CROSSING A FITNESS VALLEY AS A METASTABLE TRANSITION IN A STOCHASTIC POPULATION MODEL

ANTON BOVIER, LOREN COQUILLE, AND CHARLINE SMADI

Abstract. We consider a stochastic model of population dynamics where each individual is characterised by a trait in \( \{0, 1, \ldots, L\} \) and has a natural reproduction rate, a logistic death rate due to age or competition, and a probability of mutation towards neighbouring traits at each reproduction event. We choose parameters such that the induced fitness landscape exhibits a valley: mutant individuals with negative fitness have to be created in order for the population to reach a trait with positive fitness. We focus on the limit of large population and rare mutations at several speeds. In particular, when the mutation rate is low enough, metastability occurs: the exit time of the valley is an exponentially distributed random variable.

1. Introduction

The biological theory of adaptive dynamics aims at studying the interplay between ecology and evolution through the modelling of three basic mechanisms: heredity, mutations, and selection. It was first developed in the 1990ies, partly heuristically, by Metz, Geritz, Bolker, Pacala, Dieckmann, Law, and coauthors [39, 25, 29, 6, 7, 24].

A rigorous derivation of the theory was achieved over the last decade in the context of stochastic individual-based models, where the evolution of a population of individuals characterised by their phenotypes under the influence of the evolutionary mechanisms of birth, death, mutation, and ecological competition in an inhomogeneous "fitness landscape" is described as a measure valued Markov process. Using various scaling limits involving large population size, small mutation rates, and small mutation steps, key features described in the biological theory of adaptive dynamics, in particular the canonical equation of adaptive dynamics (CEAD), the trait substitution sequence (TSS), and the polymorphic evolution sequence (PES) were recovered, see [15, 14, 28, 16, 17, 3]. Extensions of those results for more structured populations were investigated, for example, in [47, 36].

Contrarily to the population genetics approach, individual-based models of adaptive dynamics take into account varying population sizes as well as stochasticity, which is necessary if we

\[ \text{2010 Mathematics Subject Classification.} \quad 92D25, 60J80, 60J27, 92D15, 60F15, 37N25. \]

\[ \text{Key words and phrases.} \quad \text{Eco-evolution; Birth and death process with immigration; Selective sweep; Coupling; competitive Lotka-Volterra system with mutations.} \]

A.B. acknowledges financial support from the German Research Foundation (DFG) through the Clusters of Excellence Hausdorff Center for Mathematics and ImmunoSensation, the Priority Programme SPP 1590 Probabilistic Structures in Evolution, and the Collaborative Reseach Centre CRC 1060 The Mathematics of Emergent Effects. L.C. has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) through the Exploratory Project CanDYPop and by the Swiss National Science Foundation through the grant No. P300P2_161031. This work was also partially funded by the Chair Modélisation Mathématique et Biodiversité of VEOLIA-Ecole Polytechnique-MNHN-F.X.
aim at better understanding of phenomena involving small populations, such as mutational meltdown [21], invasion of a mutant population [14], evolutionary suicide and rescue [1], population extinction time [18, 20], or recovery phenomena [4, 8].

The emerging picture allows to give the following description of the evolutionary fate of a population starting in a monomorphic initial state: first, on a fast ecological time scale, the population reaches its ecological equilibrium. Second, if mutations to types of positive invasion fitness (the invasion fitness is the average growth rate of an individual born with this trait in the presence of the current equilibrium population) are possible, these eventually happen and the population is substituted by a fitter type once a mutant trait fixates (if coexistence is not possible). This continues, and the monomorphic population moves according to the TSS (resp. the CEAD, if mutations steps are scaled to zero) until an evolutionary singularity is reached: here two types of singularities are possible: either, the singularity is stable, in the sense that no type with positive invasion fitness can be reached, or there are several directions with equal positive fitness that can be taken. In the latter case the population splits into two or more sub-populations of different types which then continue to move on until again an evolutionary singularity is reached. If the mutation probability is small enough, all this happens on a time scale of order $1/(\mu K)$, where $\mu$ is the mutation probability and $K$ is the carrying capacity, which is a measure of the maximal population size that the environment can sustain for a long time. This process goes on until all sub-populations are located in stable evolutionary singularities. At this stage, no single mutation can lead to a trait with positive invasion fitness. Nonetheless, there may be traits with positive invasion fitness that can be reached through several consecutive mutation steps [37, 22]. The purpose of the present paper is to present a precise analysis of how such an escape from a stable singularity will happen in various scaling regimes.

As we will show, three essentially different dynamics may occur. In the first one, the mutation probability is so large that many mutants (a number of order $\mu K$, where $\mu$ will denote the mutation probability) are created in a time of order 1. In this case the fixation time scale is dominated by the time needed for a successful mutant to invade (which is of order $\log 1/\mu$). The second scenario occurs if the mutation probability is smaller, but large enough so that a fit mutant will appear before the resident population dies out. In this case the fixation time scale is exponentially distributed and dominated by the time needed for the first successful mutant to be born. The last possible scenario is the extinction of the population before the fixation of the fit mutant, which occurs when the mutation probability is very small (smaller than $e^{-CK}$ for a constant $C$ to be made precise later).

In the sequel, we denote by $\mathbb{N}$ the set of integers $\{1, 2, 3, \ldots\}$ and by $\mathbb{N}_0$ the set $\mathbb{N} \cup \{0\}$. For $n, m \in \mathbb{N}_0$ such that $n \leq m$, we also introduce the notation $\llbracket n, m \rrbracket := \{n, n+1, \ldots, m\}$.

2. Model

In this paper we analyse the escape problem in a specific simple model situation that, however, captures the key mechanisms. We consider a finite trait space $\llbracket 0, L \rrbracket$ on which the population evolves. To each trait $i \in \llbracket 0, L \rrbracket$, we assign

- a clonal birth rate: $(1 - \mu) b_i \geq 0$, where $0 \leq \mu \leq 1$ is the mutation probability;
- a natural death rate: $d_i \geq 0$. 

An individual can also die from type-dependent competition. We assign to each pair \((i, j) \in \mathbb{I}_0^2\):
- a \textbf{competition kernel}: \(c_{ij} \geq 0\), where \(c_{ii}, c_{i0}, c_{iL} > 0\), for all \(i \in \mathbb{I}_0^L\).

To be able to scale the effective size of a population, the competition kernel is scaled down by the so-called \textit{carrying capacity}, \(K\), that is, the competitive pressure exerted by an individual of type \(j\) on an individual of type \(i\) is \(c_{ij}/K\). Finally, to represent mutations, we assign to each pair \((i, j) \in \mathbb{I}_0^2\):

- a \textbf{mutation kernel}: \((m_{ij})_{(i,j) \in \mathbb{I}_0^2}\) satisfying \(m_{ij} \in [0, 1]\), for all \((i, j) \in \mathbb{I}_0^2\) and \(\sum_{j \in [0, L]} m_{ij} = 1\). We will focus on two cases:

\[
m_{ij}^{(1)} = \delta_{i+1,j} \quad \text{or} \quad m_{ij}^{(2)} = \frac{1}{2}(\delta_{i+1,j} + \delta_{i-1,j}),
\]

where \(\delta_{i,j}\) is the Kronecker delta \((1 \text{ if } i = j, 0 \text{ otherwise})\).

The state of a population is an element of \(\mathbb{N}_0^L\). As we will see, before the population extinction, which is of an exponential order (see Section 3.3), the total population size has the same order as the carrying capacity \(K\). Hence it will be more convenient to study the rescaled process \(X^K = (X^K_0(t), \ldots, X^K_L(t)) = X/K \) and to think of this as an element of \(\mathbb{R}^L\). Let \(e_i\) denote the \(i\)-th unit vector in \(\mathbb{R}^L\). The generator of \(X^K\) acts on bounded measurable functions \(f: \mathbb{R}^L \to \mathbb{R}\), for all \(X^K \in (\mathbb{N}_0/K)^L\), as

\[
(L^K f)(X^K) = (1 - \mu)K \sum_{i=0}^{L} (f(X^K + c_i/K) - f(X^K))b_iX^K_i + K \sum_{i=0}^{L} (f(X^K - e_i/K) - f(X^K))(d_i + \sum_{j=0}^{L} c_{ij}X^K_j)X^K_i + \mu K \sum_{i=0}^{L} \sum_{j=0}^{L} (f(X^K + e_j/K) - f(X^K))b_jm_{ij}X^K_i.
\]

A key result, due to Ethier and Kurtz [27], is the law of large number when \(K \uparrow \infty\) (for fixed \(\mu\) and fixed time intervals), which we recall now.

**Proposition 2.1** ([27], Chapter 11, Thm 2.1). \textit{Suppose that the initial conditions converge in probability to a deterministic limit, i.e. \(\lim_{K \to \infty} X^K(0) = x(0)\). Then, for each \(T \in \mathbb{R}_+\), the rescaled process \((X^K(t), 0 \leq t \leq T)\) converges in probability, as \(K \to \infty\), to the deterministic process \(x^\mu = (x^\mu_0, \ldots, x^\mu_L)\) which is the unique solution to the following dynamical system:

\[
\frac{dx^\mu_i}{dt} = \left((1 - \mu)b_i - d_i - \sum_{j=0}^{L} c_{ij}x^\mu_j\right)x^\mu_i + \mu \sum_{j} m_{ij}b_jx^\mu_j, \quad i = 0, \ldots, L,
\]

with initial condition \(x(0)\).

There will be two important quantities associated with our processes. The equilibrium density of a monomorphic \(i\)-population is

\[
\bar{x}_i := \frac{b_i - d_i}{c_{ii}} \vee 0.
\]
The effective growth rate (or selective advantage or disadvantage) of a small mutant population with trait $i$ in a $j$-population at equilibrium, is the so-called \textit{invasion fitness}, $f_{ij}$, given by

$$f_{ij} := b_i - d_i - c_{ij} \bar{x}_j.$$ \hfill (2.5)

The importance of the above two quantities follow from the properties of the limiting competitive Lotka-Volterra system \hbox{(2.3)} with $\mu = 0$. Indeed, if we assume

$$\bar{x}_1 = \frac{b_1 - d_1}{c_{11}} > 0, \quad \text{and} \quad f_{01} < 0 < f_{10},$$ \hfill (2.6)

then the system \hbox{(2.3)} with $\mu = 0$ and $L = 1$ has a unique stable equilibrium $(x_0 = 0, x_1 = \bar{x}_1)$ and two unstable steady states $(x_0 = \bar{x}_0, x_1 = 0)$ and $(x_0 = 0, x_1 = 0)$.

We are interested in the situation where $\bar{x}_0 > 0$, $f_{i0} < 0$, $1 \leq i \leq L - 1$, $f_{i0} > 0$, and $f_{0L} < 0$. Under these assumptions, all mutants created by the initial population will initially have a negative growth rate and thus tend to die out. However, if by chance such mutants survive long enough to give rise to further mutants, etc, such that eventually an individual will reach the trait $L$, it will found a population at this trait that, with positive probability, will grow and eliminate the resident population through competition. Our purpose is to analyse precisely how this process will happen. The process that we want to describe can be seen as a manifestation of the phenomenon of \textit{metastability} (see, e.g., the recent monograph [9] and references therein). The initial population appears stable for a long time and makes repeated attempts to send mutants to the trait $L$, which will eventually be reached and take over the entire population. As we will see, this leads to several features known from metastable phenomena in other contexts: exponential laws of the transition times, fast realisation of the final "success run", and the realisation of this run by a "most likely" realisation. As usual in the context of metastability, we need a scaling parameter to make precise asymptotic statements. In our case this is the \textit{carrying capacity}, $K$, which allows to scale the population size to infinity. Apart from scaling the population size by taking $K \uparrow \infty$, we will also be interested in the limit of small mutation probabilities, $\mu = \mu_k \downarrow 0$, with possibly simultaneous time rescaling. This gives rise to essentially different asymptotics, depending on how $\mu$ tends to zero as a function of $K$.

3. Results

Before stating our main results, let us make our assumptions precise:

\textbf{Assumption 1.}

- Viability of the resident population:
  \[ \bar{x}_0 > 0. \]

- Fitness valley: All traits are unfit with respect to 0 except $L$:
  \[ f_{i0} < 0, \text{ for } i \in \llbracket 1, L - 1 \rrbracket \text{ and } f_{i0} > 0. \] \hfill (3.1)

- All traits are unfit with respect to $L$:
  \[ f_{iL} < 0, \text{ for } i \in \llbracket 0, L - 1 \rrbracket. \] \hfill (3.2)
The following fitnesses are different:

\[ f_{i0} \neq f_{j0}, \text{ for all } i \neq j, \quad (3.3) \]
\[ f_{iL} \neq f_{jL}, \text{ for all } i \neq j. \quad (3.4) \]

![Figure 1. Example of a fitness landscape satisfying Assumption 1 with \( L = 6 \). Blue curve: \( i \mapsto f_{i0} \), red curve: \( i \mapsto f_{iL} \).](image)

Note that conditions (3.3) and (3.4) are imposed in order to lighten the analysis of the deterministic system (see Lemma 6.1). Similar results are probably true without this assumptions but the proofs would be unnecessarily more technical. Similar hypotheses are made in the article [26].

Before proceeding to the statements of our results, let us show that Assumption 1 can be realised with well-chosen birth-, death-, and competition rates. A possibility is to fix birth and death rates associated to every trait to be 1 and 0, respectively. In that case, Assumption 1 will impose constraints on the competition rates \( (c_{i0})_{i \in \{1, L\}} \) and \( (c_{iL})_{i \in \{0, L-1\}} \), which must be equal to \( (1 - f_{i0})_{i \in \{1, L\}} \) and \( (1 - f_{iL})_{i \in \{0, L-1\}} \), respectively. We complete the competition matrix by taking symmetric values (except for \( c_{0L} \) and \( c_{L0} \) which are now fixed and different) and by choosing \( c_{ij} = 1 \), for all pairs \( (i, j) \in \{1, L-1\}^2 \).

3.1. **Deterministic limit** \( (K, \mu) \to (\infty, \mu), \text{ then } \mu \to 0 \). The first regime we are interested in is the case when \( \mu \) is small but does not scale with the population size. From a biological point of view, this corresponds to high mutation probabilities. Note that a similar scaling has been studied in [10] and [26]. In both papers, the context was very different since these authors considered the arrival of fitter mutants, rather than unfitter mutants, as we do here. In [10], individuals only suffer competition from the nearest neighbouring traits. In [26], an exponentially growing population of tumor cells is modelled by a Moran model, and back mutations are not considered. The analysis of the limiting deterministic system has some similarities but the concentration of the (rescaled) stochastic system is quite different due to the very different fitness hypotheses.

**Theorem 3.1.** Suppose that Assumption 1 holds. Take as initial condition

\[ x^\mu(0) = (\bar{x}_0, 0, \ldots, 0). \]

Then, for \( i \in \{0, L\} \), as \( \mu \to 0 \), uniformly on bounded time intervals,

\[ \frac{\log \left[ x^\mu_i(t \cdot \log (1/\mu)) \right]}{\log(1/\mu)} \to x_i(t), \]

(3.6)
where \( x_i(t) \) is piece-wise linear. More precisely,

(1) in the case of 1-sided mutations, \( m_{ij} = m_{ij}^{(1)} \), for \( i \in \{0, L - 1\} \),

\[
x_i(t) = \begin{cases} 
-i, & \text{for } 0 \leq t < L/f_{I0}, \\
-i - (t - L/f_{I0}) \min_{k \in \{0, 1\}} |f_k|, & \text{for } t > L/f_{I0},
\end{cases}
\]

and

\[
x_L(t) = \begin{cases} 
-L + f_{I0}t, & \text{for } 0 \leq t < L/f_{I0}, \\
0, & \text{for } t > L/f_{I0}.
\end{cases}
\]

(2) in the case of 2-sided mutations, \( m_{ij} = m_{ij}^{(2)} \): consider the sequence \( \{i_1, \ldots, i_r\} \) of “fitness records”, defined recursively by \( i_1 = 0, i_k = \min\{i \in \{0, L - 1\} : f_i < f_{i_{k-1}}\} \),

\[
x_i(t) = \begin{cases} 
-i \vee (-L - (L - i) + f_{I0}t), & \text{for } 0 \leq t < L/f_{I0}, \\
-(L - i) \vee \max_{k \in \{0, 1\}} \{-i - |f_k|(t - L/f_{I0})\} & \max_{k \in \{1, \ldots, r\}} \{-i_k - |f_k|(t - L/f_{I0})\}, & \text{for } t > L/f_{I0}.
\end{cases}
\]

Moreover,

\[
\left( x_0^f(t \log(1/\mu)), x_L^f(t \log(1/\mu)) \right) \to \begin{cases} 
(\bar{x}_0, 0), & \text{for } 0 \leq t < L/f_{I0}, \\
(0, \bar{x}_L), & \text{for } t > L/f_{I0}.
\end{cases}
\]

The shape of \( x(t) := (x_0(t), \ldots, x_L(t)) \) can be seen on Figures 3 and 4 in the 1-sided and 2-sided cases, respectively.

In the 1-sided case, the rescaled deterministic process \( x(t) \) can be explained as follows: In the first phase, the 0-population stays close to \( \bar{x}_0 \) until the \( L \)-population reaches order one. As competition between the populations of type \( i \) and \( j \) for \( i, j \neq 0 \) is negligible in comparison to competition between type \( i \) and type 0, for \( i \in \{1, L\} \), the \( i \)-population first stabilises around \( O(\mu^t) \) in a time of order \( o(1) \), then the \( L \)-population, starting from a size \( O(\mu^t) \), grows exponentially with rate \( f_{I0} \) until reaching order one (which takes a time \( L/f_{I0} \)) while the other types stay stable. Next, a swap between populations 0 and \( L \) (two-dimensional Lotka-Volterra system) is happening in a time of order \( o(1) \), and finally, for \( i \neq L \), the \( i \)-population decays exponentially from \( O(\mu^t) \) with a rate given by the lowest (negative) fitness of its left neighbours, \( \min_{k \in \{0, 1\}} |f_k| \) while the \( L \)-population approaches its equilibrium density \( \bar{x}_L \). To understand the rate of decrease during the last phase, let us consider only the 0- and 1-populations. The competition exerted by populations \( j \in \{0, L - 1\} \) on the 0- and 1-populations is negligible with respect to the competition exerted by the \( L \)-population, which has a size of order 1. As a consequence, \( x_0^f \) has a dynamics close to this of the solution to:

\[ \dot{\bar{x}}_0(t) = f_{0L}\bar{x}_0(t), \]

that is to say, \( x_0(t) \sim x_0(0)e^{f_{0L}t} \), and \( x_1 \) has a dynamics close to this of the solution to:

\[ \dot{\bar{x}}_1(t) = f_{1L}\bar{x}_1(t) + \mu\bar{x}_0(t) = f_{1L}\bar{x}_1(t) + \mu \bar{x}_0(0)e^{f_{0L}t}, \]

that is to say

\[ x_1(t) \sim x_1(0)e^{f_{1L}t} + \mu \frac{x_0(0)}{f_{0L} - f_{1L}} \left( e^{f_{0L}t} - e^{f_{1L}t} \right). \]
From this heuristics we get that

\[ x_i(t \log(1/\mu)) \sim x_i(0)\mu^{l_{iL}} + \mu \frac{x_i(0)}{[f_i L] - [f_0 L]} \left( \mu^{l_{iL}} - \mu^{l_{iL}}} \right) \]

\[ = \mu \left( C_i \mu^{l_{iL}} + \frac{C_0}{[f_i L] - [f_0 L]} \right), \quad (3.11) \]

where \( C_0 \) and \( C_1 \) are of order 1. We thus see that the leading order is \( \mu^{1+\inf([f_0 L], [f_i L])} \). Reasoning in the similar way for the other populations yields that the leading order for the variation of the \( i \)-population size \( (i \in [0, L - 1]) \) is \( \mu^{1+\inf([f_0 L], [f_0 L], \ldots, [f_i L])} \).

