1 **Competitive exclusion strengthens selection for transmissibility and increases the benefit of**

2 **recombination for within-host adaptation.**

3 Short Title: **Competitive exclusion and adaptation at multiple scales.**

4 Nathan T. Jacobs[a], Jeffrey N. Weiser[1a]

5

6 [a]Department of Microbiology, New York University Grossman School of Medicine, New York,

7 NY, USA.

8 [1]Address correspondence to Jeffrey N. Weiser, jeffrey.weiser@nyulangone.org

Abstract

Pathogens experience selection at multiple scales, given the need to transmit between hosts and replicate within them. This presents the challenge of cross-scale selective conflict when adaptations to one scale compromise fitness at another, such as mutations that improve transmissibility but make individuals less competitive within hosts. Selection operates differently at these scales, with tight transmission bottlenecks subjecting pathogen populations to genetic drift, and large population sizes within hosts enabling efficient selection for beneficial mutations. Compounding the reduction in diversity by transmission bottlenecks is the occupant-intruder competitive strategy exhibited by some pathogens, where the first variant to colonize a host prevents later arriving variants from contributing to infection, preventing immigration and turning transmission into a "founder takes all" contest. Here, we used multiple modeling approaches to examine how this behavior affects the efficiency of selection for both transmissibility and within-host fitness. We find that in the face of a trade-off, selection for transmissibility is maximized under a tight transmission bottleneck that minimizes within-host competition during colonization. While mutations with increased within-host fitness are favored during within-host replication, an occupant-intruder strategy prevents these mutants from displacing established residents and propagating across the host population, leading to their extinction if they are insufficiently transmissible. Finally, a model of competition on the scale of the host population revealed that competitive exclusion limits the propagation of mutations with improved within-host fitness, unless resident populations can incorporate alleles from intruding variants by recombination. Thus, competitive exclusion may facilitate the improvement and maintenance of pathogen transmissibility, with directional recombination allowing resident populations to mitigate the potential loss of within-host fitness imposed by this occupant-intruder strategy.

Author Summary

Transmission is a defining feature of infectious diseases, and so a better understanding of how transmissibility evolves is important for improving disease surveillance and prevention. Successful transmission is often achieved by a small number of individuals which, after establishing residency in a host, may prevent newcomers from participating in infection. Here, we use modeling to examine how competitive exclusion of challengers by resident populations affects the balance between within-host competitive ability and transmissibility. We find that competitive exclusion strengthens selection for transmissibility by disproportionately benefitting the first variant to colonize a host and preventing mutants that may be more competitive but less transmissible from displacing established residents. Competitive exclusion also limits the propagation of mutants that improve within-host fitness without reducing transmissibility, however, increasing the advantage of recombination that allows resident populations to acquire beneficial alleles from challengers. Competitive strategies that allow pathogens to "claim ownership" of hosts may thus help pathogen populations maintain transmissibility, with genetic recombination facilitating within-host adaptation through the incorporation of beneficial alleles from challengers.

Introduction

Pathogen life history is defined by replication to large population sizes within individuals hosts and the need to transmit to new ones before clearance or host death. Pathogens thus face selection at both the within- and between-host scales, presenting a challenge for evolution when adaptations at one scale reduce fitness at another. Within-host adaptation may be short-sighted, for example, favoring growth in the current host while totally compromising transmissibility [1]. And on the epidemiological scale, selection for transmissibility occurs at the interface between hosts, but mutations that increase transmissibility may be selected against during within-host growth, even

55 if the cost to within-host fitness is mild. This potential for cross-scale selective conflict raises the
56 question of how pathogens can balance both between- and within-host fitness.

57 Pathogen transmission is often characterized by tight genetic bottlenecks in which substantial
58 diversity is lost between the donor and recipient hosts [2-5]. While contraction in diversity may be
59 the result of strong purifying selection [2, 6], experimental analyses of transmission using neutral
60 markers have shown substantial loss of neutral diversity as well, suggesting that infection is often
61 established by one or a few individual virions or bacterial cells [3, 5]. The loss of genetic diversity
62 at this transmission bottleneck causes both a founder effect, wherein the entire population in a
63 recipient host descends from these few variants, and genetic drift, wherein benefical alleles may
64 be lost and deleterious alleles fixed due to the small size of the founding population.

65 Both host susceptibility and pathogen behavior can influence the size of the transmission
66 bottleneck and the balance between selection and genetic drift in transmission. If a host is
67 sufficiently resistant to colonization such that most exposures do not result in transmission, then
68 successful colonization could be mediated by rare events in which a single individual establishes
69 infection on its own. This may explain the tight bottlenecks observed in settings with repeated
70 exposures, such as household transmission in humans or between co-housed animals, as the head
71 start achieved by the first variant to transmit puts later arriving individuals at a disadvantage,
72 making it less likely that they will contribute meaningfully to infection in the recipient host [3, 4].
73 This bottleneck is wider, though, in conditions of increased host susceptibility or closer contact
74 where an exposure comprises more infectious particles, suggesting that transmission is driven
75 largely by the successes and failure of individual pathogens [7, 8].

76 After establishing infection, a pathogen population may prevent later arriving pathogens
77 from infecting the same host, in a process termed competitive exclusion or competitive exclusion.

78 This phenomenon has been described for multiple taxonomically diverse pathogens, such as
79 influenza A viruses (IAV) and *Streptococcus pneumoniae* [9, 10]. In both cases, co-infection by
80 different genotypes of the same pathogen in animal models can occur only when both arrive at the
81 same time or within several hours of each other. After 18 hours, guinea pigs inoculated with one
82 IAV variant can no longer be superinfected by a second variant, which *in vitro* experiments
83 attribute to the prevention of cellular co-infection [9]. Furthermore, once *Streptococcus*
84 *pneumoniae* in the mouse nasopharynx reach a quorum after approximately 6 hours of
85 colonization, they begin producing bacteriocins and cognate immunity proteins that later arriving
86 planktonic cells do not, leading to efficient killing of superinfecting bacteria, in a process referred
87 to as fratricide [10]. This turns transmission into a "founder takes all" contest, in which the first
88 variants to arrive become the dominant genotype in the recipient host, and prevents the founder
89 variant(s) from being displaced by a variant that is less transmissible but more fit during within-
90 host competition. This strengthens selection for transmissibility, but also prevents immigration of
91 variants with greater within-host fitness, thereby making within-host adaptation reliant on
92 mutation.

93 Once infection is established, microorganisms typically grow to large population sizes
94 within the host. In contrast to the genetic drift that may occur during transmission, selection
95 operates more efficiently in these large within-host populations, with within-host adaptation being
96 observed in many human infections with pathogens such as *Staphylococcus aureus* and *S.*
97 *pneumoniae* [11, 12]. Novel genotypes favored by selection may be generated by mutation during
98 within-host replication, or when the presence of multiple viral or bacterial genotypes within the
99 same host causes genes from one genotype to enter the background of the other. Recombination
100 or reassortment in viruses carries the additional requirement of cellular coinfection, as recombinant

genomes are generated by template switching during replication in the cell and reassortant progeny are generated by the packaging of genome segments from different sources. Gene flow between coinfecting bacterial populations, however, may be less restricted, as it can occur through multiple mechanisms such as uptake of free DNA of dead bacteria (transformation), direct exchange between cells through a pilus (conjuation), or hitchhiking of bacterial DNA in bacteriophages (transduction). These processes have been shown to accelerate adaptation in multiple pathogen systems, including Φ6 bacteriophage [13, 14], influenza A virus [15, 16], and *S. pneumoniae* [17, 18]. While competitive exclusion generally prevents superinfecting variants from replicating and contributing to infection directly, the incorporation of alleles from these newcomers by an established resident population may serve as a second method of within-host adaptation.

Here we use multiple modeling approaches to predict how interactions between founder effects at the transmission bottleneck, competitive exclusion, and gene flow affect the efficiency of selection for novel alleles that affect both within-host fitness and transmissibility. We find that competitive exclusion can strengthen selection for transmissibility by imposing tight transmission bottlenecks and preventing displacement by superinfecting variants with greater within-host fitness, and that gene flow enables resident populations to selectively incorporate alleles that improve within-host fitness, thereby balancing between- and within-host fitness.

Results

We used three models to consider how selection within hosts, during transmission between two hosts, and on the scale of the host population affect the balance of transmissibility and within-host fitness. In the first set of models shown in **Figs 1 and 2**, we use probabilistic models of population genetics to consider how the probability an allele is transmitted or reaches fixation in one host is shaped by its initial frequency, relative transmissibility or within-host fitness, and the

124 size of the transmission bottleneck. In the second model shown in **Fig 3**, we use a branching
125 process model to consider the fate of a novel allele that arises during infection, and how its relative
126 transmissibility and within-host fitness affect its chances of extinction or fixation on the host
127 population scale. In the third set of models shown in **Figs 4 and 5**, we use a compartmental
128 susceptible-infected-susceptible (SIS) model in which multiple variants compete for hosts to
129 consider the balance between transmissibility and within-host fitness on the epidemiological scale.
130 A summary of the parameters, and the models in which they are relevant, is provided in **Table 1**.

131 **The efficiency of selection in transmission and the within-host environment.**

132 Novel alleles may arise during the course of infection in a host, but must be transmitted to
133 new hosts to avoid extinction. To evaluate the probability of a novel allele being transmitted, we
134 developed a probabilistic model in which the likelihood an allele is transmitted depends on its
135 frequency, p, in the donor host, and the extent to which it improves transmissibility, expressed as
136 a selection coefficient, s_t. In **Fig 1A**, we examine how both p and s_t affect the probability of that
137 an individual variant carrying an allele will colonize the recipient host. When s_t is low, an allele is
138 likely to be transmitted only if it is nearly ubiquitous in the donor population. Even very beneficial
139 alleles with high s_t, though, must still be present at moderate frequencies to have an appreciable
140 chance of transmission.

141 New individuals entering an already colonized host are unlikely to contribute meaningfully
142 to infection, but beneficial alleles from these individuals may be incorporated into a resident
143 population through recombination. To evaluate the conditions under which this process could
144 facilitate the fixation of novel alleles, we considered the probability that a novel allele benefitting
145 within-host fitness with selection coefficient s_r would fix after being introduced to a within-host
146 population a given number of times, whether by recombination or *de novo* mutation (**Fig 1B**). We

147 observe that even slightly beneficial alleles are much more likely to reach fixation upon
148 incorporation than neutral ones, which require ~100 independent introductions to have a 50%
149 chance of fixing. The required number of introductions required for an allele to have a 50% chance
150 of fixation is shown in **Fig 1C**, revealing that alleles with selection coefficients of at least 0.347
151 are likely to fix after just one introduction event. These probabilities depend on the effective size
152 of the colonizing population and are therefore sensitive to perturbations by environmental factors
153 such as host immunity or antimicrobial therapies.

