Transmission cycles generate diversity in pathogenic viruses

Daniel Sigal[†,1,2], J.N.S. Reid[†,1] and L.M. Wahl[1,*]

[†]Contributed equally

[1]Applied Mathematics, Western University, London, Ontario, Canada N6A 5B7

[2]Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada N6A 5B7

[*]Corresponding author: lwahl@uwo.ca

Running title: Transmission cycle diversity

Abstract

We investigate the fate of *de novo* mutations that occur during the in-host replication of a pathogenic virus, predicting the probability that such mutations are passed on during disease transmission to a new host. Using influenza A virus as a model organism, we develop a life-history model of the within-host dynamics of the infection, using a multitype branching process with a coupled deterministic model to capture the population of available target cells. We quantify the fate of neutral mutations and mutations affecting five life-history traits: clearance, attachment, budding, cell death, and eclipse phase timing. Despite the severity of disease transmission bottlenecks, our results suggest that in a single transmission event, several mutations that appeared *de novo* in the donor are likely to be transmitted to the recipient. Even in the absence of a selective advantage for these mutations, the sustained growth phase inherent in each disease transmission cycle generates genetic diversity that is not eliminated during the transmission bottleneck.

Keywords: mutation, disease transmission, adaptation, influenza, life history

INTRODUCTION

Many pathogens experience population dynamics characterized by periods of rapid expansion, while a host is colonized, interleaved with extreme bottlenecks during transmission to new hosts. The effect of these transmission cycles on pathogen evolution has been well-studied, with particular focus on long-standing predictions regarding the evolution of virulence (reviewed in Alizon *et al.* 2009), conflicting pressures of within- and between-host fitness (Gilchrist and Sasaki 2002; Coombs *et al.* 2007; Day *et al.* 2011; see Mideo *et al.* 2008 for review), or broader factors affecting the evolutionary emergence of pathogenic strains (Antia *et al.* 2003; Iwasa *et al.* 2003; Reluga *et al.* 2007; Alexander and Day 2010; see Gandon *et al.* 2012 for review).

In the experimental evolution of microbial populations, the impact of population bottlenecks has also been studied in some depth, both theoretically (Bergstrom *et al.* 1999; Wahl and Gerrish 2001; Wahl *et al.* 2002) and experimentally (Burch and Chao 1999; Elena *et al.* 2001; Raynes *et al.* 2014; Lachapelle *et al.* 2015; Vogwill *et al.* 2016). While severe population bottlenecks clearly reduce genetic diversity, the period of growth between bottlenecks can have the reverse effect: generating substantial *de novo* adaptive mutations and promoting their survival (Wahl *et al.* 2002). The survival of a novel adaptive lineage is predicted to depend not only to the timing and severity of bottlenecks, but on the details of the microbial life history and the trait affected by the mutation (Alexander and Wahl 2008; Patwa and Wahl 2008; Wahl and Zhu 2015).

The effects of transmission bottlenecks on the evolution of an RNA virus have been explicitly studied in a series of experimental papers, demonstrating that severe bottlenecks (one surviving individual) reduced fitness (Duarte *et al.* 1992) despite rapid population expansion between transmission events (Duarte *et al.* 1993). The magnitude of this effect depends on both the initial fitness of the lineage (Novella *et al.* 1995) and on bottleneck severity (Novella *et al.* 1996). In theoretical work, a model of a viral quasispecies undergoing periodic transmission events predicts that

pathogens should maintain a mutation-selection balance with high virulence if the pathogen is horizontally transferred, if the bottleneck size is not too small, and if the number of generations between bottlenecks is large (Bergstrom *et al.* 1999).

Unlike the bottlenecks imposed in serial passaging, transmission bottlenecks in nature are not constrained by experimental control. Thus, key parameters such as the bottleneck size – the number of microbes initiating an infection – have proven difficult to estimate. Nonetheless experimental models (see Abel *et al.* 2015 for review), as well as recent techniques such as DNA barcoding (Varble *et al.* 2014) and sequencing of donor-recipient pairs in humans (Poon *et al.* 2016) have shed new light on this issue. In addition, we note that many human viruses – including human immunodeficiency virus, hepatitis B virus, and influenza A virus – reproduce by viral budding in the context of a potentially limited target cell population (Garoff *et al.* 1998); the survival of *de novo* mutations has not yet been predicted for this microbial life history. Thus, the effects of transmission bottlenecks on the genetic diversity of viral pathogens, that is, on the fate of *de novo* mutations, are as yet unknown.

In this contribution, we first develop a deterministic model of the within-host dynamics of early infection by a viral pathogen. We couple this to a detailed life-history model, using a branching process approach to follow the fate of specific *de novo* mutations that are either phenotypically neutral, or affect various life-history traits. These techniques allow us to predict which adaptive changes in virus life history are most likely to persist, and how the diversity of the viral sequence is predicted to change between donor and recipient. We can thus predict, for example, the rate at which *de novo* single nucleotide polymorphisms arise during the course of a single infection, and are transmitted to a subsequent host.

Throughout the paper, we will illustrate our results with parameters that have been chosen to model the life history and transmission dynamics of influenza A virus (IAV). IAV is an orthomyxovirus

(Bouvier and Palese 2008) that imposes a significant burden on global health, causing seasonal epidemics, sporadic pandemics, morbidity and mortality (Carrat and Flahault 2007). It is estimated that infection with seasonal strains of influenza results in around 36,000 deaths per year in the United States, although exact numbers are difficult to determine (Chowell *et al.* 2008).

Mathematical modelling is a well-established tool for predicting the evolution of influenza (Larson *et al.* 1976; Bocharov and Romanyukha 1994). Because of the critical importance of immune evasion in influenza, interest has focused on the adaptation of the virus in response to immune pressure, focusing on antigenic drift (Boianelli *et al.* 2015) and antigenic shift (Feng *et al.* 2011) in the global influenza pandemic (van de Sandt *et al.* 2012). Recent models, however, have specifically addressed the within-host dynamics of influenza A virus (Beauchemin *et al.* 2005; Baccam *et al.* 2006; Beauchemin and Handel 2011; Smith and Perelson 2011; Dobrovolny *et al.* 2013; Boianelli *et al.* 2015). In concert with these contributions, recent empirical work has elucidated the life history of the influenza A virus, providing quantitative estimates of parameters such as the minimum infectious dose (Varble *et al.* 2014; Poon *et al.* 2016), the size of the target cell population, and the kinetics of viral budding (Baccam *et al.* 2006; Beauchemin and Handel 2011; Pinilla *et al.* 2012). Although we now have an increasingly clear picture of the within-host life history of this important pathogen (Beauchemin and Handel 2011; Biggerstaff *et al.* 2014), estimates of the rate at which *de novo* mutations arise and are transmitted have not yet been available. Our approach allows direct access to this question.

LIFE HISTORY AND TRANSMISSION MODEL

Deterministic Model We use a system of ordinary differential equations (ODEs) to approximate the within-host dynamics during the early stages of infection by a pathogenic virus, assuming a life history that involves infection of a target cell, an eclipse phase, and finally an infectious stage.

Specifically, we propose:

$$\left.\begin{array}{lll} \text{target cells:} & \frac{dy_T}{dt} & = -\alpha y_T(t)v(t) \\ \text{infected (eclipse):} & \frac{dy_E}{dt} & = \alpha y_T(t)v(t) - (D+E)y_E(t) \\ \text{budding cells:} & \frac{dy_B}{dt} & = Ey_E(t) - Dy_B(t) \\ \text{free virus:} & \frac{dv}{dt} & = -Cv(t) + By_B(t) - \alpha y_T(t)v(t) \end{array}\right\}. \quad (1)$$

Here y_T represents susceptible target cells (in the case of influenza A virus we consider epithelial cells of the upper respiratory tract), y_E represents cells that are infected by the virus but not yet in the budding stage, y_B represents mature infected cells (infected cells that are budding), and v represents the free virus, that is, virions not attached to target cells (Baccam *et al.* 2006). Parameter B gives the rate at which budding cells produce infectious free virus; C gives the clearance rate for free virus. Infected cells die at constant rate D, while E represents the rate at which infected cells mature, leaving the eclipse phase and becoming budding cells. The parameter α gives the rate of attachment per available target cell. Thus the overall attachment rate for a virion is a function of the time-varying target cell population, and can be written $A(t) = \alpha y_T(t)$, with the corresponding mean attachment time, $A(t)^{-1}$.

A limitation of ODE approaches is that all transitions are described by exponential distributions. To relax this assumption, we introduce a sequence of k infected stages through which infected cells pass before reaching the budding stage. This 'chain of independent exponentials' allows for more realistic gamma distributions of eclipse times (Wahl and Zhu 2015). Specifically, we replace system (1) with:

$$\left.\begin{array}{lll} \text{target cells:} & \frac{dy_T}{dt} & = -\alpha y_T(t)v(t) \\ \text{eclipse stage 1:} & \frac{dy_1}{dt} & = \alpha y_T(t)v(t) - (D+kE)y_1(t) \\ \text{eclipse stage } 2...k & \frac{dy_j}{dt} & = kEy_{j-1}(t) - (D+kE)y_j(t) \quad j = 2...k \\ \text{budding:} & \frac{dy_B}{dt} & = kEy_k(t) - Dy_B(t) \\ \text{free virus:} & \frac{dv}{dt} & = -Cv(t) + By_B(t) - \alpha y_T(t)v(t) \end{array}\right\}. \quad (2)$$

When $k = 1$, this model reduces to System 1; for $k > 1$, y_1 gives the population of initially infected

cells, which pass through k eclipse stages at rate kE before budding. The transition rate kE is set such that the expected time in the eclipse phase, in total, is fixed at $1/E$ for any value of k. In the supplementary material, we also investigate a model in which the death term, D, is set to zero during the eclipse stages and only acts during the budding stage. This likewise gives a more realistic distribution for the lifetime of infected cells.

The founding virus begins as an initial population of free virus (the initial infectious dose, $v(0) = v_0$) at time $t = 0$. We do not assume that all viral particles in the founding dose are genetically identical, but we do assume that they are phenotypically identical, that is, they are described by the same parameter values in the deterministic model. As described further in the stochastic model below, we assume that disease transmission occurs at time τ during the peak viral shedding period (when the free virus population, v is at or near a peak value). For the transmission event to a new susceptible individual, a new founding population is sampled from the total viral load. In particular, each free viral particle becomes part of the infectious dose transmitted to the next individual with probability F. The value of F is computed such that for the founding virus, the expected size of the transmitted sample is v_0, that is, $F = v_0/v(\tau)$. Note that only free virions – those not yet attached to a target cell – are transferred to the next individual during transmission.

Immune responses clearly play a critical role in the within-host dynamics of viral infections, as well-documented for models of influenza A (Beauchemin and Handel 2011; Smith and Perelson 2011; Dobrovolny *et al.* 2013). In the model proposed above, innate immune mechanisms are included in the clearance rate of free virus and the death rate of infected cells. Because we use this model only until the time of peak viral shedding, which occurs 54.5 hours post infection (see parameter values, below) and before the adaptive immune response is activated (Tamura and Kurata 2004), we do not include the adaptive immune response. We address this issue further in the Discussion. Likewise, we do not include replenishment of the susceptible target cell population over the initial 54.5 hours of the infection. This is consistent with complete desquamation of the epithelium (loss of all ciliated cells) within three days post-infection in murine influenza, followed

by regeneration of the epithelial cells beginning five days post-infection (RAMPHAL *et al.* 1979).

Stochastic Life History Model To describe the lineage associated with a rare *de novo* mutation, a stochastic model is required. To gain tractability, we assume that the mutant lineage propagates in an environment for which the overall dynamics of the target cell population are driven by the deterministic system (2). Thus we treat the free virus, eclipse-phase cells and budding cells in the mutant lineage stochastically, but use the deterministic system to predict the susceptible target cell population at any time.

As in the deterministic model, free virions clear at a constant rate C or adsorb to susceptible host cells at rate $A(t)$. Note that the attachment rate of a free virion is not constant; it depends on target cell availability, such that $A(t) = \alpha y_T(t)$, where $y_T(t)$ is the target cell population predicted by system (2). Host cells enter the eclipse phase when a virion adsorbs, and exit the eclipse phase at rate E. After the eclipse phase, mature infected cells bud virions at rate B. Since budding itself does not immediately kill the host cells (GAROFF *et al.* 1998), after infection the cell is subject to a constant death rate D, or in other words the cell remains alive for an average time $1/D$.

This stochastic growth process can be described as a branching process, using a multitype probability generating function (pgf) to describe a single lineage of free virions (associated with dummy variable (x_1), infected cells (x_2), and mature cells (x_3). As derived in the Appendix, the pgf for this process, $G(t, x_1, x_2, x_3)$, satisfies:

$$\begin{aligned}\frac{\partial G}{\partial t} = &(A(t)x_2 + C - (A(t) + C)x_1)\frac{\partial G}{\partial x_1} \\ &+(-(E + D)x_2 + Ex_3 + D)\frac{\partial G}{\partial x_2} \\ &+(Bx_1x_3 + D - (D + B)x_3)\frac{\partial G}{\partial x_3}\end{aligned} \tag{3}$$

where $A(t)$, B, C, D and E are attachment, budding, clearance, cell death and eclipse maturation rates, respectively. Equation (3) captures the time evolution of the pgf, given each of these probabilistic events. As shown in the Appendix, Equation (3) can be converted to a system of ODEs using the standard method of characteristics, and is thus amenable to numerical solution. Analogous to System (2), Equation (3) can also be extended to include a chain of k infected stages before the budding stage, yielding more realistic distributions of the eclipse time.

