

Research



Cite this article: Parsons TL, Lambert A, Day T, Gandon S. 2018 Pathogen evolution in finite populations: slow and steady spreads the best. *J. R. Soc. Interface* **15**: 20180135. <http://dx.doi.org/10.1098/rsif.2018.0135>

Received: 23 February 2018
Accepted: 11 September 2018

Subject Category:
Life Sciences – Mathematics interface

Subject Areas:
biomathematics, evolution

Keywords:
epidemiology, virulence, life-history evolution, genetic drift, demographic stochasticity, Adaptive Dynamics

Author for correspondence:
Sylvain Gandon
e-mail: sylvain.gandon@cefe.cnrs.fr

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.4238543>.

Pathogen evolution in finite populations: slow and steady spreads the best

Todd L. Parsons¹, Amaury Lambert^{1,2}, Troy Day^{3,4} and Sylvain Gandon⁵

¹Laboratoire de Probabilités, Statistique et Modélisation (LPSM), Sorbonne Université, CNRS UMR 8001, Paris, France

²Center for Interdisciplinary Research in Biology (CIRB), Collège de France, PSL Research University, CNRS UMR 7241, INSERM U1050, Paris, France

³Department of Mathematics and Statistics, and ⁴Department of Biology, Queen's University, Kingston, Canada

⁵Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), Université de Montpellier–Université Paul-Valéry Montpellier–EPHE, CNRS UMR 5175, Montpellier, France

TLP, 0000-0002-2599-8415; SG, 0000-0003-2624-7856

The theory of life-history evolution provides a powerful framework to understand the evolutionary dynamics of pathogens. It assumes, however, that host populations are large and that one can neglect the effects of demographic stochasticity. Here, we expand the theory to account for the effects of finite population size on the evolution of pathogen virulence. We show that demographic stochasticity introduces additional evolutionary forces that can qualitatively affect the dynamics and the evolutionary outcome. We discuss the importance of the shape of the pathogen fitness landscape on the balance between mutation, selection and genetic drift. This analysis reconciles Adaptive Dynamics with population genetics in finite populations and provides a new theoretical toolbox to study life-history evolution in realistic ecological scenarios.

1. Introduction

Why are some pathogens virulent and harm their hosts while others have minimal effect on host fitness? Our ability to understand and predict the evolutionary dynamics of pathogen virulence has considerable implications for public-health management [1–3]. A classical explanation for pathogen virulence involves trade-offs with other pathogen life-history traits. If certain components of pathogen fitness, such as a high transmission rate or a low clearance rate, necessarily require that the pathogen incidentally increase host mortality, then virulence is expected to evolve [4]. A now classical way to develop specific predictions from this hypothesis is invasion analysis and evolutionary game theory, under assumptions that have been since formalized as Adaptive Dynamics [1,4–6]. This approach relies on the assumption that the mutation rate is small so that the epidemiological dynamics occur on a faster timescale than the evolutionary dynamics [4,7–9]. Under simple epidemiological assumptions (e.g. well-mixed population, no co-infection or superinfection with different genotypes, a single infection pathway, etc.) the evolutionarily stable level of virulence maximizes the basic reproduction ratio R_0 of the pathogen [7,10,11], but see e.g. [12–17] for more complex epidemiological scenarios.

The above-mentioned theory allows one to determine the level of virulence expected to evolve under a broad range of epidemiological scenarios but it still suffers from the fundamental shortcoming of being a deterministic theory. The number of infected individuals, however, can be very small (e.g. at the onset of an epidemic or after a vaccination campaign) and demographic stochasticity—i.e. randomness in individual mortality and reproduction [18]—is likely to affect both the epidemiological and evolutionary dynamics of the disease. If all that such stochasticity did was to introduce random noise, then the predictions of deterministic theory would likely suffice. However, several recent

studies have demonstrated that this is not the case. For example, [19,20] each used different theoretical approaches to demonstrate that finite population size tends to select for lower virulence and transmission, using perturbation series, and assuming fixed numbers of infecteds, respectively, to estimate fixation probabilities. Likewise, [21] analysed the effect of finite population size in a complex epidemiological model with unstable epidemiological dynamics and showed that finite population size could induce an evolutionary instability that may either lead to selection for very high or very low transmission.

Taken together, the existing literature presents a complex picture of the factors that drive virulence evolution and it remains unclear how all of these factors are related to one another and how they might interact. In this paper, we develop a very general theory of pathogen evolution that can be used to examine virulence evolution when the above-mentioned factors are at play. First, we use an individual-based description of the epidemiological process to derive a stochastic characterization of the evolutionary epidemiology dynamics of the pathogen. This theoretical framework is used to pinpoint the effect of finite population size on the interplay between epidemiology and evolution. Second, we analyse this model under the realistic assumption that the rate of mutation is small, so that pathogen evolution can be approximated by a sequence of fixations. We derive the probability of fixation of a mutant pathogen under both weak and strong selection regimes, and for different epidemiological scenarios. Third, we use this theoretical framework to derive the stationary distribution of pathogen virulence resulting from the balance between mutation, selection and genetic drift. This yields new predictions regarding the effect of the shape of pathogen fitness landscape and the size of the population on the long-term evolution of the pathogen. As the question of virulence evolution can be viewed as a specific example of the more general notion of life-history evolution [22,23], our results should be directly applicable to other life-history traits and other organisms as well, providing a new theoretical approach for studying life-history evolution in realistic ecological scenarios based on the principles advocated in [24]: Stochastic Adaptive Dynamics (SAD).

2. Model

We use a classical SIR epidemiological model with demography, where hosts can either be susceptible, infected or recovered. The number of each of these types of hosts is denoted by N_S , N_I and N_R , respectively. Because we are interested in the effect of demographic stochasticity the model is derived from a microscopic description of all the events that may occur in a finite—but not fixed—host population of total size $N_T = N_S + N_I + N_R$ living in an area of size n (the details of the model are given in the electronic supplementary material).

We use λ to denote the rate at which new susceptible hosts enter the population *per unit area*, and therefore the total rate is given by λn . We focus on the case of frequency-dependent transmission; i.e. new infections occur at rate $(\beta/N_T)N_S N_I$, where β is a constant quantifying the combined effects of contact rate among individuals and the probability of pathogen transmission, given an appropriate contact occurs. Note, however, that other forms of transmission

(e.g. density-dependent transmission [25]) yield qualitatively similar results [26]. We also assume that already infected hosts cannot be reinfected by another pathogen strain (i.e. no co-infections). All hosts are assumed to suffer a constant *per capita* death rate of δ , whereas infected hosts die *due to disease at per capita* rate α and they recover at *per capita* rate γ . Finally, to study pathogen evolution, we need to introduce genetic variation in the pathogen population. Therefore, we consider d pathogen strains which differ in their transmission rate β_i , and virulence α_i , with $i \in \{1, \dots, d\}$. Likewise, we use the subscripted variable N_i to denote the number of hosts infected with strain i .

The above assumptions give a continuous-time Markov process tracking the number of individuals of each type of host. To progress in the analysis, we use a diffusion approximation and work with host densities defined as $S = N_S/n$, $I_i = N_i/n$ and $N = N_T/n$ and we define the total density of infected hosts as $I = \sum_{i=1}^d I_i$. When n is sufficiently large (but finite) these variables can be approximated using a continuous state space and so this model can be described by a system of stochastic differential equations (see electronic supplementary material, §3).

2.1. Deterministic evolution

In the limit where the habitat size (and thus the host population size) becomes infinite, demographic stochasticity becomes unimportant and the epidemiological dynamics are given by the following system of ordinary differential equations:

$$\left. \begin{aligned} \dot{S} &= \lambda - \frac{\beta}{N} SI - \delta S, \\ \dot{I} &= \frac{\beta}{N} SI - (\delta + \bar{\alpha} + \gamma)I \\ \text{and} \quad \dot{N} &= \lambda - \delta N - \bar{\alpha}I. \end{aligned} \right\} \quad (2.1)$$

The bars above α and β refer to the mean of the transmission rate and the virulence distributions of the infected host population (i.e. $\bar{\alpha} = \sum_{i=1}^d \alpha_i I_i / I$). In the absence of the pathogen, the density of hosts equilibrates at $S_0 = \lambda / \delta$. A monomorphic pathogen ($d = 1$, $\beta = \beta$ and $\bar{\alpha} = \alpha$) is able to invade this equilibrium if its basic reproduction ratio, $R_0 = \beta / (\delta + \alpha + \gamma)$ is greater than one. If this condition is fulfilled, then the system reaches an endemic equilibrium, where $S_{\text{eq}} / N_{\text{eq}} = 1 / R_0$, $I_{\text{eq}} / N_{\text{eq}} = (\delta / (\delta + \gamma))(1 - 1 / R_0)$ and $N_{\text{eq}} = (\lambda(\delta + \gamma) / \delta(\beta - \alpha))R_0$.

When several strains are present in the population, the evolutionary dynamics of the pathogen can be tracked with [27,28]:

$$\dot{p}_i = p_i(r_i - \bar{r}), \quad (2.2)$$

where $p_i = I_i / I$ is the frequency of hosts infected with strain i . The quantity $r_i = \beta_i(S/N) - (\delta + \alpha_i + \gamma)$ is the instantaneous *per capita* growth rate of strain i and $\bar{r} = \sum_{i=1}^d p_i r_i$ is the average *per capita* growth rate of the infected host population. When $d = 2$ only two strains are competing (a wild-type, strain 1, and a mutant, strain 2) and the change in the frequency p_2 of the mutant strain is given by:

$$\dot{p}_2 = p_1 p_2 \left(\frac{S}{N} \Delta\beta - \Delta\alpha \right), \quad (2.3)$$

where $\Delta\beta = \beta_2 - \beta_1$ and $\Delta\alpha = \alpha_2 - \alpha_1$ are the effects of the mutation on transmission and virulence, respectively.

The above formalization can be used to understand the evolution of pathogen life history under different scenarios. First, under the classical Adaptive Dynamics assumption that the mutation rate is very small, one may use a separation of timescales where the epidemiological dynamics reach an endemic equilibrium (set by the resident pathogen, strain 1) before the introduction of a new variant (strain 2) by mutation. In this case, evolution favours the strain with the highest basic reproduction ratio, $R_{0,i} = \beta_i / (\delta + \alpha_i + \gamma)$. In other words, evolution favours strains with higher transmission rates and lower virulence. According to the trade-off hypothesis, however, transmission and virulence cannot evolve independently. For example, the within-host growth rate of pathogens is likely to affect both traits and result in a functional trade-off between transmission and virulence [4,7–9]. Under this assumption, equation (2.3) can be used to predict the evolutionary stable virulence strategy (figure 1).¹ The above model can also be used to predict virulence evolution when the evolutionary and epidemiological dynamics occur on a similar timescale [27–29]. For instance, these models can be used to understand virulence evolution during an epidemic [4,30–32]. In this case, a pathogen strain i with a lower R_0 may outcompete other strains if its instantaneous growth rate, r_i , is higher.

2.2. Stochastic evolution

Finite population size introduces demographic stochasticity and the epidemiological dynamics can be described by the following system of (Itô) stochastic differential equations:

$$\left. \begin{aligned} dS &= \left(\lambda - \frac{\bar{\beta}}{N}SI - \delta S \right) dt + \sqrt{\frac{\lambda}{n}} dB_1 - \sqrt{\frac{\delta S}{n}} dB_2 - \sqrt{\frac{\bar{\beta}SI}{nN}} dB_3, \\ dI &= \left(\frac{\bar{\beta}}{N}SI - (\delta + \bar{\alpha} + \gamma)I \right) dt + \sqrt{\frac{\bar{\beta}SI}{nN}} dB_3 - \sqrt{\frac{(\delta + \bar{\alpha})I}{n}} dB_4 - \sqrt{\frac{\gamma I}{n}} dB_5 \\ \text{and } dN &= (\lambda - \delta N - \bar{\alpha}I) dt + \sqrt{\frac{\lambda}{n}} dB_1 - \sqrt{\frac{\delta S}{n}} dB_2 - \sqrt{\frac{(\delta + \bar{\alpha})I}{n}} dB_4, \end{aligned} \right\} \quad (2.4)$$

where B_1, \dots, B_5 are independent Brownian motions. As expected, when $n \rightarrow \infty$ this set of stochastic differential equations reduces to the deterministic equations in (2.1) (n.b., both (2.4) and (2.1) require that one knows the strain frequencies, as given by (2.2) or (2.5) below, respectively, for a complete description of the dynamics).

In finite populations, the pathogen, and indeed the host population itself, are destined to extinction with probability 1. The time it takes for this to occur, however, depends critically on the parameter values. For example, in a host population infected with a monomorphic pathogen (i.e. $d = 1$), if R_0 is larger than 1 the size of the infected host population reaches a quasi-stationary distribution which is approximately normal. The mean of this distribution is of order n and its standard deviation is of order \sqrt{n} [33,34]. The extinction time from the quasi-stationary distribution increases exponentially with n [33,34], and so, in the remainder of the paper, we will assume that n is large enough so that we can focus on the dynamics conditional on non-extinction.

As in the deterministic case, one can study evolutionary dynamics by focusing on the change in strain frequencies.

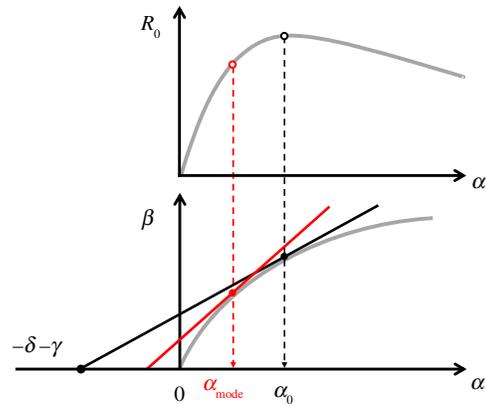


Figure 1. Schematic of the effect of finite population size on the evolution of pathogen virulence. The grey line in the top figure represents the effect of pathogen virulence, α , on R_0 (for an asymmetric fitness function). The grey line in the bottom figure represents the effect of pathogen virulence, α , on pathogen transmission, β . In the deterministic version of our model, the marginal value theorem can be used to find the optimal pathogen virulence, α_0 (dashed black arrow). In this model, optimal virulence maximizes R_0 in the absence of demographic stochasticity. Finite population size modifies selection and favours pathogen strategies with lower virulence (see equation (3.4)). The mode of the stationary distribution of pathogen virulence is indicated by a dashed red arrow, α_{mode} (see equation (3.7)). This geometrical construction indicates that finite population size is expected to favour slower strains even if they have a lower R_0 .

We obtain a stochastic differential equation analogous to (2.2) (see electronic supplementary material, S4):

$$dp_i = \left(p_i(r_i - \bar{r}) - \frac{1}{nI} p_i(v_i - \bar{v}) \right) dt + \frac{1}{\sqrt{nI}} \sum_{j=1}^d (\delta_{ij} - p_i) \sqrt{v_j p_j} dB_j, \quad (2.5)$$

where $v_i = \beta_i(S/N) + (\delta + \alpha_i + \gamma)$ is the variance in the growth rate of strain i (while $r_i = \beta_i(S/N) - (\delta + \alpha_i + \gamma)$ is the mean) and $\bar{v} = \sum_{i=1}^d p_i v_i$ is the average variance in growth rate of the infected host population. The first (advective, dt) component in equation (2.5) is analogous to (2.2). The second (diffusive, dB) component shows that finite population size (i.e. when the pathogen-infected population size, as measured by the total number of infected hosts, nI is not too large) can affect the direction of evolution. In contrast with the deterministic model, the evolutionary dynamics are not driven exclusively by the expected growth rate r_i , but also by a minimization of the variance. This effect is akin to bet-hedging theory stating that a mutant strategy with lower variance in reproduction may outcompete a

resident strategy with a higher average instantaneous growth rate [35,36]. To better understand this effect, it is particularly insightful to examine the case $d = 2$ when only two strains are competing and the change in frequency p_2 of the mutant strain is given by:

$$dp_2 = p_1 p_2 \left(\frac{S}{N} \Delta\beta \left(1 - \frac{1}{nI} \right) - \Delta\alpha \left(1 + \frac{1}{nI} \right) \right) dt + \sqrt{\frac{p_1 p_2}{nI} (p_1 v_2 + p_2 v_1)} dB. \quad (2.6)$$

The first component (the advective component) in equation (2.6) is similar to (2.3) except for the $1/nI$ terms. Those terms are due to the fact that a transmission (or a death) event of the mutant is associated with a change in the number of mutants as well as an increase (decrease) of the total infected host population size by one individual. This concomitant variation of infected host population size affects the effective change of the mutant frequency (relative to the change expected under the deterministic model where the population size is assumed to be infinite). This effect decreases the benefit associated with higher transmission and increases the cost of virulence. In the long-term, this effect (the first term in (2.5)) is thus expected to select for lower virulence. But this long-term evolutionary outcome cannot be described by an evolutionary stable state because demographic stochasticity is also expected to generate noise (the diffusion term in (2.5)). Indeed, this stochasticity (i.e. genetic drift) may lead to the invasion and fixation of strains with lower *per capita* growth rates. In the following, we fully characterize this complex evolutionary outcome with the stationary distribution of pathogen virulence under different epidemiological scenarios.

3. Results

The above theoretical framework embodied by the stochastic differential equations (2.4) and (2.5) subsume the deterministic model and can be used to study the interplay of all the relevant factors affecting virulence evolution. In the following, we will assume that pathogen mutation is rare, so that evolution can be described, as in classical Adaptive Dynamics, as a chain of fixations of new pathogen mutations. In contrast with Adaptive Dynamics, however, demographic stochasticity in the resident population may allow neutral, or even mildly deleterious, mutations to go to fixation. The analysis of the effect of finite population size requires specific ways to quantify the stochastic fate of a genotype [37]. To determine the fate of a new mutation we need to compute the probability of fixation of a mutant pathogen in a resident population. In the absence of selection, the fixation probability of a mutant allele depends only on the demography of the population. When the size of the population is fixed and equal to N the fixation probability of a neutral allele is $1/N$. When the fixation probability of a mutant is higher than neutral it indicates that the mutant is selectively favoured. This is particularly useful in many complex situations where the interplay between selection and genetic drift are difficult to disentangle like time-varying demography [38,39] or spatial structure [40]. In our model, the difficulty arises from (i) the stochastic demography of the infected host population and (ii) the fact that pathogen life-history traits feedback on the epidemiological dynamics and thus on the intensity of genetic drift.

3.1. Stationary distribution of pathogen virulence at equilibrium

Here we assume, as in the Adaptive Dynamics framework, that the pathogen mutation rate μ is so low that the mutant pathogen (strain 2) arises when the resident population (strain 1) has reached a quasi-stationary distribution tightly peaked about nI_{eq} (i.e. close to the endemic equilibrium derived in the deterministic model). The R_0 of the two strains may be written in the following way: $R_{0,2} = R_{0,1}(1 + s)$ where s measures the magnitude of selection.

When selection is strong (i.e. $s \gg 1/n$) the probability of fixation of the mutant when $N_{I_2}(0)$ mutants with $R_{0,2} > 1$ are introduced into a resident population at equilibrium is (see electronic supplementary material, §5.2):

$$U_{\text{strong}} \approx 1 - \left(\frac{R_{0,1}}{R_{0,2}} \right)^{N_{I_2}(0)} \approx N_{I_2}(0)s, \quad (3.1)$$

which may be obtained by approximating the invading strain by a branching process (see electronic supplementary material, §8.2 for a rigorous justification).

When the mutant and the resident have similar values of $R_0 > 1$ (i.e. s is of order $1/n$) selection is weak, and the derivation of the probability of fixation is a much more difficult problem. The classical population genetics approach under the assumption that population size is fixed (or is characterized by a deterministic trajectory independent of mutant frequency) is to use the diffusion equation of mutant frequency to derive the probability of fixation [38,39]. But in our model, equation (2.3) is not autonomous and is coupled with the epidemiological dynamics. To derive the probability of fixation we use a separation of timescale argument to reduce the dimension of the system (see [41] for a discussion of the approach). Indeed, if selection is weak, as $n \rightarrow \infty$, the deterministic component of the model sends the system rapidly to the endemic equilibrium, which is now a manifold of fixed points, on which coexistence is possible at all mutant frequencies. After this, it is possible to approximate the change in frequency of the mutant by tracking the dynamics of the projection of the mutant frequency on this manifold (see electronic supplementary material, §5.3). This one-dimensional system can then be used to derive the probability of fixation under weak selection. A first-order approximation in s and σ is:

$$U_{\text{weak}} \approx p + \frac{p(1-p)}{2} (nI_{\text{eq}}s + \sigma) \approx nI_2(0) \left(\frac{1}{nI_{\text{eq}}} + \frac{1}{2} \left(s + \frac{\sigma}{nI_{\text{eq}}} \right) \right), \quad (3.2)$$

where $p = I_2(0)/I_{\text{eq}}$ and $\sigma = (\beta_1 - \beta_2)/\beta_2$. The first term in (3.2) is the probability of fixation of a single neutral mutation introduced in an infected host population at the endemic equilibrium, nI_{eq} . The second term takes into account the effect due to selection. First, selection may be driven by differences in R_0 . Second, even if strains have identical R_0 (i.e. $s = 0$) selection may be driven by σ which measures the difference in transmission rate; this effect selects for lower transmission rates, and, since under weak selection the R_0 values are approximately equal, for lower virulence. Note, however, that the effect of s rapidly overwhelms the effect of σ as the infected host population size nI_{eq} becomes large (unless s is of order $1/n$). The probability of fixation

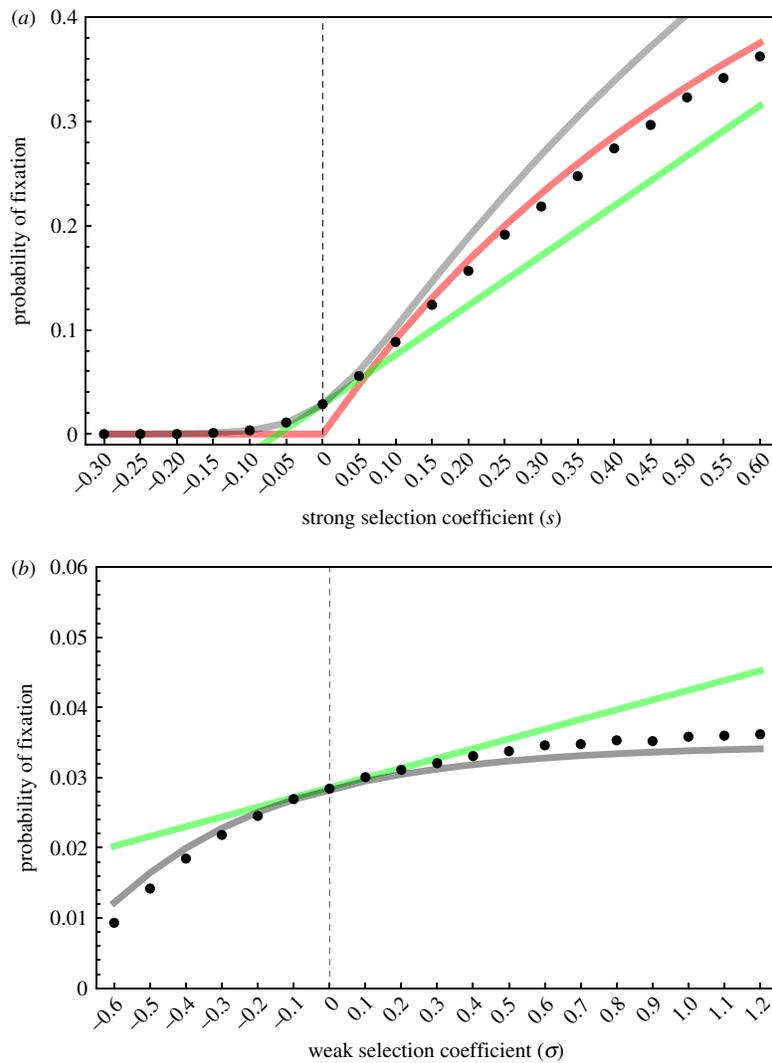


Figure 2. Probability of fixation for (a) different values of s (strong selection effect) and (b) different values of σ for fixed R_0 (weak selection effect). Simulation results for the model described in §2 are indicated with a dot, weak selection approximation is indicated with a grey line and its linear approximation (equation (3.2)) is indicated with a green line, the strong selection approximation is indicated with a red line (equation (3.1)). Parameter values of the resident population: $n = 100$, $R_0 = 4$, $\delta = 1$, $\alpha = 3$, $\gamma = 1$, $\lambda = 2$, $\beta_1 = 20$. For the simulation, a single mutant (an individual host infected with a mutant pathogen) is introduced at the endemic equilibrium set by the resident pathogen: $S_{\text{eq}} = 24$ and $I_{\text{eq}} = 35$. 10^6 simulations are realized for each parameter values and we plot the proportion of the simulations where the mutant goes to fixation. We implement strong selection by setting $\beta_1 = \beta_2(1 + s)$ so that $R_{0,1} = R_{0,2}(1 + s)$, and weak selection by setting $\beta_1 = \beta_2(1 + \sigma)$, while holding $R_{0,1} = R_{0,2}$ by setting $\alpha_2 = (\delta + \alpha_1 + \gamma)/(1 + \sigma) - \delta - \gamma$.

given in (3.2) confirms that evolution tends to push towards higher basic reproductive ratio but when the population size is small other forces may affect the evolutionary outcome. In particular, when nI_{eq} is small, strains with lower R_0 can reach fixation. Figure 2 shows the result of stochastic simulations that confirm the approximations (3.1) and (3.2) under different epidemiological scenarios, and show that our approximations already perform extremely well for populations as small as the order of 100 hosts (see electronic supplementary material, S7 for details of the simulations).

Even though the probability of fixation helps understand the interplay between selection and genetic drift it does not account for any differences in the time to fixation and it is often difficult to measure this probability experimentally as well (but see [42]). What may be more accessible is a characterization of the phenotypic state of the population across

different points in time (or in space among replicate populations)—that is, the stationary distribution of the virulence of the pathogen under the action of mutation, selection and genetic drift [43–45] (figure 3).

To derive the stationary distribution of pathogen virulence, we first need to impose a trade-off between virulence and transmission rate, setting $\beta = \beta(\alpha)$, and introduce the mutation kernel $K(\alpha_m, \alpha)$, the probability that a mutant with strategy α_m appears in a monomorphic population with strategy α . Here, we assume that this distribution has mean equal to the current resident trait value and variance v . Under the assumption that the mutation rate μ remains small, pathogen polymorphism is limited to the transient period between the introduction of a mutant and a fixation, and we may consider the (monomorphic) resident virulence as a random process evolving in time.

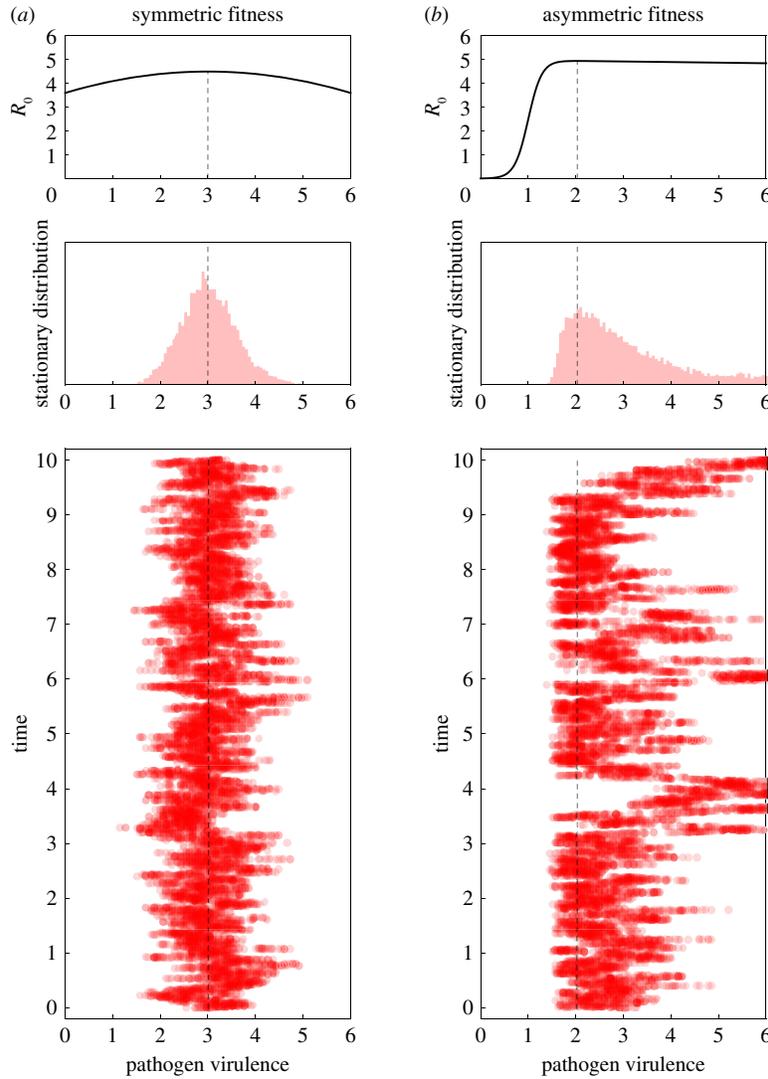


Figure 3. Dynamics of pathogen virulence across time (one time unit on the graph is 10^7 time steps in the simulation) and stationary distribution of pathogen virulence for two different fitness landscapes: (a) symmetric fitness landscape with $\beta(\alpha) = (\delta + \gamma + \alpha)R_{0,\max}(1 - w(\alpha_0 - \alpha)^2)$, $R_{0,\max} = 4.5$ and $\alpha_0 = 3$, (b) asymmetric fitness landscape with $\beta(\alpha) = 5(1 - 0.005\alpha)/(1 + \exp(7(1 - \alpha)))$. We take $K(\alpha_m, \alpha) = \begin{cases} 1 & \text{if } \alpha_m - \alpha = \pm \Delta\alpha = \frac{6}{100}, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$ The dashed vertical line indicates the position of α_0 . Other parameter values: $n = 200$, $\delta = 1$, $\alpha = 3$, $\gamma = 1$, $\lambda = 2$, $\mu = 0.001$.

The probability of fixation (3.2) accurately describes the direction of evolution, and the evolution of pathogen virulence can then be described by the following Fokker-Planck diffusion equation (see electronic supplementary material, §6):

$$\frac{\partial \psi(\alpha, t)}{\partial t} = -\frac{\mu\nu}{2} \frac{\partial}{\partial \alpha} \left[nI_{\text{eq}}(\alpha) \left(\frac{R'_0(\alpha)}{R_0(\alpha)} - \frac{1}{nI_{\text{eq}}(\alpha)} \frac{\beta'(\alpha)}{\beta(\alpha)} \right) \psi(\alpha, t) \right] + \frac{\mu\nu}{2} \frac{\partial^2 \psi(\alpha, t)}{\partial \alpha^2}, \quad (3.3)$$

where $\psi(\alpha, t)$ is the probability of observing pathogen virulence α at time t and $'$ indicates the derivative with respect to α , and we write $R_0(\alpha)$ and $I_{\text{eq}}(\alpha)$ to emphasize that these quantities depend on the resident virulence. The first term of the above equation indicates a strong deterministic trend, with $R'_0(\alpha)$ indicating a trend towards a higher basic reproduction ratio, offset by a finite population effect

proportional to $-\beta'(\alpha)$ that tends to select for lower transmission. Under the classical assumption that pathogen transmission and pathogen virulence are linked by a genetic trade-off one can ask what the level of pathogen virulence is where the advective term is zero. This trait value corresponds to the mode of the stationary distribution of pathogen virulence and is given by the following condition (see electronic supplementary material, equation S.39):

$$\beta'(\alpha) = R_0(\alpha) \left(1 + \frac{1}{nI_{\text{eq}}(\alpha) - 1} \right). \quad (3.4)$$

When the infected host population is very large (i.e. $n \rightarrow \infty$) we recover the marginal value theorem, while finite population size increases the slope $\beta'(\alpha)$ and reduces the mode of the stationary distribution (figure 1). Thus, provided the transmission–virulence trade-off function is concave, finite population size is expected to decrease virulence and

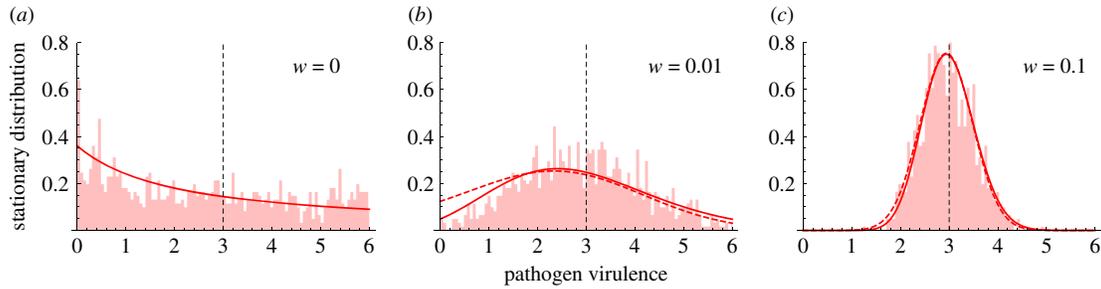


Figure 4. Stationary distribution for symmetric fitness landscapes with increasing strength of selection around the optimum with $\beta(\alpha) = (\delta + \gamma + \alpha)R_{0,\max}(1 - w(\alpha_0 - \alpha)^2)$, $K(\alpha_m, \alpha)$ as in figure 3, and $\alpha_0 = 3$ for three different values of w : (a) $w = 0$, (b) 0.01 and (c) 0.1. Note that when $w = 0$, the fitness landscape is flat, whereas with increasing w , the landscape becomes more tightly peaked at the optimum. The light red histogram indicates results of a stochastic simulation. The red line indicates the stationary distribution of the diffusion approximation (the dashed line indicates the approximation of this distribution, see (3.6)). The dashed vertical line indicates the position of α_0 . Parameter values: $n = 200$, $R_{0,\max} = 4$, $d = 1$, $\alpha = 3$, $\gamma = 1$, $\lambda = 2$, $\mu = 0.001$.

transmission rates. In other words, pathogen avirulence may be viewed as a bet-hedging strategy because even if it reduces the instantaneous growth rate r_i , the reduced variance in growth rate v_i is adaptive in finite populations.

Let us now consider the limiting case when all the pathogen strains have the same R_0 . This corresponds to a very special case where the fitness landscape is flat. The deterministic model predicts that pathogen life-history variation is neutral near the endemic equilibrium (see (2.2)). The probability of fixation (3.2) shows, however, that selection is *quasi-neutral* and favours pathogens with lower transmission and virulence rates [19,20,26,46]. The stationary distribution results from the balance between selection (pushing towards lower values of pathogen traits) and mutation (reintroducing variation). If we focus on virulence and allow variation between the minimal viable value α_{\min} and the maximal viable value α_{\max} , the stationary distribution is (see electronic supplementary material, equation S.33):

$$\psi_{\text{flat}}(\alpha) = \frac{1}{\ln((\delta + \alpha_{\max} + \gamma)/(\delta + \alpha_{\min} + \gamma))} \frac{1}{(\delta + \alpha + \gamma)}. \quad (3.5)$$

It is worth noting that this distribution is independent of the pathogen-infected population size. Indeed, near the endemic equilibrium, when pathogens have the same R_0 , the probability of fixation (3.2) is independent of infected population size. So this prediction holds even in very large populations. The time to fixation may, however, be considerably longer in large populations and the assumption that polymorphism is always reduced to the resident and a single mutant may not always hold as the population size increases. Yet, stochastic simulations confirm that (3.5) correctly predicts the stationary distribution, which is relatively insensitive to the infected population size, but varies with $\delta + \gamma$ (figure 4a).

