Appendix S1: Alternative mutation models.

***a. Constant mutation model*:** This is the classic approach where mutation is a simple Poisson process. Let $k(t,\mathbf{g})$ be the total number of mutations produced by all individuals of the genotype class **g**, over some time period $[0,t]$. $k(t,\mathbf{g})$ is distributed as a Poisson $k(t,\mathbf{g}) \sim \mathcal{P}(U\int_0^t I(t,\mathbf{g})\,dt)$. From the additivity of the Poisson distribution, the total number of mutations produced by the population is also Poisson distributed,

$$k_T(t) = \sum_{\mathbf{g}} k(t,\mathbf{g}) \sim \mathcal{P}(U\int_{\tau=0}^{t} I_T(\tau)\,d\tau). \quad . \quad \text{(A1.1)}$$

Recalling that $I_T(\tau) = \sum_{\mathbf{g}} I(\tau,\mathbf{g})$ is the total density of infected cells at any given time τ. Our derivations directly apply under this model, using the genomic mutation rate U. The total production of new mutations by the population is therefore a Poisson process with rate $\lambda(t) = \partial_t E(k_T(t))$, where $E(.)$ denotes expectation. Because any individual may produce mutant offspring in this model, the corresponding per capita rate of mutation is simply $\lambda(t)/I_T(t) = U$, from (A1.1).

b. infection – dependent mutation model: by contrast, in this model, mutations occur on a given virus conditional on this virus undergoing an infectious cycle. This time, the total number of mutations produced by genotype class **g** becomes $k(t,\mathbf{g}) \sim \mathcal{P}(\mu\, n_I(t,\mathbf{g}))$, where $n_I(t,\mathbf{g})$ is the total number of infections undergone by individuals of the genotype class **g** during the period $[0,t]$. By analogy with the *constant mutation* model, the rate of mutation specific to genotype class **g** up to time t is directly proportional to the rate of new infections by genotypic class **g**:

$$\lambda(t,\mathbf{g}) = \partial_t E(k(t,\mathbf{g})) = \mu\, E(n_I(t,\mathbf{g})) = \mu\, \beta(\mathbf{g}) S(t)\, I(t,\mathbf{g}) \quad . \quad \text{(A1.2)}$$

We may now use results from the theory of spatial Poisson processes, applied to the phenotypic space with space variable **g**. The process of mutation over the whole region of genotypic values **g** spanned by the virus population is also a Poisson process with rate equal to the sum of the rates over space (across genotype classes). The production of new mutations by the whole population is therefore a Poisson process with rate

$$\lambda_e(t) = \sum_{\mathbf{g}} \lambda(t,\mathbf{g}) = \mu\, S(t) \sum_{\mathbf{g}} \beta(\mathbf{g})\, I(t,\mathbf{g}) = \mu\, S(t)\, \bar{\beta}(t)\, I_T(t) \quad . \quad \text{(A1.3)}$$

Simulations (not shown) show that the total rate of production of new mutations is indeed given by $\lambda_e(t)$ when mutation is conditional on infection.

Finally, to compute the corresponding per – capita rate of mutation, we must account for the fact that, contrary to the *constant mutation* model, only those individuals that

generate new infections ("reproducers") can mutate. The number of such individuals is $I_T(t)\,P_I$ where P_I is the proportion of individuals that are currently undergoing an infection at time t. This proportion is given by the probability of an infected cell generating a new infection (with total rate $\bar{\beta}(t)S(t)I_T(t)$), as opposed to dying (with total rate $v\,I_T(t)$). Based on the expression in **(M1. 3)**, and focusing only on events occurring to an infected cell here, we thus have

$$P_I = \frac{\bar{\beta}(t)S(t)}{\bar{\beta}(t)S(t) + v} \quad . \qquad \text{(A1.4)}$$

The equivalent mutation rate (per individual producing mutants) is therefore

$$U_e(t) \equiv \frac{\lambda_e(t)}{P_I\,I_T(t)} = \mu\,(S(t)\,(t) + v) = \mu\,(\bar{r}(t) + 2v) \quad . \qquad \text{(A1.5)}$$

It is this mutation rate that must be used to predict the evolutionary dynamics with infection – dependent mutation; at demographic equilibrium ($\bar{r}_* = 0$), its value is simply $U_e = 2\,\mu\,v$.

Appendix S2: Equilibrium distribution of virus phenotypes, and mean transmission rate.

