# INFLUENCE OF VIRAL REPLICATION MECHANISMS ON WITHIN-HOST EVOLUTIONARY DYNAMICS

Claude Loverdo,[1,4] Miran Park,[1] Sebastian J. Schreiber,[2] and James O. Lloyd-Smith[1,3]

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095

[2]Department of Evolution and Ecology and Center for Population Biology, University of California, Davis, California 95616

[3]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892

[4]E-mail: loverdo@ucla.edu

**Viruses replicate their genomes using a variety of mechanisms, leading to different distributions of mutations among their progeny. Yet, models of viral evolution often only consider the mean mutation rate. To investigate when and how replication mechanisms impact viral evolution, we analyze the early dynamics of within-host infection for two idealized cases: when all offspring virions from an infected cell carry the same genotype, mutated or not; and when mutations occur independently across offspring virions. Other replication life histories fall between these extremes. Using branching process models, we study the probability that viral infection becomes established when mutations are lethal, and in the more general case of two strains of different fitness. For a given mean mutation rate, we show that a lineage of viruses with correlated mutations is less likely to survive than with independent mutations, but when it survives, the viral population grows faster. While this holds true for all parameter regimes, replication life history has a quantitatively significant influence on viral dynamics when stochastic effects are important and when mutations are crucial for survival—conditions typical of evolutionary escape situations.**

**KEY WORDS: Models/simulations, mutations, reproductive strategies.**

Developing a mechanistic understanding of viral evolution is a key step toward addressing important questions about antiviral drug resistance, immune escape, and adaptation to new host species or tissues (Moya et al. 2004; Pybus and Rambaut 2009). Despite important advances in modeling viral dynamics, most models are either highly detailed, which confines them to highly specific viral life histories, or are simple and general but ignore important features of viral biology. One approach to bridging these extremes is to integrate known aspects of viral life history into simple models to study when and how conclusions are modified, and consequently how much these biological details have to be taken into account for relevant modeling. For example, Pearson et al. (2011) studied how the distribution of virions produced by an infected cell modifies the extinction probability of a viral lineage. Here we expand this line of analysis to explore the

mutational dynamics that are central to viral evolution. Mutations can be deleterious or even lethal (Anderson et al. 2004; Bull et al. 2007; Martin and Gandon 2010; Manrubia et al. 2010). Yet, mutations are also crucial for viral survival in a changing environment, for instance to develop resistance to a drug (Maisnier-Patin and Andersson 2004) or to adapt to a new host species in which the virus is initially unfit (Holmes and Drummond 2007; Domingo 2010; Pepin et al. 2010). This article explores how a more realistic depiction of how mutations arise in viral replication may alter the overall dynamics of viral growth and evolution within hosts.

A fundamental aspect of virology is the study of the various mechanisms for replication of viral genomes. The seminal work of Luria (1951) used the distribution of mutants in plate assays in conjunction with mathematical models to infer the underlying replication mechanism. He introduced the distinction between

binary replication and the stamping machine. Binary replication is when the genome of the virion that initially infects a cell undergoes successive rounds of duplication, with each generation of new genomes being used as the templates for the next, for example, the dsDNA bacteriophage T2. The stamping machine mechanism is when the initial viral genome is used to make a template from which all new viral genomes are produced. The ssDNA bacteriophage $\phi$X174 studied by Denhardt and Silver (1966) exemplifies this strategy, as it is first complemented to dsDNA from which new ssDNA strands are produced. For a given mutation rate per copying step, the stamping machine mechanism leads to fewer mutations than binary replication (because each offspring genome undergoes only two copying steps in a stamping machine, whereas under binary replication each of the $N$ final offspring genomes is $\ln(N)/\ln(2)$ copying steps away from the original genome), and has been proposed to be favored by evolution to reduce the mutational load (French and Stenger 2003; Sardanyés et al. 2009; Thébaud et al. 2010).

However, the mutation rate per generation (the mean proportion of mutants produced by an infected cell) is more easily measured experimentally than the mutation rate per genome copying step and the number of such steps within an infected cell (Sanjuán et al. 2010). Is the mean mutation rate per generation sufficient to characterize the viral evolutionary dynamics? The same mean can result from different underlying processes. For example, in the pure stamping-machine model, the mutation rates when making and reading the template can be different. An example occurs in retroviruses: retrotranscriptases incorporate the information from the viral RNA to the DNA genomic material of the host, which is then used as a template by the DNA-dependent RNA polymerases of the host to synthesize the new viral RNA copies. The two steps use different enzymes, and hence are likely to have different error rates. It is difficult to assess which enzyme is responsible for most mutations based on independent measurements of the enzymes' error rates, because the measurements span several orders of magnitude and overlap widely (de Mercoyrol et al. 1992; Shaw et al. 2002; Svaroskaia et al. 2003; Kireeva et al. 2008). However, Kim et al. (1996) and O'Neil et al. (2002) show that for two experimental systems, the retrotranscriptase is responsible for at least one-third to one-half of the mutations. Because mutations to the retrotranscriptase enzyme itself may lead to higher or lower fidelity, it is likely that systems exist where the retrotranscriptase makes more errors than the RNA polymerase and vice versa. In the extreme, if the mutation rate creating the template is much larger than the mutation rate reading the template, either all the new virions produced by an infected cell are mutated or none are, that is, the presence of mutations in the virions produced by a cell is completely correlated. In the opposite extreme, each new virion has an independent chance of being a mutant, there is no correlation in the presence of mutants (Fig. 1). Thus, even when these

alternative forms of replication lead to the same mean mutation rate, the distributions of mutants can be very different.

For a given mean mutation rate, we consider two distributions: an infected cell produces virions all with the same genotype, mutated or not (hereafter named the "all-or-none" mechanism), or virions mutate independently of each other (hereafter named the "independent" mechanism) (Fig. 1). All known real-life replication mechanisms (e.g., binary replication and stamping machines with error in both steps) lead to correlation between mutants that falls between these two extremes. Studying these limits allows us to address whether the mean mutation rate is enough to describe the evolutionary dynamics, and to delimit the regimes where we need to know more about the distribution of mutants. As the differences in the distribution beyond the mean are more likely to matter when there are very few virions, we focus our study on early dynamics of viral establishment within a host. As mutations can be either deleterious or adaptive, we study the case of lethal mutations and the case of two strains with different (nonzero) fitnesses. This latter case is the foundation for studies of viral adaptation, and we emphasize its relevance for studies of evolutionary escape.

