Competitive exclusion strengthens selection for transmissibility and increases the benefit of recombination for within-host adaptation.

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

Nathan T. Jacobs^a, Jeffrey N. Weiser^{1a}

^aDepartment of Microbiology, New York University Grossman School of Medicine, New York, NY, USA.

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

9 Abstract

10 Pathogens experience selection at multiple scales, given the need to transmit between hosts
 11 and replicate within them. This presents the challenge of cross-scale selective conflict when
 12 adaptations to one scale compromise fitness at another, such as mutations that improve
 13 transmissibility but make individuals less competitive within hosts. Selection operates differently
 14 at these scales, with tight transmission bottlenecks subjecting pathogen populations to genetic drift,
 15 and large population sizes within hosts enabling efficient selection for beneficial mutations.
 16 Compounding the reduction in diversity by transmission bottlenecks is the occupant-intruder
 17 competitive strategy exhibited by some pathogens, where the first variant to colonize a host
 18 prevents later arriving variants from contributing to infection, preventing immigration and turning
 19 transmission into a “founder takes all” contest. Here, we used multiple modeling approaches to
 20 examine how this behavior affects the efficiency of selection for both transmissibility and within-
 21 host fitness. We find that in the face of a trade-off, selection for transmissibility is maximized
 22 under a tight transmission bottleneck that minimizes within-host competition during colonization.
 23 While mutations with increased within-host fitness are favored during within-host replication, an
 24 occupant-intruder strategy prevents these mutants from displacing established residents and
 25 propagating across the host population, leading to their extinction if they are insufficiently
 26 transmissible. Finally, a model of competition on the scale of the host population revealed that
 27 competitive exclusion limits the propagation of mutations with improved within-host fitness,
 28 unless resident populations can incorporate alleles from intruding variants by recombination. Thus,
 29 competitive exclusion may facilitate the improvement and maintenance of pathogen
 30 transmissibility, with directional recombination allowing resident populations to mitigate the
 31 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

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

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

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

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

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

Once infection is established, microorganisms typically grow to large population sizes within the host. In contrast to the genetic drift that may occur during transmission, selection operates more efficiently in these large within-host populations, with within-host adaptation being observed in many human infections with pathogens such as *Staphylococcus aureus* and *S. pneumoniae* [11, 12]. Novel genotypes favored by selection may be generated by mutation during within-host replication, or when the presence of multiple viral or bacterial genotypes within the same host causes genes from one genotype to enter the background of the other. Recombination 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 (conjugation), or hitchhiking of bacterial DNA in bacteriophages (transduction). These processes have been shown to accelerate adaptation in multiple pathogen systems, including $\Phi 6$ 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

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

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

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

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

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

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

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

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

Selection in the face of conflicts between scales.

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

is 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 infected with a mutant variant ($\beta_{mut} = w_{t,mut} * \beta_{wt}$). The relative R_0 of this mutant, compared to the wild-type ancestor, is the product of its relative transmissibility and within-host fitness (**Fig 3A**).

The probability that an allele is successfully transmitted depends on both its frequency in that host and its relative transmissibility, as shown in **Fig 1A**. Therefore, we explicitly modeled the within-host frequency of a novel allele as a Wright-Fisher process to calculate the average number of new mutant infections caused by the initial host in which the mutation occurred (**Fig 3B**). These simulations show a strong effect of within-host fitness, which helps variants reach higher frequencies shortly after being generated and allows them to cause an appreciable number of secondary infections even if they are less transmissible. Without such an advantage, alleles that are deleterious or even neutral with respect to within-host fitness are less likely to escape the host in which they are generated, as their initially low frequency is not increased by selection. This strong influence of within-host fitness on the initial fate of a novel allele may thus constrain the emergence of mutations that improve transmissibility and ultimately R_0 , especially ones that carry even a slight cost to within-host fitness.