154 **Selection at the transmission bottleneck in the face of a trade-off.**

155 Pathogen populations grow to large sizes within a host, but transmission is often
156 characterized by a tight bottleneck in which few individuals colonize a new host. To determine the
157 effect of transmission bottlenecks on the efficiency of selection during transmission, we considered
158 how the size of such a bottleneck affected the probability that an allele would be transmitted, as in
159 **Fig 1A**. In this model, a trade-off exists between transmissibility and within-host fitness, such that
160 increases in transmissibility come at the cost of within-host fitness. The relative fitness of a novel
161 variant is thus described by two distinct selection coefficients, s_t and s_r, denoting changes in
162 transmissibility and within-host fitness, respectively. The functional form of this trade-off parallels
163 that considered by Farahpour et al. [19] and depends on a shape parameter δ, which can vary from
164 0 (no trade-off) to 1 (absolute trade-off between transmissibility and within-host fitness) (**Fig 2A**).
165 As δ increases, increases to transmissibility carry more severe costs, which presents two challenges
166 to selection for transmissibility. First, mutant individuals that are co-transmitted with wild-type
167 ones must then compete for dominance in the new host to establish infection, and so the advantage
168 of a mutant in transmission may be outweighted by a disadvantage in competition during the early
169 stages of colonization. Second, as the mutant allele will be present at low frequency when first

170 generated, the competitive disadvantage faced by the mutant may prevent it from reaching an
171 appreciable frequency in the donor host. We explore this problem in **Fig 2B**, estimating the
172 threshold frequency that an allele with a given selection coefficient s_t must have in the donor host
173 in order to have a 50% probability of establishing colonization in the recipient host for differently
174 sized transmission bottlenecks. For wider bottlenecks, where more individuals compete during the
175 early stage of infection, the advantage of increased transmissibility is dwarfed by the disadvantage
176 of reduced within-host fitness, and transmission mutants require progressively higher frequencies
177 in the donor host to have the same 50% probability of establishing infection in the recipient host.
178 Particularly for high values of δ, the requisite frequency in the donor host approaches the within-
179 host frequency at which all individuals transmitted to the recipient host are transmission mutants,
180 meaning that the allele conferring increased transmissibility is fixed during transmission. Thus,
181 while wide transmission bottlenecks increase the probability that a mutant actually transmits to a
182 recipient host, transmission mutants with reduced within-host fitness benefit the most from tight
183 transmission bottlenecks which allow them to avoid within-host competition in the establishment
184 of new infections.

185 **Selection in the face of conflicts between scales.**

186 In order for a novel allele to emerge in the pathogen population, it must first be transmitted
187 from the host in which it was generated, and from there be propagated across the host population.
188 We model both of these processes to predict the fate of a novel allele based on its transmissibility
189 and within-host fitness relative to those of the wild-type ancestor variant. In this model, "within-
190 host fitness" refers to both competitive ability within a co-infected host ($s_r = 1 - w_r$ for $w_r < 1$, or
191 $s_r = 1 - \frac{1}{w_r}$ for $w_r > 1$), and the duration of infection when the host is infected with a mutant variant
192 ($\gamma_{mut} = \frac{\gamma_{wt}}{w_{r,mut}}$). "Transmissibility" refers to both the relative probability that a mutant individual is

193 transmitted from a co-infected host (Pr $(mut) = \frac{p_{mut}w_{t,mut}}{p_{mut}w_{t,mut} + p_{wt}}$) and the relative infectivity of hosts

194 infected with a mutant variant ($\beta_{mut} = w_{t,mut} * \beta_{wt}$). The relative R_0 of this mutant, compared to the

195 wild-type ancestor, is the product of its relative transmissibility and within-host fitness (**Fig 3A**).

196 The probability that an allele is successfully transmitted depends on both its frequency in

197 that host and its relative transmissibility, as shown in **Fig 1A**. Therefore, we explicity modeled the

198 within-host frequency of a novel allele as a Wright-Fisher process to calculate the average number

199 of new mutant infections caused by the initial host in which the mutation occurred (**Fig 3B**). These

200 simulations show a strong effect of within-host fitness, which helps variants reach higher

201 frequencies shortly after being generated and allows them to cause an appreciable number of

202 secondary infections even if they are less transmissible. Without such an advantage, alleles that

203 are deleterious or even neutral with respect to within-host fitness are less likely to escape the host

204 in which they are generated, as their initially low frequency is not increased by selection. This

205 strong influence of within-host fitness on the initial fate of a novel allele may thus constrain the

206 emergence of mutations that improve transmissibility and ultimately R_0, especially ones that carry

207 even a slight cost to within-host fitness.

208 Once a novel variant is successfully transmitted, it may still go extinct if each transmission

209 chain originiating from the initial host ends. To estimate the probability that all transmission from

210 the initial host terminates, indicating extinction, or at least one chain continues indefinitely,

211 indicating fixation of the novel allele in the pathogen population, we used a branching process

212 model (**Fig 3C**). This model calculates the probability that all of an individual's offspring

213 eventually go extinct based on the expected reproductive success of each individual, which

214 epidemiologically corresponds to the number of secondary cases caused by a single infected

215 individual, R_0 in a totally susceptible population or R_e, the effective reproductive number, in a

216 partially susceptible one. When competitive exclusion is perfectly efficient, partitioning wild-type
217 and mutant pathogens into distinct hosts, the R_e of a novel variant is simply the product of its
218 relative effects on transmissibility and infection duration, $R_e = \beta_{mut} * \gamma_{mut}$. This results in a
219 boundary in which alleles that increase transmissibility at the cost of within-host fitness or vice
220 versa may emerge only if they increase R_0. If superinfection is allowed to occur, however, than a
221 variant with greater within-host fitness may displace a less competitive resident population from
222 an already colonized host. This increases the relative importance of within-host fitness on the host
223 population scale, reducing the R_e of mutants that sacrifice within-host fitness for gains in
224 transmissibility, while increasing the R_e of mutants that have made the opposite trade by
225 prioritizing within-host fitness. By extension, this causes some mutants with decreased R_0 to have
226 $R_e > 1$ and therefore a chance of fixing, while others with increased R_0 may have $R_e < 1$, preventing
227 fixation. Competitive exclusion thus enhances selection for transmissibility on the host population
228 by 1) limiting the propagation of mutants with reduced R_0, and 2) preventing more transmissible
229 mutants with increased R_0 from being displaced by superinfection.

230 **Competition in the face of a trade-off between transmissibility and infection duration.**

231 Traits that promote pathogen transmission adversely can affect within-host survival and
232 vice versa. To determine how a trade-off between transmissibility and within-host fitness shapes
233 epidemiological fitness, we considered a SIS model in which each variant has a distinct
234 transmissibility, w_t, and within-host fitness, w_r, with the range of possible values being constrained
235 by the trade-off shown in **Fig 2A** (**Fig 4A**). Each of these parameters affects R_0 in a similar manner
236 to that in the preceding section, i.e., for a given variant i with $w_{t,i}$ and $w_{r,i}$, $\beta_i = \beta_0 * w_{t,i}$ and $\gamma i = \frac{\gamma 0}{w_{r,i}}$,
237 where β_0 and γ_0 are the maximum achievable transmissibility and minimum achievable rate of
238 clearance, respectively. The relative R_0 of a variant is then equal to the product $w_{t,i}w_{r,i}$, with the

239 maximum R_0 acheivable for a given trade-off (δ) occurring where $w_t = w_r = 1 - \delta$ (**Fig 4B**). As δ
240 increases and the trade-off becomes more concave, the value of this maximum achievable R_0
241 decreases (**Fig 4C**).

242 To determine how these variants that differ in both transmissibility and within-host fitness
243 compete on the epidemiological scale, we analyzed the dynamics of an epidemic in which 100
244 hosts are initially infected, each with a different variant having w_t from 0.01 to 1 and corresponding
245 w_r. We observe that more transmissible variants predominate during the early stages of an
246 epidemic, as they rapidly infect the available susceptible hosts, but are then displaced by the variant
247 with optimal R_0 (**Fig 4D, 4E**). The dominance of this optimal variant occurs rapidly when δ is
248 close to 0 or 1, but intermediate values, e.g. 0.3 – 0.6, allow prolonged coexistence of variants that
249 are more transmissible or persist for longer, but are ultimately suboptimal, resulting in more
250 diverse epidemics (**Fig 4F**). Similarly, the optimal variant takes longest to reach fixation at
251 intermediate values of δ (**Fig 4G**).

252 **Balancing between- and within-host fitness by competitive exclusion and recombination.**

253 The preceding results demonstrate that competitive exclusion allows a pathogen population
254 to balance transmissibility and within-host fitness in the face of a trade-off between the two, but
255 can also facilitate clonal interference in which beneficial mutations compete with each other. To
256 explore how pathogens might mitigate this detrimental effect of comepetitive exclusion, we
257 considered a SIS model in which variants differ in transmissibility and within-host fitness, but
258 where each aspect of fitness is governed by a separate locus and no trade-off exists between the
259 two traits. Additionally, mutations may occur in infected hosts, altering a variant's transmissibility
260 or within-host fitness. Finally, interaction between different variants can occur through three
261 different modes of competition. In the first, termed "displacement," a superinfecting variant with

262 greater within-host fitness may displace a less fit resident variant. In the second, termed
263 "exclusion," resident variants cannot be displaced by superinfecting variants. In the third, termed
264 "exclusion + recombination," residents cannot be displaced, but may incorporate alleles from
265 challengers through recombination. The biological mechanism of recombination would depend on
266 the agent in question, such as transformation in bacteria or reassortment in segmented viruses, but
267 it is presumed that alleles from challengers enter the background of the resident population.

268 To determine how the competitive modes described above affected a pathogen's
269 exploration of a fitness landscape, we analyzed changes in pathogen transmissibility and within-
270 host fitness over time on the epidemiological scale. The epidemic begins with some hosts infected
271 with variants that have uniformly low within-host fitness but vary in transmissibility, and some
272 hosts infected with poorly transmitting variants that differ in within-host fitness. As in **Fig 4**, we
273 observe that the most transmissible variants quickly predominate as they infect the available
274 susceptible hosts (**Fig 5A**). When displacement by superinfection is allowed, these variants are
275 readily displaced by those with greater within-host fitness, with average transmissibility
276 decreasing as average within-host fitness increases (**Fig 5B**). Emergence of the genotype with
277 maximum fitness in both dimensions is driven by mutations that improve the transmissibility of
278 the most competitive variants. Competitive exclusion allows these more transmissible variants to
279 avoid displacement by more competitive challengers and maintain their early advantage, but
280 emergence of the optimal genotype is conversely driven by mutations that improve within-host
281 fitness. This process is notably slower, as variants with greater within-host fitness can only be
282 propagated by causing new infections. The coupling of competitive exclusion and recombination,
283 however, allows the alleles conferring high within-host fitness to be incorporated onto the

284 background of the most transmissible variants, resulting in a rapid increase in the average R_0 of
285 circulating variants compared to the other two modes (**Fig 5C**).

286 ## Discussion

287 We report that cross-scale selective conflict shapes the balance between transmissibility and
288 within-host fitness in two ways. First, the high efficiency of selection for growth within the host
289 allows mutants with greater within-host fitness to predominate, even if they greatly compromise
290 transmissibility. Second, the dependence of a variant's transmission probability on its within-host
291 frequency hinders selection for mutants that improve transmissibility but not within-host fitness.
292 Left unchecked, this could allow selection to favor variants with greater within-host fitness but
293 lower R_0, but competitive exclusion prevents the propagation of these mutants on the
294 epidemiological scale. The same occupant-intruder strategy may delay within-host adaptation,
295 however, by subjecting populations to genetic drift in transmission and preventing immigration of
296 variants with greater within-host fitness. Pathogens exhibiting this occupant-intruder strategy can
297 thus adapt to the within-host environment faster by incorporating beneficial alleles from intruders.