To estimate the probability that the lineage associated with a *de novo* mutation does not survive the transmission bottleneck, we will need a pgf describing a complete cycle of in-host growth followed by a transmission bottleneck. We thus numerically integrate the pgf G, described above, from time 0 to time τ, and then compose it with a pgf describing disease transmission. To describe disease transmission, we simply assume that each free virion in the infected host is transmitted with fixed probability F, as described above. As derived in the appendix, this approach allows us to estimate the probability that a *de novo* mutation that first occurs at time t_0 is transmitted to the next host, $1 - X(t_0)$, the rate at which such "surviving" mutations arise at each time during the infection, $\nu(t_0)$, and, ultimately, the probability that a given mutation occurs *de novo* and is transmitted to the next host, $\mathcal{P}$.

Beneficial Mutations Our goal is to predict the fate of mutations that may arise *de novo* in the viral population. Although most mutations will be deleterious, we note that the virus population grows by several orders of magnitude (possibly up to seven) during a single infection, and thus deleterious mutations should be effectively purged by selection. We therefore focus in this contribution on neutral mutations (no phenotypic effect), or rare mutations that confer an adaptive advantage to the virus. For a budding virus, changes in five life history traits can confer a selective advantage: a reduction in either the cell death rate, $\tilde{D} = D - \Delta_D$, or clearance rate, $\tilde{C} = C - \Delta_C$; an increase in the attachment rate, $\tilde{\alpha} = \alpha + \Delta_\alpha$, or budding rate, $\tilde{B} = B + \Delta_B$; or an increase in the rate at which cells mature and begin budding, $\tilde{E} = E + \Delta_E$.

To estimate the probability that a beneficial mutation ultimately survives, we substitute the parameters above for the analogous parameters in the pgf $G(t, x_1, x_2, x_3)$ and numerically evaluate $G(\tau, x_1, 1, 1)$, which describes the distribution of free virions in the mutant lineage at time τ, as described in the Appendix. We then compose this function with the pgf describing disease transmission. The accuracy of these numerical solutions was verified using an individual-based Monte Carlo simulation, developed for a reduced model without target cell limitation, similar to the approach described by Patwa and Wahl (2009).

Selective Advantage Finally, in order to compare the fitness of mutations affecting different traits, we calculate the selective advantage of each mutation. Following common experimental practice, we define fitness in terms of the doubling time, that is, we assume that in the time required for the founding population to double, the mutant lineage grows by a factor of $2(1+s)$. Given the founding growth rate g, we substitute the founding doubling $t = \ln(2)/g$ into $2(1+s) = \exp(\tilde{g}t)$ to find the selective advantage of the mutant, $s = 2^{\bar{s}} - 1$, where $\bar{s} = \frac{\tilde{g}}{g} - 1$. (For the relatively small s values presented here, this definition of the selective advantage differs from the more appropriate but less commonly used $\bar{s}$ by a constant factor of $\ln 2$.)

To estimate the average growth rates, g and $\tilde{g}$, we consider a single cycle of growth, starting from a single free virus at time 0. In this case the partial derivative of G with respect to x_1, defined as $Z = \partial G(\tau, 1, 1, 1)/\partial x_1$, gives the expected number of free virions at time τ, illustrated here for the case $k = 1$ (Grimmett and Welsh 2014). The derivative was calculated numerically, and the average exponential growth rate of the free virus population is then given by $g = \ln Z/\tau$.

Parameter values for influenza A virus Parameter values were estimated where possible from the empirical and clinical literature for influenza A virus, and are displayed in Table 1. Beauchemin and Handel 2011 give a range of values for several relevant parameters, from which parameter

estimates for C, D, and E were chosen. Specifically, we take the clearance time to be 3 hours, the cell death time 25 hours, and the eclipse time 6 hours (Baccam *et al.* 2006; Beauchemin and Handel 2011).

Table 1: Parameter Estimates for Influenza A Virus

Parameter	Definition	Estimate
α	per target cell attachment rate	$\frac{2.375\times10^{-9}}{\text{hour cell}}$
$1/B$	mean time between each budding event	$\frac{19 \text{ hours}}{200 \text{ infectious virions}}$
$1/C$	mean clearance time	3 hours
$1/D$	mean cell death time	25 hours
$1/E$	mean eclipse time	6 hours
$y_T(0)$	initial number of target cells	4×10^8
$v(0) = v_0$	number of virions to initiate infection	100
k	stages in eclipse phase	30
μ	mutation rate (per site per replication)	6.7×10^{-7}

To estimate the time between each budding event, $1/B$, we first consider the total number of virions produced per cell, the "burst size". For influenza A virus, the burst size has been estimated to be between 1000-10000 virions (Stray and Air 2001). However, not all virions produced are infectious and in fact a large fraction are unable to infect a host cell; the particle to infectivity ratio for influenza A is approximately 50:1 (Martin and Helenius 1991; Roy *et al.* 2000). Taking the upper bound of the range for burst size, of the 10000 virions produced only 200 are predicted to be infectious. Recall that budding does not kill the host cell, therefore budding time depends on the eclipse and cell death times. An eclipse time of 6 hours and a cell death time of 25 hours gives a budding time of 19 hours. Therefore, the time between each infectious budding event, $1/B$, is assumed to be assumed to be 19/200 hours per infectious virion.

The number of upper respiratory epithelial cells in a healthy adult is estimated to be 4×10^8

(Baccam *et al.* 2006). Consistent with the complete desquamation of the epithelium observed in murine influenza (Ramphal *et al.* 1979), we therefore take $y_T(0) = 4 \times 10^8$. In the supplementary material we investigate the sensitivity of our main results to this value. Similarly, we assume that an infection is founded by $v_0 = 100$ virions, consistent with recent sequencing of donor-recipient pairs (Poon *et al.* 2016). However since values of 10-200 have been suggested in the literature (McCaw *et al.* 2011; Varble *et al.* 2014; Peck *et al.* 2015), we will also demonstrate results over a range of v_0 values.

To allow for realistically distributed eclipse times, we assume a gamma-distributed eclipse phase by including a sequence of k infected stages before the budding stage. As described above, the mean eclipse time, $1/E$, is set to 6 hours. The variance of the eclipse period of influenza A can then be used to estimate k. Pinilla *et al.* (2012) used a best-fit analysis for kinetic parameters of influenza A to predict a mean eclipse time of 6.6 hours, with an eclipse period standard deviation, σ, of 1.2 hours. Since the standard deviation for a gamma distribution with mean m is given by $\sigma = m/\sqrt{k}$, these values suggest that a realistic value of k is approximately 30.

We fix the attachment rate, α, such that that the peak of the free viral load occurs within the reported range for influenza A of 48 to 72 hours post-infection (Wright *et al.* 2001; Lau *et al.* 2010). The attachment rate $\alpha = 2.375 \times 10^{-9}$ per hour per cell provided in Table 1 yields a peak time of $\tau = 54.5$ hours, and implies a mean attachment time, $1/A(0)$, of just over one hour when target cells are plentiful. We assume that disease transmission is most likely at the peak viral shedding time, and thus study a transmission event that occurs at this peak time, τ. Note that when we examine the sensitivity of the model, for example when changing v_0, we leave the attachment rate α fixed. We recompute the time course $v(t)$ and assume that the transmission event occurs at the peak value of $v(t)$. The transmission time, τ, then differs slightly between cases. In no case was τ outside the empirically estimated range of 48-72 hours.

The probability that each free virion survives the bottleneck and is transmitted to the next susceptible individual is defined as F. This probability is calculated by using the peak number of free virions, $v(\tau)$, found by numerically solving model 2. As only free virions contribute to the infectious dose, the fraction of free virions surviving the bottleneck is $F = v_0/v(\tau)$, where again v_0 is the founding population size for the next infected individual.

The mutation rate for influenza A, per nucleotide per replication, has been estimated as $\mu = 2\times10^{-6}$ (Nobusawa and Sato 2006). This estimate was obtained for the IAV nonstructural gene during plaque growth, and thus does not include lethal mutations. Neglecting differences in transition and transversion rates, we divide this value by three to estimate the rate at which a specific, non-lethal nucleotide substitution occurs. We investigate the sensitivity of our results to this parameter as well.

Data Availability The authors affirm that all data necessary for confirming the conclusions of this article are represented fully within the article and its tables and figures.

RESULTS

Figure 1 illustrates the deterministic dynamics of System 2, showing the time course of the in-host influenza A infection. The free virus peaks at 54.5 hours, just after the peak in the mature (budding) cell population. Note that in this simplified model, the availability of target cells limits the infection. As described earlier, this model is only accurate while the adaptive immune response remains negligible; although we illustrate the full seven days of infection, we use only the first 54.5 hours in the subsequent analysis.

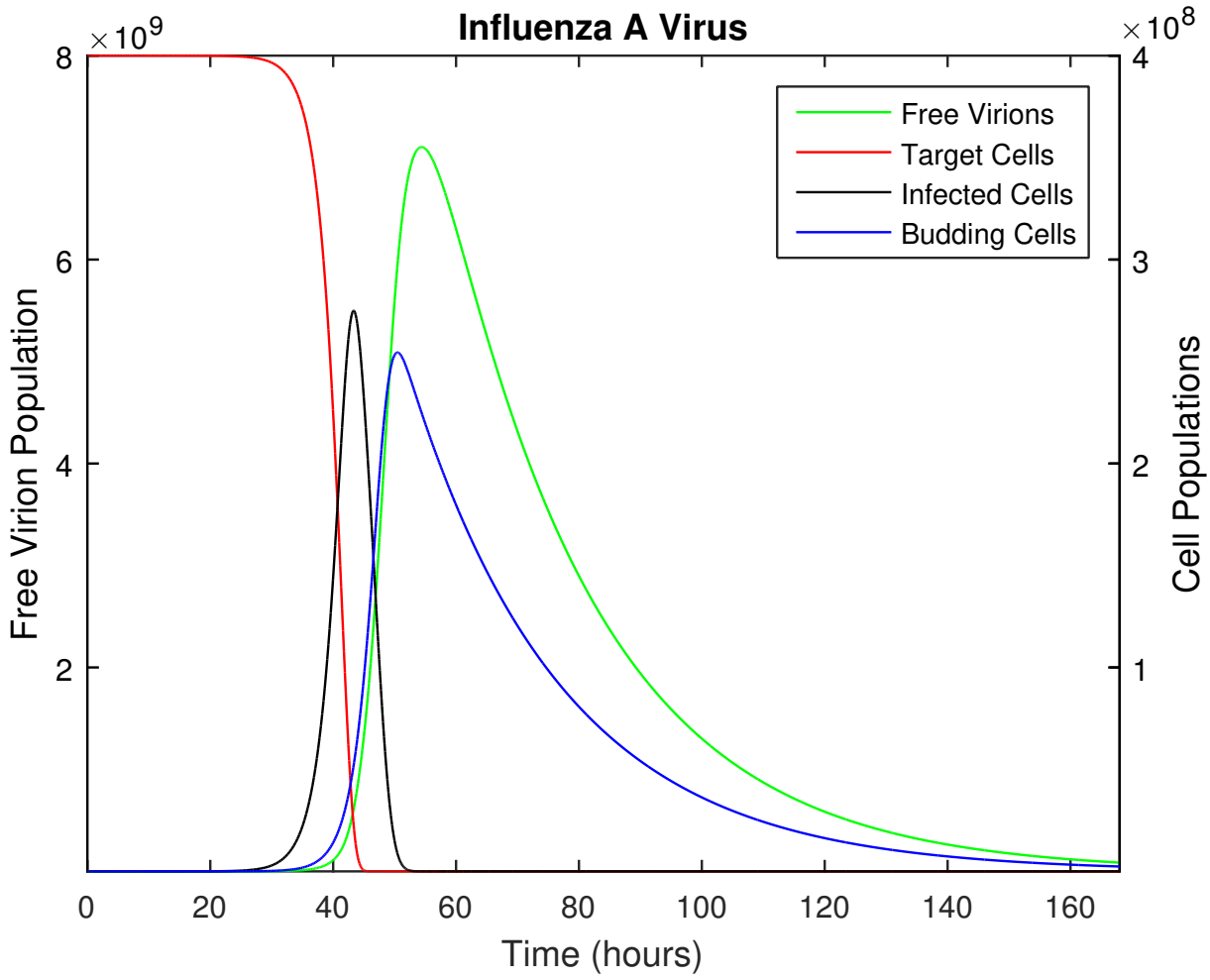

Figure 1: The time course of influenza A infection over the span of a week (168 hours). Parameter values are provided in Table 1, with the following initial conditions: 4×10^8 epithelial cells (target cells), 100 virions (initial infection dose), all other populations initially zero.

Figure 2 shows what we will refer to as the *mutation transmission rate*, that is, the probability that at least one copy of a specific mutation arises *de novo* during an infection time course, survives genetic drift and is successfully transmitted to the subsequent host. Model predictions for beneficial mutations affecting each life history trait are shown versus the selective coefficient, s; the intercept at $s = 0$ shows the prediction for neutral mutations. Here we have assumed for comparison that the baseline mutation rate is equal for all types of mutation, however the y-axis in Figure 2 scales approximately linearly with μ. In the Supplementary Material, we illustrate results for a wide range of mutation rates.