Second, we consider a general fitness landscape with a single maximum. It is possible to derive a good approximation for the stationary distribution (see electronic supplementary material, S.38):

$$\psi_{\text{approx}}(\alpha) = \frac{\beta(\alpha_0)}{\beta(\alpha)} \mathcal{N}(\alpha_0, s^2), \quad (3.6)$$

where α_0 is the virulence that maximizes R_0 , $\mathcal{N}(\alpha_0, s^2)$ is the Gaussian distribution with mean α_0 and variance $s^2 = 1/(nI_{\text{eq}}(\alpha_0)|R_0''(\alpha_0)|/R_0(\alpha_0))$, and $nI_{\text{eq}}(\alpha_0)$ is the

expected number of infected individuals at the endemic equilibrium when the virulence is α_0 . We thus see the effect of demographic stochasticity is to bias the Gaussian, putting more weight on values of the virulence below α_0 : if $R_0(\alpha_0) \geq R_0(\alpha)$, then, $\beta(\alpha_0)/\beta(\alpha) \geq (\delta + \alpha_0 + \gamma)/(\delta + \alpha + \gamma) > 1$ for $\alpha < \alpha_0$. This becomes more clear when we consider the mode and mean of the (true) stationary distribution (see electronic supplementary material, §6.3):

$$\alpha_{\text{mode}} = \alpha_0 - \frac{s^2}{\delta + \alpha_0 + \gamma} \quad (3.7)$$

and

$$\alpha_{\text{mean}} = \alpha_0 - s^2 \left(\frac{1}{\delta + \alpha_0 + \gamma} + \frac{|I'_{\text{eq}}(\alpha_0)|}{I_{\text{eq}}(\alpha_0)} + \frac{R_0'''(\alpha_0)}{|R_0''(\alpha_0)|} \right), \quad (3.8)$$

respectively. Equations (3.7) and (3.8) indicate that, as expected from the simple optimization approach used above in (3.4) and illustrated in figures 3 and 4, a lower infected population size tends to decrease pathogen virulence. However, the above derivation of the stationary distribution goes far beyond this optimization criterion. First, it accurately predicts the mode of the stationary distribution; in particular, it shows that the peakedness of the fitness landscape may affect the mode of the stationary distribution. The skew of the fitness landscape can also have huge effects on the stationary distribution (figure 3): a positive skew leads to a higher mean virulence and may thus counteract the effect of a small pathogen-infected population. In other words, whether demographic stochasticity favours lower or higher virulence also depends on the shape of the fitness landscape. Second, our analysis predicts the amount of variation one may expect to see around this mode. Unlike the criteria used to derive a single optimal strategy, our approach predicts accurately the expected variation around this mode (figures 3 and 4). Note that the population remains monomorphic most of the time (because mutation is assumed to be small) but the variance of the stationary distribution refers to the distribution of phenotypes explored through time (or through space if stochastic evolution is taking place in multiple isolated populations).

4. Discussion

Evolutionary theory has led to the development of different mathematical tools for studying phenotypic evolution in a broad diversity of ecological scenarios [47–49]. For instance, Adaptive Dynamics is a powerful theoretical framework to study life-history evolution when mutation is assumed to be rare so that demographic and evolutionary processes can be decoupled [5,6]. This analysis yields evolutionarily stable life-history strategies and captures the ultimate outcome of evolution. This approach, however, relies on the assumption that population size is infinite and that the epidemiological trajectory is deterministic. Finite population size, however, can also affect evolutionary trajectories. In particular, even the fittest genotype can be invaded by a deleterious mutant when population size is reduced. This leads to the collapse of the concept of evolutionarily stable strategy. On the other hand, population genetics allows us to consider the effect of finite population size and drift, but at the cost of assuming fixed population sizes and ignoring ecological processes (however, see, e.g. [38,50,51]). Here we build upon Stochastic Adaptive Dynamics (SAD) [43,45,49], a new theoretical framework where the evolutionary outcome of life-history evolution is studied by systematically scaling from ecologically complex individual-based stochastic models to the stationary distribution of the phenotype under mutation–selection–drift equilibrium. Under the assumption that the mutation rate is small, the long-term stochastic dynamics and equilibrium distribution can be derived from a diffusion approximation. In contrast with previous population genetics models, the present framework also allows life-history evolution to affect population size and, consequently, the amount of demographic stochasticity. In other words, this framework retains key strengths of Adaptive Dynamics but relaxes a major assumption by allowing genetic drift to affect the evolutionary outcome (see also [52], p. 1149). As such, our SAD framework is an important step towards a better integration between Adaptive Dynamics and classical population genetics.

We show that finite population size induces a selective pressure towards strains with lower variance in growth rate (but see also [35,39,46]). A simple way to understand this effect is to compare the fate of two strains with the same R_0 but with different life-history strategies. The *fast* strain is very transmissible but has a short duration of infection (e.g. because of high virulence or high clearance rate). The *slow* strain has a long duration of infection but a small transmission rate. As the two strains have the same R_0 , deterministic models predict that these two strains should coexist once common, but that neither can invade from small numbers. With finite population size, however, invasion is possible. When population size is assumed to be fixed, say N , the two strains are forced to share the same speed because when e.g. one strain infects a new host, the artificial constraint on the pathogen population size requires the death of a host infected by the other strain. By contrast, when population size is allowed to vary stochastically, the competing strains can have different speeds. The fast strain has a higher extinction rate simply because more events happen per unit of time: rare events, such as large fluctuations, will happen more regularly for the faster strain. As in Aesop's fable, 'slow and steady wins the race': when population size is allowed to vary stochastically, however, the race

has no finish line (unlike a fixed population model, you cannot hit N and 'win') and a strain can succeed only by outlasting its competitors. The advantage thus falls to the slower strain which persists by using longer infectious periods to 'wait out' periods of paucity of susceptibles, outliving the more volatile fast strain.

Previous studies [19,20] pointed out the influence of finite population size on the direction of virulence evolution, but they focused mainly on the quasi-neutral case where all the strains have the same R_0 . Humpalik *et al.* [20] did look at scenarios where strains have different R_0 , but assumed a fixed population, strong selection, and values of R_0 so large that all hosts are infected. None of these studies considered varying strengths of selection, and none provided a derivation of the stationary distribution at mutation–selection–drift equilibrium, which describes the long-term behaviour of pathogen virulence, that we believe is key to explore the interaction between finite population size and phenotypic evolution. This distribution yields testable predictions on the mean as well as other moments of the phenotypic distribution.

The approximation (3.6) shows that this distribution is moulded by two main parameters: (i) the pathogen fitness landscape, and (ii) the effective size of the infected host population. First, the pathogen fitness at the endemic equilibrium can be derived from (2.5) and depends mainly on the way R_0 varies with pathogen life-history traits. Under the classical transmission–virulence assumption, R_0 is maximized for some intermediate virulence. But the shape of the trade-off also affects the shape of the fitness landscape and in particular its symmetry. When the fitness landscape of the pathogen is symmetric, reducing the infected population size increases the variance of the stationary distribution but also decreases the mean (and the mode) of this distribution. This effect results from the selection for a reduction of the variance identified in (2.5). This is the effect that emerges in the quasi-neutral case. When the fitness landscape is flat, this may lead to an important bias towards lower virulence (figure 4). When the fitness landscape of the pathogen is asymmetric the skewness of the fitness landscape can affect the mean of the stationary distribution when the equilibrium host population size, nI_{eq} , is reduced. More specifically, negative (positive) skewness reduces (increases) the mean of the stationary distribution. It is interesting to note that classical functions used to model the trade-off between virulence and transmission tend to generate positive skewness in the fitness landscape [4,8,14]. The asymmetry of these fitness functions may thus counteract the effects of stochasticity *per se* identified in symmetric fitness landscapes. In other words, predictions on the stochastic evolutionary outcome are sensitive to the shape of genetic constraints acting on different pathogen life-history traits. This result is very similar to the deterministic effects discussed in [53,54] on the influence of asymmetric fitness landscapes on phenotypic evolution. Note, however, that the effect analysed by [54] is driven by environmental effects on phenotypes. In our model, we did not assume any environmental effects, and a given genotype is assumed to produce a single phenotype.

While we considered the standard SIR model, our approach can be generalized to consider a number of variants, including the SIRS model, the SEIR model, models with multiple exposed and infected compartments, etc. Our strong selection results for the fixation probability will apply whenever invasion implies fixation [55] (this

assumption is also necessary in general to derive (3.3); in particular, the diffusion approximation can fail if e.g. there is an evolutionary branching, such as in a model with co-infection [17]). Multi-type branching processes [56] would allow the addition of e.g. exposed classes, whereas general (non-Markovian) branching processes would allow the consideration of arbitrary distributions for the infectious period [57–59]. Weak selection results may be obtained when the exchange of stability between resident and invader results in a manifold of equilibria connecting the steady states. The structure of the SIR model considered here lends itself to computing the reduced diffusion (2.5); a general, albeit computation-heavy, method to derive the reduced equation is presented in [41]. If, on the other hand, multiple strains could coexist at a stable node or focus or a saddle point for the infinite population limit, then one would have to use large deviations theory or adapt the results of [60], respectively. An important extension would be to consider models with less variance in the infectious period than the exponential distribution considered here, as they could diminish the effects of demographic stochasticity.

We analysed the effects of demographic stochasticity induced by finite population size but environmental stochasticity may also affect evolution [36,61,62]. Environmental factors are known to have dramatic impacts on pathogen transmission and it would thus be particularly relevant to expand the current framework to account for the effects of random perturbations of the environment on pathogen evolution [63]. Indeed, although we focused our analysis on the stationary distribution at the endemic equilibrium of the classical SIR model, we can also explore the effect of demographic stochasticity on the transient evolutionary dynamics away from the endemic equilibrium, e.g. in epidemic scenarios, under bottlenecks, etc. [64].

Further, to focus on the effects of finite populations, we have considered a well-mixed population, whereas it is well known that spatial spreading can facilitate coexistence of competing strains and trade-offs between pathogen virulence and host mobility [65–67]. Moreover, spatial structure can result in smaller local effective population sizes, thus amplifying the effects of demographic stochasticity. Other factors may reduce the effective infected host population size as well. For instance, variance in transmission among infected hosts is likely to reduce the effective infected population size below nI_{eq} . One source of heterogeneity in transmissibility may be induced by public-health interventions (e.g. vaccination, drug treatments), but intrinsic behavioural or immunological heterogeneities among hosts may induce superspreading transmission routes as well [68,69]. As such,

a structured stochastic model would be an important extension.

Another possible extension of this model would be to analyse the effect of demographic stochasticity on the multi-locus dynamics of pathogens. Indeed, the interaction between genetic drift and selection is known to yield complex evolutionary dynamics resulting in the build-up of negative linkage disequilibrium between loci. But the analysis of this so-called Hill–Robertson effect is often restricted to population genetics models with fixed population size. The build-up of linkage disequilibrium in some epidemiological models has been discussed in some simulation models [70,71]. Our model provides a theoretical framework to explore the effect of finite population size on multi-locus dynamics of pathogens and to generate more accurate predictions on e.g. the evolution of drug resistance [72].

Finally, although we have presented our results in the context of pathogen evolution, it is hopefully clear that a very similar theoretical framework could be used to study other examples of life-history evolution in the context of demographic stochasticity. Current general life-history theory largely neglects the evolutionary consequences of stochasticity arising from small population sizes. Our results suggest that it would be profitable to determine what sorts of insights might be gained for life-history evolution more generally by using the type of theoretical framework developed here.

Data accessibility. Mathematica simulation code is available on request.

Authors' contributions. T.L.P., A.L., T.D. and S.G. designed the research, performed the research and wrote the paper.

Competing interests. We declare we have no competing interests.

Funding. Some of this research was performed while T.L.P. was supported by a Fondation Sciences Mathématiques de Paris postdoctoral fellowship. A.L. thanks the Center for Interdisciplinary Research in Biology (Collège de France) for funding. S.G. thanks the CNRS (PICS and PEPS MPI) for funding.

Acknowledgements. We thank four anonymous reviewers for their thorough reading and insightful comments that substantially improved the presentation of this work. S.G. thanks Gauthier Boaglio for his help in the development of the simulation code. Simulations were performed on the Montpellier Bioinformatics Biodiversity cluster.

Endnote

¹Equation (2.3) shows us that a rare mutant can invade if and only if $\Delta\beta/\Delta\alpha > N/S = R_0$ i.e. if the line through $(-\delta - \gamma, 0)$ with slope $\Delta\beta/\Delta\alpha$ lies above the curve $(\alpha, \beta(\alpha))$ at the resident virulence. If $\beta(\alpha)$ is concave, then for all α , $\Delta\beta/\Delta\alpha \leq \beta'(\alpha)$; in particular, if the line through $(-\delta - \gamma, 0)$ is tangent to the curve at α_0 , then no mutant can invade, and α_0 is an evolutionary stable strategy; see [73].

References

- Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds). 2005 *Adaptive dynamics of infectious diseases: in pursuit of virulence management*. Cambridge, UK: Cambridge University Press.
- Bull JJ, Lauring AS. 2014 Theory and empiricism in virulence evolution. *PLoS Pathog.* **10**, 1–3. (doi:10.1371/journal.ppat.1004387)
- Gandon S, Day T, Metcalf CJE, Grenfell BT. 2016 Forecasting epidemiological and evolutionary dynamics of infectious diseases. *Trends Ecol. Evol.* **31**, 776–788. (doi:10.1016/j.tree.2016.07.010)
- Frank SA. 1996 Models of parasite virulence. *Quat. Rev. Biol.* **71**, 37–78. (doi:10.1086/419267)
- Metz JAJ, Nisbet RM, Geritz SAH. 1992 How should we define 'fitness' for general ecological scenarios? *Trends Ecol. Evol.* **7**, 198–202. (doi:10.1016/0169-5347(92)90073-K)
- Geritz SAH, Kisdi G, Meszena E, Metz JAJ. 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Ecol. Evol.* **12**, 35–57. (doi:10.1023/A:1006554906681)
- Anderson RM, May RM, Anderson B. 1992 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press.
- Alizon S, Hurford A, Mideo N, Van Baalen M. 2009 Virulence evolution and the trade-off hypothesis:

- history, current state of affairs and the future. *J. Evol. Biol.* **22**, 245–259. (doi:10.1111/j.1420-9101.2008.01658.x)
9. Cressler CE, McLeod DV, Rozins C, Van Den Hoogen J, Day T. 2016 The adaptive evolution of virulence: a review of theoretical predictions and empirical tests. *Parasitology* **143**, 915–930. (doi:10.1017/S003118201500092X)
 10. Anderson RM, May RM. 1982 Coevolution of hosts and parasites. *Parasitology* **85**, 411–426. (doi:10.1017/S0031182000055360)
 11. May RM, Anderson RM. 1983 Epidemiology and genetics in the coevolution of parasites and hosts. *Proc. R. Soc. Lond. B* **219**, 281–313. (doi:10.1098/rspb.1983.0075)
 12. Nowak MA, May RM. 1994 Superinfection and the evolution of parasite virulence. *Proc. R. Soc. B* **255**, 81–89. (doi:10.1098/rspb.1994.0012)
 13. Andreasen V, Pugliese A. 1995 Pathogen coexistence induced by density-dependent host mortality. *J. Theor. Biol.* **177**, 159–165.
 14. van Baalen M, Sabelis MW. 1995 The dynamics of multiple infection and the evolution of virulence. *Am. Nat.* **146**, 881–910. (doi:10.1086/285830)
 15. Morozov A, Best A. 2012 Predation on infected host promotes evolutionary branching of virulence and pathogens' biodiversity. *J. Theor. Biol.* **307**, 29–36. (doi:10.1016/j.jtbi.2012.04.023)
 16. Cortez MH. 2013 When does pathogen evolution maximize the basic reproductive number in well-mixed host–pathogen systems? *J. Math. Biol.* **67**, 1533–1585. (doi:10.1007/s00285-012-0601-2)
 17. Lion S, Metz JAJ. 2018 Beyond R_0 maximisation: on pathogen evolution and environmental dimensions maximisation: on pathogen evolution and environmental dimensions. *Trends Ecol. Evol.* **33**, 458–473. (doi:10.1016/j.tree.2018.02.004)
 18. Lande R, Engen S, Saether B-E. 2003 *Stochastic population dynamics in ecology and conservation*. Oxford, UK: Oxford University Press.
 19. Kogan O, Khasin M, Meerson B, Schneider D, Myers CR. 2014 Two-strain competition in quasineutral stochastic disease dynamics. *Phys. Rev. E* **90**, 042149. (doi:10.1103/PhysRevE.90.042149)
 20. Humpik J, Hill AL, Nowak MA. 2014 Evolutionary dynamics of infectious diseases in finite populations. *J. Theor. Biol.* **360**, 149–162. (doi:10.1016/j.jtbi.2014.06.039)
 21. Read JM, Keeling MJ. 2007 Stochasticity generates an evolutionary instability for infectious disease. *Ecol. Lett.* **10**, 818–827. (doi:10.1111/j.1461-0248.2007.01078.x)
 22. Stearns SC. 1992 *The evolution of life histories*. Oxford, UK: Oxford University Press.
 23. Roff DA. 2002 *Life history evolution*. Oxford, UK: Oxford University Press.
 24. Doebeli M, Ispolatov Y, Simon B. 2017 Towards a mechanistic foundation of evolutionary theory. *eLife* **6**, e23804. (doi:10.7554/eLife.23804)
 25. McCallum H, Barlow N, Hone J. 2001 How should pathogen transmission be modelled? *Trends Ecol. Evol.* **16**, 295–300. (doi:10.1016/S0169-5347(01)02144-9)
 26. Parsons TL. 2012 Asymptotic analysis of some stochastic models from population dynamics and population genetics. PhD thesis, University of Toronto.
 27. Day T, Gandon S. 2006 Insights from Price's equation into evolutionary. In *Disease evolution: models, concepts, and data analyses* (eds ZL Feng, U Dieckmann, SA Levin). Providence, RI: American Mathematical Soc.
 28. Day T, Gandon S. 2007 Applying population-genetic models in theoretical evolutionary epidemiology. *Ecol. Lett.* **10**, 876–888. (doi:10.1111/j.1461-0248.2007.01091.x)
 29. Gandon S, Day T. 2007 The evolutionary epidemiology of vaccination. *J. R. Soc. Interface* **4**, 803–817. (doi:10.1098/rsif.2006.0207)
 30. Lenski RE, May RM. 1994 The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *J. Theor. Biol.* **169**, 253–265. (doi:10.1006/jtbi.1994.1146)
 31. Day T, Proulx SR. 2004 A general theory for the evolutionary dynamics of virulence. *Am. Nat.* **163**, E40–E63. (doi:10.1086/382548)
 32. Berggruber TW, Lion S, Gandon S. 2013 Evolution of suicide as a defence strategy against pathogens in a spatially structured environment. *Ecol. Lett.* **16**, 446–453. (doi:10.1111/ele.12064)
 33. Barbour AD. 1976 Quasi-stationary distributions in Markov population processes. *Adv. Appl. Prob.* **8**, 296–314. (doi:10.2307/1425906)
 34. Näsell I. 2001 Extinction and quasi-stationarity in the Verhulst logistic model. *J. Theor. Biol.* **211**, 11–27. (doi:10.1006/jtbi.2001.2328)
 35. Gillespie JH. 1974 Natural selection for within-generation variance in offspring number. *Genetics* **76**, 601–606.
 36. Frank SA, Slatkin M. 1990 Evolution in a variable environment. *Am. Nat.* **136**, 244–260. (doi:10.1086/285094)
 37. Proulx SR, Day T. 2002 What can invasion analyses tell us about evolution under stochasticity in finite populations? *Selection* **2**, 2–15. (doi:10.1556/Select.2.2001.1-2.2)
 38. Otto SP, Whitlock MC. 1997 The probability of fixation in populations of changing size. *Genetics* **146**, 723–733.
 39. Lambert A. 2006 Probability of fixation under weak selection: a branching process unifying approach. *Theor. Popul. Biol.* **69**, 419–441. (doi:10.1016/j.tpb.2006.01.002)
 40. Rousset F. 2004 *Genetic structure and selection in subdivided populations*. Princeton, NJ: Princeton University Press.
 41. Parsons TL, Rogers T. 2017 Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology. *J. Phys. A: Math. Theor.* **50**, 415601. (doi:10.1088/1751-8121/aa86c7)
 42. Gifford DR, de Visser JAGM, Wahl LM. 2012 Model and test in a fungus of the probability that beneficial mutations survive drift. *Biol. Lett.* **9**, 20120310. (doi:10.1098/rsbl.2012.0310)
 43. Champagnat N, Lambert A. 2007 Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *Ann. Appl. Prob.* **17**, 102–155. (doi:10.1214/105051606000000628)
 44. Lehmann L. 2012 The stationary distribution of a continuously varying strategy in a class-structured population under mutation–selection–drift balance. *J. Evol. Biol.* **25**, 770–787. (doi:10.1111/j.1420-9101.2012.02472.x)
 45. Debarre F, Otto SP. 2016 Evolutionary dynamics of a quantitative trait in a finite asexual population. *Theor. Popul. Biol.* **108**, 75–88. (doi:10.1016/j.tpb.2015.12.002)
 46. Parsons TL, Quince C. 2007 Fixation in haploid populations exhibiting density dependence II: the quasi-neutral case. *Theor. Popul. Biol.* **72**, 468–479. (doi:10.1016/j.tpb.2007.04.002)
 47. Parker GA, Smith JM. 1990 Optimality theory in evolutionary biology. *Nature* **348**, 27–33. (doi:10.1038/348027a0)
 48. Roff DA. 1993 *Evolution of life histories: theory and analysis*. Berlin, Germany: Springer Science & Business Media.
 49. Otto SP, Day T. 2007 *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton, NJ: Princeton University Press.
 50. Ewens WJ. 1967 The probability of survival of a new mutant in a fluctuating environment. *Heredity* **22**, 438–443. (doi:10.1038/hdy.1967.53)
 51. Kimura M, Ohta T. 1974 Probability of gene fixation in an expanding finite population. *Proc. Natl Acad. Sci. USA* **71**, 3377–3379. (doi:10.1073/pnas.71.9.3377)
 52. Waxman D, Gavrilets S. 2005 20 questions on adaptive dynamics. *J. Evol. Biol.* **18**, 1139–1154. (doi:10.1111/j.1420-9101.2005.00948.x)
 53. Mountford MD. 1968 The significance of litter-size. *J. Anim. Ecol.* **37**, 363–367. (doi:10.2307/2953)
 54. Urban MC, Bürger R, Bolnick DI. 2013 Asymmetric selection and the evolution of extraordinary defences. *Nat. Commun.* **4**, 2085. (doi:10.1038/ncomms3085)
 55. Geritz SAH. 2005 Resident–invader dynamics and the coexistence of similar strategies. *J. Math. Biol.* **50**, 67–82. (doi:10.1007/s00285-004-0280-8)
 56. Harris TE. 1963 *The theory of branching processes*. Berlin, Germany: Springer.
 57. Crump KS, Mode CJ. 1968 A general age-dependent branching process. I. *J. Math. Anal. Appl.* **24**, 494–508. (doi:10.1016/0022-247X(68)90005-X)
 58. Crump KS, Mode CJ. 1969 A general age-dependent branching process. II. *J. Math. Anal. Appl.* **25**, 8–17. (doi:10.1016/0022-247X(69)90210-8)
 59. Jagers P. 1975 *Branching processes with biological applications*. London, UK: Wiley.
 60. Kifer Y. 1981 The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point. *Isr. J. Math.* **40**, 74–96. (doi:10.1007/BF02761819)
 61. Starrfelt J, Kokko H. 2012 Bet-hedging—a triple trade-off between means, variances and

- correlations. *Biol. Rev.* **87**, 742–755. (doi:10.1111/j.1469-185X.2012.00225.x)
62. Schreiber SJ, Baer CF, Bronstein JL. 2015 Unifying within- and between-generation bet-hedging theories: an ode to J. H. Gillespie. *Am. Nat.* **186**, 792–796. (doi:10.1086/683657)
63. Nguyen A, Rajon E, Fouchet D, Pontier D, Rabinovich J, Gourbiere S, Menu F. 2015 Low virulence evolves as a bet-hedging strategy in fluctuating environment. *BioRxiv* 034041. (<https://doi.org/10.1101/034041>)
64. Parsons *et al.* (in preparation).
65. Osnas EE, Hurtado PJ, Dobson AP. 2015 Evolution of pathogen virulence across space during an epidemic. *Am. Nat.* **185**, 332–342. (doi:10.1086/679734)
66. Lion S, Gandon S. 2016 Spatial evolutionary epidemiology of spreading epidemics. *Proc. R. Soc. B* **283**, 20161170. (doi:10.1098/rspb.2016.1170)
67. Penczykowski RM, Laine A-L, Koskella B. 2016 Understanding the ecology and evolution of host–parasite interactions across scales. *Evol. Appl.* **9**, 37–52. (doi:10.1111/eva.12294)
68. Woolhouse MEJ *et al.* 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342. (doi:10.1073/pnas.94.1.338)
69. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)
70. Althaus CL, Bonhoeffer S. 2005 Stochastic interplay between mutation and recombination during the acquisition of drug resistance mutations in human immunodeficiency virus type 1. *J. Virol.* **79**, 13 572–13 578. (doi:10.1128/JVI.79.21.13572-13578.2005)
71. Fraser C. 2005 HIV recombination: what is the impact on antiretroviral therapy? *J. R. Soc. Interface* **2**, 489–503. (doi:10.1098/rsif.2005.0064)
72. Day T, Gandon S. 2012 The evolutionary epidemiology of multilocus drug resistance. *Evolution* **66**, 1582–1597. (doi:10.1111/j.1558-5646.2011.01533.x)
73. Charnov EL. 1976 Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* **9**, 129–136. (doi:10.1016/0040-5809(76)90040-X)

SUPPLEMENTARY INFORMATION
Pathogen evolution in finite populations: slow and steady spreads the best

Todd L. Parsons, Amaury Lambert, Troy Day and Sylvain Gandon

September 11, 2018

Contents

1	Introduction	3
2	A Stochastic Epidemiological Model with Multiple Pathogen Strains	3
3	Stochastic Differential Equation Formulation	5
3.1	Obtaining Equation (4)	9
4	Itô’s Formula and Derivation of Equation (5)	10
5	Probability of Fixation of a Mutant Pathogen	13
5.1	A Deterministic Limit and its Asymptotic Analysis	13
5.2	The Strong Selection Case	15
5.3	The Weak Selection Case	17
5.3.1	Computing the derivatives of π	24
5.3.2	Reduced Diffusion	26
5.3.3	Frequency Process	27
5.3.4	Results for $d = 2$	29
5.4	Reconciling the Strong and Weak Selection Results	32
6	Adaptive Dynamics	36
6.1	Stationary Distribution	40
6.2	Asymptotic Approximation to Stationary Distribution	41
6.3	Mean & Mode of the Stationary Distribution	44
6.3.1	Estimating the Mean	44
6.3.2	Estimating the Mode	45
7	Some Remarks Regarding Simulations	47
8	Some Rigorous Demonstrations	48
8.1	Proof of Proposition 2	48
8.2	Proof of Proposition 5	50
8.2.1	Macroscopic Initial Frequencies	50
8.2.2	Novel Strain in Small Number of Copies	58
8.3	Proof of Proposition 10	60

Glossary of Notation

n	system size
$\lambda^{(n)}$	immigration rate for susceptible individuals; $\lambda^{(n)} = \lambda + \mathcal{O}(\frac{1}{n})$
$\delta^{(n)}$	base mortality rate; $\delta^{(n)} = \delta + \mathcal{O}(\frac{1}{n})$
$\beta_i^{(n)}$	contact rate for strain i ; $\beta_i^{(n)} = \beta_i + \mathcal{O}(\frac{1}{n})$
$\alpha_i^{(n)}$	excess mortality for strain i ; $\alpha_i^{(n)} = \alpha_i + \mathcal{O}(\frac{1}{n})$
$\gamma_i^{(n)}$	recovery rate for strain i ; $\gamma_i^{(n)} = \gamma_i + \mathcal{O}(\frac{1}{n})$
$R_{0,i}$	$\frac{\beta_i}{\delta + \alpha_i + \gamma_i}$
$R_{0,i}^{(n)}$	$\frac{\beta_i^{(n)}}{\delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)}} = R_{0,i} (1 + \frac{r_i}{n}) + o(\frac{1}{n})$
$S^{(n)}(t)$	number of susceptible individuals at time t
$I_i^{(n)}(t)$	number of individuals infected with strain i at time t
$R^{(n)}(t)$	number of recovered individuals at time t
$N^{(n)}(t)$	total number of individuals at time t
$P_i^{(n)}(t)$	frequency of individuals infected with strain i at time t
$\mathbf{I}^{(n)}(t)$	$(I_1^{(n)}(t), \dots, I_d^{(n)}(t), N^{(n)}(t))$
$\mathbf{E}^{(n)}(t)$	$(S^{(n)}(t), I_1^{(n)}(t), \dots, I_d^{(n)}(t))$
$\bar{S}^{(n)}(t)$	density of susceptible individuals at time t
$\bar{I}_i^{(n)}(t)$	density of individuals infected with strain i at time t
$\bar{R}^{(n)}(t)$	density of recovered individuals at time t
$\bar{N}^{(n)}(t)$	total density of individuals at time t
$\bar{\mathbf{I}}^{(n)}(t)$	$(\bar{I}_1^{(n)}(t), \dots, \bar{I}_d^{(n)}(t))$
$\bar{\mathbf{E}}^{(n)}(t)$	$(\bar{S}^{(n)}(t), \bar{I}_1^{(n)}(t), \dots, \bar{I}_d^{(n)}(t), \bar{N}^{(n)}(t))$
$\hat{S}^{(n)}(t)$	density of susceptible individuals at time nt
$\hat{I}_i^{(n)}(t)$	density of individuals infected with strain i at time nt
$\hat{R}^{(n)}(t)$	density of recovered individuals at time nt
$\hat{N}^{(n)}(t)$	total density of individuals at time nt
$\hat{\mathbf{I}}^{(n)}(t)$	$(\hat{I}_1^{(n)}(t), \dots, \hat{I}_d^{(n)}(t))$
$\hat{\mathbf{E}}^{(n)}(t)$	$(\hat{S}^{(n)}(t), \hat{I}_1^{(n)}(t), \dots, \hat{I}_d^{(n)}(t), \hat{N}^{(n)}(t))$
$S(t)$	asymptotic density of susceptible individuals at time t
$I_i(t)$	asymptotic density of individuals infected with strain i at time t
$R(t)$	asymptotic density of recovered individuals at time t
$N(t)$	total asymptotic density of individuals at time t
$\mathbf{E}^{*,i}$	endemic equilibrium with resident strain i
$\mathbf{I}(t)$	$(I_1(t), \dots, I_d(t))$
$\mathbf{E}(t)$	$(S(t), I_1(t), \dots, I_d(t), N(t))$
$\hat{S}(t)$	asymptotic density of susceptible individuals in slow time limit
$\hat{I}_i(t)$	asymptotic density of individuals infected with strain i in slow time limit
$\hat{R}(t)$	asymptotic density of recovered individuals in slow time limit
$\hat{N}(t)$	asymptotic total density of individuals in slow time limit
$\hat{\mathbf{I}}(t)$	$(\hat{I}_1(t), \dots, \hat{I}_d(t))$
$\hat{\mathbf{E}}(t)$	$(\hat{S}(t), \hat{I}_1(t), \dots, \hat{I}_d(t), \hat{N}(t))$
$P_i(t)$	asymptotic frequency of individuals infected with strain i in slow time limit

1 Introduction

In this SI, we derive the results in the main text. Where suitable references exist in the literature, we keep the discussion informal, sketching how the results are obtained and referring to the appropriate references for rigorous proofs. Where they do not, we first give a heuristic derivation for a broader audience, whilst deferring the proofs to the end.

2 A Stochastic Epidemiological Model with Multiple Pathogen Strains

We consider a family of random processes $(S^{(n)}(t), I_1^{(n)}(t), \dots, I_d^{(n)}(t), R^{(n)}(t))$, indexed by a parameter n , the “system size” [31], which plays a role similar to the census population size in population genetics (see *e.g.*, [15, 11, 13]): it can be thought of as the area in which the population lives, determining the population density per unit area and the rate of immigration. Similarly to fixed-size population genetic models, we will consider the asymptotic behaviour of our model when n is large. $S^{(n)}(t)$, $I_1^{(n)}(t)$, \dots , $I_d^{(n)}(t)$, and $R^{(n)}(t)$ are the number of susceptible individuals, individuals infected with strain $i = 1, \dots, d$, and recovered individuals, respectively. We will write $N^{(n)}(t)$ for the total population size at time t , so that

$$N^{(n)}(t) = S^{(n)}(t) + I_1^{(n)}(t) + \dots + I_d^{(n)}(t) + R^{(n)}(t).$$

Using this notation, our compartmental model for the epidemic is represented graphically in Figure S.1.

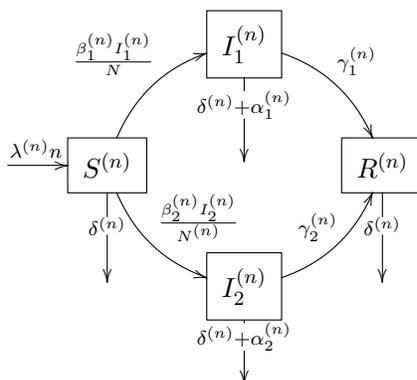


Figure S.1: Compartmental model of a two-strain SIR epidemic. Arrows indicate transitions between states and are labelled with the corresponding transition rate. Arrows into empty space indicate deaths.

Equivalently, we may describe our model as a continuous-time Markov chain \mathbf{E} taking values in \mathbb{N}^{d+2} with transition rates given in Table 1. When a transition occurs at time t , we will distinguish

Transition	Rate
$S \rightarrow S + 1$	$\lambda^{(n)}n$
$S \rightarrow S - 1$	$\delta^{(n)}S$
$S \rightarrow S - 1, I_k \rightarrow I_k + 1$	$\frac{\beta_k^{(n)}SI_k}{N}$
$I_k \rightarrow I_k - 1$	$(\delta^{(n)} + \alpha_k^{(n)})I_k$
$I_k \rightarrow I_k - 1, R \rightarrow R + 1$	$\gamma_k^{(n)}I_k$
$R \rightarrow R - 1$	$\delta^{(n)}R$

Table 1: Transition rates from the state $S^{(n)}(t) = S$, $I_k^{(n)}(t) = I_k$ and $R^{(n)}(t) = R$

between the value $\mathbf{E}(t-)$ of the Markov chain before the transition and its value $\mathbf{E}(t)$ after the transition.

All parameters given in Table 1 may depend on n , but are assumed to have a constant value to first approximation in n

$$\alpha_i^{(n)} = \alpha_i + \mathcal{O}\left(\frac{1}{n}\right), \quad \beta_i^{(n)} = \beta_i + \mathcal{O}\left(\frac{1}{n}\right), \quad \gamma_i^{(n)} = \gamma_i + \mathcal{O}\left(\frac{1}{n}\right), \quad \delta^{(n)} = \delta + \mathcal{O}\left(\frac{1}{n}\right), \quad \lambda^{(n)} = \lambda + \mathcal{O}\left(\frac{1}{n}\right) \quad (\text{S.1})$$

Simple calculations using the master equation tell us that in the absence of infected individuals, the expected value of $N^{(n)}(t)$ is

$$\mathbb{E}\left[N^{(n)}(t)\right] = e^{-\delta^{(n)}t}N^{(n)}(0) + \frac{\lambda^{(n)}}{\delta^{(n)}}\left(1 - e^{-\delta^{(n)}t}\right)n,$$

which approaches an equilibrium value of $\frac{\lambda^{(n)}}{\delta^{(n)}}n$ as $t \rightarrow \infty$. Thus, to first approximation, the total population size is proportional to n .

