



**HAL**  
open science

# Beneficial mutation-selection dynamics in finite asexual populations: a free boundary approach

Lionel Roques, Jimmy Garnier, Guillaume Martin

► **To cite this version:**

Lionel Roques, Jimmy Garnier, Guillaume Martin. Beneficial mutation-selection dynamics in finite asexual populations: a free boundary approach. *Scientific Reports*, 2017, 7, 10.1038/s41598-017-17212-5 . hal-02620234

**HAL Id: hal-02620234**

**<https://hal.inrae.fr/hal-02620234v1>**

Submitted on 25 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# SCIENTIFIC REPORTS



OPEN

## Beneficial mutation-selection dynamics in finite asexual populations: a free boundary approach

Lionel Roques<sup>1</sup>, Jimmy Garnier<sup>2</sup> & Guillaume Martin<sup>3</sup>

Using a free boundary approach based on an analogy with ice melting models, we propose a deterministic PDE framework to describe the dynamics of fitness distributions in the presence of beneficial mutations with non-epistatic effects on fitness. Contrarily to most approaches based on deterministic models, our framework does not rely on an infinite population size assumption, and successfully captures the transient as well as the long time dynamics of fitness distributions. In particular, consistently with stochastic individual-based approaches or stochastic PDE approaches, it leads to a constant asymptotic rate of adaptation at large times, that most deterministic approaches failed to describe. We derive analytic formulas for the asymptotic rate of adaptation and the full asymptotic distribution of fitness. These formulas depend explicitly on the population size, and are shown to be accurate for a wide range of population sizes and mutation rates, compared to individual-based simulations. Although we were not able to derive an analytic description for the transient dynamics, numerical computations lead to accurate predictions and are computationally efficient compared to stochastic simulations. These computations show that the fitness distribution converges towards a travelling wave with constant speed, and whose profile can be computed analytically.

Adaptation under mutation selection and drift is a central process of evolutionary biology. With the advent of experimental evolution and cancer studies, much focus on the dynamics of adaptation has shifted from sexuals to asexuals. While the former remains intensely studied and has equally important applications, direct studies of adaptation as it unfolds is often restricted to asexual model species (reviewed in<sup>1</sup>). This new focus has shed light on the complex interplay of drift, selection and mutation in asexuals, especially when multiple lineages co-segregate and compete for ultimate fixation, i.e. with clonal interference (reviewed in<sup>2</sup>).

The simplest microscopic model describing this interplay of drift selection and mutation considers a set of  $K$  haploid genotypes, in a population of constant census size  $N$ . Each genotype  $i \in [1, K]$  has Malthusian fitness  $m_i$ , which is the expected rate of change  $m_i$  (per unit time) of the log of the genotype's census size. The population size is maintained constant at size  $N$  by randomly sampling individuals after reproduction (culling of the population). There are continuous time and discrete time versions of this model (detailed e.g. in<sup>3</sup>): in the continuous time version (Moran Model), reproduction and culling occur asynchronously at exponentially distributed time intervals, in the discrete time model (Wright-Fisher model, used in our simulations), reproduction and culling of the whole population occur synchronously every generation. In both case, mutations arise as a Poisson process with rate  $U$  per unit time per individual, independently of the genotype. Single mutations produce a mutant offspring which fitness is the sum of the parental fitness and a random deviate, which distribution is independent of the genotype (we assume it has finite moments, as is classically considered). Multiple mutations have additive effect on the offspring. Thus, this classic additive model ignores epistasis for fitness (e.g. discussed in<sup>4-6</sup>). A classic 'textbook' result (e.g.<sup>7</sup>) states that the dynamics of genotype frequencies, in the discrete time model (Wright-Fisher), are well approximated by the dynamics of the continuous time version (Moran model), whenever fitnesses do not vary too much among genotypes, at any time. Overall, for both models, the distribution of Malthusian fitness within

<sup>1</sup>BioSP, INRA, 84914, Avignon, France. <sup>2</sup>CNRS – Université Savoie Mont-Blanc, LAMA, F-73000, Chambéry, France.

<sup>3</sup>CNRS, ISEM, UMR 5554, 34095, Montpellier, France. Correspondence and requests for materials should be addressed to L.R. (email: [lionel.roques@inra.fr](mailto:lionel.roques@inra.fr))

the population  $p(t, \cdot)$  admits the same limit as  $N \rightarrow +\infty$  and in continuous time. This limit is characterized by the following continuous-time nonlocal integro-differential equation (see e.g.<sup>2,8,9</sup>):

$$\partial_t p = U(J^*p - p) + (m - \bar{m}(t))p, \quad t > 0, m \in \mathbb{R} \text{ with } \bar{m}(t) = \int_{\mathbb{R}} s p(t, s) ds, \quad (1)$$

where the sign  $*$  stands for the convolution product:

$$J^*p(t, m) = \int_{\mathbb{R}} J(m - s) p(t, s) ds. \quad (2)$$

Assuming no epistasis, the distribution of mutation effects on fitness is constant and is described by the mutation kernel  $J$  (see<sup>4-6</sup> for other approaches which account for the effects of epistasis). Thus, the integral term  $U(J^*p - p)$  describes the mutation effects. The term  $(m - \bar{m}(t))p$  describes selection with  $\bar{m}$  giving the mean fitness in the population at time  $t$ .

Beneficial mutation-selection dynamics in asexual populations are characterized by a constant asymptotic rate of adaptation  $v_\infty$  at large times if the effects of mutation on fitness are non-epistatic, i.e., if these effects are additive as we assume here. This behaviour arises, over large times, in numerical simulations of the microscopic model described above (finite  $N$ ). However, the limit dynamical system described by the integro-differential Eq. (1) does not capture this behaviour (as detailed below). The asymptotic rate of adaptation  $v_\infty$  depends, among other things, on the mutation rate  $U$  and the population size  $N$ . Current theory gives approximate analytic expressions for the expected asymptotic rate of adaptation, depending on these parameters, with different conclusions and approximations for different parameter ranges (e.g.<sup>10-14</sup>). However, they only provide predictions for the long term behaviour of the system, once it reaches a steady dynamical state with a constant rate of adaptation, which typically take several thousand generations of transient “burn in” before they are reached, thus also limiting their application to empirical dynamics. On the other hand, deterministic models of the form (1), based on an infinite population size assumption (e.g.<sup>15,16</sup>) provide insight on transient behaviours (i.e., before drift starts to act<sup>6</sup>) but fail to predict a constant asymptotic rate of adaptation. A deterministic theory that would describe the transient dynamics of the fitness distribution and its convergence towards the correct stationary regime is therefore still missing.