298 A limitation of this study is that our model does not allow for variation in host susceptibility
299 to infection, which may significantly impact the early spread of a mutant. Recent theoretical work
300 by Morris et al., for example, found that infection history creates strong selection at the point of
301 transmission for antibody escape variants [20]. As escape from the first host in which a mutation
302 appears is an important first step in emergence, the probability that such an escape mutant emerges
303 would depend strongly on the local frequency of hosts with those antibodies.

304 Transmission bottlenecks have previously been studied for many infectious agents such as
305 HIV, IAV, *S. pneumoniae* [2-5, 21]. More general theoretical work predicted that transmission
306 mutants may be favored by tight bottlenecks in organisms, but that such bottlenecks would be rare

307 for directly transmitted pathogens like ones that replicate in the respiratory tract [22]. Our model
308 agrees that in the face of a growth-transmission trade-off, transmission mutants have the greatest
309 chance of establishing new infections under a tight transmission bottleneck that allows them to
310 avoid competition during colonization, but predicts that competitive exclusion may promote
311 selection for transmission even in respiratory pathogens like IAV and *S. pneumoniae*. More recent
312 theory has investigated the influence of transmission bottlenecks on the evolution of an emerging
313 pathogen, specifically with $R_0 < 1$, and the probability that a mutation prevents its extinction in the
314 human population [23, 24]. These models, like ours, find that tight transmission bottlenecks can
315 aid selection for transmissibility through a founder effect, and that wide bottlenecks can prevent
316 the emergence of mutations that improve transmissibility at the expense of within-host fitness,
317 even if they improve R_0. Our model diverges from theirs in that it considers a host-adapted
318 pathogen with $R_0 > 1$ at an endemic equilibrium with $R_e = 1$, and so mutants can have a reduced
319 R_0 that is still greater than one. This allowed us to show that cross-scale selective conflict, in
320 addition to hindering the emergence of more transmissible but less competitive variants as shown
321 by Schreiber et al., can also facilitate the emergence of mutants with greater within-host fitness
322 but ultimately lower R_0 [24].

323 In conjunction with previous empirical studies of transmission, our model underscores the
324 importance of transmission in purging variants that result from short-sighted within-host evolution
325 [25, 26]. Deleterious mutations such as defective viral genomes, for example, are consistently
326 generated *in vivo* and replicated with help from abundant cellular coinfection, but transmission is
327 often driven by individual virions, selecting against mutants that are not viable at low MOI
328 conditions [26-29]. Relaxing selection for transmission, however, selects for within-host fitness
329 without consideration for transmissibility. Serial passage experiments in mice, for example, have

330 resulted in the inactivation of genes that promote transmission at the expense of within-host fitness,
331 such as *dltB* in *S. pneumoniae* [30], or selected for accelerated growth in *Plasmodium chabaudi*
332 [31, 32]. Analogous *dlt* inactivation and evolution of more virulent *Plasmodium knowlesi* parasites
333 have also been observed during human infections, suggesting that within-host adaptation has
334 important consequences for disease severity and warranting further study of how transmission
335 affects the rate of such adaptation [11, 33]. Replication by blood-stage *Plasmodium* parasites also
336 involves significant epigenetic changes, which are reversed during passage through the mosquito,
337 further underscoring the importance of transmission in counteracting the potentially detrimental
338 effects of within-host selection [34, 35]. This reversal of short-sighted evolution by transmission
339 does require that transmissible varaints not be completely displaced during infection, suggesting
340 that maintaining diversity in a latent viral reservoir [36, 37] or by continuous genetic exchange
341 during colonization [38, 39] are important for ensuring future transmissibility.

342 The balance between transmissibility and within-host fitness parallels an invasion-persistence
343 trade-off in other organisms—some genotypes may more readily disperse and colonize new
344 hosts/territories, but are less persistent and thus fare poorly when present in the same environment
345 as less dispersable but more competitive genotypes. This trade-off is evident in several *S.*
346 *pneumoniae* genes, such as toxin pneumolysin [40, 41], capsule serotype (for example, serotype 4
347 compared to 23F) [40, 42], and the aforementioned *dltB* [30, 43], which promote transmissibility
348 at the expense of within-host fitness. Whether these genes are favored by selection and maintained
349 depends on the relative importance of transmissibility and within-host fitness, as shown by the loss
350 of *dltB* in serial passage where transmission is more certain. In *Pseudomonas aeruginosa*, selection
351 readily favors a hyperswarming phenotype that doubles the rate of dispersal at a modest 10%
352 reduction in growth rate [44]. By contrast, our results find that selection for comparable

353 transmission mutants with reduced within-host fitness is efficient only once the mutant causes at
354 least one secondary infection, and that the difficulty in escaping the initial host constrains the
355 emergence of such transmission mutants. On the epidemiological scale, the evolutionarily stable
356 genotype tends to be one with intermediate within-host fitness and transmissibility, but the initial
357 abundance of susceptible hosts at the beginning of an epidemic allows the most transmissible
358 variants to enjoy an early advantage. Analyses of spatially structured host populations reveals that
359 these more transmissible strains often dominate the expanding wavefront of an epidemic, similar
360 to the manner in which hyperswarming variants dominate the wavefront of a *Pseudomonas*
361 population [45]. While this early abundance may be transient, it plays an important role in the
362 acquisition of new alleles that benefit within-host fitness without affecting transmissibility—if
363 recombination is directional, with alleles from challengers being incorporated into the resident
364 population, and the resident population is more likely to be a more transmissible genotype, then
365 beneficial alleles from challengers are most likely to enter a more transmissible background.

366 Our results suggest an important role for directional gene flow in helping resident populations
367 mitigate the effect of genetic drift in transmission and the prevention of immigration imposed by
368 competitive exclusion. It is interesting that the same regulatory circuitry that governs competence
369 for transformation also controls competitive behaviors such as bacteriocin production in *S.*
370 *pneumoniae* [46] or expression a type VI secretion system in *Vibrio cholerae* [47]. It has been
371 suggested that this shared regulation of competitive behaviors and competence points to a role for
372 neighbor predation in the evolution of these bacteria [48], with recent studies providing empirical
373 support [46, 47]. One argument against the evolutionary benefit of transformation is that the free
374 DNA available for uptake comes from dead cells which, if killed by natural selection, are likely to
375 contribute mutations that reduce fitness [49]. This argument is supported by other studies showing

376 little evidence of recombination driving strongly selected traits such as antibiotic resistance [50],
377 or that DNA acquired through transformation is also useful as a nutrient [51]. However, our results
378 highlight the potential benefits of transformation when a population is not panmictic but structured
379 into distinct hosts—the same occupant-intruder strategy that enchances selection for
380 transmissibility also hinders selection for within-host fitness by imposing genetic drift during
381 transmission and preventing immigration, but recombination allows the incorporation of alleles
382 from challengers onto the background of established resident populations. In contrast to direct
383 competition between a resident and challenger, this directional recombination, in conjunction with
384 efficient selection due to high within-host population size, allows resident populations to
385 selectively retain alleles that improve within-host fitness. Occupant-intruder strategies and
386 recombination may thus play important roles in mitigating cross-scale selective conflicts for
387 infectious agents.

388 **Methods**

389 **Population genetics of selection at the transmission bottleneck and in within-host**
390 **populations.**

391 The probability that a given allele is transmitted depends on its frequency, p, in the population
392 attempting to colonize the susceptible host, and its selection coefficient, s_t. The probability that a
393 single individual carrying an allele is transmitted is estimated by:

394 $$\Pr(Transmission) = \frac{p}{p + (1 - s_t)(1 - p)} = \frac{p}{p + 1 - s_t - p + s_t p} = \frac{p}{1 - s_t + s_t p} \quad (1)$$

395 Kimura's fundamental prediction of allele fixation in haploid populations:

396 $$\Pr(Fixation) = \frac{1 - e^{-2Ns_r p}}{1 - e^{-2Ns_r}} \quad (2)$$

397 was used to estimate the probability that an allele entering a population by mutation or
398 recombination would eventually reach fixation [52]. N was set at a previously estimated effective

population size of 133 individuals [53], and $\frac{\lambda}{N}$ was substituted for p to represent λ individuals acquiring the allele by recombination, yielding the probability of fixation after λ introduction events:

$$\Pr\,(Fixation|\lambda) = \frac{1 - e^{-2\lambda s_r}}{1 - e^{-2Ns_r}} \quad (3)$$

Solving for λ where Pr(fixation) = 0.5 yields the required number of times an allele must be introduced to a population to have a 50% chance of fixation:

$$\lambda_{50} = \frac{\ln 2 + \ln\,(1 + e^{-2Ns_r})}{2s_r} \quad (4)$$

This equation implies that if $\lambda_{50} \leq 1$, then an allele has a at least a 50% chance of fixation following a single introduction event. Substituting 1 for λ_{50} in (4) yields the inequality

$$e^{-2Ns_r} - \frac{e^{2s_r}}{2} \leq 1 \quad (5)$$

When solved for s_r, this inequality reveals the minimum selection coefficient required for fixation to be likely after a single introduction event.

Fitness trade-offs and population bottlenecks in transmission.

When infection is established by multiple individuals, the probability of that an allele is present in the initially colonizing population depends upon the number of individuals transmitted, B, and is governed by the binomial distribution:

$$\Pr\,(Transmission|N) = \sum_{k=1}^{B} \binom{B}{k} \left(\frac{p}{1 - s_t + s_t p}\right)^k \left(\frac{(1 - s_t)(1 - p)}{1 - s_t + s_t p}\right)^{B-k} \quad (6)$$

As these initially arriving individuals replicate and compete during the early stages of infection, the probability that an allele will fix in the colonizing population is governed by Kimura's fundamental prediction of allele fixation in haploid populations [52]:

$$\Pr\,(Fixation) = \frac{1 - e^{-2Bs_r p}}{1 - e^{-2Bs_r}} \quad (7)$$

Combining (6) and (7) yields the probability that an individual carrying an allele will establish colonization:

$$\Pr\,(Colonization) = \sum_{k=1}^{N}\left[\binom{B}{k}\left(\frac{p}{1-s_t+s_t p}\right)^{k}\left(\frac{(1-s_t)(1-p)}{1-s_t+s_t p}\right)^{N-k}\frac{1-e^{-2ks_r}}{1-e^{-2Bs_r}}\right] \qquad (8)$$

Solving for p where Pr(Colonization) = 0.5 yields the required frequency in the donor population for an allele to have a 50% frequency of being transmitted to a susceptible host.

When a trade-off exists between transmissibility and within-host fitness, an allele improving transmissibility with selection coefficient s_t decreases a pathogen's within-host fitness. The extent of this decrease is governed by the equation:

$$\Delta w_r = \left(1 - s_t^{\frac{1}{-\log_2 \delta}}\right)^{-\log_2 \delta} \qquad (9)$$

resulting in the negative selection coefficient s_r:

$$s_r = 1 - \frac{1}{1-\Delta w_r} \qquad (10)$$

We then refine (8) using these context-specific selection coefficients to estimate the required frequency that an allele must have in the donor host to have a 50% probability of establishing infection in a new host:

$$\Pr\,(Colonization) = \sum_{k=1}^{B}\left[\binom{B}{k}\left(\frac{p}{1-s_t+s_t p}\right)^{k}\left(\frac{(1-s_t)(1-p)}{1-s_t+s_t p}\right)^{B-k}\frac{1-e^{-2ks_r}}{1-e^{-2Bs_r}}\right] \qquad (11)$$

Branching process model of novel allele emergence.