To interpret these results, the empirical mutation rate must be carefully considered. The rate estimate we use reflects the probability, per replication, that a specific substitution occurs at a specific nucleotide in the influenza A sequence, given that the substitution is non-lethal. Thus for example if the substitution of interest is neutral or effectively neutral, the model predicts that this substitution would occur *de novo* in the donor and be transmitted to a recipient about once in every 2000 transmission events. If the substitution of interest confers a selective advantage, the mutation transmission rate would be higher. Clearly, a large fraction of viable mutations will be deleterious and would be outcompeted before transmission; this would correspond to a lower overall mutation rate as examined in the Supplementary Material and outlined further in the Discussion.

The most striking result of Figure 2 is the predicted evolvability of influenza A during a single transmission cycle. The mutation transmission rate of one in two thousand, per substitution per site, may contribute substantial diversity since the influenza A genome is a sequence of over 13,000 nucleotides with three possible substitutions per site. We will return to the interpretation and implications of this prediction in the Discussion.

The near-overlapping lines in Figure 2 indicate that the mutation transmission rate does not vary widely across life history traits, and also illustrates the maximum selective advantage made possible by improvements to each trait. For example, clearance and cell death rates can only be reduced to zero, limiting the range of s for these traits. Although there is no upper bound on the rates of attachment or maturation to budding (eclipse rate), once these rates are effectively instantaneous, further increases do not appreciably change the growth rate, and so higher s values are also inaccessible for these traits. Similarly, increases to the budding rate cannot improve the growth rate without bound, due to target cell limitation.

Results in Figure 2 assume the default parameter set (Table 1); in particular, 100 virions are chosen at random from the free virus population and transmitted to the new host. In Figure 3, we fix the

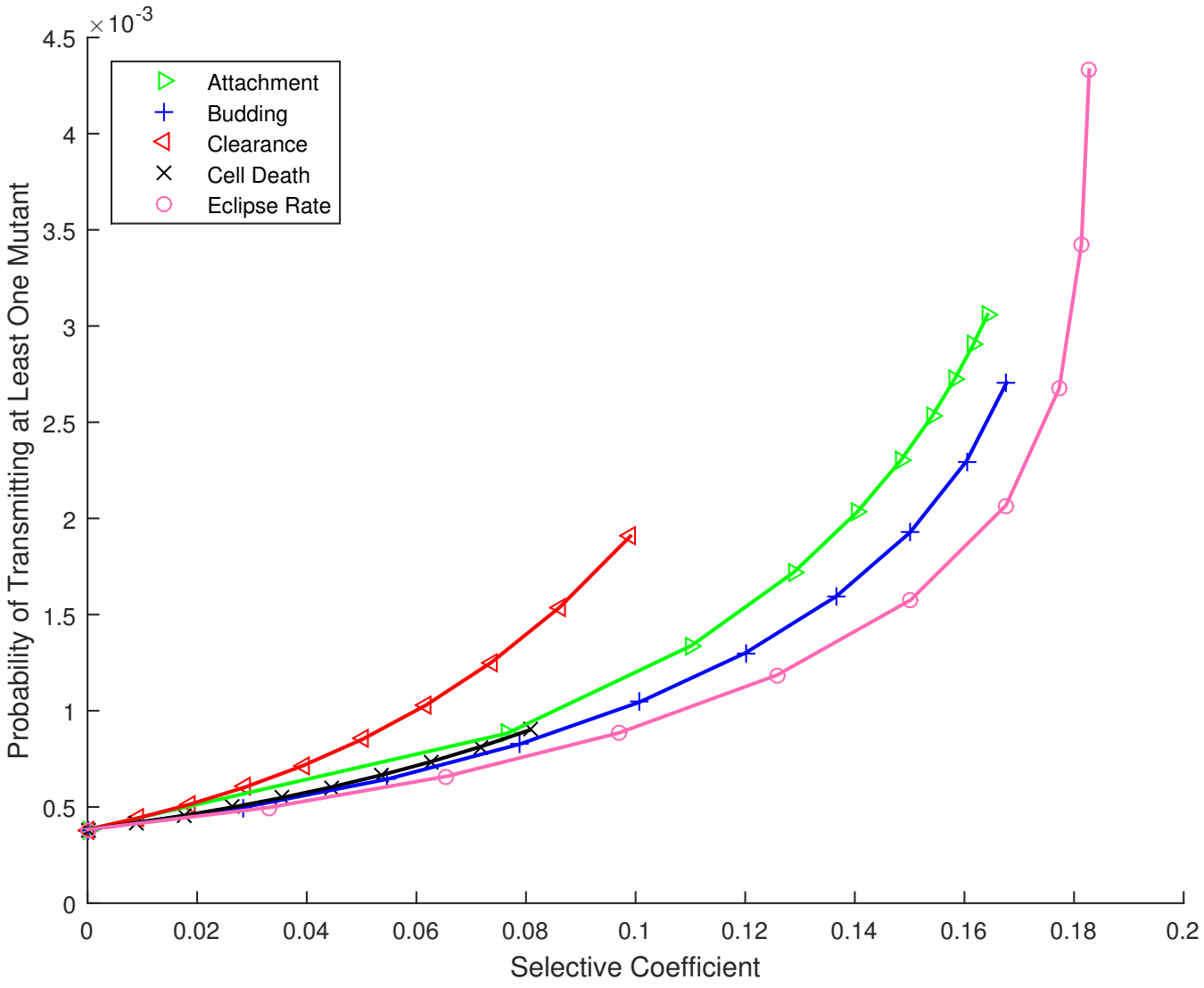

Figure 2: Probability that at least one *de novo* mutation arises during the infection time course and is passed to the next host, for mutations affecting the life history of influenza A virus, versus their selective coefficient. Parameters as provided in Table 1.

selective coefficient ($s = 0.05$) but vary the size of this transmission bottleneck. We find that the mutation transmission rate increases roughly linearly with bottleneck size.

The results above compare mutations that have equivalent effects on the overall growth rate of the virus, assuming that the underlying mutation rate is the same for all mutations. Although the question of mutational accessibility is beyond our focus, some sense of the degree to which these mutations might be physiologically achievable can be obtained by considering the relative changes

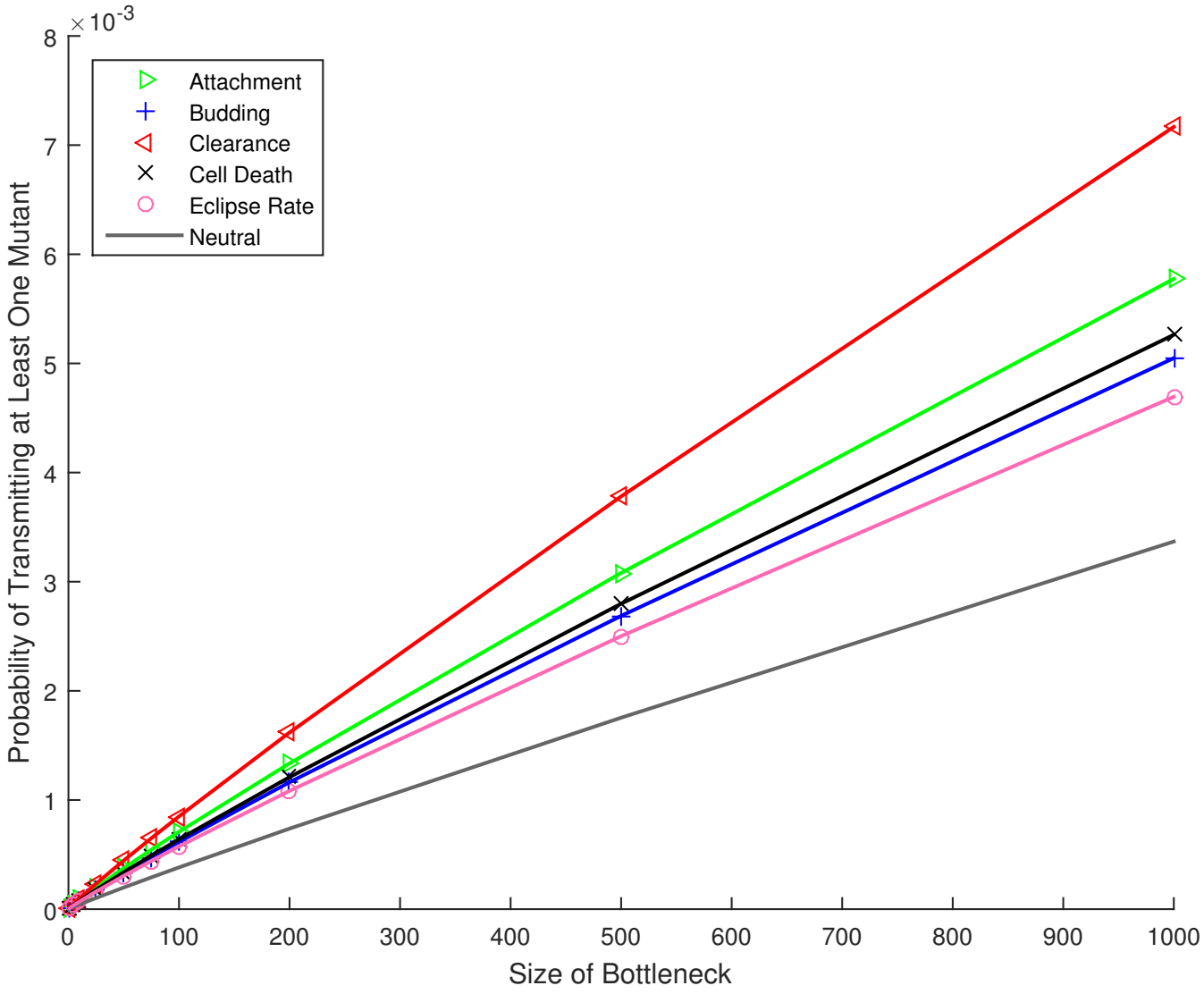

Figure 3: Probability that at least one *de novo* mutation arises during the infection time course and is passed to the next host, for mutations affecting the life history of influenza A virus, versus the number of virions in the transmission bottleneck. All mutations have a selective advantage of $s = 0.05$, except for the curve marked "neutral", for which $s = 0$. Other parameters as provided in Table 1.

required to the trait value. To this end, Figure 4 shows the relative change in each life history parameter necessary to achieve a specific increase in growth rate (selective coefficient). To incur an advantage of $s = 0.08$, for example, requires less than a 10% change in the rate at which cells leave the eclipse phase and begin budding; in contrast the attachment rate would need to double (change by over 100%) to achieve the same selective advantage. Note again that clearance and cell death rates can only be reduced by at most 100%, limiting the range of their possible effects. For the

other three traits, as described previously, beneficial mutations can produce selection coefficients in the approximate range $0 < s < 0.2$, but further rate increases produce diminishing returns and fitness saturates.

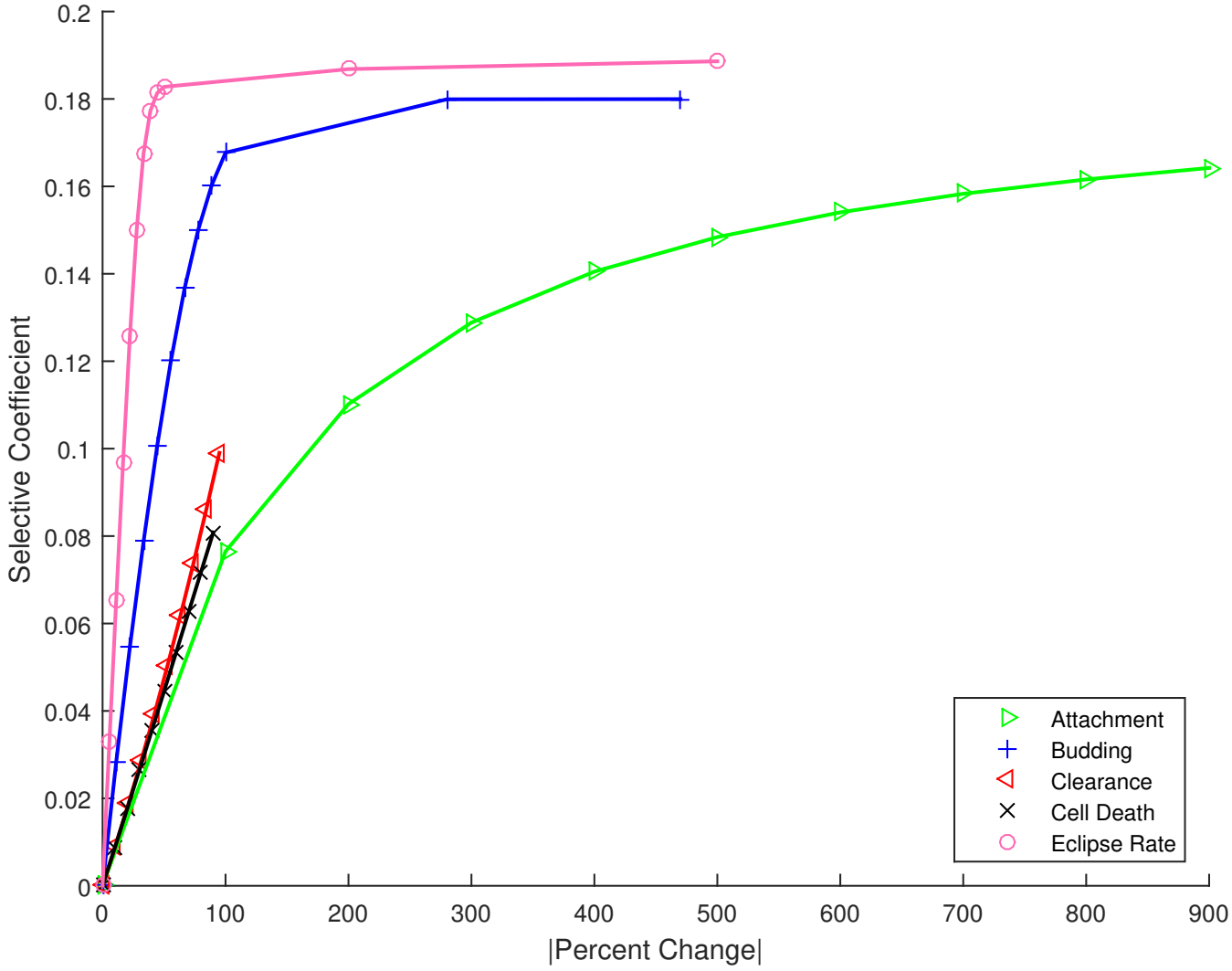

Figure 4: The change in selective coefficient achieved by a given absolute percent change in trait value, for mutations affecting the five life-history traits. For example, large changes in attachment rate would be required to achieve the same advantage as relatively small changes in eclipse timing.

