

Evolution of virus

Bell

1

2

3

4

5

6 Evolutionary dynamics of a virus in a vaccinated population

7 Graham Bell

8

9

10 Biology Department and Redpath Museum

11 McGill University

12 Montreal, Quebec H3A 1B1, Canada

13

14 Email: graham.bell@mcgill.ca

15

16 Summary

17 The progress of an epidemic in a small closed community is simulated by an agent-based model
18 which allows vaccination and variation. The attributes of the virus are governed by two genetic
19 loci: the P-locus, which determines growth, and the M-locus, which determines immune
20 characteristics. Mutation at either locus modifies the attributes of the virus and leads to
21 evolution through natural selection. For both loci the crucial variable is the potential mutation
22 supply U_{Pot} , because evolution is likely to happen when $U_{Pot} > 1$. Mutation at the P-locus causes
23 a limited increase in virulence, which may be affected by vaccine design. Mutation at the M-
24 locus may cause a qualitative shift of dynamic regime from a simple limited epidemic to a
25 perennial endemic disease by giving rise to escape mutants which may themselves mutate. A
26 broad vaccine that remains efficacious despite several mutations at the M-locus prevents this
27 shift and provides protection despite the evolution of the virus. Escape variants may
28 nevertheless arise through recombination after coinfection, and can be suppressed by timely
29 revaccination, using the prevalent strain to design the vaccine.

30

31

32

Introduction

33 Once a virus has first infected a single individual in a population of susceptible hosts, its lineage

34 may subsequently follow one of two paths: either it fails to be transmitted and soon becomes

35 extinct, or it proliferates so as to cause an epidemic. Which path is followed depends in part on

36 deterministic factors, such as its initial rate of transmission, and in part on stochastic factors,

37 because the initial survival and transmission of the virus are strongly influenced by chance. The

38 time course of an epidemic in a local population can be adequately described by the

39 mathematical theory that has been developed over the last century (reviewed by Brauer 2017).

40 If the virus is liable to vary when it replicates, however, the virus population will evolve during

41 the epidemic and as its properties change its dynamics may become more complicated (Lenski

42 & May 1994, Day & Proulx 2004, Day & Gandon 2007), as in the well-known case of influenza

43 (reviewed by Nelson & Holmes 2007). As for the virus population as a whole, the spread of any

44 new strain, beginning with a single individual arising by mutation or recombination, will be

45 influenced both by deterministic processes such as natural selection and by stochastic

46 processes such as genetic drift. The ongoing pandemic caused by the coronavirus SARS-CoV-2

47 has made the public more aware of virus evolution and more fearful of its consequences, as the

48 spread of new strains undermines efforts to control the outbreak (Day et al. 2021).

49 Novel mutations which arise during the proliferation of an ancestral strain may modify the

50 propensity to infect naïve hosts, which have never before been infected by the ancestor, or the

51 ability to infect recovered hosts, which have previously been infected but then cleared the

52 virus. The first kind, often called virulence or life-history variants, alters the extent of
53 proliferation within the host, which results in a viral titre characteristic of a particular strain.
54 This in turn affects its virulence, because a higher titre causes more severe symptoms and
55 endangers the host, and its transmissibility, because a higher titre results in more virus particles
56 being shed (May & Anderson 1983, van Baalen & Sabelis 1995). The second kind, often called
57 escape variants, alters the antigenic properties of the virus and may thereby enable it to evade
58 the immune system of the host (Gog & Grenfell 2002, Grenfell et al. 2004, Fryer & McLean
59 2011). These two approaches use different methods and have remained largely separate. The
60 evolution of virulence is usually viewed as a long-term process using quantitative population
61 genetics to model competition between many virus lineages, whereas escape is usually
62 represented as the short-term outcome of competition between two clones of virus (Gandon &
63 Day 2007).

64 Vaccination protects the population by arming individuals with an immune response against
65 virus strains which have an antigenic profile similar to the vaccine. It is an effective method of
66 reducing the incidence of the disease, and may even drive the virus extinct if the frequency of
67 unvaccinated hosts is too low to sustain an epidemic (Anderson & May 1991). It is also a
68 profound alteration of the host environment, however, and may thereby act as an agent of
69 selection on the virus population. Moreover, vaccinated individuals are not necessarily
70 equivalent to recovered individuals, either in terms of their immune response or the virulence
71 and transmissibility of any infection which succeeds in taking hold. When there are many
72 strains of virus circulating in a vaccinated population, each differing in its virulence,
73 transmissibility and antigenic properties, and each subject to stochastic as well as deterministic

74 change, then analytical solutions of conventional mathematical models based on differential
75 equations may be difficult to find (although Day et al. 2020 show how a complex situation can
76 be analysed). Numerical solutions can still be found, of course, but agent-based simulations
77 can be used instead (e.g. Roche, Drake & Rohani 2011). These can incorporate any number of
78 parameters with ease, at the expense of making it difficult to explore the parameter space
79 thoroughly and correspondingly difficult either to identify or to validate general principles. The
80 most radical approach is to abandon equations altogether, in favour of a set of rules that
81 govern how individuals move around, meet one another, and infect others or are themselves
82 infected. In this report I shall use an agent-based model based on rules rather than equations to
83 describe the evolutionary dynamics of a virus in a vaccinated population. The virus resembles
84 SARS-CoV-2 in some respects, but the model is intended only to suggest some general features
85 of virus evolution, rather than to predict the future course of the pandemic.

86

87 Methods

88 This report is based on an agent-based, quasi-realistic model of an epidemic in an imaginary small
89 town described in a previous report (Bell 2020). The program itself and a detailed account of the
90 model are included in the Supplementary Material.

91 **Variation and mutation.** The virus has two loci that govern its interaction with a host. The
92 Phenotype or P-locus governs the proliferation of the virus within the host, and thereby its
93 virulence and transmissibility. The Immunity or M-locus governs the interaction of the virus
94 with the immune system of the host. The ancestral strain responsible for the outbreak has

95 given states at the P-locus and the M-locus that determine its transmissibility, virulence and
96 immune characteristics.

97 Each locus consists of a string of given length of binary digits ('bits'), one of which is switched
98 from one state to the other when mutation occurs. Competition within the host is neglected,
99 for simplicity, so that only successful mutations are considered – those that both arise and
100 become fixed within an infected host, which then transmits the mutant strain. The effect of a
101 mutation at the P-locus is to increase viral titre by a random factor from the value in the
102 ancestral strain towards some fixed maximal value, leading to an increase, by a different
103 random amount, in both transmissibility and virulence. A mutation at the M-locus alters the
104 antigenic properties of the virus and may thereby enable it to evade the immune system of the
105 host.

106 The number of mutations at either locus is limited by the number of times that the virus is
107 transmitted, which for the ancestral strain is at most equal to the size of the host population.
108 We can then define the potential mutation supply for either locus as $U_{\text{Pot}} = HuL$, where H is the
109 number of individuals in the host population, u is the fundamental rate of mutation per site,
110 and L is the number of mutable sites at the locus. In an unvaccinated population, at least one
111 mutation will arise, on average, if $U_{\text{Pot}} > 1$. Mutation and the initial spread of a mutant strain
112 are stochastic events, however, so no mutation may occur, or, if it does, the new strain may
113 soon become extinct. In a vaccinated population the condition is more stringent because a
114 successful mutant must arise before the host population is fully vaccinated.

115 **Immune response.** In response to infection by a strain of virus with a given epitope sequence at
116 its M-locus, the host generates an antibody with the complementary sequence and stores this
117 in its immune memory. (For example, if the virus M-locus is 001011001, the host generates and
118 stores 110100110.) If the host survives and is subsequently exposed to the same strain of virus,
119 it is able to express the stored sequence and may thereby disable the virus. The effectiveness
120 of the immune response depends on how the clearance of the virus is related to the degree of
121 complementarity. The simplest mechanism is always to activate the immune response if
122 complementarity equals or exceeds some threshold value, or alternatively the probability of
123 activation might decrease continuously as some function of decreasing complementarity.