## Model and Methods
### EARLY STAGE OF INFECTION

Because we are comparing the all-or-none and independent mechanisms for a fixed mean number of mutants produced by an infected cell, our model is focused on the early dynamics of infection when stochastic effects play an important role. We assume that a virion has a fixed probability $q$ to successfully infect a cell before being degraded. In the early phase of infection, the pool of susceptible cells is not significantly depleted, so it can be considered constant, and large enough so that infection of one cell with multiple viruses can be neglected. We also assume that the impact of the immune response is constant through time, that is, we do not consider the adaptive response and associated complexities of virus–immune coevolution (Kamp 2003; Volkov et al. 2010). More generally, we assume that there are no interactions between virions, so their demographic fates are independent. Thus, we use branching processes to model viral dynamics similarly to Antia et al. (2003) or Iwasa et al. (2004).

### NUMBER OF VIRIONS PRODUCED PER INFECTED CELL

A cell successfully infected by a virus releases multiple new virions. Pearson et al. (2011) showed that the distribution of the number of virions produced per infected cell changes the probability of survival of a viral population initiated with a single infecting virion. They proposed two main models: (i) bursting, in
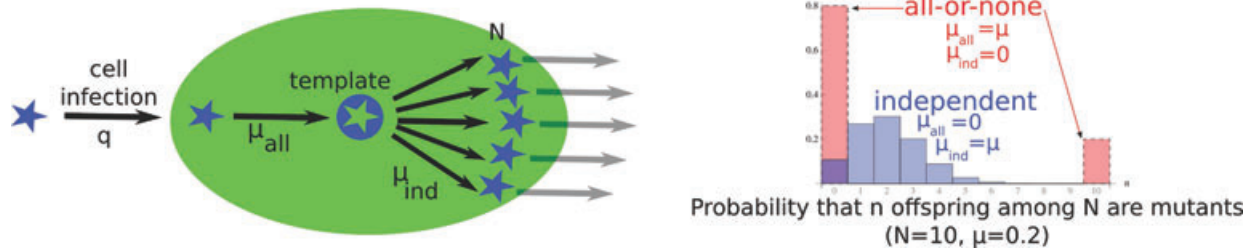
**Figure 1.** In pure stamping machine replication, the genome of a virion that successfully infects a cell (with probability $q$) is copied to make a template, and the template is read to make the genetic material for $N$ new virions. Mutations can occur in making the template ($\mu_{all}$) or reading the template ($\mu_{ind}$). We study the two limiting cases, when one of the mutation rates is negligible compared to the other. With the "all-or-none" mechanism ($\mu_{all} \gg \mu_{ind}$, so all the mutations occur when making the template), either all offspring virions are mutated or none are. With the "independent" mechanism ($\mu_{ind} \gg \mu_{all}$, so all the mutations occur when reading the template) mutations are independent from each other. The all-or-none mechanism leads to larger variance in the number of mutants than the independent mechanism.

which cells release exactly $N$ virions, and (ii) budding, where the infected cells are assumed to die at a constant rate $d$ per unit time and to produce a new virion at a constant rate $b$ per unit time, leading the number of virions produced to follow a geometric distribution with mean $N = b/d$. For simplicity, we focus on the budding case, but we show that our main conclusions also hold true for the bursting model (see Appendices S1 and S3). The range of realistic $N$ ranges at least from 10 to $5 \times 10^4$ (Todd et al. 1997; Kew et al. 2005; Parada et al. 2006; Chen et al. 2007; Wilhelm and Matteson 2008; De Paepe et al. 2010).

## VIRAL GENOTYPES AND FITNESS

Mutations link the different viral genotypes. $\mu_{ij}$ is the probability to mutate from strain $i$ to strain $j$. Each genotype is characterized by two fitness parameters: the probability $q_i$ for a virion to infect a cell, and the mean number $N_i$ of virions produced by an infected cell. Additionally, when a cell infected with strain $i$ releases one or several virions that have the mutant genotype of strain $j$, we assume that such a cell releases an average of $N_i$ virions, and that new virions bearing genotype $j$ successfully infect a cell with probability $q_j$.

## A GENERAL STOCHASTIC MODEL FOR VIRAL INVASION

We model the system using discrete-time branching process, where time is measured in generations, the cycle of infection of a cell and release of new virions. In real time, these generations may overlap, but we mainly discuss the probability that the initial virus starts a successful infection of the host, which is independent of the time scale.

To analyze the branching processes, we use generating functions which gather the information on the probabilities $p(k)$ that a virion produces $k$ virions for next generation: $g(z) = \sum_{k=0}^{\infty} p(k)z^k$. The generating function for generation $t$ is given

by the $t$-fold composition $g^t(z)$ of the original generating function. Hence, generating functions provide an easy way to understand the distribution of the viral population at any generation. In particular, the probability that the viral population is extinct at generation $t$ is given by $g^t(0)$. Moreover, standard branching process theory implies that the probability of eventual extinction $e$ is given by the smallest positive solution to $g(e) = e$ (Harris 1963). The probability for a viral lineage to survive the early steps and go to full infection is $s = 1 - e$. These results can be extended to multitype branching processes, with the generating map $g_i(\vec{z}) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_n=0}^{\infty} p_i(\vec{k})z_1^{k_1}...z_n^{k_n}$ with $p_i(\vec{k})$ the probability that one virion of strain $i$ produces a set of $k_1$ virions of type 1, $k_2$ virions of type 2, etc. (Harris 1963).

Our analysis focuses on the case of a single founding virion. As we assume that virions are independent, the dynamics for an initial dose of several virions can be obtained directly from the dynamics for each of them. For example, the probability of extinction at generation $t$ for all $N$ founders is $(g^t(0))^N$. For many viruses, the number of viral particles that pass from one host to the next in a transmission event is very small, so the stochastic effects considered here remain important (Ali et al. 2006; Kearney et al. 2009; Keele et al. 2009; Wang et al. 2010).