In this appendix we use previous results from Lande (1980) to give the equilibrium distribution of the multivariate genotypic values (**g**) at mutation – selection balance (evolutionary equilibrium), assuming a given (unknown) epidemiological equilibrium. From these results, we derive the mean fitness and transmission rate of the viral population for a given demographic state equilibrium. For the sake of simplicity, we first ignore 'true lethal' mutants ($\beta = 0$) to deal only with continuous fitness variation among mutants (***a.*** and ***b.***), and then include those true lethals (**c.**). We then use these result to derive the full epidemiological and evolutionary equilibrium state of the population: mean transmission rate ($\bar{\beta}_*$), total density of infected I_{T*} and susceptible cells S_* (***d.***).

a. Equilibrium distribution of viral phenotypic traits and transmission rate: We have seen (eq.**(4)**) that under our assumptions, Malthusian fitness is a multivariate quadratic function of **g**, which corresponds to the classic multivariate Gaussian function for absolute fitness ($r(\mathbf{g}) = \log W(\mathbf{g})$). Lande (1980) derived the dynamics of the distribution of **g** in this type of fitness landscape, under weak stabilizing selection and strong mutation, with normally distributed pleiotropic effects on all traits in vector **g** (with covariance **M**). The assumptions correspond to our model and context: weak selection was assumed to retrieve the quadratic function in eq. **(4)**, Gaussian effects of mutations on **g** was assumed, and the high mutation rate assumption is a priori valid in viruses. Note that the Lande model describes multiple unlinked loci, under the assumption of a high mutation rate per locus (Turelli, 1984). However, as we assume asexuality here, the whole genome is modeled as a single locus and the criterion of high mutation is a per genome high mutation rate, roughly $U >> E(s)$ among random mutants. In this case the equilibrium distribution of **g** is a multivariate Gaussian. Its mean at the optimum (set to $\mathbf{g} = \mathbf{0}$ without loss of generality) and its genetic variance – covariance matrix $\mathbf{G}_*$ is the solution to the equation $U\mathbf{M} = \mathbf{G}_*.\boldsymbol{\Sigma}.\mathbf{G}_*$, where $\boldsymbol{\Sigma}$ is assumed a constant.

At this stage, we must make two important points. First, in our context $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_t = S(t)\beta_o\boldsymbol{\Sigma}_\beta$ is time dependent in general (eq. **(4)**), but at epidemiological and evolutionary equilibrium it is not, as $S(t)$ sets to some unknown equilibrium value $S(t) = S_*$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_* = S_*\beta_o\boldsymbol{\Sigma}_\beta$. Second, it is always possible to consider a space basis for **g** where $\boldsymbol{\Sigma}_\beta = \mathbf{I}_n$ is the identity matrix in *n* dimensions (or $\boldsymbol{\Sigma}_* = S_*\beta_o\mathbf{I}_n$), and **M** is a diagonal matrix. Individual trait definitions in **g** are irrelevant to our problem, so we may consider any orthonormal basis for the space of genotypic values **g**. In this basis, the equilibrium covariance verifies $U\mathbf{M} = S_*\beta_o\mathbf{G}_*^2$ and thus the distribution of genetic values is an uncorrelated multivariate Gaussian:

$$\mathbf{g} \sim N\left(\mathbf{0}, \mathbf{G}_* = \sqrt{\frac{U}{S_*\beta_o}}\mathbf{M}^{1/2}\right) \quad , \tag{A2.1}$$

where the ½ exponent denotes matrix square root. The whole distribution of the transmission rate $\beta(\mathbf{g})$ can then easily be retrieved using eq.**(1)**: it is a quadratic form in Gaussian vectors ($\mathbf{g}$), well studied in statistics (Mathai and Provost, 1992). In particular its mean is (recalling that in our basis $\mathbf{\Sigma}_\beta = \mathbf{I}_n$)

$$\bar{\beta}_* = \beta_o - \frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}_\beta . \mathbf{G}_*\right) = \beta_o - \frac{1}{2}\sqrt{\frac{U}{S_*\beta_o}}\ \mathrm{tr}\left(\mathbf{M}^{1/2}\right) \quad . \qquad \text{(A2.2)}$$

b. Link with mutation fitness effects distributions in exponential phase: The above results are based on non measurable landscape parameters. By contrast, we have seen that what is typically available empirically is the distribution of fitness among single mutants, in exponential phase, which corresponds to setting $S_* = S_{max}$, and selective covariance matrix $\beta_o\, S_{max}\mathbf{I}_n$ in our diagonal basis. From their definition in eq.**(5)**, the selection coefficients of mutations $s(\boldsymbol{\varepsilon}_\mathbf{g})$ from an optimal parent genotype in this exponential phase is also a quadratic form of the vectors of mutation effects on $\mathbf{g}$, $\boldsymbol{\varepsilon}_\mathbf{g} \sim N(\mathbf{0}, \mathbf{M})$. The distribution of such a quadratic form is a gamma if and only if $\mathbf{M}$ is proportional to identity, which is the rationale for the following equivalent landscape approximation (see details in Martin and Lenormand, 2006). Matrix $\mathbf{M}$ is diagonal with n distinct diagonal terms d_i. Now, to express these non measurable parameters into measurable quantities, we approximate the real landscape by an equivalent one where s is indeed gamma, i.e. with only $n_e < n$ eigenvalues that are all equal to some constant d_e. The approximation is obtained by choosing n_e and d_e that match the two first moments of the gamma distribution. In this equivalent landscape, with $\mathbf{M} \approx d_e\mathbf{I}_{n_e}$, the distribution of mutation effects on fitness $s(\boldsymbol{\varepsilon}_\mathbf{g})$ is then a negative gamma with shape $\alpha = 1/CV^2(s) = n_e/2$ and absolute mean $\bar{s} = E(|s|) = \mathrm{tr}(\mathbf{\Sigma}_{max}.\mathbf{M}) = \beta_o\, S_{max}\, n_e d_e/2$. These two parameters being available empirically, we will use the equivalence $n_e = 2\alpha$ and $d_e = \bar{s}/(\beta_o\, S_{max}\alpha)$.