Once a novel variant is successfully transmitted, it may still go extinct if each transmission chain originating from the initial host ends. To estimate the probability that all transmission from the initial host terminates, indicating extinction, or at least one chain continues indefinitely, indicating fixation of the novel allele in the pathogen population, we used a branching process model (**Fig 3C**). This model calculates the probability that all of an individual's offspring eventually go extinct based on the expected reproductive success of each individual, which epidemiologically corresponds to the number of secondary cases caused by a single infected

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

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

Traits that promote pathogen transmission adversely can affect within-host survival and vice versa. To determine how a trade-off between transmissibility and within-host fitness shapes epidemiological fitness, we considered a SIS model in which each variant has a distinct transmissibility, w_t , and within-host fitness, w_r , with the range of possible values being constrained by the trade-off shown in **Fig 2A (Fig 4A)**. Each of these parameters affects R_0 in a similar manner 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 $\gamma_i = \frac{\gamma_0}{w_{r,i}}$, where β_0 and γ_0 are the maximum achievable transmissibility and minimum achievable rate

of clearance, respectively. The relative R_0 of a variant is then equal to the product $w_{t,i}w_{r,i}$, with the maximum R_0 achievable for a given trade-off (δ) occurring where $w_t = w_r = 1 - \delta$ (**Fig 4B**). As δ increases and the trade-off becomes more concave, the value of this maximum achievable R_0 decreases (**Fig 4C**).

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

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

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

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

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

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

Discussion

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

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

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

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

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

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

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

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

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

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

Methods

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

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

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

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

$$\Pr(\text{Fixation}) = \frac{1-e^{-2N_s r p}}{1-e^{-2N_s r}} \quad (2)$$

was used to estimate the probability that an allele entering a population by mutation or 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:

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

Solving for λ where $\Pr(\text{fixation}) = 0.5$ 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:

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

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]:

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

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

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

Solving for p where $\Pr(\text{Colonization}) = 0.5$ 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:

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

resulting in the negative selection coefficient s_r :

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

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:

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

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

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

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

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

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

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

This mutation occurs when the number of infected hosts is at an endemic equilibrium and $R_{e,WT} = 1$ 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 proportional reduction in susceptible hosts and $R_{e,MUT}$ assumes that superinfection exclusion is perfectly efficient. To determine how the ability of variants with greater within-host fitness to displace less fit ones through superinfection would affect selection, we modified $R_{e,MUT}$ as follows:

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

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

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

if relative within-host fitness is reduced ($\gamma_{MUT} > \gamma_{WT}$). In both cases, $N = 133$ individuals as before.

Based on this reproductive number $R_{e,MUT}$ and the expected number of secondary infections caused by the initial host, the extinction probability of a given chain of mutant transmission was estimated using a Galton-Watson process [54]. After the probability a given transmission chain's extinction was calculated, the probability of ultimate extinction, in which all transmission chains arising from the initial host, was calculated as:

$$\Pr(\text{Total Extinction}) = q_x \quad (16),$$

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

$$\Pr(\text{Fixation}) = 1 - \Pr(\text{Total Extinction}) = 1 - q_x \quad (17).$$

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

SIS model with invasion-persistence trade-off.

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

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

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

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:

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

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

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

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

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”):

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

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

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

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

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

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: PMC4289910.

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: PMC45061371.

4. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Luring 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: PMC5933925.

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: PMC4272616.

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: PMC3845350.

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: PMC2621345.

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: PMC4949428.

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: PMC3681746.

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: PMC6342471.

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: PMC5361624.

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: PMC5736351.