The generating function for a geometric distribution of mean $N$ is $g(z) = 1/(1 + N(1 - z))$. If a virion has a probability $q$ to infect a cell, which then releases on average $N$ virions, the generating function is $g(z) = 1 - q + q/(1 + N(1 - z))$. With the all-or-none mechanism, a virion $i$ has a probability $q_i$ to infect a cell, and makes a template $j$ with probability $\mu_{ij}$:

$$g_i^{all}(\vec{z}) = 1 - q_i + q_i \sum_j \frac{\mu_{ij}}{1 + N_i(1 - z_j)}. \tag{1}$$

In the independent case, mutations occur when the template is read:

$$g_i^{ind}(\vec{z}) = 1 - q_i + \frac{q_i}{1 + N_i\left(1 - \sum_j \mu_{ij} z_j\right)}. \qquad (2)$$

Using these generating functions, we study the survival probabilities and the distribution of the number of virions when mutants are lethal, and in the more general case of two strains of different fitnesses.

## Results

### LETHAL MUTANTS

Viral genomes must satisfy various constraints: replicating quickly, folding properly to interact with various proteins, and coding for several functional proteins within a constrained length. Consequently, mutations are usually deleterious and often catastrophic. Site-directed mutagenesis on several RNA and DNA viruses shows that the proportion of mutations that are lethal is 20–40%, and for nonlethal mutations the mean fitness reduction is about 10% (Sanjuán 2010). For this analysis, we assume that mutations are either lethal or neutral. Lethal mutants produce no offspring virions. The rate of lethal mutations $\mu$ is the rate of mutation multiplied by the proportion of lethal mutations. The overall mutation rate per genome per generation can be as low as 0.005 (e.g., bacteriophage M13), and as high as almost 1 (e.g., polio virus 1) (Sanjuán et al. 2010). Thus, the probability $\mu$ of lethal mutation for one generation can be taken anywhere from 0.001 to 0.5. The mutation load may be quite high under normal conditions, and some antiviral drugs act by increasing the mutation rate further (Anderson et al. 2004).
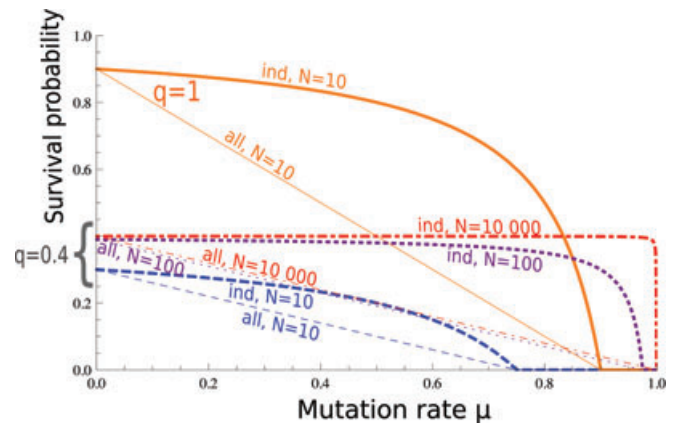
### SURVIVAL PROBABILITIES

For this scenario the basic reproductive number, that is, the mean number of viable virions produced in one generation by a virion, is $R_0 = q(1 - \mu)N$: a virion has a probability $q$ to infect a cell, which then releases an average of $N$ virions, a proportion $\mu$ of them being lethal mutants. If $\mu$ is large enough so that $R_0 \leq 1$, the virus goes extinct with certainty. However, even when $R_0 > 1$, the viral lineage may still go extinct due to stochastic effects. Solving for the survival probabilities for a viral lineage initiated with a viable virus yields

$$s_{all} = \max\left\{0, q(1-\mu) - \frac{1}{N}\right\}$$
$$= \max\left\{0, q(1-\mu)\left(1 - \frac{1}{R_0}\right)\right\} \text{ and,} \qquad (3)$$

$$s_{ind} = \max\left\{0, q - \frac{1}{N(1-\mu)}\right\} = \max\left\{0, q\left(1 - \frac{1}{R_0}\right)\right\}. \qquad (4)$$

For each mechanism, the second formulation can be easily understood: to ensure survival, a virion needs to infect a cell (probability $q$) and to produce a nonmutated template (probability $1 - \mu$ for the all-or-none mechanism, 1 for the independent mechanism).



**Figure 2.** Survival probabilities with lethal mutations. Results from equations (3) and (4). The survival probability of the viral lineage decreases when the lethal mutation rate $\mu$ increases, until reaching 0 when $R_0 \leq 1$. When all other parameters are equal, the survival probability is always higher for the independent mechanism.

Then the infected cell leads to a successful infection with a probability $1 - 1/R_0$ (a classic result for geometric distributions (Harris 1963)).

As seen from the first formulation of the survival probabilities and from Figure 2, the survival probability increases with $q$ and $N$ increasing and $\mu$ decreasing, as expected. Additionally, as easily seen from the second formulation of equations (3) and (4), $s_{ind}(1 - \mu) = s_{all}$. Viral survival is more likely with the independent mechanism than with the all-or-none mechanism, with equality when $\mu = 0$ (there is no difference between all-or-none and independent mechanisms if there is no mutation), or when $R_0 \leq 1$ (extinction is certain in any case).

This relation can be derived heuristically, as shown in Figure 3. At the first generation, the all-or-none template has undergone one round of mutation, whereas the independent template is nonmutated. At the second generation, all-or-none as well as independent templates have undergone an additional cycle of mutations, which are independent for each template. Both situations are similar, but it is as if the all-or-none lineage had started with an additional round of mutations. The survival probability with the all-or-none mechanism is equal to the survival probability with a modified independent mechanism, where the initial virion has undergone one cycle of mutations: $s_{all} = s_{ind}(1 - \mu) + \mu \times 0 = s_{ind}(1 - \mu)$, as shown previously. However, this simple relation cannot be extended to the case of two viable strains, which is considered below. The schematic in Figure 3 makes the assumption that $q$ and $N$ are equal for all viable strains. When we consider viable strains with different phenotypes, the templates may look similar but they are present in different numbers, so the relation derived here cannot be generalized.
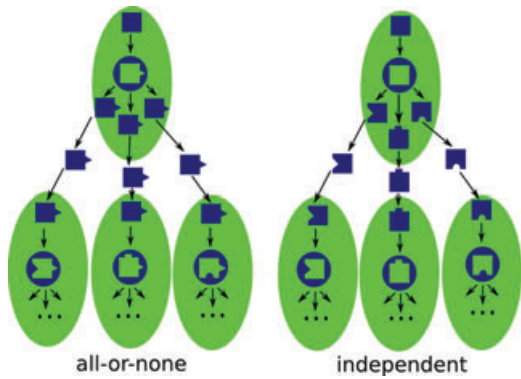
**Figure 3.** Schematic of the appearance of mutations with the all-or-none and independent mechanisms. The top green oval represents the first infected cell, and the blue square is the first infecting virion. The genome from the initial virus is copied into a template, which is then copied into new genomes that are packaged into new virions. These are released, and can infect new host cells, where the process repeats. The blue shape is modified each time a mutation could have occurred.