Now, still in this equivalent landscape, the equilibrium genotypic covariance matrix becomes $\mathbf{G}_* \approx \sqrt{U/(S_*\beta_o)\, d_e}\mathbf{I}_{n_e}$ from eq. **(A2.1)** and $\mathbf{M} \approx d_e\mathbf{I}_{n_e}$. Using the conversion from (d_e, n_e) to $(\bar{s}, \alpha)$, the mean transmission rate can be expressed approximately:

$$\bar{\beta}_v \approx \beta_o - \frac{1}{2}\sqrt{\frac{U}{S_*}}\left(\sum\nolimits_1^{n_e}\sqrt{d_e}\right) = \beta_o - \sqrt{\frac{U\,\bar{s}\,\alpha}{S_{max}\,S_*}} \quad . \qquad \text{(A2.3)}$$

The subscript v in $\bar{\beta}_v$ is for viable genotypes (see below in ***c.***). The corresponding mean Malthusian fitness of the population (viable genotypes) is then given by (from eq.**(4)**)

$$\bar{r}_v = S_*\bar{\beta}_v - v = r_o(t) - \sqrt{\frac{S_*}{S_{max}}\ U\,\bar{s}\,\alpha} \quad , \qquad \text{(A2.4)}$$

where we recall that $r_o(t) = \beta_o S_* - v$ is the fitness of the optimal genotype in the current conditions (at $S = S_*$).

c. Including 'true lethal' mutants: So far we have only considered continuous fitness variation among mutants. However, in the full model a proportion p_L of mutations are 'true lethals' in that they form a qualitatively distinct class of mutants with zero transmission (they can be seen as having infinite effect on genotypic value: $|\boldsymbol{\varepsilon_g}| \to \infty$). We can deal with this class as a separate system: the subpopulation of viable genotypes with transmission rate $\bar{\beta}_v$ and mean fitness $\bar{r}_v$ given by eqs. **(A2.3)** and **(A2.4)**, and the subpopulation of 'true lethals' with transmission rate $\beta_{nv} = 0$ and corresponding mean fitness $r_{nv} = -v$. The subscript nv here stands for 'non viable'. The evolutionary dynamics of competition and mutation between these two populations are very simple: there are only two classes with selection and unidirectional mutation from the viable to the non viable class with rate Up_L. The effects of selection and mutation on the frequency f_L of true lethals and $(1 - f_L)$ of viables are computed from the equation of selection in continuous time models (Rice, 2004). At equilibrium, if the mean fitness of the whole population is $\bar{r}_*$, we get:

$$\begin{aligned} \partial_t f_L &= f_L(r_{nv} - \bar{r}_*) + U\, p_L(1 - f_L) = 0 \\ \partial_t(1 - f_L) &= -\partial_t f_L = (1 - f_L)(\bar{r}_v - \bar{r}_*) - U\, p_L(1 - f_L) = 0 \end{aligned} \quad . \qquad \text{(A2.5)}$$

This equation admits two possible equilibrium solutions:

$$\begin{aligned} &(E1):\ f_L = 1 \text{ and } \bar{r}_* = -v \\ &(E2):\ f_L = U\frac{p_L}{v + \bar{r}_v} \text{ and } \bar{r}_* = \bar{r}_v - U\, p_L \end{aligned} \quad . \qquad \text{(A2.6)}$$

Equilibrium $(E1)$ is akin to a form of "error catastrophe" as in quasispecies theory (Eigen, 1971), where the fit class (here the viable class) fully disappears under high mutation pressure ($f_L = 1$, all genotypes are lethals). Equilibrium $(E2)$ corresponds to a population in polymorphic state with a class of true lethals at some equilibrium frequency $(0 < f_L < 1)$. The transition from $(E1)$ to $(E2)$ occurs when $f_L = U\, p_L/(v + \bar{r}_v) = 1$ which means that this error catastrophe occurs whenever $U\, p_L \geq \bar{r}_v + v$. Note that extinction will occur deterministically whenever the mean growth rate is negative, $\bar{r}_* = \bar{r}_v - U\, p_L < 0$ (equilibrium $E2$), while the error catastrophe occurs only after $\bar{r}_* = \bar{r}_v - U\, p_L \leq -v$ which is below this first threshold. Therefore we believe that equilibrium $E1$ bears little biological relevance because demographic extinction will occur before error catastrophe, as U increases, and we do not study it any further. Under the polymorphic equilibrium $(E2)$, the mean transmission is given by $\bar{\beta}_* = (1 - f_L)\bar{\beta}_v$, with $f_L$deduced from $\bar{r}_v = S_*\bar{\beta}_v - v$. The mean transmission rate in the full model is then obtained by rearranging eq.**(A2.3)**, with U replaced by $U(1 - p_L)$ the mutation rate to viable genotypes:

$$\bar{\beta}_* \approx \beta_o - \sqrt{\frac{U(1 - p_L)\,\bar{s}\,\alpha}{S_{max}\,S_*}} - \frac{Up_L}{S_*} \quad . \qquad \text{(A2.7)}$$

d. Demographic and evolutionary equilibrium for transmission and cell densities: We can now jointly solve the demographic and evolutionary equilibrium, in the full model with a portion p_L of 'true lethal' mutations. At demographic equilibrium, the susceptible host cell density verifies $S_* = v/\bar{\beta}_*$, which, once plugged into eq. **(A2.7)**, gives two solutions for the the equilibrium transmission rate. Only one of them is biologically sound, giving $\bar{\beta}_* \leq \beta_o$. We thus get the equilibrium transmission in terms of the basic parameters, as given in eq. **(6)**:

$$\bar{\beta}_* \approx \frac{\beta_o v}{v + U p_L}\left(1 + x - \sqrt{x(2+x)}\right) \quad , \quad \text{(A2.8)}$$

where $x = \frac{U\,\bar{s}\,\alpha\,(1-p_L)}{2(v+Up_L)(v+r_o)}$ is a composite parameter that increases with U. The corresponding equilibrium cell densities are given directly in the main text (eq.**(7)**).

Appendix S3: Inferring the proportion of apparent and true lethals from single mutant fitness distributions.

We have seen that a key parameter determining the efficacy of a treatment by lethal mutagenesis is the proportion p_L of 'true lethals' among all random mutations, which determines both the equilibrium titer (eqs. **(7)** and **(9)**) and the extinction threshold for U_c (eq. **(10)**). However, the proportion of lethals obtained empirically need not be equal to p_L, as some genotypes can be 'apparent lethals', with a non-zero transmission but a negative (undetectable) growth in exponential phase. This latter class of lethals is already accounted for in the model with continuous variation in transmission rate among mutants (with $p_L = 0$), so they should be discarded from the total proportion of observed lethals to make any prediction. We explain below how the proportion p_L^* of such apparent lethals can be inferred from the distribution of selection coefficients among viable mutants (the ones with detectable growth), which is the one empirically available.

Consider the set of all mutations (excluding 'true lethals') occurring on an optimal parental strain ($\mathbf{g} \approx \mathbf{0}$). Mutants' selection coefficients $s \approx S_{max}(\beta_o - \beta)$ (from eq. **(1)**) follow a negative gamma distribution with shape α and mean $E(s) = -\bar{s}$ (scale $\bar{s}/\alpha$). Recalling that $S_{max}\beta_o = r_o + v$, this implies that the proportion of apparent lethals is

$$p_L^* = P(s \geq r_o) = \frac{\Gamma\left(\alpha, \alpha\frac{r_o}{\bar{s}}\right)}{\Gamma(\alpha)} \quad , \qquad \text{(A3.1)}$$

where $\Gamma(.,.)$ is the incomplete gamma function. In the virus datasets that we looked at, the distribution of selection coefficients (denoted s_v) is given after scaling by r_o. Additionally, only viable mutants are reported (i.e. those mutants that have a positive growth rate when $= S_{max}$), so that $s_v = (r_o - r)/r_o | s \geq r_o$. The mean of the these observable s_v is a conditional mean: $E(s_v) = -\bar{s}_v = E(s/r_o)/P(s \geq r_o) = -\bar{s}/(r_o(1 - p_L^*))$, so that eq. (A3.1) can be written:

$$p_L^* = \frac{\Gamma\left(\alpha, \frac{\alpha(1-p_L^*)}{\bar{s}_v}\right)}{\Gamma(\alpha)} \quad \xrightarrow[\bar{s}_v \ll \alpha]{} \quad 0 \quad , \qquad \text{(A3.2)}$$

and p_L^* can be estimated as the solution of eq. (A3.2). However, recall that α is the shape of the distribution of β among all mutants including apparent lethals with $\beta\, S_{max} \leq v$, while the observed shape α_v of s_v is that of s among only viable mutants. The distribution of s_v is a right truncated gamma distribution which parent (untruncated) distribution has shape α and scale $\beta = \bar{s}/(\alpha\, r_o) = \bar{s}_v/(\alpha(1 - p_L^*))$. The moments of such a distribution are known (e.g. Coffey and Muller, 2000) as a function of the moments of the parent untruncated gamma distribution. We can thus express the latter as a function of the former (at least

numerically): $\hat{\alpha} = f(\alpha_v, \bar{s}_v, p_L^*)$. We can then plug this inferred value $\hat{\alpha}$ (which still depends on the unknown p_L^*) into eq. (A3.2) and finally solve the equality for p_L^*.

