Cross-scale dynamics and the evolutionary emergence of infectious diseases

Sebastian J. Schreiber[1,*], Ruian Ke[2], Claude Loverdo[3], Miran Park[4], Prianna Ahsan[4], James O. Lloyd-Smith[4,5,**]

1 Department of Evolution and Ecology, University of California, Davis, California, 95616, USA

2 Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA

3 Laboratoire Jean Perrin, UPMC, CNRS, 4 place Jussieu, Case Courrier 114, 75252 Paris Cedex 05, France

4 Department of Ecology & Evolution, University of California, Los Angeles, California, 90095, USA

5 Fogarty International Center, National Institutes of Health, Bethesda, Maryland, 20892, USA

Corresponding authors: *sschreiber@ucdavis.edu, **jlloydsmith@ucla.edu

Abstract

When emerging pathogens encounter new host species for which they are poorly adapted, they must evolve to escape extinction. Pathogens experience selection on traits at multiple scales, including replication rates within host individuals and transmissibility between individuals. We introduce and analyze a stochastic, multi-scale model linking pathogen growth and competition within individuals to transmission between individuals. Our analysis reveals a new factor that quantifies how quickly mutant strains increase in frequency when they initially appear in the infected host population. This cross-scale quantity combines with viral mutation rates, reproductive numbers, and transmission bottleneck width to determine the likelihood of evolutionary emergence, and whether evolution occurs swiftly or gradually within chains of transmission. Wider transmission bottlenecks facilitate emergence of pathogens with short-term infections, but hinder emergence of pathogens exhibiting cross-scale selective conflict and long-term infections. These results provide a framework for a new generation of evidence-based risk assessment of emergence threats.

Introduction

Emerging infectious diseases are rising in frequency and impact and are placing a growing burden on public health and world economies (1–4). Nearly all of these emergence events involve pathogens that are exposed to novel environments such as zoonotic pathogens entering human populations from non-human animal reservoirs, or human pathogens exposed to antimicrobial drugs (1). In these novel environments, pathogens may experience new selective forces occurring at multiple biological scales, leading to reduced replication rates within hosts or less efficient transmission between hosts. When these novel environments are sufficiently harsh, emergence only occurs when the pathogen adapts sufficiently quickly to avoid extinction. As genetic sequencing of pathogens becomes increasingly widespread, there are clear signs of such rapid adaptation (5–10), but we lack a cohesive framework to understand how this process might work. Theoretical studies over the past decade have shed important insights into circumstances under which this evolutionary emergence is possible, but have focused on the host-to-host transmission dynamics and treated within-host dynamics only implicitly (11–14). Here, we introduce and analyze a model explicitly linking these two biological scales and demonstrate how within-host viral competition, infection duration, between-host transmissibility, and the size of transmission bottlenecks determine the likelihood of evolutionary emergence. This analysis sheds new light on factors governing pathogen emergence, addresses long-standing questions about evolutionary aspects of emergence, and lays the foundation for making risk assessments which integrate outcomes from *in vitro* and *in vivo* experiments with findings from sequence-based surveillance in the field.

Recent empirical findings have highlighted the need for a new generation of theory on pathogen emergence, which addresses the current frontiers of dynamics within hosts and across scales. For most pathogens, and certainly for RNA viruses and single-stranded DNA viruses, individual hosts often are not dominated by single pathogen genotypes (15, 16). Furthermore, at the host population scale, pathogen allele frequencies at a given locus exhibit a range of dynamics from rapid selective sweeps for drug resistance or immune escape (17, 18) to gradually changing frequencies (19, 20). Together, these observations lead to the long-standing question of whether adaptive evolution of viruses occurs within single hosts by rapid fixation of beneficial mutants, or more slowly by a gradual shift of allele frequencies along chains of transmission (21, 22). A recent wave of studies tracking changes in within-host genetic diversity through chains of transmission among hosts (23–30) provide unique opportunities to address this question, but a theoretical framework is needed.

Empirical studies, together with analyses at broader population scales, have highlighted the crucial influence of the transmission process – and particularly the population bottleneck associated with transmission – in filtering viral diversity. The existence of transmission bottlenecks has long been recognized, and is hypothesized to play a critical role in pathogen evolution (31–33). Recent studies have shown that bottleneck widths vary considerably among pathogens and routes of transmission (34, 35), and perhaps across different phases of host adaptation (36). Narrow transmission bottleneck sizes of 1 to 2 virions are common for HIV-1 (37) and hepatitis C virus (38, 39). In contrast, deep sequencing data for patients infected with Ebola virus suggest that transmission bottleneck sizes are typically greater than 100 viral particles (40), and similarly wide transmission bottlenecks have been reported for natural transmission of influenza (25, 30, 41).

A major frontier in understanding viral adaptation is how the transmission process influences evolution at population scales. Past work has emphasized the potentially deleterious effect of genetic drift (31, 33), but a rising tide of studies reports direct selection for transmissibility. This can arise as a strong selection bias at the transmission bottleneck, where strains present at low or undetectable frequencies in the donor host are preferentially transmitted to the recipient (36, 42, 43), or it can be measured directly via experimental infection and transmission studies (20, 44–46)(though we emphasize that enhanced transmissibility is not inevitable, and depends on availability of suitable adaptive genotypes (47)). Overall transmission rates are thus determined by total viral loads, weighted by genotype-specific transmissibilities (43). Importantly, the transmissibility trait can vary independently from viral replication fitness within hosts, so there is potential for conflicts in selection across scales. Indeed, there is clear evidence that HIV-1 has certain genotypes that transmit more efficiently, but then the within-host population tends to evolve toward lower-transmission strains during an infection (43, 48, 49); a similar phenomenon has been reported for H5N1 influenza (42) and H9N2 influenza (50). In an extreme example, *Plasmodium* parasites were found to rapidly evolve resistance to an antimalarial drug, but at the cost of complete loss of transmissibility (51). Experimental evolution studies have highlighted how antagonistic pleiotropy can lead to tradeoffs between viral replication and the extracellular survival that is required for transmission (52). Together these findings contribute to a growing evidence base that cross-scale conflicts in selection may inhibit the emergence of new viral strains in many systems (reviewed in (14)).

Collectively these empirical findings highlight the need for a theory of evolutionary emergence that accounts explicitly for the within-host dynamics of competing viral strains, transmission bottlenecks, and host-to-host transmission dynamics (53). To this end, we introduce and analyze a model which integrates previous work onstochastic models of evolutionary emergence and deterministic models explicitly coupling within- and between-host dynamics (11, 13, 45, 54, 55). Our analysis allows us to address several fundamental questions about the emergence of novel pathogens: What factors limit evolutionary emergence for pathogens with different life histories? Why do some apparently ‘nearby’ adaptive mutants fail to emerge? How do bottleneck sizes influence the likelihood of emergence? Do evolutionary changes occur swiftly within individual hosts, or gradually across chains of transmission? Moreover, our analysis allows us to examine the relative importance of genetic diversity in zoonotic reservoirs versus the acquisition of new mutations following spillover into humans (56–58). Specifically, we address the long-standing question of how much is emergence risk increased if the “spillover inoculum” includes some genotypes bearing adaptive mutations for the novel host? Finally, our analysis enables us to unify findings from previous theoretical studies, and propose mechanistic interpretations of phenomenological parameters from earlier work.

Results

Our stochastic multi-scale model of evolutionary emergence follows a finite number of individuals in a large, susceptible host population exposed to a pathogen from a reservoir population (Fig 1A). Although our framework represents many types of pathogens and can be extended to any number of strains, we focus on the case of a viral pathogen with two strains: a wild-type maladapted for the novel environment and a mutant strain potentially adapted for the novel environment. Each infected individual begins with a viral load consisting of $v_w(0)$ wild-type virions and $v_m(0)$ mutant-type virions, where $v_w(0) + v_m(0)$ equals N, the size of the transmission bottleneck. The viral populations increase exponentially until saturating at day T_e of the infection with a maximal viral load K at which the viral replication rate is balanced by the viral clearance rate (Fig 1B). The wild and mutant strains increase exponentially at rates, r_w and r_m, and mutations arise with probability μ. The infectious period ends after T days. This within-host model can describe a range of viral dynamics from infections with primarily an exponential phase of viral growth to infections maintaining a stable viral load for an extended period.

At the scale of the host population, the transmission dynamics along chains of hosts is stochastic. Each infectious individual encounters a Poisson-distributed number of susceptible individuals at a rate of β individuals per day. Encounters result in a transmission event with probability $p(E)$ where $E = b_w v_w(t) + b_m v_m(t)$ is the effective viral load at the time t of transmission, b_w, b_m are the transmissibilities of the viral strains, and $p(E)$ is an increasing function of E. Our main analyses assume that the transmission function $p(E)$ is linear, but nonlinear transmission functions yield nearly identical results (Supplementary Figs S–4 through S–7). When transmission occurs, a newly infected individual is infected with N virions (the transmission bottleneck width) sampled binomially from the source individual's viral load weighted by the transmissibilities of the viral strains. After a transmission event, the newly infected individual's viral load is governed by the within-host model. By explicitly modeling the cross-scale dynamics, our model simultaneously tracks the number of infected hosts and the viral loads within each infected host (Fig 1D). The structure of our model is similar to a recent model of molecular viral evolution along transmission chains (55). However, our model accounts for transmission dynamics rather than conditioning on a chain of transmission, and explicitly accounts for the dynamics of competing viral strains.

The probability of evolutionary emergence.

We first focus on the scenario of a single individual in the host population getting infected by the wild-type strain. We assume that the mean number of individuals infected by this individual (the reproductive number R_w of the wild-type) is less than one. Hence, in the absence of mutations, there is no chance of a major outbreak (59). However,

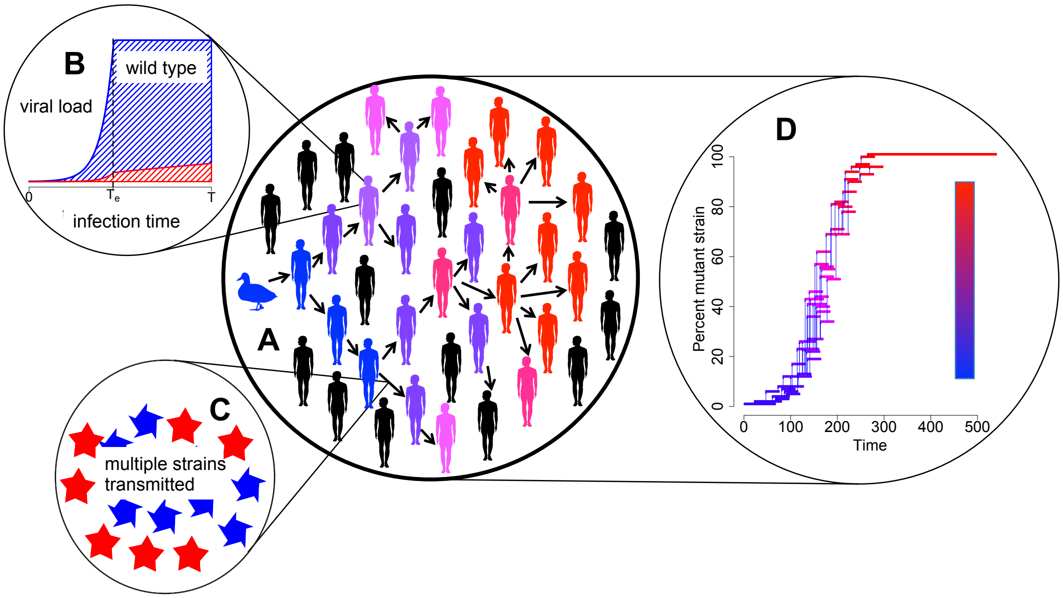


Figure 1. *The cross-scale dynamic of evolutionary emergence.* An individual is initially infected from the reservoir host population (panel A) with only the wild type viral strain (in blue). Within an infected individual (panel B), the viral load increases at an exponential rate until saturating at day T_e and ending after T days. Mutations ensure individuals typically have a mixed infection with wild-type (blue) and mutant (red) viral strains (panel B). The likelihood of transmission between individuals, and the composition of the infecting dose (panel C), depend on the size and composition of the infected individual's viral load at the time of contact, and on the transmissibility of each strain. As the infection spreads in the population (panel A), the frequency of the mutant virions among infected individuals varies (panel D) and, ultimately, determines whether evolutionary emergence occurs. In D, each horizontal line marks the infectious period of an individual whose infection was initiated with that percentage of the mutant strain and the vertical arrows represent transmission events between individuals.

if the wild-type strain produces a mutant strain whose reproductive number R_m is greater than one, there is a chance for a major outbreak. The mutant strain might have a higher reproductive number than the wild-type strain because it replicates more rapidly within the host or because it transmits more effectively to new hosts (or both). We define these within-host and between-host selective advantages as $s = r_m - r_w$ and $\tau = \log(b_m) - \log(b_w)$, respectively.

Consistent with theoretical expectations, a non-zero probability of evolutionary emergence requires the mutant's reproductive number R_m to be greater than one (Fig 2). However, the mixture of selective advantages or disadvantages of the mutant strain that give rise to $R_m > 1$ depends in a complex manner on the pathogen's life history traits, such as the duration of the infection (Fig 2A,B vs. C,D) and the transmission bottleneck width (Fig 2A,C vs. B,D). Notably, for long-term infections with a large transmission bottleneck size, the emergence probability can be effectively zero (i.e. $< 10^{-16}$) for mutant strains whose reproductive number exceeds one (white region bounded by the solid red curve in Fig 2D).

To understand these complexities, we determine the conditions under which the mutant's reproductive number R_m exceeds one, and then present analytic approximations for the emergence probability when $R_m > 1$.