When the mutation kernel  $J$  contains some beneficial mutations, due to the infinite population size assumption, the approach (1) becomes unrealistic as  $t$  becomes large<sup>16</sup>: with this approach, the rate of adaptation  $v(t) = \bar{m}'(t)$  increases exponentially fast, and may even blow-up in finite time if the kernel  $J(s)$  decays exponentially or slower as  $s \rightarrow +\infty$ . To overcome this problem, Martin and Roques<sup>6</sup> proposed a phenomenological approach which consisted in connecting the transient dynamics obtained with (1) with the asymptotic rate of adaptation  $v_\infty$ . In that respect, they computed the time  $\tau$  such that the rate of adaptation given by model (1) reaches some a priori asymptotic rate of adaptation:  $\bar{m}'(\tau) = v_\infty$ . Then, they assumed the fitness distribution to behave as a traveling wave with constant profile for  $t \geq \tau$ :  $p(t, m) = p(\tau, m - v_\infty(t - \tau))$ . This approach is accurate, but not entirely satisfying as it provides no information on the asymptotic rate of adaptation  $v_\infty$ , which has to be estimated numerically using individual-based simulations, or given by other theories.

When mutations with large effects are rare, or mutation rates sufficiently high—see<sup>14</sup> and chapter II, part 1.2 in<sup>17</sup>, we can use the following diffusion approximation of the mutation effects

$$U(J^*p - p) \simeq D\partial_{mm}^2 p - B\partial_m p, \quad (3)$$

with  $D = M_2 U/2$ ,  $B = M_1 U$  and  $M_1, M_2$  the first and second moments of  $J$ . A main advantage of this approximation is that diffusion operators are generally more tractable than convolutions as we see in the sequel. The dynamics of  $p(t, \cdot)$  can then be approached by the nonlocal PDE:

$$\partial_t p = D\partial_{mm}^2 p - B\partial_m p + (m - \bar{m}(t))p, \quad t > 0, m \in \mathbb{R} \text{ with } \bar{m}(t) = \int_{\mathbb{R}} s p(t, s) ds. \quad (4)$$

This model, with  $B = 0$  (which can be assumed without loss of generality, by considering the above equation in a frame moving with speed  $B$  to the right) has been exhaustively investigated by Alfaro and Carles<sup>15</sup> who derived an analytical characterization of the dynamics of fitness. Their approach showed that, again, the model is unrealistic as the rate of adaptation  $v(t) = \bar{m}'(t)$  ultimately becomes infinite.

As long as beneficial mutations occurs, the deterministic models (1) and (4) fail to describe the asymptotic rate of adaptation. This failure is related to the infinite speed of propagation of the support of  $p(t, m)$  which holds for the diffusion operator  $D\partial_{mm}^2 p$  or the integro-differential operator  $U(J^*p - p)$  (as long as  $J$  contains some beneficial mutations) as a consequence of the strong maximum principle (see e.g.<sup>18</sup> in the diffusion case). Because of this propagation property, positive concentrations of individuals with arbitrarily large fitnesses are created at each time, and tend to increase by selection, pulling the distribution towards larger fitnesses.

Traditionally, this unrealistic feature of many PDE models is compensated by introducing a cut-off<sup>19</sup>, i.e., a threshold value of concentration  $p_c > 0$  below which individuals cannot grow, which means here that they cannot be selected. Tsimring *et al.*<sup>20</sup> adopted this cut-off approach, replacing the selection term  $(m - \bar{m}(t))p$  by  $\chi(p - p_c)(m - \bar{m}(t))p$ , where  $\chi(\cdot)$  is the Heaviside step function. Using this approach, they focused on the large-time dynamics of fitness. They showed in particular the existence of a traveling wave solution with constant speed. The wave speed can then be computed numerically by solving a system of transcendental equations. However, as the speed depends on the threshold  $p_c$ , it gives no quantitative information on the actual expected

asymptotic rate of adaptation and on the transient dynamics in a population with given mutational characteristics.

Alternative approaches such as stochastic integro-differential or diffusion equations have also been considered by several authors<sup>9,21,22</sup> to take into account the effects of genetic drift arising from the finite size of the population. These equations are of the form (1) or (4), with an additional term  $\sqrt{p(t, m)}\eta(t, m)$  where  $\eta(t, m)$  is Gaussian white noise. Using a stochastic diffusion equation of this form, and some results in<sup>23</sup>, Neher and Hallatschek<sup>14</sup> established a simple formula for the asymptotic rate of adaptation  $v_\infty$ , depending on the population size  $N$ .

The aim of our work is to propose a deterministic framework to follow the transient and large-time dynamics of fitness distributions in finite populations, in the presence of beneficial (and deleterious) mutations. Contrarily to cut-off approaches, this new framework does not rely on the assumption that highly fit mutants do not grow if their concentration is small: here, the best fitness class remains finite at all times. This approach is inspired by the physics of ice melting, where a free boundary separates the liquid and solid phases, which is known as “the Stefan problem”<sup>24</sup>. In our case, the free boundary is a moving interface corresponding to the position of the fittest class at time  $t$ . Beyond this interface, unlike what happens in the standard diffusion case or with cut-off approaches, the distribution is exactly 0.

**Free boundary approach.** In this section, we construct a model which captures some properties of the microscopic models mentioned above, while preserving the PDE formalism of the limit system (4). In particular the support of the distribution is bounded to the right and the rate of adaptation remains finite and depends on the population size  $N$ . In that respect, we replace the problem (4) by the free boundary problem:

$$\begin{cases} \partial_t p &= D\partial_m^2 p + (m - X(t))p, \quad t > 0, m \in (-\infty, s(t)) \\ p(t, s(t)) &= 0, \quad t > 0 \\ s'(t) &= -\mu D \partial_m p(t, s(t)), \quad t > 0. \end{cases} \tag{5}$$