If one knows the values of $S^{(n)}(t)$ and $\mathbf{I}^{(n)}(t) := (I_1^{(n)}(t), \dots, I_d^{(n)}(t))$, then given one of $R^{(n)}(t)$ or $N^{(n)}(t)$ one can determine the other. For our purposes, it is more convenient to track the total population size, and consider the epidemic

$$\mathbf{E}^{(n)}(t) := \left(S^{(n)}(t), \mathbf{I}^{(n)}(t), N^{(n)}(t)\right).$$

In what follows, rather than working with $\mathbf{E}^{(n)}(t)$ we will focus on the rescaled process

$$\bar{S}^{(n)}(t) := \frac{1}{n}S^{(n)}(t), \quad \bar{I}_i^{(n)}(t) := \frac{1}{n}I_i^{(n)}(t), \quad \bar{N}^{(n)}(t) := \frac{1}{n}N^{(n)}(t),$$

and

$$\bar{\mathbf{E}}^{(n)}(t) := \left(\bar{S}^{(n)}(t), \bar{\mathbf{I}}^{(n)}(t), \bar{N}^{(n)}(t)\right).$$

$\bar{\mathbf{E}}^{(n)}(t)$ has the advantage of being a density dependent population process [21, 22, 23, 24] as generalized in [28]: the transition rates in (1) depend only on the densities $\bar{S}^{(n)}(t), \bar{\mathbf{I}}^{(n)}(t), \bar{N}^{(n)}(t)$ and not on the absolute numbers of individuals. As we discuss below, density dependent population processes have a number of nice features, including a law of large numbers and central limit

theorems.

Remark 1. To simplify our subsequent use of subscripts, we will consider $\bar{E}^{(n)}(t)$ as a process taking values in \mathbb{R}^{d+2} , the space of points

$$\mathbf{x} = (x_0, x_1, \dots, x_d, x_{d+1}),$$

and use $\bar{S}^{(n)}(t)$ and $\bar{E}_0^{(n)}(t)$, *etc.* interchangeably.

3 Stochastic Differential Equation Formulation

Here, we introduce a very convenient way of writing our Markov chain as the solution to a stochastic integral equation with the help of simple Poisson processes.

A Poisson process P is a Markov process making jumps of +1 exclusively, and such that $P(0) = 0$. A Poisson process P is called a simple Poisson process if it jumps at constant rate 1. In this case, $(P(at))$ is a Poisson process with rate a . This can be generalized by noting that $(P(\int_0^t a(s) ds))$ is a time-inhomogeneous Poisson process which jumps at rate $a(t)$ at time t . Similarly, there is a unique continuous-time Markov chain X satisfying

$$X(t) = x_0 + P \left(\int_0^t f(X(s-)) ds \right)$$

and when $X(t-) = x$, X jumps to $x + 1$ at rate $f(x)$.

Then it is not difficult to extend this (see Chapter 6, §4 in [14] for details) to our Markov process as follows:

$$\begin{aligned} S^{(n)}(t) &= S^{(n)}(0) + P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}t) \\ &\quad - P_{-\mathbf{e}_0 - \mathbf{e}_{d+1}} \left(\int_0^t \delta^{(n)} S^{(n)}(s) ds \right) - \sum_{i=1}^d P_{-\mathbf{e}_0 + \mathbf{e}_i} \left(\int_0^t \frac{\beta_i^{(n)} S^{(n)}(s) I_i^{(n)}(s)}{N^{(n)}(s)} ds \right) \\ I_i^{(n)}(t) &= I_i^{(n)}(0) + P_{-\mathbf{e}_0 + \mathbf{e}_i} \left(\int_0^t \frac{\beta_i^{(n)} S^{(n)}(s) I_i^{(n)}(s)}{N^{(n)}(s)} ds \right) \\ &\quad - P_{-\mathbf{e}_i - \mathbf{e}_{d+1}} \left(\int_0^t (\delta^{(n)} + \alpha_i^{(n)}) I_i^{(n)}(s) ds \right) - P_{-\mathbf{e}_i} \left(\int_0^t \gamma_i^{(n)} I_i^{(n)}(s) ds \right) \\ N^{(n)}(t) &= N^{(n)}(0) + P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}t) - P_{-\mathbf{e}_0 - \mathbf{e}_{d+1}} \left(\int_0^t \delta^{(n)} S^{(n)}(s) ds \right) \\ &\quad - P_{-\mathbf{e}_{d+1}} \left(\int_0^t \delta^{(n)} \left(N^{(n)}(s) - \sum_{i=1}^d I_i^{(n)}(s) - S^{(n)}(s) \right) ds \right) \\ &\quad - \sum_{i=1}^d P_{-\mathbf{e}_i - \mathbf{e}_{d+1}} \left(\int_0^t (\delta^{(n)} + \alpha_i^{(n)}) I_i^{(n)}(s) ds \right). \end{aligned}$$

where all the processes $P_{\mathbf{l}}(t)$ are independent, simple Poisson processes, indexed by the corresponding jumps, \mathbf{l} , of the Markov process $(\mathbf{E}^{(n)}(t))$ and \mathbf{e}_i is the i^{th} standard basis vector, the element

of \mathbb{R}^{d+2} with zeros everywhere except for a 1 at row i *i.e.*, an immigration event is indexed by $\mathbf{e}_0 + \mathbf{e}_{d+1}$, as it increases the number of susceptibles and the total population size by 1.

Changing variables, we get

$$\begin{aligned}
\bar{S}^{(n)}(t) &= \bar{S}^{(n)}(0) + \frac{1}{n} P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}t) \\
&\quad - \frac{1}{n} P_{-\mathbf{e}_0 - \mathbf{e}_{d+1}} \left(n \int_0^t \delta^{(n)} \bar{S}^{(n)}(s) ds \right) - \sum_{i=1}^d \frac{1}{n} P_{-\mathbf{e}_0 + \mathbf{e}_i} \left(n \int_0^t \frac{\beta_i^{(n)} \bar{S}^{(n)}(s) \bar{I}_i^{(n)}(s)}{\bar{N}^{(n)}(s)} ds \right) \\
\bar{I}_i^{(n)}(t) &= \bar{I}_i^{(n)}(0) + \frac{1}{n} P_{-\mathbf{e}_0 + \mathbf{e}_i} \left(n \int_0^t \frac{\beta_i^{(n)} \bar{S}^{(n)}(s) \bar{I}_i^{(n)}(s)}{\bar{N}^{(n)}(s)} ds \right) \\
&\quad - \frac{1}{n} P_{-\mathbf{e}_i - \mathbf{e}_{d+1}} \left(n \int_0^t (\delta^{(n)} + \alpha_i^{(n)}) \bar{I}_i^{(n)}(s) ds \right) - \frac{1}{n} P_{-\mathbf{e}_i} \left(n \int_0^t \gamma_i^{(n)} \bar{I}_i^{(n)}(s) ds \right) \\
\bar{N}^{(n)}(t) &= \bar{N}^{(n)}(0) + \frac{1}{n} P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}t) - \frac{1}{n} P_{-\mathbf{e}_0 - \mathbf{e}_{d+1}} \left(n \int_0^t \delta^{(n)} \bar{S}^{(n)}(s) ds \right) \\
&\quad - \frac{1}{n} P_{-\mathbf{e}_{d+1}} \left(n \int_0^t \delta^{(n)} \left(\bar{N}^{(n)}(s) - \sum_{i=1}^d \bar{I}_i^{(n)}(s) - \bar{S}^{(n)}(s) \right) ds \right) \\
&\quad - \sum_{i=1}^d \frac{1}{n} P_{-\mathbf{e}_i - \mathbf{e}_{d+1}} \left(n \int_0^t (\delta^{(n)} + \alpha_i^{(n)}) \bar{I}_i^{(n)}(s) ds \right).
\end{aligned}$$

This formalism is useful because it will allow us to write each r.h.s. as the sum of a deterministic trend and of a stochastic term with zero expectation.

Recall that the marginal value $P(t)$ of a simple Poisson process at time t is a Poisson random variable with parameter t . In particular, $P(t) - t$ has mean 0 and variance t . So if we write

$$\tilde{P}(t) := P(t) - t,$$

we are writing $P(t)$ as the sum of a deterministic trend t and of a stochastic term $\tilde{P}(t)$ with mean 0. If we come back to the example of the Markov process X jumping at rate $f(X)$, we can write

$$X(t) = x_0 + \int_0^t f(X(s-)) ds + M(t),$$

where we have set

$$M(t) := \tilde{P} \left(\int_0^t f(X(s-)) ds \right).$$

In addition, since the increments $\tilde{P}(t+s) - \tilde{P}(t)$ are independent of the past before t , have mean 0 and variance s , we can write the last equation in differential form

$$dX(t) = f(X(t-)) dt + dM(t),$$

with $dM(t) = U(t) - f(X(t-)) dt$, where $U(t)$ equals 1 iff P jumps at $\int_0^t f(X(s-)) ds$ and equals 0 otherwise. In particular, conditional on $X(t-) = x$, $dM(t)$ has mean 0 and variance $f(x) dt$. Thus, we also recover the infinitesimal variation of X as the sum of an infinitesimal trend in the dynamics

and of a stochastic fluctuation term with zero expectation.

Now let us return to our initial process. We adopt the same notation as previously, for example $\tilde{P}_{-e_0-e_{d+1}}(t) = P_{-e_0-e_{d+1}}(t) - t$ and

$$M_{-e_0-e_{d+1}}^{(n)}(t) := \tilde{P}_{-e_0-e_{d+1}} \left(n \int_0^t \delta^{(n)} \bar{S}^{(n)}(s) ds \right).$$

Proposition 1. *The infinitesimal variation of $\bar{\mathbf{E}}^{(n)}(t)$ can be written as the sum of an infinitesimal deterministic trend and of a stochastic fluctuation term with zero expectation:*

$$d\bar{S}^{(n)}(t) = F_0^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right) dt + \frac{1}{n} dM_{e_0+e_{d+1}}^{(n)}(t) - \frac{1}{n} dM_{-e_0-e_{d+1}}^{(n)}(t) - \frac{1}{n} \sum_{i=1}^d dM_{-e_0+e_i}^{(n)}(t)$$

$$d\bar{I}_i^{(n)}(t) = F_i^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right) dt + \frac{1}{n} dM_{-e_0+e_i}^{(n)}(t) - \frac{1}{n} dM_{-e_i-e_{d+1}}^{(n)}(t) - \frac{1}{n} dM_{-e_i}^{(n)}(t)$$

$$d\bar{N}^{(n)}(t) = F_{d+1}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right) dt + \frac{1}{n} dM_{e_0+e_{d+1}}^{(n)}(t) - \frac{1}{n} dM_{-e_0-e_{d+1}}^{(n)}(t) - \frac{1}{n} \sum_{i=1}^d dM_{-e_i-e_{d+1}}^{(n)}(t) - \frac{1}{n} dM_{-e_{d+1}}^{(n)}(t),$$

where

$$F_0^{(n)}(\mathbf{x}) = \lambda^{(n)} - \left(\sum_{i=1}^d \beta_i^{(n)} \frac{x_i}{x_{d+1}} + \delta^{(n)} \right) x_0$$

$$F_i^{(n)}(\mathbf{x}) = \left(\beta_i^{(n)} \frac{x_0}{x_{d+1}} - (\delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)}) \right) x_i$$

$$F_{d+1}^{(n)}(\mathbf{x}) = \lambda^{(n)} - \delta^{(n)} x_{d+1} - \sum_{i=1}^d \alpha_i^{(n)} x_i,$$

and $dM_{e_0+e_{d+1}}^{(n)}(t)$, $dM_{-e_0-e_{d+1}}^{(n)}(t)$, $dM_{-e_0+e_i}^{(n)}(t)$, $dM_{-e_i-e_{d+1}}^{(n)}(t)$, $dM_{-e_i}^{(n)}(t)$ and $dM_{-e_{d+1}}^{(n)}(t)$, are independent infinitesimal noise terms with mean zero and respective infinitesimal variances $n\rho_{e_0+e_{d+1}}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt$, $n\rho_{-e_0-e_{d+1}}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt$, $n\rho_{-e_0+e_i}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt$, $n\rho_{-e_i-e_{d+1}}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt$, $n\rho_{-e_i}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt$ and $n\rho_{-e_{d+1}}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt$, where

$$\rho_{e_0+e_{d+1}}^{(n)}(\mathbf{x}) = \lambda^{(n)}$$

$$\rho_{-e_0-e_{d+1}}^{(n)}(\mathbf{x}) = \delta^{(n)} x_0$$

$$\rho_{-e_0+e_i}^{(n)}(\mathbf{x}) = \beta_i^{(n)} \frac{x_0 x_i}{x_{d+1}}$$

$$\rho_{-e_i-e_{d+1}}^{(n)}(\mathbf{x}) = (\delta^{(n)} + \alpha_i^{(n)}) x_i$$

$$\rho_{-e_i}^{(n)}(\mathbf{x}) = \gamma_i^{(n)} x_i$$

$$\rho_{-e_{d+1}}^{(n)}(\mathbf{x}) = \delta^{(n)} \left(x_{d+1} - \sum_{i=0}^d x_i \right).$$

Remark 2. Note that $n\rho_i^{(n)}(\bar{\mathbf{E}}^{(n)}(t))$, is rate at which the Markov process $(\mathbf{E}^{(n)}(t))$ makes a jump \mathbf{l} .

Setting

$$\begin{aligned} \mathbf{M}^{(n)}(t) := & (\mathbf{e}_0 + \mathbf{e}_{d+1})M_{\mathbf{e}_0+\mathbf{e}_{d+1}}^{(n)}(t) - (\mathbf{e}_0 + \mathbf{e}_{d+1})M_{-\mathbf{e}_0-\mathbf{e}_{d+1}}^{(n)}(t) + \sum_{i=1}^d (\mathbf{e}_i - \mathbf{e}_0)M_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(t) \\ & - \sum_{i=1}^d (\mathbf{e}_i + \mathbf{e}_{d+1})M_{-\mathbf{e}_i-\mathbf{e}_{d+1}}^{(n)}(t) - \sum_{i=1}^d \mathbf{e}_i M_{-\mathbf{e}_i}^{(n)}(t) - \mathbf{e}_{d+1}M_{-\mathbf{e}_{d+1}}^{(n)}(t), \end{aligned} \quad (\text{S.2})$$

the result in the proposition can be written more compactly as

$$d\bar{\mathbf{E}}^{(n)}(t) = \mathbf{F}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt + \frac{1}{n}d\mathbf{M}^{(n)}(t), \quad (\text{S.3})$$

where

$$\mathbf{F}^{(n)}(\mathbf{x}) = (F_0^{(n)}(\mathbf{x}), F_1^{(n)}(\mathbf{x}), \dots, F_d^{(n)}(\mathbf{x}), F_{d+1}^{(n)}(\mathbf{x})).$$

Thus, the function $\mathbf{F}^{(n)}(\mathbf{x})$ describes the infinitesimal trend in the dynamics, whereas the terms $M_i^{(n)}(t)$ capture the de-trended fluctuations corresponding to each type of possible event. This equation is analogous to an Itô SDE, only now the driving noise is the discontinuous $\mathbf{M}^{(n)}(t)$, rather than the more familiar Brownian motion.

We note that for $i = 1, \dots, d$,

$$M_i^{(n)}(t) = M_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(t) - M_{-\mathbf{e}_i-\mathbf{e}_{d+1}}^{(n)}(t) - M_{-\mathbf{e}_i}^{(n)}(t),$$

so that $M_i^{(n)}(t)$ is independent of $M_j^{(n)}(t)$ for all $1 \leq i \neq j \leq d$, whereas

$$\mathbb{E} \left[dM_i^{(n)}(t)^2 \right] = \left(\beta_i^{(n)} \frac{S^{(n)}(t)}{N^{(n)}(t)} + (\delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)}) \right) I_i^{(n)}(t), \quad (\text{S.4})$$

two facts that will prove useful in what follows.

In particular, it makes sense to define $\mathbf{a}^{(n)}$ as the infinitesimal variance-covariance matrix of $\bar{\mathbf{E}}^{(n)}(t)$ by

$$a_{ij}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt := \text{Cov} \left[d\bar{E}_i^{(n)}(t), d\bar{E}_j^{(n)}(t) \right] = \frac{1}{n^2} \text{Cov} \left[dM_i^{(n)}(t), dM_j^{(n)}(t) \right]. \quad (\text{S.5})$$

Let us compute $\mathbf{a}^{(n)}$. Because all distinct terms in the definition (S.2) of $M^{(n)}$ are independent, all cross terms vanish. For example, for any $1 \leq i \leq d$

$$\begin{aligned} & \text{Cov} \left[d\bar{E}_0^{(n)}(t), d\bar{E}_i^{(n)}(t) \right] \\ &= \frac{1}{n^2} \text{Cov} \left[dM_{\mathbf{e}_0+\mathbf{e}_{d+1}}^{(n)}(t) - dM_{-\mathbf{e}_0-\mathbf{e}_{d+1}}^{(n)}(t) - \sum_{j=1}^d dM_{-\mathbf{e}_0+\mathbf{e}_j}^{(n)}(t), \right. \\ & \quad \left. dM_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(t) - dM_{-\mathbf{e}_i-\mathbf{e}_{d+1}}^{(n)}(t) - dM_{-\mathbf{e}_i}^{(n)}(t) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^2} \text{Cov} \left[-dM_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(t), dM_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(t) \right] \\
&= -\frac{1}{n^2} \text{Var} \left[dM_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(t) \right] = -\frac{1}{n} \rho_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(\bar{\mathbf{E}}^{(n)}(t)) dt,
\end{aligned}$$

which shows that $a_{0i}^{(n)}(\mathbf{x}) = -\frac{1}{n} \rho_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)}(\mathbf{x})$. We should not be surprised by the fact that this infinitesimal covariance is negative. Each event where a susceptible is infected by strain i has the effect of simultaneously decreasing the number of susceptibles and increasing the number of individuals infected by strain i .

By similar calculations, it is easy to see that

$$\begin{aligned}
a_{0,d+1}^{(n)}(\mathbf{x}) &= \frac{1}{n} \rho_{\mathbf{e}_0+\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}) + \frac{1}{n} \rho_{-\mathbf{e}_0-\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}), \\
a_{d+1,d+1}^{(n)}(\mathbf{x}) &= \frac{1}{n} \rho_{\mathbf{e}_0+\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}) + \frac{1}{n} \rho_{-\mathbf{e}_0-\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}) + \frac{1}{n} \sum_{i=1}^d \rho_{-\mathbf{e}_i-\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}) + \frac{1}{n} \rho_{-\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}),
\end{aligned}$$

and for any $1 \leq i \leq d$,

$$a_{i,d+1}^{(n)}(\mathbf{x}) = \frac{1}{n} \rho_{-\mathbf{e}_i-\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}).$$

Also for any $1 \leq i \neq j \leq d$, $a_{ij}^{(n)}(\mathbf{x}) = 0$ and

$$a_{jj}^{(n)}(\mathbf{x}) = \frac{1}{n} \rho_{-\mathbf{e}_0+\mathbf{e}_j}^{(n)}(\mathbf{x}) + \frac{1}{n} \rho_{-\mathbf{e}_j-\mathbf{e}_{d+1}}^{(n)}(\mathbf{x}) + \frac{1}{n} \rho_{-\mathbf{e}_j}^{(n)}(\mathbf{x}).$$

In other words, for any $0 \leq i, j \leq d+1$,

$$a_{ij}^{(n)}(\mathbf{x}) = \begin{cases} \frac{1}{n} \left(\frac{\beta_i^{(n)} x_0 x_j}{x_{d+1}} + (\delta^{(n)} + \alpha_j^{(n)} + \gamma_j^{(n)}) x_j \right) & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{S.6})$$

Equivalently,

$$\mathbf{a}^{(n)}(\mathbf{x}) = \frac{1}{n} \sum_{\mathbf{l}} \mathbf{l}^\top \mathbf{l} \rho_{\mathbf{l}}(\mathbf{x}), \quad (\text{S.7})$$

where the sum is over all possible jumps \mathbf{l} of the Markov process $(\mathbf{E}^{(n)}(t))$.

3.1 Obtaining Equation (4)

Note that similarly to Brownian motion, the infinitesimal mean of $\frac{1}{\sqrt{n}} M_{\mathbf{l}}^{(n)}(t)$ during the time interval dt is zero and its infinitesimal variance is $\rho_{\mathbf{l}}^{(n)}(\tilde{\mathbf{E}}^{(n)}(t)) dt$, whereas the jump size $\frac{1}{\sqrt{n}}$ tends to 0 as $n \rightarrow \infty$ (and thus, $\frac{1}{\sqrt{n}} M_{\mathbf{l}}^{(n)}(t)$ is approximately continuous for large n), we see that for large values of n , this noise is approximately equal to a Brownian motion with the same variance:

$$\frac{1}{\sqrt{n}} M_{\mathbf{l}}^{(n)}(t) \approx \sqrt{\rho_{\mathbf{l}}^{(n)}(\tilde{\mathbf{E}}^{(n)}(t))} dB_{\mathbf{l}}(t),$$

where all Brownian motions $B_{\mathbf{e}_0+\mathbf{e}_{d+1}}$, $B_{-\mathbf{e}_0-\mathbf{e}_{d+1}}$, $B_{-\mathbf{e}_0+\mathbf{e}_i}$, $B_{-\mathbf{e}_i-\mathbf{e}_{d+1}}$, $B_{-\mathbf{e}_i}$ and $B_{-\mathbf{e}_{d+1}}$ are independent.

This allows us to rewrite the results in Proposition 1 in the form of the following diffusion approximation for n large,

$$\begin{aligned} d\bar{S}^{(n)}(t) &\approx F_0^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right) dt + \frac{1}{\sqrt{n}} \sqrt{\rho_{\mathbf{e}_0+\mathbf{e}_{d+1}}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{\mathbf{e}_0+\mathbf{e}_{d+1}}(t) \\ &\quad - \frac{1}{\sqrt{n}} \sqrt{\rho_{-\mathbf{e}_0-\mathbf{e}_{d+1}}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_0-\mathbf{e}_{d+1}}(t) - \frac{1}{\sqrt{n}} \sum_{i=1}^d \sqrt{\rho_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_0+\mathbf{e}_i}(t) \end{aligned}$$

$$\begin{aligned} d\bar{I}_i^{(n)}(t) &\approx F_i^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right) dt + \frac{1}{\sqrt{n}} \sqrt{\rho_{-\mathbf{e}_0+\mathbf{e}_i}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_0+\mathbf{e}_i}(t) \\ &\quad - \frac{1}{\sqrt{n}} \sqrt{\rho_{-\mathbf{e}_i-\mathbf{e}_{d+1}}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_i-\mathbf{e}_{d+1}}(t) - \frac{1}{\sqrt{n}} \sqrt{\rho_{-\mathbf{e}_i}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_i}(t) \end{aligned}$$

$$\begin{aligned} d\bar{N}^{(n)}(t) &\approx F_{d+1}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right) dt + \frac{1}{\sqrt{n}} \sqrt{\rho_{\mathbf{e}_0+\mathbf{e}_{d+1}}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{\mathbf{e}_0+\mathbf{e}_{d+1}}(t) \\ &\quad - \frac{1}{\sqrt{n}} \sqrt{\rho_{-\mathbf{e}_0-\mathbf{e}_{d+1}}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_0-\mathbf{e}_{d+1}}(t) - \frac{1}{\sqrt{n}} \sum_{i=1}^d \sqrt{\rho_{-\mathbf{e}_i-\mathbf{e}_{d+1}}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_i-\mathbf{e}_{d+1}}(t) \\ &\quad - \frac{1}{\sqrt{n}} \sqrt{\rho_{-\mathbf{e}_{d+1}}^{(n)} \left(\bar{\mathbf{E}}^{(n)}(t) \right)} dB_{-\mathbf{e}_{d+1}}(t) \end{aligned}$$

(see [23] for a rigorous statement).

Setting

$$\bar{I}^{(n)}(t) := \sum_{l=1}^d \bar{I}_l^{(n)}(t),$$

so that

$$d\bar{I}^{(n)}(t) = \sum_{l=1}^d d\bar{I}_l^{(n)}(t),$$

and combining independent Brownian motions, we obtain equation (3) in the main text (*n.b.*, to simplify notation in the main text, we use S , I and R in lieu of $\bar{S}^{(n)}(t)$, $\bar{I}^{(n)}(t)$ and $\bar{N}^{(n)}(t)$).

We shall not use this diffusion approximation in the sequel, where we continue to consider the process with discrete jumps, (S.3).

4 Itô's Formula and Derivation of Equation (5)

As a first application of the SDE representation, we apply Itô's formula with jumps to our process to obtain an SDE for the proportion of each strain.

To motivate this, suppose we had a deterministic differential equation

$$\dot{\mathbf{Y}}(t) = \mathbf{f}(\mathbf{Y}(t))$$

and we let $X(t)$ be a deterministic real function of $\mathbf{Y}(t)$, say

$$X(t) := g(\mathbf{Y}(t))$$

where $g : \mathbb{R}^{d+2} \rightarrow \mathbb{R}$ is assumed to be continuously differentiable.

Then, applying the chain rule, we derive a differential equation satisfied by $X(t)$:

$$\dot{X}(t) = \sum_{j=0}^{d+1} \frac{\partial g}{\partial x_j}(\mathbf{Y}(t)) f_j(\mathbf{Y}(t))$$

or equivalently

$$X(t) = X(0) + \int_0^t \sum_{j=0}^{d+1} \frac{\partial g}{\partial x_j}(\mathbf{Y}(s)) f_j(\mathbf{Y}(s)) ds$$

The analogue of the chain rule in the fully stochastic case is the Meyer-Itô's formula (see *e.g.*, [29]).

$$\begin{aligned} X^{(n)}(t) &= X^{(n)}(0) + \int_0^t \sum_{j=0}^{d+1} \frac{\partial g}{\partial x_j}(\mathbf{E}^{(n)}(s)) F_j^{(n)}(\bar{\mathbf{E}}^{(n)}(s)) \\ &\quad + \frac{1}{2} \sum_{j,k=0}^{d+1} a_{jk}^{(n)}(\bar{\mathbf{E}}^{(n)}(s)) \frac{\partial^2 g}{\partial x_j \partial x_k}(\bar{\mathbf{E}}^{(n)}(s)) ds + \frac{1}{n} \int_0^t \sum_{j=0}^{d+1} \frac{\partial g}{\partial x_j}(\mathbf{E}^{(n)}(s)) dM_j^{(n)}(s) + \varepsilon^{(n)}(t), \end{aligned} \quad (\text{S.8})$$

where $\mathbf{a}^{(n)}(\mathbf{x})$ is the infinitesimal variance-covariance matrix of $\bar{\mathbf{E}}^{(n)}(t)$ defined in (S.6) and

$$\begin{aligned} \varepsilon^{(n)}(t) &= \sum_{s < t} g(\bar{\mathbf{E}}^{(n)}(s)) - g(\bar{\mathbf{E}}^{(n)}(s-)) - \sum_{j=0}^{d+1} \frac{\partial g}{\partial x_j}(\bar{\mathbf{E}}^{(n)}(s-)) \Delta \bar{E}_j^{(n)}(s) \\ &\quad - \frac{1}{2} \sum_{j,k=0}^{d+1} \frac{\partial^2 g}{\partial x_j \partial x_k}(\bar{\mathbf{E}}^{(n)}(s-)) \Delta \bar{E}_j^{(n)}(s) \Delta \bar{E}_k^{(n)}(s), \end{aligned} \quad (\text{S.9})$$

where the sum is over the times s of discontinuity of $\bar{\mathbf{E}}^{(n)}$. At a time t of discontinuity,

$$\Delta \bar{E}_j^{(n)}(t) := \bar{E}_j^{(n)}(t) - \bar{E}_j^{(n)}(t-)$$

denotes the magnitude of the jump in $\bar{E}_j^{(n)}$ at time t . The term $\varepsilon^{(n)}(t)$ correcting for discontinuities distinguishes the more general Meyer-Itô formula from the familiar Itô's formula for diffusions. In Section 8.1, we show that $\varepsilon^{(n)}(t) = \mathcal{O}(1/n^2)$.

Using this, we can derive Equation (5) from the main text. Let

$$\Pi_i(\mathbf{x}) = \frac{x_i}{\sum_{l=1}^d x_l}$$

so that

$$P_i^{(n)}(t) = \Pi_i(\bar{\mathbf{I}}^{(n)}(t)) = \Pi_i(\bar{\mathbf{E}}^{(n)}(t)).$$

is the proportion of the population infected with strain i . Since $P_i^{(n)}(t)$ is a deterministic function of $\bar{\mathbf{E}}^{(n)}(t)$, we can use Itô's formula (S.8) for jump processes. The following statement will be proved rigorously in Section 8.1.

Proposition 2. *The fraction of the population infected by strain i satisfies*

$$\begin{aligned} P_i^{(n)}(t) = & P_i^{(n)}(0) + \int_0^t P_i^{(n)}(s) \left(r_i^{(n)}(\bar{S}^{(n)}(s), \bar{N}^{(n)}(s)) - \sum_{j=1}^d r_j^{(n)}(\bar{S}^{(n)}(s), \bar{N}^{(n)}(s)) P_j^{(n)}(s) \right) \\ & + \frac{1}{n} \frac{1}{\sum_{l=1}^d \bar{I}_l^{(n)}(s)} P_i^{(n)}(s) \left(v_i^{(n)}(\bar{S}^{(n)}(s), \bar{N}^{(n)}(s)) - \sum_{j=1}^d v_j^{(n)}(\bar{S}^{(n)}(s), \bar{N}^{(n)}(s)) P_j^{(n)}(s) \right) ds \\ & + \frac{1}{n} \int_0^t \frac{1}{\sum_{l=1}^d \bar{I}_l^{(n)}(s)} \sum_{j=1}^d \left(\mathbb{1}_{\{i=j\}} - P_i^{(n)}(s) \right) dM_j^{(n)}(s) + \varepsilon_i^{(n)}(t), \end{aligned} \tag{S.10}$$

where

$$r_i^{(n)}(x_0, x_{d+1}) := \beta_i^{(n)} \frac{x_0}{x_{d+1}} - (\delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)})$$

gives the Malthusian growth rate of strain i , whereas

$$v_i^{(n)}(x_0, x_{d+1}) := \beta_i^{(n)} \frac{x_0}{x_{d+1}} + (\delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)})$$

is the infinitesimal variance associated with the growth of strain i . In addition, for any $T > 0$, there is a constant C such that for all $t \in [0, T]$, $\mathbb{P}\{|n^2 \varepsilon_i^{(n)}(t)| \geq C\}$ vanishes as $n \rightarrow \infty$.

To obtain Equation (5) in the main text, we omit the lower order error term $\varepsilon_i^{(n)}(t) = \mathcal{O}(1/n^2)$, recall that

$$I^{(n)}(t) = \sum_{l=1}^d I_l^{(n)}(t) = n \sum_{l=1}^d \bar{I}_l^{(n)}(t)$$

gives the total number of infectives, and observe that, similarly to the previous section,

$$\frac{1}{n} dM_i^{(n)} \approx \frac{1}{\sqrt{n}} \sqrt{v_i^{(n)}(\bar{S}^{(n)}(s), \bar{N}^{(n)}(s)) \bar{I}_i^{(n)}(t)} dB_i(t)$$

for independent Brownian motions B_1, \dots, B_d , so that

$$\frac{1}{n} \frac{1}{\sum_{l=1}^d \bar{I}_l^{(n)}(s)} dM_j^{(n)} \approx \frac{1}{\sqrt{I^{(n)}(t)}} \sqrt{v_i^{(n)}(\bar{S}^{(n)}(s), \bar{N}^{(n)}(s)) P_i^{(n)}(t)} dB_i(t).$$

5 Probability of Fixation of a Mutant Pathogen

In this section, we will be interested in the long time behaviour of our multi-strain stochastic epidemics. In particular, we tackle the problem of predicting which strains will be outcompeted and which strains will fix.

5.1 A Deterministic Limit and its Asymptotic Analysis

We begin this section with a result stating the convergence to a deterministic dynamical system as $n \rightarrow \infty$.

Proposition 3 (Theorem 2.2, [23]). *If*

$$\bar{S}^{(n)}(0) \rightarrow S(0), \quad \bar{I}_i^{(n)}(0) \rightarrow I_i(0), \quad \text{and} \quad \bar{N}^{(n)}(0) \rightarrow N(0)$$

as $n \rightarrow \infty$, then for any fixed $T > 0$, with probability 1,

$$\sup_{t \leq T} \|\bar{\mathbf{E}}^{(n)}(t) - \mathbf{E}(t)\| \rightarrow 0, \tag{S.11}$$

where $\mathbf{E}(t) := (S(t), I_1(t), \dots, I_d(t), N(t))$ is the solution to the following system of ordinary differential equations:

$$\dot{S}(t) = \lambda - \left(\sum_{i=1}^d \beta_i \frac{I_i(t)}{N(t)} + \delta \right) S(t), \tag{S.12a}$$

$$\dot{I}_i(t) = \left(\beta_i \frac{S(t)}{N(t)} - (\delta + \alpha_i + \gamma_i) \right) I_i(t), \tag{S.12b}$$

$$\dot{N}(t) = \lambda - \delta N(t) - \sum_{i=1}^d \alpha_i I_i(t), \tag{S.12c}$$

with initial conditions $S(0)$, $\mathbf{I}(0)$, and $N(0)$.

Note that the result in the previous proposition can be written more compactly as

$$\dot{\mathbf{E}} = \mathbf{F}(\mathbf{E}),$$

where

$$\mathbf{F}(\mathbf{x}) = \lim_{n \rightarrow \infty} \mathbf{F}^{(n)}(\mathbf{x}).$$

While we continue to work with the finite n fully stochastic process, the bifurcation structure of the deterministic system (S.12) will guide our analysis of the stochastic model. In particular, the steady states of this model, together with the degenerate case that arises when stability is exchanged between fixed points, give rise to two regimes that correspond to strong and weak selection in classical population genetics. To be explicit, let

$$R_{0,i} := \frac{\beta_i}{\delta + \alpha_i + \gamma_i}.$$

be the basic reproduction number of strain i . $R_{0,i}$ is the expected total number of new infections caused by a single infected individual, assuming an unlimited supply of susceptibles.

If $R_{0,i} \neq R_{0,j}$ for all $1 \leq i \neq j \leq d$, the equations (S.12) have $d + 1$ fixed points, one at $\mathbf{0}$ and one at the d equilibria where the population is infected by a single strain

$$\mathbf{E}^{*,i} := (S^{*,i}, I_1^{*,i}, \dots, I_d^{*,i}, N^{*,i}),$$

where

$$S^{*,i} := \frac{\lambda}{\delta R_{0,i}} \left(1 - \frac{\alpha_1(R_{0,i} - 1)}{\beta_1 - \alpha_1} \right), \quad I_j^{*,i} := \begin{cases} \frac{\lambda(R_{0,i} - 1)}{\beta_i - \alpha_i} & \text{if } i = j, \text{ and} \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad N^{*,i} := R_{0,i} S^*. \quad (\text{S.13})$$

When $d = 1$, it is shown in [32] when $\delta > \alpha_1$ that: if $R_{0,1} > 1$ then unique endemic equilibrium of the strain, $\bar{\mathbf{E}}^{*,1}$ is globally asymptotically stable, whereas if $R_{0,i} \leq 1$, the disease-free equilibrium $\mathbf{0}$ is globally asymptotically stable. The stability of fixed points is slightly more subtle when there is more than one strain.