## DISTRIBUTION OF THE NUMBER OF VIRIONS PRODUCED

We analyze how the distribution of the number of virions differs for the two mechanisms (see Appendix S2.2). The average number of viable virions at generation $t$ is $\langle n(t) \rangle = R_0^t$. The variance (eq. 18 in Appendix S2) is always larger for the all-or-none mechanism.

The mean number of virions $R_0^t$ is an average across all lineages including those that go extinct. Often, however, it is of interest to focus on the instances when the viral lineage survived. The average number of virions at generation $t$ conditioned on survival is $\langle n_{alive}(t) \rangle = R_0^t/s(t)$, with $s(t)$ the survival probability at generation $t$. Most extinction happens in the first few generations, because the more generations pass without extinction, the greater the number of virions, and hence extinction becomes less likely (for more detailed discussion, see Appendix S2.1.2). Thus, $s(t)$ for large $t$ can be approximated by $s(\infty)$, and consequently $\langle n_{alive,ind}(t) \rangle \simeq (1 - \mu)\langle n_{alive,all}(t) \rangle$.

A viral lineage with the all-or-none mechanism is more likely to go extinct, but when it survives, it is as if the infection started $v = -\log(1 - \mu)/\log(R_0)$ generations earlier than with the independent mechanism. The expressions for the mean, the variance, and more generally, the full distribution of the probability to observe $n$ virions conditioned on the survival of the viral lineage, tend to the same value for the all-or-none and the independent mechanisms in the limit $t$ large, provided that we replace the number of generations $t$ by $t + v$ for the independent mechanism (see Appendix S2.2 and Fig. 4). However, these effects are quantitatively significant only if the virus is marginally fit in its environment ($R_0$ slightly larger than 1), or if the lethal mutation
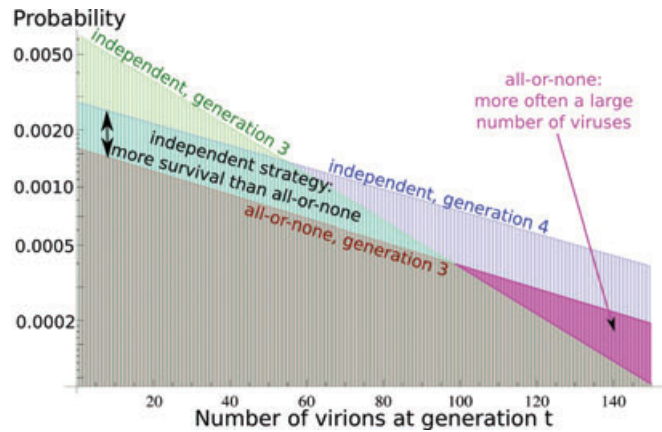


**Figure 4.** Distribution of the probability to observe $n$ viable virions at generation $t$ when mutations are lethal ($n = 0$ is not represented). At generation $t = 3$, observing a large number of virions is more likely with the all-or-none mechanism than with the independent mechanism. The parameters are $\mu = 0.5$, $N = 10$, $q = 0.4$, resulting in a head start for the all-or-none mechanism of $v = 1$ generations. Indeed, the distribution for the independent mechanism at the generation $t = 3 + v = 4$ is very similar to the distribution for the all-or-none mechanism at generation $t = 3$, except for a vertical shift that reflects the higher survival probability for the independent mechanism.

rate is high (i.e., $\mu$ is close to 1). In both of these cases, the virus is on the brink of survival, so stochastic effects matter more.

## ADAPTIVE EVOLUTION

The previous section considered viral dynamics when mutations are lethal for the virus. Now, we consider the more general case of mutations to a strain of nonzero fitness and their implications for viral adaptation. In particular, we focus on the phenomenon of evolutionary escape following an environmental shift, for example, a shift to a new host species or exposure to drug therapy, when the initial strain of the virus has a reproductive number $R_0 < 1$, and thus has to mutate to a fitter strain to persist in the new environment (Antia et al. 2003; Iwasa et al. 2004; André and Day 2005; Orr and Unckless 2008). We use the general generating functions (1) and (2) for a two-strain model, with $\mu_1$ the mutation rate from strain 1 to strain 2 and $\mu_2$ the mutation rate from strain 2 to strain 1. The model is a first step toward a more realistic fitness landscape, where several mutational steps are needed to reach the new optimal fitness (Iwasa et al. 2004; Shih et al. 2007; Weissman et al. 2009). The main results about viral survivorship and distributional dynamics are summarized below (see Appendix S3 for details).

## SURVIVAL PROBABILITIES

Explicit solutions for the survival probabilities cannot be found in general. To build an approximation, we express the survival
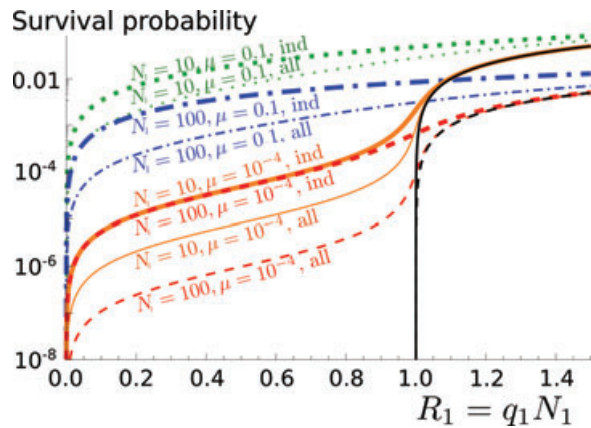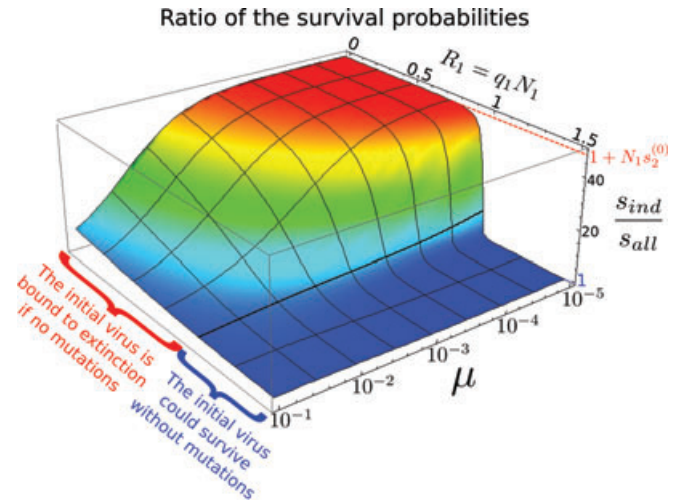
**Figure 5.** Survival probability for two strains, as a function of the basic reproductive number of the initial strain $R_1 = q_1 N_1$, where $R_1$ is changed by varying $q_1$ and keeping $N_1$ fixed. The mutant phenotype is $q_2 = 0.5$ and $N_2 = 100$. $\mu_1 = \mu_2 = \mu$. The exact solution cannot be distinguished from the approximation $s_1^{(1)}$. The black curves show $s_1^{(0)}$, the survival probability without mutations.