The free boundary  $s(t)$  separates the region where the distribution is positive with the portion of the fitness space  $[s(t), +\infty)$  where the distribution is 0. This free boundary propagates with the speed  $s'(t)$ . In the absence of selection (i.e., if  $(m - X(t))u$  is removed), the Eq. (5) corresponds to the Stefan problem<sup>24</sup> that we mentioned in the Introduction. Note that this model admits an extra-parameter  $\mu$  compared to (4). This parameter describes how the speed of the interface  $s'(t)$  and the slope of the distribution at this interface are connected. Since  $p$  is a probability distribution, its mass should equal 1 at all time, leading to the following definition of  $X(t)$ :

$$X(t) = \int_{-\infty}^{s(t)} y p(t, y) dy - \frac{1}{\mu} s'(t) = \bar{m}(t) - \frac{1}{\mu} s'(t), \quad t > 0. \tag{6}$$

To our knowledge, there is no generic theory that gives existence and uniqueness of the solution of the Cauchy problem (5)–(6) with initial condition  $p_0$  (but see<sup>25</sup> for some results on closely related equations, without the nonlocal term  $X(t)$ ). We show in Section 2.1 that for each value of the parameter  $\mu > 0$  the problem (5)–(6) admits a traveling wave solution, whose profile and speed depend on  $\mu$ . We then derive in Section 2.2 an additional relationship between the parameters in (5) and the population size  $N$ . In particular, we derive a formula for the parameter  $\mu$ —and therefore the asymptotic rate of adaptation—depending on the population size  $N$ . We compute the corresponding asymptotic rate of adaptation.

**Travelling wave solution.** We look for a travelling wave solution with constant speed  $v$ , i.e., a solution of the form  $p(t, m) = \phi(m - vt)$  and  $s(t) = vt$  with  $\phi > 0$  on  $(-\infty, 0)$  and  $\int_{-\infty}^0 \phi(y) dy = 1$ . Plugging this solution into (5)–(6) gives

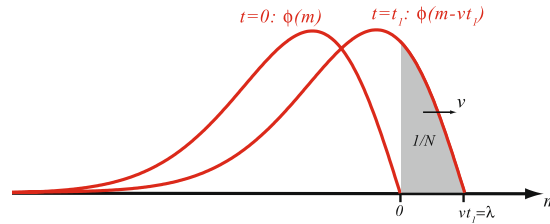
$$\begin{cases} D\phi''(z) + v\phi'(z) + \left( z - \int_{-\infty}^0 y\phi(y) dy + \frac{1}{\mu}v \right)\phi(z) = 0, \quad z \in (-\infty, 0), \\ \phi(0) = 0 \text{ and } \phi'(0) = -\frac{v}{\mu D}. \end{cases} \tag{7}$$

Then we can compute an explicit solution of this equation using the Airy function  $\text{Ai}$  which, by definition, solves the ODE  $\text{Ai}''(z) - z\text{Ai}(z) = 0$ , and is bounded in  $\mathbb{R}$ . The solution of (7) is then given by

$$\phi(z) = \frac{v}{\mu D^{2/3} \text{Ai}'(-z_0)} e^{-\frac{vz}{2D}} \text{Ai}(-z_0 - z/D^{1/3}), \tag{8}$$

where  $-z_0 \approx -2.34$  is the largest zero of the Airy function ( $\text{Ai}(-z_0) = 0$ ). Additionally, we have  $\bar{z} = \int_{-\infty}^0 y\phi(y) dy = -D^{1/3}z_0 + \frac{v}{\mu} - \frac{v^2}{4D}$ . The function  $\phi$  defined by (8) solves the problem (7), it is positive on  $(-\infty, 0)$  and it is of mass 1 if and only if

$$\mu = \frac{v}{D^{1/3} \text{Ai}'(-z_0)} \int_0^\infty e^{\frac{vz}{2D}} \text{Ai}(z - z_0) dz. \tag{9}$$



**Figure 1.** Computation of the time  $t_1$  required to establish a beneficial mutation beyond the initial best fitness class  $m = 0$  at time  $t$  ( $t = 0$  here). During the period  $(t, t + t_1)$ , the wave travels to the right with speed  $v$ . The time  $t_1$  is such that the shaded region has mass  $1/N$ , i.e., the expected number of individuals with positive fitness is equal to 1.

As the above function  $v \mapsto \mu$  is strictly increasing, equal to 0 at  $v = 0$  and converges to  $+\infty$  as  $v \rightarrow +\infty$ , this shows that for any  $\mu > 0$ , there is a unique  $v > 0$  satisfying (9). Altogether, this shows that, for any  $\mu > 0$ , the problem (5)–(6) admits a unique positive travelling wave solution  $p(t, m) = \phi(m - vt)$  satisfying  $\int_{-\infty}^0 \phi(y) dy = 1$ .

**Effect of the population size  $N$  on the model parameters.** Assume that the asymptotic rate of adaptation  $v_\infty$  has been reached, with a solution of (5)–(6) given by the travelling wave  $p(t, m) = \phi(m - vt)$  with speed  $v = v_\infty$ . In order to derive an additional relationship between  $v$ ,  $\mu$  and the population size  $N$ , we first compute the expected time  $t_1$  required to establish a beneficial mutation beyond the initial best fitness class at time  $t$ . Then, we compute the speed as the ratio between the expected increase  $\lambda$  in the best fitness class due to this beneficial mutation and the time  $t_1$  (Fig. 1).

For the sake of simplicity, and without loss of generality, we assume that  $t = 0$ , so that the best fitness class (i.e., the upper bound of the support of the distribution  $p(0, \cdot)$ ) is  $m = 0$ . At each time  $t$ , the whole wave  $p(t, m)$  is of mass 1, and corresponds to  $N$  individuals. As the wave moves with speed  $v$ , the time  $t_1$  is such that the mass between  $z = 0$  and  $z = vt_1$  is equal to  $1/N$ . In other terms:

$$N \int_0^{vt_1} \phi(m - vt_1) dm = N \int_{-vt_1}^0 \phi(z) dz = 1. \tag{10}$$