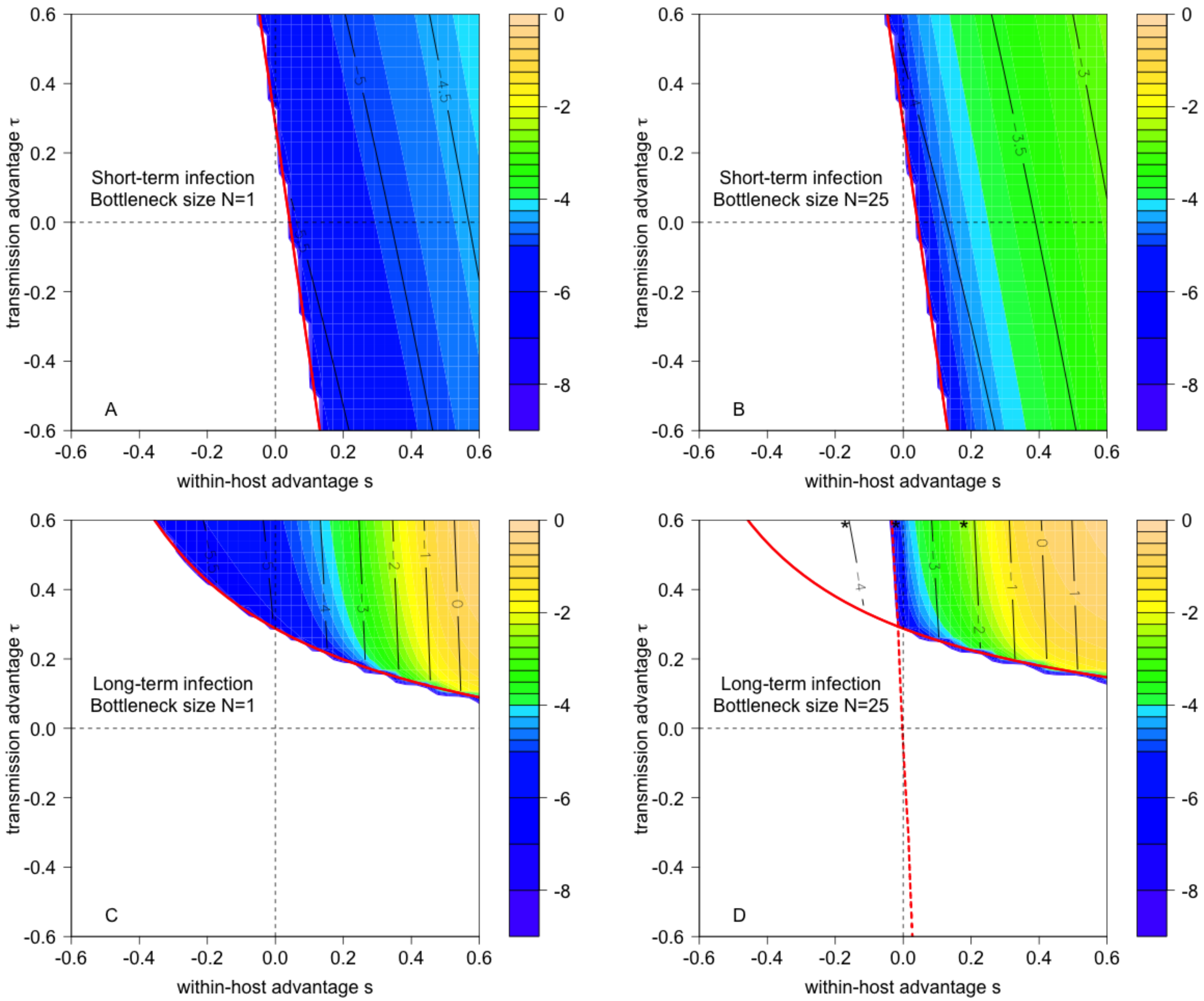


Figure 2. The joint effects of within-host and between-host selective advantage of the mutant on the probability of emergence for pathogens with short-term (A,B) and long-term (C,D) infectious periods, and with transmission bottlenecks of size $N = 1$ (A,C) and $N = 25$ (B,D). Colorations correspond to $\log_{10}$ of the emergence probability. The critical value of the mutant reproductive number R_m equaling one is drawn in solid red. Black contour lines correspond to $\log_{10}$ of the number of mutant virions transmitted by an individual initially only infected with the wild strain. In D, the critical value of the cross-scale mutant reproductive rate, $\alpha = 1$, is shown as a red dashed line and the black asterisks correspond to the τ and s values used in Fig 3. Parameters: $K = 10^7$, $\beta T = 30$, $T = 7.5$ (short-term infection) and $T = 30$ (long-term infection), b_w chosen so that $R_0 = 0.75$ for wild type, $r_w = 1.25$ and $\mu = 10^{-7}$.

Cross-scale selection and the mutant reproductive number R_m.

The reproductive numbers of the wild-type strain (R_w) and mutant strain (R_m) can be calculated as the product of the contact rate, the average transmissibility of the strain during the infectious period, and the infection duration (see

Supplementary Information). These reproductive numbers are positively correlated with the contact rate, infection duration, transmissibility, and viral per-capita growth rates. The influence of the maximal viral load K depends on the infection duration. For short-term infections, defined here as infections with a relatively short saturated phase (i.e. $T - T_e \ll T_e$), increasing K has little effect on a strain's reproductive number. For long-term infections, defined here as as infections with a long saturated phase (i.e. $T - T_e \gg T_e$), reproductive numbers increase with K.

Whether a selective advantage at either scale results in the mutant reproductive number R_m exceeding one depends on the duration of the infection (Supplementary Information). For short-term infections, the mutant reproductive number satisfies

$$\log R_m \approx \log R_w + \tau + s(T - 1/r_w). \tag{1}$$

This approximation shows that a sufficiently strong selective advantage at either scale can result in the mutant reproductive number exceeding one ($R_m > 1$) despite a selective disadvantage at the other scale (confirmed by exact calculations in Fig 2A,B). For short-term infections where viral dynamics are dominated by the exponential phase, the longer the duration of infection, the greater the influence of the within-host selective advantage compared to the between-host selective advantage (e.g., steep contours in Fig 2A).

For long-term infections, the mutant's reproductive number satisfies

$$\log R_m \approx \log R_w + \tau. \tag{2}$$

This approximation implies that a between-host selective advantage is required for evolutionary emergence (confirmed by exact calculations in Fig 2C). When viral dynamics are dominated by the saturated phase at fixed K, a within-host selective advantage has little impact on the average viral load during the infectious period of an individual solely infected with the mutant strain and, consequently, provides a minimal increase in the mutant reproductive number.

Going beyond the mutant reproductive number.

When the mutant strain has a reproductive number greater than one, there is a non-zero probability of a major outbreak that is well-approximated by the product of three terms (Supplementary Information):

$$\left\{\begin{array}{l}\text{mean size}\\ \text{of a minor}\\ \text{outbreak}\\ \text{due to the}\\ \text{wild type}\end{array}\right\} \times \left\{\begin{array}{l}\text{mean \# individu-}\\ \text{als infected with one}\\ \text{mutant virion by an}\\ \text{individual infected}\\ \text{initially with only}\\ \text{the wild type}\end{array}\right\} \times \left\{\begin{array}{l}\text{probability an}\\ \text{individual in-}\\ \text{fected with one}\\ \text{mutant virion}\\ \text{causes a major}\\ \text{outbreak}\end{array}\right\} \tag{3}$$

This expression, which can be viewed a multi-scale extension of earlier theory (11, 12), highlights three key ingredients, in addition to $R_m > 1$, for evolutionary emergence.

First, the size of the minor outbreak produced by the wild type determines the number of opportunities for the mutant strain to appear within a host. The average size of this minor outbreak equals $\frac{1}{1-R_w}$, as noted by earlier studies (11, 12). If the wild strain is badly maladapted (e.g. $R_w < 1/2$), then it is expected not to spread to multiple individuals (i.e. $\frac{1}{1-R_w} < 2$) and opportunities for transmission of mutant virions are very limited. Alternatively, if the wild strain is only slightly maladapted to the new host (e.g. $R_w = 0.95$), then, even without any mutations, the pathogen is expected to spread to many individuals (e.g. $\frac{1}{1-R_w} = 20$), thereby providing greater opportunities for evolutionary emergence. Our analysis implies that higher contact rates, within-host viral growth rates, viral transmissibility, and maximal viral loads (for long infectious periods) facilitate these larger reproductive values.

Second, the mutant strain must be transmitted successfully to susceptible individuals — the second term of our approximation (3). For an individual initially infected only with the wild-type strain, the mean number of transmission events with mutant virions equals the product of the contact rate, the infection duration, and the likelihood that a mutant virion is transmitted during a contact event, averaged over the full course of infection (Supplementary Information). The likelihood of transmitting mutant virions on the tth day of infection is proportional to the product of the transmission bottleneck width (N), the within-host frequency of the mutant strain, and the transmissibility b_m of the mutant strain. This highlights an important distinction between short-term and long-term infections. For short-term infections, there is insufficient time for the frequency of mutants to rise within a host, so transmission events with mutant virions are rare ($< 1/1,000$ for all black contour lines in Fig 2A,B). This is a key obstacle to evolutionary emergence in short-term infections. In contrast, for long-term infections where the mutant strain has a substantial within-host selective advantage, the mutant strain is transmitted frequently (e.g. the expected number of events > 1 for some contours in Fig 2C,D).

Finally, even if the mutant strain is successfully transmitted, an individual infected with the mutant strain needs to give rise to a major outbreak — the third term of equation (3). This requires the mutant strain to rise in frequency in the infected host population. A mean field analysis for larger bottleneck sizes ($N > 5$ in the simulations) reveals that mutant frequency initially grows geometrically by a factor α that equals the number of mutant virions, on average, transmitted by an individual initially infected with a single mutant virion and $N - 1$ wild type virions (Supplementary Information). We call α the "cross-scale mutant reproductive rate" as it corresponds to the number of mutant virions at the beginning of the next disease generation produced by a mutant virion in the current disease generation. If this cross-scale reproductive rate is greater than one, then each mutant virion replaces itself with more than one mutant virion in the next generation of infection and the frequency of mutant virions increases in the infected host population. If the cross-scale reproductive rate α is less than one, the frequency of mutants decreases, thereby hindering evolutionary emergence.

For short-term infections, the cross-scale mutant reproductive rate α is equal to the ratio of the reproductive numbers:

$$\alpha = \frac{R_m}{R_w} \tag{4}$$

Thus for short-term infections there is no additional condition required for emergence. Whenever the mutant reproductive number R_m exceeds one, there is a mean tendency for the mutant strain to increase in frequency once it has been successfully transmitted to susceptible individuals (i.e. $\alpha > 1$ because $R_m > 1 > R_w$). The greater the ratio R_m/R_w, the more rapid the increase in frequency.

For long-term infections, there is sufficient time for within-host selection to change the frequency of the mutant strain within a host. Larger transmission bottlenecks increase the likelihood that these changes in frequency are transmitted between hosts. For these long infectious periods and larger bottlenecks, a within-host selective disadvantage reduces the cross-scale mutant reproductive rate α (Supplementary Information):

$$\alpha \approx \exp(\tau + sT/2) \text{ for } s \text{ sufficiently small.} \tag{5}$$

Hence, the cross-scale mutant reproductive rate α may be less than one even when the mutant reproductive number R_m is greater than one. This phenomenon, which arises from the interplay of dynamics at within-host and between-host scales, moderated by the transmission bottleneck width, explains the puzzling behavior about the emergence probabilities noted earlier (the white region bounded by solid and dashed red lines in Fig 2D).

The importance of these frequency dynamics can be visualized via individual-based outbreak simulations, and cobwebbing diagrams summarizing the mean field dynamics. When the mutant reproductive number R_m is greater than one but its cross-scale mutant reproductive rate α is less than one, mutant virions may be transmitted but the resulting mixed infections are invariably taken over by purely wild-type infections (Fig 3A). Only pure mutant infections can escape this "relapse" to wild-type, and then only if the mutation rate μ is low enough that new wild-type virions are slow to appear. When the within-host selective disadvantage is weak and the between-host selective advantage is strong, the cross-scale mutant reproductive rate may be slightly greater than one and the mutant strain can drift to higher frequencies within the infected host population (Fig 3B). For large within-host selective advantages, the cross-scale mutant reproductive rate is large and the mutant strain can sweep rapidly to fixation in the infected host population (Fig 3C). Thus, in addition to revealing a new condition needed for evolutionary emergence, the cross-scale mutant reproductive rate α summarizes the conditions under which evolution occurs swiftly or gradually within chains of transmission.

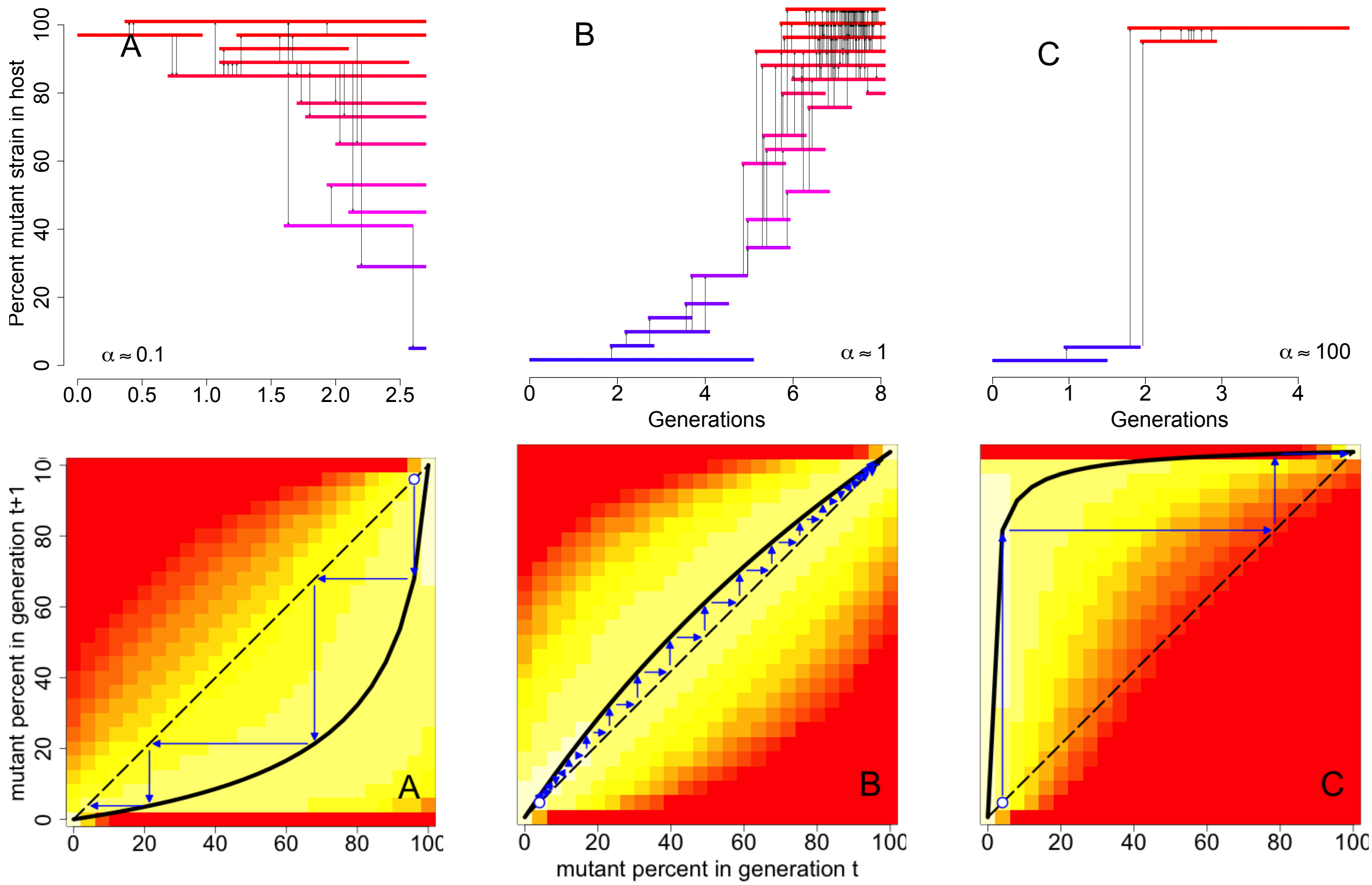


Figure 3. *Frequency dynamics of the mutant strain in the host population.* For long-term infections with moderate to large transmission bottlenecks ($N > 5$), individual-based simulations corresponding to three values of the cross-scale mutant reproductive rate α illustrate (A) the mutant strain decreasing in frequency (despite an index case initially only infected with the mutant strain) when the cross-scale mutant reproductive rate α is less than one, (B) a gradual sweep to fixation of the mutant strain when $\alpha \approx 1$, and (C) fast sweeps to fixation for large values of $\alpha > 1$. In these individual based simulations, each horizontal line marks the infectious period of an individual whose infection was initiated with that percentage of the mutant strain and the vertical arrows represent transmission events between individuals. In the bottom half of the figure, the mean field dynamics corresponding to each of the individual-based simulations are plotted as cobwebbing diagrams. The solid black curves correspond to the expected frequency of the mutant strain in the infected host population in the next generation given the frequency in the current generation. Thin blue lines indicate how the expected frequencies change across multiple generations. The colored backgrounds represent the expected number of individuals infected with a certain percentage of the mutant strain (vertical axis) by an individual with an initial percentage of the mutant strain (horizontal axis). Lighter colors correspond to higher values. Parameter values as in Fig 2D indicated with black asterisks.