Figure 2 gives the overall probability that a *de novo* mutation is generated and passed on. As described in the Methods, this value reflects the integrated probability of occurrence and survival for mutations that could first occur at any time during the infection time course. To better understand the dynamics of this process, in Figure 5 we show the predicted survival probability, the probability that the mutation survives and is transmitted to the next host, for mutations that arise at time

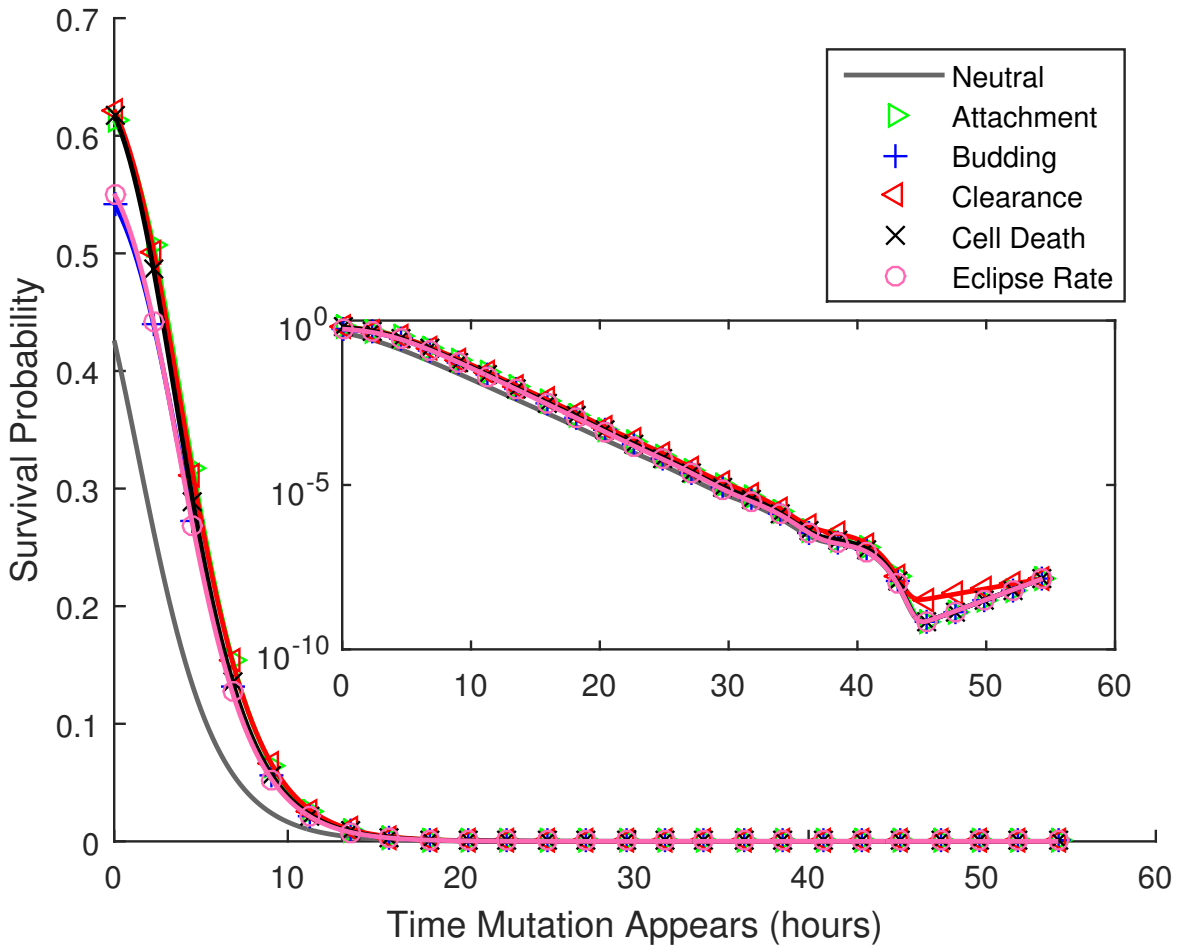

Figure 5: Given that a *de novo* mutation first occurs at time t_0 after the start of the infection, the probability that at least one copy of it is transmitted to the next host, versus t_0. All mutations have a selective advantage of $s = 0.05$, except for the curve marked "neutral", for which $s = 0$. Parameters as provided in Table 1. The inset shows the same results with in a semilog plot.

t_0 during the infection time course. Despite the fact that transmission to the next host occurs at 54.5 hours, the figure gives the impression that mutations that arise after about the first 10 hours of infection have little chance of survival.

The results in Figure 5 are mitigated, however, by the fact that many more replication events occur later during the growth phase. To investigate the rate at which surviving mutations (mutations that are transfered to the next host) first occur, we consider the product of the transmission probability for mutations that arise at each time and the number of new virions produced at that time, $By_B(t_0)$.

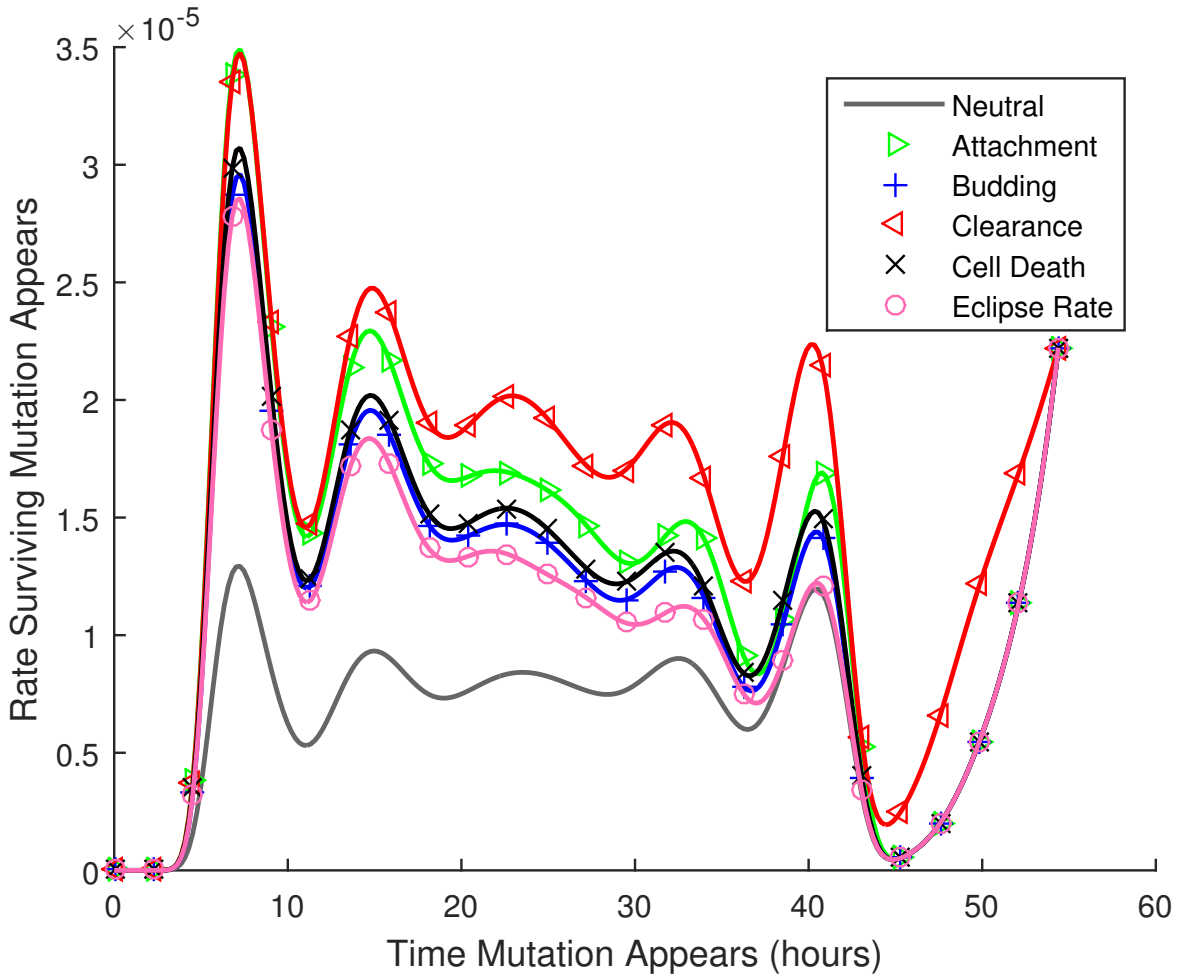

Figure 6: The rate at which transmitted mutations appear, versus the time at which they first appear. For this figure, the probabilities of being passed on to the next host illustrated in Figure 5 are multiplied by the number of new virions produced at each time, $\mu B y_B(t_0)$ from Equation 2. Thus the figure illustrates the relative numbers of ultimately transmitted mutations that occur at each time during the infection time course.

Figure 6 shows these results. The model predicts that transmitted mutations occur throughout the infection time course, except during the first few hours of infection, when very few new virions are produced, and for a brief window approximately 10 hours before the transmission event. The latter effect presumably occurs because virions produced in this window are unlikely to be free at the time of transmission (infected cells are not transmitted). The oscillations in these curves occur because the founder virions start syncronously at $t = 0$ as free virions, and must attach and complete the eclipse phase before new virions can be produced.

DISCUSSION

We develop a model of within-host pathogen evolution, and use this to predict the fate of *de novo* mutations that occur during disease transmission cycles. Using parameter values specific to influenza A virus and an estimate of the non-lethal mutation rate for IAV, our results predict that the probability that at least one copy of a *de novo* nucleotide substitution is transmitted to the subsequent host is about 5×10^{-4} per substitution per site, assuming the mutation is either neutral or beneficial. Multiplying by three possible nucleotide changes and the $\approx$ 13,600 sites in the IAV genome yields an estimate that as many as 20 sites in the founding dose for the recipient may contain substitutions that occurred *de novo* in the donor. This upper bound, however, must be corrected by the fraction of non-lethal mutations that are either neutral or beneficial. If approximately half of all non-lethal mutations are neutral or beneficial, as reported for another single-stranded RNA animal virus with a similar genome size (Sanjuán *et al.* 2004), we predict each recipient founding dose will contain about ten *de novo* substitutions. If the fraction of neutral or beneficial mutations, among non-lethal mutations, is closer to 10% (see Eyre-Walker and Keightley (2007) for review), we predict about two new substitutions per transmission. As demonstrated in Figures S2 and S7, these estimates scale directly with the underlying mutation rate and the size of the transmission bottleneck. Despite this inherent uncertainty, our results predict that a small handful of mutations occurring *de novo* in the donor will be transmitted to each recipient of IAV.

Our approach makes the simplifying assumption that the founding infectious dose in the donor is phenotypically, but not genetically uniform. Thus the predicted *de novo* mutations may occur on different genetic backgrounds circulating within the donor. Recent evidence suggests that multiple lineages are transmitted between donor-recipient pairs in IAV (Poon *et al.* 2016), and it seems unlikely that all transmitted lineages would be phenotypically identical. Thus a clear direction for future work would be to expand our approach to track multiple distinct lineages within the host, and predict the fates of mutations occurring on these backgrounds.

We can also take our estimate of (1.5×10^{-3} non-lethal substitutions per site per transmission event)×(10-50% neutral or beneficial) to predict $1.5 - 7.5 \times 10^{-4}$ substitutions per site per transmission event. These values are consistent with the observed evolutionary rate of IAV throughout a seasonal epidemic, 2×10^{-3} substitutions per site per year in the nonstructural (NS) gene (Kawaoka *et al.* 1998), if the chain of influenza transmission typically involves 3 to 13 transmission events per season.