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

26. Zwart MP, Erro E, van Oers MM, de Visser J, Vlak JM. Low multiplicity of infection in vivo results in purifying selection against baculovirus deletion mutants. *J Gen Virol.* 2008;89(Pt 5):1220-4. Epub 2008/04/19. doi: 10.1099/vir.0.83645-0. PubMed PMID: 18420800.
27. Jacobs NT, Onuoha NO, Antia A, Steel J, Antia R, Lowen AC. Incomplete influenza A virus genomes occur frequently but are readily complemented during localized viral spread. *Nat Commun.* 2019;10(1):3526. Epub 2019/08/08. doi: 10.1038/s41467-019-11428-x. PubMed PMID: 31387995; PubMed Central PMCID: PMC6684657.
28. Miyashita S, Kishino H. Estimation of the size of genetic bottlenecks in cell-to-cell movement of soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *J Virol.* 2010;84(4):1828-37. Epub 2009/12/04. doi: 10.1128/jvi.01890-09. PubMed PMID: 19955302; PubMed Central PMCID: PMC2812364.
29. Zwart MP, Elena SF. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu Rev Virol.* 2015;2(1):161-79. Epub 2016/03/10. doi: 10.1146/annurev-virology-100114-055135. PubMed PMID: 26958911.
30. Cooper VS, Honsa E, Rowe H, Deitrick C, Iverson AR, Whittall JJ, et al. Experimental Evolution In Vivo To Identify Selective Pressures during Pneumococcal Colonization. *mSystems.* 2020;5(3). Epub 2020/05/14. doi: 10.1128/mSystems.00352-20. PubMed PMID: 32398278.
31. Barclay VC, Kennedy DA, Weaver VC, Sim D, Lloyd-Smith JO, Read AF. The effect of immunodeficiency on the evolution of virulence: an experimental test with the rodent malaria *Plasmodium chabaudi*. *Am Nat.* 2014;184 Suppl 1:S47-57. Epub 2014/07/26. doi: 10.1086/676887. PubMed PMID: 25061677.
32. Barclay VC, Sim D, Chan BH, Nell LA, Rabaa MA, Bell AS, et al. The evolutionary consequences of blood-stage vaccination on the rodent malaria *Plasmodium chabaudi*. *PLoS Biol.* 2012;10(7):e1001368. Epub 2012/08/08. doi: 10.1371/journal.pbio.1001368. PubMed PMID: 22870063; PubMed Central PMCID: PMC3409122.
33. Ciucă M, Lupașco G, Negulici E, Constantinesco P. [Research on the experimental transmission of *Plasmodium malariae* to man]. *Arch Roum Pathol Exp Microbiol.* 1964;23(3):763-76. Epub 1964/09/01. PubMed PMID: 5318792.
34. Spence PJ, Brugat T, Langhorne J. Mosquitoes Reset Malaria Parasites. *PLoS Pathog.* 2015;11(7):e1004987. Epub 2015/07/03. doi: 10.1371/journal.ppat.1004987. PubMed PMID: 26133171; PubMed Central PMCID: PMC4489640.
35. Spence PJ, Jarra W, Levy P, Reid AJ, Chappell L, Brugat T, et al. Vector transmission regulates immune control of *Plasmodium* virulence. *Nature.* 2013;498(7453):228-31. Epub 2013/05/31. doi: 10.1038/nature12231. PubMed PMID: 23719378; PubMed Central PMCID: PMC3784817.
36. Brooks K, Jones BR, Dileria DA, Wilkins DJ, Claiborne DT, McNally S, et al. HIV-1 variants are archived throughout infection and persist in the reservoir. *PLoS Pathog.* 2020;16(6):e1008378. Epub 2020/06/04. doi: 10.1371/journal.ppat.1008378. PubMed PMID: 32492044.
37. Brumme ZL, Sudderuddin H, Ziemniak C, Luzuriaga K, Jones BR, Joy JB, et al. Genetic complexity in the replication-competent latent HIV reservoir increases with untreated infection duration in infected youth. *Aids.* 2019;33(2):211-8. Epub 2018/10/17. doi: 10.1097/qad.0000000000002045. PubMed PMID: 30325763; PubMed Central PMCID: PMC6298800.