The dueling effects of transmission bottlenecks.

Wider bottlenecks increase the likelihood of evolutionary emergence for pathogens with a short infectious period, but can hinder or facilitate evolutionary emergence of long-term infections (Fig 4A,B). For short-term infections, evolutionary emergence is constrained primarily by the transmission of mutant virions by individuals initially infected

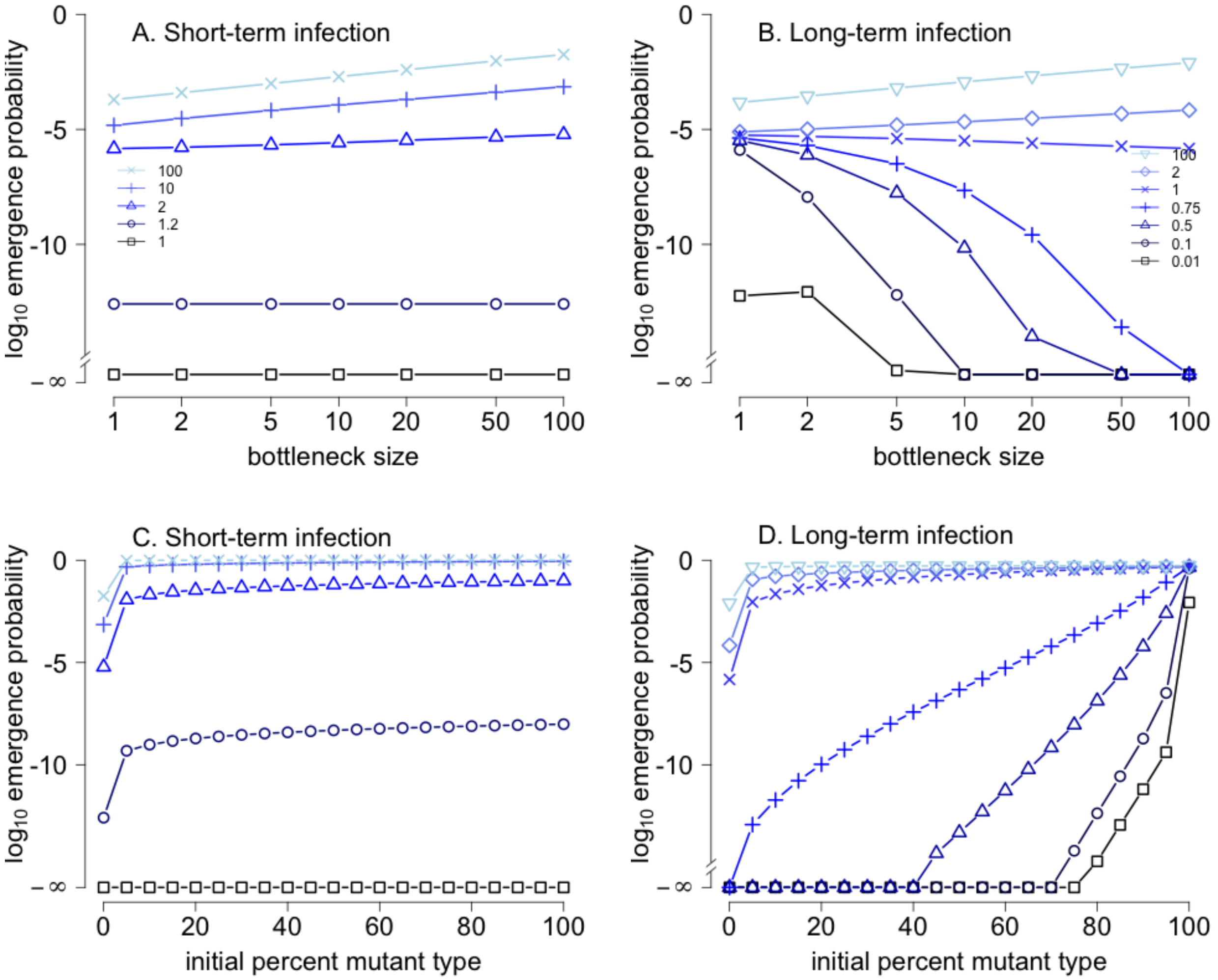


Figure 4. Effects of bottleneck size and mixed infections of the index case on evolutionary emergence of short-term and long-term infections. Different curves correspond to different tendencies, as measured by the cross-scale mutant reproductive rate α, for the mutant strain to increase in frequency in the infected host population. In (A) and (B), bottleneck size has negative effect on emergence when the cross-scale mutant reproductive rate α is less than one and a positive effect when α is greater than one. In (C) and (D), index cases initially infected with higher percentages of the mutant strain are more likely to lead to emergence. $-\infty$ corresponds to numerical values of 10^{-16} or smaller. Parameters: $K = 10^7$, $\beta T = 150$, $T = 7.5$ for short-term infections and $T = 30$ for long-term infections, b_w chosen so that $R_0 = 0.75$ for the wild strain, $r_w = 1.25$, $\tau = 1$, s chosen to achieve the α values reported in the legend, and $\mu = 10^{-7}$. $N = 25$ in (C) and (D).

with only the wild strain. Wider transmission bottlenecks alleviate this constraint, especially when the mutant strain is expected to increase rapidly within the infected population ($\alpha \gg 1$; Fig 4A). When the mutant strain rises slowly in the infected host population (α slightly greater than one), the emergence probability is insensitive to the bottleneck size, regardless of infection duration.

For long-term infections for which the mutant strain's reproductive number R_m is greater than one, but the cross-scale mutant reproductive rate α is less than one, emergence probabilities decrease sharply with bottleneck size (Fig 4B and Supplementary Information). Because a mutant reproductive number R_m greater than one requires a between-host selective advantage ($\tau > 0$) for a long-term infection, the cross-scale mutant reproductive rate α is less than one only if there is a within-host selective disadvantage ($s < 0$) so that mixed infections tend to be taken over by the wild-type. Consequently, the mutant virus can start an epidemic only when a host is infected with mutant particles only, an event that becomes increasingly unlikely for larger bottleneck sizes N.

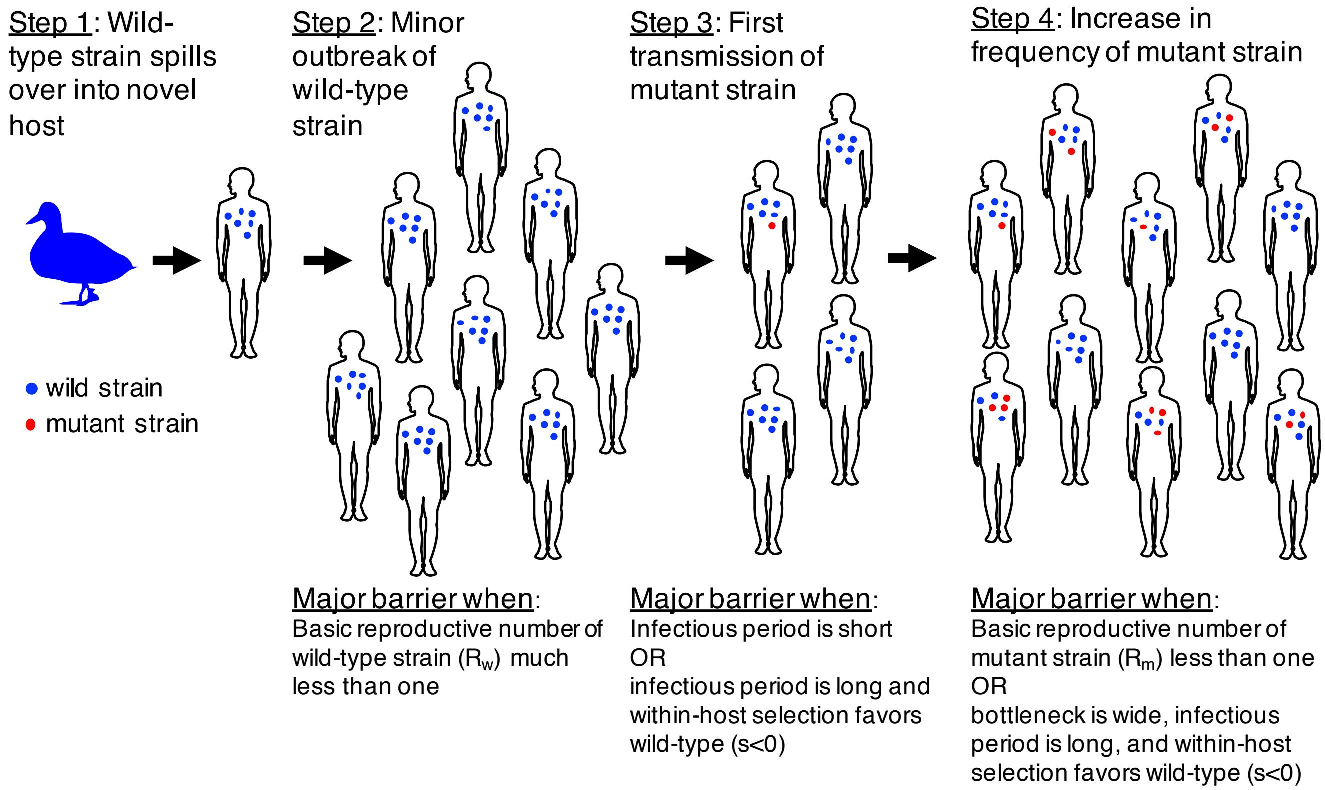
Mutant spillover events hasten evolutionary emergence.

When the mutant strain is circulating in the reservoir, the index case can begin with a mixed infection which invariably makes evolutionary emergence more likely (Fig 4C,D). For short-term infections, spillover doses that contain low or high frequencies of mutants have a roughly equal impact on emergence, and the magnitudes of these increases are relatively independent of the cross-scale mutant reproductive rate α (Fig 4C). This arises because the initial production and transmission of the mutant strain is the primary constraint on evolutionary emergence for short-term infections with $R_m > 1$ (black contours in Fig 2A,B). Consequently, mutant spillover events of any size are sufficient to overcome this constraint.

For long-term infections, the impact of mutant spillover depends on the cross-scale mutant reproductive rate α. When α is less than one, only spillover doses with high frequencies of mutants have a significant effect on emergence (i.e. bottom three curves in Fig 4D). When the cross-scale mutant reproductive rate α is greater than one, the effect mimics short-term infections and mutant spillover events of any size can substantially increase the chance of emergence (top three curves in Fig 4C,D).

Discussion

We have presented a cross-scale model for evolutionary emergence of novel pathogens, linking explicit representations of viral growth and competition within host individuals to viral transmission between individuals. This framework integrates and extends the findings of past theory on this problem by including mixed infections, explicit transmission bottlenecks, and a distinct trait of transmissibility for each viral genotype, phenomena that are highlighted by current empirical research as essential components of viral evolution. Our work identifies four steps to evolutionary emergence (Fig 5) and four ingredients that govern these steps: (i) the reproductive number of the wild type which determines the size of a minor outbreak of this strain, (ii) the rate at which individuals infected initially with the wild-type strain transmit the mutant strain, and (iii) the cross-scale mutant reproductive rate which corresponds to the mean number of mutant virions transmitted by an individual whose initial infection only included one mutant virion, and (iv) the reproductive number of the mutant strain. Prior studies (11–14) identified the importance of the two reproductive numbers and a phenomenological ‘mutation rate’, but ingredients (ii) and (iii) are new mechanistic insights arising from the cross-scale dynamics. By analyzing these ingredients of evolutionary emergence, we show


Step 1: Wild-type strain spills over into novel host
Step 2: Minor outbreak of wild-type strain
Step 3: First transmission of mutant strain
Step 4: Increase in frequency of mutant strain
● wild strain
● mutant strain
Major barrier when:
Major barrier when:
Major barrier when:

Basic reproductive number of wild-type strain (R_w) much less than one

Infectious period is short OR infectious period is long and within-host selection favors wild-type (s<0)

Basic reproductive number of mutant strain (R_m) less than one OR bottleneck is wide, infectious period is long, and within-host selection favors wild-type (s<0)

Figure 5. The major steps and barriers for evolutionary emergence.

how the probability of emergence is governed by selection pressures at within-host and between-host scales, the width of the transmission bottleneck, and the infection duration. We also map the conditions under which different broad-scale patterns are observed, from rapid selective sweeps to slower diffusion of new types. While our study has focused on within-host and between-host scales of selection, it could be generalized readily to other types of cross-scale dynamics where selection may act differently at different scales, such as within-farm and between-farm scales where genetic data have given insights into the emergence of high-pathogenicity avian influenza strains (60).

Previous studies of evolutionary emergence of pathogens (11–14) have assumed infected individuals are, at any point in time, infected primarily by a single pathogen strain. Consequently, shifts from infection with one strain to infection with another must occur abruptly, relative to other processes. Such abrupt shifts could correspond to within-host selective sweeps or, if mutant strains remain at low frequency, to rare events in which only the mutant strain is transmitted. The seminal studies (11, 12) showed that under these conditions the probability of emergence is proportional to the frequency of these events, which they bundled together into a phenomenological "mutation rate".