Although transmission bottlenecks in IAV, as in many other pathogens, can be extremely severe, our results are consistent with previous work demonstrating that the period of growth between population bottlenecks has an even greater impact (Wahl *et al.* 2002); this period of sustained population expansion promotes the survival of new mutations, as seen more generally in any growing population (Otto and Whitlock 1997). The rapid growth of influenza during early infection, from a relatively small infectious dose to peak viral loads many orders of magnitude larger, implies that neutral substitutions, or mutations conferring even a small benefit, will have ample opportunity to compete with founder strains. This further implies that the life history of influenza A should be well adapted to the disease transmission cycle in humans, in other words, selection has the opportunity to rapidly fine-tune the life histories of pathogens experiencing extreme transmission bottlenecks.

This result is consistent with previous theoretical (Bergstrom *et al.* 1999) and experimental work on viral evolution (Duarte *et al.* 1992; Duarte *et al.* 1993; Novella *et al.* 1995; Novella *et al.* 1996). The latter work focused on the loss of fitness due to population bottlenecks, but fitness could be maintained or improved when the bottleneck size was as large as five or ten individuals (Novella *et al.* 1996). Similarly, Bergstrom *et al.* 1999 predicted that viral pathogens would be well-adapted if the bottleneck size is large (or order five or ten), and the number of generations between bottlenecks is large (of order 25 or 50). The parameter values we explored for influenza A correspond to over 25 population doublings between transmission events, with bottleneck sizes of 10 to 200, and are thus consistent with a parameter regime in which the pathogen is able to

improve or maintain fitness.

The use of a specific life-history model imposes natural limits on the growth rate and thus the selective advantage that can be achieved by budding viruses. For the parameters specific to influenza A, changes to the clearance rate of the free virus or death rate of infected cells could only achieve a selective advantage of $s < 0.1$. This occurs mathematically because these rates cannot be reduced below zero; it follows intuitively because even if infected cells or virus never die or lose infectivity, growth remains limited by other processes. Mutations with larger beneficial effects, in the range $0.1 < s < 0.2$, are accessible only by reducing the eclipse phase, or through very large magnitude changes to the attachment or budding rates. Given that predicted differences in survival probability for the different traits are rather modest (Figure 5), these results suggest that small magnitude changes in the eclipse timing of influenza A will be subject to selective pressure. The limits we observe in the achievable growth rate suggest that larger effect beneficial mutations in influenza A are not only unlikely, they may not be physically possible given the life history of this virus.

We have focused this study on the in-host life history of the virus. In principle, however, a beneficial mutation could also affect the transmissibility of the lineage (parameter F), producing virions that are preferentially transferred to a new host (Handel and Bennett 2008). This would be distinct from mutations that increase viral load; mutations affecting F would increase the probability that an individual viral particle is transmitted, for example by prolonging the stability of the virion in the external environment.

These results explore mutations affecting a single trait in isolation. Clearly higher fitness could be achieved by mutations that affect several traits, if beneficial pleiotropic mutations are available. Previous work suggests that the survival probability of pleiotropic mutations typically falls between the predictions obtained for single-trait mutations of equivalent selective effect (Wahl and Zhu 2015). In addition, we have investigated the transmission of *de novo* mutations when rare. Given the

magnitude of the viral loads measured in influenza A, it is clear that multiple beneficial mutations could emerge and compete before the virus is transmitted to a new host. Thus we would expect that clonal interference and multiple mutation dynamics might come into play in describing the adaptive trajectory more fully (Desai and Fisher 2007; Desai *et al.* 2007).

A limitation of the model is that the immune response is not explicitly included as a dynamic variable. Innate immunity is activated when an infection is detected, which is usually within the first few hours of infection. Adaptive immunity, however, is activated, at the earliest, three days post-infection (Tamura and Kurata 2004). Since our model addresses early infection (up to 54.5 hours post-infection) adaptive immune effects are assumed negligible. The innate immune response, however, cannot be neglected, as its main purpose is to limit viral replication (van de Sandt *et al.* 2012). In our approach, innate immune mechanisms are included in the viral clearance and infected cell death rates, but are assumed to be constant throughout this early stage of the infection. This phenomenon has been reviewed in some detail in previous work (Smith and Perelson 2011; Boianelli *et al.* 2015; Baccam *et al.* 2006; Beauchemin and Handel 2011), from which it is clear that directly incorporating the immune response is necessary for an accurate representation of the full time course of infection (Boianelli *et al.* 2015). Even when limiting our attention to early infection only, interferon-I and natural killer cells could be included to more accurately model innate immunity (Boianelli *et al.* 2015). However, the complexity of the immune system creates a significant challenge in accurately modeling influenza A dynamics, even during this initial time period (Boianelli *et al.* 2015). In particular, many key parameters of immune kinetics remain unquantified, creating additional uncertainty (Dobrovolny *et al.* 2013).

Finally, it is well understood that antigenic drift is associated with the evolution of influenza A virus (Carrat and Flahault 2007). Antigenic drift would be formalized in our model as a reduction in the death rate of infected cells or the clearance rate of free virions, as these life history parameters would be improved by any immune evasion. In fact, Figure 2 predicts that for mutations with small selective effects ($s < 0.08$), of all possible mutations with the same selective effect, clearance

mutations are the most likely to survive when rare. Thus mutations affecting the viral clearance rate are most likely to adapt. This could shed light on the mechanisms underlying the maintenance of antigenic drift, however much remains to be understood about the complex transmission and evolutionary dynamics of influenza A virus. It is our hope that predicting the fate of *de novo* mutations affecting IAV life history is an important piece of this interesting puzzle.

ACKNOWLEDGMENTS

J.N.S.R. was funded by the Ontario Graduate Scholarship (OGS). L.M.W. is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). We thank Catherine Beauchemin for a number of insightful and helpful comments.

LITERATURE CITED

Abel, S., P. A. zur Wiesch, B. Davis, and M. Waldor, 2015 Analysis of bottlenecks in experimental models of infection. PLOS Pathogens *11*(6)**:** e1004823.

Alexander, H. and T. Day, 2010 Risk factors for the evolutionary emergence of pathogens. J. R. Soc. Interface *7*(51)**:** 1455–1474.

Alexander, H. and L. Wahl, 2008 Fixation probabilities depend on life history: fecundity, generation time and survival in a burst-death model. Evolution *62*(7)**:** 1600–1609.

Alizon, S., A. Hurfurd, N. Mideo, and M. Van Baalen, 2009 Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. J. Evol. Biol. *22*(2)**:** 245–259.

Antia, R., R. Regoes, J. Koella, and C. Bergstrom, 2003 The role of evolution in the emergence of infectious diseases. Nature **426:** 658–661.

Baccam, P., C. Beauchemin, C. A. Macken, F. G. Hayden, and A. S. Perelson, 2006 Kinetics of influenza A virus infection in humans. Journal of virology *80*(15)**:** 7590–7599.

Beauchemin, C. and A. Handel, 2011 A review of mathematical models of influenza A infections within a host or cell culture: lessons learned and challenges ahead. BMC Public Health *11*(Suppl 1)**:** S7.

Beauchemin, C., J. Samuel, and J. Tuszynski, 2005 A simple cellular automaton model for influenza A viral infections. Journal of Theoretical Biology *232*(2)**:** 223–234.

Bergstrom, C. T., P. McElhany, and L. A. Real, 1999 Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proceedings of the National Academy of Sciences of the United States of America **96:** 5095–5100.

Biggerstaff, M., S. Cauchemez, C. Reed, M. Gambhir, and L. Finelli, 2014 Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC infectious diseases *14*(1)**:** 1.

Bocharov, G. and A. Romanyukha, 1994 Mathematical model of antiviral immune response III. Influenza A virus infection. Journal of Theoretical Biology *167*(4)**:** 323–360.

Boianelli, A., V. K. Nguyen, T. Ebensen, K. Schulze, E. Wilk, N. Sharma, S. Stegemann-Koniszewski, D. Bruder, F. R. Toapanta, C. A. Guzmán, and others, 2015 Modeling Influenza Virus Infection: A Roadmap for Influenza Research. Viruses *7*(10)**:** 5274–5304.

Bouvier, N. M. and P. Palese, 2008 The biology of influenza viruses. Vaccine **26:** D49–D53.

Burch, C. L. and L. Chao, 1999 Evolution by small steps and rugged landscapes in the RNA virus phi6. Genetics **151:** 921–927.

Carrat, F. and A. Flahault, 2007 Influenza vaccine: the challenge of antigenic drift. Vaccine *25*(39)**:** 6852–6862.

Chowell, G., M. Miller, and C. Viboud, 2008 Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. Epidemiology and infection *136*(06)**:** 852–864.

Coombs, D., M. Gilchrist, and C. Ball, 2007 Evaluating the importance of within- and between-host selection pressures on the evolution of chronic pathogens. Theor. Popul. Biol. **72:** 576–591.

Day, T., S. Alizon, and N. Mideo, 2011 Bridging scales in the evolution of infectious disease life histories: theory. Evolution **65:** 3448–3461.

Desai, M. M. and D. S. Fisher, 2007 Beneficial mutation–selection balance and the effect of linkage on positive selection. Genetics *176*(3)**:** 1759–1798.

Desai, M. M., D. S. Fisher, and A. W. Murray, 2007 The speed of evolution and maintenance of variation in asexual populations. Current biology *17*(5)**:** 385–394.

Dobrovolny, H. M., M. B. Reddy, M. A. Kamal, C. R. Rayner, and C. A. Beauchemin, 2013 Assessing mathematical models of influenza infections using features of the immune response. PloS one *8*(2)**:** e57088.

Duarte, E., D. Clarke, A. Moya, E. Domingo, and J. Holland, 1992 Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. Proceedings of the National Academy of Sciences of the United States of America **89:** 6015–6019.

Duarte, E. A., D. K. Clarke, A. Moya, S. F. Elena, E. Domingo, and J. Holland, 1993 Many-trillionfold amplification of single RNA virus particles fails to overcome the Muller's ratchet effect. Journal of Virology **67:** 3620–3623.

Elena, S. F., R. Sanjuán, A. V. Bordería, and P. E. Turner, 2001 Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. Infection, Genetics and Evolution **1:** 41—48.

Eyre-Walker, A. and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nature Reviews Genetics *8*(8)**:** 610–618.

Feng, Z., S. Towers, and Y. Yang, 2011 Modeling the effects of vaccination and treatment on pandemic influenza. The AAPS journal *13*(3)**:** 427–437.

Gandon, S., M. E. Hochberg, R. D. Holt, and T. Day, 2012 What limits the evolutionary emergence of pathogens? Phil. Trans. R. Soc. Lond. B *368*(1610)**:** 20120086.

Garoff, H., R. Hewson, and D.-J. E. Opstelten, 1998 Virus maturation by budding. Microbiology and Molecular Biology Reviews *62*(4)**:** 1171–1190.

Gilchrist, M. and A. Sasaki, 2002 Modeling host-parasite coevolution: a nested approach based on mechanistic models. J. Theor. Biol. **218:** 289–308.

Grimmett, G. and D. Welsh, 2014 *Probability: an introduction*. Oxford University Press, USA.

Handel, A. and M. R. Bennett, 2008 Surviving the bottleneck: transmission mutants and of microbial populations. Genetics *180*(4)**:** 2193–2200.

Iwasa, Y., F. Michor, and M. Nowak, 2003 Evolutionary dynamics of escape from biomedical intervention. Proc. R. Soc. Lond. B **270:** 2573–2578.

Kawaoka, Y., O. T. Gorman, T. Ito, K. Wells, R. O. Donis, M. R. Castrucci, I. Donatelli, and R. G. Webster, 1998 Influence of host species on the evolution of the nonstructural (NS) gene of influenza A viruses. Virus Research *55*(2)**:** 143–156.

Lachapelle, J., J. Reid, and N. Colegrave, 2015 Repeatability of adaptation in experimental populations of different sizes. Proc. R. Soc. Lond. B *282*(1805)**:** 20143033.

Larson, E. W., J. W. Dominik, A. H. Rowberg, and G. A. Higbee, 1976 Influenza virus population dynamics in the respiratory tract of experimentally infected mice. Infection and immunity *13*(2)**:** 438–447.

Lau, L. L., B. J. Cowling, V. J. Fang, K.-H. Chan, E. H. Lau, M. Lipsitch, C. K. Cheng, P. M. Houck, T. M. Uyeki, J. M. Peiris, and others, 2010 Viral shedding and clinical illness in naturally acquired influenza virus infections. Journal of Infectious Diseases *201*(10)**:** 1509–1516.

Martin, K. and A. Helenius, 1991 Transport of incoming influenza virus nucleocapsids into the nucleus. Journal of virology *65*(1)**:** 232–244.

McCaw, J. M., N. Arinaminpathy, A. C. Hurt, J. McVernon, and A. R. McLean, 2011 A mathematical framework for estimating pathogen transmission fitness and inoculum size using data from a competitive mixtures animal model. PLoS Comput Biol *7*(4)**:** e1002026.

MIDEO, N., S. ALIZON, and T. DAY, 2008 Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases. Trends in Ecol. Evol. **23:** 511–517.

NOBUSAWA, E. and K. SATO, 2006 Comparison of the mutation rates of human influenza A and B viruses. Journal of Virology *80*(7)**:** 3675–3678.