124 **Vaccination.** In vaccinated hosts, the sequence complementary to the vaccine sequence is
125 likewise stored in immune memory. Vaccination is akin to previous infection by a virus whose
126 sequence at the M-locus corresponds to the vaccine. It will thereby confer some degree of
127 immunity to this strain of the virus: complete immunity for a highly efficacious vaccine, but only
128 partial immunity for a low-efficacy vaccine. If the host is exposed to a strain of virus whose M-
129 locus does not correspond exactly with the vaccine, its immunity will be less, to an extent that
130 depends on the specificity of the vaccine. A narrow vaccine will activate the immune response,
131 but only if it is fully complementary to an invading virus; a broad vaccine may be less effective
132 against a fully complementary virus, but might be active to some extent against a strain with
133 partial complementarity.

134 If an individual becomes infected by the virus despite vaccination or previous infection, the
135 virus population within the host will grow, as it does in naïve hosts. It may grow to a different
136 titre in naïve, recovered and vaccinated individuals, however, so that its virulence and

137 transmissibility will in general differ between these categories of host. A weak vaccine permits
138 a higher titre and thereby greater virulence and transmissibility, whereas a strong vaccine
139 suppresses the growth of the virus and reduces both virulence and transmissibility, relative to
140 naïve unvaccinated individuals.

141 **Standard parameter set.** The simulations reported here were conducted in a small town of
142 about 4000 people with a demographic and occupational profile similar to that of many
143 communities in Europe and North America. The parameters for the simulations reported here
144 were chosen so that an epidemic infects about 75 – 80% of an unvaccinated population over a
145 period of about 100 days before the virus is no longer able to propagate and becomes locally
146 extinct. The authorities may take measures to control, curtail or delay the epidemic, including
147 vaccination. A vaccination program proceeds as a specified series of cohorts identified by age
148 and occupation. In the simulations reported here, vaccination gives permanent protection and
149 all individuals are vaccinated during the program, regardless of infection history and current
150 status, although the model allows incomplete compliance and temporary immunity.

151

152 Results and Discussion

153

154 **Weak and strong vaccines.** The effects of mutation at the P-locus on virulence and
155 transmissibility are antagonistic, because hosts infected by a mutant strain with increased
156 virulence are more likely to die and thus unable to transmit the virus. The spread of a mutant
157 strain is therefore governed by the balance of positive and negative effects on the rate of

158 transmission, which is expected to lead to an optimal intermediate level of virulence (Anderson
159 & May 1982). This balance may be shifted by vaccination. In particular, the cost of virulence
160 arising from the excess death rate of infected hosts is reduced if the virus is able to grow slowly
161 in vaccinated hosts, so that selection will favour increased growth and hence increased
162 virulence. This possibility has led to the controversial conclusion that a weak vaccine may
163 actually favour the spread of more virulent strains of the virus (Gandon et al. 2001; Gandon et
164 al. 2003).

165 The simplest model is to suppose that naïve hosts are completely susceptible and recovered
166 hosts are completely immune, whereas vaccinated hosts have an intermediate level of
167 immunity. The virulence of the ancestral strain of virus in those vaccinated hosts which
168 become infected is likewise intermediate between its value in naïve and recovered hosts. We
169 then expect that greater virulence will evolve in vaccinated hosts, and consequently that
170 virulence in naïve hosts will increase as a correlated response to selection. As mutations at the
171 P-locus arise and spread by virtue of their increased transmissibility, the mean virulence of the
172 virus, over all infected hosts, will tend to increase. The maximum value of mean virulence
173 during the epidemic is shown in Figure 1 for replicate populations with different vaccination
174 states. A weak vaccine does induce the evolution of somewhat greater virulence in naïve hosts,
175 relative to a strong vaccine, but the difference between weak and strong vaccines is not
176 formally significant ($F = 3.4$; $df = 1, 38$; $0.1 > P > 0.05$), and virulence is greatest in unvaccinated
177 populations because of their greater mutation supply. Selection for increased virulence is
178 expected to be slight, because individuals who die have already had the opportunity to transmit
179 the virus (Day et al. 2020). It is also highly variable because it depends on two stochastic

180 factors: the effect of mutations on virulence and transmissibility, and their timing relative to the
181 vaccination program. The incidence of cases with severe symptoms (which may result in death)
182 is increased by mutation and reduced by vaccination, while the difference between weak and
183 strong vaccines is imperceptible (Supplementary Figure 1).

184 **Evolution of antigenic evasion.** Without variation at the M-locus, vaccination using the
185 ancestral strain as a model halts the epidemic and substantially reduces the number of cases,
186 provided that it is efficacious, universally administered and timely. Even if these conditions are
187 met, however, mutations at the M-locus may enable the virus to spread, provided they are
188 sufficiently frequent. If $U_{Pot} < 1$ it is unlikely (but not impossible) that any mutations will occur,
189 and vaccination will be effective. If $U_{Pot} > 1$ then escape mutations will usually (although not
190 always) occur, and lineages bearing these mutations can spread even in a fully vaccinated
191 population because they render the whole population susceptible: mutations of this sort create
192 their own potential mutation supply, and may thereby give rise to an indefinite series of further
193 mutations in the future. The succession of mutant strains can shift the population from one
194 dynamic regime to another (Supplementary Figure 2). In the first place, a mutant strain may
195 spread after the initial collapse of the epidemic, which will follow from vaccination or simply
196 from the reduced availability of hosts who are susceptible to the ancestral strain. As the
197 mutant strain in turn declines in abundance it can be replaced by a third mutant strain, if one
198 should occur, and in this way give rise to a more or less regular succession of disease cycles.
199 Alternatively, if $U_{Pot} \gg 1$ mutant strains might arise so frequently that several will be circulating
200 at any given time, and the virus might be maintained for a long period of time with irregular
201 low-amplitude fluctuations over time, before eventually becoming stochastically extinct. Figure

202 2 shows that the shift from a short-term regular epidemic to a longer-term evolving epidemic in
203 a vaccinated population occurs around the critical potential mutation supply of $U_{Pot} \approx 1$. The
204 shift from one dynamic regime to another is thus governed by the potential mutation supply.
205 No such shift will occur in an unvaccinated population, because escape mutations will spread
206 only in recovered individuals. Most of these will appear relatively early in the epidemic
207 (because the rate of transmission and thus the opportunity for mutation is greatest at this
208 time), before many host individuals have recovered from the infection and are immune to the
209 ancestral strain. Consequently, the density of susceptible hosts is not much greater for a
210 mutant and its selective advantage is only modest, so the lineage will spread only slowly and is
211 likely to die out. In a vaccinated population, on the other hand, the transmission of the
212 ancestral strain is blocked by the vaccine, the selective advantage of mutants is correspondingly
213 much greater, and a mutant lineage can rapidly expand to fill the ecological space provided by
214 vaccinated but susceptible hosts. The potential for vaccination-driven emergence of a vaccine-
215 resistant strain was demonstrated analytically by Scherer & McLean (2002).