**Figure 6.** Ratio of the survival probabilities $s_{ind}/s_{all}$ as a function of the reproductive number of the initial strain, $R_1 = q_1 N_1$, and the mutation rate $\mu_1 = \mu_2 = \mu$, with $N_1 = 100$ and the mutant phenotype $q_2 = 0.5$ and $N_2 = 100$. Analysis of the small $\mu$ limit predicts that this ratio tends to 1 when $R_1 > 1$ and tends to $1 + N_1 s_2^{(0)}$ when $R_1 < 1$.

probability beginning from a virion of strain 1 as a function of the survival probability beginning from a virion of strain 2, and vice versa. Then it is possible to make iterative approximations, starting from the survival probabilities in the absence of mutations ($s_i^{(0)}$ when the viral infection starts with a virion of strain $i$). This process converges to the exact values of the survival probabilities (see Appendix S3.1.3). The first step of the iteration ($s_1^{(1)}$) corresponds to neglecting back mutations ($\mu_2 = 0$), an approximation often made in the evolutionary escape literature (e.g., Iwasa et al. (2004)) In the range of parameters explored, this first step of the iteration already gives a good approximation, indistinguishable from the exact solution (Fig. 5).

If the mutant has higher fitness than the initial strain, increasing the mutation rate increases survival, as expected (Fig. 5). However, the gain is most significant when the initial strain is unfit ($R_1 = N_1 q_1 < 1$), in which case mutations are needed to rescue the virus.

The survival probability of a viral lineage with the independent mechanism is greater than or equal to that of a lineage with the all-or-none mechanism, regardless of the fitness of the founding virus (see Appendix S3.1.6). This result is quite general: it is true in any parameter regime, for any number of viral types and any viral offspring distribution. The only restriction is that it relies on the assumption that $q$ depends only on the genome in the virion and $N$ depends only on the genome of the infecting virion.

The impact of the replication mechanism on survival can be understood intuitively. Under the all-or-none mechanism, beneficial mutations arise in clusters within the same lineage so their impact is redundant. With the independent mechanism, mutations tend to appear alone, so for the same total number of mutants more lineages are rescued (Fig. S3 in Appendix S3). Beneficial

mutations have greater effect when they are spread out than when they are clustered.

However, although the survival probability is always greater for the independent mechanism, the effect can vary greatly in magnitude. To evaluate it, we study the survival probability in the limit of small mutation rates (see Appendix S3.1.4). In circumstances where the initial strain could survive without mutations ($R_1 = q_1 N_1 > 1$), the difference of this development with the survival probability without mutations is only a relatively small correction. Thus the ratio of the survival probabilities for the all-or-none and independent mechanisms tends to one in this regime (Fig. 6). However, under conditions corresponding to evolutionary escape, when the initial strain cannot survive without mutations ($R_1 = q_1 N_1 < 1$) but the mutant may rescue it ($R_2 = q_2 N_2 > 1$), we find

$$s_{1,ind} \simeq \frac{\mu_1 q_1 N_1 s_2^{(0)}}{(1 - q_1 N_1)} \simeq \left(1 + N_1 s_2^{(0)}\right) s_{1,all}. \qquad (5)$$

In the evolutionary escape regime, the survival probability is higher for the independent than for the all-or-none mechanism by a factor of $1 + N_1 s_2^{(0)}$, a potentially large difference (Fig. 6). This factor can be understood heuristically in the limit $N_1 s_2^{(0)}$ large, which will be pertinent to most scenarios with $R_2 > 1$ as considered here. In this regime, a mutation occurring via the all-or-none mechanism gives rise to a cluster of mutants with an average size of $N_1$, and leads almost surely to the survival of the one lineage where the mutation occurs. With the independent mechanism, the equivalent number of mutants would be spread among $N_1$ distinct lineages instead of one, and each of these lineages would have a survival probability of approximately $s_2^{(0)}$. Thus, the

independent mechanism leads to $N_1 s_2^{(0)}$ surviving lineages instead of one.

## DISTRIBUTION OF THE NUMBER OF VIRIONS

The overall mean number of virions is the same for the all-or-none and the independent mechanisms, but the variance is larger for the all-or-none mechanism (see exact eq. 65 and associated discussion in Appendix S3). If we focus on the surviving lineages, as they are less numerous with the all-or-none mechanism than with the independent mechanism, each of them must include a larger number of virions.

## EXTENSION TO MULTISTEP EVOLUTIONARY TRAJECTORIES

We have discussed a simple scenario where a single mutation leads to a fit genotype. In many real systems, however, several mutations are needed for a virus to adapt to some challenge (Shih et al. 2007; Bloom et al. 2010). If we consider multistep trajectories, the independent mechanism will still lead to greater viral survival than the all-or-none mechanism, but the ratio of the survival probabilities will be dominated by the mutational steps from unfit to fit strains. Indeed, in the limit of low mutation rates, the survival of fit strains is not affected by mutations, and steps between unfit strains involve very low survival probabilities, so the clustering of the all-or-none mutants has negligible effect (see Appendix S3.3). The two-strain case presented above provides the essential building block to explore more complex evolutionary trajectories.