Approaching  $\phi(z)$  by  $z\phi'(0)$  in a neighborhood of 0, we simply get

$$N \int_{-vt_1}^0 -\frac{v}{\mu D} z dz = 1, \tag{11}$$

leading to the following expression for  $t_1$ :

$$t_1 = \frac{1}{v} \sqrt{\frac{2\mu D}{Nv}}. \tag{12}$$

Second, we compute the size  $\lambda$  of the fitness step implied by the establishment of the beneficial mutation, given that it has occurred. To do so, we use the fact that we are assuming a regime of pervasive clonal interference (many mutations of small effect co-segregate), which is already required for the diffusion approximation used to describe mutation. In this context (detailed in<sup>21</sup>), the backgrounds that produce those mutant lineages destined to invade in the future, tend to be tightly packed at the very rightmost edge of the fitness distribution. In our context, this means that (i) we can assume that the fitness step is due to a beneficial mutation from a background standing at the free boundary, and (ii) the resulting mutants are sufficiently fit and close to one another that their probability of avoiding initial loss due to drift is equal among lineages. Under these two assumptions, the expected fitness increase by these mutants is simply the average fitness effect of beneficial mutations from a parent in the current best fitness class ( $m = 0$ ), so that

$$\lambda = \int_0^\infty sJ(s)ds / \left( \int_0^\infty J(s)ds \right), \tag{13}$$

( $\lambda \approx (M_2)^{1/2}$  under the diffusion approximation). We deduce that the asymptotic rate of adaptation  $v = v_\infty$  should also solve  $v = \frac{\lambda}{t_1}$ . Using formula (12), we conclude that:

$$v = \frac{2D\mu}{\lambda^2 N}. \tag{14}$$

Third, we plug the formula (14) into (9), leading to:

$$\int_0^\infty e^{\frac{\mu D^{1/3}}{\lambda^2 N} z} \text{Ai}(z - z_0) dz = \frac{\lambda^2 N \text{Ai}'(-z_0)}{2D^{2/3}}. \tag{15}$$

As the left-hand side in this expression is an increasing function of  $\mu$ , it can easily be solved numerically. Setting  $\alpha = \frac{\mu D^{1/3}}{\lambda^2 N}$ , and approaching the integral in the left-hand side of (15) by  $\exp(\alpha z_0 + \alpha^3/3)$ , we also get an approximate analytic expression for  $\mu$ : setting  $\beta = \frac{\lambda^2 N \lambda^i (-z_0)}{2D^{2/3}}$ , and  $\gamma = 12 \ln \beta + 4\sqrt{z_0^3 + 9(\ln \beta)^2}$  we get:

$$\mu = \frac{\lambda^2 N}{D^{1/3}} \left[ \frac{1}{2} \gamma^{1/3} - 2z_0 \gamma^{-1/3} \right] \approx \frac{\lambda^2 N \gamma^{1/3}}{2D^{1/3}}, \tag{16}$$

where the last approximation assumes that  $\beta \gg 1$ , i.e.,  $(\lambda/U)^{2/3} N \gg 1$ . Using Eq. (14), we get that the corresponding formula for the asymptotic rate of adaptation is:

$$v_\infty = v = (\gamma D^2)^{1/3}. \tag{17}$$

This shows that the asymptotic rate of adaptation tends to increase with  $N$ , like  $(\ln N)^{1/3}$ . This is consistent with the findings of Neher and Hallatschek<sup>14</sup>; in the framework of stochastic PDEs, they obtained the following formula:

$$v_{NH} = D^{2/3} [24 \ln(ND^{1/3})]^{1/3}. \tag{18}$$

As a byproduct of formula (17), we can compute the expected asymptotic variance in fitness  $V_\infty$  within a population. Multiplying by  $z$  the Eq. (7) satisfied by  $\phi$  and integrating by parts, we get:

$$V_\infty = \int_{-\infty}^0 z^2 \phi(z) dz - \bar{z}^2 = v_\infty - \frac{v_\infty \bar{z}}{\mu}. \tag{19}$$

Thus,  $V_\infty$  can be computed explicitly in terms of the model parameters  $\lambda, U, N, D$ , using (17) and the relationship  $v_\infty/\mu = 2D/(\lambda^2 N)$  which follows from (14).

### Validation

To check the validity of our approach, we used an individual-based, discrete time Wright-Fisher model of genetic drift, selection and mutation as a benchmark. In this model, each individual  $i = 1, \dots, N$  has Malthusian fitness  $m_i$ , with corresponding Darwinian fitness  $e^{m_i}$  (discrete time counterpart of the Malthusian fitness). At each generation,  $N$  individuals are sampled with replacement and according to their Darwinian fitnesses. This is modelled by a multinomial experiment with  $N$  trials and event probabilities  $(e^{m_1}, \dots, e^{m_N})/\sum_{i=1, \dots, N} e^{m_i}$ . Then, in this discrete time approach, the mutation step occurs simultaneously in all the individuals. It is simulated by randomly drawing, for each individual, a Poisson number of mutations, with a rate  $U$ . The effect of a unique mutation on fitness was drawn according to the distribution  $J$ , multiple mutations having additive effects on the fitness of the offspring. We considered kernels for which the diffusion approximation  $U(J^* p - p) \simeq D \partial_m^2 p$  can be made:  $J(s) \propto \exp(-|s/(\sigma\sqrt{2})|^\beta)$ , with  $\beta = 2$  (Gaussian distribution, as in<sup>14</sup>) and  $\beta = 10$  (as in<sup>13</sup>), with small variance  $2\sigma^2 \Gamma(3/\beta)/\Gamma(1/\beta) = 10^{-4}$  (same value as in<sup>14</sup>).

For mutation rates  $U$  ranging from  $10^{-4}$  to  $2 \cdot 10^{-1}$ , and for population sizes ranging from  $N = 10^2$  to  $N = 10^7$ , we simulated 100 replicate trajectories of the individual-based model, for  $t \in (0, 1000)$ . We then followed the empirical distribution of fitness in each population. This allowed us to compute the mean fitness  $\bar{m}_{num}(t)$  within each population, and its average value  $\langle \bar{m}_{num}(t) \rangle$  among the 100 replicates.