216 The outcome of selection in a vaccinated population will often be a shift in dynamic regime and
217 a substantial increase in the number of cases, relative to a comparable unvaccinated
218 population, because some host individuals will be infected twice, once by the ancestral strain
219 and again by the mutant. If several mutations occur sequentially, each will be likely to spread as
220 its predecessor declines, so that many individuals are infected several times and the total
221 number of cases greatly exceeds that of an unvaccinated population. The effect of vaccination
222 on the total number of cases when there is recurrent mutation at the M-locus is illustrated in
223 Figure 3. In some examples, where no mutation occurs until late in the vaccination program, or

224 fails to spread if it occurs earlier, the resurgence of the epidemic does not occur and
225 vaccination reduces the overall number of infections. In most cases, however, recurrent
226 mutation with $U_{Pot} \approx 1$ prolongs the epidemic and causes many more cases than would occur in
227 a comparable unvaccinated population, where the number of cases is limited by the number of
228 host individuals, provided that infection confers immunity in recovered individuals. This
229 surprising result follows from the greater selective advantage of strains which carry a mutation
230 at the M-locus in a vaccinated population.

231 **Narrow and broad vaccines.** A narrow vaccine is highly specific, being efficacious against the
232 ancestral strain of virus but giving little protection against antigenic variants. A broad vaccine is
233 less active against the ancestral strain, but provides some protection against variants, even
234 when complementarity is incomplete. The perfect vaccine would be efficacious against both
235 the ancestral strain and mutational variants, but efficacy and breadth might be incompatible.
236 Suppose that the efficacy E of a vaccine, the probability that it confers immunity to a given
237 strain of virus, declines exponentially with decreasing complementarity C , such that $E_C = E_{limit}$
238 $\exp[-k(C_{limit} - C)]$, where E_{limit} is the efficacy of the vaccine when it corresponds perfectly with
239 the M-locus of the strain ($C = C_{limit}$). A broad vaccine would then have a shallow slope (small k),
240 giving greater protection against similar strains at the expense of giving less protection against
241 matching strains (low E_{limit}). Hence, a broad vaccine might reduce the selective advantage of
242 mutations at the M-locus, and thereby prevent the epidemic from becoming cyclical or
243 perpetual; for example, Fryer & McLean (2011) suggested that broad vaccines against HIV
244 might suppress the emergence of escape mutants. This idea was investigated by manipulating
245 E_{limit} and k to produce narrow, medium and broad vaccines in populations where mutation at

246 the M-locus (but not the P-locus) was allowed. The outcome showed little effect of vaccine
247 design, but an overwhelming effect of mutation supply. In unvaccinated populations the
248 epidemic was limited, as before (Figure 4). Vaccinated populations behaved in the same way,
249 provided that few mutations occurred, since most mutations fail to spread. Above a certain
250 threshold, however, escape mutations shifted the population into a new regime with multiple
251 infections per host individual and far more cases overall. A narrow vaccine may predispose the
252 population to shift, but the effect is modest at best ($X^2 = 3.0$, $df = 1$, $0.1 > P > 0.05$), whereas the
253 effect of mutation supply is clear.

254 **Paradox of vaccination.** Simple escape mutations at the M-locus not only prolong the epidemic
255 and increase the number of cases, but may shift the dynamic regime from a limited short-term
256 epidemic to a perennial endemic disease. This outcome is not borne out by empirical data on
257 the effects of vaccination. Most vaccines remain effective for long periods of time, and in many
258 cases (e.g. measles) resistance has never evolved. The rarity of vaccine resistance is particularly
259 striking in contrast with the rapid evolution of resistance to antibiotics (Mishra et al. 2012).

260
261 This may be a consequence of mutation supply within the host: the immune reaction clears an
262 infection when the pathogen population is still small, whereas antibiotics are usually
263 administered only when the pathogen population has become large enough to cause overt
264 symptoms (Kennedy & Read 2017, 2018; Bell & MacLean 2018). Nevertheless, escape variants
265 have been reported for some vaccines, for example hepatitis B. Even in this case, however,
266 they have failed to eliminate the ancestral strain, perhaps because of a high cost of resistance
267 (Francois et al. 2001). Resistance might often be costly because viral genomes are very

268 compact and hence likely to be disrupted by random change. Reid et al. (2019) reviewed a wide
269 range of models designed to predict the consequences of vaccination against an evolving
270 pathogen, most of which referred to a specific vaccine against a particular pathogen. They
271 concluded that: “Overall, the studies of vaccines that have been in use, have trial data, or have
272 existing homologs predicted positive health outcomes despite vaccine resistance.” The few
273 exceptions require conditions which are thought to be exceptional, such as high levels of cross-
274 immunity between contemporary strains (e.g. Worby et al. 2017). Hence, real-world studies
275 overwhelmingly predict that vaccination will be beneficial, even in the long term against an
276 evolving pathogen.

277 Vaccine resistance is expected to be rare if the potential mutation supply is usually very low.
278 The rate of mutation per base pair per replication is much higher in RNA viruses than in
279 organisms with a DNA genome, so the requirement that the potential mutation supply be small
280 might seem unlikely. However, the second reason given by Kennedy & Read (2017) for the
281 rarity of vaccine resistance is the extensive redundancy of most vaccines: a variety of antibodies
282 can be produced by the host in response to the presentation of several epitopes on each of
283 several antigens by the pathogen. Consequently, an effective immune response can be
284 mounted against strains similar to, but not identical with, the ancestral strain, leading to the
285 broad cross-reactivity identified by McLean (1995) as a feature of successful vaccines. Most
286 vaccines might therefore tend to be both broad and efficacious, however unlikely such a
287 felicitous combination might seem. Consider a narrow vaccine, which is 100% efficacious
288 against the exactly corresponding sequence at the M-locus of the virus, but otherwise
289 completely inactive, so that a single mutation confers immunity. If the vaccine were somewhat

290 broader, so that it also gave 100% protection against strains that differed at any one of the sites
291 of the M-locus, then a double mutation would be required for immunity. The potential
292 mutation supply now refers to the supply of *double* mutants, which will be proportional to u^2 .
293 The vaccine now restricts the spread of the virus to a single epidemic, and no shift of dynamic
294 regime occurs (Figure 5). This does not require perfect efficacy. If the vaccine is only 70%
295 efficacious, the number of cases is usually increased only slightly, relative to a perfectly
296 efficacious vaccine. On rare occasions, however, a mutation may occur early in the epidemic
297 and drift to high frequency, such that the mutant lineage is liable to undergo a second mutation
298 before the (rather ineffective) vaccination program has eliminated the virus; it is only in this
299 case that a shift in dynamic regime will occur.

300 **Revaccination.** If mutation at the M-locus is generating escape mutants, revaccination might
301 halt the spread of a mutant strain, using the most abundant contemporary strain itself as the
302 model for the vaccine. In an unvaccinated population the epidemic will take its normal course,
303 and die out when the virus is no longer able to proliferate. If a single mutation appears before
304 a vaccination program is fully implemented the ancestral strain is eliminated by the vaccine,
305 producing a fall in the number of new cases, but the mutant strain then spreads rapidly. The
306 outcome is an increase in the total number of cases, as previously noted. Revaccination against
307 the most abundant contemporary strain usually palliates this trend, although it is only effective
308 if administered swiftly (Supplementary Figure 3). Revaccination can reduce the number of
309 excess cases if it is administered soon after the previous vaccination program but if it is delayed
310 its effect is less because the mutant strain has already reached high frequency (Figure 3).