In the 2-sided case, a modification of the order of magnitude of the \( i \)-population (for \( i \neq L \)) happens due to backward mutations. The reasoning is similar than the heuristics we have just described, except that for the decrease rate of the \( i \)-population \( (i \in [1, L]) \) to have an impact on the decrease rate of the \( (i - 1) \)-population, their respective sizes have to satisfy \( x_i/\mu \geq C/\mu \), for a positive constant \( C \). Under this condition the number of type-\( (i - 1) \) individuals produced by mutations by the type \( i \)-individuals has the same order as the type \( (i - 1) \) population size.

### 3.2. Stochastic limit \((K, \mu) \rightarrow (\infty, 0)\)

When the mutation probability is small, the dynamics and time scale of the invasion process will depend on the scaling of the mutation probability per reproductive event, \( \mu \), with respect to the carrying capacity \( K \). We will consider in this section mutation probabilities with two possible forms. Either,

\[ \mu = f(K) K^{-1/\alpha}, \quad \text{with} \quad \alpha \geq 1 \quad \text{and} \quad |\ln f(K)| = o(\ln K), \quad (3.12) \]

or

\[ \mu = o(1/K). \quad (3.13) \]

For simplicity, in Sections 3.2 and 3.3 we only consider the mutation kernel \( m^{(i)}_{ij} = \mu \delta_{i+1, j} \).

For \( v \geq 0 \) and \( 0 \leq i \leq L \), let \( T^{(K,v)}_v \) denote the first time the \( i \)-population reaches the size \([vK],

\[ T^{(K,v)}_v := \inf \{ t \geq 0, X_i(t) = [vK] \}. \quad (3.14) \]

In a time of order one, there will be of order \( K \mu^i \) mutants of type \( i \), provided that this number is larger than 1. In particular, there will be of order \( K \mu^L \) fit \( L \)-mutants at time one, if \( L/\alpha < 1 \). This is the regime of large mutation probability. In this case, the time for the \( L \)-population to hit a size of order \( K \) is of order \( \log K \). We obtain a precise estimate of this time, as well as of the time for the trait \( L \) to outcompete the other traits under the same assumptions. Let us introduce

\[ t(L, \alpha) := \frac{L}{\alpha} f_{iL} \sup \{ \left( 1 - \frac{i}{\alpha} \right) \frac{1}{|f_i L|}, 0 \leq i \leq L - 1 \}, \]

and the time needed for the populations at all sites but \( L \) to get extinct,

\[ T^{(K,\Sigma)}_0 := \inf \{ t \geq 0, \sum_{0 \leq i \leq L - 1} X_i(t) = 0 \}. \quad (3.15) \]

With this notation we have the following asymptotic result.
Theorem 3.2. Assume that (3.12) holds and that \( L < \alpha < \infty \). Then there exist two positive constants \( \varepsilon_0 \) and \( c \) such that, for every \( 0 < \varepsilon < \varepsilon_0 \),

\[
\liminf_{K \to \infty} \mathbb{P} \left((1 - ce) \frac{L}{\alpha f_{l,0}} > \frac{T_{x}^{(K,L)}}{\log K} < \frac{T_{x}^{(K,L)}}{\log K} < (1 + ce) \frac{L}{\alpha f_{l,0}}\right) \geq 1 - ce. \tag{3.16}
\]

Moreover,

\[
\frac{T_{0}^{(K,\Sigma)}}{\log K} \to t(L, \alpha), \quad \text{in probability,} \quad (K \to \infty) \tag{3.17}
\]

and there exists a positive constant \( V \) such that

\[
\limsup_{K \to \infty} \mathbb{P} \left( \sup_{i \leq \alpha} | X_{L} \left(T_{x}^{(K,L)} + t\right) - \bar{x}_{L}K | > c\varepsilon K \right) \leq ce. \tag{3.18}
\]

In other words, it takes a time of order \( t(L, \alpha) \log K \) for the \( L \)-population to outcompete the other populations and enter in a neighbourhood of its monomorphic equilibrium size \( \bar{x}_{L}K \). Once this has happened, it stays close to this equilibrium for at least a time \( e^{kV} \), where \( V \) is a positive constant.

Note that the constant \( t(L, \alpha) \) can be intuitively computed from the deterministic limit. Indeed, for \( \alpha > L \), we will prove that the system performs small fluctuations around the deterministic evolution studied above: the \( i \)-population first stabilises around \( O(K\mu) \) in a time of order one, then the \( L \)-population grows exponentially with rate \( f_{l,0} \) until reaching order \( K \) (super-critical branching process, needs a time close to \( L \log K/(\alpha f_{l,0}) \)) while the other types stay stable, the swap between populations 0 and \( L \) then takes a time of order one, and finally, for \( i \neq L \), the \( i \)-population decays exponentially from \( O(K\mu) \) to extinction with a rate given by the lowest (negative) fitness of its left neighbours (sub-critical branching process, needs a time close to \( \log K \)). Thus the time until extinction of all non-\( L \) populations is close to the constant \( (3.2) \) times \( \log K \).

Next we consider the case of small mutation probability, when \( L/\alpha > 1 \). In this case, there is no \( L \)-mutant at time one, and the fixation of the trait \( L \) happens on a much larger time scale. In this section, we are interested in the case where the mutation \( L \) goes to fixation with a probability close to one. In particular, the first \( L \)-mutant has to be born before the extinction of the population.

We define, for \( 0 < \rho < 1 \),

\[
\lambda(\rho) := \sum_{k=0}^{\infty} \frac{(2k)!}{(k-1)!(k+1)!} \rho^k (1 - \rho)^{k+1},
\]

and, for \( \lfloor \alpha \rfloor + 1 \leq i \leq L - 1 \), set \( \rho_i := b_i/(b_i + d_i + c_0\bar{x}_0) \).

Theorem 3.3. Assume that (3.12) holds, \( \alpha \notin \mathbb{N} \) and \( \alpha < L \). Then there exist two positive constants \( \varepsilon_0 \) and \( c \), and two exponential random variables \( E_- \) and \( E_+ \) with parameters

\[
(1 + ce) \bar{x}_0 b_0^{L-1} f_{l,0} b_L \prod_{i=\lfloor \alpha \rfloor + 1}^{L-1} \lambda(\rho_i) \quad \text{and} \quad (1 - ce) \bar{x}_0 b_0^{L-1} f_{l,0} b_L \prod_{i=\lfloor \alpha \rfloor + 1}^{L-1} \lambda(\rho_i),
\]
such that, for every \( \varepsilon \leq \varepsilon_0 \),
\[
\liminf_{K \to \infty} \mathbb{P} \left( E_- \leq \frac{T^{(K,L)} \vee T^{(K,\Sigma)}}{K \mu^L} \leq E_+ \right) \geq 1 - c \varepsilon.
\] (3.19)

- There exists a positive constant \( V \) such that if \( \mu \) satisfies
\[
e^{-VK} \ll K \mu \ll 1,
\]
then the same conclusion holds, with the corresponding parameters, for \( E_- \) and \( E_+ \):
\[
(1 + c \varepsilon) \bar{x}_0 \frac{f_{L0}}{b_L} \prod_{i=1}^{L-1} \lambda(\rho_i) \quad \text{and} \quad (1 - c \varepsilon) \bar{x}_0 \frac{f_{L0}}{b_L} \prod_{i=1}^{L-1} \lambda(\rho_i).
\]
Moreover, under both assumptions, there exists a positive constant \( V \) such that
\[
\limsup_{K \to \infty} \mathbb{P} \left( \sup_{t \leq e^{-VK}} \left| X_L \left( T^{(K,L)} + i \right) - \bar{x}_L K \right| > c \varepsilon K \right) \leq c \varepsilon.
\]

In the first case, the typical trajectories of the process are as follows: mutant populations of type \( i \), for \( 1 \leq i \leq \lfloor \alpha \rfloor \), reach a size of order \( K \mu^i \gg 1 \) in a time of order \( \log K \) (they are well approximated by birth-death processes with immigration and their behaviour is then close to the deterministic limit), and mutant populations of type \( i \), for \( \lfloor \alpha \rfloor + 1 \leq i \leq L \), describe a.s. finite excursions, whose a proportion of order \( \mu \) produces a mutant of type \( i+1 \). Finally, every \( L \)-mutant has a probability \( f_{L0}/b_L \) to produce a population which outcompetes all other populations. The term \( \lambda(\rho_i) \) is the expected number of individuals in an excursion of a subcritical birth and death process of birthrate \( b_i \) and death rate \( d_i + c_0 \bar{x}_0 \) excepting the first individual. Hence \( \mu \lambda(\rho_i) \) is the approximated probability for a type \( i \)-population \( (\lfloor \alpha \rfloor + 1 \leq i \leq L - 1) \) to produce a mutant of type \( i+1 \), and the overall time scale can be recovered as follows:

1. The last 'large' population is the \( \lfloor \alpha \rfloor \)-population, which reaches a size of order \( K \mu^{\lfloor \alpha \rfloor} \) after a time which does not go to infinity with \( K \).
2. The \( \lfloor \alpha \rfloor \)-population produces an excursion of an \( (\lfloor \alpha \rfloor + 1) \)-population at a rate of order \( K \mu^{\lfloor \alpha \rfloor + 1} \), which has a probability of order \( \mu \) to produce an excursion of a \( (\lfloor \alpha \rfloor + 2) \)-population, and so on,
giving the order \( K \mu^L \).

Notice that Theorem 3.3 implies that, for any mutation rate which converges to zero more slowly than \( e^{-VK}/K \), the population will cross the fitness valley with a probability tending to 1 as \( K \to \infty \). Our results thus cover a wide range of biologically relevant cases.

In fact, we believe that the results hold true as long as \( K \mu \gg \rho_0(K) \), where \( \rho_0(K) \) is the inverse of the mean extinction time of the 0-population starting at its quasi-stationary distribution (see the next section for a precise definition). However, we were not able to control precisely enough the law of \( X_0 \) before its extinction (but see [18] for results in this direction).

We also think that \( \alpha \notin \mathbb{N} \) is only a technical assumption which could be suppressed but would bring more technicalities into the proof. Namely, in this case, the \( \lfloor \alpha \rfloor \) population size would not be large but of order one and we would have to control its size more carefully.
3.3. **On the extinction of the population.** One of the key advantages of stochastic logistic birth and death processes on constant size processes when dealing with population genetics issues is that we can compare the time scale of mutation processes and the population lifetime. In particular, for the case of fitness valley crossing, we can show that if the mutation probability \( \mu \) is too small, the population will get extinct before the birth of the first mutant of type \( L \).

The quantification of the lifetime of populations with interacting individuals is a tricky question (see [18, 19] for recent results) and we are not able to determine necessary and sufficient conditions for the \( L \)-mutants to succeed in invading before the population extinction. However we will provide some bounds in the next results.

The previous theorem (Theorem 3.3) provided a wide range of mutation probabilities \( \mu \) for which the type \( L \) mutant fixates. The following theorem (Theorem 3.4) provides a small range for which the population dies before the birth of the first \( L \)-mutant. Before stating it, we introduce a parameter scaling the extinction time of the 0-population, \( \rho_0(K) := \sqrt{K \exp(-K(b_0 - d_0 + d_0 \ln(d_0/b_0)))} \). (3.20)

More precisely, it is stated in [18] that \( \mathbb{E}_{\nu}[T_{0}^{(mono)}] = 1/\rho_0(K) \), where \( \nu \) is the stationary distribution of a monomorphic 0-population, and \( T_{0}^{(mono)} \) its extinction time. We also need to introduce the two stopping times
\[
T_0 := \inf\{ t \geq 0, X_i(t) = 0, \forall 0 \leq i \leq L \} \quad \text{and} \quad B_L := \inf\{ t \geq 0, X_L(t) > 0 \},
\]

as well as the following assumption.

**Assumption 2.** The birth- and death-rates satisfy the conditions
\[
b_i < d_i, \quad 1 \leq i \leq L - 1.
\]

Then we have the following result:

**Theorem 3.4.** Suppose that Assumption [1] holds.

1. If \( K\mu \ll \rho_0(K) \), then \( \mathbb{P}(T_0 < B_L) \xrightarrow{K \to \infty} 1 \).
2. If Assumption 2 holds and \( K\mu^L \ll \rho_0(K) \), then \( \mathbb{P}(T_0 < B_L) \xrightarrow{K \to \infty} 1 \).

If \( K\mu^L \ll \rho_0(K) \) but the intermediate mutants are fitter, the pattern is less clear. For instance, one of the intermediate mutants could fix before being replaced (or not) by the type \( L \) mutant.

4. **Generalisations**

Our results can be generalised to the following settings:

- If the fitness landscape is such that coexistence is allowed between populations of traits 0 and \( L \), i.e. if \( f_{10} > 0 \) and \( f_{0L} > 0 \), then the analysis of the invasion phase is the same, but the fixation phase differs in such a way that traits 0 and \( L \) become macroscopic and stabilise around their common equilibrium \((n_0^*, n_L^*)\), the non trivial fixed point of the 2-species Lotka-Volterra system. Moreover, the unfit mutant populations stay microscopic if we assume \( f_{i,0,L} := b_i - d_i - c_{i0}n_0^* - c_{iL}n_L^* < 0 \), for all \( i = 1, \ldots, L - 1 \). In the 1-sided case, those stay of order \( K\mu^i \), while in the 2-sided case, they stay of order \( K\mu^\text{max}[i,L-i] \).

There is no complicated decay phase as in Section 6.3.2 and its stochastic analog.
If the mutation probability $\mu$ depends on the trait $i$, while still fulfilling the prescribed scalings associated to our different theorems, those still hold.

Consider the biologically relevant case (especially for cancer) where deleterious mutations accumulate until a mutant individual gathers $L$ different mutations, in which case it becomes fit. Each individual bearing $k$ mutations can then be labeled by the trait $k$. The main difference with our setting is that there are now $L!$ ways of reaching an individual of trait $L$ with a sequence of $L$ mutations. Thus, the invasion time of the population $L$ is divided by $L!$.

5. BIOLoGICAL CONTEXT

The existence of complex phenotypes often involve interactions between different genetic loci. This can lead to cases, where a set of mutations are individually deleterious but in combination confer a fitness benefit. To acquire the beneficial genotype, a population must cross a fitness valley by first acquiring the deleterious intermediate mutations. Empirical examples of such phenomena have been found in bacteria [43, 38] and in viruses [41, 30], for instance.

To model those phenomena, several authors considered the case of the sequential fixation of intermediate mutants, as it appeared to be the most likely scenario to get to the fixation of the favorable mutant [52, 49, 40], especially when the population size is small or the mutants neutral or weakly deleterious.

A scenario where a combination of mutations fixates simultaneously without the prior fixation of one intermediate mutant was first suggested by Gillespie [31]. He observed that the rate of production of fit genotypes is proportional to the population size, and because in the population genetic models the probability of fixation of a beneficial allele is independent of the population size, he deduced that the expected time for the fixation of the fit mutant decreases as population size increases. Thus it could be a likely process in the evolution of large populations. This scenario, called stochastic tunneling by Iwasa and coauthors [35], has been widely studied since then (see [13, 50, 51, 32, 33] and references therein) by means of constant size population genetic models. But the use of such models hampers taking into account several phenomena.

First, an important question is the lifetime of the population under study. If the mutation probability is too small, the population can get extinct before the appearance of the first favourable mutant. Imposing a constant (finite or infinite) population size is thus very restrictive in this respect. In the case of logistic processes that we are studying in this work, the total population size typically remains in the order of the carrying capacity $K$ during a time of order $e^{KV}$ (with $V$ a positive constant depending on the model’s parameters), before getting extinct.

Second, in population genetic models, a fitness is assigned to each type, independently of the population state. In the case of the Moran model, which is used in the series of papers we just mentioned, the probability for a given individual to be picked to replace an individual who dies is proportional to its fitness. If we want to compare our result with this setting, we have to assume:

$$b_i = b \quad \text{and} \quad |f_{ij}| = |f_{ji}|, \quad \forall 0 \leq i, j \leq L,$$

thus restricting the type of fitnesses we could take into account (see Section 3 in [11] for a detailed discussion on this topic).

Another series of papers [34, 44, 45, 42, 2] focuses on initially large populations doomed for rapid extinction (for instance cancer cells subject to chemotherapy, or viruses invading a new
host while not being adapted to it), except if they manage to accumulate mutations to produce a fit variant (for instance resistant to treatments). The authors use multi-type branching processes to model the population. This approach has the advantage to lead to explicit expressions, as the branching property makes the calculations easier, but has two main drawbacks: first it neglects interactions between individuals, whereas it is well known that they are fundamental in processes such as tumor growth; second, branching processes either go to extinction or survive forever with an exponentially growing size, which is not realistic for biological populations.

A last point we would like to comment is the possibility of back mutations. They are ignored in all papers we mentioned, usually accompanied with the argument that they would not have a macroscopic effect on the processes under consideration. However, it has been shown that, when the mutation probabilities are large enough, scenarios where some loci are subject to two successive opposite mutations are likely to be observed (for an example, see [23]). This is why we included the possibility of back mutations in the case of high mutation probabilities in Section 3.1.

6. Proof of Theorem 3.1

We give the detailed proof for \( L \) even and mention the modifications which have to be made for \( L \) odd during the proof. A key step in the proof of Theorem 3.1 is the following lemma.