Applying this method to the data available for several virus species in **Table 1**, we find that the overall proportion of apparent lethals inferred this way is very small. In other words, almost all lethals observed empirically are indeed true lethals, so that we can safely use the observed proportion of lethal mutations as a proxy for p_L.

Supplementary Method: Gillespie's Stochastic simulation algorithm with multiple genotypes and infection - dependent mutation

Following our basic assumptions, each individual genotype was defined by a vector **g**, a corresponding transmission rate $\beta(\mathbf{g})$ computed according to eq. **(1)**, and a density of cells infected by this genotype $I(t,\mathbf{g})$ at time t, the density of susceptible cells ($S(t)$) was also followed in parallel. The viral population was initialized with $I(0)$ cells (inoculum) all infected by the optimal genotype $\mathbf{g} = \mathbf{0}$ and with the density of susceptible cells set at its maximum (in the absence of infection: $S(0) = S_{max} = \lambda/\delta$).

We used Gillespie's (1977) stochastic simulation algorithm to generate exact simulations. It allows simulating the dynamics of the densities of susceptible cells and of all the genotype specific infected cells as an exact stochastic birth-death process. This scheme provides an exact simulation of the within-host demographic dynamics, but also entails the processes of selection and drift on the virus population, as they are directly generated by the dynamics of infected cell densities.

At any time, several possible events can take place (with corresponding rates per unit time given by the master equation (3)): birth or death of a susceptible cell (rate λ and $\delta\, S(t)$, respectively), cell – to – cell transmission by genotype **g** (i.e. release from the parent cell and infection of a new cell, rate $\beta(\mathbf{g})S(t)I(t,\mathbf{g})$) and death of a cell infected by genotype **g** (rate $v\, I(t,\mathbf{g})$). Let n_g be the number of genotypes in the population, there are $2\,(n_g+1)$ possible events. Denote a_i the rate of occurrence of event i at time t.The probability that event i occurs in the infinitesimal time interval $[\tau, \tau + d\tau]$ is (Gillespie, 1977)

$$P(i,\tau) = a_i e^{-\tau \sum_i a_i}\, d\tau \quad . \qquad \text{(M1. 1)}$$

Starting from a given state (determining the current value of the a_i's), the waiting time τ to the next event is drawn into an exponential distribution with rate $\sum_i a_i$ as

$$P(\tau) = \sum_i P(i,\tau) = \left(\sum_i a_i\right) e^{-\tau \sum_i a_i} \quad , \qquad \text{(M1. 2)}$$

and the identity of the next event ($i \in [1,2(n_g+1)]$) is drawn with point probability

$$P(i) = \int_{\tau=0}^{\infty} P(i,\tau)d\tau = \frac{a_i}{\sum_i a_i} \quad . \qquad \text{(M1. 3)}$$

Integrating **(M1. 1)** over all possible waiting times. At each new event, the number of cells is updated accordingly: $S(t) \to S(t)+1$ (resp. $S(t)-1$) for the birth (resp. death) of a susceptible cell, $I(t,\mathbf{g}) \to I(t,\mathbf{g})+1$ and $S(t) \to S(t)-1$, for a cell – to – cell transmission of genotype **g**, $I(t,\mathbf{g}) \to I(t,\mathbf{g})-1$ for the death of a cell infected by **g**. The new time $t+\tau$ is recorded, and the process is repeated, until reaching a maximal time or until extinction of the virus population (when $I_T(t) = \sum_{\mathbf{g}} I(t,\mathbf{g}) = 0$).

Mutation was simulated jointly with this demographic process, creating new genotypic classes ($\mathbf{g}'$ with initial cell density $I(t,\mathbf{g}') = 1$). In constant mutation rate model, a Poisson number of mutations was created from each individual of each genotype class,

every small time interval dt, with a rate accordingly equal to U. In second model with infection – dependent mutation, mutation was allowed on one individual of a given genotypic class **g**, every time the event drawn was infection by **g** (with rate $\beta(\mathbf{g})S(t)I(t,\mathbf{g})$). In this case, a Poisson number of mutations were generated from one individual of the genotypic class **g**, with rate μ, the rate of mutation per infectious cycle.