Our cross-scale analysis identifies the mechanistic counterpart to this phenomenological "mutation rate", which is the probability that an individual infected initially with the wild-type strain ends up transmitting at least one virion of the mutant strain (Step 3 in Fig 5). This quantity, which is approximated by the black contours in Fig 2, is governed chiefly by the ability of the mutant strain to reach an appreciable frequency within the host over the course of an infection. This is evident from the strong dependence on the strength of within-host selection—which surprisingly is much stronger than the dependence on the transmission advantage of mutant virions—and the higher values found for larger bottleneck widths, which favor transmission of low-frequency mutants through a straight-forward sampling effect. The duration of infection plays a crucial role, and our analysis showed that achieving this first transmission of the adaptive mutant is a key barrier to evolutionary emergence for short-term infections (Fig 2A,B). This aligns with the recent finding that potential immune-escape variants of H1N1pdm influenza, expected to have a strong fitness advantage, were present at surprisingly low frequencies in infected humans, and have been detected very rarely at the consensus level (i.e. they have failed to emerge) (61). While more investigation is needed to determine the relevant s and τ parameters for these strains, these data are consistent with the mechanism we identify whereby these variants may be adaptive but have insufficient time to reach high enough frequencies to avoid being lost in transmission bottlenecks.

Our analysis highlights an additional factor, the cross-scale mutant reproductive rate α, previously unrecognized in models neglecting within-host diversity and analyses centered on R_0 for pure infections. Even after the mutant strain has been transmitted, it needs to increase in frequency at the scale of the infected host population (Step 4 in Fig 5). Specifically, each transmitted mutant virion, on average, needs to replace itself with more than one transmitted mutant virion in the next generation of infected hosts. When this occurs, it sets up a positive feedback along chains of infections: individuals with a higher frequency of the mutant strain tend to infect more individuals, which in turn provides more opportunities to transmit, on average, higher frequencies of the mutant strain to the next generation. Conversely, when this between-generation cross-scale mutant reproductive rate is less than one, the positive feedback leads to lower and lower frequencies of the mutant strain within the infected host population. This positive feedback mechanism is stronger for wider transmission bottlenecks (≥ 5 virions in our numerical explorations), which better preserve the mutant frequency from one host to the next.

The directionality of the positive feedback is more complex, and depends on multiple factors including the infection duration and the presence or absence of cross-scale conflicts. For long-term infections, mutant frequencies can drop deterministically within a host, and hence prevent emergence, even if the mutant strain has a reproductive number greater than one. This occurs when the mutant strain has a within-host selective disadvantage and between-host selective advantage (upper left quadrant of Fig 2D); the long infectious period allows time for the within-host disadvantage to drive the mutant strain to lower frequency and, thereby, set up the positive feedback effectively preventing evolutionary emergence. In contrast, for short-term infections the mutant strain tends to rise in frequency whenever the mutant reproductive number is greater than one, because there is insufficient time for any within-host disadvantage to act. In particular, evolutionary emergence may occur despite within-host selective disadvantages, a possibility excluded by previous theory (14). Collectively these two results imply that, in the face of cross-scale conflict and wide transmission bottlenecks, longer infectious periods can inhibit, rather than facilitate (13), evolutionary emergence (Fig 2B,D).

Our cross-scale analysis enables us to address two long-standing and interrelated questions in emerging pathogen research, regarding the influence of transmission bottleneck size on emergence probability and the importance of "pre-adapted" mutations circulating in the animal reservoir (53, 56–58, 62). In both cases, the answer depends on the cross-scale mutant reproductive rate α that governs the frequency feedback. Under most circumstances, wider bottlenecks boost the probability of emergence (Fig 4A,B), because they favor the onward transmission of mutant virions when they are rare; this is particularly vital for the first transmission of mutant virions (i.e. Step 3 in Fig. 5). The exception is for long-term infections with $\alpha < 1$, such that the mutant tends to decline in frequency in the infected host population. Under these circumstances, wider bottlenecks hinder emergence by propagating reductions in the frequency of the mutant strain more efficiently from host to host (Step 4 in Fig. 5). Conventional thinking about the influence of bottlenecks on viral adaptation emphasizes fitness losses due to genetic drift and the effects of Muller's ratchet (31–33), which become more severe for narrower bottlenecks. Contrary to these negative effects of narrow bottlenecks, our findings highlight that narrower bottlenecks can aid emergence in long-term infections with a cross-scale conflict in selection (Fig 4B). Here the adaptive gain in transmissibility at population scales can be impeded by the selective disadvantage at the within-host scale, but, intriguingly, this disadvantage is neutralized by genetic drift arising from narrow bottlenecks. Given the evidence for cross-scale evolutionary conflicts for HIV-1 (43, 48, 49), our results suggest the possibility that HIV-1's narrow transmission bottleneck (37) could play a role in the emergence of novel strains (e.g. drug resistant strains).

Similar mechanisms dictate the influence of mutant viral strains circulating in the reservoir, particularly for long-term infections (Fig 4C,D). If the cross-scale mutant reproductive rate α is greater than one, so that the mutant frequency rises easily in the infected host population, then even low frequencies of mutants in the reservoir lead to substantial risk of emergence. Indeed, for long-term infections with $\alpha > 1$, emergence becomes almost certain when there are mutants in the initial spillover inoculum. Conversely, when the cross-scale mutant reproductive rate is less than one, emergence probability scales with the proportion of mutants in the initial dose, and when $\alpha \ll 1$, the initial dose must consist almost entirely of the mutant strain in order to pose any major risk. These findings yield direct lessons for the growing enterprise of conducting genetic surveillance on zoonotic pathogens in their animal reservoirs (63–65). Generally, risk to humans increases if there is any non-zero proportion of mutant viruses in the spillover inoculum, so tracking the presence of such mutants is beneficial. Surprisingly, the quantitative frequency of mutants in the initial dose has little impact on emergence probability in most scenarios, with the one exception of long-term infections with $\alpha < 1$. Collectively, these results suggest that any knowledge of the cross-scale mutant reproductive rate and mutant reproductive numbers can help to refine our goals for genetic surveillance, and that in many circumstances presence/absence detection is sufficient. Of course, a crucial prerequisite for genetic surveillance is knowledge of genotypes of concern; the integration of various research approaches to address this question, and estimate key quantities, is an on-going research challenge (66, 67).

While there are not sufficient data from past emergence events to test our model's conclusions, recent studies combining animal transmission experiments with deep sequencing have exhibited many phenomena aligned with our findings. Moncla et al. (2016) conducted deep sequencing analyses of H1N1 influenza viruses, in the context of ferret airborne transmission experiments that examined the adaptation of avian-like viruses to the mammalian host. Their results provide in-depth insights into selection within hosts and at transmission bottlenecks, for a range of mutations on genetic backgrounds that change as adaptation proceeds (i.e. equivalent to numerous separate implementations of our model of a single mutational step). They observe a fascinating range of dynamics: some mutations appeared to have α moderately above 1, exhibiting modest increases in frequency between generations, but achieved this outcome with different traits (e.g. S113N on the HA190D225D background exhibited strong within-host selection and no evident transmission advantage, while D265V showed weak within-host selection but its frequency rises in transmission). Another mutation (I187T on the 'Mut' background) appeared to have $\alpha \gg 1$ and exhibited strong selection at both scales; notably, this mutation is widespread in 17/17 human-derived isolates of the post-emergence 1918 virus, consistent with the successful and rapid emergence our model would predict. Moncla et al. also present substantial evidence of cross-scale conflict in selection, as one mutation (G225D on 'Mut' background) exhibited declining frequencies within ferrets but rose to fixation in 2/2 transmission events, while numerous mutations in the HA2 region rose in frequency within the host but were eliminated in transmission. Another study examined a set of 'gain-of-function' mutations in H5N1 influenza in ferrets, and reported a slow rise in frequency when the virus was passaged between ferrets by intranasal inoculation, then rapid fixation of these mutations during airborne transmission (20); the airborne transmission data are consistent with strong between-host selection and a high α value (though we emphasize that circulating H5N1 viruses required substantial modification to the favorable genetic background used in those experiments). Intriguingly, Moncla et al. synthesized their results with those of earlier studies (35, 42, 47) to hypothesize that the 'stringency' of the transmission bottleneck varies systematically during the course of viral adaptation, with loose bottlenecks prevailing when viruses first encounter a new host species (and perhaps again when the virus is host-adapted (30)), and much tighter bottlenecks at the key juncture in host adaptation when a genotype with greater transmissibility is available to be selected. If this hypothesis is correct, then our findings can be applied to each adaptive step independently, and may help to identify which viral traits are most crucial to adaptive steps subject to tighter or looser bottlenecks.

Our results focus on systems where there is one major rate-limiting step to emergence, and the viral population can be represented by one wild-type and one mutant strain. This is a simplification of most viral emergence problems, but will apply directly to systems where a single large-effect mutation is the primary barrier to emergence of a supercritical strain, as for Venezuelan equine encephalitis virus emerging from rodents to horses (68). While it is possible to extend our exact computations and analysis of the cross-scale mutant reproductive rate to systems with multiple mutational steps, the present analysis already provides insights into more complex evolutionary scenarios. For evolutionary trajectories that proceed through a fixed series of genotypes, the probability of emergence can be approximated by extension of our equation (3), as in previous work (11, 12, 14). If emergence requires multiple mutational steps which pass through a fitness valley, then the scale at which this valley occurs matters. A within-host fitness valley in replication rates would hinder pathogens with long-term infections and larger bottleneck widths, more than those with smaller bottlenecks. A between-host fitness valley in transmissibility could hinder evolutionary emergence of pathogens causing long-term infections more than those causing short-term infections, unless the within-host landscape is sufficiently favorable to allow traversing the valley within a single host's long-term infection. Recent studies have also highlighted the importance of considering the broader genotype space, which can reveal indirect paths that circumvent fitness valleys (69), alternative genotypes that yield similar phenotypes (36), and the costs of higher mutation rates arising from deleterious mutants (70).

Our analysis also focuses on a simple "logistic-like" model for within-host viral dynamics. This simplification allows us to study how evolutionary emergence is limited by different factors for pathogens dominated by exponential versus saturated phases of viral growth, while maintaining analytic tractability. Future important extensions would be to allow within-host fitness to alter the carrying capacity in the saturated phase, as well as identifying the relative contributions of stochastic within-host dynamics, immune responses, and host heterogeneity on viral emergence. We have assumed that the bottleneck width N is fixed for a given pathogen. This is broadly consistent with currently available data (37–39), but it will be important to explore the consequences of variation in bottleneck width arising from different routes of transmission, or possibly from changing viral loads. The computational and analytical framework developed here can be extended to account for these additional complexities. Other important extensions can explore the impact of clonal competition on emergence probabilities (71–73) or the potential for complementation to rescue pathogen strains from deep fitness valleys–a mechanism that depends on wide transmission bottlenecks (74).

Our cross-scale analysis opens the door for a new generation of integrative risk assessment models for pathogen emergence, which will integrate growing streams of data collected in laboratories and field surveillance programs (66). At present there is no framework, other than the intuition of individual scientists, to link together the discoveries from targeted experiments, massively parallel phenotypic screens, experimental evolution, clinical medicine, and field epidemiology and disease ecology. Mathematical and computational models that connect biological scales using mechanistic principles can make unique contributions to this transdisciplinary enterprise, by formally integrating diverse empirical findings and by identifying the crucial knowledge gaps to focus future research. The work presented here is a step on the path to realizing this potential.

Models and Methods

The cross-scale model with explicit within-host dynamics

The cross-scale dynamics are modeled as a continuous time, age-dependent, multi-type branching process (Fig 1). The "type" of individual corresponds to their initial viral load, and the "age" of an individual corresponds to the time since their initial infection. Within an infected host, the viral dynamics determine how the viral load and composition changes over time due to competition between strains and mutation events. Transmission events are determined by the viral load and composition of the host and, consequently, are age-dependent.

Due to ultimately large viral loads, the within-host dynamics are modeled with coupled differential equations where $v(t) = (v_w(t), v_m(t))$ denotes the vector of viral abundances:

$$\frac{dv_i}{dt} = \begin{cases} (1-\mu)r_i v_i + \mu r_j v_j & \text{if } v_w + v_m \leq K \\ (1-\mu)r_i v_i + \mu r_j v_j & \\ -\frac{v_i}{v_m+v_w}(r_w v_w + r_m v_m) & \text{else} \end{cases} \tag{6}$$

$$\text{with } i \neq j \in \{w, m\}.$$

At time $t = 0$, $v(0) = (v_w(0), v_m(0))$ corresponds to the initial viral load of an infected individual.

The number of contacts of an infected during the infectious period is Poisson distributed with mean βT. On the event of a contact at t days after becoming infected, the probability of transmission equals $p(b_w v_w(t) + b_m v_m(t))$. On the event of transmission, the probability of infecting an individual with a viral load of $\tilde{v} = (\tilde{v}_w, \tilde{v}_m)$ with $\tilde{v}_w + \tilde{v}_m = N$ equals

$$\psi(v(t), \tilde{v}) = \frac{N!}{\tilde{v}_w! \tilde{v}_m!} \left(\frac{b_w v_w(t)}{b_w v_w(t) + b_m v_m(t)}\right)^{\tilde{v}_w} \left(\frac{b_m v_m(t)}{b_w v_w(t) + b_m v_m(t)}\right)^{\tilde{v}_m}$$

Under these assumptions, during their infectious period, an infected individual of type $v(0)$ infects a Poisson distributed number of individuals with viral load $\tilde{v}$ and the mean of this distribution equals

$$F(v(0), \tilde{v}) = \int_0^T \beta \psi(v(t), \tilde{v}) p(b_w v_w(t) + b_m v_m(t))\, dt.$$

Methods

To solve the probabilities of emergence, we use the discrete-time branching process given by censusing the infected the population at the beginning of each generation of infection i.e at times 0, T, $2T$, etc. All the statistics of this process are given by the probability generating map $G : [0,1]^{N+1} \to [0,1]^{N+1}$ where $N+1$ is the number of types of initial viral loads. We index the coordinates by the initial number of mutant virions $0, 1, 2, \ldots, N$ within an infected individual and have

$$G_i(s) = \exp\left(\sum_j F((i, N-i), (j, N-j))(s_j - 1)\right).$$

By the limit theorem for multitype branching processes, the i-th coordinate of

$$q(t) = G^t(0, \dots, 0)$$

is the probability of extinction by generation t when there is initially one infected individual with initial viral load $(i, N-i)$.