Definition 1. We distinguish between two regimes of selection.

(i) The *strong selection* case, when $R_{0,1} > R_{0,i}$ for all $i > 1$, $\delta > \alpha_1$ and $R_{0,1} > 1$;

(ii) The *weak selection* case, $R_{0,1} = R_{0,i} = R_0^*$ for $i \leq m$, whilst $R_0^* > R_{0,j}$ for $m < j \leq d$.

Proposition 4. *The long term behavior of the deterministic system (S.12) differs according to the selection regime.*

(i) *In the strong selection case, the equilibrium state $\bar{\mathbf{E}}^{*,1}$ with strain 1 endemic and all other strains extinct is globally asymptotically stable from any initial condition for which $I_1(0) > 0$.*

(ii) *In the weak selection case, we arrive at a degenerate situation in which deterministic coexistence of strains $1, \dots, d$ is possible. Strains $m + 1, \dots, d$ will eventually disappear, whereas all points $\mathbf{x} \in \mathbb{R}_+^{d+2}$ such that*

$$\begin{aligned} \sum_{i=1}^m (\beta_i - \alpha_i) x_i &= \lambda(R_0^* - 1), \\ x_{m+1} &= \dots = x_d = 0, \\ x_{d+1} &= \frac{1}{\delta} \left(\lambda - \sum_{i=1}^m \alpha_i x_i \right), \\ x_{d+1} &= R_0^* x_0 \end{aligned} \quad (\text{S.14})$$

are fixed points for the system (S.12). The set Ω of such points is globally attracting, but no point in Ω is an attracting fixed point.

Proof. Point (i) follows by a direct adaptation of the result in [6]: rearranging (S.12b), we see that

$$\frac{1}{\beta_i} \frac{\dot{I}_i(t)}{I_i(t)} + \frac{1}{R_{0,i}} = \frac{S(t)}{N(t)} = \frac{1}{\beta_1} \frac{\dot{I}_1(t)}{I_1(t)} + \frac{1}{R_{0,1}},$$

so that

$$\left(\frac{I_i(t)}{I_i(0)}\right)^{\frac{1}{\beta_i}} e^{\frac{1}{R_{0,i}}t} = \left(\frac{I_1(t)}{I_1(0)}\right)^{\frac{1}{\beta_1}} e^{\frac{1}{R_{0,1}}t},$$

and, recalling that $I_1(t)$ is bounded for all $t > 0$, we see that for all $i > 1$

$$I_i(t) = I_i(0) \left(\frac{I_1(t)}{I_1(0)}\right)^{\frac{\beta_i}{\beta_1}} e^{-\beta_i \frac{R_{0,1} - R_{0,i}}{R_{0,1}R_{0,i}}t} \rightarrow 0$$

as $t \rightarrow \infty$.

The same argument shows that in the weak selection case, strains $m + 1, \dots, d$ will eventually disappear, whereas all points in Ω are fixed points. Moreover, all vectors \mathbf{u} tangent to Ω , *i.e.*, such that

$$\sum_{i=1}^m (\beta_i - \alpha_i) u_i = 0, \quad u_{m+1} = \dots = u_d = 0, \quad \text{and} \quad u_{d+1} = R_0^* u_0,$$

are all eigenvectors to the Jacobian of \mathbf{F} – evaluated at any $\mathbf{x} \in \Omega$ – corresponding to the eigenvalue 0. Thus, while Ω is a globally attracting set, no point in Ω is an attracting fixed point. \square

We now turn to the computation of the fixation probability of a novel strain in the fully stochastic system. Informed by the previous statement, we will consider two cases, strong and weak selection, where the dynamics of the process – and thus our approach to the fixation probabilities – are qualitatively different. We will then show, despite the difference in the approaches, and in the expressions for the fixation probability thereby obtained, that our two results for the fixation probability agree on all intermediate scalings, and may thus be combined (heuristically) via the method of matched asymptotic expansions, to obtain a single expression valid across all scales.

5.2 The Strong Selection Case

We begin by recalling that for the deterministic approximation (S.12) to apply, we required that $\bar{S}^{(n)}(0) \rightarrow S(0)$, $\bar{I}_i^{(n)}(0) \rightarrow I_i(0)$ and $\bar{N}^{(n)}(0) \rightarrow N(0)$ as $n \rightarrow \infty$. Unpacking this assumption, we see that

$$I_i^{(n)}(0) = nI_i(0) + o(n),$$

i.e., that a non-trivial portion of the population is already infected with strain i .

For any strain with $I_i^{(n)}(0) \ll n$, $\bar{I}_i^{(n)}(0) \rightarrow 0$, and thus $I_i(t) \equiv 0$ for all $t \leq T$, for any fixed $T > 0$: until $\mathcal{O}(n)$ individuals are infected, strain i is effectively invisible to the deterministic approximation on any finite time interval. This is not to say that the strain is absent, but rather, if we sample individuals from the population uniformly at random, the probability of sampling an individual infected with strain i is zero.

We will consider the case when a fixed number k of strain 2 individuals invade an established resident population. For our purposes, a strain i is established if it is initially present in macroscopic numbers, *i.e.*,

$$\bar{I}_i^{(n)}(0) \rightarrow I_i(0) > 0.$$

In light of the results in 5.1, we will assume that there is only a single resident strain, strain 1. We will first consider the case when the resident strain is in endemic equilibrium, and then generalise to the case when the resident strain, whilst still present in macroscopic numbers, is initially away from equilibrium.

Should the invading strain, strain 2, exceed εn individuals, for any $\varepsilon > 0$, we arrive again in the domain of applicability of the deterministic approximation ($\bar{I}_2^{(n)}(0) \rightarrow I_2(0) > \varepsilon > 0$). If the reproductive number of the invader is greater than that of the resident, $R_{0,2} > R_{0,1}$, then for n sufficiently large, the dynamics are essentially deterministic, and with high probability (*i.e.*, tending to 1 as $n \rightarrow \infty$) the process will in finite time T_ε enter an ε -neighbourhood of the fixed point $\mathbf{E}^{*,2}$ for arbitrarily small $\varepsilon > 0$. Once the process reaches this new equilibrium, we will see the resident strain is no longer viable, and subsequently disappears.

If we start at the endemic equilibrium of the resident strain 1, $\mathbf{E}^{*,1}$, then until $I_2^{(n)}$ exceeds εn , the epidemic process $\mathbf{E}^{(n)}(t)$ will remain close to that point. We thus have

$$\frac{S^{(n)}(t)}{N^{(n)}(t)} = \frac{\bar{S}^{(n)}(t)}{\bar{N}^{(n)}(t)} \approx \frac{1}{R_{0,1}},$$

and to first approximation, strain 2 has per-host transmission and clearance/mortality rates of

$$\frac{\beta_2}{R_{0,1}} \quad \text{and} \quad \delta + \alpha_2 + \gamma_2.$$

This latter is a birth and death process (see *e.g.*, [3]) which will go extinct with probability

$$q = \frac{\delta + \alpha_2 + \gamma_2}{\frac{\beta_2}{R_{0,1}}} = \frac{R_{0,1}}{R_{0,2}}$$

if $R_{0,2} > R_{0,1}$, and with probability $q = 1$ otherwise. This probability of extinction q is for a single initial individual infected with strain 2, and becomes q^k for k initial individuals. Now, a birth and death process either goes extinct or grows arbitrarily large, so with probability $1 - \frac{R_{0,2}}{R_{0,1}}$ it will eventually exceed εn .

Similarly, when we have reached a neighbourhood of $\mathbf{E}^{*,2}$ the transmission and clearance/mortality rates of strain 1 are approximately

$$\frac{\beta_1}{R_{0,2}} \quad \text{and} \quad \delta + \alpha_1 + \gamma_1.$$

Since $R_{0,2} > R_{0,1}$, this is a subcritical birth-death process which goes extinct with probability 1. Thus, invasion implies replacement, where for our purposes, the process *invades* if it exceeds εn individuals for some fixed $\varepsilon > 0$.

To summarize, we have heuristically derived

Proposition 5 (Strong Selection). *Consider a population infected with 2 strains such that $R_{0,2} > R_{0,1}$.*

- (i) [Macroscopic initial frequencies] *If $\bar{I}_2^{(n)}(0) \rightarrow I_2(0) > 0$, then strain 1 will go extinct with high probability.*
- (ii) [Novel strain in small number of copies, resident at endemic equilibrium] *Suppose that $\bar{I}_1^{(n)}(0) \rightarrow \bar{I}_1^{*,1}$, and that $I_2^{(n)}(0) = k$ for some fixed positive integer k . Then, provided $R_{0,2} > 1$, for any $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ I_1^{(n)}(t) = 0 \text{ and } \|\bar{\mathbf{E}}^{(n)}(t) - \bar{\mathbf{E}}^{*,2}\| < \varepsilon \text{ for some } t < \infty \right\} = 1 - \left(\frac{R_{0,1}}{R_{0,2}} \right)^k. \quad (\text{S.15})$$

In other words, the event that strain 1 becomes extinct asymptotically coincides with the event that strain 2 invades, which happens with a probability asymptotically equal to the probability of survival $1 - \left(\frac{R_{0,1}}{R_{0,2}} \right)^k$ of a time-homogeneous birth-death process; on this event, the system reaches in finite time the deterministic equilibrium (S.13) with only strain 2 endemic.

We give a rigorous proof of this result in Section 8.2, and compare (S.15) to fixation probabilities estimated from simulated epidemics in Figure S.4b.

5.3 The Weak Selection Case

Recall that weak selection corresponds to the case when $R_{0,i} = R_{0,j} = R_0^*$ for $1 \leq i, j \leq m$, whereas $R_0^* > R_{0,j}$ for $m < j \leq d$ (to simplify the discussion, we consider only the case where $m = d$). We hasten to clarify, however, that

$$R_{0,i}^{(n)} = \frac{\beta_i^{(n)}}{\delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)}},$$

so our assumptions (S.1) only impose that

$$R_{0,i}^{(n)} = R_0^* \left(1 + \frac{r_i}{n} \right) + o\left(\frac{1}{n}\right)$$

i.e., $R_{0,i}^{(n)}$ and $R_{0,j}^{(n)}$ are allowed to differ by $\mathcal{O}\left(\frac{1}{n}\right)$ terms; as we shall see below, this is analogous to the weak selection limit of classical population genetics, and the values r_i will appear as selection coefficients in a diffusion approximation.

In this case, we have a separation of timescales: there is a fast time-scale, in which (S.11) tells us that the stochastic process approximately follows the trajectories of (S.12) arbitrarily closely to an arbitrarily small neighbourhood of Ω . Then, as we discuss below, there is a slow-time scale, in which, having arrived at Ω , the stochastic process remains near this critical manifold.

Let

$$\begin{aligned}\hat{S}^{(n)}(t) &:= \bar{S}^{(n)}(nt) = \frac{1}{n}S^{(n)}(nt), \\ \hat{I}_i^{(n)}(t) &:= \bar{I}_i^{(n)}(nt) = \frac{1}{n}I_i^{(n)}(nt), \\ \hat{N}^{(n)}(t) &:= \bar{N}^{(n)}(nt) = \frac{1}{n}N^{(n)}(nt),\end{aligned}$$

and let

$$\hat{\mathbf{E}}^{(n)}(t) := (\hat{S}^{(n)}(t), \hat{\mathbf{I}}_i^{(n)}(t), \hat{N}^{(n)}(t)),$$

where, as before, we let $\hat{\mathbf{I}}^{(n)}(t) = (\hat{I}_1^{(n)}(t), \dots, \hat{I}_d^{(n)}(t))$, and note that

$$(\hat{S}^{(n)}(0), \hat{\mathbf{I}}^{(n)}(0), \hat{N}^{(n)}(0)) = (\bar{S}^{(n)}(0), \bar{\mathbf{I}}^{(n)}(0), \bar{N}^{(n)}(0)).$$

Here, rescaling time by n is analogous to the passage to so-called ‘‘coalescent time’’ or ‘‘generation time’’, which is used to derive the diffusion limit of the Wright-Fisher model in classical population genetics.

Recalling (S.3), the SDE for $\hat{\mathbf{E}}^{(n)}(t)$ is then

$$\begin{aligned}\hat{\mathbf{E}}^{(n)}(t) &= \hat{\mathbf{E}}^{(n)}(0) + \int_0^{nt} \mathbf{F}^{(n)}(\bar{\mathbf{E}}^{(n)}(s)) ds + \frac{1}{n}\mathbf{M}^{(n)}(nt) \\ &= \hat{\mathbf{E}}^{(n)}(0) + \int_0^t n\mathbf{F}^{(n)}(\hat{\mathbf{E}}^{(n)}(s)) ds + \frac{1}{n}\mathbf{M}^{(n)}(nt).\end{aligned}$$

Thus, in the slow time scale, the advective component is accelerated by a factor of n , causing the process to move rapidly to the critical manifold Ω ; as $n \rightarrow \infty$, this movement becomes instantaneous, and the process immediately jumps to Ω at time $t = 0$. Moreover, stochastic fluctuations away from Ω are restored instantaneously, so the process becomes ‘‘trapped’’ on Ω as $n \rightarrow \infty$. The following statement formalizes this idea using the projection $\boldsymbol{\pi}$ defined as follows. Let $\mathbf{E}(t, \mathbf{x}) = (S(t, \mathbf{x}), \mathbf{I}(t, \mathbf{x}), N(t, \mathbf{x}))$ be the solution to (S.12) with initial conditions $S(0) = x_0$, $I_i(0) = x_i$, and $N(0) = x_{d+1}$ and let

$$\boldsymbol{\pi}(\mathbf{x}) := \lim_{t \rightarrow \infty} \mathbf{E}(t, \mathbf{x}),$$

i.e., $\boldsymbol{\pi}(\mathbf{x})$ is the point on Ω at which the trajectory of (S.12) starting from \mathbf{x} meets Ω .

Proposition 6. As $n \rightarrow \infty$, $(\hat{\mathbf{E}}^{(n)}(t)) \Rightarrow (\hat{\mathbf{E}}(t))$ ¹ where the latter is a diffusion on the manifold Ω , solution to the system of stochastic differential equations

$$d\hat{\mathbf{E}}_i = \sum_{j=0}^{d+1} \frac{\partial \pi_i}{\partial x_j}(\hat{\mathbf{E}}) f_j(\hat{\mathbf{E}}) dt + \frac{1}{2} \sum_{j=0}^{d+1} \sum_{k=0}^{d+1} \frac{\partial^2 \pi_i}{\partial x_j \partial x_k}(\hat{\mathbf{E}}) a_{jk}(\hat{\mathbf{E}}) dt + \sum_{j=0}^{d+1} \sum_{k=1}^D \frac{\partial \pi_i}{\partial x_j}(\hat{\mathbf{E}}) \sigma_{jk}(\hat{\mathbf{E}}) dB_k(t) \quad (\text{S.16})$$

where the (B_k) denote $D = 3(d+1)$ independent standard Brownian motions,

$$\mathbf{f}(\mathbf{x}) := \lim_{n \rightarrow \infty} n \left(\mathbf{F}^{(n)}(\mathbf{x}) - \mathbf{F}(\mathbf{x}) \right),$$

and $\boldsymbol{\sigma}(\mathbf{x})$ is the $(d+1) \times D$ matrix

$$\begin{bmatrix} \sqrt{\lambda} & -\sqrt{\delta x_0} & -\sqrt{\frac{\beta_1 x_0 x_1}{x_{d+1}}} & 0 & 0 & -\sqrt{\frac{\beta_2 x_0 x_2}{x_{d+1}}} & \dots & 0 & -\sqrt{\frac{\beta_d x_0 x_d}{x_{d+1}}} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{\frac{\beta_1 x_0 x_1}{x_{d+1}}} & -\sqrt{(\delta + \alpha_1) x_1} & -\sqrt{\gamma_1 x_1} & 0 & \dots & \dots & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \dots & \dots & \dots & 0 & \sqrt{\frac{\beta_d x_0 x_d}{x_{d+1}}} & -\sqrt{(\delta + \alpha_d) x_d} & -\sqrt{\gamma_d x_d} & 0 \\ \sqrt{\lambda} & -\sqrt{\delta x_0} & 0 & -\sqrt{(\delta + \alpha_1) x_1} & 0 & 0 & \dots & \dots & 0 & -\sqrt{(\delta + \alpha_d) x_d} & 0 & -\sqrt{\delta(x_{d+1} - \sum_{i=0}^d x_i)} \end{bmatrix},$$

and

$$\mathbf{a}(\mathbf{x}) = \lim_{n \rightarrow \infty} n \mathbf{a}^{(n)}(\mathbf{x}) = \boldsymbol{\sigma}(\mathbf{x}) \boldsymbol{\sigma}(\mathbf{x})^\top.$$

In other words, for each i

$$d\hat{\mathbf{E}}_i(t) = \mathbf{D}\pi_i(\hat{\mathbf{E}}(t)) \mathbf{f}(\hat{\mathbf{E}}(t)) dt + \mathbf{D}\pi_i(\hat{\mathbf{E}}(t)) \boldsymbol{\sigma}(\hat{\mathbf{E}}(t)) d\mathbf{B}(t) + \frac{1}{2} \text{Tr} \left[\boldsymbol{\sigma}^\top(\hat{\mathbf{E}}(t)) \mathbf{H}\pi_i(\hat{\mathbf{E}}(t)) \boldsymbol{\sigma}(\hat{\mathbf{E}}(t)) \right] dt, \quad (\text{S.17})$$

where $\mathbf{D}\pi_i$ denotes the gradient vector of π_i , $\mathbf{H}\pi_i$ its Hessian matrix, and Tr denotes the trace operator.

This diffusion can be understood as the result of stochastic fluctuations around Ω immediately followed by a strong deterministic drift towards Ω .

As can be seen from Proposition 3, the drift pushes the process very rapidly onto Ω , so that in the limit, the process lives permanently in Ω . Now to understand the interplay between the deterministic dynamics towards Ω and the stochastic fluctuations around Ω , it is useful to think of the dynamics in two steps. Suppose that starting from a point $\hat{\mathbf{E}}(t-) \in \Omega$, the process \mathbf{E} has a jump \mathbf{l} . Then, the rescaled process $(\hat{\mathbf{E}}^{(n)}(t))$ has a jump $\frac{1}{n}\mathbf{l}$.

¹A family of random variables $\{X^{(n)}\}$ taking values in a space S is said to *converge weakly* to X if

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X^{(n)})] = \mathbb{E}[f(X)]$$

for all $f \in C(S)$; the values $\mathbb{E}[f(X)]$ completely characterise the distribution of X . Weak convergence is denoted by

$$X^{(n)} \Rightarrow X.$$

Here, S is the Skorokhod space $\mathbb{D}_{\mathbb{R}^{d+2}}[0, \infty)$ of right-continuous functions from $[0, \infty)$ to \mathbb{R}^{d+2} with left limits; the interested reader is referred to [5] for a very readable account of weak convergence on \mathbb{D} .

In a second step, it is immediately projected back to the manifold by the drift *at the new location*, so:

$$d\hat{\mathbf{E}}(t) = \boldsymbol{\pi} \left(\hat{\mathbf{E}}(t-) + \frac{1}{n} \mathbf{l} \right) - \hat{\mathbf{E}}(t-)$$

Thus, expanding the i -th component of the r.h.s. of the last equation and recalling that $\boldsymbol{\pi} \left(\hat{\mathbf{E}}(t-) \right) = \hat{\mathbf{E}}(t-)$ yields

$$\begin{aligned} d\hat{E}_i^{(n)}(t) &= \pi_i \left(\hat{\mathbf{E}}^{(n)}(t-) + \frac{1}{n} \mathbf{l} \right) - E_i^{(n)}(t-) \\ &= \frac{1}{n} \sum_j \frac{\partial \pi_i}{\partial x_j} (\hat{\mathbf{E}}^{(n)}(t-)) l_j + \frac{1}{n^2} \sum_j \sum_k \frac{1}{2} \frac{\partial^2 \pi_i}{\partial x_j \partial x_k} (\hat{\mathbf{E}}^{(n)}(t-)) l_j l_k + o\left(\frac{1}{n^2}\right). \end{aligned}$$

To determine $\mathbb{E}[d\hat{E}_i^{(n)}(t)]$ (*i.e.*, the i^{th} component of the drift in the diffusion approximation) we need only sum this over all possible jumps \mathbf{l} , weighted by their probabilities:

$$\begin{aligned} &\mathbb{E}[d\hat{E}_i^{(n)}(t)] \\ &= \sum_{\mathbf{l}} \left(\frac{1}{n} \sum_j \frac{\partial \pi_i}{\partial x_j} (\mathbf{E}^{(n)}(t-)) l_j + \frac{1}{n^2} \sum_j \sum_k \frac{1}{2} \frac{\partial^2 \pi_i}{\partial x_j \partial x_k} (\mathbf{E}^{(n)}(t-)) l_j l_k + o\left(\frac{1}{n}\right) \right) n \rho_{\mathbf{l}}(\mathbf{E}^{(n)}(t-)) d(nt) \\ &= n \sum_j \frac{\partial \pi_i}{\partial x_j} (\hat{\mathbf{E}}^{(n)}(t-)) F_j^{(n)}(\hat{\mathbf{E}}(t-)) dt + \frac{1}{2} \sum_j \sum_k \frac{\partial^2 \pi_i}{\partial x_j \partial x_k} (\hat{\mathbf{E}}^{(n)}(t-)) a_{jk}(\hat{\mathbf{E}}^{(n)}(t-)) dt, \end{aligned}$$

where, because we have rescaled time, $d(nt)$ replaces dt in probability of a jump at t . Recalling $\mathbf{F}(\hat{\mathbf{E}}(t)) = 0$ and (S.7), this yields the first term of the previous equation yields the first term of Eq (S.16).

A picture (Figure S.2) more immediately explains the emergence of the variance induced drift: unless the flow lines are parallel, jumps of identical magnitude and direction will be returned to the manifold Ω at different distances from the initial point, as one moves along the manifold:

Of course the rigorous way of obtaining the result is to use Itô's formula as done in the proof.

Proof. The weak convergence $\hat{\mathbf{E}}^{(n)} \Rightarrow \hat{\mathbf{E}}$ is proven in [19] (see [27] for an informal, applications-oriented discussion).

To characterise the limit $\hat{\mathbf{E}}$, we shall make use of $\boldsymbol{\pi}$. Unfortunately, $\boldsymbol{\pi}(\mathbf{x})$ is impossible to compute analytically, but we can still use it to obtain an SDE for $\hat{\mathbf{E}}(t)$. We first observe that if \mathbf{F} is twice-continuously differentiable, then $\boldsymbol{\pi}$ is as well [18]. The continuity of $\boldsymbol{\pi}$ then tells us that $\boldsymbol{\pi}(\hat{\mathbf{E}}^{(n)}(t)) \Rightarrow \boldsymbol{\pi}(\hat{\mathbf{E}}(t))$ as well. Since $\boldsymbol{\pi}$ has first and second derivatives, we may apply Itô's formula (see Section 4) to $\boldsymbol{\pi}(\hat{\mathbf{E}}^{(n)}(t))$:

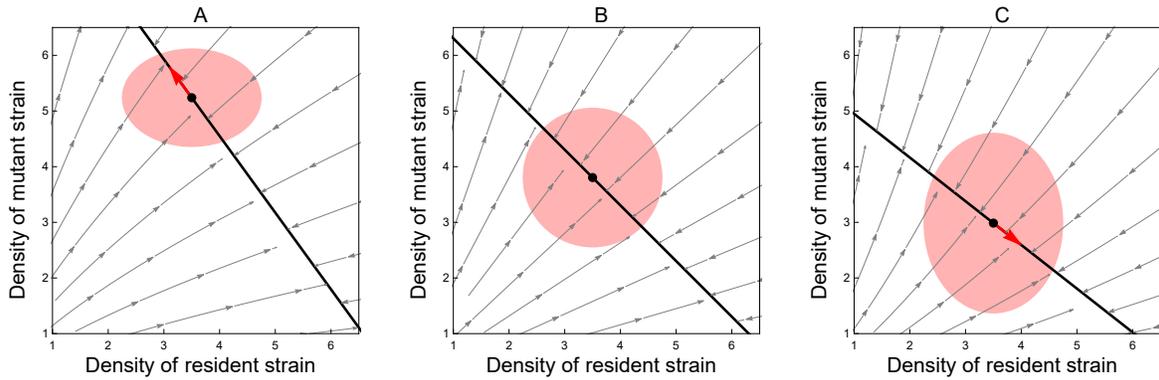


Figure S.2: Dynamics of the densities of the resident and the mutant strain in the phase plane when the two strains have the same basic reproduction number R_0 , but under three different scenarios. (A) The mutant strain has a lower virulence than the resident. (B) The two strains have the same virulence. (C) The mutant strain has a higher virulence than the resident. The deterministic trajectories are shown as grey arrows that point towards the manifold Ω (the black line). The light red ellipsoid has axes proportional to the infinitesimal variance of the jumps that displace each strain from a given point on the manifold (black dot). The combination of the effect of stochasticity and the fast deterministic return to the manifold generates a drift (red arrow) that favours the strain with the lower virulence. Parameter values of the resident: $\beta_1 = 10$, $\alpha_1 = 2$, $\delta = 0.05$, $\gamma = 0.5$. Virulence of the mutant: $\alpha_2 = 1.25, 2$, and 2.75 in A, B and C, respectively.

$$\begin{aligned}
\pi_i(\hat{\mathbf{E}}^{(n)}(t)) &= \pi_i(\hat{\mathbf{E}}^{(n)}(0)) \\
&+ \int_0^t \sum_{j=0}^{d+1} n \frac{\partial \pi_i}{\partial x_j}(\mathbf{E}^{(n)}(s)) F_j^{(n)}(\hat{\mathbf{E}}^{(n)}(s)) + \frac{1}{2} \sum_{j,k=0}^{d+1} na_{jk}^{(n)}(\hat{\mathbf{E}}^{(n)}(s)) \frac{\partial \pi_i}{\partial x_j \partial x_k}(\mathbf{E}^{(n)}(s)) ds \\
&+ \int_0^t \frac{1}{n} \sum_{j=0}^{d+1} \frac{\partial \pi_i}{\partial x_j}(\mathbf{E}^{(n)}(s)) dM_j^{(n)}(ns) + \varepsilon_i^{(n)}(nt) \quad (\text{S.18})
\end{aligned}$$

where, as before, $a_{jk}^{(n)}(\mathbf{x})$ is given by (S.6) and $\varepsilon_i^{(n)}(nt)$ is a smaller order error term.

On first inspection, it might appear that the drift term, which is multiplied by n , explodes as $n \rightarrow \infty$; however, from the definition of $\boldsymbol{\pi}$, we see that $\boldsymbol{\pi}(\mathbf{E}(t, \mathbf{x})) = \boldsymbol{\pi}(\mathbf{x})$, and thus,

$$0 = \frac{d}{dt} \Big|_{t=0} \boldsymbol{\pi}(\mathbf{E}(t, \mathbf{x})) = \sum_{j=0}^{d+1} \frac{\partial \pi_i}{\partial x_j}(\mathbf{x}) F_j(\mathbf{x}),$$

and the terms of order $\mathcal{O}(n)$ vanish identically.

We can thus replace $\mathbf{F}^{(n)}$ by $\mathbf{F}^{(n)} - \mathbf{F}$ in (S.18), leaving

$$\sum_{j=0}^{d+1} \frac{\partial \pi_i}{\partial x_j}(\mathbf{E}^{(n)}(s)) n \left(F_j^{(n)}(\hat{\mathbf{E}}^{(n)}(s)) - F_j(\hat{\mathbf{E}}^{(n)}(s)) \right),$$

which remains bounded, as our assumptions (S.1) guarantee that $\mathbf{F}^{(n)} - \mathbf{F}$ is $\mathcal{O}(\frac{1}{n})$.

Next, we recall (S.2)

$$\begin{aligned}
M^{(n)}(t) &= (\mathbf{e}_0 + \mathbf{e}_{d+1}) M_{\mathbf{e}_0 + \mathbf{e}_{d+1}}^{(n)}(t) - (\mathbf{e}_0 + \mathbf{e}_{d+1}) M_{-\mathbf{e}_0 - \mathbf{e}_{d+1}}^{(n)}(t) \\
&+ \sum_{i=1}^d (\mathbf{e}_i - \mathbf{e}_0) M_{-\mathbf{e}_0 + \mathbf{e}_i}^{(n)}(t) - \sum_{i=1}^d (\mathbf{e}_i + \mathbf{e}_{d+1}) M_{-\mathbf{e}_i - \mathbf{e}_{d+1}}^{(n)}(t) - \sum_{i=1}^d \mathbf{e}_i M_{-\mathbf{e}_i}^{(n)}(t) - \mathbf{e}_{d+1} M_{-\mathbf{e}_{d+1}}^{(n)}(t),
\end{aligned}$$

where, for example,

$$M_{-\mathbf{e}_0 - \mathbf{e}_{d+1}}^{(n)}(t) = \tilde{P}_{-\mathbf{e}_0 - \mathbf{e}_{d+1}} \left(\int_0^t \delta^{(n)} S^{(n)}(s) ds \right).$$

Thus,

$$\begin{aligned}
\frac{1}{n} M_{-\mathbf{e}_0 - \mathbf{e}_{d+1}}^{(n)}(nt) &= \tilde{P}_{-\mathbf{e}_0 - \mathbf{e}_{d+1}} \left(\int_0^{nt} \delta^{(n)} S^{(n)}(s) ds \right) \\
&= \frac{1}{n} \tilde{P}_{-\mathbf{e}_0 - \mathbf{e}_{d+1}} \left(n^2 \int_0^t \delta^{(n)} \bar{S}^{(n)}(ns) ds \right).
\end{aligned}$$

The latter is a stochastic process with jumps of order $\frac{1}{n}$ and variance

$$\int_0^t \delta^{(n)} \bar{S}^{(n)}(ns) ds = \int_0^t \delta^{(n)} \hat{E}_0^{(n)}(s) ds.$$

Thus, as $n \rightarrow \infty$, $\frac{1}{n}M_{-\mathbf{e}_0-\mathbf{e}_{d+1}}^{(n)}(nt)$ approaches a continuous stochastic process with variance

$$\int_0^t \delta \hat{E}_0(s) ds.$$

The martingale central limit theorem (see *e.g.*, [14]) tells us that the only stochastic process with these properties is a Brownian motion with the same variance,

$$\int_0^t \sqrt{\delta \hat{E}_0(s)} dB_{-\mathbf{e}_0-\mathbf{e}_{d+1}}(s).$$

(*i.e.*, $B_{-\mathbf{e}_0-\mathbf{e}_{d+1}}(t)$ is a standard Brownian motion with mean 0 and variance t).

Proceeding similarly, in the limit, we may replace all the terms $M_{\mathbf{l}}^{(n)}$ with integrals of independent Brownian motions, so that as $n \rightarrow \infty$, $\frac{1}{n}\mathbf{M}^{(n)}(nt)$ approaches

$$\int_0^t \sigma(\hat{\mathbf{E}}(s)) d\mathbf{B}(s)$$

where

$$\begin{aligned} \mathbf{B}(t) = & (B_{\mathbf{e}_0+\mathbf{e}_{d+1}}(t), B_{-\mathbf{e}_0-\mathbf{e}_{d+1}}(t), B_{-\mathbf{e}_0+\mathbf{e}_1}(t), B_{-\mathbf{e}_1-\mathbf{e}_{d+1}}(t), B_{-\mathbf{e}_1}(t), \\ & \dots, B_{-\mathbf{e}_0+\mathbf{e}_d}(t), B_{-\mathbf{e}_d-\mathbf{e}_{d+1}}(t), B_{-\mathbf{e}_d}(t), B_{-\mathbf{e}_{d+1}}(t)) \end{aligned}$$

is an ordered list of the D Brownian motions corresponding to the D noises $M_{\mathbf{l}}^{(n)}(t)$ and $\sigma(\mathbf{x})$ is as in the statement. Taking the limit as $n \rightarrow \infty$ on both sides of (S.18) and recalling that $\pi(\hat{\mathbf{E}}(t)) = \hat{\mathbf{E}}(t)$, we obtain (S.16). \square

While the drift terms seem rather mysterious, they may be interpreted geometrically. We first observe that

Proposition 7. $(D\pi)(\mathbf{x})$ is the projection onto the tangent space to Ω at \mathbf{x} , $T_{\mathbf{x}}\Omega$.

Proof. We first observe that, since $\pi(\mathbf{x}) \in \Omega$ for all \mathbf{x} , we must have

$$\pi(\pi(\mathbf{x})) = \pi(\mathbf{x}).$$

If, moreover, $\mathbf{x} \in \Omega$, we also have $\pi(\pi(\mathbf{x})) = \mathbf{x}$, so taking derivatives on left and right, using the chain rule, we have that

$$(D\pi)(\pi(\mathbf{x}))(D\pi)(\mathbf{x}) = \mathbb{I},$$

where \mathbb{I} denotes the identity matrix. Now, since $\mathbf{x} \in \Omega$, the right hand side is equal to

$$(D\pi)(\mathbf{x})(D\pi)(\mathbf{x}),$$

so we have that $(D\pi)(\mathbf{x})$ is a projection. It remains to see that it is a projection onto the tangent space. We will do so by showing it's image contains, and is contained by, the tangent space.

For the former, we recall that a vector \mathbf{X} is in the tangent space to Ω if and only if there exists a

parametric curve $\sigma_{\mathbf{x}, \mathbf{X}}(t)$ such that

- (i) $\sigma_{\mathbf{x}, \mathbf{X}}(0) = \mathbf{x}$,
- (ii) $\dot{\sigma}_{\mathbf{x}, \mathbf{X}}(0) = \mathbf{X}$, and,
- (iii) $\sigma_{\mathbf{x}, \mathbf{X}}(t) \in \Omega$ for all $t \in \mathbb{R}$.

We then have $\pi(\sigma_{\mathbf{x}, \mathbf{X}}(t)) = \sigma_{\mathbf{x}, \mathbf{X}}$, and thus

$$(\mathbf{D}\pi)(\mathbf{x})\mathbf{X} = \left. \frac{d}{dt} \right|_{t=0} \pi(\sigma_{\mathbf{x}, \mathbf{X}}(t)) = \dot{\sigma}_{\mathbf{x}, \mathbf{X}}(0) = \mathbf{X},$$

and thus $T_{\mathbf{x}}\Omega$ is in the image of $(\mathbf{D}\pi)(\mathbf{x})$.

On the other hand, since $\pi(\mathbf{x}) \in \Omega$, we have $\mathbf{F}(\pi(\mathbf{x})) = \mathbf{0}$, and again, taking derivatives using the chain rule, we have

$$(\mathbf{D}\mathbf{F})(\pi(\mathbf{x}))(\mathbf{D}\pi)(\mathbf{x}) = \mathbf{0},$$

so that if $\mathbf{x} \in \Omega$, we have $(\mathbf{D}\mathbf{F})(\mathbf{x})(\mathbf{D}\pi)(\mathbf{x}) = \mathbf{0}$ and thus

$$(\mathbf{D}\mathbf{F})(\mathbf{x})(\mathbf{D}\pi)(\mathbf{x})\mathbf{X} = \mathbf{0}$$

i.e., the image of $(\mathbf{D}\pi)(\mathbf{x})$ is contained in the kernel of $(\mathbf{D}\mathbf{F})(\mathbf{x})$, which we have already observed is $T_{\mathbf{x}}\Omega$. Thus, $\text{Im}((\mathbf{D}\pi)(\mathbf{x})) = T_{\mathbf{x}}\Omega$. \square

Thus, the drift vector $(\mathbf{D}\pi)(\mathbf{x})\mathbf{f}(\mathbf{x})$ from (S.17) is the projection of the vector $\mathbf{f}(\mathbf{x})$ onto the tangent space to Ω . This is an immediate consequence of the strong drift: in the absence of constraints, the process would move (on average) in the direction of this vector, whose components are the relative fitness of each strain, multiplied by the density of that strain. However, density limitation prevents unlimited growth, confining the process to the manifold Ω , and thus the direction of motion to the tangent plane, and the strains experience a drift that is the best approximating vector to their unconstrained growth rates.