NOVELLA, I. S., S. F. ELENA, A. MOYA, E. DOMINGO, and J. J. HOLLAND, 1995 Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. Journal of Virology **69:** 2869–2872.

NOVELLA, I. S., S. F. ELENA, A. MOYA, E. DOMINGO, and J. J. HOLLAND, 1996 Repeated transfer of small RNA virus populations leading to balanced fitness with infrequent stochastic drift. Mol Gen Genet **252:** 733–738.

OTTO, S. P. and M. C. WHITLOCK, 1997 The probability of fixation in populations of changing size. Genetics *146*(2)**:** 723–733.

PATWA, Z. and L. WAHL, 2008 Fixation probabilities for lytic viruses: the attachment-lysis model. Genetics **180:** 459–470.

PATWA, Z. and L. WAHL, 2009 The impact of host-cell dynamics on the fixation probability for lytic viruses. Journal of Theoretical Biology *259*(4)**:** 799–810.

PECK, K. M., C. H. CHAN, and M. M. TANAKA, 2015 Connecting within-host dynamics to the rate of viral molecular evolution. Virus Evolution *1*(1)**:** vev013.

PINILLA, L. T., B. P. HOLDER, Y. ABED, G. BOIVIN, and C. A. BEAUCHEMIN, 2012 The H275Y neuraminidase mutation of the pandemic A/H1N1 influenza virus lengthens the eclipse phase and reduces viral output of infected cells, potentially compromising fitness in ferrets. Journal of virology *86*(19)**:** 10651–10660.

POON, L. L., T. SONG, R. ROSENFELD, X. LIN, M. B. ROGERS, B. ZHOU, and R. SEBRA *et al.*, 2016 Quantifying influenza virus diversity and transmission in humans. Nature Genetics **48:** 195–200.

RAMPHAL, R., W. FISCHLSCHWEIGER, J. W. SHANDS JR, and P. A. SMALL JR, 1979 Murine influenzal tracheitis: A model for the study of influenza and tracheal epithelial repair. American

Review of Respiratory Disease **120:** 1313–1324.

Raynes, Y., A. Halstead, and P. Sniegowski, 2014 The effect of population bottlenecks on mutation rate evolution in asexual populations. J. Evol. Biol. **27:** 161–169.

Reluga, T., R. Meza, D. Walton, and A. Galvani, 2007 Reservoir interactions and disease emergence. Theor. Popul. Biol. **72:** 400–408.

Roy, A.-M. M., J. S. Parker, C. R. Parrish, and G. R. Whittaker, 2000 Early stages of influenza virus entry into Mv-1 lung cells: involvement of dynamin. Virology *267*(1)**:** 17–28.

Sanjuán, R., A. Moya, and S. F. Elena, 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proceedings of the National Academy of Sciences of the United States of America *101*(22)**:** 8396–8401.

Smith, A. and A. S. Perelson, 2011 Influenza A virus infection kinetics: quantitative data and models. Wiley Interdisciplinary Reviews: Systems Biology and Medicine *3*(4)**:** 429–445.

Stray, S. J. and G. M. Air, 2001 Apoptosis by influenza viruses correlates with efficiency of viral mRNA synthesis. Virus research *77*(1)**:** 3–17.

Tamura, S.-I. and T. Kurata, 2004 Defense mechanisms against influenza virus infection in the respiratory tract mucosa. Jpn J Infect Dis *57*(6)**:** 236–47.

van de Sandt, C. E., J. H. Kreijtz, and G. F. Rimmelzwaan, 2012 Evasion of influenza A viruses from innate and adaptive immune responses. Viruses *4*(9)**:** 1438–1476.

Varble, A., R. A. Albrecht, S. Backes, M. Crumiller, N. M. Bouvier, D. Sachs, A. García-Sastre, and others, 2014 Influenza A virus transmission bottlenecks are defined by infection route and recipient host. Cell host & microbe *16*(5)**:** 691–700.

Vogwill, T., R. L. Phillips, D. R. Gifford, and R. C. MacLean, 2016 Divergent evolution peaks under intermediate population bottlenecks during bacterial experimental evolution. Proc. R. Soc. Lond. B *283*(1835)**:** 20160749.

Wahl, L. and P. J. Gerrish, 2001 The probability that beneficial mutations are lost in populations with periodic bottlenecks. Evolution *55*(12)**:** 2606–2610.

Wahl, L., P. J. Gerrish, and I. Saika-Voivod, 2002 Evaluating the impact of population bottlenecks in experimental evolution. Genetics *162*(2)**:** 961–971.

Wahl, L. and A. D. Zhu, 2015 Survival probability of beneficial mutations in bacterial batch culture. Genetics *200*(1)**:** 309–320.

Wright, P. F., R. G. Webster, and others, 2001 Orthomyxoviruses. Fields virology **1:** 1533–1579.

APPENDIX

Let $p_{lmn}(t)$ be the probability that l free virions, m infected cells, and n mature cells exist in the focal lineage at time t, and let $A(t)$ denote the time-dependent per virion attachment rate, which depends on the available target cells, $y_T(t)$, as predicted in the deterministic model 1. Parameters B, C, and D represent the budding, clearance and cell death rates, while E denotes the rate at which cells exit the eclipse phase and begin budding. Although the stochastic model follows the mutant lineage, for simplicity we will use A as opposed to $\tilde{A}$, *etc.*, throughout the Appendix. Also for notational clarity we illustrate the case $k = 1$. Taking into account the stochastic events of attachment, budding, clearance, cell death and cell maturation, is is straightforward to demonstrate that the probability generation function (pgf) describing the time evolution of the lineage must satisfy:

$$\begin{aligned} G(t+\Delta t, \overrightarrow{x}) = G(t, \overrightarrow{x}) &+ \sum_{l,m,n} p_{lmn}(t) lC\Delta t[-x_1^l x_2^m x_3^n + x_1^{l-1} x_2^m x_3^n] + \\ &\sum_{l,m,n} p_{lmn}(t) lA(t)\Delta t[-x_1^l x_2^m x_3^n + x_1^{l-1} x_2^{m+1} x_3^n] + \\ &\sum_{l,m,n} p_{lmn}(t) mE\Delta t[-x_1^l x_2^m x_3^n + x_1^l x_2^{m-1} x_3^{n+1}] + \\ &\sum_{l,m,n} p_{lmn}(t) mD\Delta t[-x_1^l x_2^m x_3^n + x_1^l x_2^{m-1} x_3^n] + \\ &\sum_{l,m,n} p_{lmn}(t) nB\Delta t[-x_1^l x_2^m x_3^n + x_1^{l+1} x_2^m x_3^n] + \\ &\sum_{l,m,n} p_{lmn}(t) nD\Delta t[-x_1^l x_2^m x_3^n + x_1^l x_2^m x_3^{n-1}] \end{aligned} \tag{4}$$

Taking the limit as $\Delta t \to 0$, Equation 4 yields the following linear partial differential equation:

$$\begin{aligned} \frac{\partial G}{\partial t} = &(A(t)x_2 + C - (A(t)+C)x_1)\frac{\partial G}{\partial x_1} \\ &+(-(E+D)x_2 + Ex_3 + D)\frac{\partial G}{\partial x_2} \\ &+(Bx_1x_3 + D - (D+B)x_3)\frac{\partial G}{\partial x_3} \end{aligned} \tag{5}$$

Equation 5 can be converted to a system of ordinary differential equations using the standard method of characteristics, which yields the following system of ordinary differential equations

$$\left.\begin{aligned} \frac{dx_1}{dT} &= A(t)x_2 + C - (A(t)+C)x_1 \\ \frac{dx_2}{dT} &= -(E+D)x_2 + Ex_3 + D \\ \frac{dx_3}{dT} &= Bx_1x_3 + D - (D+B)x_3 \\ \frac{dt}{dT} &= -1 \end{aligned}\right\} .$$

This system can be solved numerically to determine the value of G at time τ, given the known initial condition corresponding to a single free virion at time t_0, $G(t_0, x_1, x_2, x_3) = x_1$. For convenience we let $\mathcal{G}(t_0, x_1) = G(\tau, x_1, 1, 1)$ under this initial condition.

The function $\mathcal{G}(t_0, x_1)$, then, gives the distribution of free virions at time τ, just before disease transmission, given the lineage began with a single virion at time t_0. Composing this with the pgf of the bottleneck process, we obtain $\mathcal{G}(t_0, 1 - F + Fx_1)$ as the pgf describing the distribution of free virions transmitted to a new host (given that one new host is infected). The probability that a given lineage, that arose at time t_0, is *n*ot transmitted to the new host is obtained by evaluating at $x_1 = 0$:

$$X(t_0) = \mathcal{G}(t_0, 1 - F) \ .$$

We then use this to compute the expected rate at which surviving mutant strains appear at time t_0, where "surviving" means the lineage will be transfered to the next host:

$$\nu(t_0) = \mu \left(1 - X(t_0)\right) B \, y_B(t_0)$$

where μ is the probability that the mutation of interest occurs, per new virion produced. We use this to compute $\mathcal{S}$, the expected number of times that the mutation of interest occurs *de novo*, over

the course of the infection, and survives to be transmitted to the next host:

$$\mathcal{S} = \int_0^\tau \nu(t_0) dt_0 \,.$$

Consider dividing the time interval $(0, \tau)$ such that $\delta t = \tau/N$ and $t_i = i\delta t$. In this case for small δt, the quantity $\nu(t_0)\delta t$ approximates the probability that a surviving mutation occurs during time interval $(t_0, t_0 + \delta t)$. This allows us to compute $\mathcal{P}$, the probability that the mutation of interest occurs *de novo* during the course of the infection and is transmitted to the new host:

$$\mathcal{P} = 1 - \lim_{N\to\infty} \prod_{i=0}^{N-1} \left(1 - \nu(t_i)\frac{\tau}{N}\right)$$

which by product integration can be succinctly expressed as:

$$\mathcal{P} = 1 - e^{-\mathcal{S}} \,.$$

We also compute the expected number of mutant virions transmitted to the recipient host, $\mathcal{N}$. We do this by first computing $\partial_x \mathcal{G}(t_0, x)|_{x=1}$, which gives the expected number of mutant virions at time τ, given that a mutant virion was produced at time t_0. We multiply this value by the number of mutant virions being produced at time t_0, $\mu B y_B(t_0)$, and integrate from 0 to τ, to get the total expected number of mutant virions at time τ. Multiplying by the bottleneck fraction, F, gives the expected number of mutant virions transmitted to the recipient host:

$$\mathcal{N} = F \int_0^\tau \mu B y_B(t_0) \cdot \partial_x \mathcal{G}(t_0, x)|_{x=1} dt_0 \,.$$

Note that $\mathcal{S}$ and $\mathcal{N}$ differ because each *de novo* mutation produces a lineage that could in principle contribute more than one virion to the recipient.

Supplementary Figures

S1. Expected number of successful *de novo* occurrences, $\mathcal{S}$

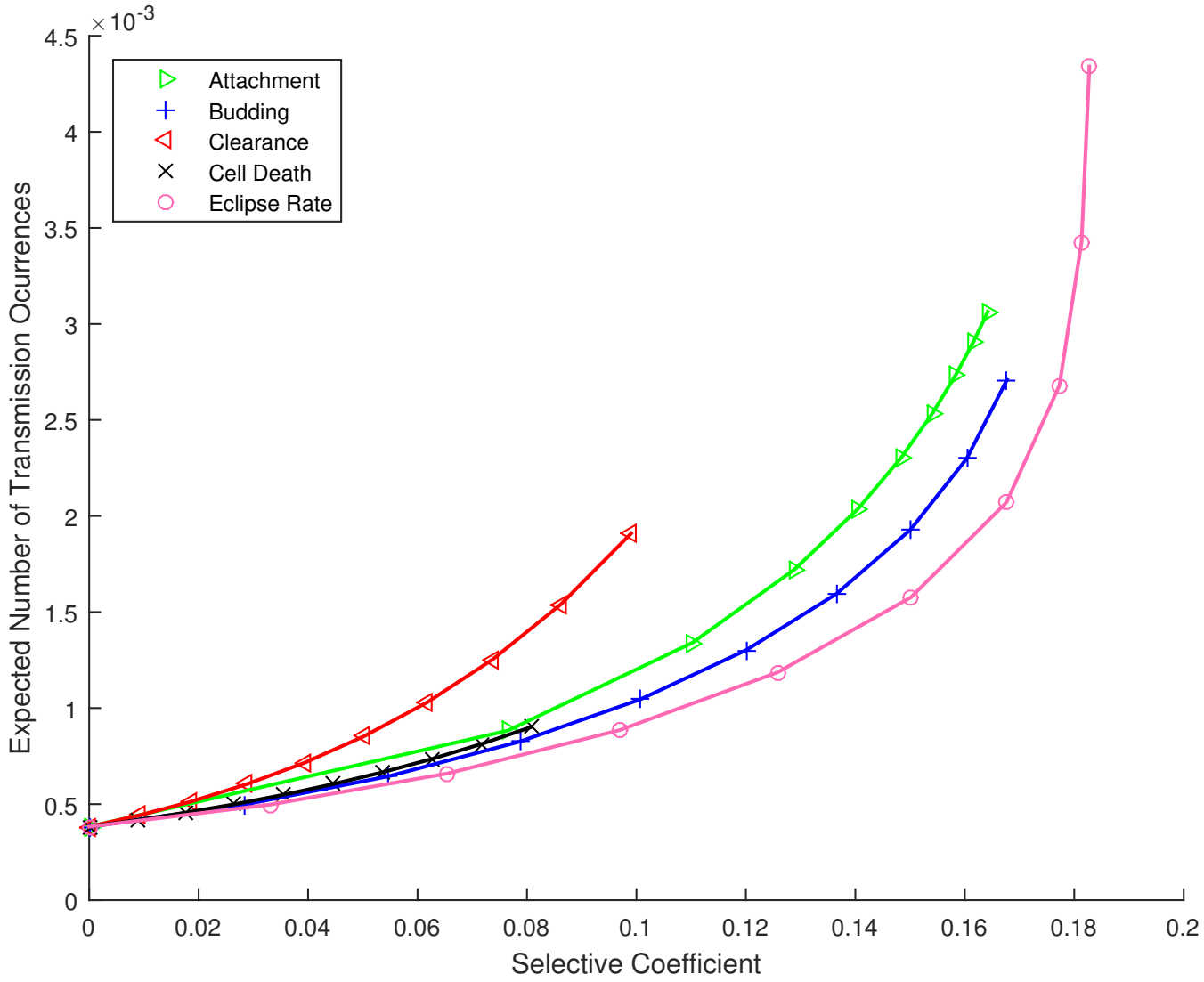

Figure S1: The number of times that a given mutation is expected to arise *de novo*, during a single infection time course, and produce a lineage that is transmitted to the next infected individual ($\mathcal{S}$, as described in the Appendix), versus the selective coefficient. This quantity scales linearly with μ, the mutation rate.