**Lemma 6.1.** Let \( \zeta \in \{1, 2\} \), \((b_0, \ldots, b_L) \in (\mathbb{R}^+)^{L+1}\), \((\ell_0, \ldots, \ell_L) \in (\mathbb{R}^+)^{L+1}\), \((p_0, \ldots, p_L) \in (\mathbb{R}^+)^{L+1}\) and \((f_0, \ldots, f_L) \in \mathbb{R}^{L+1}\) such that

\[
f_i \neq f_j, \quad \text{for all } i \neq j. \tag{6.1}
\]

Let

\[
M_\zeta(\mu, L) := \begin{pmatrix}
  f_0 - b_0 \mu & 0 & 0 & 0 & 0 \\
  \frac{\zeta}{\zeta} b_0 & f_1 - b_1 \mu & 0 & 0 & 0 \\
  0 & \frac{\zeta}{\zeta} b_1 & f_2 - b_2 \mu & 0 & 0 \\
  0 & 0 & \ddots & \ddots & 0 \\
  0 & 0 & 0 & \frac{\zeta}{\zeta} b_{L-1} & f_L - b_L \mu
\end{pmatrix}. \tag{6.2}
\]

Then the solution to the linear system

\[
\frac{dy}{dt} = M_\zeta(\mu, L)y, \tag{6.3}
\]

with initial condition

\[
y(0) = (\ell_0 \mu^{p_0}, \ldots, \ell_L \mu^{p_L}), \tag{6.4}
\]

satisfies

\[
\lim_{\mu \to 0} \frac{\log(y_i(t \log(1/\mu)))}{\log(1/\mu)} = \frac{-m_i(t)}{p_f} := -\min_{\gamma, \alpha \in [0, L]} \{i - \gamma + p_\gamma - t f_\alpha\}, \tag{6.5}
\]

with the convention \( p_0 = \infty \), for \( p \geq 0 \).
Proof. Under assumption (6.1), the matrix $M$ in (6.2) is diagonalisable for $\mu$ small enough: it can be checked that $M = S D S^{-1}$ with

$$D = ((f_i - b_i \mu) \delta_{ij})_{0 \leq i,j \leq L},$$

$$S = \left( \left( \frac{\xi}{\mu} \right) \right)^{L-i} \prod_{k=i+1}^{L} (f_j - f_k) + \mu(b_k - b_i) \right) \mathbf{1}_{[i\leq j]} \right)_{0 \leq i,j \leq L},$$

$$S^{-1} = \left( \left( \frac{\mu}{\xi} \right) \right)^{L-j} \prod_{k=j+1}^{L} (f_i - f_k) + \mu(b_k - b_i) \right) \mathbf{1}_{[i\leq j]} \right)_{0 \leq i,j \leq L}.$$

The solution to the system (6.3) can then be written in the form

$$y(t) = \exp(tM_\xi)y(0) = S \exp(tD)S^{-1}y(0),$$

which reads in coordinates, for $i = 0, \ldots, L$,

$$y_i(t) = \sum_{\alpha, \gamma = 0}^{L} S_{\alpha \gamma} e^{(f_{\alpha} - b_{\mu})t} \frac{\mu^\gamma}{S_{\alpha \gamma}} \sum_{\gamma \leq \alpha \leq L} \mu^{-\gamma} e^{(f_{\alpha} - b_{\mu})t} \cdot C_{i \alpha} C'_{\gamma \gamma} \cdot \ell_{\gamma}.$$

Thus

$$y_i(t \log(1/\mu)) = \sum_{\gamma \leq \alpha \leq L} \mu^{-\gamma} e^{(f_{\alpha} - b_{\mu})t} \cdot (C_{i \alpha} C'_{\gamma \gamma} \cdot \ell_{\gamma}).$$

As $\mu$ tends to zero, the sum is dominated by the term with the smallest exponent of $\mu$, by which definition is $m_i(t)$, defined in (6.5). Thus there exists a constant $C > 0$, such that

$$y_i(t \log(1/\mu)) = C \mu^{m_i(t)} (1 + o(1)),$$

which implies the assertion of (6.5) and concludes the proof of the lemma. \hfill \Box

6.1. Before the swap.

6.1.1. Time interval $0 \leq t \leq T_{L^{-1}}$. If $m_{ij} = m_{ij}^{(2)}$ let $\zeta = 2$ and

$$\tau_{L^{-1}}(\epsilon, \mu) = \inf \{ t : \exists i \in [0, L] \text{ s.t. } x^\mu_i(t) > \mu^{-\epsilon} \} \wedge \inf \{ t : |x^\mu_0(t) - \bar{x}_0| > \epsilon \} \wedge \inf \{ t : x^\mu_L(t) > \mu^{L-2+\epsilon} \},$$

while if $m_{ij} = m_{ij}^{(1)}$ let $\zeta = 1$ and

$$\tau_{L^{-1}}(\epsilon, \mu) = \inf \{ t : \exists i \in [0, L] \text{ s.t. } x^\mu_i(t) > \mu^{-\epsilon} \} \wedge \inf \{ t : |x^\mu_0(t) - \bar{x}_0| > \epsilon \} \wedge \inf \{ t : x^\mu_L(t) > \epsilon \}.$$

and define

$$T_{L^{-1}} := \lim_{\epsilon \to 0} \lim_{\mu \to 0} \frac{\tau_{L^{-1}}(\epsilon, \mu)}{\log(1/\mu)}.$$

There exists a finite $C$ such that on the time interval $[0, \tau_{L^{-1}}(\epsilon, \mu)]$, for $i \in [0, L]$,

$$\frac{dx^\mu_i}{dt} \geq (f_0 - C \epsilon)x^\mu_i + \mu \left( \frac{b_{i-1}}{\xi} x^\mu_{i-1} - b_i x^\mu_i \right).$$
Hence, by the Gronwall lemma, \( x^\nu \) is bigger than the solution to \( \frac{dy}{dt} = M_\epsilon y \) with \( f_i = f_{i0} - C \epsilon \). Applying Lemma 6.1 with \( y(0) = (\bar{x}_0, 0, \ldots, 0) \) and thus \( \ell_0 = \bar{x}_0, p_0 = 0, \ell_i = 0 \) for \( i \neq 0 \), we get, using (3.1), for \( \epsilon \) small enough and \( t > 0 \),

\[
\lim_{\mu \to 0} \frac{\log(x_\nu^\mu(t \log(1/\mu)))}{\log(1/\mu)} \geq \begin{cases} -i - C \epsilon t, & \text{for } i = 0, \ldots, L - 1, \\ -L + t(f_{i,0} - C \epsilon), & \text{for } i = L. \end{cases}
\]

(6.16)

On the other hand, on the same time interval, we have, for some positive \( C \),

\[
\frac{dx_i^\mu}{dt} \leq (f_{i0} + C \epsilon)x_i^\mu + \mu \left( \frac{b_{i-1} - b_i}{\mu^2} x_{i-1}^\mu - b_i x_i^\mu \right) + E_i,
\]

(6.17)

where, until \( \tau_{L-1} \), with \( \kappa := \sup b_i / 2 \),

\[
(E, \zeta) = \begin{cases} (\kappa \mu \cdot (\mu^{1-\epsilon}, \mu^{2-\epsilon}, \ldots, \mu^{L-1-\epsilon}, \mu^{L-2+\epsilon}, 0), 2) & \text{if } m_{ij} = m_{ij}^{(2)}, \\ ((0, 0, \ldots, 0), 1) & \text{if } m_{ij} = m_{ij}^{(1)}. \end{cases}
\]

Again by the Gronwall lemma, \( x^\mu \) is smaller than the solution to \( \frac{dy}{dt} = M_\epsilon y + E \), where the \( f_i \) in \( M_\epsilon \) are given by \( f_i = f_{i0} + C \epsilon \). The variation of parameters method yields

\[
y(t) = e^{M_\epsilon} (y(0) + \int_0^t e^{-M_\epsilon s} ds) E
\]

\[
= e^{M_\epsilon} y(0) + S \int_0^t e^{(t-s)D} ds S^{-1} E
\]

\[
= e^{M_\epsilon} y(0) + S \left( \frac{e^{(f_i - b_i) \delta_{ij}}}{f_i - b_i \mu} \right) S^{-1} E - S \left( \frac{1}{f_i - b_i \mu} \right) S^{-1} E.
\]

(6.18)

Now we compute the order of magnitude of each term as in (6.10) in the proof of Lemma 6.1 and show that the two terms in (6.18) involving \( E \) are negligible with respect to the main term. Set

\[
e_1(t) := S \left( e^{(f_i - b_i \mu) \delta_{ij}} \right) S^{-1} E, \quad e_2 := S \left( \delta_{ij} \right) S^{-1} E.
\]

(6.19)

In the case \( m_{ij} = m_{ij}^{(2)} \) we have, for \( i \neq L \), from Lemma 6.1 that

\[
(e_1(t \log(1/\mu)) \vee e_2)_i = O(\mu^{\min_{\mu \in [0,1]} [t-\gamma+(2+\epsilon)0]_1} < \mu^{\gamma t}) = O(\mu^{\gamma t}),
\]

and if \( i = L \) we get

\[
(e_1(t \log(1/\mu)) \vee e_2)_L = O(\mu^{L-(f_{i0} + C \epsilon) + \epsilon}) = o(\mu^{L-(f_{i0} + C \epsilon)}).
\]

(6.20)

Consequently, proceeding as for the lower bounding ODE, we get

\[
\lim_{\mu \to 0} \frac{\log(x_\nu^\mu(t \log(1/\mu)))}{\log(1/\mu)} \leq \lim_{\mu \to 0} \frac{\log(y(t \log(1/\mu)))}{\log(1/\mu)}
\]

\[
= \begin{cases} -i + C \epsilon t, & \text{for } i = 0, \ldots, L - 1, \\ -L + t(f_{i,0} + C \epsilon), & \text{for } i = L. \end{cases}
\]

(6.22)
Finally observe that, as the only growing population is the one with trait \( L \),

\[
T_{L-1} = \lim_{\varepsilon \to 0} \lim_{\mu \to 0} \frac{\tau_{L-1}(\varepsilon, \mu)}{\log(1/\mu)} = \begin{cases} 2/f_{L0}, & \text{for } m_{ij} = m_{ij}^{(2)}, \\ L/f_{L0}, & \text{for } m_{ij} = m_{ij}^{(1)} \end{cases} \tag{6.23}
\]

In the case \( m_{ij} = m_{ij}^{(1)} \) the proof continues directly with Subsection 6.2

6.1.2. Time interval \( T_{L-1} \leq t \leq T_{L-2} \). Let \( m_{ij} = m_{ij}^{(2)} \) and

\[
\tau_{L-2}(\varepsilon, \mu) = \inf \{ t : \exists i \in [0, L-1] \text{ s.t. } x_i^\mu(t) > \mu^{1-\varepsilon} \}
\wedge \inf \{ t : |x_i^\mu(t) - \bar{x}_i| > \varepsilon \} \wedge \inf \{ t : x_i^{\mu L}(t) > \mu^{L-3+\varepsilon} \} \wedge \inf \{ t : x_i^{\mu L}(t) > \mu^{L-4+\varepsilon} \}
\]

and define

\[
T_{L-2} := \lim_{\varepsilon \to 0} \lim_{\mu \to 0} \frac{\tau_{L-2}(\varepsilon, \mu)}{\log(1/\mu)} \tag{6.25}
\]

There exists a positive \( C \) such that on the time interval \([\tau_{L-1}(\varepsilon, \mu), \tau_{L-2}(\varepsilon, \mu)]\),

\[
\frac{dx_i^\mu}{dt} \geq (f_{i0} - C\varepsilon)x_i^\mu + \mu \left( \frac{b_{i-1}}{2} x_{i-1}^{\mu} 1_{[i=L-1]} + \frac{b_{i+1}}{2} x_{i+1}^{\mu} 1_{[i=L-1]} - b_i x_i^\mu \right) \tag{6.26}
\]

Hence by the Gronwall lemma, and notations 6.3, \( x_i^\mu \) is bigger than the solution to

\[
\frac{dy}{dt} = \begin{pmatrix} M_{\text{left}}(L-2) & 0 \\ 0 & M_{\text{right}}(1) \end{pmatrix} y =: M'(L-2, 1)y \tag{6.27}
\]

where \( M_{\text{left}}(L-2) = M_2(L-2) \) with \( f_i = f_{i0} - C\varepsilon \) and

\[
M_{\text{right}} = \begin{pmatrix} f_{L-1,0} - C\varepsilon & \mu b_L \\ 0 & f_{L0} - C\varepsilon \end{pmatrix} \tag{6.28}
\]

Applying then twice Lemma 6.1 once with \( M_{\text{left}}(L-2) \) and \( y_{\text{left}} = (y_0, \ldots, y_{L-2}) \) and once with \( M_{\text{right}}(1) \) (treated as \( M(1) \) with “reversed indices”, i.e. \( f_i, b_i \) replaced by \( f_{L-i}, b_{L-i} \) and \( y_{\text{right}} = (y_{L-1}, y_L) \), with

\[
y(0) = (\bar{x}_0, \mu, \mu^2, \ldots, \mu^{L-1}, \mu^{L-2}) \tag{6.29}
\]

up to \( o_{\varepsilon}(1) \) terms in the powers of \( \mu \) due to the range of possible initial conditions coming from the previous phase (those however do not change anything to the calculations), we get

\[
\lim_{\mu \to 0} \frac{\log(x_i^\mu(t \log(1/\mu)))}{\log(1/\mu)} \geq \begin{cases} -i - C\varepsilon t, & \text{for } i = 0, \ldots, L-2, \\ -(L-1) + t(f_{L0} - C\varepsilon), & \text{for } i = L-1, \\ -(L-2) + t(f_{L0} - C\varepsilon), & \text{for } i = L. \end{cases} \tag{6.30}
\]

On the other hand, we have the upper bound

\[
\frac{dx_i^\mu}{dt} \leq (f_{i0} + C\varepsilon)x_i^\mu + \mu \left( \frac{b_{i-1}}{2} x_{i-1}^{\mu} 1_{[i=L-1]} + \frac{b_{i+1}}{2} x_{i+1}^{\mu} 1_{[i=L-1]} - b_i x_i^\mu \right) + E_i \tag{6.31}
\]

where until \( \tau_{L-2} \) we have

\[
E = \mu \cdot (\mu^{1-\varepsilon}, \mu^{2-\varepsilon}, \ldots, \mu^{L-2-\varepsilon}, \mu^{L-3+\varepsilon}, \mu^{L-2+\varepsilon}, \mu^{L-3+\varepsilon}) \tag{6.32}
\]

By the Gronwall lemma, \( x_i^\mu \) is smaller than the solution to \( \frac{dy}{dt} = M'y + E \) with \( f_i = f_{i0} + C\varepsilon \). Using the same method as above (variation of constants in the two blocks), we get (6.30) also as
an upper bound, with \( f_{L_0} - C \varepsilon \) replaced by \( f_{L_0} + C \varepsilon \), and \(-i - C \varepsilon t\) replaced by \(-i + C \varepsilon t\). Finally observe that,
\[
T_{L-2}^- = \lim_{\varepsilon \to 0} \lim_{\mu \to 0} \frac{\tau_{L-2}(\varepsilon, \mu)}{\log(1/\mu)} = \frac{4}{f_{L_0}}. \tag{6.33}
\]

6.1.3. Induction until \( T_{L/2}^- \). In this section if \( L \) is odd, then \( L/2 \) has to be replaced by \([L/2]\). For \( k \in \{3, \ldots, L/2\} \) we treat the time interval \( T_{L-k}^- \leq t \leq T_{L-(k+1)}^- \). Let \( m_{ij} = m_{ij}^{(2)} \) and
\[
\tau_{L-k}^-(\varepsilon, \mu) = \inf\{t : \exists i \in \left[0, L - k + 1\right] \text{ st } x_i^\mu(t) > \mu^{1-\varepsilon}\}
\]
with \( \inf\{t : |x_0^\mu(t) - \bar{x}_0| > \varepsilon\} \) and \( \inf\{t : \exists j \in \left[1, k\right] \text{ st } x_j^\mu(t) > \mu^{(L-k+j-2)\varepsilon}\}. \tag{6.34}
\]
and define
\[
T_{L-k}^- := \lim_{\varepsilon \to 0} \lim_{\mu \to 0} \frac{\tau_{L-k}^-(\varepsilon, \mu)}{\log(1/\mu)}. \tag{6.35}
\]
For \( t \in [\tau_{L-k}^-(\varepsilon, \mu), \tau_{L-(k-1)}^-(\varepsilon, \mu)] \) we have the lower bound
\[
\frac{dx_i^\mu}{dt} \geq (f_{l_0} - C \varepsilon)x_i^\mu + \mu \left( \frac{b_{l-1} - x_{i-1}^\mu}{2} 1_{\{i < L-k+1\}} + \frac{b_{l+1} - x_{i+1}^\mu}{2} 1_{\{i \geq L-k+1\}} - b_i x_i^\mu \right). \tag{6.36}
\]
Hence, by the Gronwall lemma, \( x^\mu \) is bigger than the solution to
\[
\frac{dy}{dt} = \begin{pmatrix} M_{\text{left}}(L - k) & 0 \\ 0 & M_{\text{right}}(k - 1) \end{pmatrix} y \quad \text{where} \quad M_{\text{left}}(L - k) = M_2(L - k) \quad \text{with} \quad f_i = f_{l_0} - C \varepsilon \quad \text{and}
\]
\[
M_{\text{right}}(k) = \begin{pmatrix} f_{L-k,0} - C \varepsilon & \frac{\mu}{2} b_{L-k+1} & \cdots & \frac{\mu}{2} b_L \\ 0 & f_{L-1,0} - C \varepsilon & \cdots & f_{L_0} - C \varepsilon \end{pmatrix}. \tag{6.38}
\]
Applying twice Lemma [6.1] once with \( M_{\text{left}} \) and \( y_{\text{left}} = (y_0, \ldots, y_{L-k-1}) \) and once with \( M_{\text{right}} \) (treated as \( M(k) \) with “reversed indices”, \( f_i, b_i \) replaced by \( f_{k-i}, b_{k-i} \)) and \( y_{\text{right}} = (y_{L-k}, \ldots, y_L) \), with
\[
y(0) = (\bar{x}_0, \mu, \mu^2, \ldots, \mu^{L-k}, \mu^{L-k+1}, \mu^{L-k-1}, \ldots, \mu^{L-2k}) \tag{6.39}
\]
up to \( o_\varepsilon(1) \) terms in the powers of \( \mu \) due to the range of possible initial conditions coming from the previous phase, we get
\[
\lim_{\mu \to 0} \frac{\log(x_i^\mu(t \log(1/\mu)))}{\log(1/\mu)} \geq \begin{cases} -i - C \varepsilon t, & \text{for } i = 0, \ldots, L - k, \\
-i + j - 1 + t(f_{l_0} - C \varepsilon), & \text{for } i = L - k + j, j = 1, \ldots, k. \end{cases} \tag{6.40}
\]
On the other hand, we have the upper bound
\[
\frac{dx_i^\mu}{dt} \leq (f_{l_0} + C \varepsilon)x_i^\mu + \mu \left( \frac{b_{l-1} - x_{i-1}^\mu}{2} 1_{\{i < L-k+1\}} + \frac{b_{l+1} - x_{i+1}^\mu}{2} 1_{\{i \geq L-k+1\}} - b_i x_i^\mu \right) + E_i, \tag{6.41}
\]
where on the time interval \([\tau_{L-k}^-(\varepsilon, \mu), \tau_{L-(k-1)}^-(\varepsilon, \mu)]\) we have
\[
E = \mu \cdot (\mu^{1-\varepsilon}, \mu^{2-\varepsilon}, \ldots, \mu^{L-k+1-\varepsilon}, \mu^{L-k-1+\varepsilon}, \mu^{L-k-2+\varepsilon}, \ldots, \mu^{L-1-(k-1)+\varepsilon}). \tag{6.42}
\]
By the Gronwall lemma, \( x^\mu \) is thus smaller than the solution to \( \frac{dx}{dt} = M' y + E \) with \( f_i = f_{i0} + C_\varepsilon \).