5.3.1 Computing the derivatives of π

To complete our derivation of the equations for the limiting process $\hat{\mathbf{E}}(t)$, we must compute the derivatives of the π_i .

Proposition 8. *Let $\mathbf{x} \in \Omega$ and $0 \leq i \leq d+1$. The first partial derivatives of π_i at \mathbf{x} are given by $\frac{\partial \pi_i}{\partial x_k} = 0$ if $k = 0$ or $d+1$, otherwise by*

$$\frac{\partial \pi_i}{\partial x_k} = \mathbb{1}_{\{i=k\}} - \frac{(\beta_k - \alpha_k)\beta_i x_i}{\sum_{j=1}^d (\beta_j - \alpha_j)\beta_j x_j}.$$

The second partial derivatives π_i at \mathbf{x} are given for any k, n both different from 0 and $d+1$, by:

$$\frac{\partial^2 \pi_i}{\partial x_k \partial x_n} = \frac{\beta_i}{\sum_{j=1}^d (\beta_j - \alpha_j)\beta_j x_j} \left(-(\beta_k - \alpha_k)\mathbb{1}_{\{n=i\}} - (\beta_n - \alpha_n)\mathbb{1}_{\{k=i\}} \right)$$

$$+ \frac{(\beta_k - \alpha_k)(\beta_n - \alpha_n)x_i}{\sum_{j=1}^d (\beta_j - \alpha_j)\beta_j x_j} \left(\beta_k + \beta_n + \beta_i - \frac{\sum_{j=1}^d (\beta_j - \alpha_j)\beta_j^2 x_j}{\sum_{j=1}^d (\beta_j - \alpha_j)\beta_j x_j} \right).$$

Proof. We recall that under the weak selection hypothesis,

$$\dot{I}_i(t) = \beta_i \left(\frac{S(t)}{N(t)} - \frac{1}{R_0^*} \right) I_i(t),$$

so that

$$\frac{dI_i}{dI_j} = \frac{\beta_i I_i}{\beta_j I_j}.$$

We can solve this to obtain

$$\frac{1}{\beta_i} \ln \left(\frac{I_i(t)}{I_i(0)} \right) = \frac{1}{\beta_j} \ln \left(\frac{I_j(t)}{I_j(0)} \right),$$

for all i, j , *i.e.*,

$$\frac{1}{\beta_i} \ln \left(\frac{I_i(t, \mathbf{x})}{x_i} \right) = \frac{1}{\beta_j} \ln \left(\frac{I_j(t, \mathbf{x})}{x_j} \right),$$

and, taking the limit as $t \rightarrow \infty$,

$$\frac{1}{\beta_i} \ln \left(\frac{\pi_i(\mathbf{x})}{x_i} \right) = \frac{1}{\beta_j} \ln \left(\frac{\pi_j(\mathbf{x})}{x_j} \right). \quad (\text{S.19})$$

Taking derivatives, we then have

$$\begin{aligned} \frac{1}{\beta_i} \left(\frac{1}{\pi_i} \frac{\partial \pi_i}{\partial x_k} - \frac{1}{x_i} \mathbb{1}_{\{k=i\}} \right) &= \frac{1}{\beta_j} \left(\frac{1}{\pi_j} \frac{\partial \pi_j}{\partial x_k} - \frac{1}{x_j} \mathbb{1}_{\{k=j\}} \right), \\ \frac{1}{\beta_i \pi_i} \frac{\partial \pi_i}{\partial x_0} &= \frac{1}{\beta_j \pi_j} \frac{\partial \pi_j}{\partial x_0} \\ \frac{1}{\beta_i \pi_i} \frac{\partial \pi_i}{\partial x_{d+1}} &= \frac{1}{\beta_j \pi_j} \frac{\partial \pi_j}{\partial x_{d+1}}, \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\beta_i} \left(-\frac{1}{\pi_i^2} \frac{\partial \pi_i}{\partial x_k} \frac{\partial \pi_i}{\partial x_n} + \frac{1}{\pi_i} \frac{\partial^2 \pi_i}{\partial x_k \partial x_n} + \frac{1}{x_i^2} \mathbb{1}_{\{k=i\}} \mathbb{1}_{\{n=i\}} \right) \\ = \frac{1}{\beta_j} \left(-\frac{1}{\pi_j^2} \frac{\partial \pi_j}{\partial x_k} \frac{\partial \pi_j}{\partial x_n} + \frac{1}{\pi_j} \frac{\partial^2 \pi_j}{\partial x_k \partial x_n} + \frac{1}{x_j^2} \mathbb{1}_{\{j=k\}} \mathbb{1}_{\{j=n\}} \right). \end{aligned}$$

Moreover, using (S.14) we have that

$$\sum_{i=1}^d (\beta_i - \alpha_i) \pi_i(\mathbf{x}) = \lambda (R_0^* - 1),$$

so that

$$\sum_{i=1}^d (\beta_i - \alpha_i) \frac{\partial \pi_i}{\partial x_k} = \sum_{i=1}^d (\beta_i - \alpha_i) \frac{\partial \pi_i}{\partial x_0} = \sum_{i=1}^d (\beta_i - \alpha_i) \frac{\partial \pi_i}{\partial x_{d+1}} = 0$$

and

$$\sum_{i=1}^d (\beta_i - \alpha_i) \frac{\partial^2 \pi_i}{\partial x_k \partial x_n} = 0.$$

Together, these equations give us systems of linear equations that may be solved for the various derivatives of $\pi_i(\mathbf{x})$. To illustrate, consider $\frac{\partial \pi_i}{\partial x_0}$; from the above, we have that

$$0 = \sum_{i=1}^d (\beta_i - \alpha_i) \frac{\partial \pi_i}{\partial x_0} = (\beta_1 - \alpha_1) \frac{\partial \pi_1}{\partial x_0} + \sum_{i=2}^d (\beta_i - \alpha_i) \frac{\beta_i \pi_i}{\beta_1 \pi_1} \frac{\partial \pi_1}{\partial x_0} = \left(\sum_{i=1}^d (\beta_i - \alpha_i) \frac{\beta_i \pi_i}{\beta_1 \pi_1} \right) \frac{\partial \pi_1}{\partial x_0},$$

whence $\frac{\partial \pi_1}{\partial x_0} = 0$, and thus $\frac{\partial \pi_i}{\partial x_0} = 0$ for all i . Proceeding in the same manner, we find $\frac{\partial \pi_i}{\partial x_{d+1}} = 0$ as well, and thus that all second derivatives of $\pi_i(\mathbf{x})$ involving x_0 or x_{d+1} vanish identically, whilst

$$\frac{\partial \pi_i}{\partial x_k} = \frac{\pi_i}{x_i} \mathbb{1}_{\{i=k\}} - \frac{\pi_k}{x_k} \frac{(\beta_k - \alpha_k) \beta_i \pi_i}{\sum_{j=1}^d (\beta_j - \alpha_j) \beta_j \pi_j}.$$

We shall only need to evaluate these for $\mathbf{x} \in \Omega$, where $\hat{\mathbf{E}}(t)$ is trapped. For such \mathbf{x} , the first derivatives simplify to the expression given in the statement, since $\boldsymbol{\pi}(\mathbf{x}) = \mathbf{x}$ for $\mathbf{x} \in \Omega$. Similar calculations lead to the second partial derivatives. \square

5.3.2 Reduced Diffusion

We can use the results of the previous section to provide semi-explicit expressions for the SDE satisfied by $\hat{\mathbf{E}}$ and displayed in Proposition 6.

Proposition 9. *Unlike the full stochastic SIR model, the weak selection limit $\hat{\mathbf{E}}$ can be completely characterised by a system of equations that depend only on the variables $\hat{I}_1, \dots, \hat{I}_d$:*

$$d\hat{I}_i = s_i(\mathbf{I}(t)) \hat{I}_i(t) dt + \frac{1}{\sqrt{R_0^*}} \sum_{k=1}^d \left(\mathbb{1}_{\{i=k\}} - \frac{(\beta_k - \alpha_k) \beta_i \hat{I}_i(t)}{\sum_{j=1}^d (\beta_j - \alpha_j) \beta_j \hat{I}_j(t)} \right) \sqrt{2\beta_k \hat{I}_k(t)} dB_k(t). \quad (\text{S.20})$$

where

$$s_i(\mathbf{x}) = \hat{s}_i(\mathbf{x}) - \frac{\beta_i x_i}{\sum_{j=1}^d (\beta_j - \alpha_j) \beta_j x_j} \sum_{j=1}^d (\beta_j - \alpha_j) \hat{s}_j(\mathbf{x}) \quad (\text{S.21})$$

for

$$\hat{s}_i(\mathbf{x}) := \frac{\beta_i}{R_0^*} \left(r_i - \frac{1}{\sum_{j=1}^d (\beta_j - \alpha_j) \beta_j x_j} \left(2(\beta_i - \alpha_i) - \frac{\sum_{j=1}^d (\beta_j - \alpha_j)^2 \beta_j x_j}{\sum_{j=1}^d (\beta_j - \alpha_j) \beta_j x_j} \right) \right), \quad (\text{S.22})$$

and $dB_1(t), \dots, dB_D(t)$ are independent Brownian motions.

Proof. In the previous section, we observed that $\frac{\partial \pi_i}{\partial x_0} = \frac{\partial \pi_i}{\partial x_{d+1}} = 0$, and thus any second partial derivative with respect to x_0 or x_{d+1} vanishes as well. Moreover, for $i = 1, \dots, d$,

$$\begin{aligned} n \left(F_i^{(n)}(\mathbf{x}) - F_i(\mathbf{x}) \right) &= n \left(\beta_i^{(n)} \left(\frac{x_0}{x_{d+1}} - \frac{1}{R_{0,i}^{(n)}} \right) x_i - \beta_i \left(\frac{x_0}{x_{d+1}} - \frac{1}{R_0^*} \right) x_i \right) \\ &\rightarrow -\frac{\beta_i r_i x_i}{R_0^*}, \end{aligned}$$

Moreover, for $\mathbf{x} \in \Omega$, we have

$$\frac{x_0}{x_{d+1}} = \frac{1}{R_0^*} \left(= \frac{\delta + \alpha_j + \gamma_j}{\beta_j} \right),$$

so that, from (S.6), we obtain in the limit $n \rightarrow \infty$

$$a_{jk}(\mathbf{x}) = \begin{cases} \frac{2\beta_j x_j}{R_0^*} & \text{if } j = k, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, for $1 \leq j, k \leq d$, $\sigma_{jk}(\mathbf{x})$ depends on x_0 or x_{d+1} only via the ratio $\frac{x_0}{x_{d+1}}$ which is identically equal to $\frac{1}{R_0^*}$ on $\mathbf{x} \in \Omega$.

Substituting these and the first derivatives into (S.16) allows us to complete the description of the weak limit $\hat{\mathbf{E}}(t)$, exploiting the fact that the triples of Brownian motions $B_{-e_0+e_i}$, $B_{-e_i-e_{d+1}}$ and B_{-e_j} and $B_{-e_0+e_j}$, $B_{-e_j-e_{d+1}}$ and B_{-e_i} are independent for $i \neq j$ to combine each triple into a single Brownian motion. \square

5.3.3 Frequency Process

Repeating the argument of Section 4, we can use the functions Π_i to finding an equation for the frequency of strain i ,

$$P_i(t) = \frac{\hat{I}_i(t)}{\sum_{j=1}^d \hat{I}_j(t)}$$

where, because the limiting process is a diffusion, the standard Itô formula applies. We omit the lengthy calculations this entails, and present simply the result.

For our process $\mathbf{P}(t)$, we find that

$$\begin{aligned} dP_i(t) &= b_i(\mathbf{P}(t)) dt + \frac{1}{\sqrt{R_0^*}} \frac{1}{\sqrt{I_e(\mathbf{P}(t))}} \frac{1}{\sum_{j=1}^d (\beta_j - \alpha_j) \beta_j P_j(t)} \\ &\times \sum_{j=1}^d (\mathbb{1}_{\{i=j\}} - P_i(t)) \sum_{k=1}^d (\beta_k - \alpha_k) \left(\beta_k P_k(t) \sqrt{2\beta_j P_j(t)} dB_j(t) - \beta_j P_j(t) \sqrt{2\beta_k P_k(t)} dB_k(t) \right). \end{aligned} \tag{S.23}$$

where

$$b_i(\mathbf{p}) := p_i \left(s_i(I_e(\mathbf{p})\mathbf{p}) - \sum_{m=1}^d s_d(I_e(\mathbf{p})\mathbf{p})p_d \right),$$

for $\mathbf{s}(\mathbf{x})$ as defined by (S.21) and (S.22), and where, if $\mathbf{p} \in \Delta_d$ corresponds to the point $\mathbf{x} \in \Omega$, *i.e.*,

$$p_i := \frac{x_i}{\sum_{j=1}^d x_j},$$

then

$$I_e(\mathbf{x}) = \sum_{j=1}^d x_j.$$

Writing

$$p_i := \frac{x_i}{\sum_{j=1}^d x_j} = \frac{x_i}{I_e(\mathbf{x})},$$

and recalling that

$$\lambda(R_0^* - 1) = \sum_{i=1}^d (\beta_i - \alpha_i)x_i = \sum_{i=1}^d (\beta_i - \alpha_i)I_e(\mathbf{x})p_i$$

we see that we can explicitly express I_e as a function of \mathbf{p} :

$$I_e(\mathbf{p}) = \frac{\lambda(R_0^* - 1)}{\sum_{i=1}^d (\beta_i - \alpha_i)p_i}.$$

Remark 3. The notation above has been deliberately chosen to evoke the Wright-Fisher diffusion in population genetics, with $s_i(\mathbf{p})$ and $I_e(\mathbf{p})$ a frequency-dependent selection coefficient and an effective population size, respectively. To understand the motivation for the latter, note that

$$I_e = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^d I_j^{(n)}(t),$$

so that $I_e(\mathbf{x})$ is an idealization of the total density of infected individuals (which is itself a random variable) when the diffusion limit is at the point $\mathbf{x} \in \Omega$, or, equivalently, when the frequencies of the various strains is \mathbf{p} . In principle, additional population structure and the corresponding sampling effects could be taken into account via an “effective infected population size” much as effective population sizes are used in the Wright-Fisher model.

Remark 4. As before, vector $\mathbf{b}(\mathbf{p})$ may be interpreted geometrically as the projection of the vector

$$\begin{bmatrix} s_1(I_e(\mathbf{p})\mathbf{p})p_1 \\ \vdots \\ s_d(I_e(\mathbf{p})\mathbf{p})p_d \end{bmatrix}$$

onto the simplex Δ_d .

5.3.4 Results for $d = 2$

If we have $d = 2$ strains, then, since $P_1(t) + P_2(t) = 1$, it is sufficient to consider the frequency of the invading strain, strain 2. Writing $P(t) := P_2(t)$, the results of the previous section tell us that the generator of $P(t)$ ² is

$$\mathcal{L}f(p) = b(p)f'(p) + \frac{1}{2}a(p)f''(p),$$

where

$$\begin{aligned} b(p) &:= \frac{1}{R_0^*} (r_2 - r_1) \frac{\beta_2 \beta_1 ((\beta_2 - \alpha_2)p + (\beta_1 - \alpha_1)(1-p))}{((\beta_2 - \alpha_2)\beta_2 p + (\beta_1 - \alpha_1)\beta_1(1-p))} p(1-p) \\ &- \frac{1}{R_0^*} \frac{1}{I_e(p)} \frac{\beta_2 \beta_1 ((\beta_2 - \alpha_2)p + (\beta_1 - \alpha_1)(1-p)) (\beta_2 p + \beta_1(1-p))}{((\beta_2 - \alpha_2)\beta_2 p + (\beta_1 - \alpha_1)\beta_1(1-p))^2} \times (((\beta_2 - \alpha_2)\beta_2 - (\beta_1 - \alpha_1)\beta_1) \\ &- ((\beta_2 - \alpha_2) - (\beta_1 - \alpha_1)) \left(\frac{\beta_2 \beta_1 ((\beta_2 - \alpha_2)p + (\beta_1 - \alpha_1)(1-p))}{((\beta_2 - \alpha_2)\beta_2 p + (\beta_1 - \alpha_1)\beta_1(1-p))} - 2(\beta_2 p + \beta_1(1-p)) \right)) p(1-p) \end{aligned}$$

and

$$a(p) := \frac{2}{R_0^*} \frac{1}{I_e(p)} \frac{\beta_2 \beta_1 ((\beta_2 - \alpha_2)p + (\beta_1 - \alpha_1)(1-p))^2 (\beta_2 p + \beta_1(1-p))}{((\beta_2 - \alpha_2)\beta_2 p + (\beta_1 - \alpha_1)\beta_1(1-p))^2} p(1-p).$$

The generator allows us to compute many quantities of interest for the process $P(t)$. In particular, if $h(p)$ is the probability of fixation of strain 1 given $P(0) = p$, then $h(p)$ satisfies the boundary problem

$$\begin{aligned} \mathcal{L}h(p) &= 0 \\ h(0) &= 0 \\ h(1) &= 1 \end{aligned}$$

²The generator of $\mathbf{P}(t)$ is the operator on the space of continuous functions on the d -simplex

$$\Delta_d = \left\{ \mathbf{p} : \sum_{i=1}^d p_i = 1 \right\}$$

defined by

$$\mathcal{L}f(\mathbf{p}) := \lim_{t \downarrow 0} \frac{\mathbb{E}[f(\mathbf{P}(t)) | \mathbf{P}(0) = \mathbf{p}] - f(\mathbf{p})}{t}.$$

We recall that if the diffusion process $\mathbf{P}(t)$ has SDE

$$d\mathbf{P}(t) = \mathbf{b}(\mathbf{P}(t)) dt + \boldsymbol{\varsigma}(\mathbf{P}(t)) d\mathbf{B}(t),$$

then

$$\mathcal{L}f(\mathbf{p}) = \sum_{i=1}^d b_i(\mathbf{p}) \frac{\partial f}{\partial p_i} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d a_{ij}(\mathbf{p}) \frac{\partial^2 f}{\partial p_i \partial p_j}$$

where $\mathbf{a}(\mathbf{x}) = \boldsymbol{\varsigma}(\mathbf{x})\boldsymbol{\varsigma}(\mathbf{x})^\top$ is the variance-covariance matrix for $d\mathbf{P}(t)$, and the probability density function for $\mathbf{P}(t)$, say $f(t, \mathbf{p})$, satisfies the Kolmogorov backward equation

$$\frac{\partial}{\partial t} f(t, \mathbf{p}) = \mathcal{L}f(t, \mathbf{p}).$$

(see *e.g.*, [15, 11, 13]). This may be solved to give

$$h(p) = \frac{\int_0^p e^{-2 \int \frac{b(q)}{a(q)} dq} dq}{\int_0^1 e^{-2 \int \frac{b(q)}{a(q)} dq} dq}. \quad (\text{S.24})$$

Let $\tilde{h}(p)$ be the numerator of this fraction. Substituting the expressions for $a(p)$ and $b(p)$ and some simplification yields

$$\begin{aligned} \tilde{h}(p) := \int_0^p \frac{(\beta_2 q + \beta_1(1-q))^{-\frac{(\beta_2+\beta_1)((\beta_2-\alpha_2)-(\beta_1-\alpha_1))}{\beta_2\alpha_1-\beta_1\alpha_2}} ((\beta_2-\alpha_2)q + (\beta_1-\alpha_1)(1-q))^{2-\frac{((\beta_2-\alpha_2)\beta_2-(\beta_1-\alpha_1)\beta_1)}{\beta_2\alpha_1-\beta_1\alpha_2}}}{((\beta_2-\alpha_2)\beta_2 q + (\beta_1-\alpha_1)\beta_1(1-q))} \\ \times e^{-(r_2-r_1) \int_0^q I_e(u) \frac{((\beta_2-\alpha_2)\beta_2 u + (\beta_1-\alpha_1)\beta_1(1-u))}{((\beta_2-\alpha_2)u + (\beta_1-\alpha_1)(1-u))(\beta_2 u + \beta_1(1-u))} du} dq, \end{aligned}$$

whereas

$$h(p) = \frac{\tilde{h}(p)}{\tilde{h}(1)}.$$

For ease of notation, we will write

$$\phi(q) = \int_0^q I_e(u) \frac{((\beta_2-\alpha_2)\beta_2 u + (\beta_1-\alpha_1)\beta_1(1-u))}{((\beta_2-\alpha_2)u + (\beta_1-\alpha_1)(1-u))(\beta_2 u + \beta_1(1-u))} du$$

and

$$g(q) = \frac{(\beta_2 q + \beta_1(1-q))^{-\frac{(\beta_2+\beta_1)((\beta_2-\alpha_2)-(\beta_1-\alpha_1))}{\beta_2\alpha_1-\beta_1\alpha_2}} ((\beta_2-\alpha_2)q + (\beta_1-\alpha_1)(1-q))^{2-\frac{((\beta_2-\alpha_2)\beta_2-(\beta_1-\alpha_1)\beta_1)}{\beta_2\alpha_1-\beta_1\alpha_2}}}{((\beta_2-\alpha_2)\beta_2 q + (\beta_1-\alpha_1)\beta_1(1-q))},$$

so that

$$\tilde{h}(p) = \int_0^p g(q) e^{-(r_2-r_1)\phi(q)} dq.$$

We can evaluate this expression numerically, but we will be particularly interested in a number of special cases, when we can obtain analytical approximations to $\tilde{h}(p)$.

- (i) When $r_2 = r_1$ (or, more generally, when $R_{0,i} = R_0^* (1 + o(\frac{1}{n}))$) we can give an explicit closed form for $\tilde{h}(p)$, and thus $h(p)$:

$$\begin{aligned} \tilde{h}(p) \propto \frac{(\beta_2 p + \beta_1(1-p))^{1-\frac{(\beta_2+\beta_1)((\beta_2-\alpha_2)-(\beta_1-\alpha_1))}{\beta_2\alpha_1-\beta_1\alpha_2}} ((\beta_2-\alpha_2)p + (\beta_1-\alpha_1)(1-p))^{3-\frac{((\beta_2-\alpha_2)\beta_2-(\beta_1-\alpha_1)\beta_1)}{\beta_2\alpha_1-\beta_1\alpha_2}}}{((\beta_2-\alpha_2)p + (\beta_1-\alpha_1)(1-p))} \\ - \frac{(\beta_2 + \beta_1)^{1-\frac{(\beta_2+\beta_1)((\beta_2-\alpha_2)-(\beta_1-\alpha_1))}{\beta_2\alpha_1-\beta_1\alpha_2}} ((\beta_2-\alpha_2) + (\beta_1-\alpha_1))^{3-\frac{((\beta_2-\alpha_2)\beta_2-(\beta_1-\alpha_1)\beta_1)}{\beta_2\alpha_1-\beta_1\alpha_2}}}{((\beta_2-\alpha_2) + (\beta_1-\alpha_1))}. \quad (\text{S.25}) \end{aligned}$$

This expression is not, however, especially illuminating.

- (ii) We shall principally be interested in the case when p is small, in which case we can Taylor

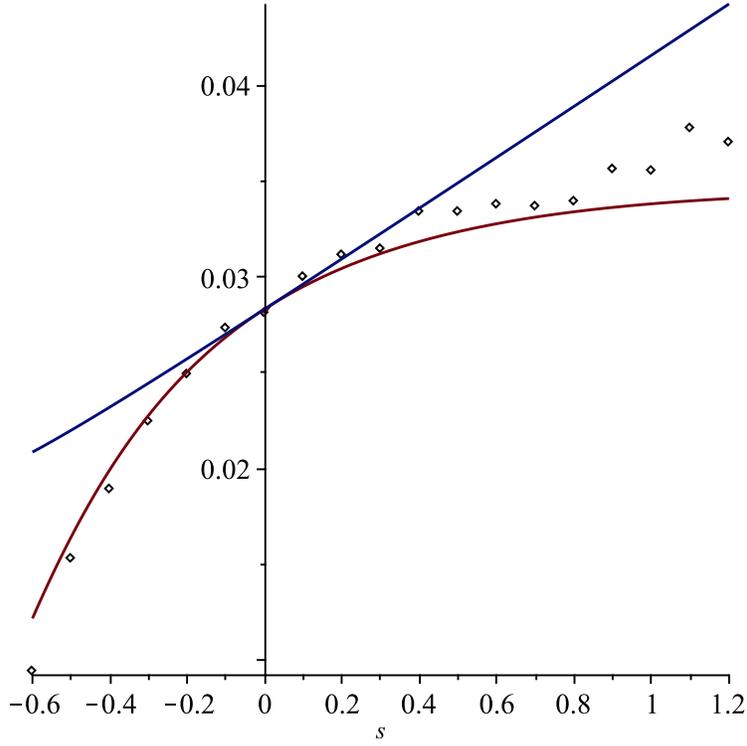


Figure S.3: We compare (S.25) (red curve) and its order p^2 approximation (S.26) (blue line) with fixation probabilities for a single mutant invader obtained via simulating the Markov chain with rates given by Table 1 (black diamonds). We vary β_2 by setting $\beta_1 = \beta_2(1 + \sigma)$, while holding $R_{0,1} = R_{0,2}$ by setting $\alpha_2 = \frac{\delta + \alpha_1 + \gamma}{1 + \sigma} - \delta - \gamma$. The other parameters are fixed at $n = 100$, $R_0^* = 4$, $\lambda = 2$, $\delta = 1$, $\beta_1 = 20$, and $\alpha_1 = 3$.

expand $\tilde{h}(p)$ as

$$\begin{aligned}
 \tilde{h}(p) &= \tilde{h}(0) + \tilde{h}'(0)p + \frac{1}{2}\tilde{h}''(0)p^2 + \mathcal{O}(p^3) \\
 &= g(0)p + \frac{1}{2}(g'(0) + g(0)(r_2 - r_1)\phi'(0))p^2 + \mathcal{O}(p^3) \\
 &= g(0) \left(p + \frac{1}{2} \left(\frac{g'(0)}{g(0)} + (r_2 - r_1)I_e(0) \right) p^2 \right) + \mathcal{O}(p^3)
 \end{aligned}$$

Unfortunately, this does not yield an estimate of the normalising constant, $\tilde{h}(1)$. To obtain this, we consider the case when $\beta_2 - \beta_1$ and $\alpha_2 - \alpha_1$ are small. While this is a restrictive assumption, it will allow us to consider the long-term evolution in the framework of adaptive dynamics, where mutational changes are assumed to be very small. To this end, we introduce σ and θ such that

$$\beta_2 = \beta_1(1 + \sigma) \quad \text{and} \quad \alpha_2 = \alpha_1(1 + \theta).$$

We then have that

$$\frac{g'(0)}{g(0)} = \sigma + \mathcal{O}(2), \frac{g(p)}{g(0)} = 1 - \sigma p + \mathcal{O}(2), \phi'(p) = -I_e(p) + \mathcal{O}(2),$$

and

$$\begin{aligned} \frac{\tilde{h}(1)}{g(0)} &= \int_0^1 (1 - \sigma p) e^{-(r_2 - r_1) \int_0^p I_e(q) dq} dp + \mathcal{O}(2) \\ &= 1 - \frac{\sigma}{2} - (r_2 - r_1) I_e(0) + \mathcal{O}(2), \end{aligned}$$

where $\mathcal{O}(2)$ is used to denote terms of order $\mathcal{O}(\sigma^2)$, $\mathcal{O}(\theta^2)$, or $\mathcal{O}(\sigma\theta)$.

Then, to order p^2 , the fixation probability is

$$h(p) := \frac{\tilde{h}(p)}{\tilde{h}(1)} = p + \frac{1}{2}(\sigma + (r_2 - r_1)I_e(0))p(1 - p) + \mathcal{O}(2), \quad (\text{S.26})$$

which may be written informally in terms of the original parameters as

$$h(p) := \frac{\tilde{h}(p)}{\tilde{h}(1)} = p + \frac{1}{2} \left(\frac{\beta_1 - \beta_2}{\beta_2} + \left(1 - \frac{R_{0,1}^{(n)}}{R_{0,2}^{(n)}} \right) n I_e(0) \right) p(1 - p) + \mathcal{O}(2)$$

to lowest order. Here we have used

$$r_2 - r_1 = n \frac{R_{0,2}^{(n)} - R_{0,1}^{(n)}}{R_0^*} + o(1) = n \frac{R_{0,2}^{(n)} - R_{0,1}^{(n)}}{R_{0,2}^{(n)}} \left(1 + \frac{r_2}{n} + o(1) \right) + o(1).$$

In practice, we are most interested in the case when a single individual carries the mutant strain, so $p = \frac{1}{nI_e(0)} = \frac{1}{I^{(n)}(0)}$. While our proofs – which assume that p is independent of n , and thus that the number of invading individuals is proportional to n – do not justify taking this value for p , we find that the expression for the fixation probability obtained by taking $p = \frac{1}{I^{(n)}(0)}$, which to lowest order is

$$\frac{1}{I^{(n)}(0)} + \frac{1}{2} \left(\frac{1}{I^{(n)}(0)} \frac{\beta_1 - \beta_2}{\beta_2} + 1 - \frac{R_{0,1}^{(n)}}{R_{0,2}^{(n)}} \right), \quad (\text{S.27})$$

agrees extremely well with simulations (Figure S.4c) – an example of the so-called “unreasonable effectiveness of mathematics” [33] – and will use it to investigate the long term evolution of the virulence in Section 6.

5.4 Reconciling the Strong and Weak Selection Results

On first inspection, our expressions for the strong and weak selection limits have little in common. In Section 5.2, we saw that if $R_{0,i}^{(n)} \rightarrow R_{0,i}$ and $R_{0,2} \neq R_{0,1}$, then, if $I_2^{(n)}(0) \rightarrow I_2(0)$, the fixation

probability was

$$\begin{cases} 1 - \left(\frac{R_{0,1}}{R_{0,2}}\right)^{I_2(0)} & \text{if } R_{0,2} > R_{0,1}, \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

On the other hand, in Section 5.3, we assume that

$$R_{0,i}^{(n)} = R_0^* \left(1 + \frac{r_i}{n}\right) + o\left(\frac{1}{n}\right)$$

and $\bar{I}^{(n)}(0) = \frac{1}{n}I_2^{(n)}(0) \rightarrow I_2(0)$, and derive a quite different appearing expression for the fixation probability.

More generally, we might consider the intermediary scalings: let $1 \ll \kappa_n \ll n$, and suppose that

$$R_{0,i}^{(n)} = R_0^* \left(1 + \frac{r_i}{\kappa_n}\right) + o\left(\frac{1}{\kappa_n}\right)$$

whilst

$$\lim_{n \rightarrow \infty} \frac{I_2^{(n)}(0)}{\kappa_n} = \iota.$$

Substituting these into our expression for strong selection, we see that, provided $r_2 > r_1$, we have that the probability of fixation is

$$1 - \left(\frac{\left(1 + \frac{r_1}{\kappa_n}\right) + o\left(\frac{1}{\kappa_n}\right)}{\left(1 + \frac{r_2}{\kappa_n}\right) + o\left(\frac{1}{\kappa_n}\right)} \right)^{\kappa_n \frac{I_2^{(n)}(0)}{\kappa_n}} \rightarrow 1 - e^{-(r_2 - r_1)\iota}$$

as $n \rightarrow \infty$ (*n.b.*, that when $\kappa_n \gg n$, $\frac{I_2^{(n)}(0)}{\kappa_n} \rightarrow 0$, so trivially the probability of fixation is 0).

On the other hand, we can begin with our expression for the fixation probability under weak selection, which written in terms of the original parameters $R_{0,2}^{(n)}$ and $R_{0,1}^{(n)}$, was proportional to

$$\begin{aligned} \tilde{h}(p) = \int_0^p \frac{(\beta_2 q - \beta_1(1-q))^{-\frac{(\beta_2 + \beta_1)((\beta_2 - \alpha_2) - (\beta_1 - \alpha_1))}{\beta_2 \alpha_1 - \beta_1 \alpha_2}} ((\beta_2 - \alpha_2)q + (\beta_1 - \alpha_1)(1-q))^{2 - \frac{((\beta_2 - \alpha_2)\beta_2 - (\beta_1 - \alpha_1)\beta_1)}{\beta_2 \alpha_1 - \beta_1 \alpha_2}}}{((\beta_2 - \alpha_2)\beta_2 q + (\beta_1 - \alpha_1)\beta_1(1-q))} \\ \times e^{-n \left(\frac{R_{0,2}^{(n)} - R_{0,1}^{(n)}}{R_0^*} + o\left(\frac{1}{n}\right) \right) \frac{((\beta_2 - \alpha_2)\beta_2 q + (\beta_1 - \alpha_1)\beta_1(1-q))}{\beta_2 \beta_1 ((\beta_2 - \alpha_2)q + (\beta_1 - \alpha_1)(1-q)) (\beta_2 q + \beta_1(1-q))}} dq \end{aligned}$$

Replacing $R_{0,2}^{(n)}$ and $R_{0,1}^{(n)}$ with the intermediary scalings, our expression becomes

$$\begin{aligned} \int_0^p \frac{(\beta_2 q - \beta_1(1-q))^{-\frac{(\beta_2 + \beta_1)((\beta_2 - \alpha_2) - (\beta_1 - \alpha_1))}{\beta_2 \alpha_1 - \beta_1 \alpha_2}} ((\beta_2 - \alpha_2)q + (\beta_1 - \alpha_1)(1-q))^{2 - \frac{((\beta_2 - \alpha_2)\beta_2 - (\beta_1 - \alpha_1)\beta_1)}{\beta_2 \alpha_1 - \beta_1 \alpha_2}}}{((\beta_2 - \alpha_2)\beta_2 q + (\beta_1 - \alpha_1)\beta_1(1-q))} \\ \times e^{-\frac{n}{\kappa_n} (r_2 - r_1 + o(1)) \int I_e(q) \frac{((\beta_2 - \alpha_2)\beta_2 q + (\beta_1 - \alpha_1)\beta_1(1-q))}{\beta_2 \beta_1 ((\beta_2 - \alpha_2)q + (\beta_1 - \alpha_1)(1-q)) (\beta_2 q + \beta_1(1-q))}} dq \end{aligned}$$

We can find a large n asymptotic expression for this probability using a pair of lemmas:

Lemma 1. *Suppose that $\phi(x)$ and $g(x)$ are an increasing continuously differentiable function and a continuous function on $[a, b]$ ($-\infty < a < b < \infty$) respectively, and that $\phi'(a) \neq 0$. Then,*

$$\lim_{M \rightarrow \infty} \frac{\int_a^b g(x) e^{-M\phi(x)} dx}{\frac{g(a) e^{-M\phi(a)}}{M\phi'(a)}} = 1.$$

Proof. Fix $\varepsilon > 0$ such that $\phi'(a) > \varepsilon$. Using Taylor's theorem, we may write

$$\phi(x) = \phi(a) + \phi'(a)(x - a) + R(x)(x - a),$$

where $R(x) \rightarrow 0$ as $x \rightarrow a$. Fix $\delta > 0$ such that

$$|R(x)| < \varepsilon \quad \text{and} \quad |g(x) - g(a)| < \varepsilon$$

for all x such that $x - a < \delta$. Finally, choose $\eta > 0$ such that $\phi(x) > \phi(a) + \eta$ for all x such that $x - a \geq \delta$ and B such that $|g(x)| < B$ for all $x \in [a, b]$. Then,

$$\begin{aligned} \int_a^b g(x) e^{-M\phi(x)} dx &= e^{-M\phi(a)} \int_a^b g(x) e^{-M(\phi(x) - \phi(a))} dx \\ &= e^{-M\phi(a)} \left(\int_a^{a+\delta} g(x) e^{-M(\phi(x) - \phi(a))} dx + \int_{a+\delta}^b g(x) e^{-M(\phi(x) - \phi(a))} dx \right). \end{aligned}$$

Now,

$$\left| \int_{a+\delta}^b g(x) e^{-M(\phi(x) - \phi(a))} dx \right| \leq \int_{a+\delta}^b B e^{-M\eta} dx \rightarrow 0$$

as $M \rightarrow \infty$, whereas

$$\begin{aligned} (g(a) - \varepsilon) \int_a^{a+\delta} e^{-M(\phi'(a) + \varepsilon)(x-a)} dx &\leq \int_a^{a+\delta} g(x) e^{-M(\phi(x) - \phi(a))} dx \\ &\leq (g(a) + \varepsilon) \int_a^{a+\delta} e^{-M(\phi'(a) - \varepsilon)(x-a)} dx. \end{aligned}$$

Now, letting $y = M(x - a)$, we have

$$\int_a^{a+\delta} e^{-M(\phi'(a) - \varepsilon)(x-a)} dx = \frac{1}{M} \int_0^{M\delta} e^{-(\phi'(a) - \varepsilon)y} dy$$

whilst

$$\int_0^{M\delta} e^{-(\phi'(a) - \varepsilon)y} dy = \frac{1}{\phi'(a) - \varepsilon} \left(1 - e^{-(\phi'(a) - \varepsilon)M\delta} \right) \rightarrow \frac{1}{\phi'(a) - \varepsilon}$$

as $M \rightarrow \infty$, and similarly for the lower bound.