**Stationary rate of adaptation and distribution of fitness.** The stationary rate of adaptation  $v_{num}$  was estimated by computing the slope given by a linear fit of  $\langle \bar{m}_{num}(t) \rangle$  for  $t \in (700, 1000)$ . The accuracy of the analytical approximation  $v_\infty$  in Eq. (17) was checked by computing the ratio  $v_\infty/v_{num}$ . In the case  $\beta = 2$  (Gaussian mutation kernel), our formula was compared to that in<sup>14</sup>, see Eq. (18). In the case  $\beta = 10$ , our formula was compared to that in<sup>13</sup>. More precisely, for beneficial mutation kernels of the form:

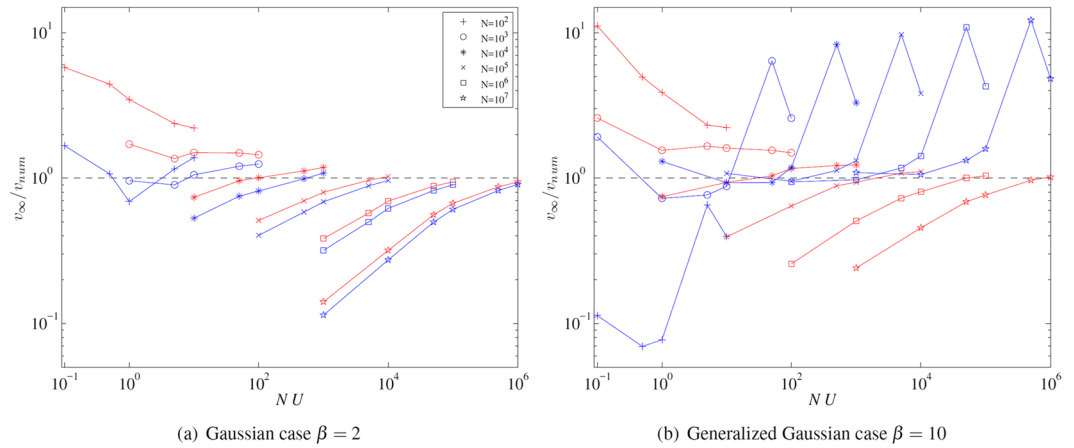
$$\rho(x) = \frac{1}{\sigma} \frac{e^{-(s/\sigma)^\beta}}{\Gamma(1 + 1/\beta)}, \tag{20}$$

with  $\beta \gg 1$  the following approximation was derived in<sup>13</sup> (formula (21) in<sup>13</sup>):

$$v_G = \frac{2\sigma^2 \ln(N\sigma\sqrt{\ln(N\sigma)})}{\ln\left(\frac{\sigma\sqrt{\beta}}{U_b} \sqrt{\ln\left(\frac{\sigma\sqrt{\beta}}{U_b}\right)}\right)^2} \left( 1 + \frac{\ln\left(\frac{\sigma\sqrt{\beta}}{U_b}\right)}{4 \ln(N\sigma)} \right)^{-2}, \tag{21}$$

where  $U_b = U/2$  is the rate of beneficial mutations.

The results are presented in Fig. 2(a,b). In the case  $\beta = 2$  (Fig. 2a) we observe that our formula (17) and the formula (18) of Neher and Hallatschek<sup>14</sup> give accurate results in the sense that  $v_\infty, v_{NH}$  and  $v_{num}$  have the same order of magnitude, for all values of  $N, U$ . The asymptotic rate of adaptation  $v_{num}$  is overestimated for small  $NU$ , while it tends to be underestimated for larger  $NU$ . Our formula is slightly more accurate than that in<sup>14</sup> for  $NU > 100$ , whereas it tends to be less accurate for smaller values of  $NU$ . For  $N \geq 10^4$ , we observe a clear improvement of the accuracy of formula (17) as  $U$  increases.



**Figure 2.** (a) (Gaussian case  $\beta=2$ ) Red curves: relative error  $v_\infty/v_{num}$  between the asymptotic rate of adaptation  $v_\infty$  given by our theory (17) and the empirical asymptotic rate of adaptation  $v_{num}$  given by individual-based simulations, depending on the population size  $N$  and the mutation rate  $U$ . Blue curves: relative error  $v_{NH}/v_{num}$ , where  $v_{NH}$  is given by (18), corresponding to the theory in<sup>14</sup> (b). (Generalized Gaussian case  $\beta=10$ ) Red curves: same legend as in panel a. Blue curves: relative error  $v_G/v_{num}$ , where  $v_G$  is given by (21), corresponding to the theory of<sup>13</sup>.

In the case  $\beta=10$  (Fig. 2b), the accuracy of the formula (17) and its behaviour in terms of  $U$  and  $N$  are comparable to the case  $\beta=2$ . Compared to the formula  $v_G$  (21), ours tends to be more accurate for large  $U$  ( $U > 10^{-2}$ ) and less accurate for small  $U$  ( $U < 10^{-2}$ ).

The accuracy of the formula (19) for the stationary variance in fitness  $V_\infty$  was tested with the same model parameters. The obtained results are comparable in terms of accuracy and of dependence of the accuracy with respect to model parameters, to those described above for  $v_\infty$ . These results are presented in Supplementary Fig. S1.

Regarding the full asymptotic distribution of fitness, we compared the analytic description (8), with parameters  $\mu$  and  $\nu$  given by (16)–(17), with empirical distributions obtained by individual-based simulations at time  $t=500$ , averaged over the 100 replicates. This comparison was carried out with four sets of parameters  $N$ ,  $U$ , corresponding to values  $v_\infty/v_{num}$  close to 1 in Fig. 2. The results presented in Fig. 3 show that, in the four cases considered here, our theory provides a very precise description of the expected asymptotic distribution.

**Trajectories of adaptation.** One of the main points of this work is to derive full trajectories of adaptation, connecting the initial state (here, a Dirac distribution at  $m=0$ ) with the stationary regime of adaptation given by the travelling wave described in Section 2.1. Using the formula (16) for setting the value of  $\mu$  in the free boundary dynamical system (5), we were able to get a numerical solution  $p(t, m)$ , using a standard finite difference method (Matlab<sup>®</sup> source code is provided in Supplementary File S3). We compared the trajectories of the mean fitness  $\bar{m}(t)$  and of the expected variance in fitness  $V(t) = \int_{\mathbb{R}} s^2 p(t, s) ds - \bar{m}^2(t)$  with the empirical mean  $\langle \bar{m}_{num}(t) \rangle$  and of the variance in fitness given by individual-based simulations with a Gaussian mutation kernel ( $\sigma^2 = 10^{-4}$ ). All of the simulations were carried-out during the time-interval  $t \in (0, 1000)$ . Solving the Eq. (5) needed about 0.5 second, whereas it took about 18 hours to compute the empirical mean of  $\bar{m}(t)$ , based on 100 replicate populations with  $N=10^6$  individuals, and 6 days with  $N=10^7$  (on an Intel(R) Xeon(R) CPU E5-2637 v3 @ 3.50 GHz).