Using the same method as above (variation of the constant in the two blocks), we get (6.40) also as an upper bound, with \( f_{i0} - C_\varepsilon \) replaced by \( f_{i0} + C_\varepsilon \), and \(-i - C_\varepsilon t \) replaced by \(-i + C_\varepsilon t \). Finally, observe that

\[
T_{L-k}^- = \lim_{\varepsilon \to 0} \lim_{\mu \to 0} \frac{T_{L-k}^-(\varepsilon, \mu)}{\log(1/\mu)} = \frac{2k}{f_{i0}}.
\] (6.43)

6.2. The swap. Let \( m_{ij} = m_{ij}^{(2)} \) and

\[
\tau^\varepsilon(\varepsilon, \mu) = \inf \{ t : \exists i \in [0, L/2] \text{ s.t. } x_i^\mu(t) > \mu^{1-\varepsilon} \} \wedge \inf \{ t : x_0^\mu(t) < \varepsilon \} \wedge \inf \{ t : x_1^\mu(t) > \mu^{1-\varepsilon} \} \wedge \inf \{ t : x_L^\mu(t) > \bar{x}_L - \varepsilon \},
\] (6.44)

or \( m_{ij} = m_{ij}^{(1)} \) and

\[
\tau^\varepsilon(\varepsilon, \mu) = \inf \{ t : \exists i \in [0, L - 1] \text{ s.t. } x_i^\mu(t) > \mu^{1-\varepsilon} \} \wedge \inf \{ t : x_0^\mu(t) < \varepsilon \} \wedge \inf \{ t : x_L^\mu(t) > \bar{x}_L - \varepsilon \},
\] (6.45)

For \( t \in [\tau_{L/2}^-(\varepsilon, \mu), \tau^\varepsilon(\varepsilon, \mu)] \) and \( \chi \in [0, L] \) we have the lower bounds

\[
\frac{dx_i^\mu}{dt} \geq (b_\chi - d_\chi - c_{i0}x_0^\mu - c_{iL}x_L^\mu - C_\mu)x_i^\mu - C'_\mu 1^{\varepsilon},
\] (6.46)

and the upper bounds

\[
\frac{dx_i^\mu}{dt} \leq (b_\chi - d_\chi - c_{i0}x_0^\mu - c_{iL}x_L^\mu)x_i^\mu + C'_\mu 1^{\varepsilon}.
\] (6.47)

Let \((x_0, x_L)\) denote the solution of the unperturbed system, i.e. of

\[
\frac{dx_i^\mu}{dt} = (b_\chi - d_\chi - c_{i0}x_0^\mu - c_{iL}x_L^\mu)x_i^\mu.
\] (6.48)

By (3.1) and (3.2), we know that this system has a unique stable equilibrium \((0, \bar{x}_L)\). Moreover, the time needed to enter an \( \varepsilon \)-neighbourhood of this equilibrium from initial conditions \((\bar{x}_0 - \varepsilon, \varepsilon)\) is of order \( O(1) \). Using the Gronwall lemma, (6.46) and (6.47) imply that on any compact time interval \((x_0^\mu, x_L^\mu) \to (x_0, x_L)\). Moreover, for all \( \mu \) small enough, the system \((x_i^\mu, x_L^\mu)\) has a stable equilibrium that converges to \((0, \bar{x}_L)\), as \( \mu \downarrow 0 \).

For the populations \((x_0^\mu, \ldots, x_{L-1}^\mu)\), we have, for \( t \in [\tau_{L/2}^-(\varepsilon, \mu), \tau^\varepsilon(\varepsilon, \mu)] \), the lower bounds

\[
\frac{dx_i^\mu}{dt} \geq (b_i - d_i - c_{i0}\bar{x}_0 - c_{iL}\bar{x}_L - C_\varepsilon)x_i^\mu + \mu \left( b_{i-1} x_{i-1}^\mu 1_{[i \leq L/2]} + b_{i+1} x_{i+1}^\mu 1_{[i \geq L/2]} - b_i x_i^\mu \right),
\] (6.49)

in the case \( m_{ij} = m_{ij}^{(2)} \) and

\[
\frac{dx_i^\mu}{dt} \geq (b_i - d_i - c_{i0}\bar{x}_0 - c_{iL}\bar{x}_L - C_\varepsilon)x_i^\mu + \mu \left( b_{i-1} x_{i-1}^\mu - b_i x_i^\mu \right),
\] (6.50)

in the case \( m_{ij} = m_{ij}^{(1)} \). We have decoupled traits 0 and \( L \) from traits 1, \ldots, \( L - 1 \). We still have to show that the functions \((x_i^\mu, i \in [1, L - 1]) \) stay smaller than \( \mu^{1-\varepsilon} \). By the Gronwall lemma, the following holds:
(1) in the case \( m_{ij} = m_{ij}^{(1)} \), the solution \( (x_1^\mu, \ldots, x_{L-1}^\mu) \) is smaller than the solution to

\[
\frac{dy}{dt} = M_1(L - 2)y
\]

with \( f_i = b_i - d_i - c_i\tilde{x}_0 - c_iL\tilde{x}_L - C\varepsilon < 0 \) and initial conditions:

\[
y(0) = (\mu, \mu^2, \ldots, \mu^{L-1}).
\]

up to \( O(1) \) terms in the powers of \( \mu \) due to the range of possible initial conditions coming from the previous phase. Applying Lemma 6.1, we get

\[
\lim_{\mu \to 0} \frac{\log(x_i^\mu(t \log(1/\mu)))}{\log(1/\mu)} \geq -i + t \sup_{i \leq n} f_i, \quad \text{for } i = 1, \ldots, L - 1.
\]

But we just mentioned that the swap has a duration of order 1. Thus, the \( t \) to be considered is negligible with respect to 1, and

\[
\lim_{t \to 0} \lim_{\mu \to 0} \frac{\log(x_i^\mu(t \log(1/\mu)))}{\log(1/\mu)} \geq -i, \quad \text{for } i = 1, \ldots, L - 1.
\]

(2) in the case \( m_{ij} = m_{ij}^{(2)} \), the solution \( (x_1^\mu, \ldots, x_{L-1}^\mu) \) is thus smaller than the solution to

\[
\frac{dy}{dt} = \begin{pmatrix} M_{\text{left}}(L/2 - 2) & 0 \\ 0 & M_{\text{right}}(L/2 - 1) \end{pmatrix} y =: M'(L/2 - 2, L/2 - 1)y
\]

with \( f_i = b_i - d_i - c_i\tilde{x}_0 - c_iL\tilde{x}_L - C\varepsilon < 0 \) and initial conditions

\[
y(0) = (\mu, \mu^2, \ldots, \mu^{L/2 + 1}, \mu^{L/2}, \mu^{L/2-1} \ldots, \mu).
\]

up to \( O(1) \) terms in the powers of \( \mu \) due to the range of possible initial conditions coming from the previous phase. Here, if \( L \) is odd, then the initial condition has to be replaced by

\[
y(0) = (\mu, \mu^2, \ldots, \mu^{L/2+1}, \mu^{L/2}, \mu^{L/2-1} \ldots, \mu).
\]

and matrix dimensions have to be modified accordingly, but the proof stays the same. Applying Lemma 6.1 twice (in the two blocks), and letting \( t \) go to 0 as the swap has a duration of order 1, we get

\[
\lim_{t \to 0} \lim_{\mu \to 0} \frac{\log(x_i^\mu(t \log(1/\mu)))}{\log(1/\mu)} \geq -\min[i, L - i], \quad \text{for } i = 1, \ldots, L - 1.
\]

On the other hand, we have the upper bound

\[
\frac{dx_i^\mu}{dt} \leq F x_i^\mu + \mu^{1-\varepsilon},
\]

with some \( F > 0 \). Thus, by the Gronwall lemma,

\[
x_i^\mu(t) \leq \mu^{-\varepsilon}, \quad \text{for } t \leq \frac{\varepsilon}{F} \log(1/\mu) \text{ and for } i = 1, \ldots L/2.
\]

and similarly for \( i = L/2, \ldots, L - 1 \) (no population changes its order of magnitude of more than \( \varepsilon \) during any time of order \( O(1) \)). We deduce that, for \( i = 1, \ldots, L - 1 \),

\[
\lim_{\mu \to 0} \frac{\log(x_i^\mu(t \log(1/\mu)))}{\log(1/\mu)} \leq \begin{cases} -\min[i, L - i] & \text{if } m_{ij} = m_{ij}^{(2)} \\ -i & \text{if } m_{ij} = m_{ij}^{(1)} \end{cases}
\]
The duration of the swap vanishes (on the time scale $\log(1/\mu)$) in the limit $\mu \to 0$. We thus have $T^* = T^{*-}_{L/2}$.

6.3. **After the swap.**

6.3.1. **Case $m_{ij} = m^{(1)}_{ij}$**. Let

$$\tau^*(\varepsilon, \mu) = \inf \{ t : \exists i \in [1, L - 1] \text{ s.t. } x_i^\varepsilon(t) > \mu^{i-2} \} \land \inf \{ t : |x_i^\varepsilon(t) - \bar{x}_L| > \varepsilon \}. \quad (6.62)$$

For $t \in [\tau^*(\varepsilon, \mu), \tau^*(\varepsilon, \mu)]$, we have the lower bound

$$\frac{dx_i^\varepsilon}{dt} \geq (f_{iL} - C\varepsilon) x_i^\varepsilon + \mu \left( b_{i-1} x_{i-1}^\varepsilon - b_i x_i^\varepsilon \right). \quad (6.63)$$

Hence, by the Gronwall lemma, and notations (6.3), $x_i^\varepsilon$ is bigger than the solution to $\frac{dy}{dt} = M_1 y$ with $f_i = f_{iL} - C\varepsilon$. Applying Lemma 6.1 with $y(0) = (\varepsilon, \mu, \ldots, \mu^{L-1}, \bar{x}_L - \varepsilon)$,

$$\text{up to } o_\varepsilon(1) \text{ terms in the powers of } \mu \text{ due to the range of possible initial conditions coming from the previous phase (and thus } p_i = i - L\delta_{iL} \text{), we get, using Assumption 1}

\begin{align*}
\lim_{\mu \to 0} \frac{\log(x_i^\varepsilon(t \log(1/\mu)))}{\log(1/\mu)} &\geq -\min_{a \leq i} \{ i - L\delta_{iL} - t(f_{al} - C\varepsilon) \} \\
&= -i + L\delta_{iL} - t \min_{a \in [0, L]} |f_{al}| + tC\varepsilon. \quad (6.65)
\end{align*}

In the same way we get the corresponding upper bound with $f_{al} - C\varepsilon$ replaced by $f_{al} + C\varepsilon$.

6.3.2. **Case $m_{ij} = m^{(2)}_{ij}$**. In this phase the system cannot be approximated by a piece-wise block-triangular linear system anymore. Let us study the ODE followed by the rescaled process. Let

$$\tau^*(\varepsilon, \mu) = \inf \{ t > T^* : \exists i \in [0, L - 1] : x_i^\varepsilon(t) > \varepsilon \} \land \inf \{ t : |x_i^\varepsilon(t) - \bar{x}_L| > \varepsilon \}. \quad (6.66)$$

For $t \in [\tau^*(\varepsilon, \mu), \tau^*(\varepsilon, \mu)]$ we have the lower bound

$$\frac{dx_i^\varepsilon}{dt} \geq (f_{iL} - C\varepsilon) x_i^\varepsilon + \mu \left( b_{i-1} x_{i-1}^\varepsilon + b_{i+1} x_{i+1}^\varepsilon - b_i x_i^\varepsilon \right). \quad (6.67)$$

and a similar upper bound where $f_{iL} - C\varepsilon$ is replaced by $f_{iL} + C\varepsilon$. Let

$$\bar{x}_i^\varepsilon := \frac{\log \left[ x_i^\varepsilon \left( t \cdot \log \left( \frac{1}{\mu} \right) \right) \right]}{\log(1/\mu)}. \quad (6.68)$$

We thus have

$$\frac{d\bar{x}_i^\varepsilon}{dt} \geq f_{iL} - C\varepsilon - \mu + \frac{b_{i-1}}{2} \mu^{1+\bar{x}_i^\varepsilon - \bar{x}_{i-1}^\varepsilon} + \frac{b_{i+1}}{2} \mu^{1+\bar{x}_i^\varepsilon - \bar{x}_{i+1}^\varepsilon} \quad (6.69)$$

and a similar upper bound, with initial condition (we reset the time of the swap to 0 from now on):

$$\bar{x}_i^\varepsilon(0) = \left( \frac{\log \varepsilon}{\log(1/\mu)}, -1, -2, \ldots, -L/2, -L/2 + 1, \ldots, -1, \frac{\log(\bar{x}_L - \varepsilon)}{\log(1/\varepsilon)} \right). \quad (6.70)$$
up to $o_{x}(1)$ terms due to the range of possible initial conditions coming from the previous phase. Here, if $L$ is odd, then the initial condition has to be replaced by

$$\tilde{x}^{\mu}(0) = \left(\frac{\log \epsilon}{\log(1/\mu)}, -1, -2, \ldots, -\lfloor L/2 \rfloor, -\lfloor L/2 \rfloor, \ldots, -1, \frac{\log(x_L - \epsilon)}{\log(1/\mu)}\right),$$

(6.71)

but the proof idea stays the same. Let $\delta > 0$ and $T^{\mu}_{\text{out}}(\delta, \mu) := \inf\{t > 0 : \tilde{x}^{\mu}_{t} > \tilde{x}^{\mu}_{t-1} - (1 - \delta) \text{ or } \tilde{x}^{\mu}_{t} > \tilde{x}^{\mu}_{t+1} - (1 - \delta)\}$. Then, for $t \in [0, T^{\mu}_{\text{out}}]$, that is, when $\tilde{x}^{\mu}_{t}$ is above one of its neighbours minus $1 - \delta$, then, for $\mu$ and $\delta$ small enough, the slope of $\tilde{x}^{\mu}_{t}$ is prescribed by the fitness of trait $i$ with respect to trait $L$ (up to a multiple of $\epsilon$). Indeed, by (6.69):

$$\frac{d\tilde{x}^{\mu}_{t}}{dt} \geq f_{L} - C\epsilon - \mu \geq f_{L} - 2C\epsilon,$$

(6.72)

$$\frac{d\tilde{x}^{\mu}_{t}}{dt} \leq f_{L} + C\epsilon - \mu + \frac{b_{L-1}}{2}\mu^{\delta} + \frac{b_{L+1}}{2}\mu^{\delta} \leq f_{L} + 2C\epsilon.$$

(6.73)

Let $T^{\text{out}}(\delta, \mu) := \inf\{t > T^{*} : \tilde{x}^{\mu}_{t} < \tilde{x}^{\mu}_{t-1} - (1 + \delta) \text{ or } \tilde{x}^{\mu}_{t} < \tilde{x}^{\mu}_{t+1} - (1 + \delta)\}$, we call it the exit time of the security region. Let us show that, for $\mu$ small enough, we have $T^{\text{out}} = \infty$, for all $i \in \{0, \ldots, L\}$. Assume by contradiction that $\inf\{T^{\text{out}}_{i}, i \in \{0, \ldots, L\}\} < \infty$.

Among the indices $i$ achieving the infimum, consider the one such that $\tilde{x}^{\mu}_{T^{\text{out}}}(\tilde{x}^{\mu}_{T^{\text{out}}})$ is maximal, that is $\tilde{x}^{\mu}_{T^{\text{out}}}$ is the highest population among those which exit the security region first. By continuity of the solutions, at $t = T^{\text{out}}^{\mu}$ we have $\min\{1 + \tilde{x}^{\mu}_{i} - \tilde{x}^{\mu}_{i+1}, 1 + \tilde{x}^{\mu}_{i} - \tilde{x}^{\mu}_{i-1}\} = -\delta$. Suppose that $\tilde{x}^{\mu}_{T^{\text{out}}}$ exits its security region by falling below its left neighbour minus one, i.e:

$$1 + \tilde{x}^{\mu}_{i} - \tilde{x}^{\mu}_{i-1} = -\delta \quad \text{and} \quad 1 + \tilde{x}^{\mu}_{i} - \tilde{x}^{\mu}_{i+1} > -\delta,$$

(6.74)

the two other possibilities (right neighbour or both) are similar. By (6.69), for $\mu$ small enough:

$$\frac{d\tilde{x}^{\mu}_{i}}{dt}(T^{\text{out}}_{i}) \geq f_{i} - C\epsilon - \mu + \min\left\{\frac{b_{i-1}}{2}, \frac{b_{i+1}}{2}\right\}\mu^{\delta} = O(\mu^{\delta}),$$

(6.75)

$$\frac{d\tilde{x}^{\mu}_{i-1}}{dt}(T^{\text{out}}_{i}) = f_{i-1} - C\epsilon - \mu + \frac{b_{i-2}}{2}\mu^{1+\tilde{x}^{\mu}_{i-1} - \tilde{x}^{\mu}_{i-2}} + \frac{b_{i+1}}{2}\mu^{2+\delta} \leq \frac{b_{i-2}}{2}\mu^{\delta} + \frac{b_{i+1}}{2}\mu^{2+\delta} = O(\mu^{\delta}).$$

(6.76)

where the upper bounds in (6.76) come from the assumption that $\tilde{x}^{\mu}_{i}$ is the highest population among these exiting their security region. Indeed, if in (6.76) we had $1 + \tilde{x}^{\mu}_{i-1} - \tilde{x}^{\mu}_{i-2} = -\delta$ then, by definition $\tilde{x}^{\mu}_{i-1}$ would exit its security region, thus we would have $\tilde{x}^{\mu}_{T^{\text{out}}_{i}} \leq \tilde{x}^{\mu}_{i}$, which contradicts (6.74). Thus, $1 + \tilde{x}^{\mu}_{i-1} - \tilde{x}^{\mu}_{i-2} > -\delta$. The equations (6.75) and (6.76) imply that the derivative $\frac{d}{dt}(\tilde{x}^{\mu}_{i} - \tilde{x}^{\mu}_{i-1})(T^{\text{out}})$ is as large as needed. Thus again by the continuity of the solutions, this implies the existence of some $t^{*} < t$ such that $\tilde{x}^{\mu}_{i}(t^{*}) < \tilde{x}^{\mu}_{i-1}(t^{*}) - (1 + \delta)$. Hence $t^{*} < T^{\text{out}}$.