A branching process was used to predict the fate of a novel allele based on its effect on transmissibility and within-host fitness. For a given set of initial conditions, this model calculates the probability that a chain of transmission goes extinct, indicating loss of the novel allele, or continues indefinitely, indicating fixation. In this model, changes in transmissibility affect the infectivity of an infected host (Relative transmissibility = $\frac{\beta_{MUT}}{\beta_{WT}}$), while changes in within-host

441 fitness affect the duration of infection (Relative within-host fitness = $\frac{\gamma_{WT}}{\gamma_{MUT}}$), and so the R_0 of a
442 mutant relative to the WT ancestor is the product of its relative transmissibility and within-host
443 fitness:

444 $$\frac{R_{0,MUT}}{R_{0,WT}} = \frac{\beta_{MUT} * \gamma_{WT}}{\beta_{WT} * \gamma_{MUT}} \qquad (12)$$

445 At the within-host scale, changes in transmissibility also affect an allele's relative probability of
446 being sampled during transmission, as in (1), and changes in within-host fitness affect the within-
447 host frequency of the allele over time. To predict the number of secondary infections caused by a
448 mutant after mutation occurs in the initial host, we simulated the within-host dynamics of infection
449 using a Wright-Fisher process. The mutation arose at a specified time during infection, which
450 varied across a uniform distribution from 12 hours to 29.5 days for an infection lasting 30 days,
451 then fluctuated based on its selection coefficient:

452 $$s_r = 1 - \frac{\gamma_{MUT}}{\gamma_{WT}} \text{ for } \gamma_{MUT} < \gamma_{WT}; s_r = 1 - \frac{\gamma_{WT}}{\gamma_{MUT}} \text{ for } \gamma_{MUT} > \gamma_{WT} \quad (13)$$

453 and a generation time of 4 hours [53], for the duration of infection, 30 days. 10 independent
454 simulations per time-of-mutation were conducted. The relative probability of mutant
455 transmission over the course of infection was calculated according to (1), then averaged across
456 all tested mutation times to estimate the number of secondary cases caused by a given mutant
457 after its generation in single host.

458 This mutation occurs when the number of infected hosts is at an endemic equilibrium and
459 $R_{e,WT} = 1$ due to a reduction in susceptible hosts, so $S_e = \frac{1}{R_0}$, $I_e = 1 - \frac{1}{R_0}$, and $R_{e,MUT} = \frac{\beta_{MUT} * \gamma_{WT}}{\beta_{WT} * \gamma_{MUT}}$. This
460 proportional reduction in susceptible hosts and $R_{e,MUT}$ assumes that superinfection exclusion is
461 perfectly efficient. To determine how the ability of variants with greater within-host fitness to
462 displace less fit ones through superinfection would affect selection, we modified $R_{e,MUT}$ as follows:

463 $$R_{e,MUT} = \frac{\beta_{MUT} * \gamma_{WT}}{\beta_{WT} * \gamma_{MUT}} + I_e \frac{\beta_{MUT}}{\beta_{WT}} \frac{1 - e^{-2s_r}}{1 - e^{-2Ns_r}};\ s_r = 1 - \frac{\gamma_{MUT}}{\gamma_{WT}} \quad (14)$$

464 if relative within-host fitness is improved ($\gamma_{MUT} < \gamma_{WT}$), or:

465 $$R_{e,MUT} = \frac{\beta_{MUT} * \gamma_{WT}}{\beta_{WT} * \gamma_{MUT}} - I_e \frac{1 - e^{-2s_r}}{1 - e^{-2Ns_r}};\ s_r = 1 - \frac{\gamma_{WT}}{\gamma_{MUT}} \quad (15)$$

466 if relative within-host fitness is reduced ($\gamma_{MUT} > \gamma_{WT}$). In both cases, $N = 133$ individuals as before.
467 Based on this reproductive number $R_{e,MUT}$ and the expected number of secondary infections caused
468 by the initial host, the extinction probability of a given chain of mutant transmission was estimated
469 using a Galton-Watson process [54]. After the probability a given transmission chain's extinction
470 was calculated, the probability of ultimate extinction, in which all transmission chains arising from
471 the initial host, was calculated as:

472 $$\Pr(\text{Total Extinction}) = q^x \quad (16),$$

473 where q is the probability of extinction of a transmission chain arising from a single host, and x is
474 the number of mutant infections caused by the initial host where the mutant arises. Finally, the
475 probability that at least one transmission chain continued indefinitely, indicating fixation of the
476 mutant in the population, was given by:

477 $$\Pr(\text{Fixation}) = 1 - \Pr(\text{Total Extinction}) = 1 - q^x \quad (17).$$

478 It is important to note that this model is sensitive to the R_0 of the wild-type ancestor, which here
479 was set at 1.5.

480 **SIS model with invasion-persistence trade-off.**

481 To investigate how trade-offs between transmissibility and within-host fitness affect the selection
482 for both traits on the scale of the host population, we define a Susceptible-Infected-Susceptible
483 (SIS) model using a system of ordinary differential equations (ODEs):

484 $$\frac{dS}{dt} = -\sum_{i=1}^{N} \beta_i S I_i + \sum_{i=1}^{N} \gamma_i I_i \quad (18)$$

$$\frac{dI_i}{dt} = \beta_i S I_i - \gamma_i I_i \qquad (19)$$

I_i indicates the number of individuals infected with pathogen variant i, which has a transmission fitness w_t and corresponding w_r as calculated by the following:

$$w_r = (1 - w_t^{\frac{1}{-\log_2(1-\delta)}})^{-\log_2(1-\delta)} \qquad (19)$$

A pathogen's w_t and w_r affect its transmissibility and the duration of infection, respectively:

$$\beta_i = w_t \beta_0 \qquad (20)$$

$$\gamma_i = \frac{\gamma_0}{w_r} \quad (21)$$

where β_0 is the maximum achievable transmission coefficient, and γ_0 is the minimum achievable rate of clearance. The transmission fitness of each type, w_t, ranges from 0.01 to 0.99 in increments of 0.01, for a total of 99 types. The system was initialized with 1 infected individual of each type, and 9,901 susceptible individuals, for a total population size of 10,000 hosts. The dynamics were then realized using the R package *deSolve*.

SIS model with within-host competition, competitive exclusion, and recombination.

To determine how superinfection exclusion and recombination affect selection for both transmissibility and within-host fitness, we define an SIS model in which the two traits are governed by independent loci, with I_{tr} denoting an individual infected with a pathogen of $w_t = t$ and $w_r = r$. The change in susceptible individuals is the same as (18):

$$\frac{dS}{dt} = -\sum_{r=1}^{N}\sum_{t=1}^{N}\beta_t S I_{tr} + \sum_{r=1}^{N}\sum_{t=1}^{N}\gamma_r I_{tr} \quad (22)$$

The change in each class of infected individuals follows one of three ecological regimes, referred to herein as "displacement," "exclusion," and "exclusion + recombination." Where resident pathogens may be displaced by superinfecting pathogens that are more fit ("Displacement"):

$$\frac{dI_{tr}}{dt} = \beta_t S I_{tr} - \gamma_r I_{tr} + \sum_{i=1}^{N}\sum_{j=1}^{r-1}\beta_t I_{tr} I_{ij}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} - \sum_{i=1}^{N}\sum_{j=r+1}^{N}\beta_i I_{ij} I_{tr}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} + \mu I_{t,r-1}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} - \mu I_{t,r}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} \quad (23)$$

In the case where superinfection exclusion occurs, and resident pathogens exclude superinfecting individuals:

$$\frac{dI_{t,r}}{dt} = \beta_t S I_{t,r} - \gamma_r I_{t,r} + \mu I_{t,r-1}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} - \mu I_{t,r}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} \quad (24)$$

Finally, in the case where superinfection exclusion occurs, but DNA from challengers is available for uptake by the resident population:

$$\frac{dI_{tr}}{dt} = \beta_t S I_{tr} - \gamma_r I_{tr} + \sum_{i=1}^{N}\sum_{j=1}^{r-1}\beta_i I_{ir} I_{tj}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} - \sum_{i=1}^{N}\sum_{j=r+1}^{N}\beta_i I_{ij} I_{tr}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} + \mu I_{t,r-1}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} - \mu I_{t,r}\frac{1-e^{-2Ns_r p}}{1-e^{-2Ns_r}} \quad (25)$$

Possible w_t and w_r values ranged from 0.1 to 1 in increments of 0.1, as well as 0.01. The system was initialized with 2,000 hosts: 1,895 susceptible hosts, 5 hosts infected with pathogens of each w_t type and $w_r = 0.1$, and 5 hosts infected with pathogens of each w_r type with $w_t = 0.1$, so that the alleles conferring efficient transmission and replication are present initially, but in different hosts. The dynamics were then realized using the R package *deSolve*.

Acknowledgements

The authors thank John Lees for helpful discussion and review of the manuscript.

References

1. Levin BR, Bull JJ. Short-sighted evolution and the virulence of pathogenic microorganisms. Trends Microbiol. 1994;2(3):76-81. Epub 1994/03/01. doi: 10.1016/0966-842x(94)90538-x. PubMed PMID: 8156275.
2. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, et al. HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science. 2014;345(6193):1254031. Epub 2014/07/12. doi: 10.1126/science.1254031. PubMed PMID: 25013080; PubMed Central PMCID: PMCPMC4289910.

3. Kono M, Zafar MA, Zuniga M, Roche AM, Hamaguchi S, Weiser JN. Single Cell Bottlenecks in the Pathogenesis of Streptococcus pneumoniae. PLoS Pathog. 2016;12(10):e1005887. Epub 2016/10/13. doi: 10.1371/journal.ppat.1005887. PubMed PMID: 27732665; PubMed Central PMCID: PMCPMC5061371.

4. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes constrain the within and between host evolution of influenza virus. Elife. 2018;7. Epub 2018/04/24. doi: 10.7554/eLife.35962. PubMed PMID: 29683424; PubMed Central PMCID: PMCPMC5933925.

5. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, et al. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. Cell Host Microbe. 2014;16(5):691-700. Epub 2014/12/03. doi: 10.1016/j.chom.2014.09.020. PubMed PMID: 25456074; PubMed Central PMCID: PMCPMC4272616.

6. Wilker PR, Dinis JM, Starrett G, Imai M, Hatta M, Nelson CW, et al. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. Nat Commun. 2013;4:2636. Epub 2013/10/24. doi: 10.1038/ncomms3636. PubMed PMID: 24149915; PubMed Central PMCID: PMCPMC3845350.

7. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS Pathog. 2009;5(1):e1000274. Epub 2009/01/24. doi: 10.1371/journal.ppat.1000274. PubMed PMID: 19165325; PubMed Central PMCID: PMCPMC2621345.

8. Frise R, Bradley K, van Doremalen N, Galiano M, Elderfield RA, Stilwell P, et al. Contact transmission of influenza virus between ferrets imposes a looser bottleneck than respiratory droplet transmission allowing propagation of antiviral resistance. Sci Rep. 2016;6:29793. Epub 2016/07/20. doi: 10.1038/srep29793. PubMed PMID: 27430528; PubMed Central PMCID: PMCPMC4949428.