S2. Expected number of transmitted mutant virions, $\mathcal{N}$

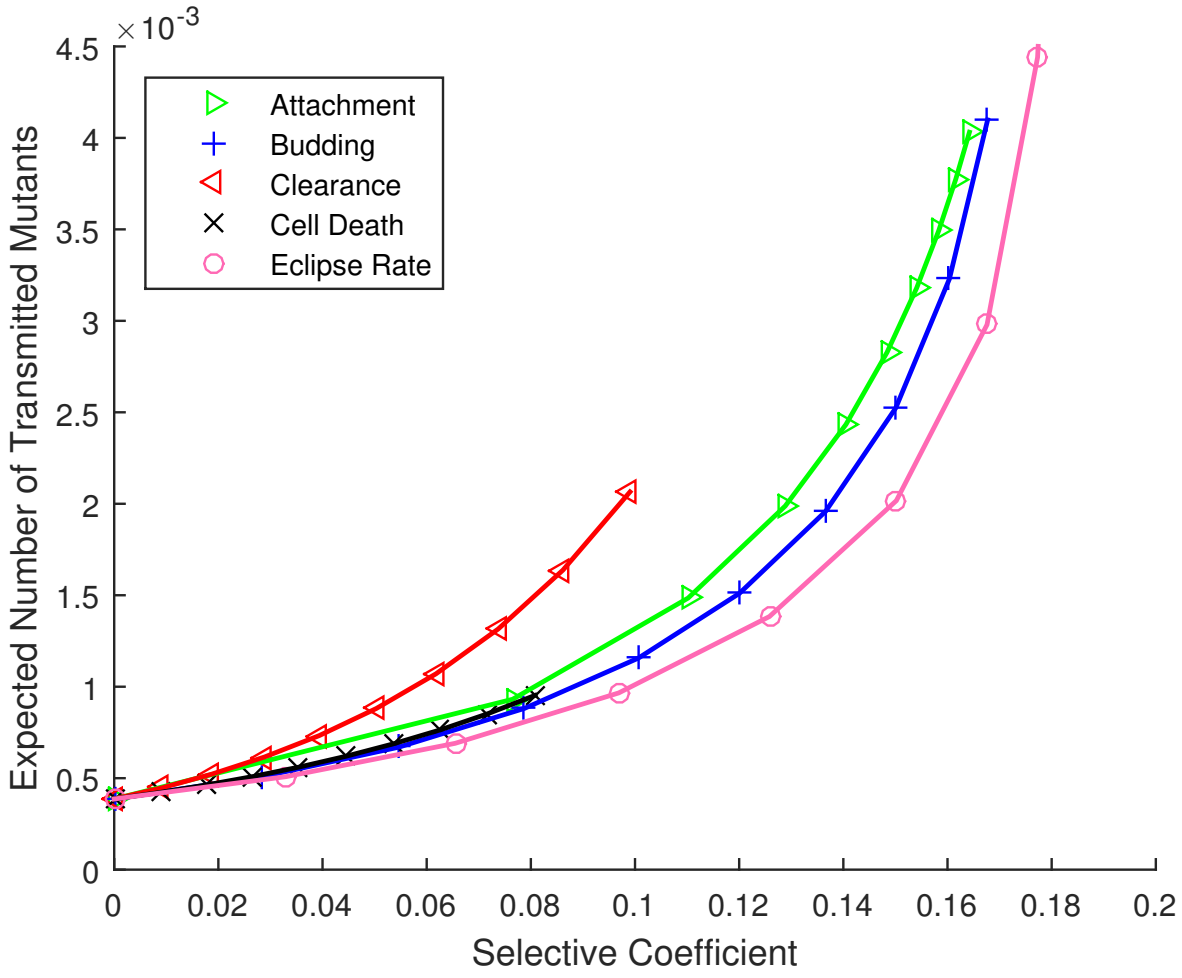

Figure S2: The expected number of virions transmitted to the recipient that have arisen *de novo* during a single infection time course in the donor ($\mathcal{N}$, as described in the Appendix), versus the selective coefficient.

S3. Effect of non-exponential cell lifetimes: Deterministic time course

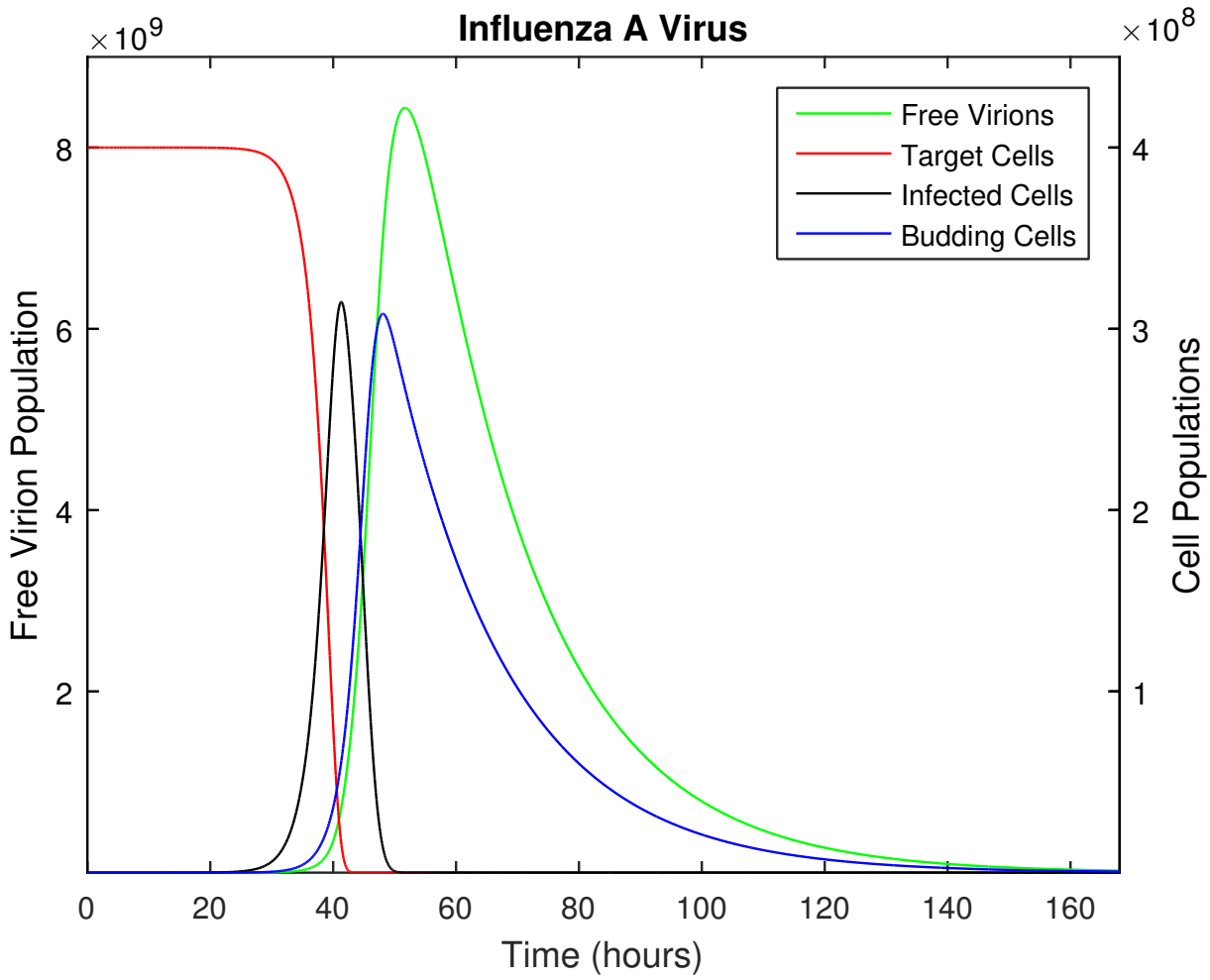

Figure S3: The time course of influenza A infection over the span of a week (168 hours). This figure is analogous to Figure 1, except that the cell death rate, D, has been set to zero during the eclipse stages, and increased during the budding phase such that the mean infected cell lifetime is unchanged. Other parameters as provided in Table 1. The infection time course is relatively insensitive to these changes in the distribution of infected cell lifetimes.

S4. Effect of non-exponential cell lifetimes: Transmission probability

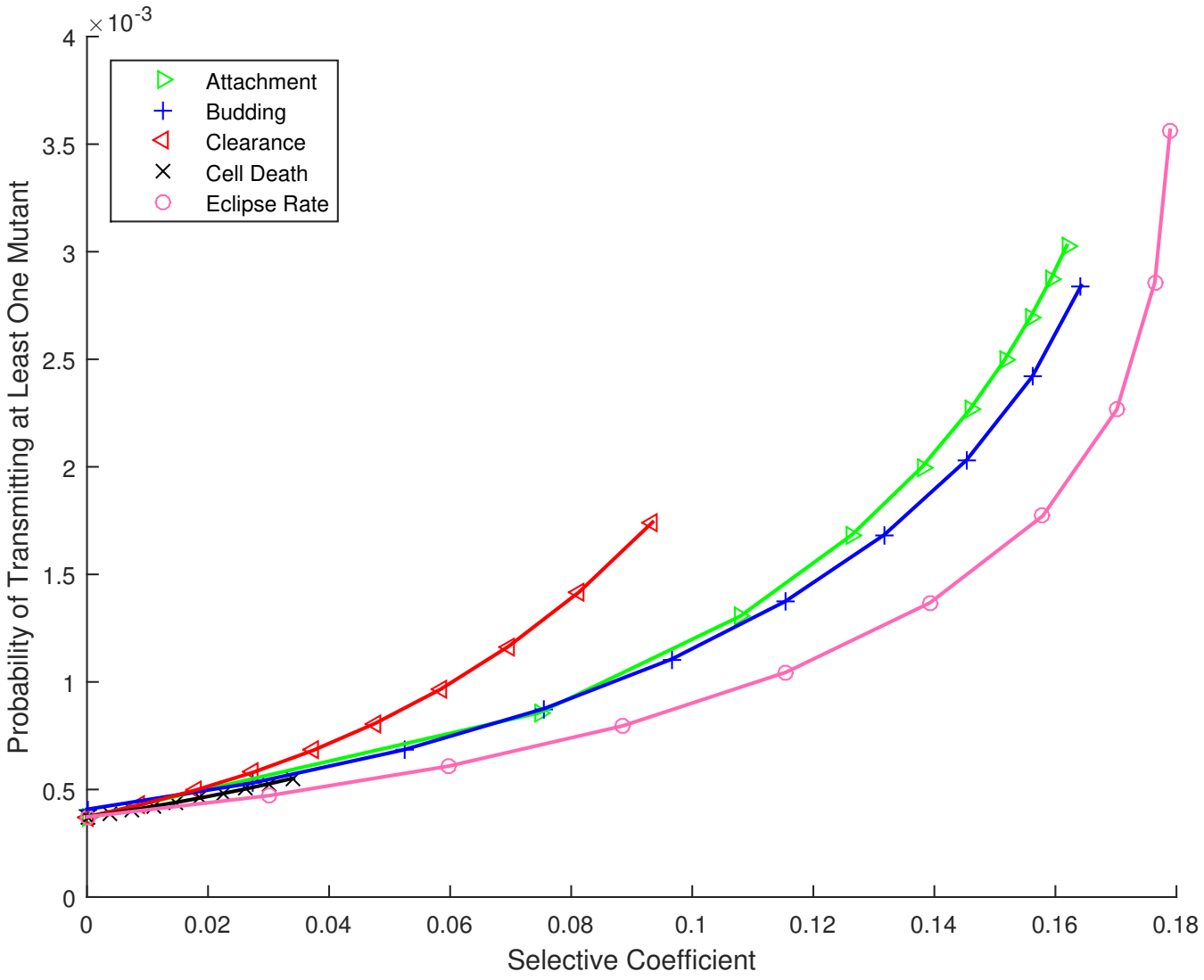

Figure S4: The probability that at least one *de novo* mutation arises during the infection time course and is passed to the next host, $\mathcal{P}$, versus the selective coefficient, s. This figure is analogous to Figure 2, except that the cell death rate, D, has been set to zero during the eclipse stages, and increased during the budding phase to yield the same mean infected cell lifetime. Other parameters as provided in Table 1. We find that the transmission of *de novo* mutations is insensitive to these changes in the distribution of infected cell lifetimes.