This allows us to describe the limit of $\tilde{x}^{\mu}_{i}$ as $\mu \to 0$. A helpful example is given on Figure 4. First, as $f_{LL} = 0$, equation (6.72) and the previous reasoning imply that until $T^{*}(\epsilon, \mu)$:

$$\frac{d\tilde{x}^{\mu}_{L}}{dt} \geq -C\epsilon \quad \text{and} \quad \frac{d\tilde{x}^{\mu}_{L}}{dt} \leq +C\epsilon,$$

(6.77)

which implies that $\tilde{x}^{\mu}_{L} \to 0$ (take the limits $\epsilon \to 0$ after $\mu \to 0$). Now the initial condition (6.70) and Assumption 1 imply that $\tilde{x}^{\mu}_{L} \to -(L - i)$, for $i = L/2, \ldots, L - 1$. Indeed, $\tilde{x}^{\mu}_{L}$ is close to 0, and
\( x_i(0) = -(L - i) \) for those indices, so the only possibility to maintain a difference of less than one with their nearest neighbours and having a negative fitness \( f_{i,L} \) is to stay constant. The shape of the first \( L/2 \)-coordinates of the process is less trivial to formulate: each \( x_i^\mu \) behaves piecewise linearly in the limit \( \mu \to 0 \) and given the sequence \( (f_{0L}, \ldots, f_{L-1,L}) \) one can construct the successive slopes by following the rule “\( x_i^\mu \) tries to decay with slope \( f_{iL} \) while being at distance at most 1 of \( x_{i-1}^\mu \) and \( x_{i+1}^\mu \); if it is not possible then it stays parallel to the largest of its neighbours, either \( x_{i-1}^\mu \) or \( x_{i+1}^\mu \).”

More precisely, consider the sequence \( \{i_1, \ldots, i_L\} \) of “fitness records” defined recursively by \( i_1 = 0, i_k = \min\{i \in [0, L - 1] : f_{iL} < f_{i_{k-1}L}\} \). Then the previous reasoning implies that, for any \( \varepsilon > 0 \), as \( \mu \to 0 \), the process \( \langle \tilde{x}^\mu(t) \rangle_{t>0} \), starting with initial condition \((6.70)\), stays in an \( \varepsilon \)-neighbourhood of the deterministic process \( x(t) \) given by:

\[
x_i(t) = -(L - i) \vee \max_{k \in [0, L]} \{-i - |f_{iL}|t\} \vee \max_{k \in [1, L]} \{-i_k - |i - i_k| - |f_{i_kL}|t\}. \tag{6.78}
\]

Once again, Figure 4 provides a helpful example to compute the formula.

\[\square\]

7. Proofs of Theorems 3.2 and 3.3

In this section we focus on mutation probabilities scaling as a negative power of \( K \) times a function whose fluctuations are much smaller than the fluctuations of a power of \( K \) (recall (3.12)).

7.1. Poisson representation. In the vein of Fournier and Méléard [28], we represent the population process in terms of Poisson measures. Let \( (Q^{(b)}_k, Q^{(m)}_k, Q^{(d)}_k, 0 \leq k \leq L) \) be independent Poisson random measures on \( \mathbb{R}^2_+ \) with intensity \( dsd\theta \), and recall that \( (e_i)_{0 \leq i \leq L} \) is the canonical basis of \( \mathbb{R}^{L+1} \). We decompose on possible jumps that may occur: births without mutation, birth with mutation and deaths of individuals. For simplicity, we write

\[
d^i_k(x) = D^i_k(x)x_i = \left(d_i + \sum_{j=0}^L c_{ij} K \right) x_i, \tag{7.1}
\]

for the total death rate of the sub-population \( i \). Recall that in this regime, we only consider the mutation kernel \( m_{ij}^{(1)} = \mu \delta_{i+1,j} \). The process \( X \) admits the following representation. For every real-valued function \( h \) on \( \mathbb{R}^{L+1}_+ \) such that \( h(X(t)) \) is integrable,

\[
h(X(t)) = h(X(0)) + \sum_{k=0}^L \int_0^{t'} \int_{\mathbb{R}^+} (h(X(s^-) + e_k) - h(X(s^-))) \mathbf{1}_{\theta \leq 1 - \rho b_k x_k(x)} Q^{(b)}_k(ds, d\theta) + \sum_{k=1}^L \int_0^{t'} \int_{\mathbb{R}^+} (h(X(s^-) + e_k) - h(X(s^-))) \mathbf{1}_{\theta \leq \rho b_{k-1} x_{k-1}(x)} Q^{(m)}_k(ds, d\theta) + \sum_{k=0}^L \int_0^{t'} \int_{\mathbb{R}^+} (h(X(s^-) - e_k) - h(X(s^-))) \mathbf{1}_{\theta \leq d_k^i(x(s^-)) x_k(x)} Q^{(d)}_k(ds, d\theta). \tag{7.2}
\]

Let us now introduce a finite subset of \( \mathbb{N} \) containing the equilibrium size of the 0-population,

\[
I^K_\varepsilon := \left[ K\left(\tilde{x}_0 - 2\varepsilon \frac{\sup_{1 \leq i \leq L} c_{0i}}{c_{00}}\right), K\left(\tilde{x}_0 + 2\varepsilon \frac{\sup_{1 \leq i \leq L} c_{0i}}{c_{00}}\right) \right] \cap \mathbb{N}, \tag{7.3}
\]

CROSSING A FITNESS VALLEY 21
and the stopping times $T^K_e$ and $S^K_e$, which denote respectively the hitting time of $[\varepsilon K]$ by the total mutant population ($X_1 + \ldots + X_L$) and the exit time of $I^K_e$ by the resident 0-population,

$$T^K_e := \inf \{ t \geq 0, \sum_{i \leq i \leq L} X_i(t) = [\varepsilon K] \}, \quad S^K_e := \inf \{ t \geq 0, X_0(t) \notin I^K_e \}. \quad (7.4)$$

As shown in [14], we know that as long as the total mutant population size is smaller than $\varepsilon K$, the resident population size stays close to its monomorphic equilibrium with a probability close to 1 (see Lemma [A.1]). This is a fundamental property of the population process, as it implies that the populations live in an almost constant environment and are subject to an almost constant competitive pressure from other individuals, $c_0 x_0$. This will allow us to couple $i$-population sizes ($1 \leq i \leq L - 1$) with subcritical branching processes with migration $X_i^{(-)}$ and $X_i^{(+)}$ to control their dynamics. Moreover, after the first growing phase for the $L$-population, if the sum of the $1$- to $(L - 1)$-mutant population sizes stays smaller than $\varepsilon K$ whereas the $L$-mutant population size exceeds the size $\varepsilon K$, the 0 and $L$ populations will behave as if they were the only ones in competition. As a consequence, the remaining time needed for the $L$-population to replace the 0-population will be close to $\log K/|f_{00}|$ (see, for instance, [14] and later in this paper for a precise statement). Hence, the first step consists in estimating the time needed for the mutant population to hit the size $[\varepsilon K]$. There are essentially two possibilities:

- Either $K \mu^L \gg 1$; in this case there is a (large) number of order $K \mu^L$ of $L$-type individuals. Hence the outcome is similar to a large resident population producing recurrently favourable mutants, studied in details in [46]. The fixation time of the trait $L$ will be of order $\log K$, and we will provide couplings with appropriate birth-death processes (without competition) with immigration to control the subpopulation sizes.

- Or $K \mu^L \ll 1$; in this case some of the mutant-population size dynamics consist in small excursions separated with periods with no individual. Indeed, the $i$-population with $i \leq \alpha$ is again well approximated by a birth-death process (without competition) with immigration, which is close to the deterministic limit, while, for the $i$-population with $i > \alpha$, the immigration term is not large enough and the population is well described, at each arrival of a single mutant, by a subcritical birth-death process. Each of the excursions of the sum of populations $i \in [\alpha, L - 1]$ has the same probability to produce a $L$-mutant which may generate a large population and invade. In this case, the time to invasion will be close to a geometric random variable, with a mean of order $1/(K \mu^L)$ much larger than $\log K$.

### 7.2. Proof of Theorem 3.2

The time needed for the favourable mutation to invade the population is very dependent on the mutation probability per reproductive event, $\mu$.

To study the case when $K \mu^L \gg 1$, we will couple each population size $X_i$, $0 \leq i \leq L - 1$ with two processes such that, for every $0 \leq i \leq L - 1$ and $t \leq T^K_e \wedge S^K_e$,

$$X_i^{(-)}(t) \leq X_i(t) \leq X_i^{(+)}(t), \quad a.s. \quad (7.5)$$

By definition of the population process in (7.2) and of the stopping times $T^K_e$ and $S^K_e$ in (7.4), the following processes satisfy (7.5):

$$X_0^{(\pm)}(t) = K \left( \bar{x}_0 \pm 2\varepsilon \sup_{1 \leq i \leq L} \frac{c_0(t)}{c_0} \right) =: \bar{x}_0^{(\pm)} K, \quad (7.6)$$
and, for $1 \leq i \leq L - 1$,
\[
X_i^{(-)}(t) = \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq (1-\mu) b_i X_i'(s)} Q_k(b)(ds, d\theta) + \int_0^t \int_{\mathbb{R}_+} 1_{\theta \geq \mu b_{i-1} X_i'(s)} Q_k^m(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq (d_i + \epsilon \sup_{1 \leq j \leq L} c_{ij}) X_i'(s)} Q_k^d(ds, d\theta),
\]
where we used the same Poisson measures as in (7.2). Note that from this representation, we get directly the classical semi-martingale decomposition for $X_i^{(-)}$ and $X_i^{(+)}$: for $* \in \{-, +\}$,
\[
X_i^{(*)}(t) = M_i^{(*)}(t) + A_i^{(*)}(t),
\]
where $M_i^{(*)}$ is a square integrable martingale and $A_i^{(*)}$ is a finite variation process, namely
\[
M_i^{(*)}(t) = \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq (1-\mu) b_i X_i'(s)} (Q_k(b)(ds, d\theta) - ds d\theta) + \int_0^t \int_{\mathbb{R}_+} 1_{\theta \geq \mu b_{i-1} X_i'(s)} (Q_k^m(ds, d\theta) - ds d\theta) - \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq (d_i + \epsilon \sup_{1 \leq j \leq L} c_{ij}) X_i'(s)} (Q_k^d(ds, d\theta) - ds d\theta),
\]
and the same expression for $M_i^{(*)}$ and $A_i^{(*)}$ by replacing the $(-)$’s by $(+)$’s and the terms
\[
d_i + c_{i0} x_0^{(*)} + \epsilon \sup_{1 \leq j \leq L} c_{ij}
\]
by
\[
d_i + c_{i0} x_0^{(-)}.
\]
Finally, we recall the expression of the quadratic variation of $M_i^{(*)}$,
\[
\langle M_i^{(*)} \rangle_t = \mu b_{i-1} \int_0^t X_i^{(*)}(s) ds + (1-\mu)b_i + d_i + c_{i0} x_0^{(*)} + \epsilon \sup_{1 \leq j \leq L} c_{ij} \int_0^t X_i^{(*)}(s) ds,
\]
and the one of $M_i^{(*)}$ is obtained by similar modifications as before.

Let us now introduce, for $1 \leq k \leq L - 1$, the following notations:
\[
s_{k0}^{(*)} := (1-\mu)b_k - d_k - c_{k0} x_0^{(*)}, \quad \text{and} \quad s_{k0}^{(-)} := (1-\mu)b_k - d_k - c_{k0} x_0^{(-)} + \epsilon \sup_{1 \leq j \leq L} c_{kj},
\]
as well as, for $* \in \{-, +\}:
\[
x_k^{(*)} := (1 \cdot \epsilon)^{k} b_0 \ldots b_{k-1} x_0^{(*)} \mu^k s_{10}^{(*)} \ldots s_{k0}^{(*)} \text{ and } \epsilon^{(k)} := \frac{|\ln \epsilon|}{s_{k0}^{(-)}}.
\]
Notice that $s_{k_0}^{(+)} \leq s_{k_0}^{(-)}$ and that $s_{k_0}^{(+)}$ and $s_{k_0}^{(-)}$ are positive, for $\varepsilon$ small enough, by Assumption (A2).

**Lemma 7.1.** Let $(t_K)_{K \in \mathbb{N}}$ be a sequence that tends to infinity, and let $\eta(K)$ be defined as follows:

$$
\eta(K) := \frac{1}{K\mu} + \sup_{1 \leq i \leq L-1} e^{-\varepsilon t_i^*_K}.
$$

(7.14)

Then there exist an event, $\mathcal{E}_K^*$, and $C < \infty$, independent of $K$, such that, for all $\varepsilon > 0$,

$$
\mathbb{P}(\mathcal{E}_K^*) \geq 1 - C\frac{\eta(K)}{\varepsilon^2},
$$

(7.15)

and that, for every $1 \leq i \leq L - 1$,

$$
X_i^{(-)}(t) \leq X_i^{(-)}(s), \quad t_i^{(1)} + \ldots + t_i^{(i)} \leq t \quad \text{a.s. on } \mathcal{E}_K^*,
$$

and

$$
X_i^{(+)}(s) \leq X_i^{(+)}(t), \quad 0 \leq s \leq t_K \quad \text{a.s. on } \mathcal{E}_K^*.
$$

Notice that by assumption, $L \geq 2$ and $K\mu^L \gg 1$. Hence, $K\mu \gg 1$ and $\eta(K) \to 0$ as $K \to \infty$.

**Proof.** From (7.10) and Definition (7.13), we get, for $* \in \{-, +\}$, that

$$
\frac{d}{dt} \mathbb{E}[X_1^{(*)}(t)] = b_0x_0^{(*)}K\mu - s_{10}^{(*)}\mathbb{E}[X_1^{(*)}(t)].
$$

This equation can be solved and this yields:

$$
\mathbb{E}[X_1^{(*)}(t)] = \frac{b_0x_0^{(*)}K\mu}{s_{10}^{(*)}} \left(1 - e^{-s_{10}^{(*)}t}\right).
$$

In particular, for every $t \geq 0$,

$$
\mathbb{E}[X_1^{(*)}(t)] \leq \frac{b_0x_0^{(*)}K\mu}{s_{10}^{(*)}},
$$

(7.16)

and, for every $t \leq t_i^{(1)}$,

$$
\mathbb{E}[X_1^{(*)}(t)] \geq \frac{b_0x_0^{(*)}K\mu}{s_{10}^{(*)}} (1 - e).
$$

(7.17)

Similarly, we get

$$
\frac{d}{dt} \mathbb{E}[(X_1^{(*)})^2(t)] = b_0x_0^{(*)}K\mu + \left(2b_0x_0^{(*)}K\mu + 2(1 - \mu)b_1 + s_{10}^{(*)}\right) \mathbb{E}[X_1^{(*)}(t)] - 2s_{10}^{(*)}\mathbb{E}[(X_1^{(*)})^2(t)]
$$

$$
= b_0x_0^{(*)}K\mu \left(1 + \left(2b_0x_0^{(*)}K\mu + 2(1 - \mu)b_1 + s_{10}^{(*)}\right) \frac{1 - e^{-s_{10}^{(*)}t}}{s_{10}^{(*)}}\right) - 2s_{10}^{(*)}\mathbb{E}[(X_1^{(*)})^2(t)].
$$

(7.18)

This equation has an explicit solution which is given by

$$
\mathbb{E}[(X_1^{(*)})^2(t)] = \left(\frac{b_0x_0^{(*)}K\mu}{s_{10}^{(*)}}\right)^2 \left[ 1 + \frac{(1 - \mu)b_1 + s_{10}^{(*)}}{b_0x_0^{(*)}K\mu} \right] (1 - e^{-2s_{10}^{(*)}t}) + 2 \left(1 + \frac{(1 - \mu)b_1}{b_0x_0^{(*)}K\mu}\right) \left(e^{-s_{10}^{(*)}t} - e^{-2s_{10}^{(*)}t}\right).
$$

(7.19)
As a consequence, we deduce that

\[
\text{Var}[X^{(s)}_1(t)] = \left( \frac{b_0x^{(s)}_{10}}{s^{(s)}_{10}} \right)^2 \left[ (1 - \mu)b_1 + s^{(s)}_{10} \right] \frac{1}{b_0x^{(s)}_{10} K \mu}
+ 2 \left( 2 + \frac{(1 - \mu)b_1}{b_0x^{(s)}_{10} K \mu} \right) e^{-s^{(s)}_{10}} + \left( \frac{(1 - \mu)b_1 + s^{(s)}_{10}}{b_0x^{(s)}_{10} K \mu} \right) \left[ 1 + \frac{(1 - \mu)b_1}{b_0x^{(s)}_{10} K \mu} \right] e^{-2s^{(s)}_{10}}. \tag{7.20}
\]

Hence, applying Doob’s martingale inequality to the sub-martingale

\[
Y^{(s)}(t) := \left( \frac{X^{(s)}_1(t)}{\mathbb{E}[X^{(s)}_1(t)]} - 1 \right)^2,
\]

we get, for every positive \( t_K \),

\[
\mathbb{P} \left( \sup_{t \leq t_K} Y^{(s)}(s) > \varepsilon^2 \right) \leq \frac{\text{Var}[X^{(s)}_1(t_K)]}{\varepsilon^2 \mathbb{E}^2[X^{(s)}_1(t_K)]} \leq C \frac{\eta(K)}{\varepsilon^2}, \tag{7.21}
\]

where \( C \) is a finite constant. Now let us introduce the events

\[
A_e^{(1,K)} := \left\{ X^{(s)}_1(s) \leq (1 + \varepsilon)^2 \frac{b_0x^{(s)}_{10} K}{s^{(s)}_{10}}, \forall s \leq t_K \right\},
\]

and

\[
B_e^{(1,K)} := \left\{ X^{(s)}_1(s) \geq (1 - \varepsilon)^2 \frac{b_0x^{(s)}_{10} K}{s^{(s)}_{10}}, \forall s \leq t_K \right\}.
\]

Equations (7.16), (7.17) and (7.21) yield

\[
\mathbb{P} \left( A_e^{(1,K)} \right) \geq 1 - C \frac{\eta(K)}{\varepsilon^2} \quad \text{and} \quad \mathbb{P} \left( B_e^{(1,K)} \right) \geq 1 - C \frac{\eta(K)}{\varepsilon^2}.
\]

The idea is then to use the bound on \( X_1 \) to bound \( X_2 \), and so on. More precisely, recall (7.12) and (7.13). By definition, on the event \( A_e^{(1,K)} \), for every \( t \leq t_K \),

\[
X^{(s)}_2(t) \leq \int_{t}^{t'} \int_{\mathbb{R}^+} \mathbf{1}_{\theta \leq (1 - \mu)b_2 x^{(s)}_2(s)} Q^{(b)}_2(ds,d\theta) + \int_{t}^{t'} \int_{\mathbb{R}^+} \mathbf{1}_{\theta \leq \mu b_{11} s^{(s)}_{11} K} Q^{(m)}_2(ds,d\theta)
- \int_{t}^{t'} \int_{\mathbb{R}^+} \mathbf{1}_{\theta \leq -(1 - \mu)b_2 x^{(s)}_2(s)} Q^{(d)}_2(ds,d\theta), \tag{7.22}
\]

and, on the event \( B_e^{(1,K)} \), for every \( t \leq t_K \),

\[
X^{(s)}_2(t) \geq \int_{t}^{t} \int_{\mathbb{R}^+} \mathbf{1}_{\theta \leq (1 - \mu)b_2 x^{(s)}_2(s)} Q^{(b)}_2(ds,d\theta) + \int_{t}^{t} \int_{\mathbb{R}^+} \mathbf{1}_{\theta \geq \mu b_{11} s^{(s)}_{11} K} Q^{(m)}_2(ds,d\theta)
- \int_{t}^{t} \int_{\mathbb{R}^+} \mathbf{1}_{\theta \leq -(1 - \mu)b_2 x^{(s)}_2(s)} Q^{(d)}_2(ds,d\theta). \tag{7.23}
\]

Hence, by applying arguments similar as before, we get the existence of two events \( A_e^{(2,K)} \) and \( B_e^{(2,K)} \) satisfying

\[
\mathbb{P} \left( A_e^{(2,K)} | A_e^{(1,K)} \cap B_e^{(1,K)} \right) \geq 1 - C \frac{\eta(K)}{\varepsilon^2} \quad \text{and} \quad \mathbb{P} \left( B_e^{(2,K)} | A_e^{(1,K)} \cap B_e^{(1,K)} \right) \geq 1 - C \frac{\eta(K)}{\varepsilon^2}.
\]
such that, for $t \leq t_K$, 
\[ X_2^{(+)}(t) \leq (1 + \varepsilon) \frac{b_1 x_1^{(+)} K \mu}{s_2^{(+)} s_{20}} = (1 + \varepsilon)^2 \frac{b_0 b_1 x_0^{(+)} K \mu^2}{s_{10}^2 s_{20}} \]
and, for $t^{(1)}_\varepsilon + t^{(2)}_\varepsilon \leq t \leq t_K$, 
\[ X_2^{(-)}(t) \geq (1 - \varepsilon)^2 \frac{b_1 x_1^{(-)} K \mu}{s_2^{(-)} s_{20}} = (1 - \varepsilon)^4 \frac{b_0 b_1 x_0^{(-)} K \mu^2}{s_{10}^2 s_{20}}. \]

We end the proof of Lemma 7.1 by iterating the procedure. \qed

We have now the tools needed to prove Theorem 3.2.