## *Discussion*

Viruses replicate in infected cells using various mechanisms, leading to different distributions of the number of offspring virions that bear mutations. This basic virological phenomenon is often overlooked when modeling viral adaptation. To investigate whether and when this matters, we analyzed a stochastic model of early infection within a host for a given mean mutation rate with the two limiting replication mechanisms, all-or-none (mutations occurring at the beginning of replication, such that all offspring virions from an infected cell have the same genotype, whether mutated or not) and independent (mutations occurring toward the end of the replication, consequently independent for each new virion). Under most circumstances, the mean mutation rate is enough to describe the dynamics, so simple model formulations are sufficient. However, when mutations are crucial for viral survival, the details of replication can substantially influence the evolutionary dynamics.

## WHEN DOES REPLICATION MECHANISM MATTER?

Mutations can be either beneficial or deleterious. Although both types of mutations are simultaneously involved in viral survival (Alexander and Day 2010), we focused for simplicity on two cases—lethal mutations and mutations between two viable strains—and proved that in both cases the survival probability is always smaller for the all-or-none mechanism. As the overall mean number of virions produced is the same for both mechanisms, the surviving viral lineages produce more virions with the all-or-none than with the independent mechanism, as if they had a head start, and the overall variance in the number of virions produced is larger. Heuristically, more variation means a greater risk of extinction of the lineage. If mutations are lethal, their effects when spread out are buffered by the presence of viable virions. If the mutations are beneficial, their effective frequency is smaller when clustered, so fewer lineages receive the survival benefit. These effects could be summarized by the adage "don't put all your eggs in one basket." The viral lineage does better by hedging its bets, in the sense that reducing variation in individual fitness increases population level fitness. The benefit of reducing variation recapitulates a classic theme in population genetics (Gillespie 1974; Frank and Slatkin 1990) and life-history evolution (Cohen 1966; Childs et al. 2010) but also appears in analyses of viral invasions at multiple scales (Lloyd-Smith et al. 2005; Pearson et al. 2011).

Beyond proving that the survival probability is always smaller for the all-or-none mechanism, we determined when the difference is quantitatively significant. For mutations between viable strains, the difference can be large when a strain not fit enough to survive without mutations ($R_0 < 1$) is rescued by a fitter mutant ($R_0 > 1$). For lethal mutants, survival probabilities significantly differ when the viable strain's $R_0$ is close to 1 or when the lethal mutation rate is high. The common property of these regimes is that mutations play a crucial role in survival, and consequently stochasticity is important. Indeed, we did not expect significant differences otherwise, because deterministic models accounting for mean behavior only are equivalent for the two replication mechanisms.

These guidelines show when viral replication mechanisms have significant effects on evolutionary dynamics, and hence determine whether these mechanisms should be considered in models addressing various problems. For instance, our analysis of lethal mutations could be applicable to the study of lethal mutagenesis (i.e., increase of the mutation rate to levels where the amount of deleterious mutations is high enough to threaten the viability of the viral population), which is thought to be the mechanism of action of some antiviral drugs (Anderson et al. 2004). These drugs are often applied when the number of viruses is already very large, justifying the use of models considering the mean mutation rate only (Bull et al. 2007; Martin and Gandon 2010). However, in some settings, such as the prophylactic use of mutagenic drugs, or their use in combination with other drugs that reduce viral population size (Iranzo et al. 2011), stochastic effects may

be important so consideration of replication mechanism will be appropriate.

Our analysis of adaptive mutations focused on the dynamics of evolutionary escape, for which stochastic effects are always important. We thus consider how assumptions in existing models relate to the mechanisms considered here. In the theoretical literature on evolutionary escape, the infected unit can be a cell within a host, or more often a host within a population (where each host is assumed to be infected with a single genotype of virus). Mutations are most often modeled as occurring at the moment of transmission, which is equivalent to independent mutants in our model (Antia et al. 2003; Iwasa et al. 2004). Mutations have also been considered as occurring during the course of infection within a host, and sweeping instantaneously to fixation (André and Day 2005). Then all subsequent secondary infections caused by this host are of the mutant strain, which is closer to the all-or-none mechanism. However, in most real systems the frequency of a mutation rises more slowly within a host, so hosts are simultaneously infected with numerous viral strains. If the viral population bottleneck at transmission is small (i.e., very few viruses successfully pass from one host to the next), the same infected host can transmit different strains to different secondary hosts, analogous to the independent mechanism. By the arguments presented here, such models assuming immediate mutation sweeps may underestimate the probability of evolutionary escape. An important goal for future work is to develop models for viral evolution at the host population scale that account for the complexity of viral evolution within host individuals.

## ROBUSTNESS OF CONCLUSIONS TO ALTERNATIVE LIFE HISTORIES

Our model describes the phenotype of each strain by two parameters: $q$, the probability for a virion to successfully infect a cell, and $N$, the mean number of virions produced by an infected cell. We assumed the simplest genotype–phenotype map: $q$ takes the value associated with the genome inside the virion, and $N$ depends on the genotype of the virion that entered the cell. However, virions are composed of genomic material packaged with proteins drawn from a pool of proteins translated from viral genomes in the parent cell. When a mutation occurs, the genotype–phenotype map can be blurred for the first generation, because the viral genomes in the parent cell are not all of the same genotype. In the case of lethal mutations, this may slightly alter the timing of when the lethal trait is expressed, but the mutant lineage is doomed in any case, so the exact assumptions about $q$ and $N$ do not change the results. In the case of beneficial mutants, these details may modify the comparison between the all-or-none and independent mechanisms, but in all cases our core result holds that the regime where modelers should account for details of the replication process is when mutations are crucial for survival.

We have studied the limiting cases of mutants among offspring virions that are either independent from each other (independent mechanism) or completely correlated (all-or-none mechanism). Our model is based on the stamping machine, in which a template is made from the genome of the initial virion and is then copied to produce new genomes, with the mutation rate of one of these steps considered negligible compared to the other. In real stamping machine processes mutations occur in both steps. For viruses that copy their genome by binary replication (Luria 1951), a mutation in the first round of genome duplication causes half of the virions produced by the cell to be mutants, approaching the all-or-none case, whereas a mutation in the last round of duplication is equivalent to an independent mutation. In some viruses, replication could combine binary and stamping machine mechanisms (Chao et al. 2002). In all these situations, the correlation between mutants falls in between the all-or-none and independent mechanisms. Our study of these limits enables us to identify the regimes where these biological differences are significant.