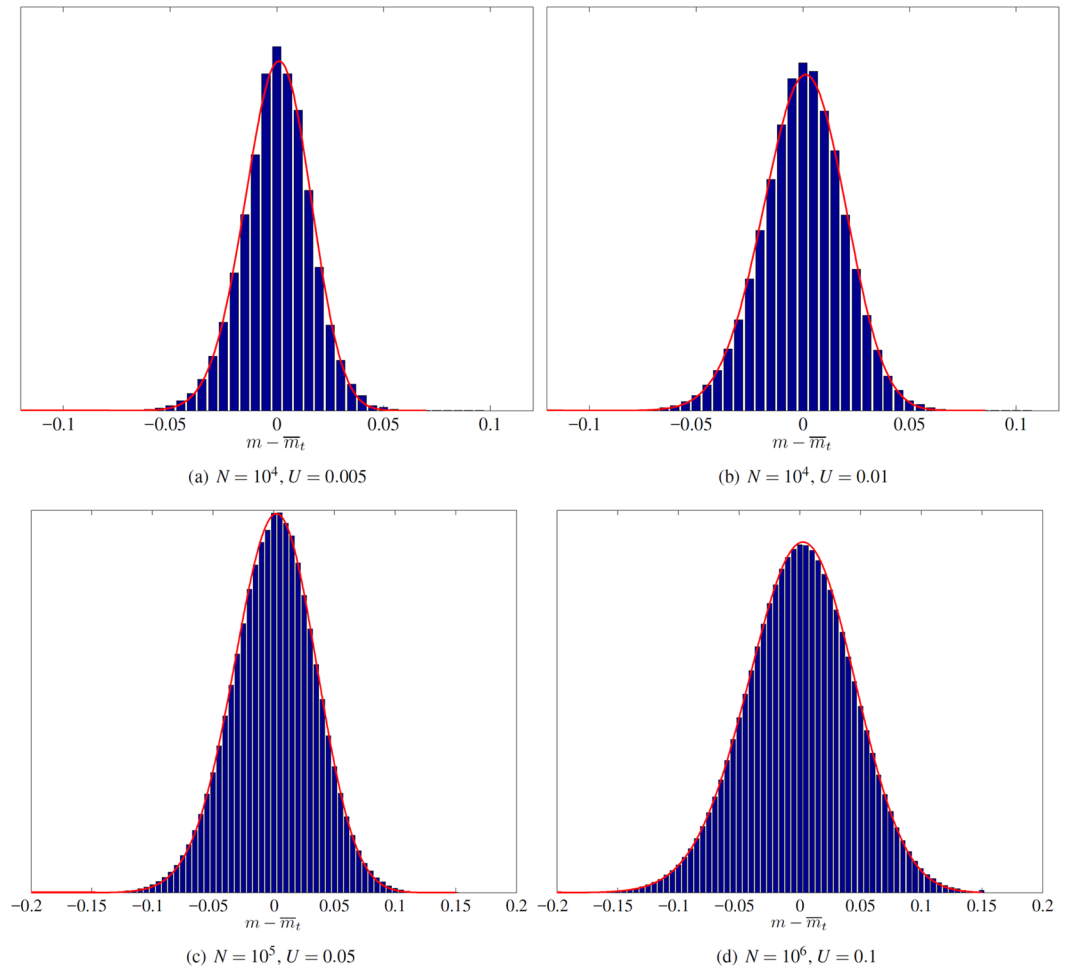
We depicted four trajectories of adaptation in Fig. 4, corresponding to values  $v_\infty/v_{num}$  close to 1 in Fig. 2. Contrarily to other deterministic approaches including beneficial mutations<sup>15,16</sup>, the rate of adaptation  $\nu(t) = \bar{m}'(t)$  converges towards a constant value. Although it was expected, this does not directly follow from the existence of a travelling wave with constant speed  $v_\infty$ , as this travelling wave could have been unstable. Additionally, the approach captures the emergence of a transient maximum in variance, which was not the case with the approach proposed in.<sup>6</sup>

As a byproduct of formula (17), we can compute the characteristic time  $t_q$  that it takes to reach a proportion  $q$  of the speed  $v_\infty$ . For the classical diffusion problem (without free boundary), starting from a clonal population at  $m=0$ , we get  $\bar{m}(t) = Dt^{2/3}$ . Thus,  $\bar{m}'(t_q) = qv_\infty$  implies

$$t_q = qv_\infty/(2D). \quad (22)$$

The accuracy of this formula is also illustrated in Fig. 4.

The numerical solution of Eqs (5)–(6) also provides the expected dynamics of the full distribution of fitness. Convergence towards a traveling wave is illustrated in Supplementary Fig. S2. The accuracy of the fitness distribution predicted by the free boundary approach (5)–(6) with parameter  $\mu$  given by (16) is illustrated in Supplementary Video 4; it is compared to an empirical distribution obtained by individual-based simulation, with  $N=10^4$  and  $U=0.01$ . The distribution predicted by the standard integro-differential model (1) (same Gaussian kernel as in the individual-based model) is illustrated in Supplementary Video 5 for comparison.



**Figure 3.** Asymptotic distribution of fitness. Blue histogram: empirical distribution of fitness, centered at 0 and averaged over 100 replicate populations at time  $t = 500$ , with a Gaussian kernel  $J$  with variance  $\sigma^2 = 10^{-4}$ . Red line: analytic expression (8), with parameters  $\mu$  and  $\nu$  given by (16)–(17).