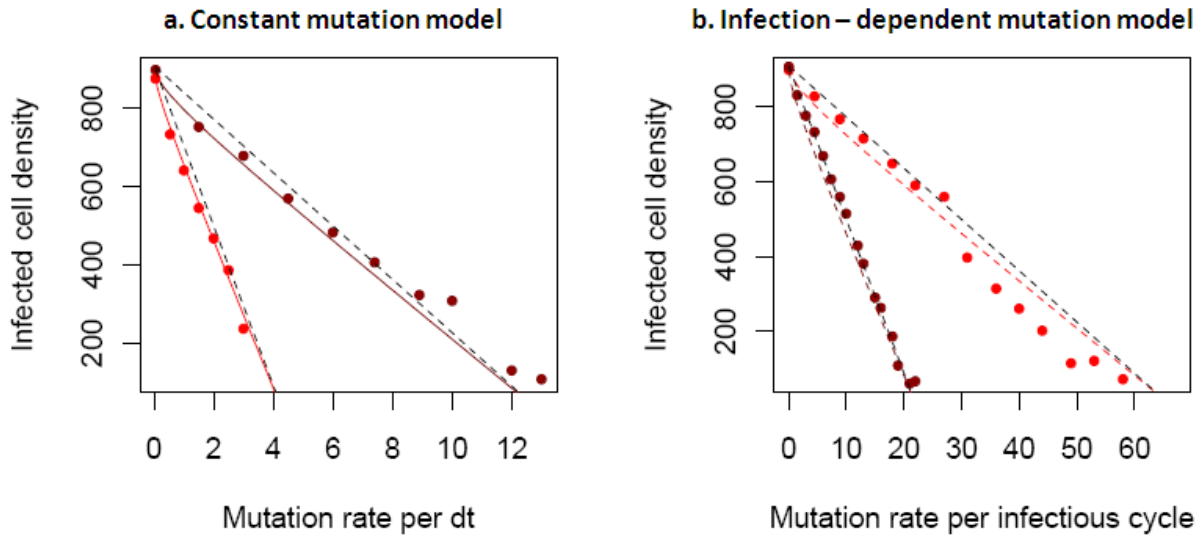
Apart from this difference, the two models are equivalent in terms of mutation effects, we drew a proportion p_L of lethal mutants from the mutant pool (according to a binomial sampling) and set the transmission rate to $\beta \to 0$ for the corresponding cells. We updated the genotypes of the remaining $(1-p_L)$ viable mutants by drawing multivariate Gaussian deviates with covariance **M** and setting $\mathbf{g}' = \mathbf{g} + \mathbf{dg}$ for each mutant, and we updated the transmission rate $\beta(\mathbf{g}')$ of each corresponding cell according to eq. **(1)**. Selection between genotypes is then naturally generated by the demographic algorithm which corresponds to the master eq.**(3)**.

Following Martin & Lenormand (2006), the covariance matrices **M** (mutation effects) and $\mathbf{\Sigma}_\beta$ (selection on transmission rates, eq. **(1)**) were drawn as Wishart deviates, a null model of phenotypic covariance matrix. Finally $\mathbf{\Sigma}_\beta$, and **M** were then scaled so as to have a given value of the parameters $\bar{s}$ and α of the distribution of mutation fitness effects in exponential phase (i.e. for S_{max}, see eq. **(5)**). For a given value of epidemiological parameters (λ, δ, v), S_{max} was set to λ/δ. Then, for a given value of the wild – type r_o (growth rate in exponential phase). The transmission rate of the optimal genotype was then computed as $\beta_o = (r_o + v)/S_{max}$, by definition. The values of both genetic and epidemiological parameters were set according to typical orders of magnitude from the literature (Nowak and May, 2000, Sanjuàn et al., 2004).