**Proof of Theorem 3.2.** From (7.5) and Lemma 7.1 we know that the $L$-population has a size of order $K \mu^L$ after a time of order $\ln(1/\varepsilon)$, for $\varepsilon$ small enough (not scaling with $K$). The proof of the asymptotics

\[ \liminf_{K \to \infty} \mathbb{P} \left( \frac{1 - c \varepsilon}{\alpha f_{10}} \frac{L}{\log K} < \frac{t^{(K,L)}_\varepsilon}{\log K} < \frac{1 + c \varepsilon}{\alpha f_{10}} \right) \geq 1 - c \varepsilon. \]

follows this of Lemma 6.1 in [46]. To end the proof of Theorem 3.2, we have one more step needed.

First we need to show that once the $L$-population size has reached the value $\varepsilon K$, the rescaled populations $X^K_L$ and $X^K_0$ behave as if they were the only ones in competition and follow a dynamics close to the solutions to (2.3) with $L = 1, \mu = 0$ and initial conditions satisfying
\[ x_L(0) = \varepsilon \quad \text{and} \quad |x_0(0) - \bar{x}_0| \leq 2 \varepsilon \sup_{1 \leq i \leq L} c_{10}/c_{00}. \]

This stays true until a time when $X^K_L$ is close to its monomorphic equilibrium size $\bar{x}_L$ and $X^K_0$ is smaller than $\varepsilon^2$. During this time interval, the $i$-population sizes, for $1 \leq i \leq L - 1$, do not evolve a lot. More precisely, there exists a constant $\varepsilon_0$ such that, for $\varepsilon \leq \varepsilon_0$ and $1 \leq i \leq L - 1$, with a probability close to one, $\mu_i^{+\varepsilon} \leq X^K_i(t) \leq \mu_i^{-\varepsilon}$, when $t$ describes an interval with a duration of order 1, which is the time needed for the rescaled population sizes $(X^K_0, X^K_L)$ to hit the set $((0, \varepsilon^2], [\bar{x}_L - \varepsilon, \bar{x}_L + \varepsilon])$ from an initial state close to $(\bar{x}_0, \varepsilon)$. To prove that, the idea is to show that the total population size stays of order $K$, and as a consequence with a probability close to one, we can find a positive $\delta$ such that $-\delta \mathbb{E}[X^K_i(t)] \leq d \mathbb{E}[X^K_i(t)]/dt \leq \delta \mathbb{E}[X^K_i(t)]$ (for rigorous arguments, see the proof of Lemma 10 in [53]). This leads to the following rigorous statement: there exist a positive $\varepsilon_0$ and a function $f : x \mapsto f(x) \in (0, x^2)$ such that, for $\varepsilon \leq \varepsilon_0$, there exist a stopping time $U^K_\varepsilon$ and an event $\mathcal{E}$ such that 
\[ \frac{U^K_\varepsilon f_{10}}{\log K} \to 1, \quad \text{in probability,} \quad \mathbb{P}(\mathcal{E}) \geq 1 - \varepsilon, \]
and almost surely on $\mathcal{E}$,
\[ f(\varepsilon) < X^K_0(U^K_\varepsilon) < \varepsilon^2, \quad |X^K_L(U^K_\varepsilon) - \bar{x}_L| \leq \varepsilon, \quad K \mu^\varepsilon < X^K_i(U^K_\varepsilon)/\mu^i < K \mu^{-\varepsilon}, \quad 1 \leq i \leq L - 1. \]

Second, we need to approximate the time for the $i$-populations ($0 \leq i \leq L - 1$) to get extinct after the time $U^K_\varepsilon$. Let us define two stopping times 
\[ V^K_\varepsilon := \inf \left\{ t \geq U^K_\varepsilon, |X^K_i(t) - \bar{x}_L| > 2 \varepsilon \right\}, \]
and
\[ W^K_e := \inf \left\{ t \geq U^K_e, \sum_{0 \leq i \leq L-1} X^i(t) > \varepsilon \right\}. \]

We will prove the following property: there exist \( \varepsilon_0, C, V > 0 \) such that, for \( \varepsilon \leq \varepsilon_0 \),
\[ \liminf_{K \to \infty} \mathbb{P} \left( e^{KV} < \left( W^K_e \land V^K_e \right) \right) \geq 1 - o_e(1), \]
(7.26)
where \( o_e(1) \) is a function of \( \varepsilon \) which goes to 0 as \( \varepsilon \) goes to 0. This will allow us to couple the \( i \)-population sizes \( (0 \leq i \leq L - 1) \) with sub-critical birth and death processes with inhomogeneous immigration in order to approximate their time to extinction.

To prove (7.26), we need to control the dynamics of two types of populations: first, the \( i \)-populations sizes, with \( 0 \leq i \leq L - 1 \), which are counterselected and whose initial size is smaller than \( O(\varepsilon^2 K) \), and second the \( L \)-population size.

Let us show that with a probability converging to 1 as \( K \to \infty \), \( W^K_e < V^K_e \). To this aim, notice that on the time interval \([0, V^K_e]\), the death rate of the \( i \)-population \( (0 \leq i \leq L - 1) \) satisfies:
\[ d_i + \sum_{j=0}^{L} \frac{c_{ij}}{K} X_j \geq d_i + c_{iL}(\bar{x}_L - 2\varepsilon). \]

Moreover we know that almost surely on the event \( E \), we have \( X_0(U^K_e) \leq \varepsilon^2 K \). Hence if we introduce, for \( 0 \leq i \leq L - 1 \) and \( k \in \mathbb{N} \), the notation
\[ T_k^{(X_i)} := \inf \{ t \geq U^K_e, X_i(t) = k \}, \]
and apply (A.5), we can compare the \( i \)-population process to a subcritical birth-death process with the effective death rate given above and obtain
\[ \mathbb{P} \left( T^{X_0}_{eK} < V^K_e \big| E \right) \leq \frac{\left( (d_0 + c_{0L}(\bar{x}_L - 2\varepsilon))/b_0 \right)^{\varepsilon^2 K} - 1}{\left( (d_0 + c_{0L}(\bar{x}_L - 2\varepsilon))/b_0 \right)^{\varepsilon^2 K_1} - 1} \leq \frac{b_0}{d_0 + c_{0L}(\bar{x}_L - 2\varepsilon)} \varepsilon^{K(1 - \varepsilon)} \geq 1 - C\varepsilon, \]
(7.27)
for any constant \( C \), \( K \) large enough and \( \varepsilon \) small enough.

Let us denote by \( M_{01} \) the number of type 1 mutants produced by type 0-individuals during the time interval \([U^K_e, V^K_e]\). From (7.2), we have:
\[ M_{01} = \int_{U^K_e}^{V^K_e} \pi_{(X_0)} d\theta. \]

Moreover, consider all the possible ordering of \( T^{(X_0)}_{eK}, V^K_e, T^{(X_0)}_{0} \) and \( \ln K/ \sqrt{\varepsilon} \), we get
\[ \mathbb{P} \left( \left\{ T^{(X_0)}_{eK} < V^K_e \land T^{(X_0)}_{0} \right\} \cup \left\{ V^K_e \land T^{(X_0)}_{0} < T^{(X_0)}_{eK} \land \frac{\ln K}{\sqrt{\varepsilon}} \right\} \cup \left\{ \frac{\ln K}{\sqrt{\varepsilon}} < V^K_e \land T^{(X_0)}_{0} < T^{(X_0)}_{eK} \right\} \right) = 1. \]
Hence, using the Markov inequality, \((7.27)\), as well as the fact that a subcritical branching process takes a time of order \(\ln K\) to get extinct (see \((A.4)\)), we get that
\[
\mathbb{P}(M_{01} > K\mu \ln K|\mathcal{E}) \leq \frac{(eK)(b_1\mu)(\ln K/\sqrt{\epsilon})}{K\mu \ln K} + \mathbb{P}\left(T_{X_0} < V^K|\mathcal{E}\right) + \mathbb{P}\left(\frac{\ln K}{\sqrt{\epsilon}} < V^K \land T^{(X_0)}_{0} < T^\epsilon_{X_0}|\mathcal{E}\right) = o_\epsilon(1).
\]
(7.28)

Applying again \((A.5)\), we get that each mutant of type 1 produced by a type 0 individual will generate a type 1 population whose size has a probability to reach \(\epsilon/\mu \ln K\) bounded by
\[
\left(\frac{b_1}{d_1 + c_{1\ell}(\bar{x}_L - 2\epsilon)}\right)^{e/\mu \ln K - 1}.
\]
We deduce that
\[
\mathbb{P}(T^{(X_1)}_{\epsilon K} < V^K_e|\mathcal{E}) \leq \mathbb{P}(M_{01} > K\mu \ln K|\mathcal{E}) + K\mu \ln K\left(\frac{b_1}{d_1 + c_{1\ell}(\bar{x}_L - 2\epsilon)}\right)^{e/\mu \ln K - 1} = o_\epsilon(1).
\]
We reiterate the reasoning for the other counter-selected mutant populations \((i\text{-populations with } 2 \leq i \leq L - 1)\) to conclude
\[
\mathbb{P}\left(W^K_e < V^K_e \mid \mathcal{E}\right) = o_\epsilon(1).
\]
(7.29)

By a direct application of Lemma \([A.1]\) we get the existence of a positive constant \(V\) such that:
\[
\liminf_{K \to \infty} \mathbb{P}\left(e^{KV} < V^K_e \leq W^K_e, \mathcal{E}\right) = 1.
\]
(7.30)

Using \((7.29)\) and \((7.30)\), we get:
\[
\liminf_{K \to \infty} \mathbb{P}\left(e^{KV} < (W^K_e \land V^K_e) \mid \mathcal{E}\right) \geq \liminf_{K \to \infty} \mathbb{P}\left(e^{KV} < V^K_e \leq W^K_e \mid \mathcal{E}\right) = \liminf_{K \to \infty} \mathbb{P}\left(e^{KV} < V^K_e \mid \mathcal{E}\right) \mathbb{P}\left(V^K_e \leq W^K_e \mid \mathcal{E}\right) = 1 - o_\epsilon(1).
\]
(7.31)

This proves \((7.26)\), as we recall that \(\mathbb{P}(\mathcal{E}) \geq 1 - \epsilon\).

We may now approximate the growth rates of the \(i\text{-population sizes } (0 \leq i \leq L - 1)\) during the time interval \([U^K_e, V^K_e \land W^K_e]\). For \(0 \leq i \leq L - 1\), let us introduce, for \(\ast \in \{-, +\}\)
\[
-\sigma_i^{(\ast)} := b_i(1 - \mu) - d_i - c_{ik}(\bar{x}_L \ast 2\epsilon) - 1_{|i| = -}\sup_{0 \leq k \leq L-1} c_{ik}\epsilon,
\]
(7.32)

where \(\ast \in \{-, +\} \setminus \ast\). Notice that, for \(\epsilon\) small enough the \((\sigma^{(\ast)}_i)_{0 \leq i \leq L-1}\) are pairwise distinct by the fourth point of Assumption I. We will consider such an \(\epsilon\) in the rest of the proof, to ensure not to divide by 0. Notice also that Equation \((7.32)\) ensures that there exists \(C > 0\) such that, for \(\epsilon\) small enough,
\[
0 < |f_{ik}| - CE < \sigma^{(+)}_i < |f_{ik}| < \sigma^{(-)}_i < |f_{ik}| + CE.
\]
From the definition of the process \(X\) in \((7.2)\) and from \((7.25)\), we get that almost surely on the event \(\mathcal{E}\) and for \(0 \leq i \leq L - 1\),
\[
P^{(\ast)}_i(t) \leq X_i(U^K_e + t) \leq P^{(+)}_i(t), \quad \forall U^K_e \leq U^K_e + t \leq V^K_e \land W^K_e
\]
(7.33)
where, for $t \geq 0$ and $* \in \{-, +\}$,
\[
P_i^{(*)}(t) = X_i(U^K_e) + \int_{U^K_e}^{U^K_e + t} \int_{\mathbb{R}^+} 1_{\theta \leq (1 - \mu b_{\beta_L} \xi_i(x))} Q_i^{(0)}(ds, d\theta) + \int_{U^K_e}^{U^K_e + t} \int_{\mathbb{R}^+} 1_{\theta \leq (1 - \mu b_{\beta_L} \xi_i(x))} Q_i^{(m)}(ds, d\theta)
\]
\[\quad - \int_{U^K_e}^{U^K_e + t} \int_{\mathbb{R}^+} 1_{\theta \leq (1 - \mu b_{\beta_L} \xi_j(x))} Q_i^{(0)}(ds, d\theta),
\]
(7.34)
where we recall that by convention $b_{-1} = 0$.

To find a lower bound of the extinction time of the unfit mutant population size, let us introduce
\[
\beta_L := \left\{ k \in \{0, L - 1\} \text{ such that } \frac{|f_{jL}|}{k/\alpha} = \inf_{0 \leq j \leq L - 1} \left\{ \frac{|f_{jL}|}{1 - j/\alpha} \right\} \right\}.
\]
(7.35)
We will see that the $\beta_L$-population is the one which takes the longest time to get extinct, and drives the time to extinction of the whole mutant-population. Recalling (7.25), we know that on the event $E$ the size at time $U^K_e$ of the $\beta_L$-population is
\[
C(\varepsilon, K)K^{\mu^\varepsilon}, \quad \text{with } \mu^\varepsilon \leq C(\varepsilon, K) \leq \mu^\varepsilon.
\]
From (7.33) and (7.34), we see that almost surely on $E$ and on the time interval $[U^K_e, V^K_e \wedge W^K_e]$, the $\beta_L$-population size is larger than a sub-critical birth and death process with initial state $C(\varepsilon, K)K^{\mu^\varepsilon}$, individual birth rate $b_{\beta_L}(1 - \mu)$, and individual death rate $b_{\beta_L}(1 - \mu) + \sigma^{(\varepsilon)}_{\beta_L}$. Applying Equation A.3, we deduce that
\[
\liminf_{K \to \infty} \mathbb{P} \left( \inf \left\{ t \geq 0, X_{\beta_L}(U^K_e + t) = 0 \right\} \geq \left( 1 - \frac{\beta_L}{\alpha} \right) \frac{(1 - \varepsilon)}{\sigma^{(\varepsilon)}_{\beta_L}} \ln K \right) \geq 1 - \varepsilon.
\]
(7.36)

The last step of the proof consists in finding a bound for $\mathbb{E}[P_i^{(*)}(t)]$ for large $t$, to show that the total unfit mutant population size will take a time of order at most $(1 + l\varepsilon)(1 - \beta_L/\alpha) \ln K/\sigma^{(\varepsilon)}_{\beta_L}$, for some positive $l$ (to be made precised later, see (7.41)) to get extinct. To simplify notations, let us introduce, for $0 \leq i \leq L - 1$ and $* \in \{-, +\}$,
\[
\tilde{f}_i^{(*)} := \inf \{ \sigma_j^{(*)}, 0 \leq j \leq i \}.
\]
(7.37)
We will see that the mutant population whose size decreases the slowest will be the leading term and scale the time needed for all but the $L$ populations to get extinct. To prove that, we will now show by induction that there exists $\varepsilon_0 > 0$ and a sequence of positive functions $(g_0 : x \mapsto x^2, g_1, ..., g_{L-1})$, such that, for every $0 \leq i \leq L - 1$, $\varepsilon \leq \varepsilon_0$ and $t \geq 0$,
\[
\mathbb{E}[P_i^{(*)}(t)] \leq g_i(\varepsilon)K^{\mu^\varepsilon}e^{-\tilde{f}_i^{(*)}}.
\]
(7.38)
For $i = 0$, from definitions (7.32), (7.34) and property (7.25), we get
\[
\mathbb{E}[P_0^{(*)}(t)] \leq \varepsilon^2Ke^{-\sigma_0^{(*)}} = \varepsilon^2Ke^{-\tilde{f}_0^{(*)}}.
\]
Let us assume that \((7.38)\) holds for every \(i\) such that \(0 \leq i < i_0 < L - 1\). Then from \((7.32), (7.34)\) and the induction hypothesis, for \(t \geq 0\),
\[
\frac{d}{dt} \mathbb{E}[p_{i_{0}+1}^{(i)}(t)] \leq -\sigma_{i_{0}+1}^{(i)} \mathbb{E}[P_{i_{0}+1}^{(i)}(t)] + \mu b_{i_{0}} \mathbb{E}[P_{i_{0}}^{(i)}(t)]
\]
\[
\leq -\sigma_{i_{0}+1}^{(i)} \mathbb{E}[P_{i_{0}+1}^{(i)}(t)] + \mu b_{i_{0}} g_{i_{0}}(\varepsilon) K \mu^{i_{0}-\varepsilon} e^{-\sigma_{i_{0}}^{(i)} t}
\]
\[
= -\sigma_{i_{0}+1}^{(i)} \mathbb{E}[P_{i_{0}+1}^{(i)}(t)] + b_{i_{0}} g_{i_{0}}(\varepsilon) K \mu^{i_{0}+1-\varepsilon} e^{-\sigma_{i_{0}}^{(i)} t}.
\] (7.39)