311 **Recombination.** The selection of more highly transmissible variants arising by mutation at the
312 P-locus has only a modest effect on the dynamics of the epidemic, but might be more
313 consequential if they were linked with escape mutants at the M-locus. The probability of
314 stochastic loss of a novel mutation at the M-locus, for example, will be less if it has become
315 linked to a highly transmissible variant caused by mutation at the P-locus. Linkage might arise
316 in two ways: through sequential mutation, or through coinfection leading to intergenic
317 recombination. In principle, recombination is likely to be the more important if coinfection is
318 much more frequent than the rate of mutation per locus. To investigate this possibility,
319 populations with different combinations of mutation and recombination can be set up: neither
320 mutation nor recombination; mutation at the P-locus alone; mutation at the M-locus alone;
321 mutation at both loci without recombination; and mutation at both loci with recombination.
322 The outcome of the experiment is shown in Figure 6. The first three treatments yield results
323 similar to those already described. With neither mutation nor recombination, the epidemic
324 follows a normal course, infecting a substantial proportion of the population before being
325 knocked down by vaccination and then dying out; mutations at the P-locus which increase
326 virulence and transmissibility cause a modest rise in the number of cases, whereas mutations at
327 the M-locus may lead to a new dynamic regime in which the virus evolves to evade the vaccine,
328 many individuals are infected several times and the total number of cases exceeds the
329 population size. When there is mutation at both loci, but no coinfection, linkage must come
330 about as the consequence of a mutation in one locus occurring in a lineage which already bears
331 a mutation at the other. This causes a modest increase in the average number of cases over
332 the course of the epidemic, relative to a population in which mutation occurs only at the M-

333 locus. If coinfection is permitted, linkage may arise either by sequential mutation or by
334 recombination. This causes a further increase in the average number of cases, suggesting an
335 effect of recombination in addition to the effect of sequential mutation alone. The marginal
336 (additional) effects of mutation and recombination can then be calculated from the five
337 treatment combinations:

338	No mutation or recombination	Phe- Imm- Rec-	scaled to 1
339	Mutation at P-locus only	Phe+ Imm- Rec-	1.27
340	Mutation at M-locus only	Phe- Imm+ Rec-	4.38
341	Interaction of P-locus and M-locus	Phe+ Imm+ Rec-	1.23
342	Recombination	Phe+ Imm+ Rec+	1.21

343
344 These marginal effects largely depend on whether the dynamic regime is changed by the
345 treatment rather than a smooth increase in the average number of cases. The experiment
346 illustrates the overriding qualitative effect of escape mutations at the M-locus and the modest
347 but appreciable quantitative contribution of mutations at the P-locus, either alone or in
348 combination with mutations at the M-locus. It is possible that the principal effect of
349 recombination, when coinfection is much more frequent than successful mutation, is to
350 facilitate the transition from a regular limited epidemic to a long-term endemic infestation.

351 Conclusions

352 Virus populations will evolve swiftly when there is genetic variation for immune characteristics
353 or within-host growth. The maximum quantity of variation is the potential mutation supply,

354 which is the maximum number of mutations in a simple epidemic: $U_{Pot} = HuL$. This suggests
355 that there is a threshold at about $U_{Pot} = 1$, above which mutations are likely to occur, leading to
356 the evolution of the virus population. The simulations based on an agent-based model broadly
357 support the analytical results of previous theory.

358 Virulence tends to increase when $U_{Pot} > 1$ for the P-locus, and this may be modulated by vaccine
359 design. The total number of cases during the epidemic cannot exceed the number of host
360 individuals, however, and will usually be substantially less. Any increase in morbidity will be
361 limited, especially when a vaccination program has been implemented.

362 Mutation at the M-locus may have a qualitatively different outcome. Escape mutations at the
363 M-locus not only prolong the epidemic and increase the number of cases, but may shift the
364 dynamic regime from a limited short-term epidemic to a perennial endemic disease. Both
365 regimes are examples of evolutionary rescue, because a population that would otherwise
366 become extinct is perpetuated by natural selection through the spread of mutants each of
367 which has a positive rate of increase. Simulations confirm that a shift of dynamic regime is likely
368 to occur when the potential mutation supply $U_{Pot} > 1$. This is because each successful mutation
369 creates its own potential mutation supply and can generate as many new mutations as the
370 ancestral strain. If escape mutations which evade vaccines based on the ancestral virus become
371 established then the epidemic may be prolonged almost indefinitely, unless it can be halted by
372 prompt revaccination. A vaccine which is both broad and efficacious will prevent this shift
373 because only multiply-mutant or recombinant strains will evade the immune response, and it
374 seems likely that most vaccines have this combination of properties.

375 The classic case of a pathogen that has repeatedly evolved vaccine resistance through mutation
376 and recombination is influenza A virus (reviewed by Taubenberger & Kash 2010). SARS-Cov-2 is
377 also a single-stranded RNA virus, and it is possible that escape mutations will occur. The
378 evolution of SARS-Cov-2 up to February 2021 has been reviewed by Rochman et al. (2021).
379 Altmann, Boyton & Beale (2021) have reviewed the evidence (to March 2021) that mutations
380 affecting the spike protein of SARS-Cov-2 cause changes in viral titre or immune recognition or
381 both. All the variants of concern involve several mutations in the spike protein and must have
382 arisen by sequential mutation or recombination or both. Several variants that have spread
383 recently are resistant to plasma from recovered or vaccinated individuals (Wang et al. 2021),
384 and deletions that cause some degree of resistance arise repeatedly in infected patients
385 (McCarthy et al. 2021). The crucial statistic is the frequency of reinfection, which is poorly
386 defined (but not vanishingly small) for SARS-Cov-2 (Boyton & Altmann 2021).

387 Although deterministic processes such as mutation supply and natural selection can be
388 recognised, stochastic processes are also important and may be decisive. All the experiments
389 described in this report show a great deal of scatter, with extreme outliers in some cases and
390 qualitatively different outcomes in others, even when the initial state of the population, the
391 mutation rate and the vaccination schedule are identical. The small size of the host population
392 is responsible for some of this scatter, but the initial fate of a mutant is influenced by the small
393 size (initially a single individual) of the mutant lineage itself, independently of the overall size of
394 the population (see Day et al. 2020, Supplementary Material, equation S8), while the schedule
395 and phenotypic effects of mutations are intrinsically stochastic. In some cases, the effect of a
396 treatment (such as vaccination) or a condition (such as mutation) are quite clear, but in others

397 the outcome is dominated by stochastic processes and the consistent effect of an intervention
398 is difficult to discern. Day et al. (2020) make a similar point on the basis of an analytical
399 mathematical model.

400 More broadly, the dynamics of an epidemic, after the passage of the initial strain, are governed
401 by evolutionary processes involving variation and natural selection. In order to evaluate these
402 processes the most urgent preliminary task is to estimate the potential mutation supply for loci
403 which govern the immune properties and within-host growth of the virus. Practical
404 recommendations which do not include expert guidance on the potential for evolutionary
405 change during the course of an epidemic will be incomplete.

406

407 Acknowledgments. I am grateful for comments on an earlier version of this article by Austin
408 Burt and Craig MacLean. My research is supported by a Discovery Grant from the Natural
409 Sciences and Engineering Research Board of Canada (Grant RGPIN/6945-2013).

410

411

412 **Figure legends**

413 Figure 1. Evolution of virulence in populations with mutation at the P-locus. Plots show the
414 maximum value during the epidemic of mean virulence in naïve (never infected) hosts, for 20
415 replicate populations with identical initial composition but different vaccination status. The
416 upper limit of variation is set at 3x the initial level of virulence. The values of virulence an

417 Figure 2. Dynamics of vaccinated populations with different rates of mutation at the M-locus. E
418 denotes a regular single epidemic; R recurrent cyclical epidemics; Q long-term infestation without clear
419 cyclical structure. The x-axis is the potential mutation supply U_{Pot} per locus, with the vertical line
420 showing $U_{Pot} \approx 1$ as an average for a vaccinated population in which no mutations occur. The horizontal
421 broken line is the population size. The plot shows 6 replicate populations at each of 10 mutation rates;
422 points have been jiggled slightly for clarity.

423 Figure 3. Effect of vaccination and revaccination on the number of cases under recurrent
424 mutation at the M-locus. Treatments for each of 20 replicate populations have identical initial
425 state; different replicates have same parameter values but different initial states. Ranked by
426 number of cases in the vaccinated population. Vaccination beginning 6 weeks after inoculation,
427 in three cohorts spaced by 3 weeks each; revaccination occurs 10 weeks (short lag) or 16 weeks
428 (long lag) after initial vaccination.