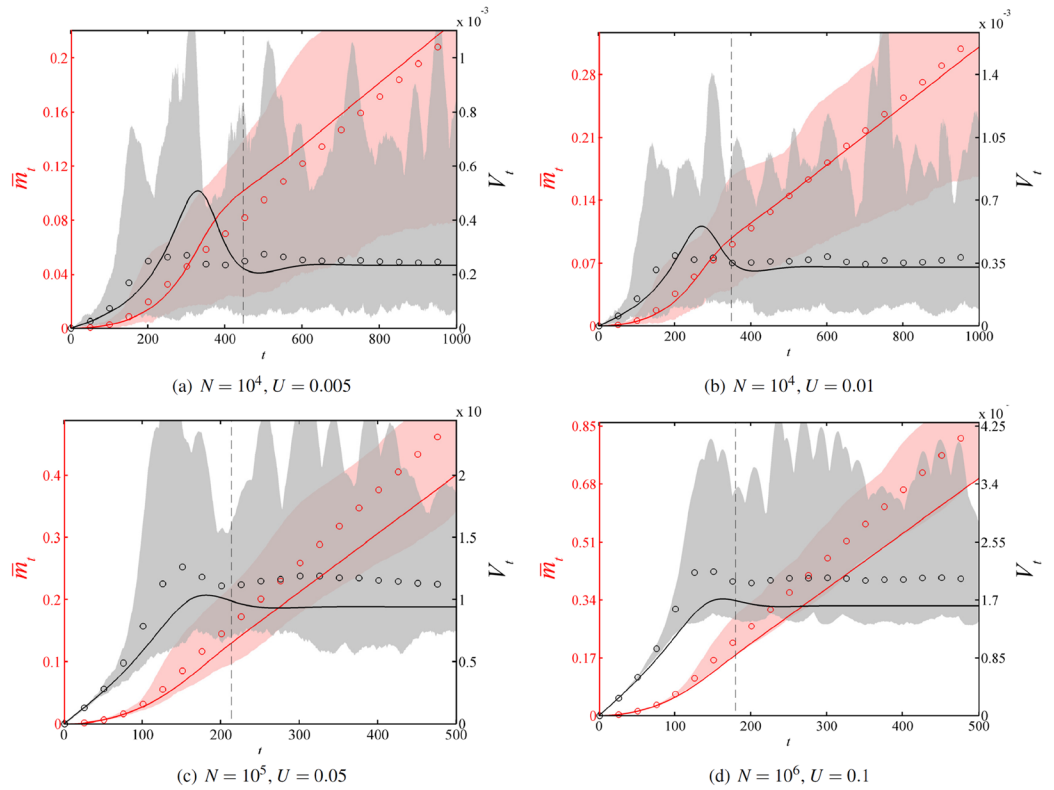
## Discussion

Using an analogy with an ice melting problem, we proposed a deterministic PDE framework which captures the entire dynamics of fitness distribution in the presence of beneficial mutations with non-epistatic effect. Our approach allows for an approximate treatment of genetic drift, while retaining the useful properties of previous deterministic approximations of these dynamics: a full description of the fitness distribution, at all times, even in the presence of multiple co-segregating alleles. On the other hand, as in existing stochastic models, the introduction of the free boundary, driven by stochastic effects at the “nose” of the fitness distribution, leads to a constant asymptotic rate of adaptation, which order is consistent with that given by stochastic individual-based simulations, even with small population sizes (e.g.,  $N = 10^3$ ).

This framework led to new analytic formulas for the asymptotic rate of adaptation  $v_\infty$  and for the asymptotic distribution of fitness  $\phi$ . The formula (17) for  $v_\infty$  depends on the population size  $N$  like  $(\ln(N))^{1/3}$ , which is consistent with previous results obtained in the framework of stochastic PDEs<sup>14,23</sup>. The accuracy of this formula is comparable to that in<sup>13</sup> and<sup>14</sup> for a wide range of population sizes and mutation rates. Yet, the description of genetic drift, in the present approach, is much more simplified than in these previous stochastic PDE models. That both methods yield similar accuracy for long term asymptotic behaviour, suggests that the heuristic used here is sufficient to deal with genetic drift in the free boundary framework.

Regarding the trajectories of adaptation, our theory predicts that the fitness distribution converges towards a travelling wave with speed  $v_\infty$ , whose profile corresponds to the asymptotic distribution of fitness  $\phi$  mentioned above. The numerical simulations of Section 3.2 indicate that, as soon as the formula for  $v_\infty$  is accurate, our approach also successfully captures the transient dynamics of the expected mean fitness and the expected variance in fitness. Although we were not able to derive an analytic description for the transient dynamics, the numerical computation of the solution of our free boundary model was very fast compared to stochastic simulation models. It should therefore be more efficient for computationally intensive studies, in particular for parameter estimation (e.g., estimation of the mutation rate) and model validation, based on experimental data.





**Figure 4.** Mean fitness  $\bar{m}(t)$  and variance trajectories  $V(t)$ . Plain lines: values of  $\bar{m}(t)$  (in red) and  $V(t) = \int_{\mathbb{R}} s^2 p(t, s) ds - \bar{m}^2(t)$  (in black) given by numerically solving (5)–(6). Circles: empirical mean fitness  $\langle \bar{m}_{num}(t) \rangle$  and variance given by individual-based simulations averaged over 100 populations, with a Gaussian kernel  $J$  with variance  $\sigma^2 = 10^{-4}$ . Shaded regions: 99% confidence intervals for the mean fitness (in red) and the variance (in gray). The vertical dashed line corresponds to the characteristic time  $t_q$  given by formula (22), such that a proportion  $q = 0.95$  of the asymptotic rate of adaptation is reached.

Free boundary problems have already been developed to analyze the dynamics of other  $N$ -particle systems<sup>26–28</sup>, such as  $N$ -Branching Random Walks and  $N$ -Branching Brownian Motions introduced by Brunet and Derrida<sup>19,29</sup>. In these systems, all of the particles reproduce at the same rate, independently of their positions, and the leftmost particle is removed at each birth event. These assumptions naturally lead to free boundary problems in the limit  $N \rightarrow +\infty$ , with a free boundary situated on the left (opposite side of the direction of propagation). Our approach is fundamentally different due to the nature of the  $N$ -particle system that we consider. In the Wright-Fisher model of selection and genetic drift (i) the “drift” step consists in a uniform sampling of individuals (or particles) to form the next generation, leading to a limit system which is not a free boundary problem; (ii) the selection step implies that individuals with larger fitness (i.e., larger value on the  $m$ -axis) have more offspring on average, by definition of fitness. This leads to a limit system which does not capture the qualitative dynamics of the finite particle system, even for  $N$  large (finite vs infinite speed of propagation), although the limit system does give a good approximation of the dynamics over a limited period of time, which increases as  $N$  gets larger, see<sup>6</sup>. Therefore, the free boundary model that we constructed here is more a “mesoscopic” or intermediate scale model than an infinite  $N$  limit of a  $N$ -particle system. The model captures some properties of the finite particle system (upper bound to the support of the distribution, finite rate of adaptation, dependence of this rate with respect to  $N$ ) while preserving the PDE formalism of the limit system. As  $N \rightarrow +\infty$ , its properties become analogous to the initial PDE limit system ( $v_\infty \rightarrow +\infty$  and  $s'(t) \rightarrow +\infty$ ), emphasizing its mesoscopic nature.