A cell successfully infected by a virus releases new virions, in numbers which can follow different distributions (Pearson et al. 2011). In the main text, we assume that infected cells release virions and die at fixed rates, resulting in a geometric distribution of the number of virions released. But our main conclusions about the comparison of the all-or-none versus independent mechanisms remain valid if cells release exactly $N$ virions (see Appendix S1 for lethal mutations and Appendix S3 for the two-strain case). These two cases are thought to describe the majority of realistic viral life histories (Pearson et al. 2011).

We also assume that the fates of separate virions are independent of each other. Experimentally, this does not hold in all cases (Vignuzzi et al. 2006; Zwart et al. 2009). This could arise from the host immune response, competition for a limited pool of susceptible host cells, or infection of host cells by multiple viral strains. However, these interactions are likely to require a large number of virions, whereas our model is focused on the stochastic effects during early infection, when there are few viruses. In the longer term, the viral load may be limited by the depletion of susceptible cells or the immune response (Saenz et al. 2010). If most transmission between hosts happens late in the infection, when the viral load is saturated, what would matter most is whether the initial challenge dose led to a full blown infection or went extinct in the early stochastic infection phase. Thus in this case, the all-or-none mechanism leads to less transmission.

## POSSIBLE EXPERIMENTS

While the empirical literature on viral mutation rates is growing (Sanjuán et al. 2010), the distribution of mutants numbers is very infrequently characterized. Traditionally cloning studies have been the sole approach to this problem (Luria 1951; Denhardt and Silver 1966; Chao et al. 2002), but these are labor

intensive, need a visible phenotypic effect of mutations, and are complicated by mutations during the amplification process. However, the rapid advance of sequencing technologies may allow for easier characterization of the distribution of mutations in the near future (Wright et al. 2011).

If characterizing the distribution of mutants were easier, we could test our prediction that—all else equal—correlated mutants lead to lower survival probability for a viral lineage than independent mutants. The survival probability of a virus introduced in a cell culture could be measured either in the presence of lethal mutations or in the evolutionary escape regime starting from a strain with one nonlethal but highly deleterious mutation. As many factors can obscure comparative studies among viruses, a better approach would be to study one virus using the stamping machine mechanism with the possibility to manipulate the mutation rates of each step independently. Retroviruses are ideal candidates for these experiments, as the two clearly identified replication steps use two different enzymes. Mutations in the retrotranscriptase gene are known to influence error rates (Svaroskaia et al. 2003), and mutagenic drugs could be used to manipulate mutation rates further. Subject to the ability to confirm that other parameters have not been affected by these manipulations, such an experimental setup could be used to test the core predictions of our analysis. The results, together with better characterization of mutant distributions for viruses of interest, would advance our knowledge of how, and when, viral replication strategies affect evolutionary dynamics.

## LITERATURE CITED

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

Ali, A., H. Li, W. L. Schneider, D. J. Sherman, S. Gray, D. Smith, and M. J. Roossinck. 2006. Analysis of genetic bottlenecks during horizontal transmission of Cucumber mosaic virus. J. Virol. 80:8345–8350.

Anderson, J., R. Daifuku, and L. Loeb. 2004. Viral error catastrophe by mutagenic nucleosides. Annu. Rev. Microbiol. 58:183–205.

André, J.-B. and T. Day. 2005. The effect of disease life history on the evolutionary emergence of novel pathogens. Proc. R. Soc. Lond. B 272:1949–1956.

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

Bloom, J. D., L. I. Gong, and D. Baltimore. 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. Science 328:1272–1275.

Bull, J. J., R. Sanjuán, and C. O. Wilke. 2007. Theory of lethal mutagenesis for viruses. J. Virol. 81:2930–2939.

Chao, L., C. U. C. Rang, and L. E. L. Wong. 2002. Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage ϕ 6. J. Virol. 76:3276.

Chen, H. Y., M. Di Mascio, A. S. Perelson, D. D. Ho, and L. Zhang. 2007. Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques. Proc. Natl. Acad. Sci. USA 104:19079–19084.

Childs, D. Z., C. J. E. Metcalf, and M. Rees. 2010. Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants. Proc. R. Soc. Lond. B 277:3055–3064.

Cohen, D. 1966. Optimizing reproduction in a randomly varying environment. J. Theor. Biol. 12:119–129.

de Mercoyrol, L., Y. Corda, C. Job, and D. Job. 1992. Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed B-DNA to left-handed Z-DNA. FEBS J. 206:49–58.

De Paepe, M., S. De Monte, L. Robert, A. B. Lindner, and F. Taddei. 2010. Emergence of variability in isogenic Escherichia coli populations infected by a filamentous virus. PLoS One 5:e11823.

Denhardt, D. T. and R. B. Silver. 1966. An analysis of the clone size distribution of ΦX174 mutants and recombinants. Virology 30:10–19.

Domingo, E. 2010. Mechanisms of viral emergence. Vet. Res. 41:38.

Frank, S. and M. Slatkin. 1990. Evolution in a variable environment. Am. Nat. 136:244–260.

French, R. and D. C. Stenger. 2003. Evolution of wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. Annu. Rev. Phytopathol. 41:199–214.

Gillespie, J. 1974. Natural selection for within-generation variance in offspring number. Genetics 76:601–606.

Harris, T. E. 1963. The theory of branching processes. Dover Phoenix Editions, Mineola, NY.

Holmes, E. C. and A. J. Drummond. 2007. The evolutionary genetics of viral emergence. Curr. Top. Microbiol. Immunol. 315:51–66.

Iranzo, J., C. Perales, E. Domingo, and S. C. Manrubia. 2011. Tempo and mode of inhibitor–mutagen antiviral therapies : A multidisciplinary approach. Proc. Natl. Acad. Sci. USA 108:16008–16013.

Iwasa, Y., F. Michor, and M. A. Nowak. 2004. Evolutionary dynamics of invasion and escape. J. Theor. Biol. 226:205–214.

Kamp, C. 2003. A quasispecies approach to viral evolution in the context of an adaptive immune system. Microbes Infection 5:1397–1405.

Kearney, M., F. Maldarelli, W. Shao, J. B. Margolick, E. S. Daar, J. W. Mellors, V. Rao, J. M. Coffin, and S. Palmer. 2009. Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. J. Virol. 83:2715–2727.

