al.* Antigenic waves of virus–immune coevolution. *Proceedings*
496 *of the National Academy of Sciences* 2021;**118**:e2103398118.
- 497 20. Corey L, Beyrer C, Cohen MS *et al.* SARS-CoV-2 Variants in Patients with Immunosuppression.
498 *New England Journal of Medicine* 2021;**385**:562–6.
- 499 21. Borges V, Isidro J, Cunha M *et al.* Long-Term Evolution of SARS-CoV-2 in an Immunocompromised
500 Patient with Non-Hodgkin Lymphoma. *mSphere* 2021;**6**, DOI: 10.1128/msphere.00244-21.
- 501 22. Weigang S, Fuchs J, Zimmer G *et al.* Within-host evolution of SARS-CoV-2 in a
502 immunosuppressed COVID-19 patient as a source of immune escape variants. *Nature*
503 *Communications* 2021;**12**:6405.
- 504 23. Kemp SA, Collier DA, Datir RP *et al.* SARS-CoV-2 evolution during treatment of chronic infection.
505 *Nature* 2021;**592**:277–82.
- 506 24. Álvarez H, Ruiz-Mateos E, Juiz-González PM *et al.* SARS-CoV-2 Evolution and Spike-Specific CD4+
507 T-Cell Response in Persistent COVID-19 with Severe HIV Immune Suppression. *Microorganisms*
508 2022;**10**:143.
- 509 25. Wilkinson SAJ, Richter A, Casey A *et al.* Recurrent SARS-CoV-2 Mutations in Immunodeficient
510 Patients. *medRxiv* 2022:2022.03.02.22271697.
- 511 26. Hill V, du Plessis L, Peacock TP *et al.* The origins and molecular evolution of SARS-CoV-2 lineage
512 B.1.1.7 in the UK. *bioRxiv* 2022:2022.03.08.481609.

- 513 27. Martin DP, Lytras S, Lucaci AG *et al.* Selection analysis identifies unusual clustered mutational
514 changes in Omicron lineage BA.1 that likely impact Spike function. *bioRxiv* 2022, DOI:
515 10.1101/2022.01.14.476382.
- 516 28. Gog JR, Grenfell BT. Dynamics and selection of many-strain pathogens. *Proceedings of the*
517 *National Academy of Sciences* 2002;**99**.
- 518 29. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *The*
519 *Journal of Chemical Physics* 2001;**115**:1716–33.
- 520 30. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus*
521 *Evolution* 2018;**4**, DOI: 10.1093/ve/vex042.
- 522 31. Khare S, Gurry C, Freitas L *et al.* GISAID’s Role in Pandemic Response. *China CDC Weekly*
523 2021;**3**:1049–51.
- 524 32. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID’s innovative contribution to
525 global health. *Global Challenges* 2017;**1**:33–46.
- 526 33. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality.
527 *Eurosurveillance* 2017;**22**, DOI: 10.2807/1560-7917.ES.2017.22.13.30494.
- 528 34. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics*
529 2020;**69**, DOI: 10.1002/cpbi.96.
- 530 35. Yu G, Lam TT-Y, Zhu H *et al.* Two Methods for Mapping and Visualizing Associated Data on
531 Phylogeny Using *Ggtree*. *Molecular Biology and Evolution* 2018;**35**:3041–3.
- 532 36. Yu G, Smith DK, Zhu H *et al.* ggtree: an R package for visualization and annotation of phylogenetic
533 trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;**8**:28–
534 36.
- 535 37. Yan L, Neher RA, Shraiman BI. Phylodynamic theory of persistence, extinction and speciation of
536 rapidly adapting pathogens. *Elife* 2019;**8**, DOI: 10.7554/eLife.44205.
- 537 38. Zinder D, Bedford T, Gupta S *et al.* The Roles of Competition and Mutation in Shaping Antigenic
538 and Genetic Diversity in Influenza. *PLoS Pathogens* 2013;**9**:e1003104.
- 539 39. Ashby B, Watkins E, Lourenço J *et al.* Competing species leave many potential niches unfilled.
540 *Nature Ecology & Evolution* 2017;**1**:1495–501.
- 541 40. Eden JS, Chisholm RH, Bull RA *et al.* Persistent infections in immunocompromised hosts are
542 rarely sources of new pathogen variants. *Virus Evolution* 2017;**3**, DOI: 10.1093/ve/vex018.
- 543 41. van Egeren D, Novokhodko A, Stoddard M *et al.* Controlling long-term SARS-CoV-2 infections is
544 important for slowing viral evolution. 2021, DOI: 10.1101/2021.04.10.21255251.
- 545 42. Kissler SM, Fauver JR, Mack C *et al.* Viral dynamics of acute SARS-CoV-2 infection and
546 applications to diagnostic and public health strategies. *PLoS Biology* 2021;**19**, DOI:
547 10.1371/journal.pbio.3001333.
- 548 43. Ashby B, Gupta S, Buckling A. Effects of epistasis on infectivity range during host-parasite
549 coevolution. *Evolution (N Y)* 2014;**68**:2972–82.

- 550 44. Paterson S, Vogwill T, Buckling A *et al.* Antagonistic coevolution accelerates molecular evolution.
551 *Nature* 2010;**464**:275–8.
- 552 45. Hall AR, Scanlan PD, Buckling A. Bacteria-Phage Coevolution and the Emergence of Generalist
553 Pathogens. *The American Naturalist* 2011;**177**:44–53.
- 554 46. Meyer JR, Dobias DT, Weitz JS *et al.* Repeatability and Contingency in the Evolution of a Key
555 Innovation in Phage Lambda. *Science (1979)* 2012;**335**:428–32.
- 556 47. Smith EE, Buckley DG, Wu Z *et al.* Genetic adaptation by *Pseudomonas aeruginosa* to the airways
557 of cystic fibrosis patients. *Proceedings of the National Academy of Sciences* 2006;**103**:8487–92.
- 558 48. Smith CA, Yates ChristianA, Ashby B. Critical weaknesses in shielding strategies for COVID-19.
559 *PLOS Global Public Health* 2022.
- 560 49. Neufeld Z, Khataee H, Czirok A. Targeted adaptive isolation strategy for COVID-19 pandemic.
561 *Infect Dis Model* 2020;**5**, DOI: 10.1016/j.idm.2020.04.003.
- 562 50. van Bunnik BAD, Morgan ALK, Bessell PR *et al.* Segmentation and shielding of the most
563 vulnerable members of the population as elements of an exit strategy from COVID-19 lockdown.
564 *Philosophical Transactions of the Royal Society B: Biological Sciences* 2021;**376**, DOI:
565 10.1098/rstb.2020.0275.
- 566 51. Mahy M, Marsh K, Sabin K *et al.* HIV estimates through 2018. *AIDS* 2019;**33**:S203–11.
567