Applying the method of variation of parameters, we get, for every \(t \geq 0\),
\[
\mathbb{E}[P_{i_{0}+1}^{(i)}(t)] \leq \mathbb{E}[P_{i_{0}+1}^{(i)}(0)] e^{-\sigma_{i_{0}}^{(i)} t} + \frac{b_{i_{0}} g_{i_{0}}(\varepsilon) K \mu^{i_{0}+1-\varepsilon}}{\sigma_{i_{0}+1}^{(i)} - \sigma_{i_{0}}^{(i)}} \left(e^{-\sigma_{i_{0}}^{(i)} t} - e^{-\sigma_{i_{0}}^{(i)} t}\right),
\] (7.40)

where the last inequality is a consequence of \((7.25)\). Hence, the \(i_{0} + 1\)-population satisfies \((7.38)\), with \(\sigma_{i_{0}+1}^{(i)} = \sigma_{i_{0}}^{(i)} \) and \(K\), according to the definition \((7.37)\), and
\[
g_{i_{0}}(\varepsilon) = 1 + \frac{2b_{i_{0}} g_{i_{0}}(\varepsilon)}{\sigma_{i_{0}+1}^{(i)} - \sigma_{i_{0}}^{(i)}}.
\]

Moreover, let us introduce \(l > 0\) such that, for \(\varepsilon\) small enough and for \(0 \leq i \leq L - 1\),
\[
\frac{1 - i/\alpha + \varepsilon/\alpha}{\tau_{i}^{(i)}} < (1 + le) \frac{1 - \beta L/\alpha}{\sigma_{\beta L}^{(i)}}
\] (7.41)

(which is possible according to the definitions \((7.32)\) and \((7.35)\)) and define
\[
s_{K} := \left(1 + le\right) \frac{1 - \beta L/\alpha}{\sigma_{\beta L}^{(i)}} \ln \left(K \mu^{\beta L}\right).
\]

Then, applying \((7.33), (7.36)\) and the Markov inequality, we get:
\[
\mathbb{P} \left(\exists i \in [0, L - 1], X_{i}(s_{K}) \geq 1 \mid \mathcal{E}\right) \leq \sum_{0 \leq i \leq L - 1} \mathbb{P} \left(X_{i}(s_{K}) \geq 1 \mid \mathcal{E}\right)
\]
\[
\leq \sum_{0 \leq i \leq L - 1} \mathbb{P} \left(P_{i}^{(i)}(s_{K}) \geq 1 \mid \mathcal{E}\right) \leq \sum_{0 \leq i \leq L - 1} \mathbb{E} \left(P_{i}^{(i)}(s_{K}) \mid \mathcal{E}\right) \leq \sum_{0 \leq i \leq L - 1} g_{i}(\varepsilon) K \mu^{i_{0}-\varepsilon} e^{-\sigma_{i_{0}}^{(i)}/K}
\]
\[
= \sum_{0 \leq i \leq L - 1} g_{i}(\varepsilon) \exp \left[\frac{\sigma_{i}^{(i)}}{\sigma_{i_{0}}^{(i)}} \left(1 + le\right) \ln \left(K \mu^{\beta L}\right) - \frac{\sigma_{i}^{(i)}}{\sigma_{i_{0}}^{(i)}} \ln \left(K \mu^{\beta L}\right)\right]
\]
\[
= \sum_{0 \leq i \leq L - 1} g_{i}(\varepsilon) \exp \left[\frac{\sigma_{i}^{(i)}}{\sigma_{i_{0}}^{(i)}} \left(1 - i/\alpha + \varepsilon/\alpha\right) - \frac{\sigma_{i}^{(i)}}{\sigma_{i_{0}}^{(i)}} \ln \left(1 + o(1)\right)\right],
\] (7.42)

where we used \((3.12)\) in the last line. According to the definition of \(l\), the last term goes to 0 as \(K\) goes to infinity.

Combining \((7.24), (7.26), (7.36)\), and \((A.1)\) ends the proof of Theorem 3.2.  \(\square\)
7.3. **Proof of Theorem 3.3.** Assume first that (3.12) holds and that \( \alpha \notin \mathbb{N} \). Theorem 3.3 addresses the case where \( \alpha \mu^2 \) is small. Only the \( [\alpha] \) first mutant populations will have a large size, as
\[
K \mu^{[\alpha]} = \int_{(\alpha]} (K) K^{1-[\alpha]/\alpha} \to \infty, \quad K \to \infty,
\]
\[
K \mu^{[\alpha]+1} = \int_{(\alpha]+1} (K) K^{1-([\alpha]+1)/\alpha} \to 0, \quad K \to \infty.
\]
For \([\alpha] + 1 \leq i \leq L - 1\), the \( i \)-mutant population sizes will describe excursions until a successful \( L \)-individual is created. By successful \( L \)-individual we mean a mutant \( L \) which generates a population out-competing the other populations. Here again the key idea will be couplings with birth and death processes without competition.

Let us denote by \( T^{(i)} \) (see definition in (7.44)) the birth time of the \( i \)-th mutant of type \( ([\alpha] + 1) \) descended from an individual of type \( [\alpha] \) and by \( X^{(i)} \) the type \( ([\alpha] + 1) \)-population generated by this individual. Then, we use the lexicographic order to number the \( k \)-mutant populations, with \([\alpha] + 2 \leq k \leq L\) (see Figure 2 for an illustration). More precisely,
- For \( j \geq 1 \), \( X^{(i)}_j \) is the \( ([\alpha] + 2) \)-population generated by the \( j \)th \( ([\alpha] + 2) \)-mutant produced by an individual of type \( ([\alpha] + 1) \) belonging to the population \( X^{(i)}_0 \)
- For \( j, k \geq 1 \), \( X^{(i)}_{jk} \) is the \( ([\alpha] + 3) \)-population generated by the \( k \)th \( ([\alpha] + 3) \)-mutant produced by an individual of type \( ([\alpha] + 2) \) belonging to the population \( X^{(i)}_j \).

As we will see along the proof, a mutant population of type \( i \) produces typically no \((i + 1)\)-mutant, one \((i + 1)\)-mutant with a probability of order \( \mu \), and more than one \((i + 1)\)-mutant with a probability of order \( \mu^2 \). The law of all trees can be approximated by the law of an inhomogeneous sub-critical Galton-Watson process, and the trees are approximately independent. Hence we will be able to approximate the probability for the \( X^{(i)}_j \) populations \((i \geq 1)\) to generate a successful mutant \( L \) by a common probability, and the time needed for a successful \( L \)-mutant to appear will be close to an exponential random variable with mean one divided by this probability.

Recall the definition of the process \( X \) in (7.2). Then, if we define the process
\[
R(t) := \int_0^t \int_{\mathbb{R}^+} 1_{\theta \leq \mu b_{[\alpha]} X^{(i)}(r)} Q^{(m)}_{[\alpha]+1}(ds, d\theta),
\]
(7.43)
the stopping time \( T^{(i)} \) which is the birth time of the \( i \)-th \( ([\alpha] + 1) \)-mutant produced by an \([\alpha]\)-individual can be expressed as
\[
T^{(i)} := \inf\{t \geq 0, R(t) \geq i\}.
\]
(7.44)
In particular, from (7.5), we get, for every \( t \leq T^K_{\epsilon} \wedge S^K_{\epsilon} \),
\[
R^{-}(t) \leq R(t) \leq R^{+}(t), \quad a.s.,
\]
(7.45)
where
\[
R^{(\pm)}(t) := \int_0^t \int_{\mathbb{R}^+} 1_{\theta \leq \mu b_{[\alpha]} X^{(i)}(r)} Q^{(m)}_{[\alpha]+1}(ds, d\theta),
\]
(7.46)
and we can also define, for \(* \in \{+,-\} \),
\[
T^{(i,*)} := \inf\{t \geq 0, R^{(*)}(t) \geq i\}.
\]

Let \( u_K \) be a sequence such that
\[
u_K \gg \ln K \quad \text{and} \quad \mu^{[\alpha]+1} K u_K \to 0.
\]
Using Lemma 7.1 and Markov Inequality, we get

\[
\mathbb{P}\left(T^{(1,-)} \leq u_K|E^K_\varepsilon\right) \leq \mathbb{P}\left(R^{-}(u_K) \geq 1|E^K_\varepsilon\right) \leq \mu b_1 \mathbb{E}\left[\int_0^{u_K} X^{(s)}_{[\alpha]}(s)ds\bigg|\mathcal{E}^K_\varepsilon\right]
\]

\[
\leq (1 + \varepsilon)^{[\alpha]} \frac{b_0 ... b_{[\alpha]} x^{(s)}_{[\alpha]} \mu_{[\alpha]+1}}{s_{[\alpha]}^{(s)} \cdots s_{[\alpha]}^{(s)}} K u_K \to 0, \quad (K \to \infty).
\]  

(7.47)

Moreover, the time needed for the first \(L\)-mutant to be born is of order \((K \mu L)^{-1}\), which is much larger than \(\ln K\), for large \(K\). This implies that we can consider the process after time \(t^{(1)}_e + ... + t^{([\alpha])}_e\), which has been defined in (7.13), and thus on the event \(E^K_\varepsilon\) satisfying \(\mathbb{P}(E^K_\varepsilon) \to 1, K \to \infty\),

\[
x^{(-)}_{[\alpha]} K \leq X^{(s)}_{[\alpha]}(s) \leq x^{(s)}_{[\alpha]} K, \quad t^{(1)}_e + ... + t^{([\alpha])}_e \leq s \leq t_K \wedge T^K_\varepsilon \wedge S^K_\varepsilon \quad \text{a.s. on } E^K_\varepsilon,
\]

where \(t_K\) is any sequence tending to infinity, as \(K\) tends to \(\infty\). Thus, if we introduce

\[
D^K_\varepsilon := E^K_\varepsilon \cap \left\{R^{-}\left(t^{(1)}_e + ... + t^{([\alpha])}_e\right) = 0\right\},
\]

\[
\text{Figure 2. Numbering of the successive sub-populations: in this example, the first } ([\alpha] + 1)\text{-mutant generates the } ([\alpha] + 1)\text{-population } X^{(1)}_0. \text{ One individual of the population } X^{(1)}_0 \text{ gives birth to an } ([\alpha] + 2)\text{-mutant, which generates the } ([\alpha] + 2)\text{-population } X^{(1)}_1. \text{ One individual of the population } X^{(1)}_1 \text{ gives birth to an } ([\alpha] + 3)\text{-mutant, which generates the } ([\alpha] + 3)\text{-population } X^{(1)}_1. \text{ The population } X^{(1)}_1 \text{ gets extinct without giving birth to any } ([\alpha] + 4)\text{-individual. The second } ([\alpha] + 1)\text{-mutant produced by an } [\alpha]\text{-individual generates the } ([\alpha] + 1)\text{-population } X^{(2)}_0. \text{ Two individuals of the population } X^{(2)}_0 \text{ give birth to an } ([\alpha] + 2)\text{-mutant. These mutants generate the } ([\alpha] + 2)\text{-populations } X^{(2)}_1 \text{ and } X^{(2)}_2, \text{ respectively. And so on with the notations previously introduced.}
and, for $* \in \{-, +\}$,

$$R_{i,j}^{(*)}(t) := \int_0^t 1_{(\theta \leq \text{sub}_{\text{type}(j)^{(*)}})}(\theta)K_i O_{i,j}^{(m)}(ds,d\theta), \quad (7.48)$$

we get that $P(D^K_e) \to 1$, as $K \to \infty$ and

$$R_{i,j}^{(*)}(t) \leq R(t) \leq R_{i,j}^{(*)}(t), \quad \text{for all} \quad t \leq T^K_e \wedge S^K_e, \quad \text{a.s. on} \quad D^K_e. \quad (7.49)$$

Notice that $R_{i,j}^{(*)}$ is a Poisson process with parameter $\mu b_{\text{type}(j)^{(*)}}$. Following the ideas developed in Section 7.2, we may couple each $k$-mutant population ($[\alpha] + 1 \leq k \leq L - 1$) with two birth and death processes, independent conditionally on their birth time. We will not detail the couplings as the ideas have already been developed and the notations are tedious, but we nevertheless state rigorously the resulting properties. Let us denote by $(T_{i,j}^{(i)} ; i \in \cup N_i^R, i \in \mathbb{N})$ the time of appearance of the populations ($X_i^{(i)}, i \in \cup N_i^R, i \in \mathbb{N}$). For all $j \in \cup N_i^R, i \in \mathbb{N}$, $T_{i,j}^{(i)} := \inf \{t \geq 0, X_i^{(i)}(t) \geq 1\}$. Then we introduce birth and death processes ($X_i^{(i),*}, i \in \cup N_i^R, i \in \mathbb{N}, * \in \{-, +\}$) with birth and death rates

$$\left( (b_{i,j}, (1-\mu)b_{i,j} + \sigma_{i,j}^{(*)}), i \in \cup N_i^R, i \in \mathbb{N}, * \in \{-, +\} \right),$$

where the $\sigma^{(*)}$'s have been defined in (7.32),

$$t(i) := [\alpha] + || + 1,$$

and $||$ is the number of terms in $j$ (for instance $||11221|| = 5$).

These processes can be constructed in such a way that, for all $j \in \cup N_i^R, i \in \mathbb{N}$,

$$X_i^{(i),*(t)} \leq X_i^{(i)}(t) + Y_i^{(i)}(t) \leq X_i^{(i),*}(t), \quad t \leq T^K_e \wedge S^K_e \quad \text{a.s. on} \quad D^K_e, \quad (7.50)$$

where $Y_i^{(i)}(t)$ is the number of mutants of type ($[\alpha] + || + 2$) produced by the $X_i^{(i)}$ population (which is of type ($[\alpha] + || + 1$)) until time $t$. Recall that among the offsprings produced by the population $X_i^{(i)}$, a fraction $(1-\mu)$ is constituted by newborn individuals of type $[\alpha] + || + 1$, and a fraction $\mu$ by new born individuals of type $[\alpha] + || + 2$, and that at each birth event the probability to have a mutation is independent from the past.

Moreover, conditionally on $(T_{i,j}^{(i)} ; i \in \cup N_i^R, i \in \mathbb{N})$, the pairs of processes ($X_i^{(i),*}, X_i^{(i),*}$), $i \in \cup N_i^R, i \in \mathbb{N}$ are independent. This last point will allow us to approximate the law of the random trees $T^{(i)} := X_0^{(i)} \cup_{\text{type}(i)} X_{V(i)}^{(i)}, \quad i \in \mathbb{N}$ (an example is depicted in Figure 2) by the same law, and independently. Indeed from Equation (7.50), we get that $(T^{(i)}_{i,j}) \in \mathbb{N}$, can be coupled with two inhomogeneous birth and death processes, whose law is well known and easy to study. This will be the object of the end of the proof.

We will now consider each tree $T^{(i)} (i \in \mathbb{N})$ with root $X_0^{(i)}$ independently, and approximate its probability to end with a $L$-mutant sub-population. For simplicity we write $0 = 0$.

Consider a vertex $X_i^{(i)}, i \in \{0\} \cup \mathbb{N}_1$ of the tree $T^{(i)}$. Thanks to the coupling (7.50) we will be able to approximate the probability for this vertex to have no child, one child or more than one child. On the event $D^K_e$ and before the time $T^K_e \wedge S^K_e$, the law of the number of offspring produced by the $X_i^{(i)}$ population is smaller (resp. larger) than the number of offspring produced by a population initiated by one individual, with individual birth rate $b_{i,j}$ and individual death
rate \((1 - \mu) b_{ij} + \sigma_{ij}^{(+)}\) (resp. \((1 - \mu) b_{ij} + \sigma_{ij}^{(-)}\)). Moreover, each offspring is a mutant of type \((t(i) + 1)\) with probability \(\mu\), and is a clone with probability \((1 - \mu)\). Hence

\[
P(X^{(i)}_{j} \text{-pop produces 1 mutant}) \leq \sum_{n=0}^{\infty} p^{(b_{ij},(1-\mu)b_{ij}+\sigma_{ij}^{(\pm)})(n)} n \mu(1 - \mu)^{n-1} \leq \mu e^{(b_{ij},b_{ij}+\sigma_{ij}^{(\pm)})},
\]

where \(p^{(-)}\) and \(e^{(-)}\) are defined in Lemma A.3. Similarly, for \(K\) large enough,

\[
P(X^{(i)}_{j} \text{-pop produces 1 mutant}) \geq \sum_{n=0}^{\infty} p^{(b_{ij},(1-\mu)b_{ij}+\sigma_{ij}^{(\pm)})(n)} n(1 - \mu)^{n-1/2} \geq 1/4 + \mu \sum_{n=0}^{\infty} p^{(b_{ij},(1-\mu)b_{ij}+\sigma_{ij}^{(\pm)})(n)} n(1 - \mu)^{n-1/2} \geq n \mu \sum_{n=0}^{\infty} (1 - \mu)^{n-1} \geq \mu(1 - 2 \sqrt{\mu}) \left[ e^{(b_{ij},(1-\mu)b_{ij}+\sigma_{ij}^{(\pm)})} - \mu \right],
\]

where \(|X^{(i)}_{j}|\) denotes the number of offspring (mutants or clones) produced by the population \(X^{(i)}_{j}\). But Cauchy-Schwarz and Markov inequalities yield

\[
E^2 \left[ 1_{|X^{(i)}_{j}| \geq \mu^{-1/2} + 1} |X^{(i)}_{j}| \right] \leq P \left[ |X^{(i)}_{j}| \geq \mu^{-1/2} + 1 \right] \geq \sqrt{\mu} E \left[ |X^{(i)}_{j}| \right] E \left[ |X^{(i)}_{j}|^2 \right] = O(\sqrt{\mu}),
\]

as the two expectations are finite according to (A.6).

Adding (A.3), we may conclude that, as \(K\) goes to infinity,

\[
P \left( X^{(i)}_{j} \text{-pop produces 1 mutant} \right) = e^{(b_{ij},b_{ij}+\sigma_{ij}^{(\pm)})} \mu(1 + O(\varepsilon)).
\]

Using again coupling (7.50) and (A.6), we get that

\[
P \left( X^{(i)}_{j} \text{-pop produces at least 2 mutants} \right) \leq \sum_{n=0}^{\infty} p^{(b_{ij},(1-\mu)b_{ij}+\sigma_{ij}^{(\pm)})(n)} n(n-1) \mu^2 \leq O(\mu^2).
\]

From the last computations, we can infer that, for \(i \geq 1\), the probability for the tree \(T^{(i)}\) to produce an \(L\)-mutant is, for large \(K\),

\[
\mu^{L-1}[\varepsilon] \left( \prod_{k=0}^{L-1} e^{(b_{ij},b_{ij}+\sigma_{ij}^{(\pm)})} \right) (1 + O(\varepsilon)).
\]

Indeed, the probability for each vertex to produce one child is of order \(\mu\), and the probability to produce at least two children is of order \(\mu^2\). Since there is only a finite number of possible mutations, independent of \(\mu\), this implies that the probability for the tree \(T^{(i)}\) to have at least one vertex with two children and end with an \(L\) individual is of order \(\mu^{L-1}[\varepsilon]\), which is negligible compared to \(\mu^{L-1}[\varepsilon]\). Moreover, we know that each \(L\)-mutant has a probability close to \(f^{(i)}/b_{L}\) to generate a population whose size hits the value \(\varepsilon K\), and once this size is reached, the time needed for the \(L\)-population to outcompete the other populations and hit its equilibrium size is of order \(\ln K\) (see for instance [14]), which is negligible with respect to the time needed for the successful
We want to bound the probability that at least one type 1 individual born from a type 0 individual is replaced by the 0-population.