Since $\varepsilon > 0$ can be chosen arbitrarily small, the result follows. \square

Lemma 2. Let $\phi(x)$, $g(x)$, and $[a, b]$ be as above. Then,

$$\lim_{M \rightarrow \infty} \frac{\int_a^{a+\frac{X}{M}} g(x) e^{-M\phi(x)} dx}{\frac{g(a)e^{-M\phi(a)}}{M\phi'(a)}} = 1 - e^{-\phi'(a)X}.$$

Proof. By direct computation, we have

$$\begin{aligned} \int_a^{a+\frac{X}{M}} g(x) e^{-M\phi(x)} dx &= \int_0^X g\left(a + \frac{y}{M}\right) e^{-M\phi\left(a + \frac{y}{M}\right)} dy \\ &= \frac{1}{M} \int_0^X g\left(a + \frac{y}{M}\right) e^{-M\left(\phi(a) + \phi'(a)\frac{y}{M} + R\left(a + \frac{y}{M}\right)\frac{y}{M}\right)} dy \end{aligned}$$

so that as $M \rightarrow \infty$,

$$\frac{\int_a^{a+\frac{X}{M}} g(x) e^{-M\phi(x)} dx}{\frac{g(a)e^{-M\phi(a)}}{M\phi'(a)}} \rightarrow \phi'(a) \int_0^X e^{-\phi'(a)y} dy.$$

The result follows. □

To apply the lemmas here, we take $a = 0$ and $b = 1$, $M_n = \frac{n}{\kappa_n}$, and, as before,

$$g(p) = \frac{(\beta_2 p - \beta_1(1-p))^{-\frac{(\beta_2 + \beta_1)((\beta_2 - \alpha_2) - (\beta_1 - \alpha_1))}{\beta_2 \alpha_1 - \beta_1 \alpha_2}} ((\beta_2 - \alpha_2)p + (\beta_1 - \alpha_1)(1-p))^{-\frac{((\beta_2 - \alpha_2)\beta_2 - (\beta_1 - \alpha_1)\beta_1)}{\beta_2 \alpha_1 - \beta_1 \alpha_2}}}{((\beta_2 - \alpha_2)\beta_2 p + (\beta_1 - \alpha_1)\beta_1(1-p))}$$

whereas we now take a slightly different definition for $\phi(p)$, which now has an $o(1)$ correction:

$$\phi(p) = (r_2 - r_1 + o(1)) \int I_e(p) \frac{((\beta_2 - \alpha_2)\beta_2 p + (\beta_1 - \alpha_1)\beta_1(1-p))}{\beta_2 \beta_1 ((\beta_2 - \alpha_2)p + (\beta_1 - \alpha_1)(1-p)) (\beta_2 p + \beta_1(1-p))} dp$$

so that

$$\phi'(0) \rightarrow I_e(0)(r_2 - r_1)$$

as $n \rightarrow \infty$. Then, using Lemma 1 we conclude that $\tilde{h}(1)$ is asymptotically equivalent to

$$\frac{g(0)e^{-M_n\phi(0)}}{M_n\phi'(0)}.$$

Now, to consider the numerator when we start with $I_2^{(n)}(0) \sim \iota\kappa_n$ individuals of the invading strain, we recall that

$$p = \lim_{n \rightarrow \infty} P_2^{(n)}(0) = \lim_{n \rightarrow \infty} \frac{I_2^{(n)}(0)}{I_2^{(n)}(0) + I_1^{(n)}(0)} = \lim_{n \rightarrow \infty} \frac{I_2^{(n)}(0)}{I_e(P_2^{(n)}(0))n}$$

so, to apply Lemma 2, we will take

$$X := X^{(n)} = M_n \frac{I_2^{(n)}(0)}{I_e(P_2^{(n)}(0))n} = \frac{I_2^{(n)}(0)}{I_e(P^{(n)}(0))\kappa_n}$$

i.e., so that $\frac{X^{(n)}}{M_n} \sim p$.

We note that $P^{(n)}(0) \propto \frac{\kappa_n \iota}{n} \rightarrow 0$ as $n \rightarrow \infty$, so

$$\lim_{n \rightarrow \infty} X^{(n)} = \frac{\iota}{I_e(0)}.$$

Thus, applying Lemma 2, we have

$$\tilde{h} \left(\frac{I_2^{(n)}(0)}{I_e(P_2^{(n)}(0))n} \right) \sim \frac{g(0)e^{-M_n \phi(0)}}{M_n \phi'(0)} \left(1 - e^{-(r_2 - r_1)\iota} \right),$$

and the probability of fixation obtained from the weak selection expression is again asymptotic to

$$1 - e^{-(r_2 - r_1)\iota}.$$

While this is not a rigorous proof, it does demonstrate heuristically that the weak and strong selection expressions for the fixation probability agree to first order when applied across the intermediate selective regimes. In particular, we can use the method of matched asymptotic expansions (see *e.g.*, [17, 20]) to combine our two solutions into a single expression valid across all scales, by summing the expressions for strong and weak selection and subtracting their common limit, where all are expressed in the unscaled (*i.e.*, strong selection parameters):

$$\left[1 - \left(\frac{R_{0,1}^{(n)}}{R_{0,2}^{(n)}} \right)^{I_2^{(n)}(0)} + h \left(\frac{I_2^{(n)}(0)}{I_e(0)n} \right) - \left(1 - e^{\left(\frac{R_{0,1}^{(n)}}{R_{0,2}^{(n)}} - 1 \right) I_2^{(n)}(0)} \right) \right]^+, \quad (\text{S.28})$$

where $[x]^+ = \max\{x, 0\}$ and we have used that

$$\frac{R_{0,1}^{(n)}}{R_{0,2}^{(n)}} - 1 = \frac{1 + \frac{r_1}{\kappa_n}}{1 + \frac{r_2}{\kappa_n}} - 1 = \frac{1}{\kappa_n} \frac{r_1 - r_2}{1 + \frac{r_2}{\kappa_n}} = \frac{1}{\kappa_n} (r_1 - r_2) + \mathcal{O}\left(\frac{1}{\kappa_n^2}\right),$$

so that

$$(r_1 - r_2)\iota \sim (r_1 - r_2) \frac{I_2^{(n)}(0)}{\kappa_n} \sim \left(\frac{R_{0,1}^{(n)}}{R_{0,2}^{(n)}} - 1 \right) I_2^{(n)}(0).$$

We illustrate how these approximations compare to a simulated epidemic in Figure S.4.

6 Adaptive Dynamics

Using the expressions for the fixation probability derived above, we can use the framework of adaptive dynamics to investigate the long-term evolution of strains. In what follows, we give an informal discussion of the derivation of the canonical *diffusion* for the process, a generalisation of the canonical equation of adaptive dynamics which allows us to consider the influence of random drift on phenotypic evolution. We refer the reader to [26] for a more extensive discussion

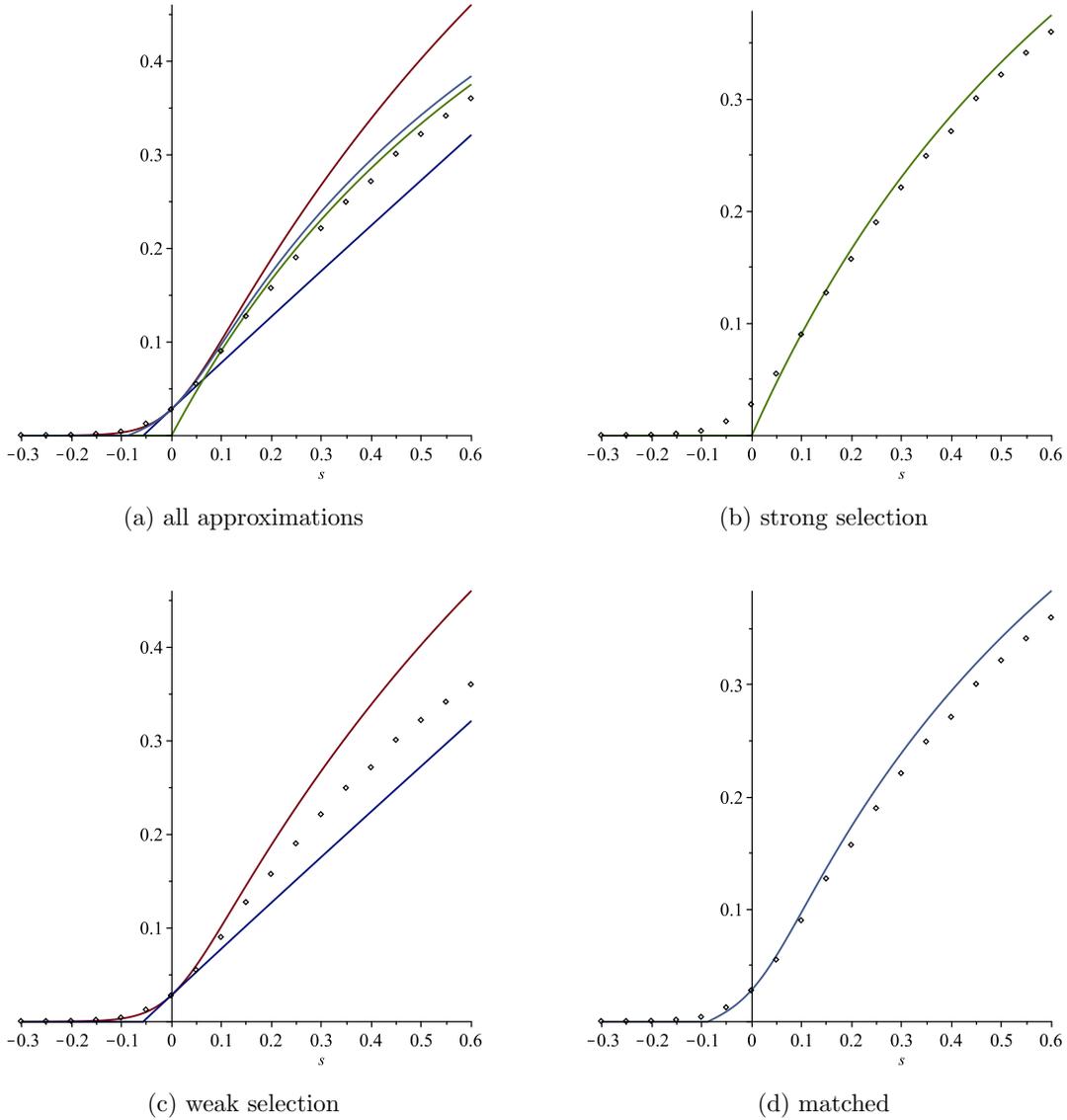


Figure S.4: We compare our various approximations to fixation probabilities obtained by Markov chain simulations (black diamonds), where we assume a single mutant invading a resident population at equilibrium. We have implemented selection on the invading strain by setting $\beta_2 = \beta_1(1 + s)$, so that $R_{0,2}^{(n)} = R_{0,1}^{(n)}(1 + s)$. The other parameters are fixed at $n = 100$, $R_{0,1}^{(n)} = R_0^* = 4$, $\lambda = 2$, $\delta = 1$, $\beta_1 = 20$, and $\alpha_1 = 3$. (b) shows strong selection approximation (S.15) (green curve), (c) shows the weak selection approximation (S.24) (red curve) and its second order approximation (S.24) (blue line), (d) shows the matched asymptotic approximation (S.28), whereas (a) overlays all the approximations for comparison. As would be expected, the strong and weak approximations do well in their corresponding parameter regimes, but poorly elsewhere, whereas the matched asymptotic provides a compromise, performing worse than the weak or strong approximations at the respective extremes, but interpolating between them for intermediate values of s .

aimed at a biological audience and to [9] for a mathematically rigorous derivation of the canonical equation.

We briefly recall the assumptions of adaptive dynamics in the context of our epidemic models; throughout, we assume that a novel mutant strain with virulence α' invades a population which is at the endemic equilibrium with a resident strain α .

We assume a tradeoff between transmissibility and virulence, so that the contact rate of a strain depends on its virulence according to some fixed function $\beta(\alpha)$. For our numerical investigations, we take

$$\beta(\alpha) = (\delta + \alpha + \gamma)(R_{0,\max} - w(\alpha - \alpha_0)^2),$$

where w is a parameter that determines the “flatness” of the fitness landscape.

The reproductive number is a function of the virulence,

$$R_0(\alpha) = \frac{\beta(\alpha)}{\delta + \alpha + \gamma}.$$

We will assume that there is a value, α_0 , for the virulence that maximises $R_0(\alpha)$.

Under these assumptions, the density of individuals infected with the resident strain at the endemic equilibrium is

$$I_{\text{eq}}(\alpha) \sim \frac{\lambda(R_0(\alpha) - 1)}{\beta(\alpha) - \alpha}$$

$I_{\text{eq}}(\alpha)$ is non-zero on a range $(\alpha_{\min}, \alpha_{\max})$; outside of this range, $R_0(\alpha) \leq 1$, and the pathogen goes extinct (when R_0 is independent of the virulence, $\alpha_{\min} = 0$; in this case, as we will see below, we must fix $\alpha_{\max} < \infty$ to have a probability distribution. The choice of α_{\max} is arbitrary).

We then have, using (S.27), that the fixation probability of a mutant strain of virulence α' arising in a single individual in a population in which a strain of virulence α' , is

$$S(\alpha, \alpha') \sim \frac{1}{nI_{\text{eq}}(\alpha)} + \frac{1}{2} \left(\frac{1}{nI_{\text{eq}}(\alpha)} \frac{\beta(\alpha) - \beta(\alpha')}{\beta(\alpha')} + 1 - \frac{R_0(\alpha)}{R_0(\alpha')} \right) + \mathcal{O}(|\alpha - \alpha'|^2).$$

To introduce the evolutionary dynamics, we assume that mutations occur in individuals with virulence α at a per-capita rate $\epsilon\eta(\alpha)$, where $\epsilon > 0$ is a dimensionless parameter that we will take to 0. This will ensure that, with high probability, fixation occurs before a second novel mutation can arise. The population is thus assumed to be monomorphic (*i.e.*, all individuals have the same virulence) between invasion events. We will assume that given a mutation occurs in an individual α , the offspring has virulence α' with probability $K(\alpha, \alpha' - \alpha)$ where K has mean 0 and finite variance:

$$\int_{-\infty}^{\infty} K(\alpha, z) z^k dz = \begin{cases} 0 & \text{if } k = 1, \\ \nu(\alpha) & \text{if } k = 2, \text{ and} \\ \mathcal{O}(1) & \text{otherwise.} \end{cases}$$

We now pass from the individual based model to the trait substitution sequence [25, 10]: we have seen that whenever a new strain arises, either the mutant or resident strain will rapidly go extinct.

Until a new mutant arises, the population will be composed entirely of individuals of the surviving strain. Let $A_\epsilon(t)$ be a random variable giving the virulence of the strain that survived the last competition event prior to time t . The population is thus entirely composed of the strain $A_\epsilon(t)$ except for times t in the short intervals when two strains are competing. If we pass to a “mutational time scale”, $\frac{t}{\epsilon}$, as $\epsilon \rightarrow 0$ the duration of these intervals shrinks to 0, and we are left with a process in which novel mutations either fix or disappear instantly, so that the population is only observed with a single strain at equilibrium.

Formally, as $\epsilon \rightarrow 0$, $A_\epsilon(\frac{t}{\epsilon}) \Rightarrow A(t)$, a continuous time Markov chain that jumps from virulence α to α' when a strain of virulence α' successfully invades a population with resident virulence α [9]. The process $A(t)$ has generator (recall, the generator is the operator \mathcal{L} defined by $\mathcal{L}f(\alpha) := \frac{d}{dt} \Big|_{t=0} \mathbb{E}[f(A(t)) | A(0) = \alpha]$):

$$\mathcal{L}f(\alpha) = n\eta(\alpha)I_{eq}(\alpha) \int_{\alpha_{\min}}^{\alpha_{\max}} K(\alpha, \alpha')s(\alpha, \alpha') (f(\alpha') - f(\alpha)) d\alpha'$$

n.b., If $\alpha' \notin [\alpha_{\min}, \alpha_{\max}]$, then the mutant cannot fix.

We now consider the limit of small mutations, and, following [8] assume that mutations will be of order $\mathcal{O}(\epsilon)$, where ϵ is a dimensionless parameter which we will take to 0; this limit of small mutational effects allows us to ignore terms of order $\mathcal{O}(|\alpha - \alpha'|^2)$ in the fixation probability. In particular, we introduce a rescaled kernel

$$K_\epsilon(\alpha, z) = \frac{1}{\epsilon} K\left(\alpha, \frac{z}{\epsilon}\right),$$

so that $K_\epsilon(\alpha, z) dz$ is the probability the mutant has virulence $\alpha + \epsilon z$. Let $A^\epsilon(t)$ be the process defined as $A(t)$ above, but with the kernel K replaced by the kernel K_ϵ .

Now, consider the time rescaled process $\hat{A}_\epsilon(t) := A^\epsilon\left(\frac{t}{\epsilon^2}\right)$; this has generator

$$\hat{\mathcal{L}}_\epsilon f(\alpha) = \frac{n}{\epsilon^2} \eta(\alpha) I_{eq}(\alpha) \int_{\alpha_{\min}}^{\alpha_{\max}} K_\epsilon(\alpha, \alpha' - \alpha) s(\alpha, \alpha') (f(\alpha') - f(\alpha)) d\alpha'.$$

In Section 8.3, we prove

Proposition 10. *As $\epsilon \rightarrow 0$, $\hat{A}_\epsilon(t)$ converges to a limiting diffusion $\hat{A}(t)$ with advective coefficient*

$$\mu(\alpha) = n\eta(\alpha)I_{eq}(\alpha)\nu(\alpha)\frac{\partial}{\partial\alpha'}s(\alpha, \alpha) = \frac{\eta(\alpha)\nu(\alpha)}{2} \left(nI_{eq}(\alpha)\frac{R'_0(\alpha)}{R_0(\alpha)} - \frac{\beta'(\alpha)}{\beta(\alpha)} \right) \quad (\text{S.29})$$

and diffusion coefficient

$$\sigma^2(\alpha) = n\eta(\alpha)I_{eq}(\alpha)\nu(\alpha)s(\alpha, \alpha) = \eta(\alpha)\nu(\alpha). \quad (\text{S.30})$$

and generator

$$\hat{\mathcal{L}}f(\alpha) = \mu(\alpha)f'(\alpha) + \frac{1}{2}\sigma^2(\alpha)f''(\alpha).$$

for $f \in C^2[\alpha_{\min}, \alpha_{\max}]$ such that $f'(\alpha_{\min}) = f'(\alpha_{\max}) = 0$.

The process $\hat{A}(t)$ is our canonical diffusion.

Remark 5. The condition that $f'(\alpha_{\min}) = f'(\alpha_{\max}) = 0$ corresponds to a zero-flux (reflecting) boundary condition at α_{\min} and α_{\max} for the corresponding forward equation [16]:

$$\begin{aligned} \int_{\alpha_{\min}}^{\alpha_{\max}} g(\alpha) \hat{\mathcal{L}} f(\alpha) d\alpha &= \int_{\alpha_{\min}}^{\alpha_{\max}} g(\alpha) \mu(\alpha) f'(\alpha) d\alpha + \frac{1}{2} \int_{\alpha_{\min}}^{\alpha_{\max}} g(\alpha) \sigma^2(\alpha) f''(\alpha) d\alpha \\ &= \left(g(\alpha_{\min}) \mu(\alpha_{\min}) - (g(\alpha_{\min}) \sigma^2(\alpha_{\min}))' \right) f(\alpha_{\min}) - \left(g(\alpha_{\max}) \mu(\alpha_{\max}) - (g(\alpha_{\max}) \sigma^2(\alpha_{\max}))' \right) f(\alpha_{\max}) \\ &\quad - g(\alpha_{\min}) \sigma^2(\alpha_{\min}) f(\alpha_{\min}) f'(\alpha_{\min}) + g(\alpha_{\max}) \sigma^2(\alpha_{\max}) f(\alpha_{\max}) f'(\alpha_{\max}) \\ &\quad - \int_{\alpha_{\min}}^{\alpha_{\max}} (g(\alpha) \mu(\alpha))' f(\alpha) d\alpha + \frac{1}{2} \int_{\alpha_{\min}}^{\alpha_{\max}} (g(\alpha) \sigma^2(\alpha))'' f(\alpha) d\alpha, \end{aligned}$$

which equals $\int_{\alpha_{\min}}^{\alpha_{\max}} \hat{\mathcal{L}}^* g(\alpha) f(\alpha) d\alpha$ for arbitrary f such that $f'(\alpha_{\min}) = f'(\alpha_{\max}) = 0$ if and only if the flux $g(\alpha) \mu(\alpha) - (g(\alpha) \sigma^2(\alpha))'$ vanishes at α_{\min} and α_{\max} .

6.1 Stationary Distribution

Using (S.29) and (S.30), we may compute the stationary distribution ψ of $\hat{A}(t)$; this stationary distribution describes the long-term behaviour of the virulence, after any “memory” of the initial state has been lost. Given any subset $\mathcal{A} \subset (\alpha_{\min}, \alpha_{\max})$, $\psi(\mathcal{A})$ gives the proportion of time that the virulence is in the set \mathcal{A} , or equivalently, the probability that at some random sampling time t , the virulence takes a value in \mathcal{A} . ψ is characterised by the relation

$$\int_{\alpha_{\min}}^{\alpha_{\max}} \hat{\mathcal{L}} f(\alpha) \psi(d\alpha) = 0.$$

In particular, if $\psi(d\alpha)$ has a density, which, in a slight abuse of notation, we write as $\psi(\alpha)$,

$$\hat{\mathcal{L}}^* \psi = 0,$$

where $\hat{\mathcal{L}}^*$, defined by

$$\hat{\mathcal{L}}^* f(\alpha) = -\frac{d}{d\alpha} [\mu(\alpha) f(\alpha)] + \frac{1}{2} \frac{d^2}{d\alpha^2} [\sigma^2(\alpha) f(\alpha)],$$

is the adjoint operator to $\hat{\mathcal{L}}$ (and thus,

$$\frac{d}{dt} f(\alpha, t) = \hat{\mathcal{L}}^* f(\alpha, t)$$

is the Fokker-Planck equation for the probability density of $\hat{A}(t)$, $f(\alpha, t)$).

Thus,

$$\psi(\alpha) = \frac{1}{Z} \frac{1}{\sigma^2(\alpha)} e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha} d\alpha \tag{S.31}$$

where $Z = \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{1}{\sigma^2(\alpha)} e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha} d\alpha$ is a normalising constant.

From the previous, we have

$$\frac{2\mu(\alpha)}{\sigma^2(\alpha)} = nI_{\text{eq}}(\alpha) \frac{R'_0(\alpha)}{R_0(\alpha)} - \frac{\beta'(\alpha)}{\beta(\alpha)}. \quad (\text{S.32})$$

Unfortunately, we can only compute its integral analytically in the case of a “flat landscape”, when $R_0(\alpha) \equiv R_0$, independently of α . In this case, we have $\beta(\alpha) = R_0(\delta + \alpha + \gamma)$, $\beta'(\alpha) = R_0$, and

$$\frac{2\mu(\alpha)}{\sigma^2(\alpha)} = -\frac{\beta'(\alpha)}{\beta(\alpha)} = -\frac{1}{\delta + \alpha + \gamma}$$

which has integral $-\ln(\delta + \alpha + \gamma)$, so that

$$\frac{1}{\sigma^2(\alpha)} e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha} d\alpha = \frac{1}{\sigma^2(\alpha)} \frac{1}{\delta + \alpha + \gamma}.$$

In particular, in the case when $\eta(\alpha)$ and $\nu(\alpha)$ (and thus $\sigma^2(\alpha)$) are constants independent of α , then we can integrate this to obtain Z and thus a closed expression for the stationary distribution:

$$\psi(\alpha) = \frac{1}{\ln\left(\frac{\delta + \alpha_{\max} + \gamma}{\delta + \alpha_{\min} + \gamma}\right)} \frac{1}{(\delta + \alpha + \gamma)}. \quad (\text{S.33})$$

Because this is a zero-flux solution, α_{\min} and α_{\max} may be chosen arbitrarily (we have a zero-flux boundary condition). We take $\alpha_{\min} = 0$; α_{\max} must be finite to ensure that Z is finite.

In the next section we will show how one may obtain an analytical approximation in the large n limit.

6.2 Asymptotic Approximation to Stationary Distribution

The stationary distribution (S.31) lends itself to an approximation by Laplace’s method (see *e.g.*, [12]), which tells us that if $\phi(x)$ is a twice differentiable function with a unique local maximum attained at $x_0 \in (a, b)$, and $g(x)$ is continuous, then

$$\int_a^b g(x) e^{n\phi(x)} dx = \sqrt{-\frac{2\pi}{n\phi''(x_0)}} g(x_0) e^{n\phi(x_0)} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

(To order $\frac{1}{n}$, one has

$$\begin{aligned} \int_a^b g(x) e^{n\phi(x)} dx &= \sqrt{-\frac{2\pi}{n\phi''(x_0)}} e^{n\phi(x_0)} \left(g(x_0) \right. \\ &\quad \left. + \frac{1}{n} \left(\frac{1}{2} \frac{g''(x_0)}{\phi''(x_0)} + \frac{1}{8} \frac{g^{(4)}(x_0)}{(\phi''(x_0))^2} + \frac{1}{2} \frac{g'(x_0)\phi^{(3)}(x_0)}{(\phi''(x_0))^2} + \frac{5}{24} \frac{g(x_0)(\phi^{(3)}(x_0))^2}{(\phi''(x_0))^3} \right) + \mathcal{O}\left(\frac{1}{n^2}\right) \right); \end{aligned} \quad (\text{S.34})$$

see *e.g.*, [4]).

We can apply this to our stationary distribution by first observing from (S.32) that

$$\frac{1}{\sigma^2(\alpha)} e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha} = g(\alpha) e^{n\phi(\alpha)}$$

for

$$\phi(\alpha) = \int I_{\text{eq}}(\alpha) \frac{R'_0(\alpha)}{R_0(\alpha)} d\alpha \quad (\text{S.35})$$

and

$$g(\alpha) = \frac{1}{\sigma^2(\alpha)\beta(\alpha)},$$

and we can thus apply Laplace's formula to compute Z .

To find our α_0 , we note that by assumption, we have chosen α_{\min} and α_{\max} so that $\frac{\lambda(R_0(\alpha)-1)}{\beta(\alpha)-\alpha} > 0$ for all $\alpha \in (\alpha_{\min}, \alpha_{\max})$. Thus, all values of α_0 such that $\phi'(\alpha_0) = 0$ satisfy $R'_0(\alpha_0) = 0$.

On the other hand, we recall that $R_0(\alpha) = \frac{\beta(\alpha)}{\delta + \alpha + \gamma}$, so that

$$R'_0(\alpha) = \frac{\beta'(\alpha)}{\delta + \alpha + \gamma} - \frac{\beta(\alpha)}{(\delta + \alpha + \gamma)^2}, \quad (\text{S.36})$$

and thus also $R_0(\alpha_0) = \beta'(\alpha_0)$.

We next observe that

$$R''(\alpha_0) = \frac{\beta''(\alpha_0)}{\delta + \alpha_0 + \gamma}. \quad (\text{S.37})$$

which depends on the choice of tradeoff function $\beta(\alpha)$. We note briefly that this is a local maximum if and only if $\beta''(\alpha_0) < 0$. In particular, if we assume that $R_0(\alpha)$ has a unique global maximum, then it must occur at α_0 . We will henceforth make this assumption (*n.b.*, we don't have to assume this in general, but in applying Laplace's method, we require that α_0 be a global maximum; more generally we could partition $(\alpha_{\min}, \alpha_{\max})$ into intervals containing a single local maxima).

We then have

$$\phi''(\alpha) = I'_{\text{eq}}(\alpha) \frac{R'_0(\alpha)}{R_0(\alpha)} + I_{\text{eq}}(\alpha) \left(\frac{R''_0(\alpha)}{R_0(\alpha)} - \left(\frac{R'_0(\alpha)}{R_0(\alpha)} \right)^2 \right)$$

so that

$$\phi''(\alpha_0) = I_{\text{eq}}(\alpha_0) \frac{R''_0(\alpha_0)}{R_0(\alpha_0)}.$$

Thus,

$$Z = \int_{\alpha_{\min}}^{\alpha_{\max}} \frac{1}{\sigma^2(\alpha)} e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha} d\alpha = \sqrt{\frac{2\pi}{n I_{\text{eq}}(\alpha_0) \frac{|R''_0(\alpha_0)|}{R_0(\alpha_0)}}} e^{n\phi(\alpha_0)} \left(\frac{1}{\sigma^2(\alpha_0)\beta(\alpha_0)} + \mathcal{O}\left(\frac{1}{n}\right) \right).$$

We then have

$$\frac{1}{Z} \frac{1}{\sigma^2(\alpha)} e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha} = \frac{\sigma^2(\alpha_0)\beta(\alpha_0)}{\sigma^2(\alpha)\beta(\alpha)} \frac{1}{\sqrt{\frac{2\pi}{nI_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}}}} e^{n(\phi(\alpha) - \phi(\alpha_0))} \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Next, recalling that $\phi'(\alpha_0) = 0$, a Taylor expansion gives

$$n(\phi(\alpha) - \phi(\alpha_0)) = \frac{1}{2} n (\phi''(\alpha_0)(\alpha - \alpha_0)^2 + \mathcal{O}((\alpha - \alpha_0)^3)).$$

Now, for α close to α_0 , $(\alpha - \alpha_0)^3$ will be quite small, so we can locally approximate our full stationary distribution by a process that is almost a Gaussian with mean α_0 and variance

$$\frac{1}{nI_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}},$$

except for a pre-factor of $\frac{\sigma^2(\alpha_0)\beta(\alpha_0)}{\sigma^2(\alpha)\beta(\alpha)}$, which skews the distribution:

$$\begin{aligned} \frac{1}{Z} \frac{1}{\sigma^2(\alpha)} e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha} &\approx \frac{\sigma^2(\alpha_0)\beta(\alpha_0)}{\sigma^2(\alpha)\beta(\alpha)} \frac{1}{\sqrt{\frac{2\pi}{nI_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}}}} e^{-\frac{1}{2} n I_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)} (\alpha - \alpha_0)^2} \\ &= \frac{\eta(\alpha_0)}{\eta(\alpha)} \frac{\nu(\alpha_0)}{\nu(\alpha)} \frac{\beta(\alpha_0)}{\beta(\alpha)} \frac{1}{\sqrt{\frac{2\pi}{nI_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}}}} e^{-\frac{1}{2} n I_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)} (\alpha - \alpha_0)^2}. \end{aligned}$$

Again assuming that $\eta(\alpha)$ and $\nu(\alpha)$ are constants independent of α , this simplifies to

$$\psi_{\text{approx}}(\alpha) = \frac{\beta(\alpha_0)}{\beta(\alpha)} \frac{1}{\sqrt{\frac{2\pi}{nI_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}}}} e^{-\frac{1}{2} n I_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)} (\alpha - \alpha_0)^2}. \quad (\text{S.38})$$

Remark 6. Applying (S.34) with $f(\alpha)g(\alpha)$ in place of $g(\alpha)$ for an arbitrary differentiable function $f(\alpha)$, we may estimate the error in integrating $f(\alpha)$ versus the true and approximate densities for the stationary distribution:

$$\begin{aligned} &\left| \frac{1}{Z} \int_{\alpha_{\min}}^{\alpha_{\max}} f(\alpha)g(\alpha) e^{-n \int \phi(\alpha)} d\alpha - \frac{1}{\sqrt{\frac{2\pi}{n|\phi''(\alpha_0)|}}} g(\alpha_0) \int_{\alpha_{\min}}^{\alpha_{\max}} f(\alpha)g(\alpha) e^{-n \frac{\phi''(\alpha_0)}{2} (\alpha - \alpha_0)^2} d\alpha \right| \\ &= \frac{1}{n} \left| \frac{1}{2} \frac{f'(\alpha_0)\phi'''(\alpha_0)}{(\phi''(\alpha_0))^2} - \frac{1}{2} \frac{f(\alpha_0)g''(\alpha_0)}{g(\alpha_0)|\phi''(\alpha_0)|} \right| + \mathcal{O}\left(\frac{1}{n^2}\right) \end{aligned}$$

From this, see that in the bounded Lipschitz metric on probability measures, the difference between

the true stationary approximation $\psi(d\alpha)$ and the Laplace approximation $\psi_{\text{approx}}(d\alpha)$ satisfies

$$\lim_{n \rightarrow \infty} d_{\text{BL}}(\psi, \psi_{\text{approx}}) \leq \frac{1}{2} \left| \frac{\phi'''(\alpha_0)}{(\phi''(\alpha_0))^2} \right| + \frac{1}{2} \left| \frac{g''(\alpha_0)}{g(\alpha_0)|\phi''(\alpha_0)|} \right|.$$

Unfortunately, we cannot similarly bound the total variation distance: for any $M > 0$ the function $f(\alpha) := e^{-M(\alpha - \alpha_{\min})^2}$ satisfies

$$\sup_{\alpha \in [\alpha_{\min}, \alpha_{\max}]} |f(\alpha)| \leq 1,$$

but, since

$$\sup_{\alpha \in [\alpha_{\min}, \alpha_{\max}]} |f(\alpha)| = \sqrt{2M} e^{-\frac{1}{2}},$$

the bound

$$\frac{1}{n} \left| \frac{1}{2} \frac{f'(\alpha_0)\phi'''(\alpha_0)}{(\phi''(\alpha_0))^2} - \frac{1}{2} \frac{f(\alpha_0)g''(\alpha_0)}{g(\alpha_0)|\phi''(\alpha_0)|} \right|$$

may be made arbitrarily large.

6.3 Mean & Mode of the Stationary Distribution

We observe that, whilst the stationary distribution is closely related to a Gaussian centred at α_0 , the value of the virulence that maximises $R_0(\alpha)$, the full stationary distribution does not have mean α_0 , and α_0 is not the most probable value of the virulence.