Supplementary Table: model notations

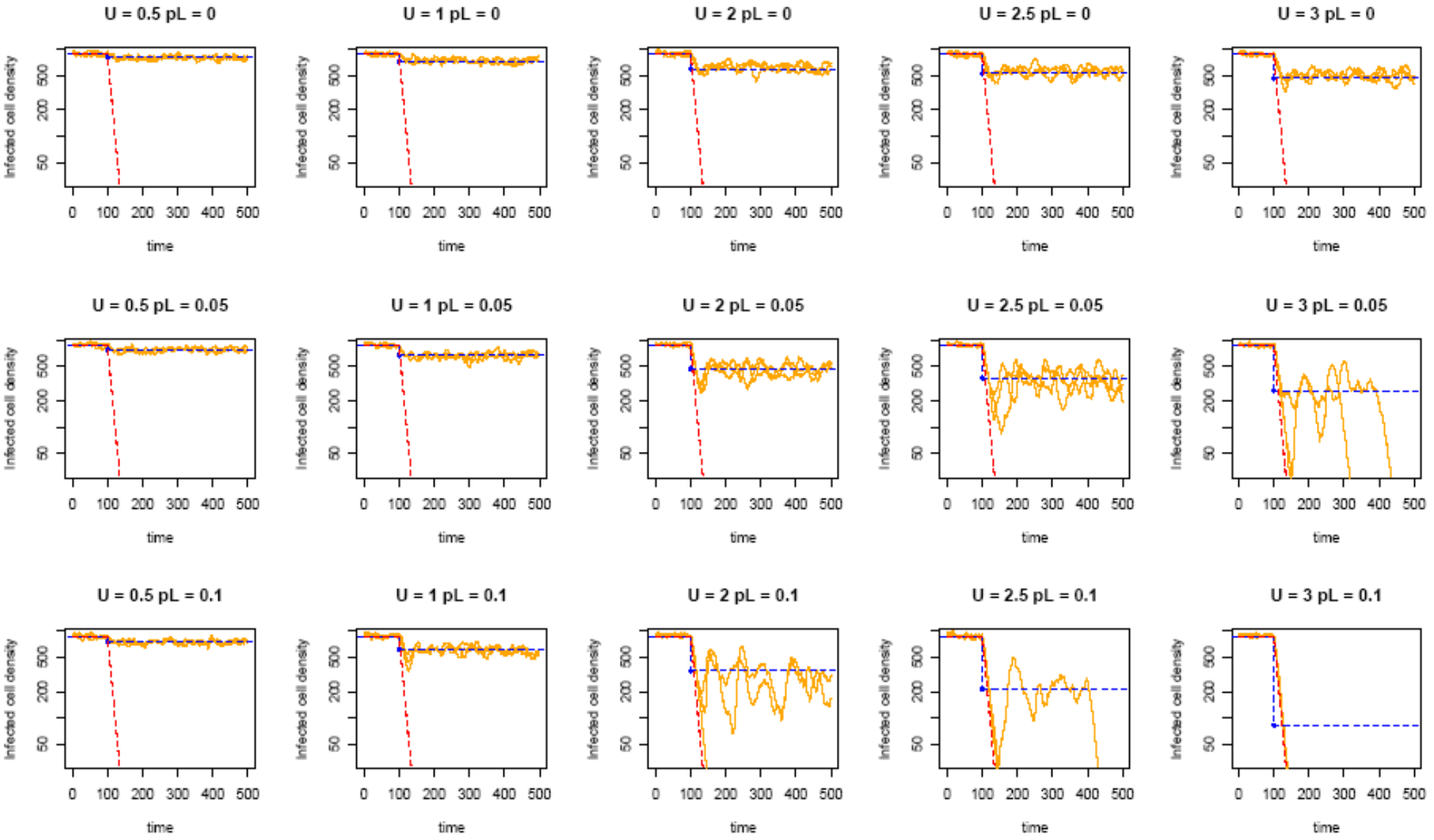
Epidemiological parameters and link to the model of virus dynamics	
λ	Mean birth rate of susceptible cells
δ	Mean per-cell death rate of susceptible cells
β_v	Transmission rate of the viral particles from the free stage into cells
k_v	Burst size of the virus
μ_v	Death rate of viral particles in free stage (blood, etc.)
v	Mean per-cell death rate of infected cells (virulence)
$\beta(\mathbf{g}), \beta_o$	Cell – to – cell transmission rate of viral genotype **g**, or of optimal genotype (resp.): $\beta = \beta_v k_v/\mu_v$
$r(t, \mathbf{g})$	Malthusian fitness of viral genotype **g** at time t
$r_o(t), r_o$	Exponential growth rate of the optimal genotype at time t, or at maximal host cell density (resp.)
$I(t, \mathbf{g})$	Density of cells infected by genotype **g** at time t
$S(t), S_*$	Density of susceptible cells at time t or at equilibrium (resp.)
$I_T(t), I_{T*}$	Total density of infected cells at time t or at equilibrium (resp.)
$\bar{\beta}(t), \bar{\beta}_*$	Mean transmission rate at time t or at equilibrium (resp.)
Mutational and genetic parameters	
U (U_e)	genomic mutation rate per unit time (or effective one, infection-dependent mutation model)
μ	Genomic mutation rate per infectious cycle (infection – dependent mutation model)
U_c , μ_c	Critical mutation rate for viral extinction (per unit time or per infection, resp.)
$\bar{s}$	Average effect (absolute value) of mutations on Malthusian fitness
α	shape of the distribution of mutation fitness effects ($= 1/CV(s)^2$)
p_L, p_L^*	Proportion of 'true' and 'apparent' lethals (resp.) among all random mutations
$\bar{r}_*$	mean Malthusian fitness at mutation – selection balance
$\mathbf{g}$	Multivariate phenotype of a given virus (alternative bases)
$\boldsymbol{G}(t), \mathbf{G}_*$	Genetic variance – covariance matrix of virus phenotypes, at time t or at equilibrium (resp.)

Supplementary Figure 1: Effect of the mean fitness effect of mutations on the equilibrium virus titer. Same figure as **Figure 2** but with $p_L = 0$ and with two possible values of the mean fitness effect of random mutations $\bar{s}$ measured at maximal susceptible host cell density ($\bar{s} = 0.15$ or 0.05, $p_L = 0$). We see that the model accurately captures the effect of mutational parameters $(U, \bar{s})$ on the efficacy of mutagenesis (NB: here the x-axis directly gives the mutation rate per unit time or per infection, not the relative rate as in **Figure 2**).