9. Marshall N, Priyamvada L, Ende Z, Steel J, Lowen AC. Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. PLoS Pathog. 2013;9(6):e1003421. Epub 2013/06/21. doi: 10.1371/journal.ppat.1003421. PubMed PMID: 23785286; PubMed Central PMCID: PMCPMC3681746.

10. Shen P, Lees JA, Bee GCW, Brown SP, Weiser JN. Pneumococcal quorum sensing drives an asymmetric owner-intruder competitive strategy during carriage via the competence regulon. Nat Microbiol. 2019;4(1):198-208. Epub 2018/12/14. doi: 10.1038/s41564-018-0314-4. PubMed PMID: 30546100; PubMed Central PMCID: PMCPMC6342471.

11. Lees JA, Kremer PHC, Manso AS, Croucher NJ, Ferwerda B, Serón MV, et al. Large scale genomic analysis shows no evidence for pathogen adaptation between the blood and cerebrospinal fluid niches during bacterial meningitis. Microb Genom. 2017;3(1):e000103. Epub 2017/03/30. doi: 10.1099/mgen.0.000103. PubMed PMID: 28348877; PubMed Central PMCID: PMCPMC5361624.

12. Young BC, Wu CH, Gordon NC, Cole K, Price JR, Liu E, et al. Severe infections emerge from commensal bacteria by adaptive evolution. Elife. 2017;6. Epub 2017/12/20. doi: 10.7554/eLife.30637. PubMed PMID: 29256859; PubMed Central PMCID: PMCPMC5736351.

13. Chao L, Tran TT, Tran TT. The advantage of sex in the RNA virus phi6. Genetics. 1997;147(3):953-9. Epub 1998/02/07. PubMed PMID: 9383044; PubMed Central PMCID: PMCPMC1208270.

576 14. Poon A, Chao L. Drift increases the advantage of sex in RNA bacteriophage Phi6.
577 Genetics. 2004;166(1):19-24. Epub 2004/03/17. doi: 10.1534/genetics.166.1.19. PubMed PMID:
578 15020402; PubMed Central PMCID: PMCPMC1470714.
579 15. Arai Y, Ibrahim MS, Elgendy EM, Daidoji T, Ono T, Suzuki Y, et al. Genetic
580 Compatibility of Reassortants between Avian H5N1 and H9N2 Influenza Viruses with Higher
581 Pathogenicity in Mammals. J Virol. 2019;93(4). Epub 2018/11/23. doi: 10.1128/jvi.01969-18.
582 PubMed PMID: 30463961; PubMed Central PMCID: PMCPMC6363993.
583 16. Pu J, Sun H, Qu Y, Wang C, Gao W, Zhu J, et al. M Gene Reassortment in H9N2 Influenza
584 Virus Promotes Early Infection and Replication: Contribution to Rising Virus Prevalence in
585 Chickens in China. J Virol. 2017;91(8). Epub 2017/02/06. doi: 10.1128/jvi.02055-16. PubMed
586 PMID: 28148803; PubMed Central PMCID: PMCPMC5375663.
587 17. Engelmoer DJ, Donaldson I, Rozen DE. Conservative sex and the benefits of
588 transformation in Streptococcus pneumoniae. PLoS Pathog. 2013;9(11):e1003758. Epub
589 2013/11/19. doi: 10.1371/journal.ppat.1003758. PubMed PMID: 24244172; PubMed Central
590 PMCID: PMCPMC3828180.
591 18. Lattar SM, Wu X, Brophy J, Sakai F, Klugman KP, Vidal JE. A Mechanism of
592 Unidirectional Transformation, Leading to Antibiotic Resistance, Occurs within Nasopharyngeal
593 Pneumococcal Biofilm Consortia. mBio. 2018;9(3). Epub 2018/05/17. doi: 10.1128/mBio.00561-
594 18. PubMed PMID: 29764945; PubMed Central PMCID: PMCPMC5954218.
595 19. Farahpour F, Saeedghalati M, Brauer VS, Hoffmann D. Trade-off shapes diversity in eco-
596 evolutionary dynamics. Elife. 2018;7. Epub 2018/08/18. doi: 10.7554/eLife.36273. PubMed
597 PMID: 30117415; PubMed Central PMCID: PMCPMC6126925.
598 20. Morris DH, Petrova VN, Rossine FW, Parker E, Grenfell BT, Neher RA, et al. Asynchrony
599 between virus diversity and antibody selection limits influenza virus evolution. bioRxiv.
600 2020:2020.04.27.064915. doi: 10.1101/2020.04.27.064915.
601 21. Orton RJ, Wright CF, King DP, Haydon DT. Estimating viral bottleneck sizes for FMDV
602 transmission within and between hosts and implications for the rate of viral evolution. Interface
603 Focus. 2020;10(1):20190066. Epub 2020/01/04. doi: 10.1098/rsfs.2019.0066. PubMed PMID:
604 31897294; PubMed Central PMCID: PMCPMC6936012.
605 22. Handel A, Bennett MR. Surviving the bottleneck: transmission mutants and the evolution
606 of microbial populations. Genetics. 2008;180(4):2193-200. Epub 2008/10/16. doi:
607 10.1534/genetics.108.093013. PubMed PMID: 18854584; PubMed Central PMCID:
608 PMCPMC2600951.
609 23. Bushman M, Antia R. A general framework for modelling the impact of co-infections on
610 pathogen evolution. J R Soc Interface. 2019;16(155):20190165. Epub 2019/06/27. doi:
611 10.1098/rsif.2019.0165. PubMed PMID: 31238835; PubMed Central PMCID:
612 PMCPMC6597765.
613 24. Schreiber SJ, Ke R, Loverdo C, Park M, Ahsan P, Lloyd-Smith JO. Cross-scale dynamics
614 and the evolutionary emergence of infectious diseases. bioRxiv. 2020:066688. doi:
615 10.1101/066688.
616 25. Visser PB, Brown DJ, Brederode FT, Bol JF. Nematode transmission of tobacco rattle virus
617 serves as a bottleneck to clear the virus population from defective interfering RNAs. Virology.
618 1999;263(1):155-65. Epub 1999/11/02. doi: 10.1006/viro.1999.9901. PubMed PMID: 10544090.
619 26. Zwart MP, Erro E, van Oers MM, de Visser J, Vlak JM. Low multiplicity of infection in
620 vivo results in purifying selection against baculovirus deletion mutants. J Gen Virol. 2008;89(Pt
621 5):1220-4. Epub 2008/04/19. doi: 10.1099/vir.0.83645-0. PubMed PMID: 18420800.

622 27. Jacobs NT, Onuoha NO, Antia A, Steel J, Antia R, Lowen AC. Incomplete influenza A
623 virus genomes occur frequently but are readily complemented during localized viral spread. Nat
624 Commun. 2019;10(1):3526. Epub 2019/08/08. doi: 10.1038/s41467-019-11428-x. PubMed
625 PMID: 31387995; PubMed Central PMCID: PMCPMC6684657.
626 28. Miyashita S, Kishino H. Estimation of the size of genetic bottlenecks in cell-to-cell
627 movement of soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding
628 up selection of variations in trans-acting genes or elements. J Virol. 2010;84(4):1828-37. Epub
629 2009/12/04. doi: 10.1128/jvi.01890-09. PubMed PMID: 19955302; PubMed Central PMCID:
630 PMCPMC2812364.
631 29. Zwart MP, Elena SF. Matters of Size: Genetic Bottlenecks in Virus Infection and Their
632 Potential Impact on Evolution. Annu Rev Virol. 2015;2(1):161-79. Epub 2016/03/10. doi:
633 10.1146/annurev-virology-100114-055135. PubMed PMID: 26958911.
634 30. Cooper VS, Honsa E, Rowe H, Deitrick C, Iverson AR, Whittall JJ, et al. Experimental
635 Evolution In Vivo To Identify Selective Pressures during Pneumococcal Colonization. mSystems.
636 2020;5(3). Epub 2020/05/14. doi: 10.1128/mSystems.00352-20. PubMed PMID: 32398278.
637 31. Barclay VC, Kennedy DA, Weaver VC, Sim D, Lloyd-Smith JO, Read AF. The effect of
638 immunodeficiency on the evolution of virulence: an experimental test with the rodent malaria
639 Plasmodium chabaudi. Am Nat. 2014;184 Suppl 1:S47-57. Epub 2014/07/26. doi:
640 10.1086/676887. PubMed PMID: 25061677.
641 32. Barclay VC, Sim D, Chan BH, Nell LA, Rabaa MA, Bell AS, et al. The evolutionary
642 consequences of blood-stage vaccination on the rodent malaria Plasmodium chabaudi. PLoS Biol.
643 2012;10(7):e1001368. Epub 2012/08/08. doi: 10.1371/journal.pbio.1001368. PubMed PMID:
644 22870063; PubMed Central PMCID: PMCPMC3409122.
645 33. Ciucă M, Lupaşco G, Negulici E, Constantinesco P. [Research on the experimental
646 transmission of Plasmodium malariae to man]. Arch Roum Pathol Exp Microbiol. 1964;23(3):763-
647 76. Epub 1964/09/01. PubMed PMID: 5318792.
648 34. Spence PJ, Brugat T, Langhorne J. Mosquitoes Reset Malaria Parasites. PLoS Pathog.
649 2015;11(7):e1004987. Epub 2015/07/03. doi: 10.1371/journal.ppat.1004987. PubMed PMID:
650 26133171; PubMed Central PMCID: PMCPMC4489640.
651 35. Spence PJ, Jarra W, Levy P, Reid AJ, Chappell L, Brugat T, et al. Vector transmission
652 regulates immune control of Plasmodium virulence. Nature. 2013;498(7453):228-31. Epub
653 2013/05/31. doi: 10.1038/nature12231. PubMed PMID: 23719378; PubMed Central PMCID:
654 PMCPMC3784817.
655 36. Brooks K, Jones BR, Dilernia DA, Wilkins DJ, Claiborne DT, McInally S, et al. HIV-1
656 variants are archived throughout infection and persist in the reservoir. PLoS Pathog.
657 2020;16(6):e1008378. Epub 2020/06/04. doi: 10.1371/journal.ppat.1008378. PubMed PMID:
658 32492044.
659 37. Brumme ZL, Sudderuddin H, Ziemniak C, Luzuriaga K, Jones BR, Joy JB, et al. Genetic
660 complexity in the replication-competent latent HIV reservoir increases with untreated infection
661 duration in infected youth. Aids. 2019;33(2):211-8. Epub 2018/10/17. doi:
662 10.1097/qad.0000000000002045. PubMed PMID: 30325763; PubMed Central PMCID:
663 PMCPMC6298800.
664 38. Cowley LA, Petersen FC, Junges R, Jimson DJM, Morrison DA, Hanage WP. Evolution
665 via recombination: Cell-to-cell contact facilitates larger recombination events in Streptococcus
666 pneumoniae. PLoS Genet. 2018;14(6):e1007410. Epub 2018/06/14. doi:

667 10.1371/journal.pgen.1007410. PubMed PMID: 29897968; PubMed Central PMCID:
668 PMCPMC6016952.
669 39. Marks LR, Reddinger RM, Hakansson AP. High levels of genetic recombination during
670 nasopharyngeal carriage and biofilm formation in Streptococcus pneumoniae. mBio. 2012;3(5).
671 Epub 2012/09/28. doi: 10.1128/mBio.00200-12. PubMed PMID: 23015736; PubMed Central
672 PMCID: PMCPMC3448161.
673 40. Kuipers K, Lokken KL, Zangari T, Boyer MA, Shin S, Weiser JN. Age-related differences
674 in IL-1 signaling and capsule serotype affect persistence of Streptococcus pneumoniae
675 colonization. PLoS Pathog. 2018;14(10):e1007396. Epub 2018/11/01. doi:
676 10.1371/journal.ppat.1007396. PubMed PMID: 30379943; PubMed Central PMCID:
677 PMCPMC6231672.
678 41. Zafar MA, Wang Y, Hamaguchi S, Weiser JN. Host-to-Host Transmission of
679 Streptococcus pneumoniae Is Driven by Its Inflammatory Toxin, Pneumolysin. Cell Host Microbe.
680 2017;21(1):73-83. Epub 2017/01/13. doi: 10.1016/j.chom.2016.12.005. PubMed PMID:
681 28081446; PubMed Central PMCID: PMCPMC5267320.
682 42. Zafar MA, Hamaguchi S, Zangari T, Cammer M, Weiser JN. Capsule Type and Amount
683 Affect Shedding and Transmission of Streptococcus pneumoniae. mBio. 2017;8(4). Epub
684 2017/08/24. doi: 10.1128/mBio.00989-17. PubMed PMID: 28830943; PubMed Central PMCID:
685 PMCPMC5565965.
686 43. Zafar MA, Hammond AJ, Hamaguchi S, Wu W, Kono M, Zhao L, et al. Identification of
687 Pneumococcal Factors Affecting Pneumococcal Shedding Shows that the dlt Locus Promotes
688 Inflammation and Transmission. mBio. 2019;10(3). Epub 2019/06/20. doi: 10.1128/mBio.01032-
689 19. PubMed PMID: 31213554; PubMed Central PMCID: PMCPMC6581856.
690 44. Deforet M, Carmona-Fontaine C, Korolev KS, Xavier JB. Evolution at the Edge of
691 Expanding Populations. Am Nat. 2019;194(3):291-305. Epub 2019/09/26. doi: 10.1086/704594.
692 PubMed PMID: 31553215; PubMed Central PMCID: PMCPMC7206977.
693 45. Osnas EE, Hurtado PJ, Dobson AP. Evolution of pathogen virulence across space during
694 an epidemic. Am Nat. 2015;185(3):332-42. Epub 2015/02/13. doi: 10.1086/679734. PubMed
695 PMID: 25674688; PubMed Central PMCID: PMCPMC4432791.
696 46. Wholey W-Y, Kochan TJ, Storck DN, Dawid S. Coordinated Bacteriocin Expression and
697 Competence in Streptococcus pneumoniae Contributes to Genetic Adaptation through Neighbor
698 Predation. PLOS Pathogens. 2016;12(2):e1005413. doi: 10.1371/journal.ppat.1005413.
699 47. Matthey N, Stutzmann S, Stoudmann C, Guex N, Iseli C, Blokesch M. Neighbor predation
700 linked to natural competence fosters the transfer of large genomic regions in Vibrio cholerae. Elife.
701 2019;8. Epub 2019/09/04. doi: 10.7554/eLife.48212. PubMed PMID: 31478834; PubMed Central
702 PMCID: PMCPMC6783263.
703 48. Veening JW, Blokesch M. Interbacterial predation as a strategy for DNA acquisition in
704 naturally competent bacteria. Nat Rev Microbiol. 2017;15(10):621-9. Epub 2017/07/12. doi:
705 10.1038/nrmicro.2017.66. PubMed PMID: 28690319.
706 49. Redfield RJ. Evolution of bacterial transformation: is sex with dead cells ever better than
707 no sex at all? Genetics. 1988;119(1):213-21. Epub 1988/05/01. PubMed PMID: 3396864; PubMed
708 Central PMCID: PMCPMC1203342.
709 50. Lehtinen S, Chewapreecha C, Lees J, Hanage WP, Lipsitch M, Croucher NJ, et al.
710 Horizontal gene transfer rate is not the primary determinant of observed antibiotic resistance
711 frequencies in <em>Streptococcus pneumoniae</em&gt. Science Advances.
712 2020;6(21):eaaz6137. doi: 10.1126/sciadv.aaz6137.

713 51. Redfield RJ. Evolution of natural transformation: testing the DNA repair hypothesis in
714 Bacillus subtilis and Haemophilus influenzae. Genetics. 1993;133(4):755-61. Epub 1993/04/01.
715 PubMed PMID: 8462839; PubMed Central PMCID: PMCPMC1205397.
716 52. Kimura M. On the probability of fixation of mutant genes in a population. Genetics.
717 1962;47(6):713-9. Epub 1962/06/01. PubMed PMID: 14456043; PubMed Central PMCID:
718 PMCPMC1210364.
719 53. Li Y, Thompson CM, Trzciński K, Lipsitch M. Within-host selection is limited by an
720 effective population of Streptococcus pneumoniae during nasopharyngeal colonization. Infect
721 Immun. 2013;81(12):4534-43. Epub 2013/10/02. doi: 10.1128/iai.00527-13. PubMed PMID:
722 24082074; PubMed Central PMCID: PMCPMC3837969.
723 54. Watson HW, Galton F. On the Probability of the Extinction of Families. The Journal of the
724 Anthropological Institute of Great Britain and Ireland. 1875;4:138-44. doi: 10.2307/2841222.
725

726

Table 1 — Summary of parameters.

Parameter	Meaning	Value
s_t	Change in transmissibility, defines relative sampling probability at transmission.	Varied in **Figs 1 – 2**, calculated from w_t in **Fig 3** .
s_r	Change in within-host fitness, defines outcome of within-host competition.	Varied in **Figs 1 – 2**, calculated from w_r in **Figs 3 – 5**.
w_t	Transmission fitness, used for calculating s_t and scaling β_0.	Varied in **Figs 3 – 5**.
w_r	Within-host fitness, used for calculating s_r and scaling γ_0.	Varied in **Figs 3 – 5**.
B	Transmission bottleneck size, number of individuals sampled during transmission.	Varied in **Fig 2**.
N	Effective size of within-host population.	133 individuals [53] in **Figs 1 – 5**.
δ	Shape of trade-off between transmissibility and within-host fitness.	Varied from 0 (no trade-off) to 0.9 (strong trade-off) in **Figs 2 and 4**.
β_0	Baseline infectivity of colonized host.	0.05 host^{-1} day^{-1} in **Figs 3 – 5**, variant β_i scaled by w_t.
γ_0	Baseline rate of clearance.	0.033 day^{-1} in **Figs 3 – 5**, variant γ_I scaled by w_r.
μ	Rate of mutation.	10^{-9} hour^{-1} in **Fig 5**.
R_0	Average number of secondary infections caused by one infected host.	1.5 in **Fig 3**.
R_e	Effective number of secondary infections caused by one infected host.	1 (WT) or $w_t w_r$ (mutant) in **Fig 3**.

730 **Figure 1 — Minority alleles are likely lost during transmission even if beneficial, but alleles**
731 **that improve within-host fitness are readily selected during infection.** The efficiencies of
732 selection for a mutant allele at the transmission bottleneck and following its introduction into a
733 within-host population were considered.

734 A. The probability of transmission was calculated for an allele improving transmissibility with
735 selection coefficient s_t and frequency p in the donor host. White dashed line traces the
736 boundary where Pr(transmission) = 0.5.
737 B. The probability of fixation was calculated for a novel allele improving within-host fitness
738 with selection coefficient s_r that is introduced into a within-host population λ times. Gray
739 dashed line represents the probability of fixation of a neutral allele.
740 C. The number of independent introduction events required for an allele with varying s_r to
741 have a 50% chance of fixation was calculated. Dashed line denotes the minimum selection
742 coefficient for an allele to have a 50% chance of fixation after just one introduction.
743

Figure 2 — When a trade-off exists between transmissibility and within-host fitness, selection for more transmissible variants is most effective with a tight transmission bottleneck. The effect of transmission bottlenecks on the efficiency of selection for more transmissible variants was considered for various trade-offs between transmissibility and within-host fitness.

A. The shape parameter delta governs the functional form of the trade-off, resulting in a neutral ($\delta = 0$), convex ($0 < \delta < 0.5$), or concave ($\delta > 0.5$) relationship. At higher δ, increases in transmissibility carry greater costs to within-host fitness.
B. The threshold frequency that a mutant with increased transmissibility (expressed by selection coefficient s_t) must have in the donor host to have a 50% probability of establishing infection in a recipient host was calculated. For a given transmission bottleneck width (B), B individuals are sampled from the donor host based on their transmissibility, which then compete for dominance in the recipient host based on their within-host fitness. Gray dashed lines in each facet represent a wide transmission bottleneck of 100 individuals.

Figure 3 — Mutants with increased within-host fitness but reduced transmissibility are likely to be transmitted initially, but competitive exclusion prevents the fixation of mutations that ultimately decrease R_0. For a pathogen population at an endemic equilibrium ($R_e = 1$), the fate of a novel allele with a given effect on both transmissibility and within-host fitness was considered.

A. The R_0 of a mutant, relative to the wild-type ancestor, was calculated based on its relative transmissibility and within-host fitness. Gray solid line represents the boundary below which mutants have reduced R_0. Regions bordered by the gray solid line and a white dashed line represent mutations that increase R_0 by increasing within-host fitness at a cost to transmissibility or vice versa.
B. The number of secondary mutant infections caused by the host in which a mutation arises was calculated for the mutations shown in a).
C. The probability of fixation is shown for mutants that have an $R_e > 1$. As in A), the gray line denotes the boundary below which mutants have a lower R_0 than wild-type. Red line denotes the boundary below which mutants have $R_0 < 1$. In the facet labeled "Displacement," a variant with greater within-host fitness may superinfect and displace a less fit resident population, while in the facet labeled "Exclusion," competitive exclusion is perfectly efficient and later arriving challengers cannot contribute to infection.

Figure 4 — In the face of a transmission-duration trade-off, competitive exclusion allows the early dominance of the most transmissible variants and ultimate fixation of the variant with optimal balance between transmissibility and infection duration. An SIS model was used to analyze the effect of a trade-off between transmissibility and within-host fitness (as measured by infection duration) on competition between variants on the epidemiogoical scale.

A. The functional form of the trade-off, as in **2a**, is shown, denoting all possible trait pairs for a given shape parameter δ. Intersection with dashed line ($w_r = w_t = 1 - \delta$) denotes the position of the optimal trait pair for each value of δ.
B. The relationship between R_0, relative to the maximum achievable R_0 without a trade-off, and transmissibility is shown for a range of δ values.
C. The maximum achievable R_0, relative to the maximum possible without a trade-off, is shown for a range of δ values.
D. The prevalences of variants with differing transmissibility are shown during the beginning of an epidemic that begins with 100 hosts, each infected with a different variant. The color of each line represents the transmissibility (lighter color = higher w_t) of a given variant, and δ is shown above each facet. For simplicity, only variants with w_t in (0.01, 0.1, 0.2…0.9, 0.99) are shown.
E. As in **D.**, but all variants are shown for an extended duration.
F. The effective diversity, as measured by Hill's N_2, of the epidemic at day 100 is shown for a range of δ values.
G. The length of time taken for the variant with optimal R_0 to outcompete other variants and reach 99% frequency is shown for a range of δ values.