For the numerical work, we used linear, logarithmic, and saturating functions for the transmission probability function p. All gave similar results but we present the linear case as most analytical results were derived for this case. To compute the extinction probabilities, we iterated the generation map G for $2,000$ generations. For the individual based simulations, we solved the within-host differential equations using matrix exponentials and renormalizing these exponentials when the viral load reached the value K. Between host transmission events where determined by a time-dependent Poisson process with rate function $p(b_w v_w(t) + b_m v_m(t))$ and mulitnomial sampling was used to determine the initial viral load of an infected individual.

Acknowledgements. This work was supported by U.S. National Science Foundation Grants EF-0928987 and DMS-1313418 to SJS, and EF-0928690 to JLS.

References

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451:990–993.
2. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367:2864–2871.
3. Morse SS, Mazet JAK, Woolhouse M, Parrish CR, Carroll D, Karesh WB, et al. Prediction and prevention of the next pandemic zoonosis. The Lancet. 2012;380:1956–1965.
4. Howard CR, Fletcher NF. Emerging virus diseases: can we ever expect the unexpected? Emerging Microbes and Infections. 2012;1:e46.
5. Steel J, Lowen AC, Mubareka S, P P. Transmission of Influenza Virus in a Mammalian Host Is Increased by PB2 Amino Acids 627K or 627E/701N. PLOS Pathogens. 2009;5:e1000252.
6. Lowder BV, Guinane CM, Ben Zakour NL, Weinert LA, Conway-Morris A, Cartwright RA, et al. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. Proceedings of the National Academy of Sciences. 2009;106:19545–19550.

7. Xu L, Bao L, Deng W, Zhu H, Li F, Chen T, et al. Rapid adaptation of avian H7N9 virus in pigs. Virology. 2014;452-453:231–236.

8. Jonges M, Welkers MR, Jeeninga RE, Meijer A, Schneeberger P, Fouchier RA, et al. Emergence of the virulence-associated PB2 E627K substitution in a fatal human case of highly pathogenic avian influenza virus A (H7N7) infection as determined by Illumina ultra-deep sequencing. Journal of virology. 2014;88(3):1694–1702.

9. Zhu W, Shu Y. Genetic tuning of avian influenza A (H7N9) virus promotes viral fitness within different species. Microbes and Infection. 2015;17:118 – 122.

10. Lam T, Zhou B, Wang J, Chai Y, Shen Y, Chen X, et al. Dissemination, divergence and establishment of H7N9 influenza viruses in China. Nature. 2015;522:102–105.

11. Antia R, Regoes RR, Koella JC, Bergstrom CT. The role of evolution in the emergence of infectious diseases. Nature. 2003;426:658–661.

12. Iwasa Y, Michor F, Nowak MA. Evolutionary dynamics of invasion and escape. Journal of Theoretical Biology. 2004;226:205–214.

13. André JB, Day T. The effect of disease life history on the evolutionary emergence of novel pathogens. Proceedings of the Royal Society B: Biological Sciences. 2005;272:1949–1956.

14. Park M, Loverdo C, Schreiber SJ, Lloyd-Smith JO. Multiple scales of selection influence the evolutionary emergence of novel pathogens. Philosophical Transactions of the Royal Society B: Biological Sciences. 2013;368:20120333.

15. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303:327–332.

16. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nature Reviews Genetics. 2009;10:540–550.

17. Pennings PS, Kryazhimskiy S, Wakeley J. Loss and recovery of genetic diversity in adapting populations of HIV. PLOS Genetics. 2014;10:e1004000.

18. Lim SG, Cheng Y, Guindon S, Seet BL, Lee LY, Hu P, et al. Viral quasi-species evolution during hepatitis Be antigen seroconversion. Gastroenterology. 2007;133:951–958.

19. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiology and Molecular Biology Reviews. 2012;76:159–216.

20. Linster M, van Boheemen S, de Graaf M, Schrauwen EJ, Lexmond P, Mänz B, et al. Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. Cell. 2014;157:329–339.

21. Lemey P, Rambaut A, Pybus OG. HIV evolutionary dynamics within and among hosts. AIDS Reviews. 2006;8:125–140.

22. Holmes EC. Viral evolution in the genomic age. PLOS Biology. 2007;5:2104–2105.

23. Murcia PR, Baillie GJ, Daly J, Elton D, Jervis C, Mumford JA, et al. Intra-and interhost evolutionary dynamics of equine influenza virus. Journal of Virology. 2010;84:6943–6954.

24. Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, White PA. Contribution of intra-and interhost dynamics to norovirus evolution. Journal of Virology. 2012;86:3219–3229.

25. Hughes J, Allen RC, Baguelin M, Hampson K, Baillie GJ, Elton D, et al. Transmission of equine influenza virus during an outbreak is characterized by frequent mixed infections and loose transmission bottlenecks. PLOS Pathogens. 2012;8:e1003081.

26. Stack JC, Murcia PR, Grenfell BT, Wood JL, Holmes EC. Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. Proceedings of the Royal Society B: Biological Sciences. 2012;.
27. Morelli MJ, Wright CF, Knowles NJ, Juleff N, Paton DJ, King DP, et al. Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. Veterinary Research. 2013;44:1–15.
28. Orton RJ, Wright CF, Morelli MJ, Juleff N, Theibaud G, Knowles NJ, et al. Observing micro-evolutionary processes of viral populations at multiple scales. Philosophical Transactions of the Royal Society B: Biological Sciences. 2013;368.
29. Vrancken B, Rambaut A, Suchard MA, Drummond A, Baele G, Derdelinckx I, et al. The Genealogical Population Dynamics of HIV-1 in a Large Transmission Chain: Bridging within and among Host Evolutionary Rates. PLOS Computational Biology. 2014;10:e1003505.
30. Poon LLM, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, et al. Quantifying influenza virus diversity and transmission in humans. Nature Genetics. 2016;48:195–200.
31. Duarte E, Clarke D, Moya A, Domingo E, Holland J. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. Proceedings of the National Academy of Sciences. 1992;89:6015–6019.
32. Bergstrom CT, McElhany P, Real LA. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proceedings of the National Academy of Sciences. 1999;96:5095–5100.
33. Novella IS, Quer J, Domingo E, Holland JJ. Exponential fitness gains of RNA virus populations are limited by bottleneck effects. Journal of Virology. 1999;73:1668–1671.
34. Gutierrez S, Michalakis Y, Blanc S. Virus population bottlenecks during within-host progression and host-to-host transmission. Current Opinion in Virology. 2012;2:546–555.
35. 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, and Microbe. 2014;16:691–700.
36. Moncla LH, Zhong G, Nelson CW, Dinis JM, Mutschler J, Hughes AL, et al. Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus. Cell, Host, and Microbe. 2016;19:169–180.
37. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proceedings of the National Academy of Sciences. 2008;105:7552–7557.
38. Wang GP, Sherrill-Mix SA, Chang KM, Quince C, Bushman FD. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. Journal of Virology. 2010;84:6218–6228.
39. Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, et al. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. PLOS Pathogens. 2011;7:e1002243.
40. Emmett KJ, Lee A, Khiabanian H, Rabadan R. High-resolution genomic surveillance of 2014 Ebolavirus using shared subclonal variants. PLOS Currents Outbreaks. 2015;1.
41. Murcia PR, Hughes J, Battista P, Lloyd L, Baillie GJ, Ramirez-Gonzalez RH, et al. Evolution of an Eurasian avian-like influenza virus in naive and vaccinated pigs. PLOS Pathogens. 2012;8:e1002730.
42. 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. Nature Communications. 2013;4.
43. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, et al. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science. 2014;345:1254031.

44. Hurt AC, Nor'e SS, McCaw JM, Fryer HR, Mosse J, McLean AR, et al. Assessing the viral fitness of oseltamivir-resistant influenza viruses in ferrets using a competitive-mixtures model. Journal of Virology. 2010;84:9427–9438.
45. McCaw JM, Arinaminpathy N, Hurt AC, McVernon J, McLean AR. A mathematical framework for estimating pathogen transmission fitness and inoculum size using data from a competitive mixtures animal model. PLOS Computational Biology. 2011;7:e1002026.
46. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, Shinya K, et al. Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. Nature. 2012;486:420–428.
47. Zaraket H, Baranovich T, Kaplan BS, Carter R, Song M, Paulson JC, et al. Mammalian adaptation of influenza A (H7N9) virus is limited by a narrow genetic bottleneck. Nature Communications. 2015;6:6553.
48. Shaw GM, Hunter E. HIV transmission. Cold Spring Harbor perspectives in medicine. 2012;2:a006965.
49. Alizon S, Fraser C. Within-host and between-host evolutionary rates across the HIV-1 genome. Retrovirology. 2013;10.
50. Sorrell EM, Wan H, Araya Y, Song H, Perez DR. Minimal molecular constraints for respiratory droplet transmission of an avian–human H9N2 influenza A virus. Proceedings of the National Academy of Sciences. 2009;106:7565–7570.
51. Goodman CD, Siregar JE, Mollard V, Vega-Rodríguez J, Syafruddin D, Matsuoka H, et al. Parasites resistant to the antimalarial atovaquone fail to transmit by mosquitoes. Science. 2016;352:349–353.
52. Wasik BR, Bhushan A, Ogbunugafor CB, Turner PE. Delayed transmission selects for increased survival of vesicular stomatitis virus. Evolution. 2015;69:117–125.
53. Gog JR, Pellis L, Wood JL, McLean AR, Arinaminpathy N, Lloyd-Smith JO. Seven challenges in modeling pathogen dynamics within-host and across scales. Epidemics. 2015;10:45–48.
54. Mideo N, Alizon S, Day T. Linking within-and between-host dynamics in the evolutionary epidemiology of infectious diseases. Trends in ecology and evolution. 2008;23:511–517.
55. Peck KM, Chan CHS, Tanaka MM. Connecting within-host dynamics to the rate of viral molecular evolution. Virus Evolution. 2015;1:vev013.
56. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, et al. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiology and Molecular Biology Reviews. 2008;72:457–470.
57. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human. viruses: discovery and emergence. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367:2864–2871.
58. Orr HA, Unckless RL. The population genetics of evolutionary rescue. PLOS Genetics. 2014;10:e1004551.
59. Diekmann O, Heesterbeek JAP. Mathematical epidemiology of infectious diseases. vol. 146. Wiley, Chichester; 2000.
60. Fusaro A, Tassoni L, Milani A, Hughes J, Salviato A, Murcia PR, et al. Unexpected Inter-farm Transmission Dynamics During a Highly Pathogenic Avian Influenza Epidemic. Journal of Virology. 2016;90:6401–11.
61. Dinis JM, Florek NW, Fatola OO, Moncla LH, Mutschler JP, Charlier OK, et al. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. Journal of Virology. 2016;90:3355–3365.
62. Pepin KM, Lass S, Pulliam JRC, Read AF, Lloyd-Smith JO. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nature Reviews Microbiology. 2010;8:802–813.

63. Russell CA, Fonville JM, Brown AE, Burke DF, Smith DL, James SL, et al. The potential for respiratory droplet transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science. 2012;336:1541–1547.

64. Davis CT, Chen L, Pappas C, Stevens J, Tumpey TM, Gubareva LV, et al. Use of highly pathogenic avian influenza A (H5N1) gain-of-function studies for molecular-based surveillance and pandemic preparedness. MBio. 2014;5:e02431–14.

65. Watanabe T, Zhong G, Russell CA, Nakajima N, Hatta M, Hanson A, et al. Circulating avian influenza viruses closely related to the 1918 virus have pandemic potential. Cell, Host, and Microbe. 2014;15:692–705.

66. Russell CA, Kasson PM, Donis RO, Riley S, Dunbar J, Rambaut A, et al. Improving pandemic influenza risk assessment. eLife. 2014;3:e03883.

67. Geoghegan JL, Senior AM, Di Giallonardo F, Holmes EC. Virological factors that increase the transmissibility of emerging human viruses. Proceedings of the National Academy of Sciences. 2016;113:4170–4175.

68. Anishchenko M, Bowen RA, Paessler S, Austgen L, Greene IP, Weaver SC. Venezuelan encephalitis emergence mediated by a phylogenetically predicted viral mutation. Proceedings of the National Academy of Sciences of the United States of America. 2006;103:4994–4999.

69. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. Adaptation in protein fitness landscapes is facilitated by indirect paths. eLife. 2016;5:e16965.

70. Loverdo C, Lloyd-Smith JO. Evolutionary invasion and escape in the presence of deleterious mutations. PLoS ONE. 2013;8:e68179.

71. Strelkowa N, Lässig M. Clonal interference in the evolution of influenza. Genetics. 2012;192:671–682.

72. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. Genetica. 1998;102:127–144.

73. Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proceedings of the National Academy of Sciences. 2012;109:4950–4955.

74. Ke R, Aaskov J, Holmes EC, Lloyd-Smith JO. Phylodynamic analysis of the emergence and epidemiological impact of transmissible defective dengue viruses. PLOS Pathogens. 2013;9:e1003193.

Supplementary Information

Derivation of the Single Strain Reproductive Numbers

When the within-host dynamics only exhibit exponential growth (i.e. $N\exp(r_iT) < K$) and there is a linear transmission function, the basic reproductive numbers equal

$$R_i = \beta b_i N \frac{\exp(r_iT) - 1}{r_i} \text{ for } i = w, m. \tag{S–1}$$

When the within-host dynamics saturate (i.e. $N\exp(r_iT) > K$), the basic reproductive number equals

$$R_i = \beta b_i \left(\frac{K - N}{r_i} + KT_s \right) \tag{S–2}$$

where $T_e = \log(K/N)/r_i$ is the length of exponential phase and $T_s = T - T_e$ is the length of saturated phase.

We derive two approximations of R_i under the assumption that s is small, $\exp(r_iT) \gg 1$, and $K \gg N$. First, assume that the infection is short-term in which case $T_e = T$. Then provided s is sufficiently small to ensure that the mutant type doesn't saturate, R_i are given by (S–1). The log ratio, provided $\exp(r_iT) \gg 1$, satisfies

$$\begin{aligned}
\log \frac{R_m}{R_w} &\approx \log \frac{\beta b_m N \exp((r_w + s)T)/(r_w + s)}{\beta b_w N \exp(r_wT)/r_w} \\
&= \tau + sT + \log \frac{r_w}{r_w + s} \\
&\approx \tau + s(T - 1/r_w)
\end{aligned}$$

which yields (1) in the main text.