To end the proof of Theorem 3.3 let us consider the case when \( \mu \ll 1/K \). From Lemma A.1 we know that, for \( \epsilon \) small enough, there exists a positive \( V \) such that with high probability, the size of a monomorphic 0-population stays at a distance smaller than \( \epsilon K \) from its equilibrium size \( \bar{n}_0 K \) during a time larger than \( e^{KV} \). As a consequence, if \( K \mu \gg e^{\theta K} \), the 0-population will produce a large number of 1-mutants during the time interval \([0, e^{\theta K}]\), with a rate very close to \( b_0 \bar{n}_0 K \mu \). Hence the proof is very similar to the previous proof, where the \([\alpha]\)-population is replaced by the 0-population.

8. Proofs of Section 3.3

8.1. Proof of Theorem 3.4 point 2. Recall from (7.2) that the process \( X_0 \) admits the following Poisson representation:

\[
X_0(t) = [\bar{x}_0 K] + \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq (1-\mu) b_0 Y_0(s)} Q_0^{(b)}(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq D_0^{b}(X_0(s))} Q_0^{(d)}(ds, d\theta),
\]

where \( D_0^{b}(X) \) is defined in (7.1). Thus, if we introduce the process \( Y_0 \) via

\[
Y_0(t) = [\bar{x}_0 K] + \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq \rho_0 Y_0(s)} Q_0^{(b)}(ds, d\theta) - \int_0^t \int_{\mathbb{R}_+} 1_{\theta \leq D_0^{b}(X_0(s))} Q_0^{(d)}(ds, d\theta),
\]

we get that, almost surely, \( X_0(t) \leq Y_0(t) \), for all \( t \geq 0 \). Now consider a time \( \nu_k \) such that

\[
\frac{1}{\rho_0(K)} \ll \nu_k \ll \frac{1}{K \mu^L}, \quad K \to \infty,
\]

where \( \rho_0(K) \) was defined in (3.20). If we apply inequality (3.7) of [18] to the process \( Y_0 \), we get:

\[
d_{TV} \left( \mathbb{P}(Y_0(\nu_k) \leq .), \delta_0(.) \right) \to 0, \quad K \to \infty,
\]

where \( d_{TV} \) is the total variation distance. This implies

\[
\mathbb{P}(X_0(\nu_k) > 0) \to 0, \quad K \to \infty.
\]

Hence to prove Theorem 3.4 point 2 it is enough to show that \( \mathbb{P}(B_L < \nu_k) \to 0 \). Notice that from (8.1) we have, for every positive \( t \),

\[
\frac{d}{dt} \mathbb{E}[X_0(t)] \leq \mathbb{E}\left[\left( b_0 - d_0 - \frac{c_0 \theta}{K} X_0(t) \right)X_0(t) \right] \leq (b_0 - d_0)\mathbb{E}[X_0(t)] - \frac{c_0 \theta}{K} \mathbb{E}^2[X_0(t)].
\]

Thus, for all \( t \geq 0 \), we have \( \mathbb{E}[X_0(t)] \leq \bar{x}_0 K \). Next we bound the expectation of the total number \( \Xi_1 \) of type 1 individuals generated by type 0 individuals by mutations before the time \( \nu_k \):

\[
\mathbb{E}[\Xi_1] \leq \int_0^{\nu_k} b_0 \mu \mathbb{E}[X_0(s)] ds \leq b_0 \bar{x}_0 K \mu \nu_k.
\]

We want to bound the probability that at least one type 1 individual born from a type 0 individual before time \( \nu_k \) has a line of descent containing a type \( L \) individual. Denote by \( \xi_i \) the event that
the $i$th type 1 individual born from a type 0 individual before time $v_K$ has a descendant of type $L$ at any time in the future. We see that
\[ P(B_L < v_K) = P\left( \bigcup_{i \in \Xi} \xi_i \right) = \mathbb{E}\left[ P\left( \bigcup_{i \in \Xi} \xi_i | \Xi \right) \right] \leq \mathbb{E}\left[ \sum_{i \in \Xi} P(\xi_i | \Xi) \right]. \]

But recall that by Assumption 2, for $1 \leq i \leq L - 1$, $b_i < d_i$. Hence using (7.52), we see that the probability of the events $(\xi_i)_{1 \leq i \leq \Xi}$ can be bounded independently of $\Xi$ by
\[ 2 \left( \prod_{1 \leq i \leq L-1} e^{(b,d_i)} \right) \mu^{L-1}. \]

This yields
\[ P(B_L < v_K) \leq b_0 \bar{x} v_K \left( \prod_{1 \leq i \leq L-1} e^{(b,d_i)} \right) \mu^L \to 0. \]

Adding (8.2) ends the proof.

8.2. **Proof of Theorem 3.4 point 1.** We introduce $v_K$ such that $\frac{1}{\rho(K)} \ll v_K \ll \frac{1}{K\mu}$. Then (8.3) and Markov Inequality ensure that with a probability close to 1, no type 1 mutant is produced before the population extinction. As a consequence, no type $L$ mutant is produced. This ends the proof.

**A. Technical results**

The next Lemma quantifies the time spent by a birth and death process with logistic competition in a vicinity of its equilibrium size. It is stated in [14] Theorem 3(c).

**Lemma A.1.** Let $b, d, c$ be in $\mathbb{R}_+$ such that $b - d > 0$. Denote by $(W_t)_{t \geq 0}$ a density dependent birth and death process with birth rate $bn$ and death rate $(d + cn/K)n$, where $n \in \mathbb{N}_0$ is the current state of the process and $K \in \mathbb{N}$ is the carrying capacity. Fix $0 < \eta_1 < (b - d)/c$ and $\eta_2 > 0$, and introduce the stopping time
\[ S_K = \inf \left\{ t \geq 0 : W_t \notin \left[ \left( \frac{b - d}{c} - \eta_1 \right)K, \left( \frac{b - d}{c} + \eta_2 \right)K \right] \right\}. \]

Then, there exists $V > 0$ such that, for any compact subset $C$ of $[(b - d)/c - \eta_1, (b - d)/c + \eta_2]$,\n\[ \lim_{K \to \infty} \sup_{K \in C} \mathbb{P}_k(S_K < e^{KV}) = 0. \]  

(A.1)

Let us now recall some results on hitting times of a birth and death process. The first, third, and last statements can be found in [12]. The second statement is a consequence of the first statement.

**Lemma A.2.** Let $Z = (Z_t)_{t \geq 0}$ be a birth and death process with individual birth and death rates $b$ and $d$. For $i \in \mathbb{Z}_+$, $T_i = \inf\{t \geq 0, Z_t = i\}$ and $\mathbb{P}_i$ (resp. $\mathbb{E}_i$) is the law (resp. expectation) of $Z$ when $Z_0 = i$. Then

- If $d \neq b \in \mathbb{R}_+$, for every $i \in \mathbb{Z}_+$ and $t \geq 0$,
  \[ \mathbb{P}_i(T_0 \leq t) = \left( \frac{d(1 - e^{(d-b)t})}{b - de^{(d-b)t}} \right)^i. \]  

(A.2)
• If $0 < b < d$ and $Z_0 = N$, the following convergence holds:

$$T_0/\log N \xrightarrow{N \to \infty} (d - b)^{-1}, \text{ in probability.} \quad (A.3)$$

• If $0 < d < b$, on the non-extinction event of $Z$, which has a probability $1 - (d/b)^{Z_0}$, the following convergence holds:

$$T_N/\log N \xrightarrow{N \to \infty} (b - d)^{-1}, \text{ a.s.} \quad (A.4)$$

• If $0 < b < d$, and $(i, j, k) \in \mathbb{N}^3$ such that $j \in (i, k)$,

$$\mathbb{P}(T_k < T_i) = \frac{1}{(d/b)^{j-i} - 1}. \quad (A.5)$$

The last result of this Appendix concerns the size distribution of the total number of individuals in a sub-critical birth and death process. We refer the reader to [48] (Theorem 3.13 applied to the case when $X$ is a geometric random variable with parameter $d/(b + d)$) or [12] for the proof of the two first points. The last one is just a consequence of the Mean Value Theorem.

**Lemma A.3.** Let us consider a birth and death process with individual birth rate $b > 0$ and individual death rate $d > 0$ satisfying $b < d$. Let $Z$ denote the total number of births during an excursion of this process initiated with one individual. Then, for $k \geq 0$,

$$p^{(b,d)}(k) := \mathbb{P}(Z = k) = \frac{(2k)!}{k!(k + 1)!} \left( \frac{b}{d + b} \right)^k \left( \frac{d}{d + b} \right)^{k+1}.$$  

In particular,

$$e^{(b,d)} := \mathbb{E}[Z] = \sum_{k=0}^{\infty} \frac{(2k)!}{(k-1)!(k + 1)!} \left( \frac{b}{d + b} \right)^k \left( \frac{d}{d + b} \right)^{k+1}.$$  

Moreover, there exist two positive constants $c$ and $\varepsilon_0$ such that, for every $\varepsilon \leq \varepsilon_0$, if $0 < d_i < b_i$ and $|b_i - d_i| \leq \varepsilon$, $i \in \{1, 2\}$, then

$$|e^{(b_1,d_1)} - e^{(b_2,d_2)}| \leq c\varepsilon. \quad (A.8)$$

**References**

[1] D. Abu Awad and S. Billiard. The double edged sword: The demographic consequences of the evolution of self-fertilization. *Evolution*, 71(5):1178–1190, 2017.

[2] H. Alexander. Conditional distributions and waiting times in multitype branching processes. *Adv. Appl. Probab.*, 45(3):692–718, 2013.

[3] M. Baar, A. Bovier, and N. Champagnat. From stochastic, individual-based models to the canonical equation of adaptive dynamics in one step. *Ann. Appl. Probab.*, 27(2):1093–1170, 2017.

[4] M. Baar, L. Coquille, H. Mayer, M. Hölzel, M. Rogava, T. Tüting, and A. Bovier. A stochastic model for immunotherapy of cancer. *Scientific Reports*, 6(1):24169, 2016.

[5] S. Billiard and C. Smadi. The interplay of two mutations in a population of varying size: a stochastic eco-evolutionary model for clonal interference. *Stochastic Process. Appl.*, 127(3):701–748, 2017.

[6] B. Bolker and S. W. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.*, 52(3):179–197, 1997.

[7] B. M. Bolker and S. W. Pacala. Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. *Am. Nat.*, 153(6):575–602, 1999.
Figure 3. Graph of $x(t)$ in the 1-sided case $m_{ij} = m_{ij}^{(1)}$ for $L = 6$ and $f_{60} = 1, (f_{06}, f_{16}, f_{26}, f_{36}, f_{46}, f_{56}) = (-5, -1, -0.25, -1.5, -2, -0.05)$, which is the fitness landscape depicted in Figure 1.

Figure 4. Graph of $x(t)$ in the 2-sided case $m_{ij} = m_{ij}^{(2)}$ for $L = 6$ and $f_{60} = 1, (f_{06}, f_{16}, f_{26}) = (-5, -1, -0.25)$, which is (compatible with) the fitness landscape depicted in Figure 1.

[8] A. Bovier, L. Coquille, and R. Neukirch. The recovery of a recessive allele in a mendelian diploid model. J. Math. Biol., https://doi.org/10.1007/s00285-018-1240-z, 2018.
[9] A. Bovier and F. den Hollander. Metastability: A potential-theoretic approach, volume 351 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2015.
[10] A. Bovier and S.-D. Wang. Trait substitution trees on two time scales analysis. Markov. Proc. Rel. Fields, 19(4):607–642, 2013.
[11] R. Brink-Spalink and C. Smadi. Genealogies of two linked neutral loci after a selective sweep in a large population of stochastically varying size. Adv. Appl. Probab., 49(1):279–326, 2017.
[12] T. Britton and E. Pardoux. Stochastic epidemics in a homogeneous community. in preparation, 2017.

[13] A. J. Carter and G. P. Wagner. Evolution of functionally conserved enhancers can be accelerated in large populations: a population–genetic model. Proc. Royal Soc. London B: Biol. Sciences, 269(1494):953–960, 2002.

[14] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. Stochastic Process. Appl., 116(8):1127–1160, 2006.

[15] N. Champagnat, R. Ferrière, and G. Ben Arous. The canonical equation of adaptive dynamics: A mathematical view. Selection, 2:73–83., 2001.

[16] N. Champagnat, R. Ferrière, and S. Méléard. From individual stochastic processes to macroscopic models in adaptive evolution. Stochastic Models, 24(suppl. 1):2–44, 2008.

[17] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. Probab. Theor. Rel. Fields, 151(1-2):45–94, 2011.

[18] J.-R. Chazottes, P. Collet, and S. Méléard. Sharp asymptotics for the quasi-stationary distribution of birth-and-death processes. Probab. Theor. Rel. Fields, 164(1-2):285–332, 2016.

[19] J.-R. Chazottes, P. Collet, and S. Méléard. On time scales and quasi-stationary distributions for multitype birth-and-death processes. arXiv preprint arXiv:1702.05369, 2017.

[20] C. Coron, M. Costa, H. Leman, and C. Smadi. A stochastic model for speciation by mating preferences. J. Math. Biol., 76(6):1421–1463, 2018.

[21] C. Coron, S. Méléard, E. Porcher, and A. Robert. Quantifying the mutational meltdown in diploid populations. Am. Nat., 181(5):623–636, 2013.

[22] M. C. Cowperthwaite, J. Bull, and L. A. Meyers. From bad to good: Fitness reversals and the ascent of deleterious mutations. PLoS Comp. Biol., 2(10):e141, 2006.

[23] M. A. DePristo, D. L. Hartl, and D. M. Weinreich. Mutational reversions during adaptive protein evolution. Molec. Biol. Evol., 24(8):1608–1610, 2007.

[24] Dieckmann and R. Law. Moment approximations of individual-based models. In U. Dieckmann, R. Law, and J. A. J. Metz, editors, The geometry of ecological interactions: simplifying spatial complexity, pages 252–270. Cambridge University Press, 2000.

[25] U. Dieckmann and R. Law. The dynamical theory of coevolution: a derivation from stochastic ecological processes. J. Math. Biol., 34:579–612., 1996.

[26] R. Durrett and J. Mayberry. Traveling waves of selective sweeps. Ann. Appl. Probab, 21(2):699–744, 2011.

[27] S. N. Ethier and T. G. Kurtz. Markov processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

[28] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. Ann. Appl. Probab., 14(4):1880–1919, 2004.

[29] S. A. Geritz, J. A. Metz, É. Kisdi, and G. Meszéna. Dynamics of adaptation and evolutionary branching. Physical Review Letters, 78(10):2024, 1997.

[30] C. Giachetti and J. J. Holland. Altered replicase specificity is responsible for resistance to defective interfering particle interference of an sdi-mutant of vesicular stomatitis virus. J. Virol., 62(10):3614–3621, 1988.

[31] J. H. Gillespie. Molecular evolution over the mutational landscape. Evolution, 38(5):1116–1129, 1984.

[32] C. S. Gokhale, Y. Iwasa, M. A. Nowak, and A. Traulsen. The pace of evolution across fitness valleys. J. Theor. Biol., 259(3):613–620, 2009.

[33] H. Haeno, Y. E. Maruvka, Y. Iwasa, and F. Michor. Stochastic tunneling of two mutations in a population of cancer cells. PLoS one, 8(6):e65724, 2013.

[34] Y. Iwasa, F. Michor, and M. A. Nowak. Evolutionary dynamics of invasion and escape. J. Theor. Biol., 226(2):205–214, 2004.

[35] Y. Iwasa, F. Michor, and M. A. Nowak. Stochastic tunnels in evolutionary dynamics. Genetics, 166(3):1571–9, 2004.

[36] H. Leman. Convergence of an infinite dimensional stochastic process to a spatially structured trait substitution sequence. Stoch. Partial Differ. Equ. Anal. Comput., 4(4):791–826, 2016.

[37] R. E. Lenski, C. Ofría, R. T. Pennock, and C. Adami. The evolutionary origin of complex features. Nature, 423(6936):139–144, 2003.
[38] S. Maisnier-Patin, O. G. Berg, L. Liljas, and D. I. Andersson. Compensatory adaptation to the deleterious effect of antibiotic resistance in salmonella typhimurium. *Mol. Microbiol.*, 46(2):355–366, 2002.

[39] J. A. Metz, S. A. Geritz, G. Meszéna, F. J. Jacobs, and J. S. Van Heerwaarden. *Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction*. WP-95-099, 1995.

[40] F. B. G. Moore and S. J. Tonsor. A simulation of Wright’s shifting-balance process: migration and the three phases. *Evolution*, 48(1):69–80, 1994.

[41] F. B-G. Moore and S. J. Tonsor. A simulation of Wright’s shifting-balance process: migration and the three phases. *Evolution*, 48(1):69–80, 1994.

[42] P. J. O’Hara, S. T. Nichol, F. M. Horodyski, and J. J. Holland. Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. *Cell*, 36(4):915–924, 1984.

[43] S. Sagitov and M. C. Serra. Multitype Bienaymé–Galton–Watson processes escaping extinction. *Adv. Appl. Probab.*, 41(1):225–246, 2009.

[44] S. J. Schrag, V. Perrot, and B. R. Levin. Adaptation to the fitness costs of antibiotic resistance in escherichia coli. *Proc. Royal Soc. London B: Biological Sciences*, 264(1386):1287–1291, 1997.

[45] M. C. Serra. On the waiting time to escape. *J. Appl. Probab.*, 43(1):296–302, 2006.

[46] M. C. Serra and P. Haccou. Dynamics of escape mutants. *Theor. Pop. Biol.*, 72(1):167–178, 2007.

[46] C. Smadi. The effect of recurrent mutations on genetic diversity in a large population of varying size. *Act. Appl. Math.*, 149(1):11–51, 2017.

[47] V. C. Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probab. Statist.*, 12:345–386, 2008.

[48] R. van der Hofstad. *Random graphs and complex networks*, volume 43 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2016.

[49] M. J. Wade and C. J. Goodnight. Wright’s shifting balance theory: an experimental study. *Science*, 253(5023):1015, 1991.

[50] D. M. Weinreich and L. Chao. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution*, 59(6):1175–82, 2005.

[51] D. B. Weissman, M. M. Desai, D. S. Fisher, and M. W. Feldman. The rate at which asexual populations cross fitness valleys. *Theor. Pop. Biol.*, 75(4):286–300, 2009.

[52] S. Wright. Factor interaction and linkage in evolution. *Proc. Royal Soc. London B: Biological Sciences*, 162(986):80–104, 1965.

**Institut für Angewandte Mathematik, Rheinische Friedrich-Wilhelms-Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany.**

*E-mail address*: bovier@uni-bonn.de

**Univ. Grenoble Alpes, CNRS, Institut Fourier, F-38000 Grenoble, France**

*E-mail address*: loren.coquille@univ-grenoble-alpes.fr

**Irstea, UR LISC, Laboratoire d’Ingénierie pour les Systèmes Complexes, 9 avenue Blaise Pascal-CS 20085, 63178 Aubière, France and Complex Systems Institute of Paris Ile-de-France, 113 rue Nationale, Paris, France**

*E-mail address*: charline.smadi@irstea.fr