6.3.1 Estimating the Mean

Applying Laplace's method with $g(\alpha)$ replaced by $\alpha g(\alpha)$ allows us to estimate the mean of the stationary distribution; to lowest order, we have

$$\int_{\alpha_{\min}}^{\alpha_{\max}} \alpha g(\alpha) e^{n\phi(\alpha)} d\alpha = \sqrt{-\frac{2\pi}{n\phi''(\alpha_0)}} \alpha_0 g(\alpha_0) e^{n\phi(\alpha_0)} (1 + \mathcal{O}(\frac{1}{n})),$$

so that, normalising by our prior estimate of Z , we find that the mean is α_0 to order $\mathcal{O}(\frac{1}{n})$. To observe the effects of a finite population size, we can use higher order corrections to Laplace's method to obtain the $\mathcal{O}(\frac{1}{n})$ terms in both the integral above and in Z ; we omit the calculations, but remark that the mean can then be shown to be

$$\begin{aligned} \alpha_0 - \frac{1}{n} \left(\frac{g'(\alpha_0)}{g(\alpha_0)\phi''(\alpha_0)} - \frac{\phi'''(\alpha_0)}{2(\phi''(\alpha_0))^2} \right) + o(\frac{1}{n}) \\ = \alpha_0 + \frac{1}{n I_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}} \left(\frac{g'(\alpha_0)}{g(\alpha_0)} - \frac{\phi'''(\alpha_0)}{2\phi''(\alpha_0)} \right) + o(\frac{1}{n}) \end{aligned}$$

where

$$\begin{aligned}\frac{g'(\alpha_0)}{g(\alpha_0)} &= -\frac{\eta'(\alpha_0)}{\eta(\alpha_0)} - \frac{\nu'(\alpha_0)}{\nu(\alpha_0)} - \frac{\beta'(\alpha_0)}{\beta(\alpha_0)} \\ &= -\frac{\eta'(\alpha_0)}{\eta(\alpha_0)} - \frac{\nu'(\alpha_0)}{\nu(\alpha_0)} - \frac{1}{\delta + \alpha_0 + \gamma},\end{aligned}$$

which simplifies to $-\frac{1}{\delta + \alpha_0 + \gamma}$ in the case when η and ν are independent of α , and

$$\frac{\phi'''(\alpha_0)}{\phi''(\alpha_0)} = \frac{R_0'''(\alpha_0)}{R_0''(\alpha_0)} - \frac{\beta'(\alpha_0) - 1}{\beta(\alpha_0) - \alpha_0} = \frac{I'_{\text{eq}}(\alpha_0)}{I_{\text{eq}}(\alpha_0)} - \frac{R_0'''(\alpha_0)}{|R_0''(\alpha_0)|}.$$

We note that this order $\mathcal{O}(\frac{1}{n})$ term is proportional to the variance of the best-fit Gaussian of the previous section. Note also that

$$I'_{\text{eq}}(\alpha) = \frac{\lambda R_0'(\alpha)}{\beta(\alpha) - \alpha} - \frac{\lambda R_0(\alpha)(\beta'(\alpha) - 1)}{(\beta(\alpha) - \alpha)^2},$$

so that $I'_{\text{eq}}(\alpha_0) = -\frac{\lambda R_0(\alpha_0)(R_0(\alpha_0) - 1)}{(\beta(\alpha_0) - \alpha_0)^2} < 0$.

One may similarly show that the variance and skewness of the stationary distribution are

$$\frac{1}{n} \frac{1}{\phi''(\alpha_0)} + o(\frac{1}{n}) = \frac{1}{n I_{\text{eq}}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}} + o(\frac{1}{n})$$

and

$$\frac{1}{n} \frac{3g'(\alpha_0)}{g(\alpha_0)(\phi''(\alpha_0))^2} + o(\frac{1}{n})$$

respectively. We note that if $\eta'(\alpha_0) \geq 0$ and $\nu'(\alpha_0) \geq 0$ (for example, if both rates are independent of α), then the stationary distribution has negative skew.

6.3.2 Estimating the Mode

We begin by observing the density function of the stationary distribution (S.31) may be written as

$$e^{\int \frac{2\mu(\alpha)}{\sigma^2(\alpha)} d\alpha - \ln \frac{\sigma^2(\alpha)}{Z}} d\alpha$$

and thus has its maximum where

$$-\frac{2\mu(\alpha)}{\sigma^2(\alpha)} = \frac{d}{d\alpha} \ln \frac{\sigma^2(\alpha)}{Z},$$

or equivalently, for the value of α^* such that

$$n\phi'(\alpha^*) = -\frac{g'(\alpha^*)}{g(\alpha^*)}$$

where $\phi(\alpha)$ is given by (S.35).

When $\sigma^2(\alpha)$ is constant (*i.e.*, the variance of the mutation kernel is independent of the resident variance) then this reduces to

$$\frac{2\mu(\alpha^*)}{\sigma^2} = 0,$$

and thus $\mu(\alpha^*) = 0$. Recalling (S.29), this tells us that

$$\frac{1}{\delta + \alpha^* + \gamma} - \left(1 - \frac{1}{nI_{\text{eq}}(\alpha^*)}\right) \frac{\beta'(\alpha^*)}{\beta(\alpha^*)} = 0,$$

and thus

$$\beta'(\alpha^*) = R_0(\alpha^*) \left(1 + \frac{1}{nI_{\text{eq}}(\alpha^*) - 1}\right). \quad (\text{S.39})$$

On the other hand, (S.36) tells us that

$$\frac{R'_0(\alpha)}{R_0(\alpha)} = \frac{\beta'(\alpha)}{\beta(\alpha)} - \frac{1}{\delta + \alpha + \gamma},$$

so that

$$\frac{R'_0(\alpha^*)}{R_0(\alpha^*)} = \frac{1}{nI_{\text{eq}}(\alpha^*)} \frac{\beta'(\alpha^*)}{\beta(\alpha^*)} > 0,$$

since $\beta(\alpha)$ is increasing. In particular, since $R'_0(\alpha)$ is maximized at α_0 , we see immediately that $\alpha^* < \alpha_0$.

Even in this special case, we cannot solve for α^* exactly. Instead we will seek a perturbative solution to

$$\phi'(\alpha^*) = -\frac{1}{n} \frac{g'(\alpha^*)}{g(\alpha^*)} \quad (\text{S.40})$$

in the general case. We already know that $\phi'(\alpha_0) = 0$; we thus seek a solution of the form

$$\alpha^* = \alpha_0 + \sum_{i=1}^{\infty} \frac{\alpha_i}{n^i}.$$

Substituting this into (S.40) and Taylor expanding right and left, we find that

$$\phi'(\alpha_0) + \frac{1}{n} \phi''(\alpha_0) \alpha_1 = -\frac{1}{n} \frac{g'(\alpha_0)}{g(\alpha_0)} + o\left(\frac{1}{n}\right)$$

i.e., that

$$\alpha^* = \alpha_0 - \frac{1}{n} \frac{g'(\alpha_0)}{g(\alpha_0)\phi''(\alpha_0)} + o\left(\frac{1}{n}\right) = \alpha_0 + \frac{1}{nI_{\text{eq}}(\alpha_0) \frac{|R''_0(\alpha_0)|}{R_0(\alpha_0)}} \frac{g'(\alpha_0)}{g(\alpha_0)} + o\left(\frac{1}{n}\right)$$

which may be expanded using the expression for $\frac{g'(\alpha_0)}{g(\alpha_0)}$ given in the previous sections. In particular,

when when $\eta(\alpha)$ and $\nu(\alpha)$ (and thus $\sigma^2(\alpha)$) are constants independent of α , we have that

$$\alpha^* = \alpha_0 - \frac{1}{nI_{eq}(\alpha_0) \frac{|R_0''(\alpha_0)|}{R_0(\alpha_0)}} \frac{1}{\delta + \alpha_0 + \gamma} + o\left(\frac{1}{n}\right), \quad (\text{S.41})$$

so that to first order, the modal virulence is the virulence maximizing $R_0(\alpha)$ less the product of the variance of the best-fit Gaussian and the expected infectious period when the virulence is α_0 .

7 Some Remarks Regarding Simulations

For all simulations, we simulate the Markov process described in Table 1.

For Figure 2, we consider a single individual infected with the mutant strain with the initial number of susceptibles, infecteds and total individuals set equal to the closest integers to nS_{eq} , nI_{eq} and nN_{eq} , where (S_{eq}, I_{eq}, N_{eq}) is the (deterministic) endemic equilibrium with only the resident strain present. No further mutations are allowed to occur. For each value of s and σ , 10^6 simulations were allowed to continue until fixation of the mutant or resident, and we plot the fraction in which the mutant fixed.

For Figures 3, 4 & 5, following Section 6, we impose a tradeoff between β and α , here taking

$$\beta(\alpha) = (\delta + \alpha + \gamma)(R_{0,\max} - w(\alpha - \alpha_0)^2).$$

The simulations were started with $nN_{eq}(\alpha_0)$ individuals, $nI_{eq}(\alpha_0)$ infected with a resident strain with virulence α_0 – the virulence maximizing $R_0(\alpha)$ – and $nS_{eq}(\alpha_0)$ susceptibles (as previously, $(S_{eq}(\alpha_0), N_{eq}(\alpha_0), I_{eq}(\alpha_0))$ is the endemic equilibrium for the deterministic limit with a resident strain of virulence α_0 .) With probability μ per time-step, an individual was selected uniformly at random from the infected population, and the virulence of the mutant pathogen was selected according to the kernel

$$K(\alpha_m, \alpha) = \begin{cases} \frac{1}{2} & \text{if } \alpha_m - \alpha = \pm\Delta\alpha, \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

for $\Delta\alpha = \frac{6}{100}$, $\alpha_{\min} = 0$, and $\alpha_{\max} = 6$. There are thus 100 discrete values for the virulence (*n.b.*, by assumption, for $\alpha = \alpha_{\min}$, mutations $\Delta\alpha$ occur with probability $\frac{1}{2}$, whereas the process remains at α_{\min} with probability $\frac{1}{2}$, and similarly for $\alpha = \alpha_{\max}$) which are the bins of the histogram in Figures 3, 4 & 5.

In [7], it is shown that μ has to be of the order of $\mathcal{O}\left(\frac{1}{n \ln n}\right)$ or smaller to ensure that mutations occur sufficiently rarely that little or no clonal interference occurs, and such that the resident population is in a small neighbourhood of the equilibrium of the deterministic limit when a mutation occurs. In Figures 3& 4, we take $n = 200$, so $n \ln n \approx 1060$ and take $\mu = \frac{1}{1000}$ to be in the appropriate range. A small amount of clonal interference was observed, but it did not contribute significantly to the histogram. For selection, taking $w = 0$ forces all strains to have equal R_0 , ensuring weak selection, whereas for $w > 0$,

$$|R_0(\alpha) - R_0(\alpha')| \approx |R_0'(\alpha)(\alpha - \alpha')| = 2wR_{0,\max}|\alpha_0 - \alpha||\alpha - \alpha'| \leq 1.44w,$$

so selection is always of order $\mathcal{O}(\frac{1}{n})$ for $w = 0.01$, whereas when $w = 0.1$, one could see strong selection when $|\alpha_0 - \alpha|$ is large. Despite this, our analytical results continue to accurately predict the stationary distribution in Figures 4C & 5, where selection need not be weak (α is far from α_0).

8 Some Rigorous Demonstrations

8.1 Proof of Proposition 2

Substituting f with Π_i in Itô's formula (S.8) for jump processes yields

$$\begin{aligned} P_i^{(n)}(t) = P_i^{(n)}(0) &+ \int_0^t \sum_{j=1}^d \frac{\partial \Pi_i}{\partial x_j}(\bar{\mathbf{I}}^{(n)}(s)) F_j^{(n)}(\bar{\mathbf{E}}^{(n)}(s)) + \frac{1}{2} \sum_{j,k=1}^d a_{jk}^{(n)}(\bar{\mathbf{E}}^{(n)}(s)) \frac{\partial \Pi_i}{\partial x_j \partial x_k}(\bar{\mathbf{I}}^{(n)}(s)) ds \\ &+ \frac{1}{n} \int_0^t \sum_{j=1}^d \frac{\partial \Pi_i}{\partial x_j}(\bar{\mathbf{I}}^{(n)}(s)) dM_j^{(n)}(s) + \varepsilon_i^{(n)}(t), \end{aligned} \quad (\text{S.42})$$

where $\varepsilon_i^{(n)}(t)$ can be expressed thanks to Equation (S.9) as

$$\begin{aligned} \varepsilon_i^{(n)}(t) = \sum_{s < t} \Pi_i(\bar{\mathbf{I}}^{(n)}(s)) - \Pi_i(\bar{\mathbf{I}}^{(n)}(s-)) &- \sum_{j=1}^d \frac{\partial \Pi_i}{\partial x_j}(\bar{\mathbf{I}}^{(n)}(s-)) \Delta \bar{I}_j^{(n)}(s) \\ &- \frac{1}{2} \sum_{j,k=1}^d \frac{\partial \Pi_i}{\partial x_j \partial x_k}(\bar{\mathbf{I}}^{(n)}(s-)) \Delta \bar{I}_j^{(n)}(s) \Delta \bar{I}_k^{(n)}(s). \end{aligned}$$

Now some elementary computations yield

$$\frac{\partial \Pi_i}{\partial x_j} = \frac{1}{\sum_{l=1}^d x_l} \left(\mathbb{1}_{\{i=j\}} - \frac{x_i}{\sum_{l=1}^d x_l} \right) \quad \text{and} \quad \frac{\partial \Pi_i}{\partial x_j \partial x_k} = -\frac{1}{\left(\sum_{l=1}^d x_l \right)^2} \left(\mathbb{1}_{\{i=j\}} + \mathbb{1}_{\{i=k\}} - 2 \frac{x_i}{\sum_{l=1}^d x_l} \right),$$

where $\mathbb{1}_{\{i=j\}}$ is equal to 1 if $i = j$ and 0 otherwise. Substituting these and (S.6) into (S.42) yields after some simplification Equation (S.10) in Proposition 2.

Now it remains to prove that $n^2 \varepsilon_i^{(n)}(t)$ is uniformly bounded with high probability. Taylor's theorem tells us that

$$\begin{aligned} \Pi_i(\bar{\mathbf{I}}^{(n)}(s)) - \Pi_i(\bar{\mathbf{I}}^{(n)}(s-)) &- \sum_{j=1}^d \frac{\partial \Pi_i}{\partial x_j}(\bar{\mathbf{I}}^{(n)}(s-)) \Delta \bar{I}_j^{(n)}(s) - \frac{1}{2} \sum_{j,k=1}^d \frac{\partial \Pi_i}{\partial x_j \partial x_k}(\bar{\mathbf{I}}^{(n)}(s-)) \Delta \bar{I}_j^{(n)}(s) \Delta \bar{I}_k^{(n)}(s) \\ &= \sum_{j,k=1}^d g_{ij}(\bar{\mathbf{I}}^{(n)}(s), \bar{\mathbf{I}}^{(n)}(s-)) \Delta \bar{I}_j^{(n)}(s) \Delta \bar{I}_k^{(n)}(s), \end{aligned}$$

where the functions $g_{ij}(\mathbf{x}, \mathbf{y})$ satisfy

$$\lim_{\mathbf{x} \rightarrow \mathbf{y}} g_{ij}(\mathbf{x}, \mathbf{y}) = 0$$

uniformly on compact sets.

Now, recalling

$$\begin{aligned} \bar{I}_i^{(n)}(t) &= \bar{I}_i^{(n)}(0) + \frac{1}{n} P_{-\mathbf{e}_0 + \mathbf{e}_i} \left(n \int_0^t \frac{\beta_i^{(n)} \bar{S}^{(n)}(s) \bar{I}_i^{(n)}(s)}{\bar{N}^{(n)}(s)} ds \right) \\ &\quad - \frac{1}{n} P_{-\mathbf{e}_i - \mathbf{e}_{d+1}} \left(n \int_0^t (\delta^{(n)} + \alpha_i^{(n)}) \bar{I}_i^{(n)}(s) ds \right) - \frac{1}{n} P_{-\mathbf{e}_i} \left(n \int_0^t \gamma_i^{(n)} \bar{I}_i^{(n)}(s) ds \right), \end{aligned}$$

we see that $\Delta \bar{I}_i^{(n)}(s)$ is non-zero only at the jump-times of the Poisson processes and are always of magnitude $\frac{1}{n}$. In particular, since $f_i(\mathbf{x})$ is smooth outside of a neighbourhood of $\mathbf{0}$, we can conclude that g_{ij} is bounded above by a constant multiple of $\|\mathbf{x} - \mathbf{y}\|$; this allows us to conclude that $|g_{ij}(\bar{\mathbf{I}}^{(n)}(s), \bar{\mathbf{I}}^{(n)}(s-))| \leq \frac{C}{n}$ and

$$|\varepsilon_i^{(n)}(t)| \leq \frac{C}{n} \sum_{s < t} \sum_{j, k=1}^d |\Delta \bar{I}_j^{(n)}(s)| |\Delta \bar{I}_k^{(n)}(s)|.$$

Further, $|\Delta \bar{I}_j^{(n)}(s)| |\Delta \bar{I}_k^{(n)}(s)|$ is non-zero if some pair of processes in $P_{-\mathbf{e}_0 + \mathbf{e}_j}, P_{-\mathbf{e}_j - \mathbf{e}_{d+1}}, P_{-\mathbf{e}_j}$, and $P_{-\mathbf{e}_0 + \mathbf{e}_k}, P_{-\mathbf{e}_k - \mathbf{e}_{d+1}}, P_{-\mathbf{e}_k}$ jump simultaneously; if $j \neq k$, the Poisson processes are independent, and probability of such an event in an interval $[t, t + \Delta t)$ is $\mathcal{O}(\Delta t^2)$, and thus tends to 0 if as $\Delta t \rightarrow 0$ *i.e.*, $|\Delta \bar{I}_j^{(n)}(s)| |\Delta \bar{I}_k^{(n)}(s)| \neq 0$ if and only if $j = k$. Moreover, the processes $P_{-\mathbf{e}_0 + \mathbf{e}_j}, P_{-\mathbf{e}_0 - \mathbf{e}_{d+1}}$ and $P_{-\mathbf{e}_j}$ are also independent, and thus cannot jump simultaneously, so that $|\Delta \bar{I}_j^{(n)}(s)| |\Delta \bar{I}_k^{(n)}(s)| \neq 0$ (and is thus equal to $\frac{1}{n^2}$) at exactly the jump times of these Poisson processes; *i.e.*,

$$\begin{aligned} \sum_{s < t} \sum_{j, k=1}^d |\Delta \bar{I}_j^{(n)}(s)| |\Delta \bar{I}_k^{(n)}(s)| &= \frac{1}{n^2} P_{-\mathbf{e}_0 + \mathbf{e}_i} \left(n \int_0^t \frac{\beta_i^{(n)} \bar{S}^{(n)}(s) \bar{I}_i^{(n)}(s)}{\bar{N}^{(n)}(s)} ds \right) \\ &\quad + \frac{1}{n^2} P_{-\mathbf{e}_i - \mathbf{e}_{d+1}} \left(n \int_0^t (\delta^{(n)} + \alpha_i^{(n)}) \bar{I}_i^{(n)}(s) ds \right) \\ &\quad + \frac{1}{n^2} P_{-\mathbf{e}_i} \left(n \int_0^t \gamma_i^{(n)} \bar{I}_i^{(n)}(s) ds \right) \end{aligned}$$

We seek an upper bound on this quantity. To that end, we begin by observing that $\bar{S}^{(n)}(t)$ and each $\bar{I}_i^{(n)}(t)$ is bounded above by $\bar{N}^{(n)}(t)$, and that

$$\bar{N}^{(n)}(t) \leq \bar{N}^{(n)}(0) + \frac{1}{n} P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}t),$$

and, since this Poisson process is increasing in t , we have that for $t \leq T$,

$$\bar{N}^{(n)}(t) \leq \bar{N}^{(n)}(0) + \frac{1}{n} P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}T).$$

Now,

$$\mathbb{E} \left[\frac{1}{n} P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}T) \right] = \lambda^{(n)}T,$$

and, applying Chebyshev's inequality, we see that for any $C > 0$,

$$\begin{aligned} \mathbb{P} \left\{ \left| \frac{1}{n} P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}T) - \lambda^{(n)}T \right| > C \right\} &= \mathbb{P} \left\{ |P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}T) - n\lambda^{(n)}T| > Cn \right\} \\ &\leq \frac{\mathbb{E} \left[(P_{\mathbf{e}_0 + \mathbf{e}_{d+1}}(n\lambda^{(n)}T) - n\lambda^{(n)}T)^2 \right]}{C^2 n^2} = \frac{\lambda^{(n)}T}{C^2 n} \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$. Thus, for any fixed $T > 0$, $\bar{N}^{(n)}(t)$ is bounded above and below on $[0, T]$ by *e.g.*, $\bar{N}^{(n)}(0) + \lambda^{(n)}T \pm 1$, with probability that approaches one as n tends to infinity.

Thus, for example,

$$\int_0^t \frac{\beta_i^{(n)} \bar{S}^{(n)}(s) \bar{I}_i^{(n)}(s)}{\bar{N}^{(n)}(s)} ds \leq \frac{\beta_i^{(n)} (\bar{N}^{(n)}(0) + \lambda^{(n)}T + 1)^2 T}{\bar{N}^{(n)}(0) + \lambda^{(n)}T - 1},$$

and we may proceed exactly as above to conclude that for $t \leq T$,

$$\frac{1}{n} P_{-\mathbf{e}_0 + \mathbf{e}_i} \left(n \int_0^t \frac{\beta_i^{(n)} \bar{S}^{(n)}(s) \bar{I}_i^{(n)}(s)}{\bar{N}^{(n)}(s)} ds \right)$$

is bounded above with probability approaching 1 as $n \rightarrow \infty$, and similarly for the other Poisson processes, from which we conclude that there exists some constant C' such that

$$|\varepsilon_i^{(n)}(t)| \leq \frac{C'}{n^2}$$

with high probability.

8.2 Proof of Proposition 5

In this section, we will make the heuristic argument of Section 5.2 rigorous using the technique of coupling (see [2] for a very good introduction): we start by constructing birth and death processes that bound $I_1^{(n)}(t)$ above and below provided $\bar{S}^{(n)}(t)$ and $\bar{N}^{(n)}(t)$ remain within ε of the endemic equilibrium, and such that the upper and lower bounds approach one another as $\varepsilon \rightarrow 0$. Finally, we show that the probability that $\bar{S}^{(n)}(t)$ and $\bar{N}^{(n)}(t)$ depart a ε -neighbourhood of the endemic equilibrium before strain 2 has either successfully invaded or gone extinct goes to 0 as $n \rightarrow \infty$. Since ε is arbitrary, we recover the naïve branching process result.

8.2.1 Macroscopic Initial Frequencies

In this section we prove the following extinction of part (i) of Proposition 5, where the population is infected with $d \geq 2$ strains.

Proposition 11. *Suppose that $R_{0,1} > R_{0,i}$ for all $i > 1$. If $\bar{I}_1^{(n)}(0) \rightarrow I_1(0) > 0$, then all strains $i > 1$ will go extinct with high probability.*

In light of the results in Section 5.1, we might, without loss of generality, assume that at time $t = 0$, the process is in some neighbourhood of endemic fixed point $\bar{\mathbf{E}}^{\star,1} = (\bar{S}^{\star,1}, \bar{I}_1^{\star,1}, \dots, \bar{I}_d^{\star,1}, \bar{N}^{\star,1})$, as defined by (S.13).

In particular, fix $\varepsilon > 0$ such that

$$1 < \frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} < \frac{1}{R_{0,1}} = \frac{\bar{S}^{\star,1}}{\bar{N}^{\star,1}} < \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon} < \frac{1}{R_{0,2}} < \frac{1}{R_{0,3}} < \dots < \frac{1}{R_{0,d}}.$$

For reasons that will become transparent below, we will assume that

$$\|\bar{\mathbf{E}}^{(n)}(0) - \bar{\mathbf{E}}^{\star,1}\| < B\varepsilon.$$

for some constant $0 < B < 1$ that will be determined later.

Let

$$\tau_\varepsilon^{(n)} := \inf \left\{ t : \|\bar{\mathbf{E}}^{(n)}(t) - \bar{\mathbf{E}}^{\star,1}\| > \varepsilon \right\},$$

where we adopt the convention that $\tau_\varepsilon^{(n)} = \infty$ if $\|\bar{\mathbf{E}}^{(n)}(t) - \bar{\mathbf{E}}^{\star,1}\| < \varepsilon$ for all t .

Provided $t < \tau_\varepsilon^{(n)}$, we have that

$$\frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} < \frac{\bar{S}^{(n)}(t)}{\bar{N}^{(n)}(t)} < \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon}.$$

Next, fix $\eta > 0$ sufficiently small that

$$(\beta_i + \eta) \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon} - (\delta + \alpha_i + \gamma_i - 3\eta) = \beta_i \left(\frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon} - \frac{1}{R_{0,i}} \right) + \eta \left(3 + \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon} \right) < 0$$

Since $I_i^{(n)}(0) \rightarrow I_i(0)$ for $i > 1$ and $\beta_i^{(n)} \rightarrow \beta_i$, *etc.* we can assume that n is sufficiently large that

$$|\beta_i^{(n)} - \beta_i| < \eta, \quad |\delta^{(n)} - \delta| < \eta, \quad |\alpha_i^{(n)} - \alpha_i| < \eta, \quad \text{and} \quad |\gamma_i^{(n)} - \gamma_i| < \eta$$

for all $i > 1$.

Thus, for $t < \tau_\varepsilon^{(n)}$, the per-infective transmission rate for strain i satisfies

$$(\beta_i - \eta) \frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} < \frac{\beta_i^{(n)} \mathcal{S}^{(n)}(t-)}{N^{(n)}(t-)} < (\beta_i + \eta) \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon}$$

whereas the total per-infective rate of removal of strain i satisfies

$$\delta + \alpha_i + \gamma_i + 3\eta > \delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)} > \delta + \alpha_i + \gamma_i - 3\eta.$$

Lemma 3. *Provided $t < \tau_\varepsilon^{(n)}$, the number of infectives of strain i is stochastically smaller³ than the birth and death process $Z_i^+(t)$ with $Z_i^+(0) = I_i^{(n)}(0)$ and birth and death rates*

$$(\beta_i + \eta) \frac{\bar{S}^{*,1} + \varepsilon}{\bar{N}^{*,1} - \varepsilon} \quad \text{and} \quad \delta + \alpha_i + \gamma_i - 3\eta,$$

and stochastically greater than the birth and death process $Z_i^-(t)$ with $Z_i^-(0) = I_i^{(n)}(0)$ and birth and death rates

$$(\beta_i - \eta) \frac{\bar{S}^{*,1} - \varepsilon}{\bar{N}^{*,1} + \varepsilon} \quad \text{and} \quad \delta + \alpha_i + \gamma_i + 3\eta.$$

Proof. It suffices to construct coupled versions of $I_i^{(n)}(t)$, $Z_i^+(t)$ and $Z_i^-(t)$ such that

$$Z_i^+(t) \geq I_i^{(n)}(t) \geq Z_i^-(t).$$

We will do so inductively, at each step constructing the processes up to the next among the aggregated jump times of all three processes, which we denote

$$0 = \tau_0 < \tau_1 < \tau_2 < \dots$$

For our underlying probability space, we assume sequences of independent rate 1 exponential random variables \mathcal{E}_k and independent uniformly distributed random variables \mathcal{U}_k on $[0, 1]$, for $k = 1, 2, \dots$

Suppose that $I_i^{(n)}(t)$, $Z_i^+(t)$ and $Z_i^-(t)$ have been constructed up to τ_k (trivially true for $k = 0$). Note that

$$\rho_{k+1} := (\beta_i + \eta) \frac{\bar{S}^{*,1} + \varepsilon}{\bar{N}^{*,1} - \varepsilon} Z_k^+(\tau_k) + (\delta + \alpha_i + \gamma_i + 3) Z_k^+(\tau_k)$$

is an upper bound on the combined rate of all transitions for all three processes. Set

$$\tau_{k+1} = \frac{\mathcal{E}_{k+1}}{\rho_{k+1}},$$

so τ_{k+1} is a rate ρ_{k+1} exponential random variable. Next, for $t < \tau_{k+1}$ we set

$$I_i^{(n)}(t) = I_i^{(n)}(\tau_k), \quad Z_i^+(t) = Z_i^+(\tau_k) \quad \text{and} \quad Z_i^-(t) = Z_i^-(\tau_k).$$

³Given random variables X and X' , we say that X is *stochastically smaller* than X' , denoted $X \preceq X'$ if

$$\mathbb{P}\{X' \geq x\} \geq \mathbb{P}\{X \geq x\}.$$

Similarly, a stochastic process X is stochastically smaller than the process X' if $S(t) \preceq X'(t)$ for all $t \geq 0$. One defines *stochastically greater* analogously.

Finally, we set

$$\begin{aligned}
& (I_i^{(n)}(\tau_{k+1}), Z_i^+(\tau_{k+1}), Z_i^-(\tau_{k+1})) \\
= & \begin{cases} (I_i^{(n)}(\tau_k)+1, Z_i^+(\tau_k)+1, Z_i^-(\tau_k)+1) & \text{if } \mathcal{U}_{k+1} \leq \frac{(\beta_i-\eta) \frac{\bar{S}^{*,1}-\varepsilon}{\bar{N}^{*,1}+\varepsilon}}{\rho_{k+1}} \\ (I_i^{(n)}(\tau_k)+1, Z_i^+(\tau_k)+1, Z_i^-(\tau_k)) & \text{if } \frac{(\beta_i-\eta) \frac{\bar{S}^{*,1}-\varepsilon}{\bar{N}^{*,1}+\varepsilon}}{\rho_{k+1}} \leq \mathcal{U}_{k+1} < \frac{\beta_i \frac{\bar{S}^{(n)}(t)}{\bar{N}^{(n)}(t)}}{\rho_{k+1}} \\ (I_i^{(n)}(\tau_k)+1, Z_i^+(\tau_k), Z_i^-(\tau_k)) & \text{if } \frac{\beta_i \frac{\bar{S}^{(n)}(t)}{\bar{N}^{(n)}(t)}}{\rho_{k+1}} < \mathcal{U}_{k+1} \leq \frac{(\beta_i+\eta) \frac{\bar{S}^{*,1}+\varepsilon}{\bar{N}^{*,1}-\varepsilon}}{\rho_{k+1}} \\ (I_i^{(n)}(\tau_k)-1, Z_i^+(\tau_k)-1, Z_i^-(\tau_k)-1) & \text{if } \frac{(\beta_i+\eta) \frac{\bar{S}^{*,1}+\varepsilon}{\bar{N}^{*,1}-\varepsilon}}{\rho_{k+1}} < \mathcal{U}_{k+1} \leq \frac{(\beta_i+\eta) \frac{\bar{S}^{*,1}+\varepsilon}{\bar{N}^{*,1}-\varepsilon} + \delta + \alpha_i + \gamma_i - 3\eta}{\rho_{k+1}} \\ (I_i^{(n)}(\tau_k), Z_i^+(\tau_k)-1, Z_i^-(\tau_k)-1) & \text{if } \frac{(\beta_i+\eta) \frac{\bar{S}^{*,1}+\varepsilon}{\bar{N}^{*,1}-\varepsilon} + \delta + \alpha_i + \gamma_i - 3\eta}{\rho_{k+1}} < \mathcal{U}_{k+1} \leq \frac{(\beta_i+\eta) \frac{\bar{S}^{*,1}+\varepsilon}{\bar{N}^{*,1}-\varepsilon} + \delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)}}{\rho_{k+1}} \\ (I_i^{(n)}(\tau_k), Z_i^+(\tau_k), Z_i^-(\tau_k)-1) & \text{if } \frac{(\beta_i+\eta) \frac{\bar{S}^{*,1}+\varepsilon}{\bar{N}^{*,1}-\varepsilon} + \delta^{(n)} + \alpha_i^{(n)} + \gamma_i^{(n)}}{\rho_{k+1}} < \mathcal{U}_{k+1} \leq 1. \end{cases}
\end{aligned}$$

It is readily verified that the resulting processes have the correct jump rates. \square

Remark 7. Note that given $\varepsilon > \varepsilon_1 > 0$, choosing $\eta > 0$ as before and $\eta_1 > 0$ analogously, we can similarly construct processes $Z_i^{+,1}(t)$ and $Z_i^{-,1}(t)$ such that

$$Z_i^+(t) \succeq Z_i^{+,1}(t) \succeq I_i^{(n)}(t) \succeq Z_i^{-,1}(t) \succeq Z_i^-(t),$$

etc. We shall apply this with a decreasing sequence of values $\varepsilon_n > 0$ below.

Now, our choice of η ensures that $Z_i^+(t)$ is subcritical for all $i > 1$. In particular, setting

$$\mu_i^+ := (\beta_i + \eta) \frac{\bar{S}^{*,1} + \varepsilon}{\bar{N}^{*,1} - \varepsilon} - (\delta + \alpha_i + \gamma_i - 3\eta) < 0,$$

we have

$$\mathbb{E} [Z_i^+(t)] = Z_i^+(0) e^{\mu_i^+ t},$$

and, for any sequence $t_n > \frac{1}{|\mu_i^+|} \ln n$, for all $i > 1$, we have, using Markov's inequality

$$\mathbb{P} \{Z_i^+(t_n) \geq 1\} \leq \mathbb{E} [Z_i^+(t_n)] \leq B \varepsilon n e^{\mu_i^+ t_n} \rightarrow 0$$

as $n \rightarrow \infty$. Thus, if we show that $\mathbb{P} \left\{ \tau_\varepsilon^{(n)} > t_n \right\} \rightarrow 1$ as $n \rightarrow \infty$, we can conclude that all strains $i > 1$ vanish after before t_n with high probability.

To this end, we start by defining

$$\tau_{\varepsilon,i}^{(n)} := \inf \left\{ t : |\bar{I}_i^{(n)}(t) - \bar{I}_i^{*,1}| \geq \varepsilon n \right\},$$

for $i = 1, \dots, d$. We define $\tau_{\varepsilon,0}^{(n)}$ and $\tau_{\varepsilon,d+1}^{(n)}$ similarly, replacing \bar{I}_i by \bar{S} or \bar{N} respectively in the above

definition (again, $\tau_{\varepsilon,i}^{(n)} = \infty$ should the respective process never exceed εn). We then have

$$\tau_{\varepsilon}^{(n)} = \min_i \tau_{\varepsilon,i}^{(n)}.$$

We continue with a classical result for birth and death processes: for $i > 1$, let

$$a_i = \frac{\delta + \alpha_i + \gamma_i - 3\eta}{(\beta_i + \eta) \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon}} > 1.$$

Then, a simple calculation shows that

$$\mathbb{E} \left[a_i^{Z_i^+(t)} \middle| Z_i^+(s) \right] = a_i^{Z_i^+(s)}$$

i.e., $a_i^{Z_i^+(t)}$ is a martingale. Let

$$\tau_{0,i}^{(n)} = \inf \{ t : Z_i^+(t) = 0 \} \quad \text{and} \quad \tau_i^{(n)} = \min \{ \tau_{0,i}^{(n)}, \tau_{\varepsilon,i}^{(n)} \}.$$

We saw above that $\tau_{0,i}^{(n)}$ and thus $\tau_i^{(n)}$ are with high probability bounded above by any sequence $t_n > \frac{1}{\min_{i>1} |\mu_i^+|} \ln n$. Now,

$$a_i^{Z_i^+(0)} = \mathbb{E} \left[a_i^{Z_i^+(\tau_i^{(n)})} \right] = a_i^{\lfloor \varepsilon n \rfloor} \mathbb{P} \left\{ \tau_{0,i}^{(n)} > \tau_{\varepsilon,i}^{(n)} \right\} + \left(1 - \mathbb{P} \left\{ \tau_{0,i}^{(n)} > \tau_{\varepsilon,i}^{(n)} \right\} \right)$$

i.e.,

$$\mathbb{P} \left\{ \tau_{0,i}^{(n)} > \tau_{\varepsilon,i}^{(n)} \right\} = \frac{a_i^{Z_i^+(0)} - 1}{a_i^{\lfloor \varepsilon n \rfloor} - 1} \geq \frac{a_i^{\lfloor B\varepsilon n \rfloor} - 1}{a_i^{\lfloor \varepsilon n \rfloor} - 1},$$

which converges to 0 as $n \rightarrow \infty$ for any $B < 1$. We thus have

$$\mathbb{P} \left\{ \tau_{\varepsilon,i}^{(n)} > t_n \right\} \rightarrow 1$$

as $n \rightarrow \infty$.