Now assume that the infection is long-term in which case $T_e < T$, and that if $s < 0$, $|s|$ is sufficiently small to ensure that the mutant type also saturates before time T. Then the basic reproductive numbers R_i are given by (S–2). If $K \gg N$, then

$$\begin{aligned}
\log \frac{R_m}{R_w} &\approx \log \frac{\beta b_m K \left(\frac{1}{r_w+s} + T - \frac{1}{r_w+s} \log \frac{K}{N} \right)}{\beta b_w K \left(\frac{1}{r_w} + T_s \right)} \\
&= \tau + \log \frac{\frac{r_w}{r_w+s} + Tr_w - \frac{r_w}{r_w+s} \log \frac{K}{N}}{1 + r_wT_s}.
\end{aligned}$$

Assume that $|s| \ll r_w$. Then

$$\log \frac{R_m}{R_w} \approx \tau + \log \frac{1 - \frac{s}{r_w} + Tr_w - (1 - \frac{s}{r_w}) \log \frac{K}{N}}{1 + r_wT_s}.$$

As $T = T_s + T_e$ and

$$(1 - \frac{s}{r_w}) \log \frac{K}{N} = \frac{r_w - s}{r_w} \log \frac{K}{N} = (r_w - s)T_e,$$

it follows that

$$\log \frac{R_m}{R_w} \approx \tau + \log \left(1 + \frac{s}{r_w} \frac{r_w T_e - 1}{1 + r_w T_s} \right).$$

As $\log(1 + x) \approx x$ for small x and $|s| \ll r_w$ by assumption, we obtain

$$\log \frac{R_m}{R_w} \approx \tau + \frac{s}{r_w} \frac{r_w T_e - 1}{r_w T_s + 1}.$$

Equation (2) in the main text follows in the case that $T_s \gg T_s$ in which case the second term is approximately zero.

Derivation of the Emergence Probability Approximation

For small mutation likelihood μ, we derive a mathematically explicit version of the approximation (3) for the emergence probability from the main text. As stated in the main text, this approximation is given by the product of three terms: the expected number of secondary, wild-type cases produced during a fade-out, the mean number of individuals infected with mutant virions by an individual initially infected only with the wild-type, and the probability of emergence from an individual infected with a single mutant virion. As noted in the main text, the first term is given by $\frac{1}{1-R_w}$. The second term requires more work. To derive an analytic approximation for this term, notice that the mean number of individuals infected with ℓ mutant virions by an individual only infected with the wild type equals

$$\int_0^T \beta p \left(v_w(t) b_w + v_m(t) b_m \right) \psi((v_w(t), v_m(t)), (N - \ell, \ell)) dt$$

where $v_w(t), v_m(t)$ is the solution of the within host viral dynamics with $v_w(0) = N, v_m(0) = 0$, and $\psi((v_w, v_m), (\tilde{v}_w, \tilde{v}_m))$ is the probability of an individual with viral load (v_w, v_m) infecting an individual with a viral load of $(\tilde{v}_w, \tilde{v}_m)$ where $\tilde{v}_w + \tilde{v}_m = N$. The solution $(v_m(t), v_m(t))$ is given by

$$v_i(t) = V_i(t) \text{ if } V_w(t) + V_m(t) \le K \text{ and } K \frac{V_i(t)}{V_w(t) + V_m(t)} \text{ otherwise}$$

where $V_w(t), V_m(t)$ are the solutions to

$$\begin{aligned} \frac{dV_w}{dt} &= r_w(1 - \mu)V_w + r_m \mu V_m \qquad V_w(0) = N \\ \frac{dV_m}{dt} &= r_m(1 - \mu)V_m + r_w \mu V_w \qquad V_m(0) = 0. \end{aligned}$$

Ignoring back mutations (i.e. setting $r_m\mu = 0$ and $r_m(1-\mu)$ to r_m), the solutions for $V_w(t), V_m(t)$ are approximately

$$V_w(t) \approx N\exp(r_w t)$$
$$V_m(t) \approx \mu N \frac{r_w}{r_w - r_m}\left(\exp(r_w t) - \exp(r_m t)\right)$$

if $r_w \neq r_m$, and

$$V_m(t) \approx \mu N r_w t \exp(r_w t)$$

if $r_w = r_m = r$. Since $\frac{x}{a+x} \approx x/a$ to first order near 0, the weighted frequency, $x_m(t)$, of mutant strain is approximately

$$x_m(t) = \frac{b_m V_m(t)}{b_w V_w(t) + b_m V_m(t)} \approx \exp(\tau)\mu r_w \frac{\exp(st)-1}{s} \text{ where } s = r_m - r_w \text{ and } \frac{b_m}{b_w} = \exp(\tau)$$

if $s \neq 0$, and

$$x_m(t) \approx \mu r_w t \exp(\tau)$$

if $s = 0$. We have

$$\psi((v_w(t), v_m(t)), (N-\ell, \ell)) \quad = \quad \frac{N!}{\ell!(N-\ell)!} x_w(t)^{N-\ell} x_m(t)^{\ell}.$$

For $\ell \geq 2$, these terms are of order μ^2 and therefore will be ignored. Hence, the only term of interest is $\ell = 1$:

$$\psi((v_w(t), v_m(t)), (N-1, 1)) \approx \mu N r_w \exp(\tau)\frac{\exp(st)-1}{s} \text{ if } s \neq 0 \text{ and } \mu N r_w t \exp(\tau) \text{ otherwise}$$

We also can approximate (assuming p is differentiable)

$$p(b_w V_w(t) + b_m V_m(t)) \approx p(b_w V_w(t)) + O(\mu) = p(b_w \min\{K, N\exp(r_w t)\}) + O(\mu)$$

We drop the $O(\mu)$ term as it will only lead to higher order terms in the approximation.

Putting all of this together gives the following approximation for the mutant force of infection

$$\beta \int_0^T \mu p(b_w \min\{K, N\exp(r_w t)\}) N r_w \exp(\tau) \frac{\exp(st)-1}{s} dt \qquad \text{(S–3)}$$

if $s \neq 0$, and

$$\beta \int_0^T \mu p(b_w \min\{K, N\exp(r_w t)\}) N r_w \exp(\tau) t dt \qquad \text{(S–4)}$$

if $s = 0$.

In the case of a linear transmission function $p(x) = x$, we can write down explicit expressions for (S–3) and (S–4). There are two cases to consider. First suppose that $N\exp(r_w T) \leq K$ i.e. the infection is short-term. Then, integrating and simplifying yields the following approximation for the mutant force of infection

$$\beta\mu N^2 b_w \frac{r_w}{s}\exp(\tau)\left(\frac{\exp((r_w+s)T)-1}{r_w+s} - \frac{(\exp(r_w T)-1}{r_w}\right).$$

Assuming $r_w \gg s$ (and thus $r_w/(r_w+s) \simeq 1 - s/r_w$),

$$\beta\mu N^2 b_w \exp(\tau)\left(\underbrace{\frac{\exp(sT)-1}{s}}_{\approx \text{ (for small } s\text{) } (1+sT/2)T} \exp(r_w T) - \frac{\exp((r_w+s)T)-1}{r_w}\right) \tag{S–5}$$

if $s \neq 0$ and

$$\beta\mu N^2 b_w \exp(\tau)\left(T\exp(r_w T) - \frac{\exp(r_w T)-1}{r_w}\right)$$

if $s = 0$. Now rather than writing out the entire expression for the case $N\exp(r_w T) \geq K$, lets write down things for when the time in the saturated phase is much, much longer than the time in the exponential phase. Then, integrating and simplifying yields the following approximation for the mutant force of infection

$$\beta\mu N K b_w r_w \exp(\tau)\underbrace{\left(\frac{\exp(sT)-1}{s} - T\right)/s}_{\approx \text{ (small } s\text{) } (1+sT/3)T^2/2} \tag{S–6}$$

if $s \neq 0$, and

$$\beta\mu N K b_w r_w \exp(\tau) T^2/2$$

otherwise.

Putting this all together, (3) for an short-term (respectively long-term) infection with $s \neq 0$ becomes the product of $\frac{1}{1-R_w}$, (S–5) (respectively (S–6)), and the probability of an outbreak starting with one individual infected with $N-1$ wild type virions and 1 mutant type virions. The final probability term can be calculated exactly using the generating functions described in the *Models and Methods* section of the main text. Fig S–1 illustrates the effectiveness of this approximation, and Fig S–2 plots the the error in the approximation.

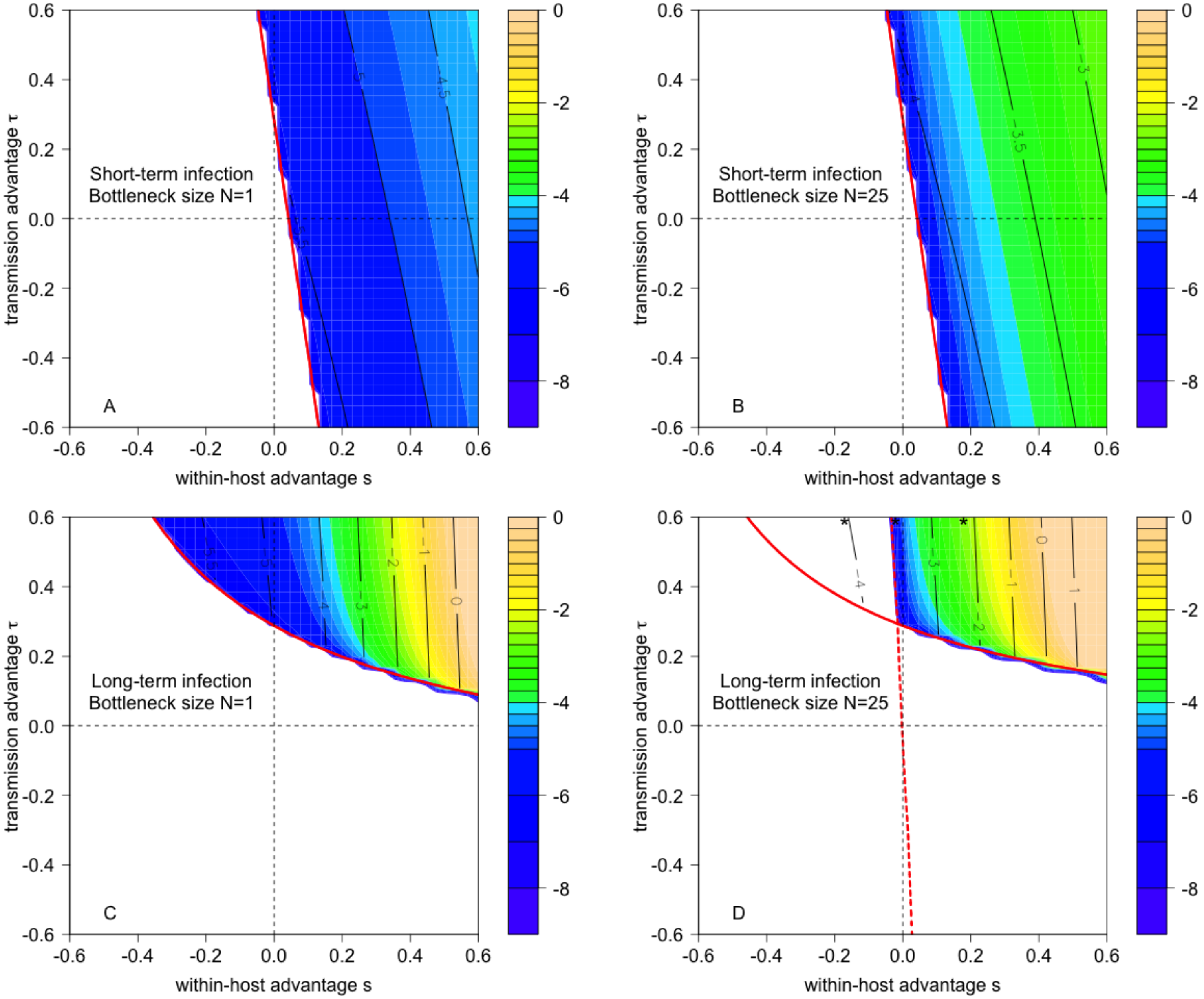


Figure S–1. The analytic approximation based on (3) for the exact computations of the emergence probabilities in Fig 2 in the main text.

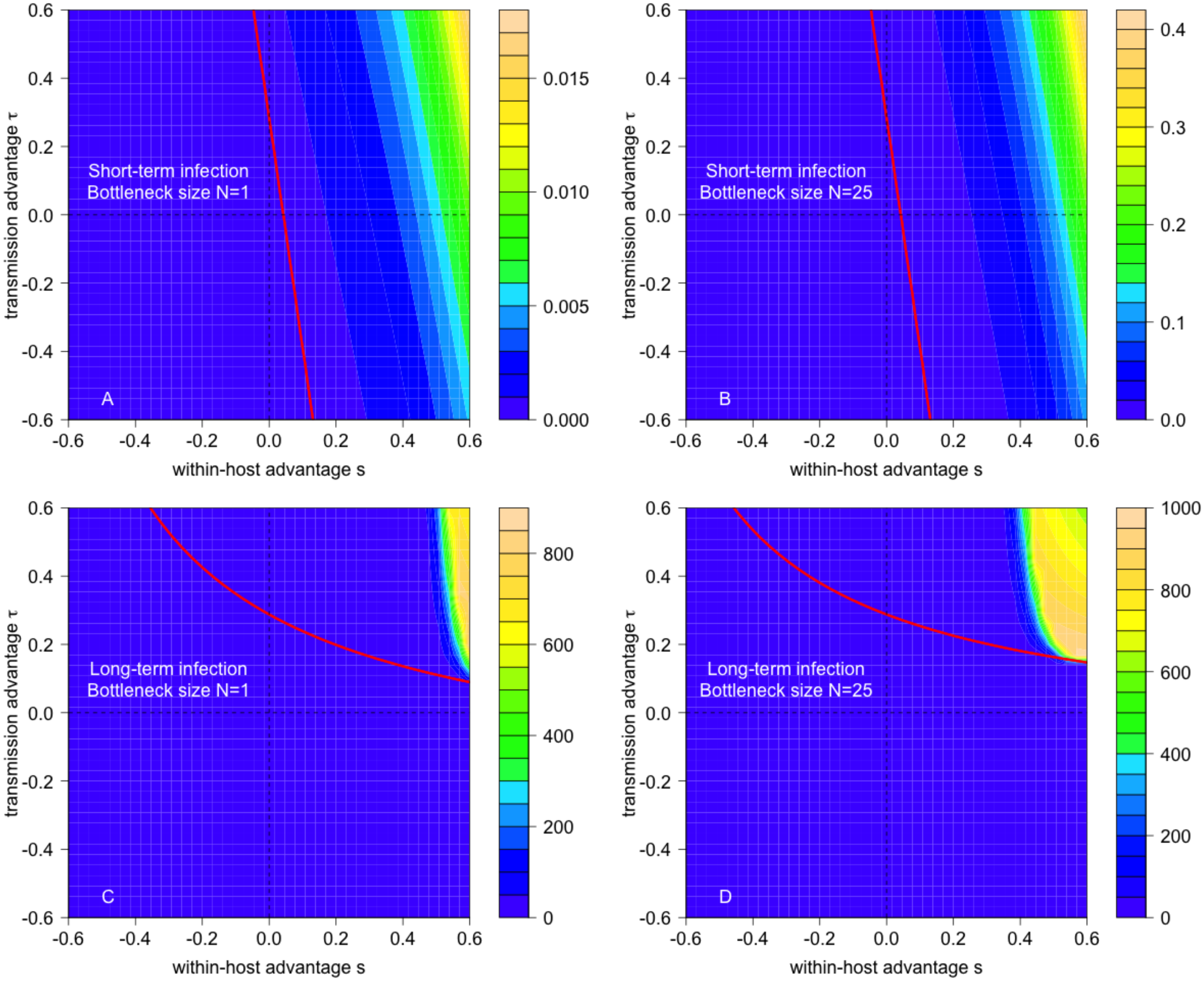


Figure S–2. Contour plots of 1,000 × the absolute value in the difference between the analytic approximation from (3) and the exact computations for the emergence probabilities. Parameters as in Fig 2 in the main text.