429 Figure 4. The effect of vaccine design and mutation supply on the number of cases during an
430 epidemic. The plot shows 20 independent replicate populations for each treatment. The x-axis
431 shows all mutations, whether or not they spread.

432 Figure 5. Effect of a broad efficacious vaccine. Individuals are supposed to be immune only if
433 the vaccine is completely complementary to the strain they are exposed to, so that a single
434 mutation at the I-locus of the virus suffices for infection, or that they are immune if the vaccine
435 differs by no more than one site from the strain they are exposed to, so that a double mutation
436 at the M-locus of the virus is required for infection. The values plotted are the total number of
437 cases during the epidemic, for pairs of populations (20 replicate pairs) with identical initial
438 state. Solid circles refer to a vaccine with 100% efficacy, open triangles with 70% efficacy. The
439 solid line is the line of equality. The dashed line marks the approximate population size of 4000
440 individuals.

441 Figure 6. The effect of mutation and recombination on the total number of infections during an
442 epidemic in a vaccinated population. Mutation either occurs or not at the P-locus (with probability $2 \times$
443 10^{-4} per bit per replication) and the M-locus (with probability 8×10^{-5} per bit per replication), and
444 recombination occurs after coinfection (with probability 0.01 per transmission) or not. Parameters other
445 than the occurrence of mutation or recombination have the same value in all runs. Each of the five
446 treatments has 20 independent replicates, ranked in order of increasing number of cases in each.

447

448

449

References

450

Altmann, D.M., Boyton, R.J. and Beale, R., 2021. Immunity to SARS-CoV-2 variants of

451

concern. *Science*, 371(6534), pp.1103-1104.

452

Anderson, R. M. & May, R. M. 1982 Coevolution of hosts and parasites. *Parasitology* **85**, 411-

453

426.

454

Anderson, R.M. and May, R.M., 1992. *Infectious diseases of humans: dynamics and control*.

455

Oxford University Press.

456

Bell, G. and MacLean, C., 2018. The search for 'evolution-proof' antibiotics. *Trends in*

457

Microbiology, 26(6), pp.471-483.

458

Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S., Angeletti, S. and Ciccozzi, M., 2020. The

459

2019-new coronavirus epidemic: evidence for virus evolution. *Journal of medical*

460

virology, 92(4), pp.455-459.

461

Boyton, R.J. and Altmann, D.M., 2021. Risk of SARS-CoV-2 reinfection after natural

462

infection. *The Lancet*, 397(10280), pp.1161-1163.

463

Brauer, F., 2017. Mathematical epidemiology: Past, present, and future. *Infectious Disease*

464

Modelling, 2(2), pp.113-127.

465

Day, T. and Gandon, S., 2007. Applying population-genetic models in theoretical evolutionary

466

epidemiology. *Ecology Letters*, 10(10), pp.876-888.

467

Day, T., Gandon, S., Lion, S. and Otto, S.P., 2020. On the evolutionary epidemiology of SARS-

468

CoV-2. *Current Biology*, 30(15), pp.R849-R857.

- 469 Day, T. and Proulx, S.R., 2004. A general theory for the evolutionary dynamics of virulence. *The*
470 *American Naturalist*, 163(4), pp.E40-E63.
- 471 François, G., Kew, M., Van Damme, P., Mphahlele, M.J. and Meheus, A., 2001. Mutant hepatitis
472 B viruses: a matter of academic interest only or a problem with far-reaching
473 implications?. *Vaccine*, 19(28-29), pp.3799-3815.
- 474 Fryer, H.R. and McLean, A.R., 2011. Modelling the spread of HIV immune escape mutants in a
475 vaccinated population. *PLoS computational biology*, 7(12), p.e1002289.
- 476 Gandon, S. and Day, T., 2007. The evolutionary epidemiology of vaccination. *Journal of the*
477 *Royal Society Interface*, 4(16), pp.803-817.
- 478 Gandon, S., Mackinnon, M. J., Nee, S. & Read, A. F. 2001. Imperfect vaccines and the evolution
479 of pathogen virulence. *Nature* **414**, 751–756.
- 480 Gandon, S., Mackinnon, M., Nee, S. & Read, A.F. 2003. Imperfect vaccination: some
481 epidemiological and evolutionary consequences. *Proceedings of the Royal Society of London B*
482 270: 1129-1136.
- 483 Gog, J.R. and Grenfell, B.T., 2002. Dynamics and selection of many-strain pathogens.
484 *Proceedings of the National Academy of Sciences*, 99(26), pp.17209-17214.
- 485 Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L., Daly, J.M., Mumford, J.A. and Holmes, E.C.,
486 2004. Unifying the epidemiological and evolutionary dynamics of
487 pathogens. *science*, 303(5656), pp.327-332.

- 488 Kennedy, D.A. and Read, A.F., 2017. Why does drug resistance readily evolve but vaccine
489 resistance does not?. *Proceedings of the Royal Society B: Biological Sciences*, 284(1851),
490 p.20162562.
- 491 Kennedy, D.A. and Read, A.F., 2018. Why the evolution of vaccine resistance is less of a concern
492 than the evolution of drug resistance. *Proceedings of the National Academy of
493 Sciences*, 115(51), pp.12878-12886.
- 494 Lenski, R.E. and May, R.M., 1994. The evolution of virulence in parasites and pathogens:
495 reconciliation between two competing hypotheses. *Journal of theoretical biology*, 169(3),
496 pp.253-265.
- 497 May, R.M. and Anderson, R.M., 1983. Epidemiology and genetics in the coevolution of parasites
498 and hosts. *Proceedings of the Royal society of London. Series B. Biological sciences*, 219(1216),
499 pp.281-313.
- 500 McCarthy, K.R., Rennick, L.J., Nambulli, S., Robinson-McCarthy, L.R., Bain, W.G., Haidar, G. and
501 Duprex, W.P., 2021. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody
502 escape. *Science*, 371(6534), pp.1139-1142.
- 503 McLean AR. 1995 Vaccination, evolution and changes in the efficacy of vaccines: a theoretical
504 framework. *Proc. R. Soc. Lond. B* 261, 389–393.
- 505 Mishra, R.P., Oviedo-Orta, E., Prachi, P., Rappuoli, R. and Bagnoli, F., 2012. Vaccines and
506 antibiotic resistance. *Current opinion in microbiology*, 15(5), pp.596-602.

- 507 Nelson, M.I. and Holmes, E.C., 2007. The evolution of epidemic influenza. *Nature reviews*
508 *genetics*, 8(3), pp.196-205.
- 509 Reid, M.C., Peebles, K., Stansfield, S.E., Goodreau, S.M., Abernethy, N., Gottlieb, G.S., Mittler,
510 J.E. and Herbeck, J.T., 2019. Models to predict the public health impact of vaccine resistance: A
511 systematic review. *Vaccine*, 37(35), pp.4886-4895.
- 512 Roberts, M.G. and Heesterbeek, J.A.P., 2003. *Mathematical models in epidemiology* (Vol. 215).
513 EOLSS.
- 514 Roche, B., Drake, J.M. and Rohani, P., 2011. An Agent-Based Model to study the epidemiological
515 and evolutionary dynamics of Influenza viruses. *BMC bioinformatics*, 12(1), pp.1-10.
- 516 Rochman, N.D., Wolf, Y.I., Faure, G., Mutz, P., Zhang, F. and Koonin, E.V., 2021. Ongoing global
517 and regional adaptive evolution of sars-cov-2. *Proceedings of the National Academy of*
518 *Sciences*, 118(29).
- 519 Scherer, A. and McLean, A., 2002. Mathematical models of vaccination. *British Medical*
520 *Bulletin*, 62(1), pp.187-199.
- 521 Taubenberger, J.K. and Kash, J.C., 2010. Influenza virus evolution, host adaptation, and
522 pandemic formation. *Cell host & microbe*, 7(6), pp.440-451.
- 523 van Baalen, M. and Sabelis, M.W., 1995. The dynamics of multiple infection and the evolution
524 of virulence. *The American Naturalist*, 146(6), pp.881-910.