S5. Effect of available target cells: Deterministic time course

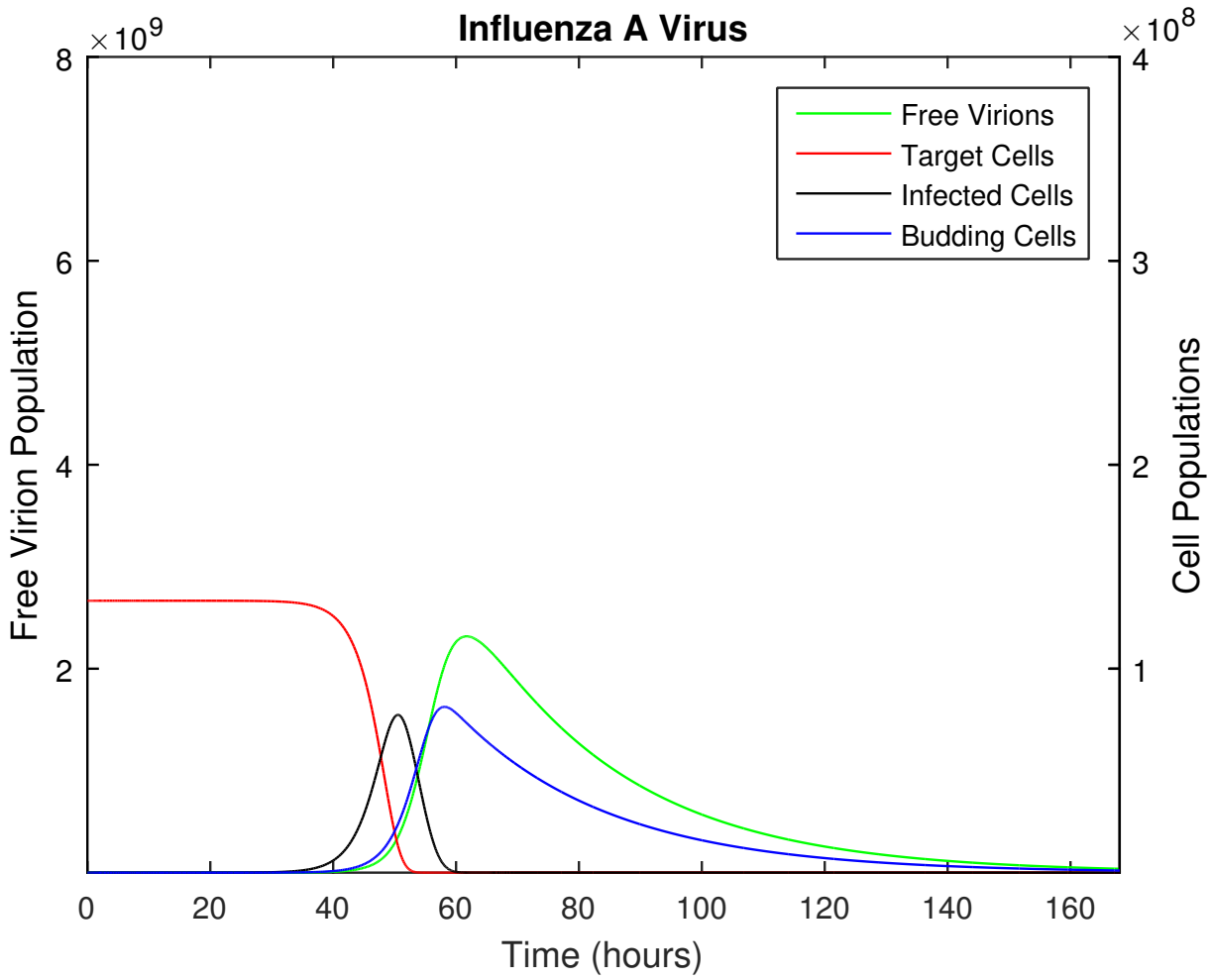

Figure S5: The time course of influenza A infection over the span of a week (168 hours). This figure is analogous to Figure 1, except that the initial target cell population, $y_T(0)$, has been reduced by a factor of 3. Although complete desquamation is the expected outcome of the infection, it is possible that spatial considerations might spare a fraction of the epithelial cells in the upper respiratory tract; we therefore included this case in sensitivity analysis. Other parameters as provided in Table 1. Comparing with Figure 1, the magnitude of the infection is scaled and the dynamics are slightly delayed. However this has little impact on the probability of transmission of a mutation (see Figure S6).

S6. Effect of number of available target cells: Transmission probability

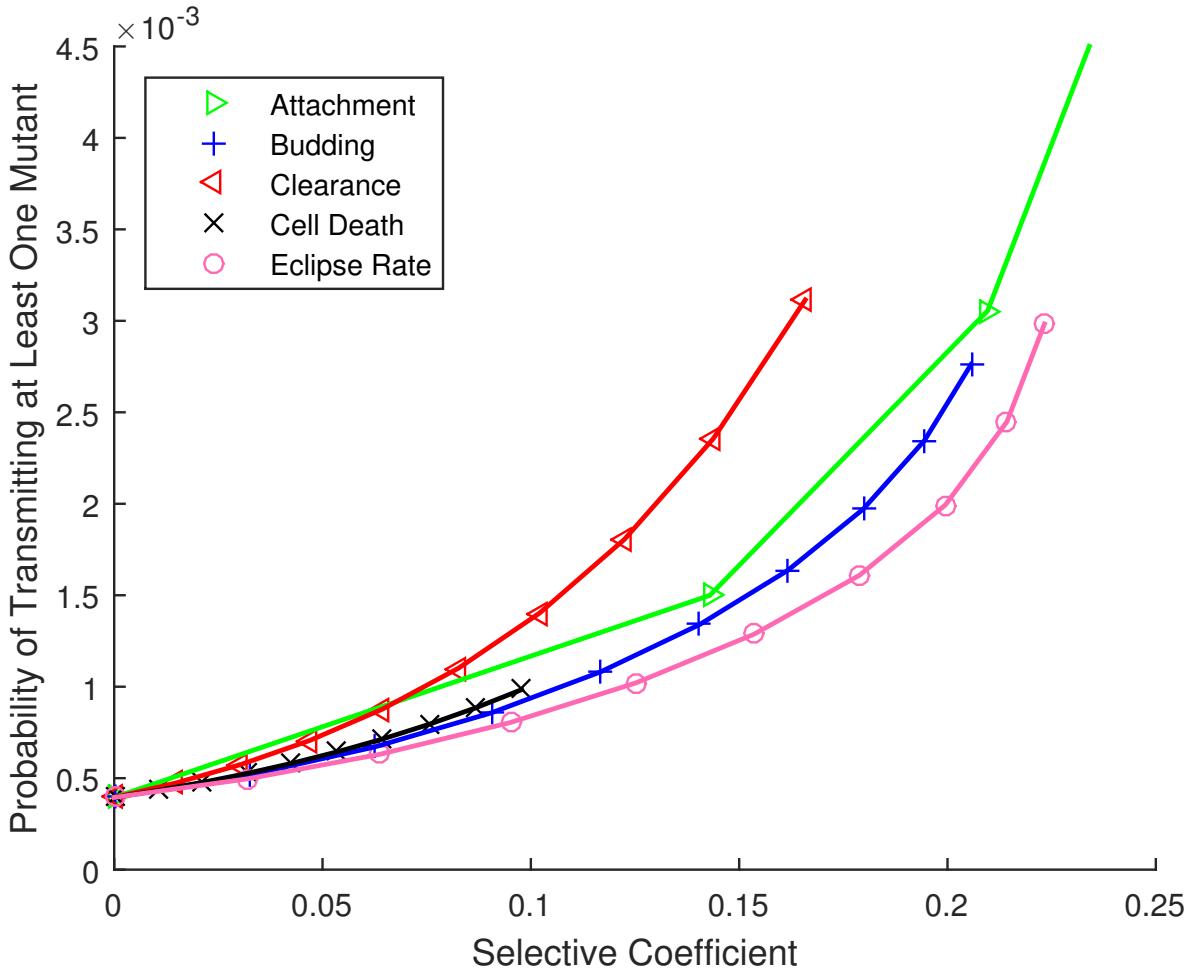

Figure S6: The probability that at least one *de novo* mutation arises during the infection time course and is passed to the next host, $\mathcal{P}$, versus the selective coefficient, s. This figure is analogous to Figure 2, except that the initial target cell population, $y_T(0)$, has been reduced by a factor of 3. Other parameters are as provided in Table 1. We find that the transmission of *de novo* mutations is insensitive to the initial number of available target cells.

S7. Effect of varying mutation rate

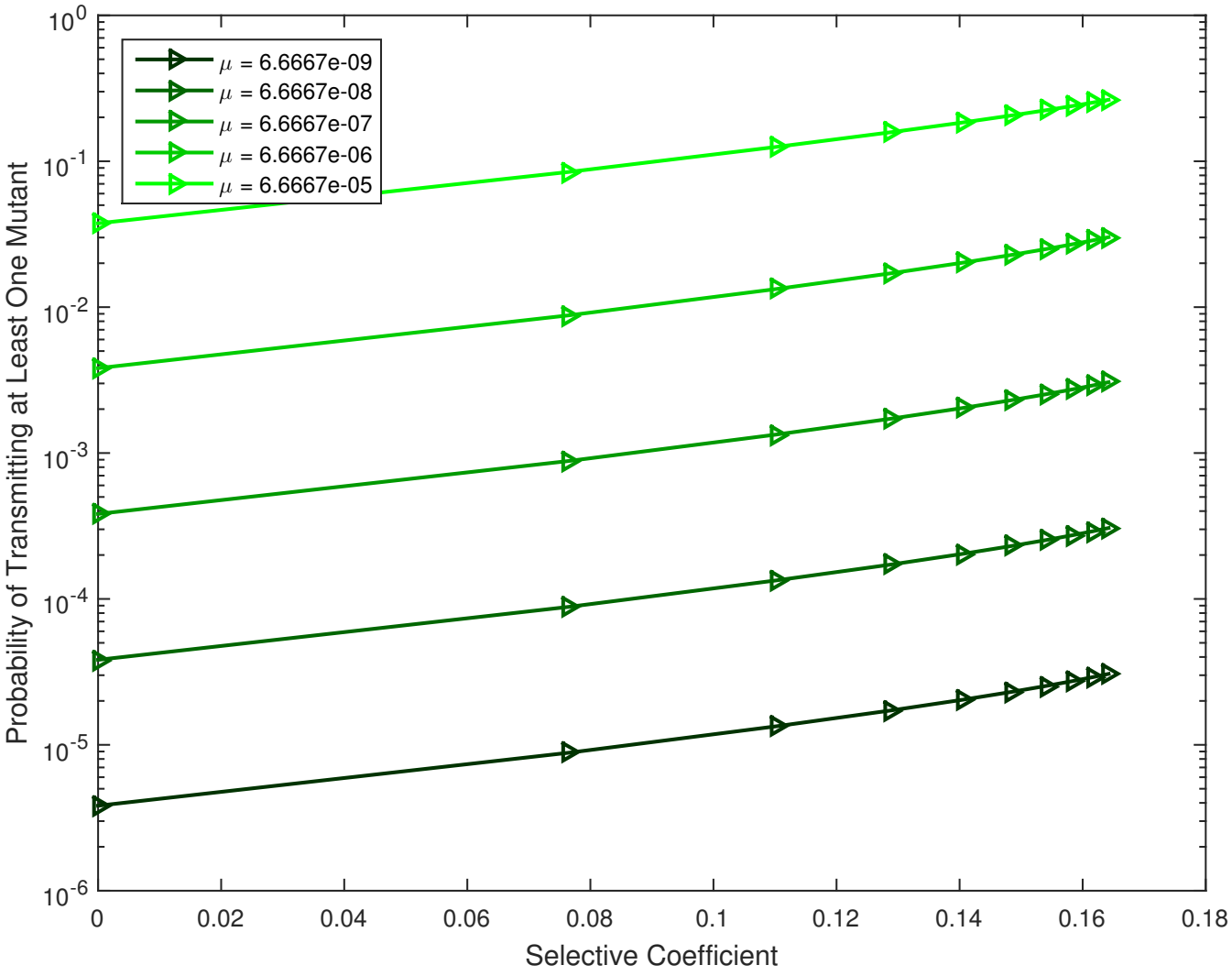

Figure S7: The effect of mutation rate on the probability of transmission. The probability that at least one copy of a *de novo* mutation is transmitted to the next host is plotted against the selective coefficient, for a mutation that increases the viral attachment rate. The mutation rate per replication event, μ, is varied. Other parameters as provided in Table 1.