Derivation of the Mean Field Frequency Dynamics

To understand how the viral composition of infected individuals change across generations, we derive a mean field approximation for the dynamics of the mean mutant viral load at the beginning of each generation of disease spread. To this end, we define a map from $h : [0,1] \to [0,1]$ where $x \in [0,1]$ represents the current mean mutant viral load in the population at the beginning of the infectious period and $h(x)$ is the mean at the beginning of infectious period in the next generation. Our derivation of this mean field dynamic is done in the limit of large $N \uparrow \infty$ and $\mu \downarrow 0$. None-the-less, as shown by the dashed red line in Fig 2D, this approximation works quite well away from this limit.

We begin by approximating the mean initial mutant viral count in individuals infected by an individual with $V_w(0) = N - \ell$ and $V_m(0) = \ell$. Recall, the force of infection for producing individuals initially infected with j mutant viral particles is given by

$$F((N-\ell,\ell),(N-j,j)) = \beta \int_0^T p(\sum_i b_i v_i(t)) \frac{N!}{(N-j)!j!} y(t)^j (1-y(t))^j \, dt$$

where $y(t) = \frac{b_m v_m(t)}{b_w v_w(t) + b_m v_m(t)}$ is the within-host frequency of the mutant strain, and $(v_w(t), v_m(t))$ is the solution of the within-host viral dynamics with initial condition $v_w(0) = N - \ell$, $v_m(0) = \ell$. Weighting this term by j and summing over j yields the expected number of mutant viral particles in an individual infected by our type $(N-\ell,\ell)$ infected individual:

$$\beta \int_0^T p(\sum_i b_i v_i(t)) \sum_j j \frac{N!}{(N-j)!j!} y(t)^j (1-y(t))^j \, dt = \beta \int_0^T p(\sum_i b_i v_i(t)) N y(t) dt$$

Now if we let $x = \ell/N$ denote the initial fraction, then dividing the previous integral by the net number of viral particles infecting new individuals yields our desired update rule

$$h(x) := \frac{\beta \int_0^T p(\sum_i b_i v_i(t)) N y(t) dt}{\beta \int_0^T p(\sum_i b_i v_i(t)) N dt} = \frac{\int_0^T p(\sum_i b_i v_i(t)) y(t) dt}{\int_0^T p(\sum_i b_i v_i(t)) dt} \tag{S–7}$$

Note that $h(x)$ is a function of x as the solution of $(V_w(t), V_m(t))$ depends on its initial condition $V_w(0) = (1-x)N, V_m(0) = xN$.

The points $x = 0$ and $x = 1$ are fixed points for h corresponding to a mutant-free and wild-type-free states.

Stability of the fixed point $x = 0$ is determined by

$$\begin{aligned} h'(0) = & \frac{\int_0^T \frac{\partial}{\partial x} p(\sum_i b_i v_i(t)) y(t) dt \; \int_0^T p(\sum_i b_i v_i(t)) dt \Big|_{x=0}}{\left(\int_0^T p(\sum_i b_i v_i(t)) dt\right)^2 \Big|_{x=0}} \\ & - \frac{\int_0^T p(\sum_i b_i v_i(t)) dt \int_0^T p(\sum_i b_i v_i(t)) y(t) dt \Big|_{x=0}}{\left(\int_0^T p(\sum_i b_i v_i(t)) dt\right)^2 \Big|_{x=0}} \\ = & \frac{\beta \int_0^T \frac{\partial}{\partial x} p(\sum_i b_i v_i(t)) y(t) dt \Big|_{x=0}}{\beta \int_0^T p(b_w v_w(t)) dt} \approx \frac{\beta \int_0^T p(\sum_i b_i v_i(t)) y(t) dt \Big|_{x=1/N}}{\beta \int_0^T p(b_w v_w(t)) dt} \end{aligned}$$

for $N \gg 1$. $h'(0)$ corresponds to α described in the main text and the final expression has the verbal interpretation given in the main text.

In the special case of a linear transmission function, $p(x) = x$, we get the simplified expression

$$h(x) = \frac{\int_0^T b_m v_m(t) dt}{\int_0^T \sum_i b_i v_i(t) dt} \tag{S–8}$$

where

$$\begin{aligned} v_w(t) &= N(1-x)e^{r_w t} \text{ if } v_w(t) + v_m(t) \le K, \text{ otherwise } K \frac{(1-x)e^{r_w t}}{(1-x)e^{r_w t} + x e^{r_m t}} \\ v_m(t) &= N x e^{r_m t} \text{ if } v_w(t) + v_m(t) \le K, \text{ otherwise } K \frac{x e^{r_m t}}{(1-x)e^{r_w t} + x e^{r_m t}} \end{aligned}$$

Carrying out the integration, in general, is complicated by the fact that the time at which $V_w(t) + V_m(t) = K$ has no explicit formula when $s \neq 0$ and, in general, this saturation time will depend on x.

In the special case of short-term infections (i.e. there is only exponential growth), we get

$$h(x) = \frac{\eta x}{(\eta - 1)x + 1}$$

where

$$\eta = \exp(\tau) \frac{r_w}{\exp(r_w T) - 1} \frac{\exp(r_m T) - 1}{r_m} = \frac{R_m}{R_w}$$

α is defined as $h'(0)$, which here is equal to η, thus for short-term infections, $\alpha = R_m / R_w$.

Since $R_w < 1$ by assumption and $R_m > 1$ is necessary for emergence, we always have $\alpha > 1$ and so the frequency dependent dynamics at the scale of the host population can not significantly impede emergence.

Now, lets consider the more difficult case of a long-term infection with a saturated phase to the within-host viral dynamics. Then

$$h'(0) = b_m \frac{\int_0^T \frac{\partial v_m(t)}{\partial x}\Big|_{x=0} dt}{\int_0^T b_w v_w(t)\Big|_{x=0} dt}.$$

For x close to 0, we have the time at which the dynamics saturate, T_e, is given approximately by

$$T_e \approx \frac{1}{r_w} \log \frac{K}{N}$$

in which case

$$\frac{\partial v_m(t)}{\partial x}\Big|_{x=0} \approx \begin{cases} N \exp(r_m t) & \text{for } t \leq T_e \\ K \exp(st) & \text{else.} \end{cases}$$

Let $T_s = T - T_e$ and assume that $T \gg T_e, \widetilde{T}_e$ where $\widetilde{T}_e$ is the length of the exponential phase for an individual infected only with the mutant strain. Then

$$\begin{aligned} \alpha = h'(0) &\approx e^{\tau} \frac{0 + K \frac{\exp(sT)-1}{s}}{0 + KT} = e^{\tau} \frac{\exp(sT) - 1}{sT} \\ &\approx e^{\tau}(1 + sT/2) \end{aligned}$$

as claimed in the main text.

Estimating the Probability of Emergence when $\alpha < 1$

When α is less than 1, the frequency of mutant virus decrease in an infected host, and consequently, even if the adapted virus may emerge, the probability of emergence is very low, and even lower when the bottleneck size, N, increases. Here, we provide an approximation for the emergence probability when $\alpha < 1$, which explains why the probability of emergence decreases dramatically when N increases.

When the outbreak starts, the first individual is infected with wild-type only. When $s < 0$, the mutation-selection balance can be reached relatively quickly, and for s negative enough, the proportion of mutant is small. So the probability to transmit at least one mutant is roughly equal to the probability to transmit one mutant, which is $N \exp(\tau) r_w \mu / |s|$ where $r_w \mu / |s|$ is the proportion of the mutant type, and $\exp(\tau)$ is its relative transmissibility. Then, if s is small enough, then the reproductive number of an individual with a mixed transmission is close to R_w of the wild-type. Thus, the number of transmissions in a wild-type outbreak can be used $(R_w/(1 - R_w))$. For an individual infected with a mixed infection, what will lead to emergence are the contacts for which only the mutant is transmitted. The number of such contacts is:

$$\beta \int_0^T (b_w V_w + b_m V_m) \frac{(b_m V_m)^N}{(b_w V_w + b_m V_m)^N} dt. \tag{S–9}$$

This can be re-written as:

$$\beta \int_0^T b_w V_w \frac{(\exp(\tau) V_m/V_w)^N}{(1+\exp(\tau)V_m/V_w)^{N-1}} dt. \tag{S–10}$$

As most cases of mixed infection will be cases started with a mix of one mutant and $N-1$ wild-type viral particles, $V_m/V_w = \exp(st)/(N-1)$. Thus previous expression is equal to:

$$\beta \int_0^T b_w \min\{K, (N-1)\exp(r_w t)\} \frac{(\exp(\tau+st)/(N-1))^N}{(1+\exp(\tau+st)/N-1)^{N-1}} dt. \tag{S–11}$$

Last, an individual infected with mutant viruses alone has to lead to a successful outbreak, which happens at approximately the same probability than in the case with no back mutations, with probability p_m. So overall, the approximation will be:

$$\begin{aligned} p_{emergence} = {} & \frac{R_w}{1-R_w} \frac{\mu \exp(\tau) r_w}{-s} p_m \beta b_w \frac{N}{N-1} \\ & \times \int_0^T \min\{K, (N-1)\exp(r_w t)\} \underbrace{\exp(\tau+st) \left(\frac{\exp(\tau+st)/(N-1)}{1+\exp(\tau+st)/(N-1)} \right)^{N-1}}_{=:f(t,N)} dt. \end{aligned} \tag{S–12}$$

Now we can ask, which parts of this expression depend on N? The mutant reproductive number $R_w = b_w \int_0^T \min\{K, N\exp(r_w t)\} dt$ is independent from N, because we have chosen b_w to keep R_w the same for all N values. Thus $b_w N/(N-1) \int_0^T \min\{K, (N-1)\exp(r_w t)\} dt$ is almost independent from N. Therefore, most of the dependence of $b_w N/(N-1) \int_0^T \min\{K, (N-1)\exp(r_w t)\} f(t,N) dt$ with N stems from the dependence of $f(t,N)$ with N. Since $a \mapsto a/(1+a)$ is an increasing function bounded above by 1 for positive a, the expression

$$\exp(\tau+st)/(N-1)/(1+\exp(\tau+st)/(N-1))$$

decreases when N increases. As $N \mapsto (a/(1+a))^{N-1}$ is a decreasing function of $N \geq 1$ for $a > 0$, we get that the probability of emergence decreases at least exponentially with the bottleneck size, as claimed in the main text. Fig S–3 illustrates that these approximations work especially when s is sufficiently negative.

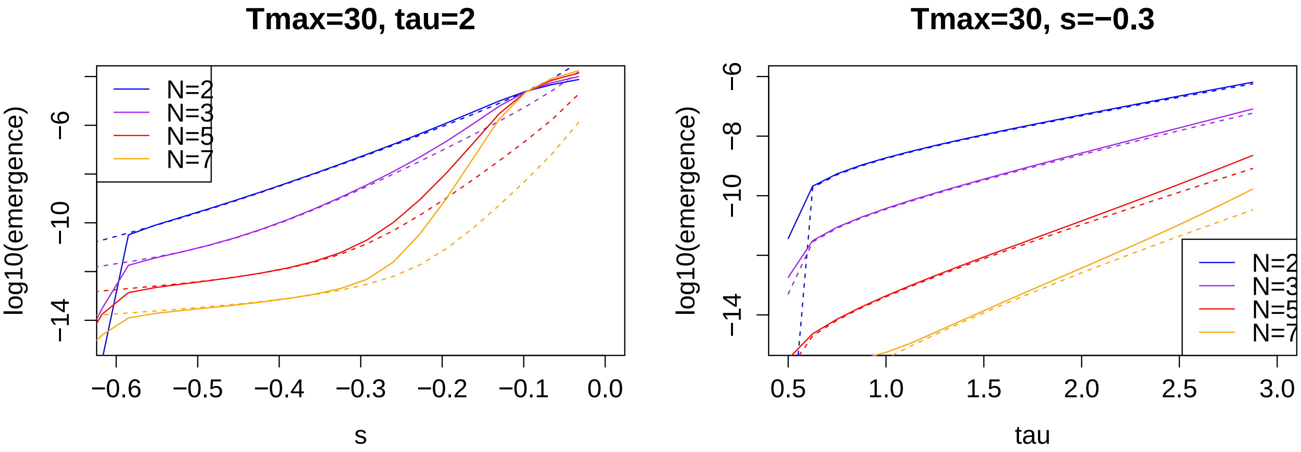


Figure S–3. Approximation (S–12) (dashed lines) vs. numerical resolution of the generating maps (R code) (solid lines)

Numerics with Nonlinear Transmission Functions

To explore the robustness of our numerical results to the assumption of a linear transmission function, we redid our numerical analysis with two non-linear transmission functions $p(x) = 1 - \exp(-x)$ in Fig S–4 and $p(x) = \log(1 + x)$ in Fig S–6. Differences between the emergence probabilities for the nonlinear and linear transmission functions are shown in Figs S–5 and S–7. As these figures demonstrate, we nearly get the same results.