Evolution of virus

Bell

525 Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D.

526 and Graham, B.S., 2021. Antibody resistance of SARS-CoV-2 variants B. 1.351 and B. 1.1.

527 7. *Nature*, 593(7857), pp.130-135.

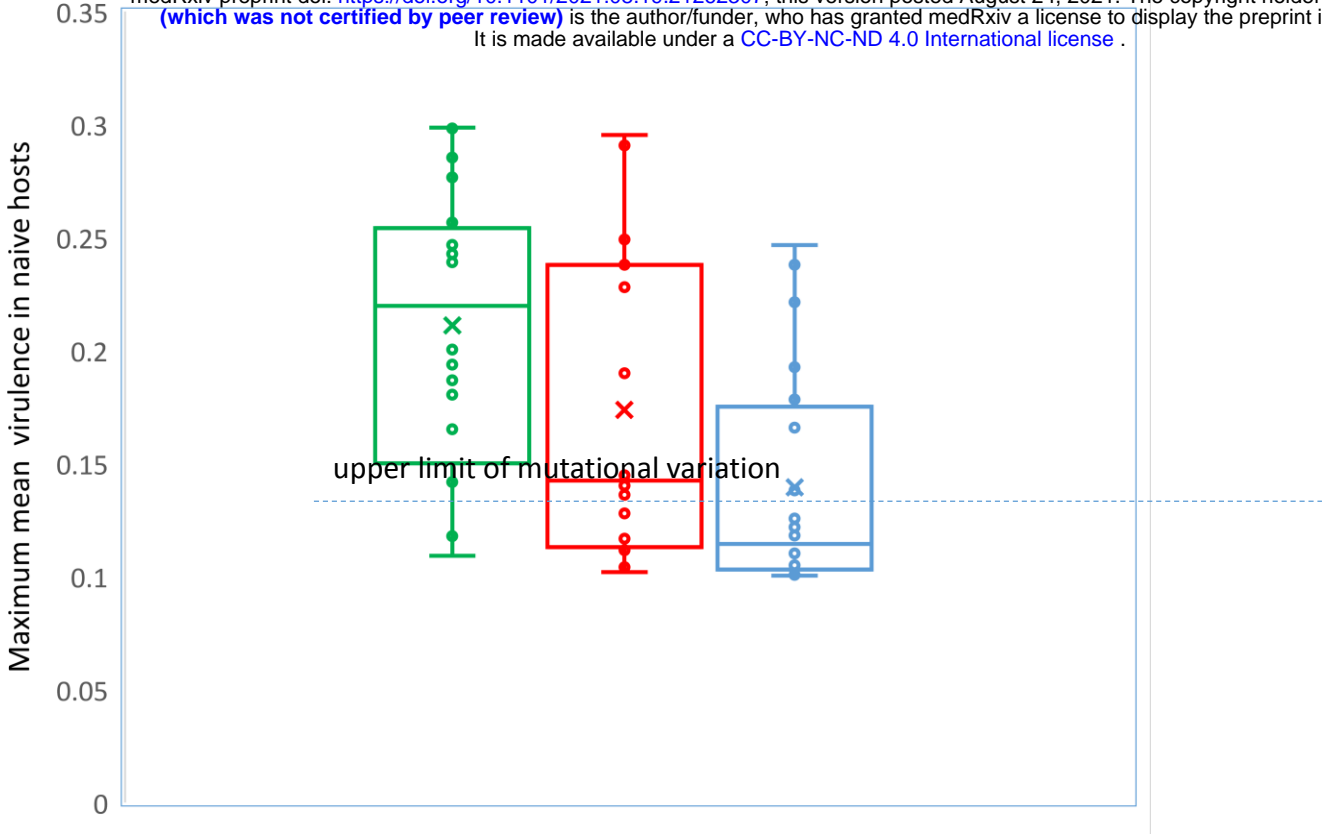
528 Worby, C.J., Wallinga, J., Lipsitch, M. and Goldstein, E., 2017. Population effect of influenza

529 vaccination under co-circulation of non-vaccine variants and the case for a bivalent A/H3N2

530 vaccine component. *Epidemics*, 19, pp.74-82.