The approach developed here assumes no epistasis, leading to unbounded fitness trajectories. Saturating fitness trajectories (sublinear increase) are typically observed in long term evolution experiments (e.g. reviewed in<sup>30</sup>). These patterns are consistent with pervasive “diminishing returns” epistasis<sup>31</sup>, i.e., combinations of beneficial mutations are less advantageous than their summed effects would predict. Such trajectories can be generated if one assumes that there is a fitness optimum,  $m^*$ . In such case, by definition, beneficial mutations cannot go beyond the optimal fitness, thus the diffusion term  $D\partial_m^2 p$  in Eq. (5) must be replaced by a “context-dependent” mutation term, i.e., a mutation term which depends on the current fitness  $m$ . Examples of such operators include integral terms of the form  $U \int_{\mathbb{R}} J_y(m - y)p(t, y) dy - Up(t, m)$ , where  $J_y$  is the mutation kernel, given the fitness  $y$  of the parent (the kernel is then supported in  $(-\infty, m^* - y]$ ). However, to the best of our knowledge, there is no theory for dealing with such nonlocal operators in a free boundary framework.

Finally, we believe that the formal approach which was developed here may be generalized to other frameworks, for the derivation of deterministic approximations of a finite population systems. For instance, to compute correction terms for the speed of propagation of traveling wave solutions of Fisher-KPP type equations, as in<sup>19,32</sup>.

## References

- Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**, 457–469 (2003).
- Sniegowski, P. D. & Gerrish, P. J. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1255–1263 (2010).
- Lambert, A. Population dynamics and random genealogies. *Stochastic Models* **24**, 45–163 (2008).
- Kryazhimskiy, S., Tkačik, G. & Plotkin, J. B. The dynamics of adaptation on correlated fitness landscapes. *Proceedings of the National Academy of Sciences* **106**, 18638–18643 (2009).
- Good, B. H. & Desai, M. M. The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics* **85**, 177–190 (2015).
- Martin, G. & Roques, L. The non-stationary dynamics of fitness distributions: Asexual model with epistasis and standing variation. *Genetics* **204**, 1541–1558 (2016).
- Rice, S. H. *Evolutionary theory: mathematical and conceptual foundations* (Sinauer Associates, 2004).
- Gerrish, P. J., Colato, A., Perelson, A. S. & Sniegowski, P. D. Complete genetic linkage can subvert natural selection. *Proceedings of the National Academy of Sciences* **104**, 6266–6271 (2007).
- Desai, M. M. & Fisher, D. S. The balance between mutators and nonmutators in asexual populations. *Genetics* **188**, 997–1014 (2011).
- Gerrish, P. J. & Lenski, R. E. The fate of competing beneficial mutations in an asexual population. *Genetica* **102**, 127–144 (1998).
- Rouzine, I. M., Wakeley, J. & Coffin, J. M. The solitary wave of asexual evolution. *Proceedings of the National Academy of Sciences* **100**, 587–592 (2003).
- Desai, M. M. & Fisher, D. S. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798 (2007).
- Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O. & Desai, M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences* **109**, 4950–4955 (2012).
- Neher, R. A. & Hallatschek, O. Genealogies of rapidly adapting populations. *Proc Natl Acad Sci USA* **110**, 437–442 (2013).
- Alfaro, M. & Carles, R. Explicit solutions for replicator–mutator equations: Extinction versus acceleration. *SIAM Journal on Applied Mathematics* **74**, 1919–1934 (2014).
- Gil, M.-E., Hamel, F., Martin, G. & Roques, L. Mathematical properties of a class of integro-differential models from population genetics. *SIAM J. Appl. Math.* **77**, 1536–1561 (2017).
- Roques, L. *Modèles de réaction–diffusion pour l’écologie spatiale* (Editions Quae, 2013).
- Protter, M. H. & Weinberger, H. F. *Maximum Principles in Differential Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1967).
- Brunet, E. & Derrida, B. Shift in the velocity of a front due to a cut-off. *Physical Review E* **56**, 2597–2604 (1997).
- Tsimring, L. S., Levine, H. & Kessler, D. A. RNA virus evolution via a fitness–space model. *Physical review letters* **76**, 4440–4443 (1996).
- Hallatschek, O. The noisy edge of traveling waves. *Proc Natl Acad Sci USA* **108**, 1783–1787 (2011).
- Good, B. H. & Desai, M. M. Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. *Theoretical population biology* **85**, 86–102 (2013).
- Cohen, E., Kessler, D. A. & Levine, H. Front propagation up a reaction rate gradient. *Physical Review E* **72**, 066126 (2005).
- Stefan, J. Über die theorie der eisbildung. *Monatshfte Mat Phys* **1**, 1–6 (1890).
- Du, Y. & Guo, Z. The Stefan problem for the Fisher–KPP equation. *Journal of Differential Equations* **253**, 996–1035 (2012).
- Durrett, R., Remenik, D. *et al.* Brunet–Derrida particle systems, free boundary problems and Wiener–Hopf equations. *The Annals of Probability* **39**, 2043–2078 (2011).
- Groisman, P. & Jonckheere, M. Front propagation and quasi-stationary distributions: the same selection principle? *arXiv preprint arXiv 1304.4847* (2013).
- De Masi, A., Ferrari, P. A., Presutti, E. & Soprano-Loto, N. Hydrodynamics of the  $N$ -BBM process. *arXiv preprint arXiv 1707.00799* (2017).
- Brunet, E. & Derrida, B. Effect of microscopic noise on front propagation. *Journal of Statistical Physics* **103**, 269–282 (2001).
- Couce, A. & Tenaillon, O. A. The rule of declining adaptability in microbial evolution experiments. *Frontiers in genetics* **6**, 99 (2015).
- Martin, G., Elena, S. F. & Lenormand, T. Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nature genetics* **39**, 555–560 (2007).
- Brunet, E., Derrida, B., Mueller, A. H. & Munier, S. Phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Physical Review E* **73**, 056126 (2006).

## Acknowledgements

The research leading to these results has received funding from the French Agence Nationale de la Recherche within the projects NONLOCAL (ANR-14-CE25-0013) and MECC (ANR-13-ADAP-0006) and from the European Research Council with the project MESOPROBIO (ERC-2014-STG). This work was also funded by INRA grant “MEDIA”.

## Author Contributions

J.G., G.M. and L.R. conceived the model and wrote the paper, J.G. and L.R. carried out the mathematical analysis, L.R. carried out the numerical computations. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-17212-5>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017