For the remaining three values $\tau_{\varepsilon,i}^{(n)}$, $i = 0, 1, d+1$, we take a different approach, as the values $\bar{S}^{\star,1}$, $\bar{I}_1^{\star,1}$, and $\bar{N}^{\star,1}$ are all non-zero and a branching process approach is no longer appropriate. Instead, we recall the SDE representation of our process (Proposition 1).

Now $\bar{\mathbf{E}}^{\star,1}$ is a stable fixed point for the dynamical system $\dot{\bar{\mathbf{E}}} = \mathbf{F}(\bar{\mathbf{E}})$, so we may write

$$\mathbf{F}(\bar{\mathbf{E}}) = A(\bar{\mathbf{E}} - \bar{\mathbf{E}}^{\star,1}) + \mathbf{G}(\bar{\mathbf{E}} - \bar{\mathbf{E}}^{\star,1}),$$

where $A := \mathbf{DF}(\bar{\mathbf{E}}^{\star,1})$, the Jacobian of $\mathbf{F}(\mathbf{x})$ evaluated at the resident endemic equilibrium, is a stable matrix and

$$\|\mathbf{G}(\bar{\mathbf{E}})\| \leq M \|\bar{\mathbf{E}}\|^2$$

for some fixed $M > 0$.

Now, let

$$\Xi^{(n)}(t) := \bar{E}^{(n)}(t) - \bar{E}^{*,1}.$$

Then, using Duhamel's principle, we have that

$$\begin{aligned} \Xi^{(n)}(t) &= e^{tA}\Xi^{(n)}(0) + \int_0^t e^{(t-s)A} \mathbf{G} \left(\Xi^{(n)}(s) \right) ds \\ &\quad + \frac{1}{n} \int_0^t e^{(t-s)A} (\mathbf{e}_0 + \mathbf{e}_{d+1}) dM_{\mathbf{e}_0 + \mathbf{e}_{d+1}}^{(n)}(t) - \frac{1}{n} \int_0^t e^{(t-s)A} (\mathbf{e}_0 + \mathbf{e}_{d+1}) dM_{-\mathbf{e}_0 - \mathbf{e}_{d+1}}^{(n)}(t) \\ &\quad + \frac{1}{n} \sum_{i=1}^d \int_0^t e^{(t-s)A} (\mathbf{e}_i - \mathbf{e}_0) dM_{-\mathbf{e}_0 + \mathbf{e}_i}^{(n)}(t) - \frac{1}{n} \sum_{i=1}^d \int_0^t e^{(t-s)A} (\mathbf{e}_i + \mathbf{e}_{d+1}) dM_{-\mathbf{e}_i - \mathbf{e}_{d+1}}^{(n)}(t) \\ &\quad - \frac{1}{n} \sum_{i=1}^d \int_0^t e^{(t-s)A} \mathbf{e}_i dM_{-\mathbf{e}_i}^{(n)}(t) - \frac{1}{n} \int_0^t e^{(t-s)A} \mathbf{e}_{d+1} dM_{-\mathbf{e}_{d+1}}^{(n)}(t). \quad (\text{S.43}) \end{aligned}$$

Let $\alpha_1, \dots, \alpha_d$ denote the eigenvalues of A . It is a standard result (see *e.g.*, [30]) that, given any $\alpha < \min\{-\Re(\alpha_j) : \Re(\alpha_j) < 0\}$, there exists a constant C , depending on α such that, when restricted to E_s , we have

$$\|e^{(t-r)A}\| \leq Ce^{-\alpha(t-r)}.$$

Thus,

$$\begin{aligned} \|\Xi^{(n)}(t)\| &\leq Ce^{-\alpha t} \|\Xi^{(n)}(0)\| + \int_0^t Ce^{-\alpha(t-s)} M \|\Xi^{(n)}(s)\|^2 ds \\ &\quad + \frac{1}{n} \int_0^t Ce^{-\alpha(t-s)} dM_{\mathbf{e}_0 + \mathbf{e}_{d+1}}^{(n)}(t) + \frac{1}{n} \int_0^t Ce^{-\alpha(t-s)} dM_{-\mathbf{e}_0 - \mathbf{e}_{d+1}}^{(n)}(t) \\ &\quad + \frac{1}{n} \sum_{i=1}^d \int_0^t Ce^{-\alpha(t-s)} dM_{-\mathbf{e}_0 + \mathbf{e}_i}^{(n)}(t) + \frac{1}{n} \sum_{i=1}^d \int_0^t Ce^{-\alpha(t-s)} dM_{-\mathbf{e}_i - \mathbf{e}_{d+1}}^{(n)}(t) \\ &\quad + \frac{1}{n} \sum_{i=1}^d \int_0^t Ce^{-\alpha(t-s)} dM_{-\mathbf{e}_i}^{(n)}(t) + \frac{1}{n} \int_0^t Ce^{-\alpha(t-s)} dM_{-\mathbf{e}_{d+1}}^{(n)}(t). \end{aligned}$$

Now, fix a sequence $\ln n \ll t_n \ll n$, we observe that

$$\begin{aligned} \mathbb{P} \left\{ \tau_\varepsilon^{(n)} < t_n \right\} &= \mathbb{P} \left\{ \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \|\Xi^{(n)}(t)\| \geq \varepsilon \right\} \\ &\leq \mathbb{P} \left\{ Ce^{-\alpha t} \|\Xi^{(n)}(0)\| \geq \frac{\varepsilon}{5 + 3d} \right\} + \mathbb{P} \left\{ \sup_{t \leq \tau_\varepsilon^{(n)}} \int_0^t Ce^{-\alpha(t-s)} M \|\Xi^{(n)}(s)\|^2 ds \geq \frac{\varepsilon}{5 + 3d} \right\} \\ &\quad + \mathbb{P} \left\{ \frac{1}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t Ce^{-\alpha(t-s)} dM_{\mathbf{e}_0 + \mathbf{e}_{d+1}}^{(n)}(s) \geq \frac{\varepsilon}{5 + 3d} \right\} \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P} \left\{ \frac{1}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t C e^{-\alpha(t-s)} dM_{-e_0 - e_{d+1}}^{(n)}(s) \geq \frac{\varepsilon}{5+3d} \right\} \\
& + \sum_{i=1}^d \mathbb{P} \left\{ \frac{1}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t C e^{-\alpha(t-s)} dM_{-e_0 + e_i}^{(n)}(s) \geq \frac{\varepsilon}{5+3d} \right\} \\
& + \sum_{i=1}^d \mathbb{P} \left\{ \frac{1}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t C e^{-\alpha(t-s)} dM_{-e_i - e_{d+1}}^{(n)}(s) \geq \frac{\varepsilon}{5+3d} \right\} \\
& + \sum_{i=1}^d \mathbb{P} \left\{ \frac{1}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t C e^{-\alpha(t-s)} dM_{-e_i}^{(n)}(s) \geq \frac{\varepsilon}{5+3d} \right\} \\
& \quad + \mathbb{P} \left\{ \frac{1}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t C e^{-\alpha(t-s)} dM_{-e_{d+1}}^{(n)}(s) \geq \frac{\varepsilon}{5+3d} \right\}
\end{aligned}$$

Now,

$$C e^{-\alpha t} \|\Xi^{(n)}(0)\| \leq C B \varepsilon$$

and

$$\sup_{t \leq \tau_\varepsilon^{(n)}} \int_0^t C e^{-\alpha(t-s)} M \|\Xi^{(n)}(s)\|^2 ds \leq \frac{C M}{\alpha} \varepsilon^2.$$

Thus, provided we choose $B < \frac{1}{C(5+3d)}$ and $\varepsilon < \frac{\alpha}{CM(5+3d)}$, then

$$\mathbb{P} \left\{ C e^{-\alpha t} \|\Xi^{(n)}(0)\| \geq \frac{\varepsilon}{5+3d} \right\} = \mathbb{P} \left\{ \sup_{t \leq \tau_\varepsilon^{(n)}} \int_0^t C e^{-\alpha(t-s)} M \|\Xi^{(n)}(s)\|^2 ds \geq \frac{\varepsilon}{5+3d} \right\} = 0,$$

Finally, we turn to the integrals $\int_0^t e^{-\alpha(t-s)} dM_t^{(n)}(s)$. Recall that each of the integrators $M_t^{(n)}(t)$ takes the form

$$\tilde{P}_t \left(n \int \Lambda_t(\bar{\mathbf{E}}^{(n)}(t)) ds \right),$$

for some continuous function Λ_t . We will thus prove the generic lemma:

Lemma 4. *Let P be a Poisson process, $\Lambda : \mathbb{R}^{d+2} \rightarrow \mathbb{R}$ be continuous, and let*

$$M^{(n)}(t) = \tilde{P} \left(n \int \Lambda(\bar{\mathbf{E}}^{(n)}(t)) ds \right).$$

then for any $\alpha > 0$, any constant C , any sequence $t_n \ll n$, and $\tau_\varepsilon^{(n)}$ as above, we have

$$\mathbb{P} \left\{ \frac{C}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t e^{-\alpha(t-s)} dM^{(n)}(s) > R \right\} \rightarrow 0$$

as $n \rightarrow \infty$, for any fixed $R > 0$.

Proof. We start by observing that for $t \leq \tau_\varepsilon^{(n)}$, $\bar{\mathbf{E}}^{(n)}(t) \in \overline{B_\varepsilon(\bar{\mathbf{E}}^{*,1})}$, a compact set, and thus

$$\Lambda(\bar{\mathbf{E}}^{(n)}(t)) \leq \bar{\Lambda}$$

for a constant $\bar{\Lambda} > 0$ depending on Λ , ε , and $\bar{\mathbf{E}}^{*,1}$.

$$\begin{aligned} \mathbb{P} \left\{ \frac{C}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t e^{-\alpha(t-s)} dM^{(n)}(s) > R \right\} &= \mathbb{P} \left\{ \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} e^{-\alpha t} \int_0^t e^{\alpha s} dM^{(n)}(s) > \frac{nR}{C} \right\} \\ &\leq \mathbb{P} \left\{ \sup_{t \leq t_n} e^{-\alpha t} \int_0^{t \wedge \tau_\varepsilon^{(n)}} e^{\alpha s} dM^{(n)}(s) > \frac{nR}{C} \right\} \\ &\leq \sum_{k=0}^{t_n-1} \mathbb{P} \left\{ \sup_{k < t \leq k+1} e^{-\alpha t} \int_0^{t \wedge \tau_\varepsilon^{(n)}} e^{\alpha s} dM^{(n)}(s) > \frac{nR}{C} \right\} \\ &\leq \sum_{k=0}^{t_n-1} \mathbb{P} \left\{ \sup_{k < t \leq k+1} e^{-\alpha k} \int_0^{t \wedge \tau_\varepsilon^{(n)}} e^{\alpha s} dM^{(n)}(s) > \frac{nR}{C} \right\} \\ &= \sum_{k=0}^{t_n-1} \mathbb{P} \left\{ \sup_{k < t \leq k+1} \int_0^{t \wedge \tau_\varepsilon^{(n)}} e^{\alpha s} dM^{(n)}(s) > \frac{nR}{C} e^{\alpha k} \right\} \\ &\leq \sum_{k=0}^{t_n-1} \mathbb{P} \left\{ \sup_{t \leq k+1} \int_0^{t \wedge \tau_\varepsilon^{(n)}} e^{\alpha s} dM^{(n)}(s) > \frac{nR}{C} e^{\alpha k} \right\}. \end{aligned}$$

Now, applying Doob's inequality,

$$\begin{aligned} \mathbb{P} \left\{ \sup_{t \leq k+1} \int_0^{t \wedge \tau_\varepsilon^{(n)}} e^{\alpha s} dM^{(n)}(s) > \frac{nR}{C} e^{\alpha k} \right\} &\leq \frac{C^2}{n^2 R^2} e^{-2\alpha k} \mathbb{E} \left[\left(\int_0^{k+1 \wedge \tau_\varepsilon^{(n)}} e^{\alpha s} dM^{(n)}(s) \right)^2 \right] \\ &= \frac{C^2}{n^2 R^2} e^{-2\alpha k} \mathbb{E} \left[\int_0^{k+1 \wedge \tau_\varepsilon^{(n)}} e^{2\alpha s} n \Lambda(\bar{\mathbf{E}}^{(n)}(s)) ds \right] \leq \frac{C^2}{n R^2} e^{-2\alpha k} \mathbb{E} \left[\int_0^{k+1} e^{2\alpha s} \bar{\Lambda} ds \right] \\ &\leq \frac{C^2 \bar{\Lambda}}{2\alpha n R^2} e^{-2\alpha k} (e^{2\alpha(k+1)} - 1) \leq \frac{C^2 \bar{\Lambda}}{2\alpha n R^2} e^{2\alpha} \end{aligned}$$

Thus,

$$\mathbb{P} \left\{ \frac{C}{n} \sup_{t \leq \tau_\varepsilon^{(n)} \wedge t_n} \int_0^t e^{-\alpha(t-s)} dM^{(n)}(s) > R \right\} \leq \frac{C^2 \bar{\Lambda}}{2\alpha n R^2} e^{2\alpha} t_n \rightarrow 0$$

as $n \rightarrow \infty$. □

Applying the lemma, we have

$$\mathbb{P} \left\{ \tau_\varepsilon^{(n)} < t_n \right\} \rightarrow 0$$

and, with high probability, all strains $i > 1$ will vanish before t_n , and moreover, during this time, the process $\bar{\mathbf{E}}^{(n)}(t)$ will remain in $\overline{B_\varepsilon(\bar{\mathbf{E}}^{*,1})}$.

8.2.2 Novel Strain in Small Number of Copies

We now consider the possibility that a new strain invades an established population. From the results of the previous section, we see that generically, the population will eventually be in an ε -neighbourhood of the fixed point $\bar{\mathbf{E}}^{\star,1}$ for arbitrary $\varepsilon > 0$, and that $I_i^{(n)}(t) \equiv 0$ for all $i > 1$.

If the new strain has reproductive number less than $R_{0,1}$, then the arguments of the preceding section apply directly, and we can conclude that the novel strain will very rapidly go extinct.

If, however, it has higher reproductive number, the invading strain now has a non-zero probability of establishing itself and replacing the resident strain. In this section, we adapt the techniques above to deal with this case (by the above, we may take $d = 2$).

We start by defining the quantities $\tau_{\varepsilon,i}^{(n)}$ and $\tau_\varepsilon^{(n)}$ as before, with the understanding that $\tau_{\varepsilon,1}^{(n)} = \infty$ if the new strain goes extinct before hitting εn .

We now fix $\varepsilon > 0$ such that

$$\frac{1}{R_{0,1}} < \frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} < \frac{1}{R_{0,2}} = \frac{\bar{S}^{\star,1}}{\bar{N}^{\star,1}} < \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon},$$

and $\eta > 0$ sufficiently small that

$$\begin{aligned} \mu_2^+ &:= (\beta_2 + \eta) \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon} - (\delta + \alpha_2 + \gamma_2 - 3\eta) \\ &> \mu_2^- := (\beta_2 - \eta) \frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} - (\delta + \alpha_2 + \gamma_2 + 3\eta) \\ &= \beta_2 \left(\frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} - \frac{1}{R_{0,2}} \right) + \eta \left(\frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} - 3 \right) > 0, \end{aligned}$$

and suppose that n is sufficiently large that $|\beta_2^{(n)} - \beta_2| < \eta$, *etc.*

Again, provided $t < \tau_\varepsilon^{(n)}$, we have that

$$\frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} < \frac{\bar{S}^{(n)}(t)}{\bar{N}^{(n)}(t)} < \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon},$$

and thus, if $Z^+(t)$ and $Z^-(t)$ are birth and death processes with birth and death rates

$$(\beta_2 + \eta) \frac{\bar{S}^{\star,1} + \varepsilon}{\bar{N}^{\star,1} - \varepsilon} \quad \text{and} \quad \delta + \alpha_2 + \gamma_2 - 3\eta$$

and

$$(\beta_2 - \eta) \frac{\bar{S}^{\star,1} - \varepsilon}{\bar{N}^{\star,1} + \varepsilon} \quad \text{and} \quad \delta + \alpha_2 + \gamma_2 + 3\eta$$

respectively, then both $Z^+(t)$ and $Z^-(t)$ are supercritical with Malthusian parameters μ_2^+ and μ_2^- , respectively, and

$$Z^-(t) < I_2^{(n)}(t) < Z^+(t)$$

stochastically for $t < \tau_\varepsilon^{(n)}$.

Now, set

$$\bar{q}_2 := \frac{\delta + \alpha_2 + \gamma_2 - 3\eta}{(\beta_2 + \eta) \frac{\bar{S}^{*,1} + \varepsilon}{N^{*,1} - \varepsilon}} < \underline{q}_2 := \frac{\delta + \alpha_2 + \gamma_2 + 3\eta}{(\beta_2 - \eta) \frac{\bar{S}^{*,1} - \varepsilon}{N^{*,1} + \varepsilon}} < 1.$$

Classical results for birth and death processes (e.g., [1]) tell us that $Z^+(t)$ and $Z^-(t)$ will hit 0 in finite time with probability $\bar{q}_2^{Z^+(0)}$ and \underline{q}_2 respectively, and will grow indefinitely otherwise, and, moreover, that there exist random variables \bar{W} and \underline{W} taking values on $[0, \infty)$ such that

$$\mathbb{P}\{\bar{W} = 0\} = \bar{q}_2^{Z^+(0)} \quad \text{and} \quad \mathbb{P}\{\underline{W} = 0\} = \underline{q}_2^{Z^-(0)}$$

and

$$e^{-\mu_2^+ t} Z^+(t) \rightarrow \bar{W} \quad \text{and} \quad e^{-\mu_2^- t} Z^-(t) \rightarrow \underline{W} \quad (\text{S.44})$$

both almost surely and in L^2 .

Now, as before, fix $\ln n \ll t_n \ll n$. Taking logarithms in (S.44), we see that for almost all $\omega \notin \{\bar{W} = 0\}$, we have

$$\frac{\ln Z^-(\omega, t_n)}{t_n} \rightarrow \mu_2^-.$$

Now, if $Z^-(t_n) \leq \varepsilon n$ then the right hand side converges to 0. Moreover, if $I_2^{(n)}(t) \leq \varepsilon n$, then necessarily $Z^-(t_n) \leq \varepsilon n$, and we conclude that

$$\limsup_{n \rightarrow \infty} \mathbb{P}\left\{t_n < \tau_\varepsilon^{(n)}\right\} \leq \mathbb{P}\{\underline{W} = 0\} = \underline{q}_2^{Z^-(0)}.$$

In particular, if we can establish that with high probability $\tau_\varepsilon^{(n)} = \tau_{\varepsilon,2}^{(n)}$, then this gives a lower bound on the probability that the novel strain invades, as the latter is the probability that $\tau_\varepsilon^{(n)}$ is finite.

Now,

$$\mathbb{P}\left\{\tau_\varepsilon^{(n)} < \tau_{\varepsilon,2}^{(n)}\right\} = \mathbb{P}\left\{\tau_\varepsilon^{(n)} < \tau_{\varepsilon,2}^{(n)}; \tau_\varepsilon^{(n)} < t_n\right\} + \mathbb{P}\left\{\tau_\varepsilon^{(n)} < \tau_{\varepsilon,2}^{(n)}; \tau_\varepsilon^{(n)} > t_n\right\}.$$

We have already established that the latter is bounded above by $\underline{q}_1^{Z^-(0)}$ as $n \rightarrow \infty$. The former follows almost exactly as the proof that $\mathbb{P}\left\{\tau_\varepsilon^{(n)} < t_n\right\} \rightarrow 0$ of the previous section; $\bar{\mathbf{E}}^{*,1}$ is now a hyperbolic fixed point rather than a stable fixed point, and A has a positive eigenvalue, but the stable manifold of $\bar{\mathbf{E}}^{*,1}$ coincides with the subset of \mathbb{R}^{d+2} with $x_1 = 0$. In particular, we may decompose

$$\mathbb{R}^{d+2} = E_S \oplus E_X,$$

where $E_S = \{x_1 = 0\}$ and E_X are A invariant subspaces, corresponding to the sum of the generalised eigenspaces for eigenvalues with negative and positive real parts respectively. We will write P_S and P_X for the corresponding projections (i.e., P_S has image E_S and kernel E_X , and oppositely for P_X), and note that both P_S and P_X commute with A . Proceeding exactly as previously, we define

$$\Xi^{(n)}(t) := \bar{\mathbf{E}}^{(n)}(t) - \bar{\mathbf{E}}^{*,1}$$

and let $\Xi_S^{(n)}(t) = P_S \Xi^{(n)}(t)$ denote its projection onto the stable subspace. Then, if

$$\tau_{\varepsilon,S}^{(n)} = \tau_{\varepsilon,0}^{(n)} \wedge \tau_{\varepsilon,2}^{(n)} \wedge \tau_{\varepsilon,3}^{(n)}$$

then

$$\mathbb{P} \left\{ \tau_{\varepsilon}^{(n)} < \tau_{\varepsilon,2}^{(n)}; t_n < \tau_{\varepsilon}^{(n)} \right\} = \mathbb{P} \left\{ \tau_{\varepsilon,S}^{(n)} = \tau_{\varepsilon}^{(n)}; t_n < \tau_{\varepsilon}^{(n)} \right\} = \mathbb{P} \left\{ \sup_{t \leq \tau_{\varepsilon}^{(n)} \wedge t_n} \|\Xi_S^{(n)}(t)\| \geq \varepsilon \right\}.$$

Applying the arguments of the previous section, using the equation for $\Xi_S^{(n)}(t)$ obtained by letting P_S act on both sides of (S.43), one obtains almost identically that

$$\mathbb{P} \left\{ \sup_{t \leq \tau_{\varepsilon}^{(n)} \wedge t_n} \|\Xi_S^{(n)}(t)\| \geq \varepsilon \right\} \rightarrow 0$$

as $n \rightarrow \infty$.

Now, note that for any $\omega \in \{\overline{W} = 0\}$, we must have $Z^+(\omega, t) < \varepsilon n$ for all t , for some sufficiently large n , so $\omega \in \{\tau_{\varepsilon}^{(n)} < \tau_{\varepsilon,2}^{(n)}\}$. Thus we have

$$\overline{q}_1^{Z^+(0)} = \mathbb{P}\{\overline{W} = 0\} \leq \liminf_{n \rightarrow \infty} \mathbb{P} \left\{ \tau_{\varepsilon}^{(n)} < \tau_{\varepsilon,1}^{(n)} \right\}$$

Finally, we notice that as $\varepsilon \rightarrow 0$ (and thus $\eta \rightarrow 0$ also,) both \overline{q}_2 and \underline{q}_2 approach

$$q_2 := \frac{\delta + \alpha_2 + \gamma_2}{\beta_2 \frac{\overline{S}^{*,1}}{N^{*,1}}} = \frac{R_{0,1}}{R_{0,2}}.$$

Since the choice of ε was arbitrary in our definition of invasion, we conclude that the probability of successful invasion of the new strain 1 is

$$1 - \left(\frac{R_{0,1}}{R_{0,2}} \right)^{I_2(0)}. \quad (\text{S.45})$$

Once this has happened, as we note above, Kurtz's deterministic approximation is applicable, and with high probability, the system will approach any arbitrarily small neighbourhood of the endemic fixed point for the new strain 1, at which point, by the argument above, the former resident strain, strain 2, goes extinct with probability approaching 1 as $n \rightarrow \infty$.

8.3 Proof of Proposition 10

We start with a lemma:

Lemma 5. *Suppose that $\alpha \in [\alpha_{\min}, \alpha_{\max}]$. Then,*

$$\int_{\alpha_{\min}}^{\alpha_{\max}} K_{\varepsilon}(\alpha, \alpha' - \alpha)(\alpha' - \alpha)^k d\alpha' = \begin{cases} \varepsilon^k \int_0^{\infty} K(\alpha, z)z^k dz + o(\varepsilon^k) & \text{if } \alpha = \alpha_{\min}, \\ \varepsilon^k \int_{-\infty}^0 K(\alpha, z)z^k dz + o(\varepsilon^k) & \text{if } \alpha = \alpha_{\max}, \text{ and} \\ \int_{-\infty}^{\infty} K(\alpha, z)z^k dz + o(\varepsilon^k) & \text{if } \alpha_{\min} < \alpha < \alpha_{\max} \end{cases}$$

Proof. Making a change of variable, we have

$$\begin{aligned} \int_{\alpha_{\min}}^{\alpha_{\max}} K_{\varepsilon}(\alpha, \alpha' - \alpha)(\alpha' - \alpha)^k d\alpha' &= \int_{\alpha_{\min}}^{\alpha_{\max}} K\left(\alpha, \frac{\alpha' - \alpha}{\varepsilon}\right) (\alpha' - \alpha)^k \frac{d\alpha'}{\varepsilon} \\ &= \varepsilon^k \int_{\frac{\alpha_{\min} - \alpha}{\varepsilon}}^{\frac{\alpha_{\max} - \alpha}{\varepsilon}} K(\alpha, z)z^k dz, \end{aligned}$$

and $\frac{\alpha_{\min} - \alpha}{\varepsilon}$ and $\frac{\alpha_{\max} - \alpha}{\varepsilon}$ tend to $-\infty$ and $+\infty$ respectively as $\varepsilon \rightarrow 0$, unless $\alpha = \alpha_{\min}$ or $\alpha = \alpha_{\max}$. By assumption, the integrals converge, so the tails must vanish as $\varepsilon \rightarrow 0$. \square

Taylor expanding $s(\alpha, \alpha')(f(\alpha') - f(\alpha))$ in α' about α , we have

$$\begin{aligned} \hat{\mathcal{L}}_{\varepsilon} f(\alpha) &= \frac{n}{\varepsilon^2} \eta(\alpha) I_{\text{eq}}(\alpha) \int_{\alpha_{\min}}^{\alpha_{\max}} K_{\varepsilon}(\alpha, \alpha' - \alpha) \left[s(\alpha, \alpha)(\alpha - \alpha')f'(\alpha) \right. \\ &\quad \left. + \left(\frac{\partial s}{\partial \alpha'}(\alpha, \alpha)f'(\alpha) + \frac{1}{2}s(\alpha, \alpha')f''(\alpha) \right) (\alpha - \alpha')^2 + \dots \right] d\alpha' \end{aligned}$$

From the Lemma, we have

$$\begin{aligned} \hat{\mathcal{L}}_{\varepsilon} f(\alpha_{\min}) &= \frac{n}{\varepsilon} \eta(\alpha_{\min}) I_{\text{eq}}(\alpha_{\min}) s(\alpha_{\min}, \alpha_{\min}) f'(\alpha_{\min}) \int_0^{\infty} K(\alpha_{\min}, z)z dz \\ &\quad + \left(n\eta(\alpha_{\min}) I_{\text{eq}}(\alpha_{\min}) \frac{\partial s}{\partial \alpha'}(\alpha_{\min}, \alpha_{\min}) f'(\alpha_{\min}) \right. \\ &\quad \left. + \frac{n}{2} \eta(\alpha_{\min}) I_{\text{eq}}(\alpha_{\min}) s(\alpha_{\min}, \alpha_{\min}) f''(\alpha_{\min}) \right) \int_0^{\infty} K(\alpha_{\min}, z)z^2 dz + o(\varepsilon), \\ \hat{\mathcal{L}}_{\varepsilon} f(\alpha_{\max}) &= \frac{n}{\varepsilon} \eta(\alpha_{\max}) I_{\text{eq}}(\alpha_{\max}) s(\alpha_{\max}, \alpha_{\max}) f'(\alpha_{\max}) \int_{-\infty}^0 K(\alpha_{\max}, z)z dz \\ &\quad + \left(n\eta(\alpha_{\max}) I_{\text{eq}}(\alpha_{\max}) \frac{\partial s}{\partial \alpha'}(\alpha_{\max}, \alpha_{\max}) f'(\alpha_{\max}) \right. \\ &\quad \left. + \frac{n}{2} \eta(\alpha_{\max}) I_{\text{eq}}(\alpha_{\max}) s(\alpha_{\max}, \alpha_{\max}) f''(\alpha_{\max}) \right) \int_{-\infty}^0 K(\alpha_{\max}, z)z^2 dz + o(\varepsilon), \end{aligned}$$

and

$$\hat{\mathcal{L}}_{\varepsilon} f(\alpha) = n\eta(\alpha) I_{\text{eq}}(\alpha) \nu(\alpha) \frac{\partial s}{\partial \alpha'}(\alpha, \alpha) f'(\alpha) + \frac{n}{2} \eta(\alpha) I_{\text{eq}}(\alpha) \nu(\alpha) s(\alpha, \alpha) f''(\alpha) + o(\varepsilon).$$

for $\alpha \in (\alpha_{\min}, \alpha_{\max})$.

We first note that to have a well-posed (finite) limit as $\varepsilon \rightarrow 0$, we must have $f'(\alpha_{\min}) = f'(\alpha_{\max}) = 0$, so that the first term, which is otherwise of order $\mathcal{O}(\varepsilon^{-1})$, vanishes.

To deal with the discontinuities at α_{\min} and α_{\max} , suppose we have f such that $f'(\alpha_{\min}) = f'(\alpha_{\max}) = 0$. Choose $h \in C^2[\alpha_{\min}, \alpha_{\max}]$ such that $h'(\alpha_{\min}) = \frac{1}{2}f''(\alpha_{\min})\frac{\int_{-\infty}^0 K(\alpha_{\min}, z)z^2 dz}{\int_0^{\infty} K(\alpha_{\min}, z)z dz}$ and $h'(\alpha_{\max}) = \frac{1}{2}f''(\alpha_{\max})\frac{\int_0^{\infty} K(\alpha_{\max}, z)z^2 dz}{\int_{-\infty}^0 K(\alpha_{\max}, z)z dz}$, and let $f_\varepsilon = f + \varepsilon h$, then

$$\begin{aligned}\hat{\mathcal{L}}_\varepsilon f(\alpha_{\min}) &= \frac{n}{2}\eta(\alpha_{\min})I_{\text{eq}}(\alpha_{\min})\nu(\alpha_{\min})s(\alpha_{\min}, \alpha_{\min})f''(\alpha_{\min}) + o(\varepsilon), \\ \hat{\mathcal{L}}_\varepsilon f(\alpha_{\max}) &= \frac{n}{2}\eta(\alpha_{\max})I_{\text{eq}}(\alpha_{\max})\nu(\alpha_{\max})s(\alpha_{\max}, \alpha_{\max})f''(\alpha_{\max}) + o(\varepsilon),\end{aligned}$$

so that $\hat{\mathcal{L}}_\varepsilon f_\varepsilon(\alpha) \rightarrow \hat{\mathcal{L}}f(\alpha)$, $f_\varepsilon(\alpha) \rightarrow f(\alpha)$, for all $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, and weak convergence follows by Corollary 8.7 in Chapter 4 of [14]

References

- [1] K. B. Athreya and P. Ney. *Branching processes*. Springer-Verlag, Berlin, 1972.
- [2] F. Ball. Coupling methods in epidemic theory. In D. Mollison, editor, *Epidemic Models: Their Structure and Relation to Data*, pages 34–52. Cambridge University Press, 1995.
- [3] M. S. Bartlett. *Stochastic processes*. Cambridge University Press, 1955.
- [4] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, New York, 1978.
- [5] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.
- [6] H. J. Bremermann and H. R. Thieme. A competitive exclusion principle for pathogen virulence. *Journal of Mathematical Biology*, 27(2):179–190, 1989.
- [7] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.*, 116(8):1127–1160, 2006.
- [8] N. Champagnat, R. Ferrière, and G. Ben Arous. The canonical equation of adaptive dynamics: a mathematical view. *Selection*, 2:71–81, 2001.
- [9] N. Champagnat and A. Lambert. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *Ann. Appl. Prob.*, 17(1):102–155, 2007.
- [10] U. Dieckmann and R. Law. The dynamical theory of coevolution: A derivation from stochastic ecological processes. *J. Math. Biol.*, 34:579–612, 1996.
- [11] R. Durrett. *Probability Models for DNA Sequence Evolution*. Springer, New York, 2nd edition, 2009.
- [12] A. Erdélyi. *Asymptotic expansions*. Dover Publications, 1956.

- [13] A. M. Etheridge. Some mathematical models from population genetics. In J. Picard, editor, *École d'Été de Probabilités de Saint-Flour, XXXIX—2009*, volume 2012 of *Lecture Notes in Math.*, Berlin Heidelberg, 2011. Springer.
- [14] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley and Sons, New York, 1986.
- [15] W. J. Ewens. *Mathematical Population Genetics*. Springer-Verlag,, Berlin/New York, 1979.
- [16] W. Feller. Diffusion processes in one dimension. *Trans. Amer. Math. Soc.*, 77(1):1–31, 1954.
- [17] E. J. Hinch. *Perturbation Methods*. Cambridge University Press, Cambridge, 1991.
- [18] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, New York, 1974.
- [19] G. S. Katzenberger. Solutions of a stochastic differential equation forced onto a manifold by a large drift. *Ann. Probab.*, 19(4):1587–1628, 1991.
- [20] J. Kevorkian and J. D. Cole. *Multiple Scale and Singular Perturbation Methods*. Springer-Verlag New York Inc., New York, 1996.
- [21] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Prob.*, 7(1):pp. 49–58, 1970.
- [22] T. G. Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Prob.*, 8(2):344–356, 1971.
- [23] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes Appl.*, 6(3):223 – 240, 1978.
- [24] T. G. Kurtz. *Approximation of Population Processes*, volume 36. Society for Industrial and Applied Mathematics, Philadelphia, 1981.
- [25] J. A. J. Metz, R. M. Nisbet, and S. A. H. Geritz. How should we define ‘fitness’ for general ecological scenarios? *Trends Ecol. Evol.*, 7(6):198–202, 1992.
- [26] S. P. Otto and T. Day. *A biologist’s guide to mathematical modeling in ecology and evolution*. Princeton University Press, 2007.
- [27] T. L. Parsons and T. Rogers. Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology. *J. Phys. A: Math. Theor.*, 50(41):415601, 2017.
- [28] P. K. Pollett. On a model for interference between searching insect parasites. *J. Austral. Math. Soc. Ser. B*, 32(02):133–150, 1990.
- [29] P. E. Protter. *Stochastic Integration and Differential Equations – a new approach*. Springer, Berlin, 2nd edition, 2004.
- [30] G. Teschl. *Ordinary differential equations and dynamical systems*, volume v. 140 of *Graduate studies in mathematics*. American Mathematical Society, Providence, R.I., 2012.

- [31] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, San Diego, Oxford, London, 1992.
- [32] C. Vargas-De-León. On the global stability of SIS, SIR and SIRS epidemic models with standard incidence. *Chaos, Solitons & Fractals*, 44(12):1106–1110, 2011.
- [33] E. P. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Commun. Pure Appl. Math*, 13(1):1–14, 1960.