531

Bell



540
541
542
543
544
545
546

Initial value

Unvaccinated

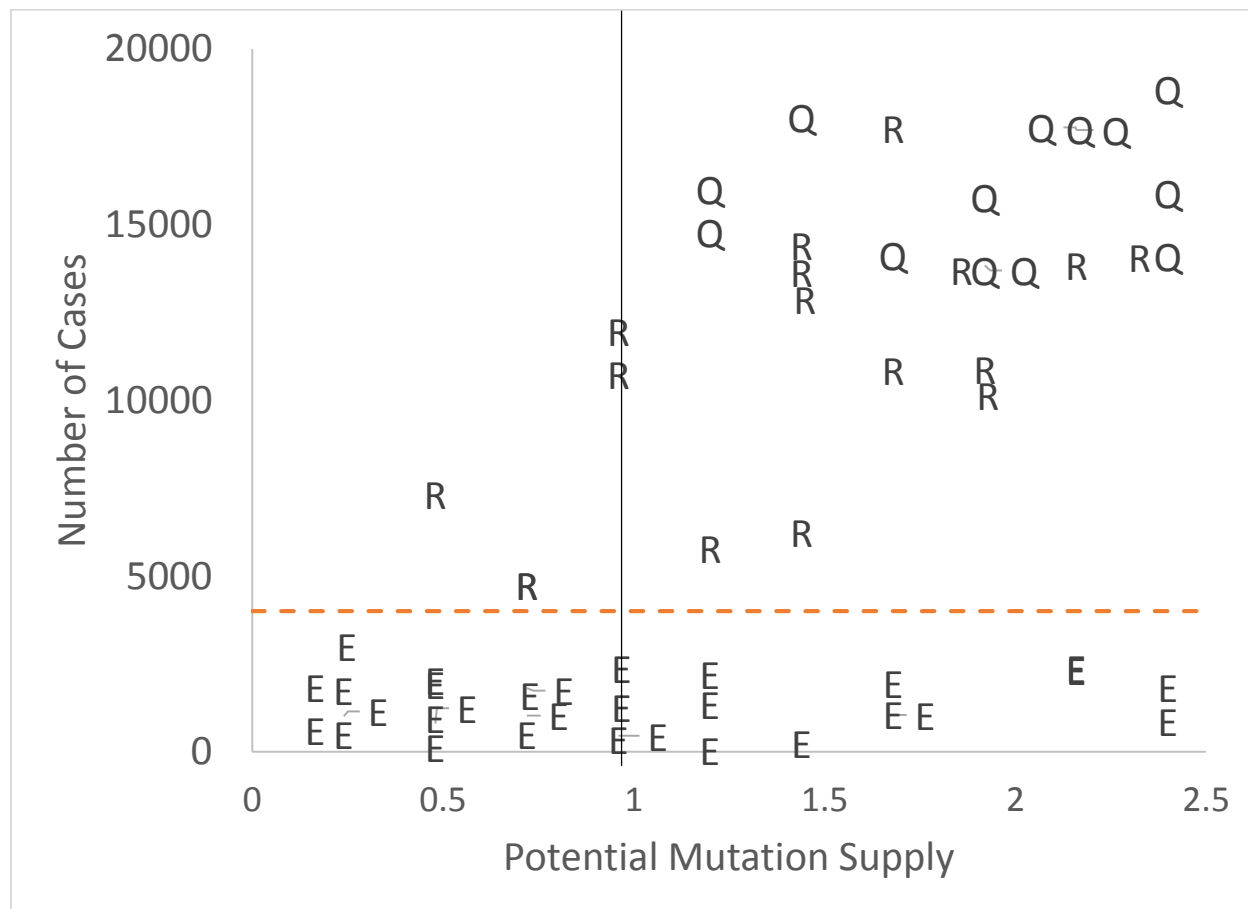
Vaccinated
High efficacy

Vaccinated
Low efficacy

547

548 Figure 2

549



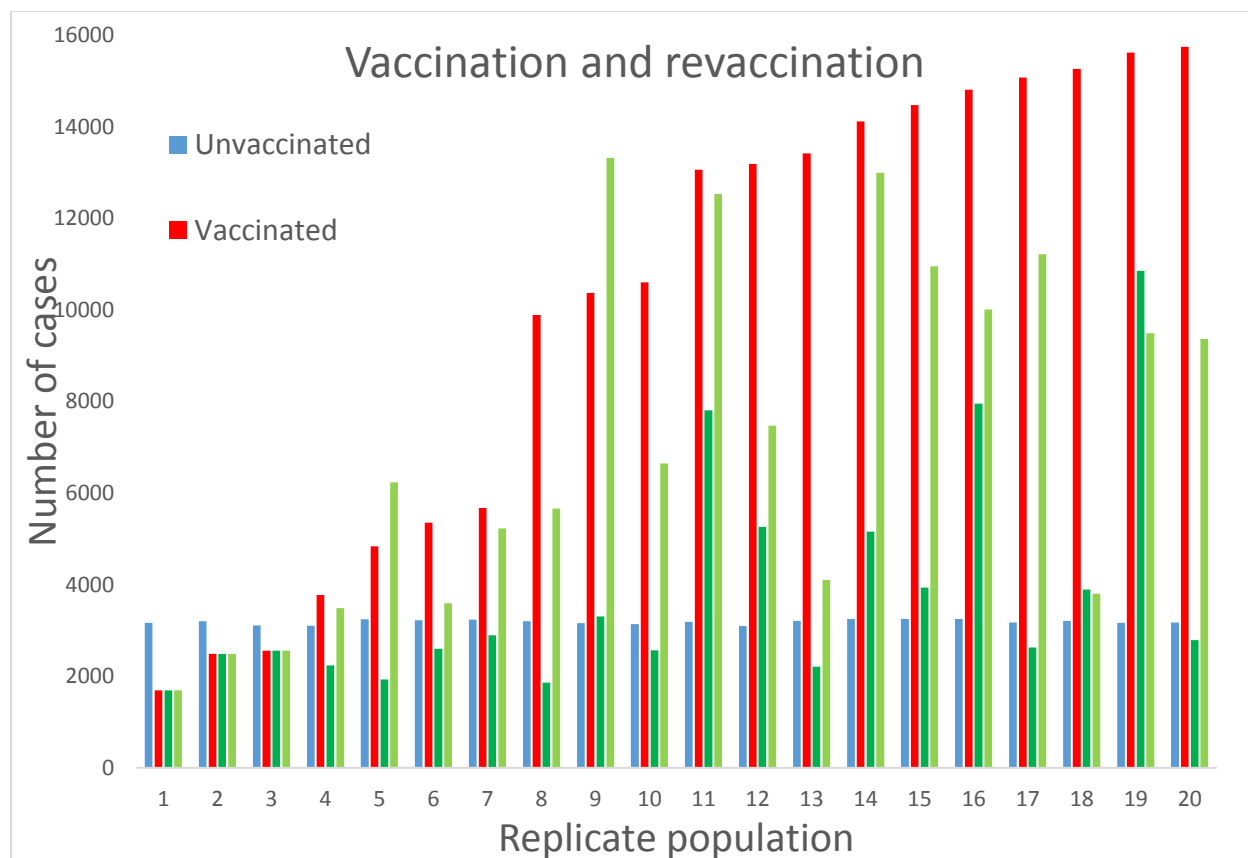
550

551

552

553 Figure 3