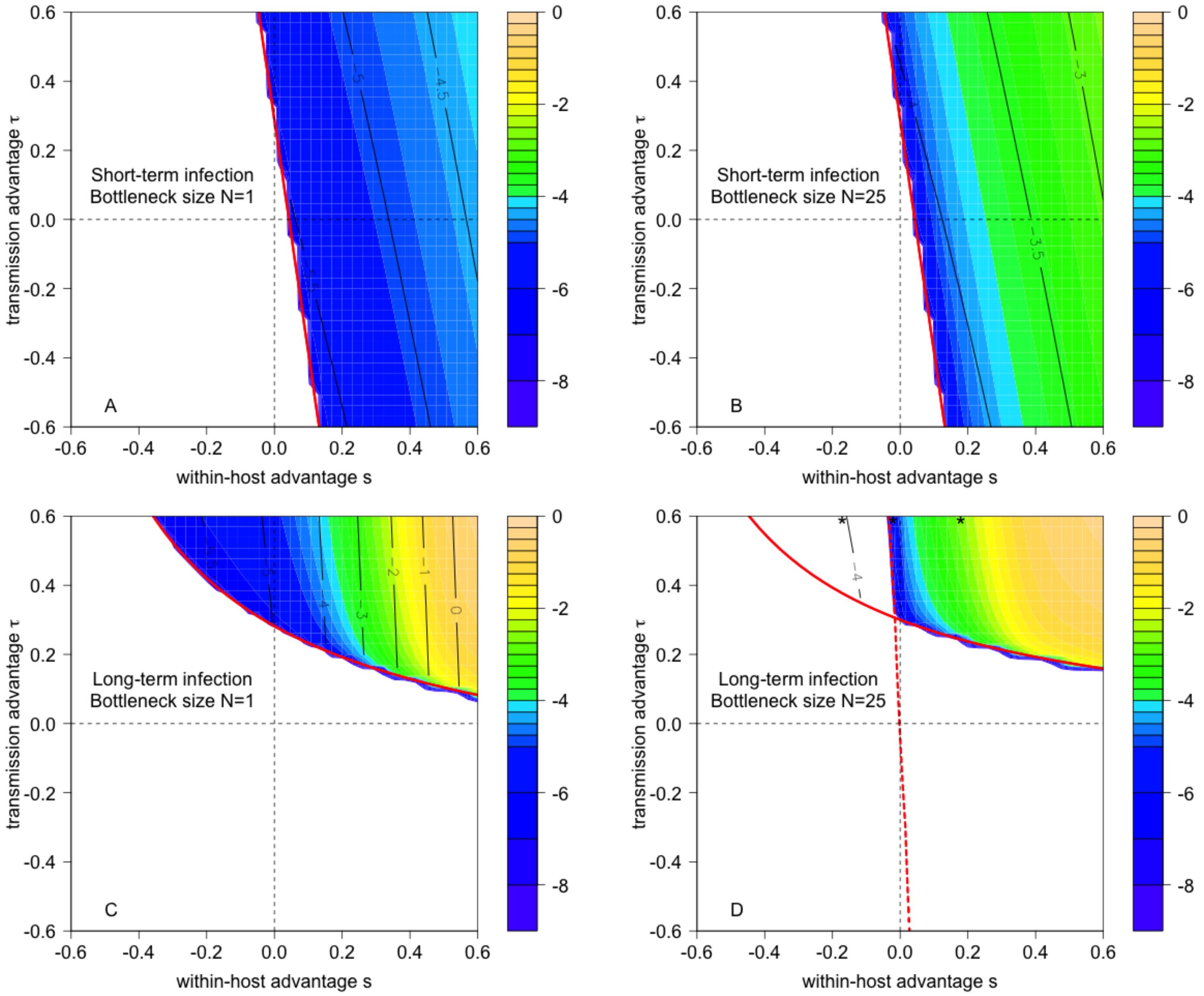


Figure S–4. Emergence probabilities for the transmission function $p(x) = 1 - \exp(-x)$ with all other parameters as indicated in Fig 2.

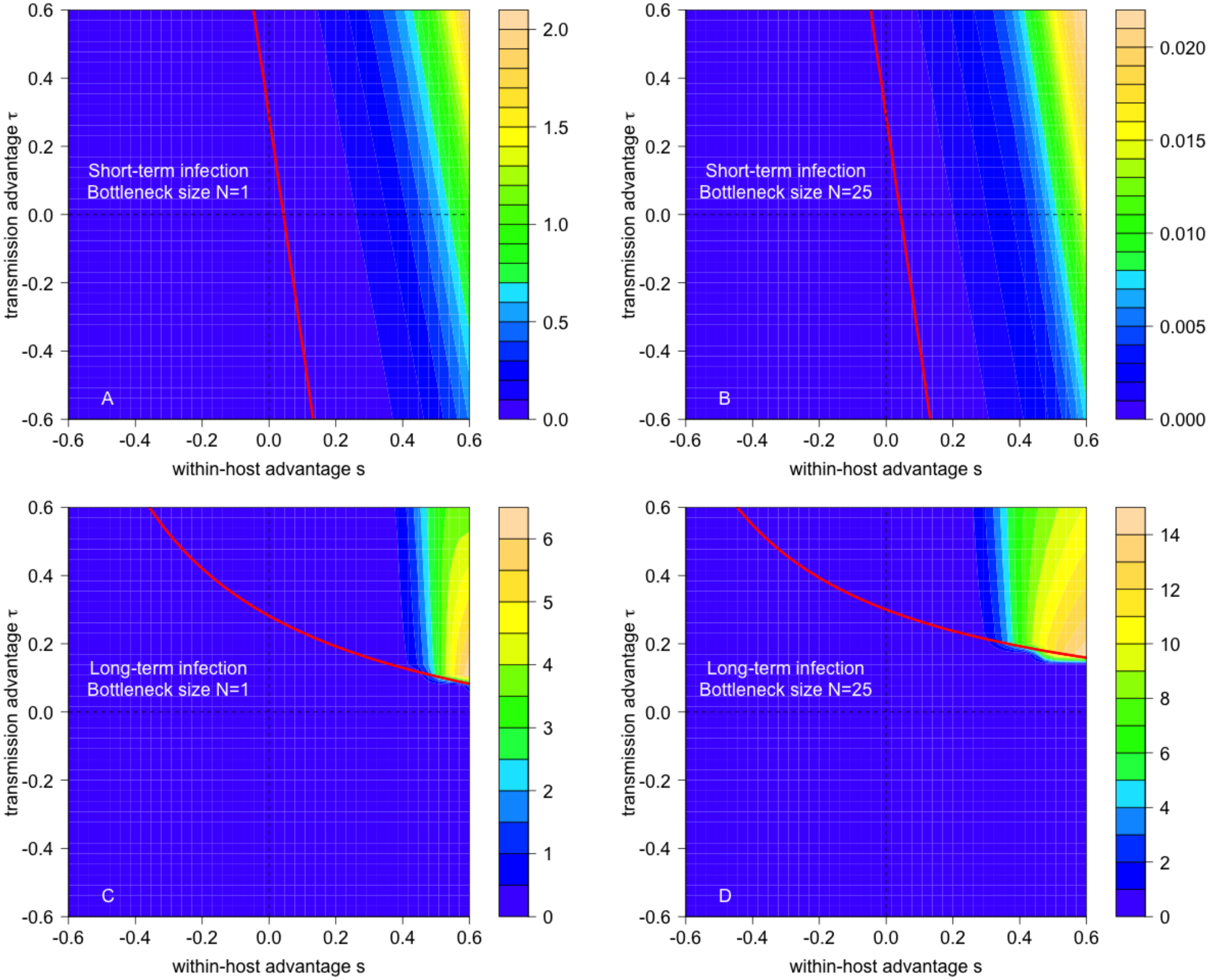


Figure S–5. Contour plots of $1,000 \times$ the absolute value of the difference between the emergence probabilities for the transmission functions $p(x) = 1 - \exp(-x)$ and $p(x) = x$. Parameters as indicated in Fig 2.

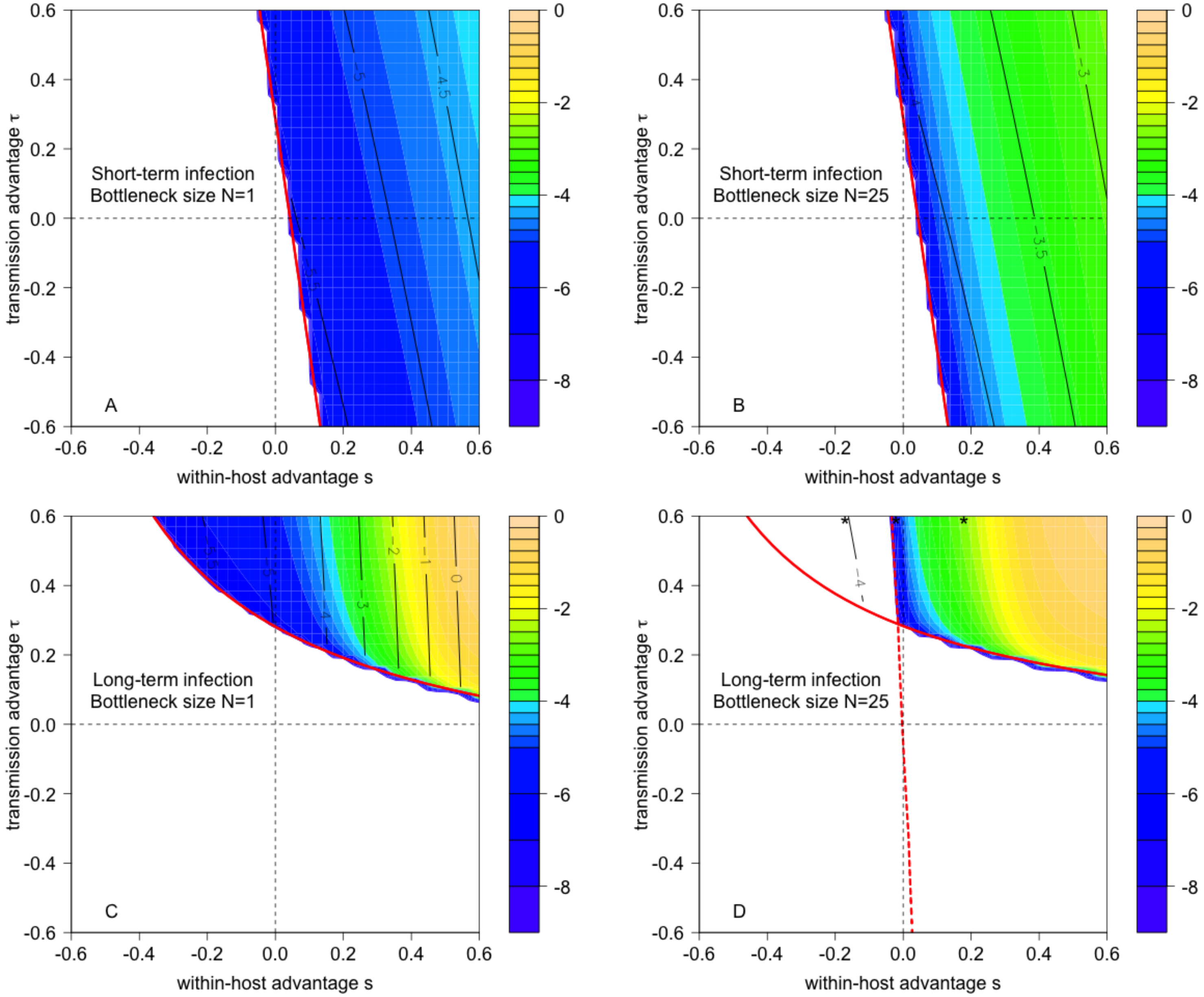


Figure S–6. Emergence probabilities for the transmission function $p(x) = \log(1 + x)$ with all other parameters as indicated in Fig 2.

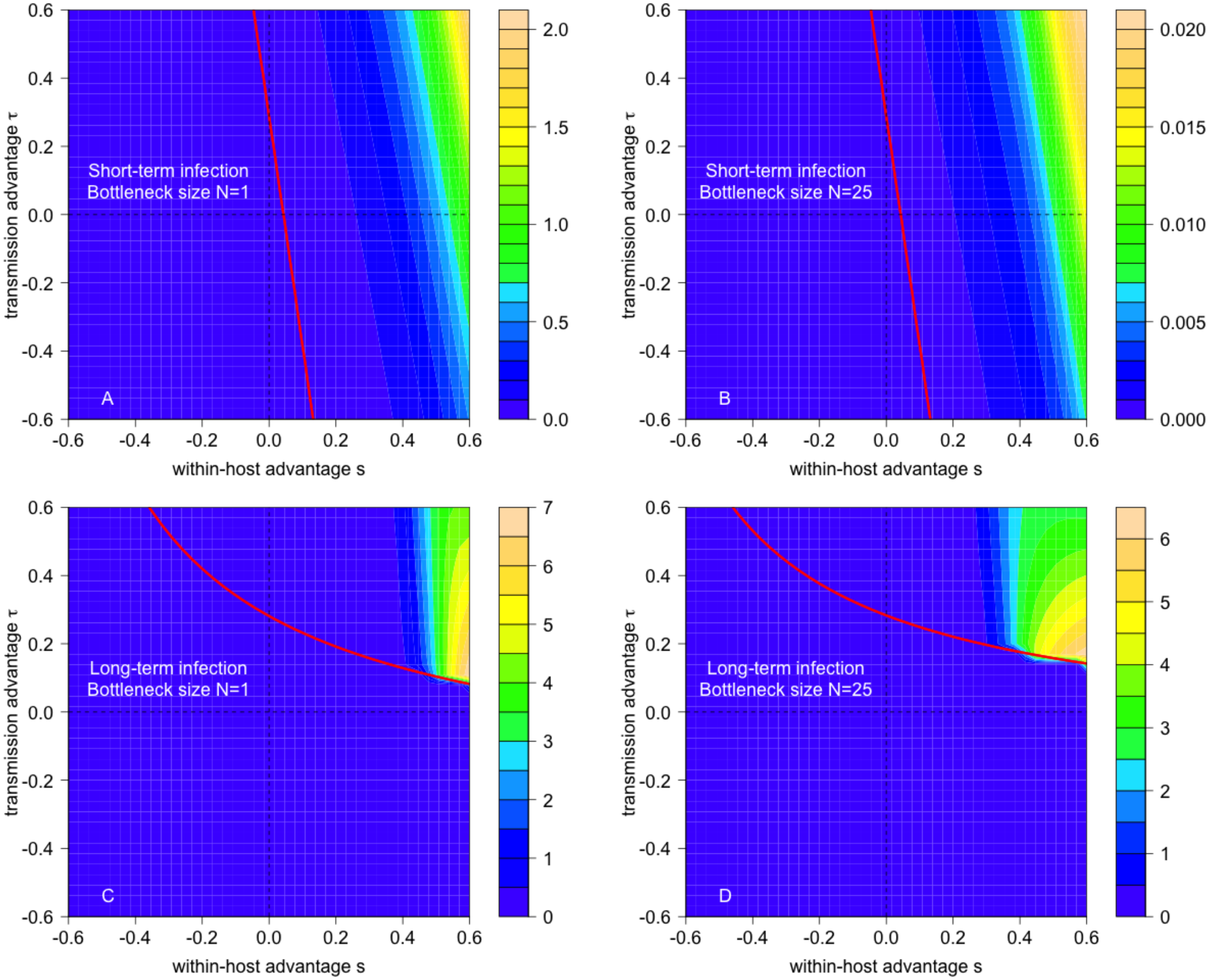


Figure S–7. Contour plots of $1,000 \times$ the absolute value of the difference between the emergence probabilities for the transmission functions $p(x) = \log(1 + x)$ and $p(x) = x$. Parameters as indicated in Fig 2.