38. Cowley LA, Petersen FC, Junges R, Jimson DJM, Morrison DA, Hanage WP. Evolution via recombination: Cell-to-cell contact facilitates larger recombination events in *Streptococcus pneumoniae*. *PLoS Genet.* 2018;14(6):e1007410. Epub 2018/06/14. doi: 10.1371/journal.pgen.1007410. PubMed PMID: 29897968; PubMed Central PMCID: PMC6016952.
39. Marks LR, Reddinger RM, Hakansson AP. High levels of genetic recombination during nasopharyngeal carriage and biofilm formation in *Streptococcus pneumoniae*. *mBio.* 2012;3(5). Epub 2012/09/28. doi: 10.1128/mBio.00200-12. PubMed PMID: 23015736; PubMed Central PMCID: PMC3448161.
40. Kuipers K, Lokken KL, Zangari T, Boyer MA, Shin S, Weiser JN. Age-related differences in IL-1 signaling and capsule serotype affect persistence of *Streptococcus pneumoniae* colonization. *PLoS Pathog.* 2018;14(10):e1007396. Epub 2018/11/01. doi: 10.1371/journal.ppat.1007396. PubMed PMID: 30379943; PubMed Central PMCID: PMC6231672.
41. Zafar MA, Wang Y, Hamaguchi S, Weiser JN. Host-to-Host Transmission of *Streptococcus pneumoniae* Is Driven by Its Inflammatory Toxin, Pneumolysin. *Cell Host Microbe.* 2017;21(1):73-83. Epub 2017/01/13. doi: 10.1016/j.chom.2016.12.005. PubMed PMID: 28081446; PubMed Central PMCID: PMC5267320.
42. Zafar MA, Hamaguchi S, Zangari T, Cammer M, Weiser JN. Capsule Type and Amount Affect Shedding and Transmission of *Streptococcus pneumoniae*. *mBio.* 2017;8(4). Epub 2017/08/24. doi: 10.1128/mBio.00989-17. PubMed PMID: 28830943; PubMed Central PMCID: PMC5565965.
43. Zafar MA, Hammond AJ, Hamaguchi S, Wu W, Kono M, Zhao L, et al. Identification of Pneumococcal Factors Affecting Pneumococcal Shedding Shows that the *dlt* Locus Promotes Inflammation and Transmission. *mBio.* 2019;10(3). Epub 2019/06/20. doi: 10.1128/mBio.01032-19. PubMed PMID: 31213554; PubMed Central PMCID: PMC6581856.
44. Deforet M, Carmona-Fontaine C, Korolev KS, Xavier JB. Evolution at the Edge of Expanding Populations. *Am Nat.* 2019;194(3):291-305. Epub 2019/09/26. doi: 10.1086/704594. PubMed PMID: 31553215; PubMed Central PMCID: PMC67206977.
45. Osnas EE, Hurtado PJ, Dobson AP. Evolution of pathogen virulence across space during an epidemic. *Am Nat.* 2015;185(3):332-42. Epub 2015/02/13. doi: 10.1086/679734. PubMed PMID: 25674688; PubMed Central PMCID: PMC4432791.
46. Wholey W-Y, Kochan TJ, Storck DN, Dawid S. Coordinated Bacteriocin Expression and Competence in *Streptococcus pneumoniae* Contributes to Genetic Adaptation through Neighbor Predation. *PLOS Pathogens.* 2016;12(2):e1005413. doi: 10.1371/journal.ppat.1005413.
47. Matthey N, Stutzmann S, Stoudmann C, Guex N, Iseli C, Blokesch M. Neighbor predation linked to natural competence fosters the transfer of large genomic regions in *Vibrio cholerae*. *Elife.* 2019;8. Epub 2019/09/04. doi: 10.7554/eLife.48212. PubMed PMID: 31478834; PubMed Central PMCID: PMC6783263.
48. Veening JW, Blokesch M. Interbacterial predation as a strategy for DNA acquisition in naturally competent bacteria. *Nat Rev Microbiol.* 2017;15(10):621-9. Epub 2017/07/12. doi: 10.1038/nrmicro.2017.66. PubMed PMID: 28690319.
49. Redfield RJ. Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? *Genetics.* 1988;119(1):213-21. Epub 1988/05/01. PubMed PMID: 3396864; PubMed Central PMCID: PMC1203342.

50. Lehtinen S, Chewapreecha C, Lees J, Hanage WP, Lipsitch M, Croucher NJ, et al. Horizontal gene transfer rate is not the primary determinant of observed antibiotic resistance frequencies in *Streptococcus pneumoniae*. *Science Advances*. 2020;6(21):eaaz6137. doi: 10.1126/sciadv.aaz6137.
51. Redfield RJ. Evolution of natural transformation: testing the DNA repair hypothesis in *Bacillus subtilis* and *Haemophilus influenzae*. *Genetics*. 1993;133(4):755-61. Epub 1993/04/01. PubMed PMID: 8462839; PubMed Central PMCID: PMC1205397.
52. Kimura M. On the probability of fixation of mutant genes in a population. *Genetics*. 1962;47(6):713-9. Epub 1962/06/01. PubMed PMID: 14456043; PubMed Central PMCID: PMC1210364.
53. Li Y, Thompson CM, Trzeciński K, Lipsitch M. Within-host selection is limited by an effective population of *Streptococcus pneumoniae* during nasopharyngeal colonization. *Infect Immun*. 2013;81(12):4534-43. Epub 2013/10/02. doi: 10.1128/iai.00527-13. PubMed PMID: 24082074; PubMed Central PMCID: PMC3837969.
54. Watson HW, Galton F. On the Probability of the Extinction of Families. *The Journal of the Anthropological Institute of Great Britain and Ireland*. 1875;4:138-44. doi: 10.2307/2841222.

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 .

A

B

C

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

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

A **B**

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 δ 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.

A **B** **C** **D** **F** **E** **G**

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 epidemiological 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.

A

B

C

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.