554



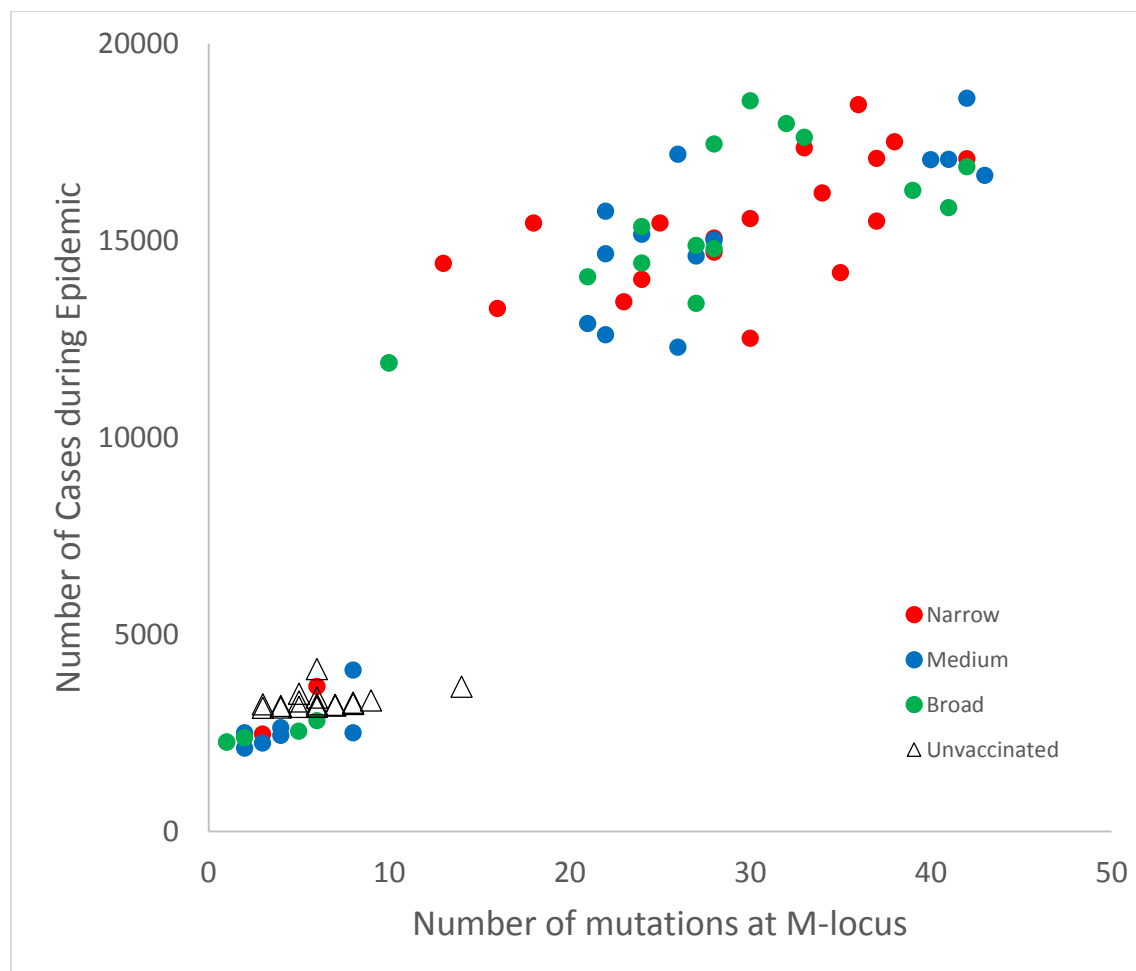
555

556

557

558 Figure 4

559



560

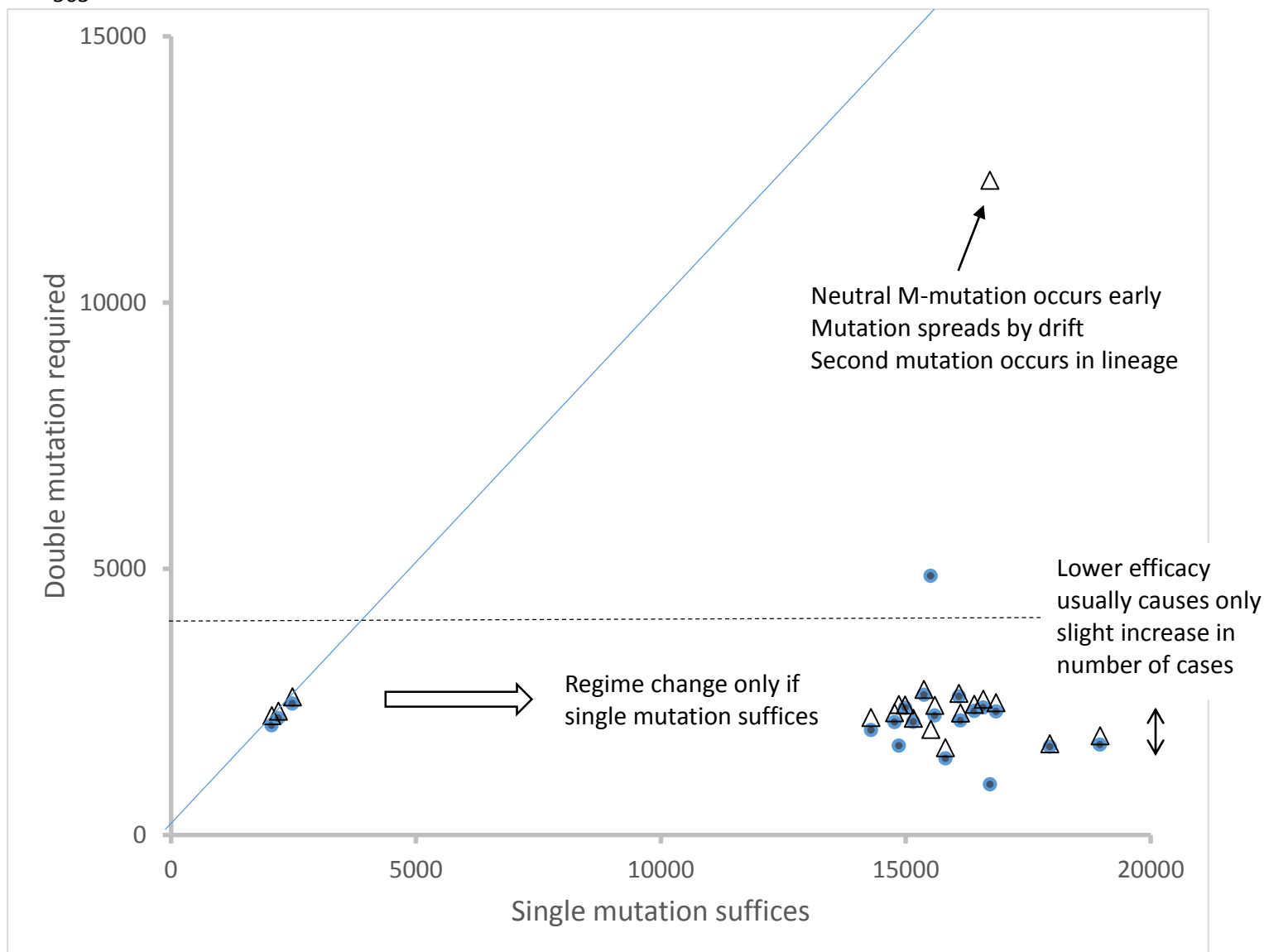
561

562

563 Figure 5

564

565



584

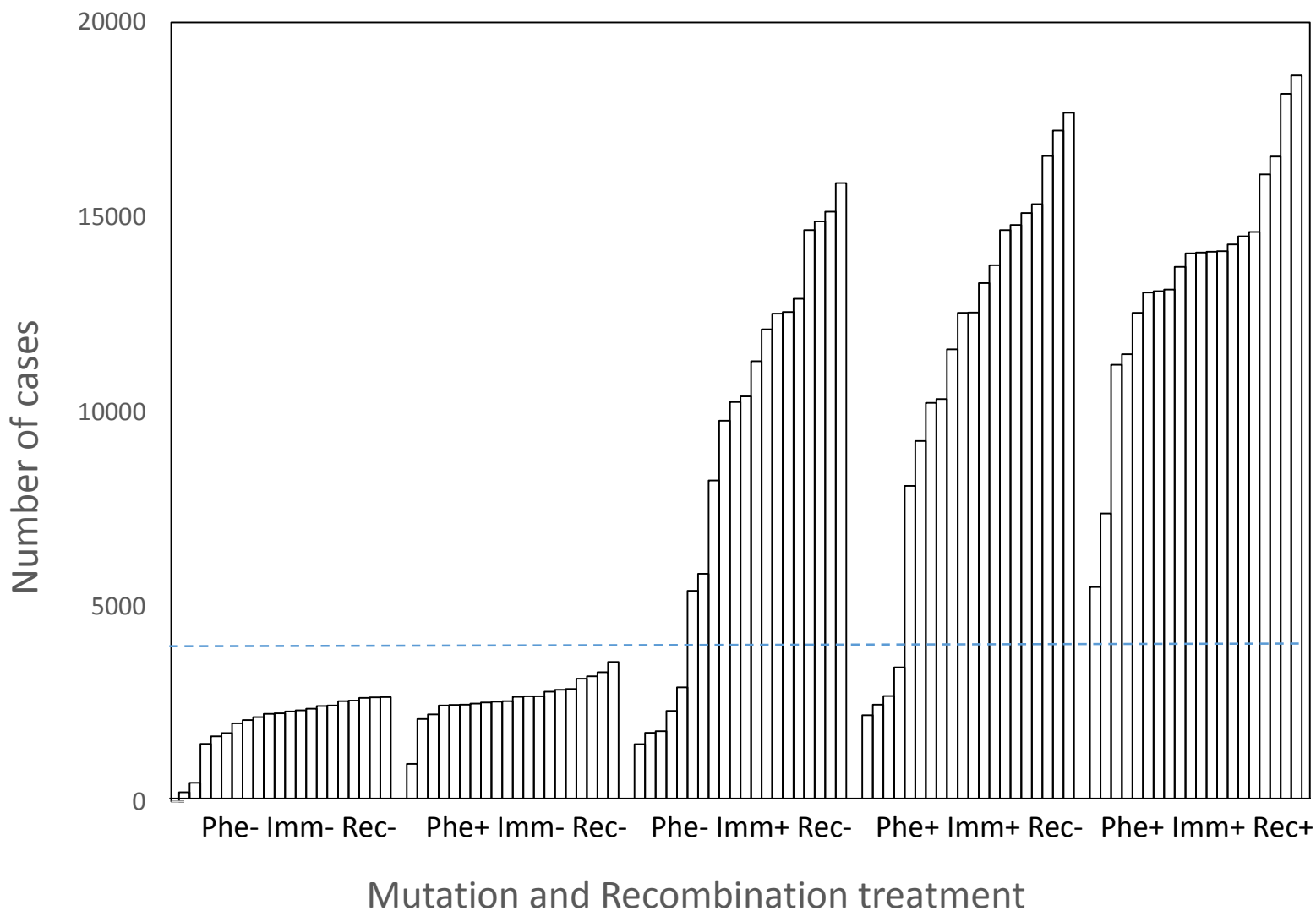
585

586

587

588 Figure 6

589



604