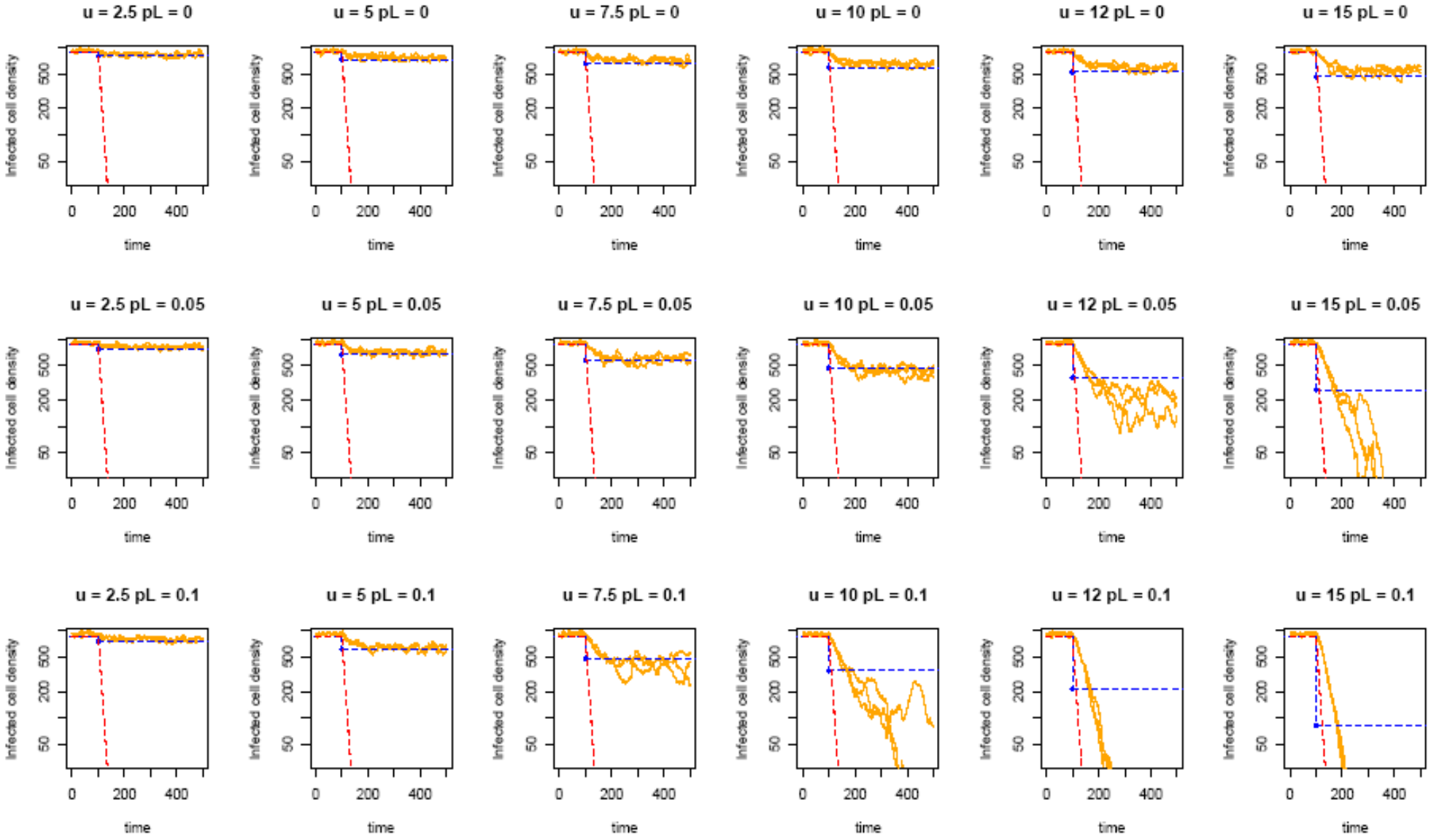


Supplementary figure 2: Examples of stochastic dynamics of infected cell density. Here we give the temporal dynamics of the infected cell density (y-axis, log-scale) for different values of the mutational parameters (U_e and p_L given on the graphs), from the same simulations as used in **Figure 2**. The dashed blue lines give the expected equilibrium host cell density before and after the onset of treatment by the mutagen, and the red dashed line gives the exponential dynamics to extinction $I(t) = I_o e^{-v\,t}$ expected when there is a lag between the onsets of the demographic and evolutionary equilibria (see text). We see that a population may go to extinction even at $U \leq U_c$ while other replicates stay close to the predicted equilibrium density. We see also that when a population goes to extinction, it does so roughly according to an exponential decrease (parallel to the red dashed line), at least in the constant mutation model (**a.**). The stochastic decrease to extinction seems to be exponential at higher rate in the infection-dependent mutation model (**b.**).

a. Constant mutation model (U = mutation rate per unit time)



b. Infection-dependent mutation model (u = mutation rate per infection)



Supplementary Figure 3: Effect of the mutational parameters on the extinction threshold. This contour plot shows the value of the critical mutation rate for extinction (U_c, eq. **Error! Reference source not found.**) as a function of the fitness effect of non-lethal mutations (product $\alpha\ \bar{s}$) and of the proportion of 'true' lethal utations (p_L). The delimiting values of U_c are indicated on the graph, and the dashed line gives the corresponding exponential approximation (right-hand side of eq. **Error! Reference source not found.**). In this example, the viral growth rate of the optimal genotype was $r_o = 2$ per unit time.