Keele, B. F., H. Li, G. H. Learn, P. Hraber, E. E. Giorgi, T. Grayson, C. Sun, Y. Chen, W. W. Yeh, N. L. Letvin, et al. 2009. Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. J. Exp. Med. 206:1117–1134.

Kew, O. M., R. W. Sutter, E. M. de Gourville, W. R. Dowdle, and M. A. Pallansch. 2005. Vaccine-derived polioviruses and the endgame strategy for global polio eradication. Annu. Rev. Microbiol. 59:587–635.

Kim, T., R. A. Mudry, C. A. Rexrode, and V. K. Pathak. 1996. Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. J. Virol. 70:7594–7602.

Kireeva, M. L., Y. A. Nedialkov, G. H. Cremona, Y. A. Purtov, L. Lubkowska, F. Malagon, Z. F. Burton, J. N. Strathern, and M. Kashlev. 2008. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. Mol. Cell 30:557–566.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz. 2005. Super-spreading and the effect of individual variation on disease emergence. Nature 438:355–359.

Luria, S. 1951. The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. Cold Spring Harbor symposia on quantitative biology. Vol. 16, P. 463. Cold Spring Harbor Laboratory Press. Available at: http://symposium.cshlp.org/content/16/463.short.

Maisnier-Patin, S. and D. I. Andersson. 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res. Microbiol. 155:360–369.

Manrubia, S. C., E. Domingo, and E. Lázaro. 2010. Pathways to extinction: beyond the error threshold. Philos. Trans. R. Soc. Lond. B 365:1943–1952.

Martin, G. and S. Gandon. 2010. Lethal mutagenesis and evolutionary epidemiology. Philos. Trans. R. Soc. Lond. B 365:1953–1963.

Moya, A., E. C. Holmes, and F. González-Candelas. 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev. Microbiol. 2:279–288.

O'Neil, P. K., G. Sun, H. Yu, Y. Ron, J. P. Dougherty, and B. D. Preston. 2002. Mutational analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. J. Biol. Chem. 277:38053–38061.

Orr, H. and R. Unckless. 2008. Population extinction and the genetics of adaptation. Am. Nat. 172:160–169.

Parada, V., G. J. Herndl, and M. G. Weinbauer. 2006. Viral burst size of heterotrophic prokaryotes in aquatic systems. J. Mar. Biol. Assoc. U.K. 86:613.

Pearson, J. E., P. Krapivsky, and A. S. Perelson. 2011. Stochastic theory of early viral infection: continuous versus burst production of virions. PLoS Comput. Biol. 7:e1001058.

Pepin, K. M., S. Lass, J. R. C. Pulliam, A. F. Read, and J. O. Lloyd-Smith. 2010. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nat. Rev. Microbiol. 8:802–813.

Pybus, O. G. and A. Rambaut. 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat. Rev. Genet. 10:540–550.

Saenz, R. A., M. Quinlivan, D. Elton, S. Macrae, A. S. Blunden, J. A. Mumford, J. M. Daly, P. Digard, A. Cullinane, B. T. Grenfell, et al. 2010. Dynamics of influenza virus infection and pathology. J. Virol. 84:3974–3983.

Sanjuán, R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos. Trans. R. Soc. Lond. B 365:1975–1982.

Sanjuán, R., M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw. 2010. Viral mutation rates. J. Virol. 84:9733–9748.

Sardanyés, J., R. V. Solé, and S. F. Elena. 2009. Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. J. Virol. 83:12579–12589.

Shaw, R. J., N. D. Bonawitz, and D. Reines. 2002. Use of an in vivo reporter assay to test for transcriptional and translational fidelity in yeast. J. Biol. Chem. 277:24420–24426.

Shih, A. C.-C., T.-C. Hsiao, M.-S. Ho, and W.-H. Li. 2007. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. Proc. Natl. Acad. Sci. USA 104: 6283–6288.

Svaroskaia, E. S., S. R. Cheslok, W. H. Zhang, W. S. Hu, and V. K. Pathak. 2003. Retroviral mutation rates and reverse transcriptase fidelity. Front. Biosci. 8:d117–d134.

Thébaud, G., J. Chadoeuf, M. J. Morelli, J. W. McCauley, and D. T. Haydon. 2010. The relationship between mutation frequency and replication strategy in positive-sense single-stranded RNA viruses. Proc. R. Soc. Lond. B 277:809–817.

Todd, S., J. S. Towner, and B. L. Semler. 1997. Translation and replication properties of the human rhinovirus genome in vivo and in vitro. Virology 229:90–97.

Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439:344–348.

Volkov, I., K. M. Pepin, J. O. Lloyd-Smith, J. R. Banavar, and B. T. Grenfell. 2010. Synthesizing within-host and population-level selective pressures on viral populations: the impact of adaptive immunity on viral immune escape. J. R. Soc. Interface 7:1311–1318.

Wang, G. P., S. A. Sherrill-Mix, K.-M. Chang, C. Quince, and F. D. Bushman. 2010. Hepatitis C virus transmission bottlenecks analyzed by deep sequencing. J. Virol. 84:6218–6228.

Weissman, D. B., M. M. Desai, D. S. Fisher, and M. W. Feldman. 2009. The rate at which asexual populations cross fitness valleys. Theor. Popul. Biol. 75:286–300.

Wilhelm, S. W. and A. R. Matteson. 2008. Freshwater and marine virioplankton: a brief overview of commonalities and differences. Freshwater Biol. 53:1076–1089.

Wright, C. F., M. J. Morelli, G. Thébaud, N. J. Knowles, P. Herzyk, D. J. Paton, D. T. Haydon, and D. P. King. 2011. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. J. Virol. 85: 2266–2275.

Zwart, M. P., L. Hemerik, J. S. Cory, J. A. G. M. de Visser, F. J. J. A. Bianchi, M. M. Van Oers, J. M. Vlak, R. F. Hoekstra, and W. Van der Werf. 2009. An experimental test of the independent action hypothesis in virus-insect pathosystems. Proc. R. Soc. Lond. B 276:2233–2242.

Associate Editor: S. Gandon

## Supporting Information

The following supporting information is available for this article:

**Appendix S1:** Lethal mutants (burst model)
**Appendix S2:** Lethal mutants (budding model)
**Appendix S3:** Adaptive evolution

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.