Figure 5 — An occupant-intruder strategy for excluding challengers selects for transmissibility, while recombination facilitates the incorporation of alleles that improve within-host fitness. An SIS model was used to examine the evolution of both transmissibility and within-host fitness when no trade-off exists between the two, during an epidemic that begins with some hosts carrying more transmissible variants with low within-host fitness, and some carrying poorly transmissible variants with high within-host fitness. Three different modes of competition govern interactions between two infected hosts: 1) "Displacement," in which a variant with greater within-host fitness may superinfect and displace a less fit variant through competition, 2) "Exclusion," in which superinfecting variants are efficiently killed and displacement is impossible, and 3) "Exclusion + Recombination," in which superinfecting variants are killed, but the resident population may incorporate within-host fitness alleles from challengers by recombination.

A. The prevalences of variants is shown over time for three different modes of competition. Lines are colored by the relative R_0 of a variant compared to the R_0 of the optimal genotype with maximum transmissibility and within-host fitness.
B. The average transmissibility (dashed lines) and within-host fitness (solid lines) of circulating variants are shown over time for each mode of competition.
C. The average R_0 of circulating variants, relative to the maximum achievable R_0, is shown over time for each mode of competition.

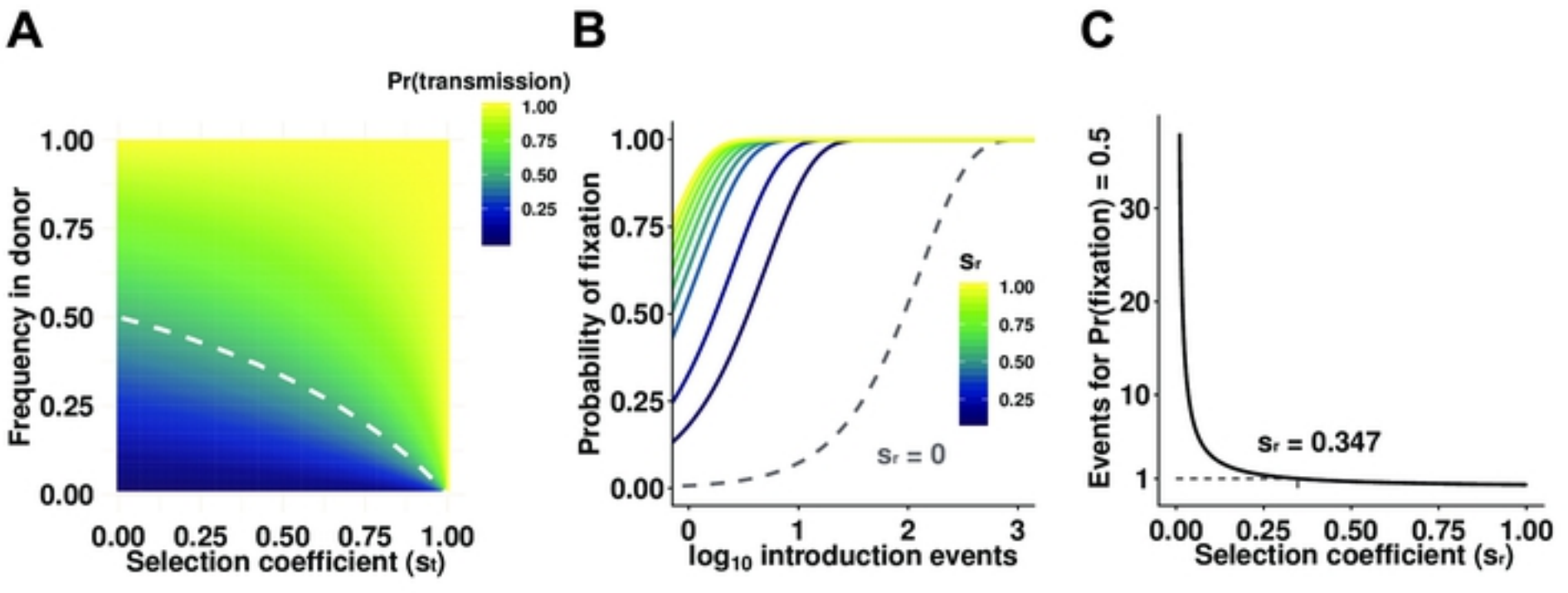
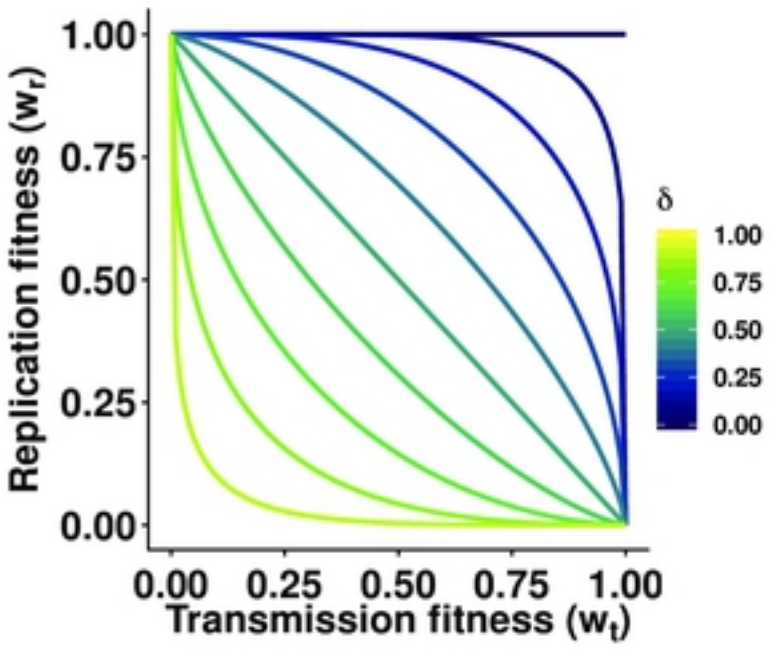

A
Pr(transmission)
1.00
0.75
0.50
0.25
Frequency in donor
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
Selection coefficient (st)
B
Probability of fixation
1.00
0.75
0.50
0.25
0.00
0 1 2 3
log10 introduction events
sr
1.00
0.75
0.50
0.25
sr = 0
C
Events for Pr(fixation) = 0.5
30
20
10
1
sr = 0.347
0.00 0.25 0.50 0.75 1.00
Selection coefficient (sr)

Fig 1


Replication fitness (w_r)
1.00
0.75
0.50
0.25
0.00
0.00 0.25 0.50 0.75 1.00
Transmission fitness (w_t)
δ
1.00
0.75
0.50
0.25
0.00

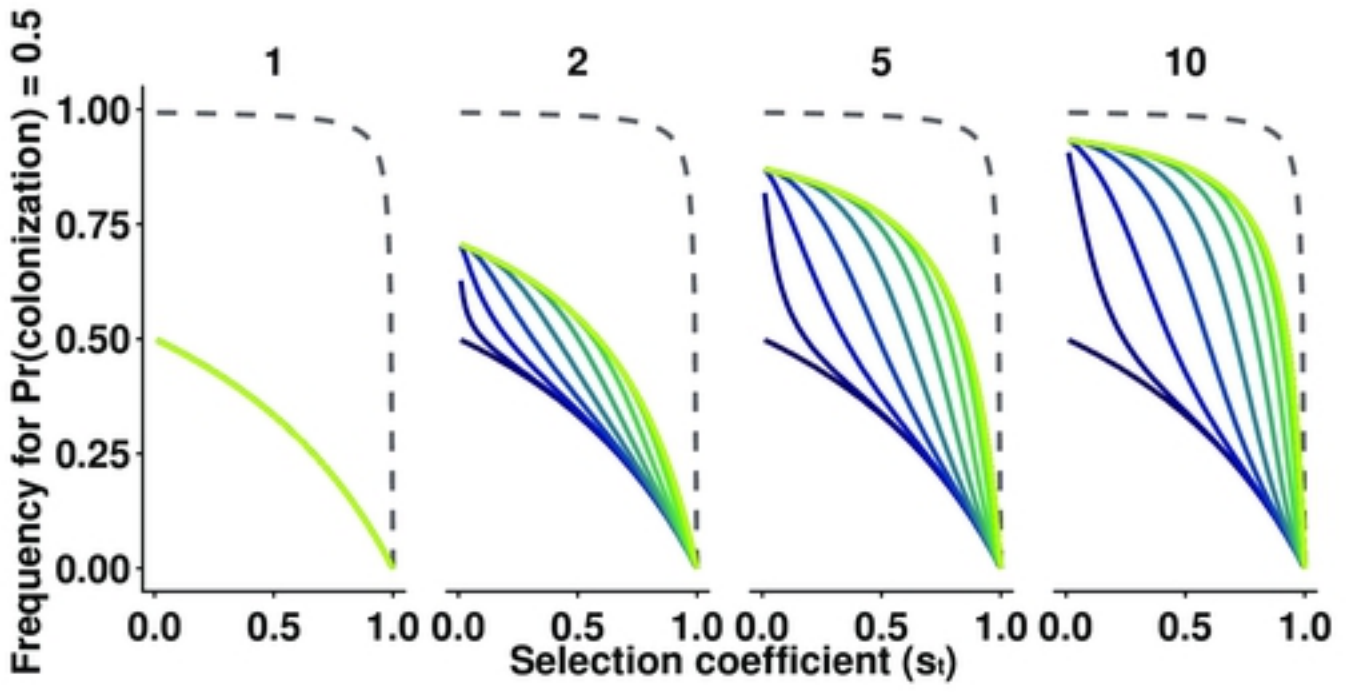

Frequency for Pr(colonization) = 0.5
1
2
5
10
1.00
0.75
0.50
0.25
0.00
0.0 0.5 1.0
Selection coefficient (s_t)

Fig 2

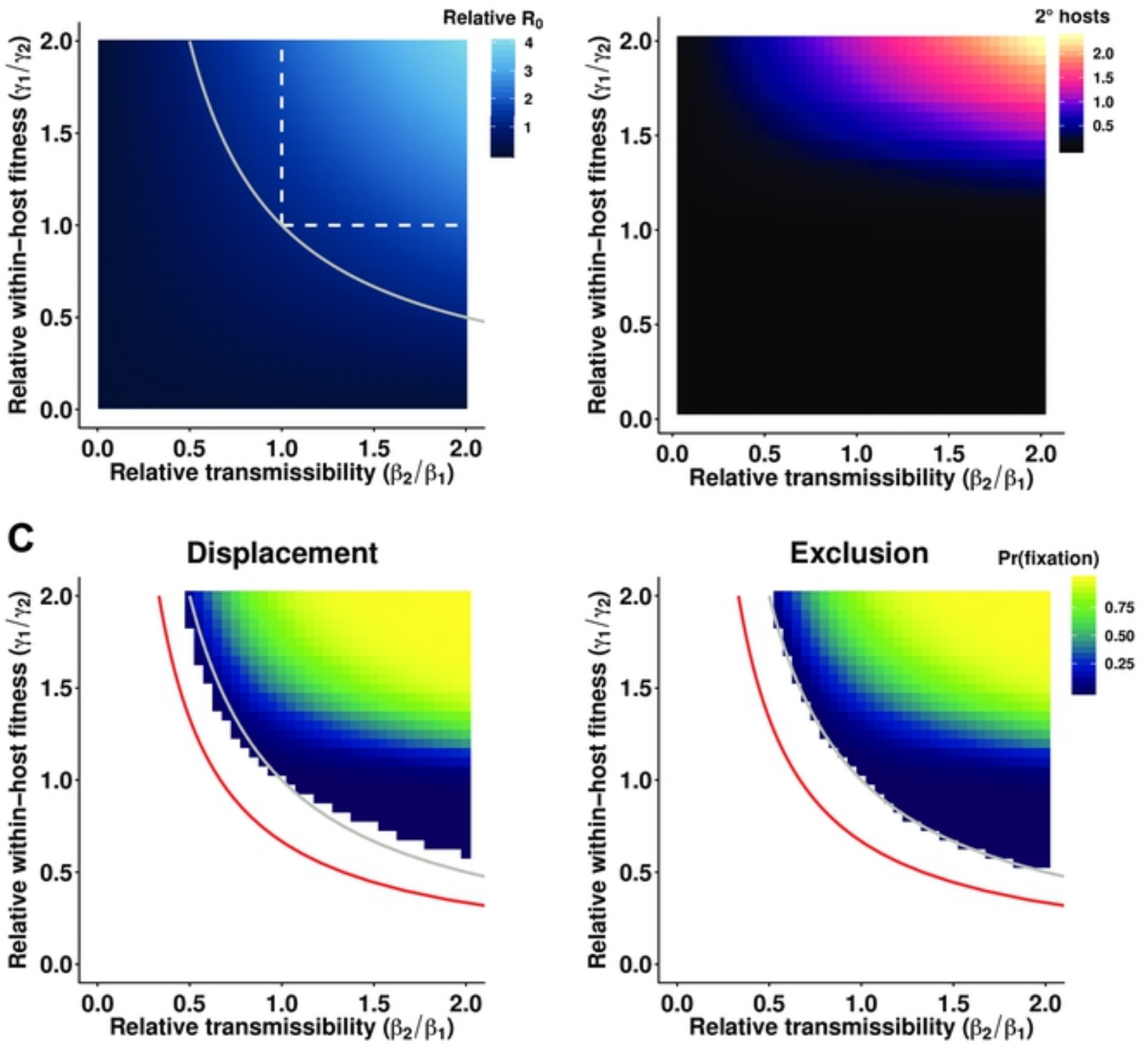

Relative R_0
4
3
2
1
2° hosts
2.0
1.5
1.0
0.5
C
Displacement
Exclusion
Pr(fixation)
0.75
0.50
0.25
Relative within–host fitness (γ_1/γ_2)
Relative transmissibility (β_2/β_1)
0.0
0.5
1.0
1.5
2.0

Fig 3

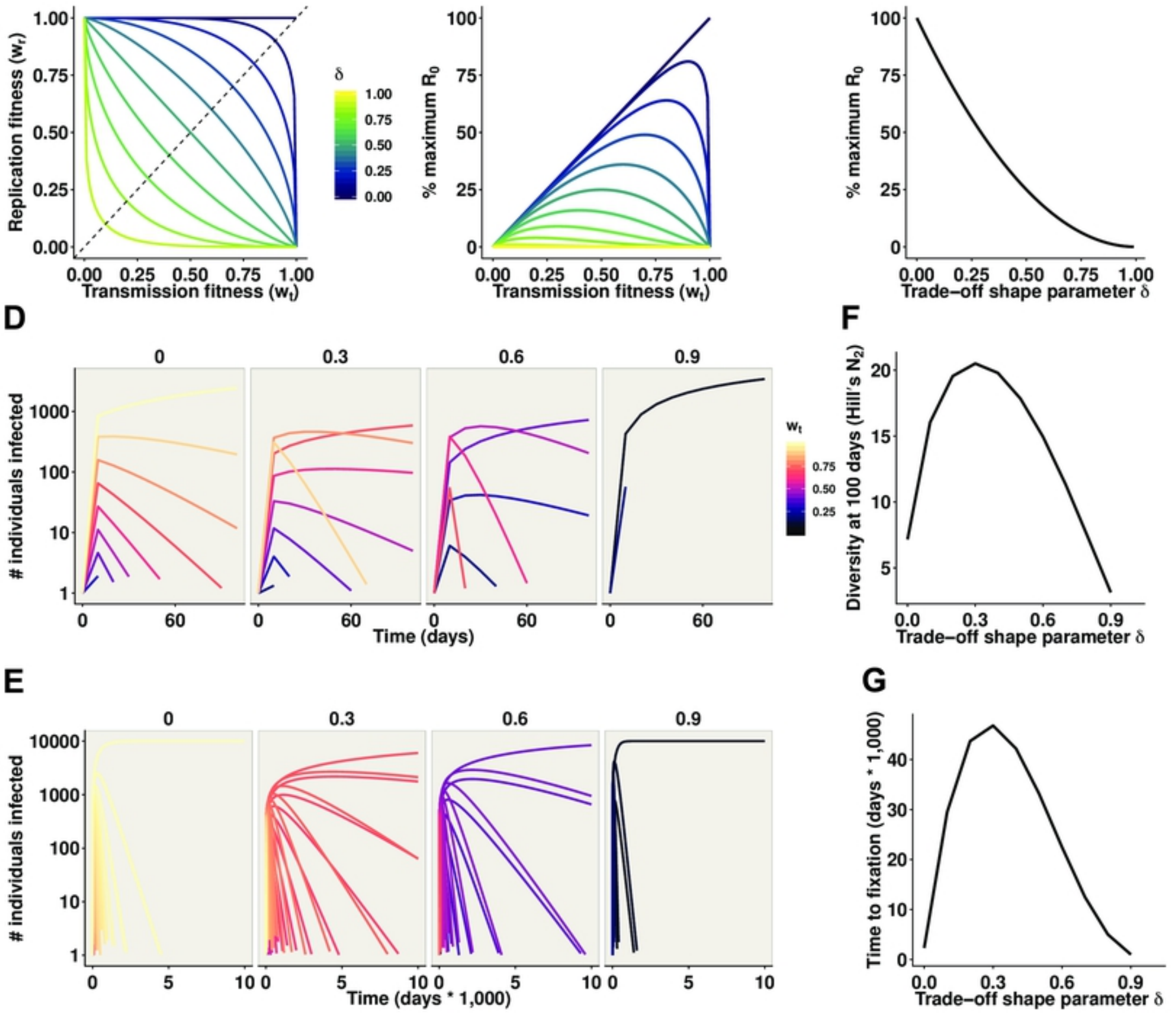

Replication fitness (w_r)
Transmission fitness (w_t)
δ
% maximum R_0
Trade–off shape parameter δ
D
E
F
G
individuals infected
Time (days)
Time (days * 1,000)
w_t
Diversity at 100 days (Hill's N_2)
Time to fixation (days * 1,000)

Fig 4

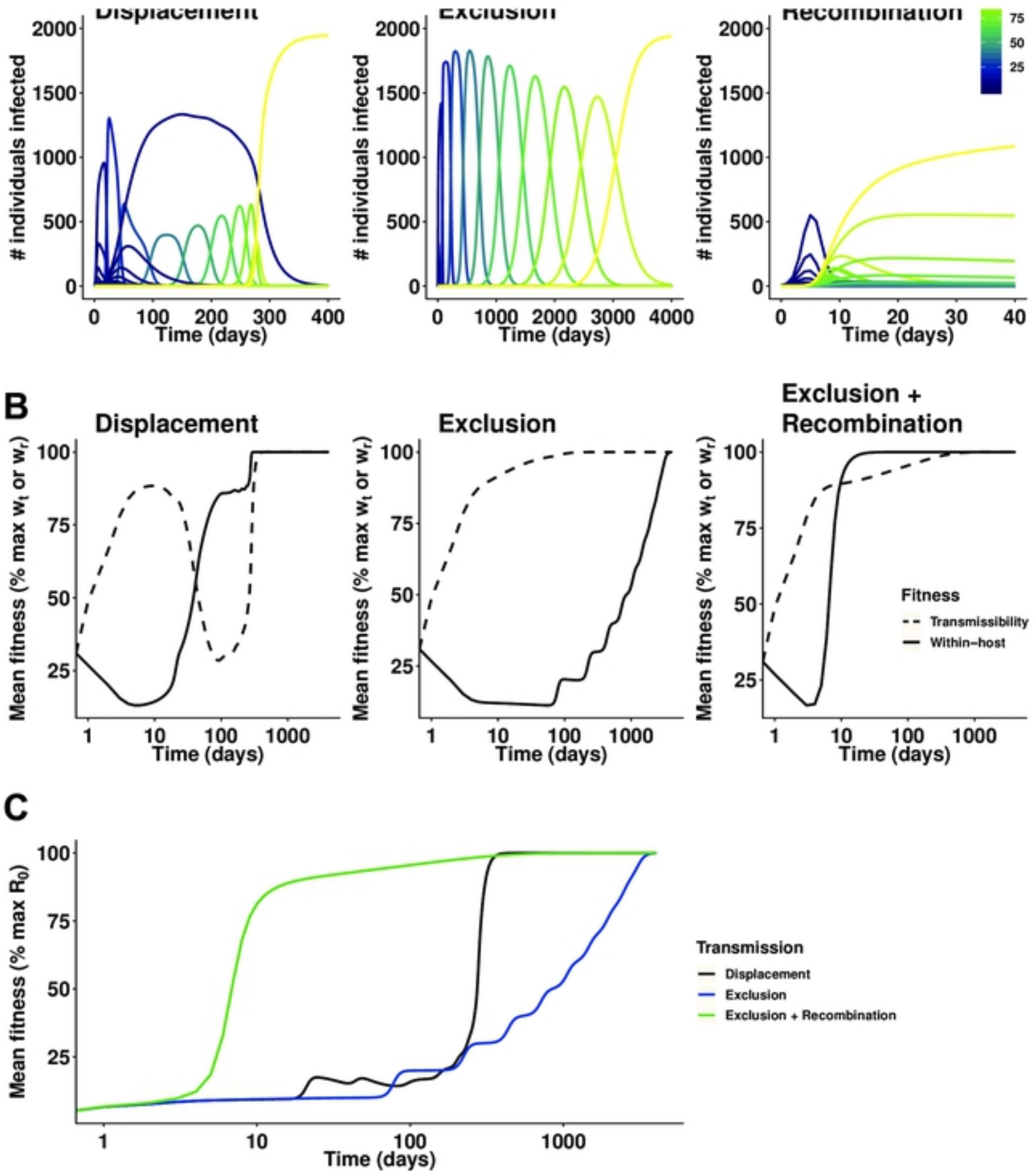

Displacement
Exclusion
Recombination
individuals infected
Time (days)
75
50
25
B
Displacement
Exclusion
Exclusion + Recombination
Mean fitness (% max w_t or w_r)
Fitness
Transmissibility
Within-host
C
Mean fitness (% max R_0)
Transmission
Displacement
Exclusion
Exclusion + Recombination

Fig 5