The Annals of Applied Probability
2019, Vol. 29, No. 6, 3541–3589
https://doi.org/10.1214/19-AAP1487

CROSSING A FITNESS VALLEY AS A METASTABLE TRANSITION IN A STOCHASTIC POPULATION MODEL[1]

BY ANTON BOVIER[*,2], LOREN COQUILLE[†,3] AND CHARLINE SMADI[‡,§]

*Rheinische Friedrich-Wilhelms-Universität Bonn**, Univ. Grenoble Alpes*[†],
Irstea[‡] *and Complex Systems Institute of Paris Ile-de-France*[§]

We consider a stochastic model of population dynamics where each individual is characterised by a trait in $\{0, 1, \dots, L\}$ 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 modeling of three basic mechanisms: heredity, mutations and competition. It was first developed in the 1990s, partly heuristically, by Metz, Geritz, Bolker, Pacala, Dieckmann, Law and coauthors [6, 7, 24, 25, 29, 39].

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

---

Received January 2018; revised February 2019.

[1] Supported in part by the Chair *Modélisation Mathématique et Biodiversité* of VEOLIA-Ecole Polytechnique-MNHN-F.X.

[2] Supported in part by the German Research Foundation (DFG) under Germany's Excellence Strategy, EXC2151, 390873048 (*Hausdorff Center for Mathematics* and *ImmunoSensation*), the Priority Programme SPP 1590 *Probabilistic Structures in Evolution* and the Collaborative Research Centre CRC 1060 *The Mathematics of Emergent Effects*.

[3] Supported in part 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.

MSC2010 subject classifications. 92D25, 60J80, 60J27, 92D15, 60F15, 37N25.

Key words and phrases. Eco-evolution, birth and death process with immigration, selective sweep, coupling, competitive Lotka–Volterra system with mutations.

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 [3, 14–17, 28]. Extensions of those results for more structured populations were investigated, for example, in [36, 47].

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 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 us 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 further 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 subpopulations 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 μ 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 subpopulations 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 [22, 37]. Our purpose is to present a precise analysis of how such an escape from a stable singularity happens 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 μK) 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 $\{1, 2, 3, \dots\}$, by $\mathbb{N}_0$ the set $\mathbb{N} \cup \{0\}$, and by $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ the set of nonnegative real numbers. For $n, m \in \mathbb{N}_0$ such that $n \leq m$, we also introduce the notation $[\![n, m]\!] := \{n, n+1, \dots, m\}$.

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 $[\![0, L]\!]$ on which the population evolves. To each trait $i \in [\![0, L]\!]$, 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 $(i, j) \in [\![0, L]\!]^2$

- a *competition kernel*: $c_{ij} \geq 0$, where $c_{ii}, c_{i0}, c_{iL} > 0$, for all $i \in [\![0, L]\!]$.

To be able to scale the effective size of a population, the competition kernel is scaled down by the so-called *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 $(i, j) \in [\![0, L]\!]^2$:

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

$$m_{ij}^{(1)} = \delta_{i+1,j} \quad \text{or} \quad m_{ij}^{(2)} = \frac{1}{2}(\delta_{i+1,j} + \delta_{i-1,j}), \tag{2.1}$$

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

We denote the stochastic process with the above mechanisms by X. The state of a population is an element of ${\mathbb{N}_0}^{L+1}$. 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 $X^K = (X_0^K(t), \dots, X_L^K(t)) = X/K$ and to think of this as an element of $\mathbb{R}^{L+1}$. Let e_i denote the ith unit vector in $\mathbb{R}^{L+1}$. The generator of X^K acts on bounded measurable functions $f : \mathbb{R}_+^{L+1} \to \mathbb{R}$, for all

$X^K \in (\mathbb{N}_0/K)^{L+1}$, as

$$
\begin{aligned}
(L^{(K)}f)(X^K) &= (1-\mu)K\sum_{i=0}^{L}(f(X^K+\mathrm{e}_i/K)-f(X^K))b_iX_i^K \\
&\quad + K\sum_{i=0}^{L}(f(X^K-\mathrm{e}_i/K)-f(X^K))\left(d_i+\sum_{j=0}^{L}c_{ij}X_j^K\right)X_i^K \qquad (2.2)\\
&\quad + \mu K\sum_{i=0}^{L}\sum_{j=0}^{L}(f(X^K+\mathrm{e}_j/K)-f(X^K))b_im_{ij}X_i^K.
\end{aligned}
$$

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

PROPOSITION 2.1 ([27], Chapter 11, Theorem 2.1). *Suppose that the initial conditions converge in probability to a deterministic limit, that is,* $\lim_{K\to\infty}X^K(0)=x(0)$. *Then, for each* $T\in\mathbb{R}_+$, *the rescaled process* $(X^K(t), 0\le t\le T)$ *converges in probability, as* $K\to\infty$, *to the deterministic process* $x^\mu=(x_0^\mu,\dots,x_L^\mu)$ *which is the unique solution to the following dynamical system*:

$$
\frac{dx_i^\mu}{dt}=\left((1-\mu)b_i-d_i-\sum_{i=0}^{L}c_{ij}x_j^\mu\right)x_i^\mu+\mu\sum_j m_{ji}b_jx_j^\mu,\quad i=0,\dots,L, \qquad (2.3)
$$

with initial condition $x(0)$.

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

$$
\bar{x}_i := \frac{b_i-d_i}{c_{ii}}\vee 0. \qquad (2.4)
$$

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 *invasion fitness*, f_{ij}, given by

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

The importance of the above two quantities follows from the properties of the limiting competitive Lotka–Volterra system (2.3) with $\mu=0$. Namely, if we assume

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

then the system (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 $\bar{x}_0 > 0$, $f_{i0} < 0$, $1 \le i \le L-1$, $f_{L0} > 0$ and $f_{0L} < 0$. Under these assumptions, all mutants created by the initial population 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, such that eventually an individual will reach the trait L, it will find 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 happens. The process that we want to describe can be seen as a manifestation of the phenomenon of *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 *carrying capacity*, K, which allows us to scale the population size to infinity. Apart from scaling the population size by taking $K \uparrow \infty$, we are also 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 μ tends to zero as a function of K.

3. Results.

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

ASSUMPTION 1.

- Viability of the resident population: $\bar{x}_0 > 0$.
- Fitness valley: All traits are unfit with respect to 0 except L:

$$f_{i0} < 0 \quad \text{for } i \in [\![1, L-1]\!] \text{ and } f_{L0} > 0. \tag{3.1}$$

- All traits are unfit with respect to L:

$$f_{iL} < 0 \quad \text{for } i \in [\![0, L-1]\!]. \tag{3.2}$$

- The following fitnesses are different:

$$f_{i0} \neq f_{j0} \quad \text{for all } i \neq j, \tag{3.3}$$

$$f_{iL} \neq f_{jL} \quad \text{for all } i \neq j. \tag{3.4}$$

See Figure 1.

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

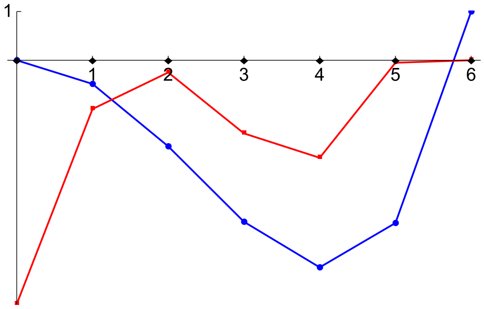

FIG. 1. *Example of a fitness landscape satisfying Assumption* 1 *with* $L = 6$. *Blue curve*: $i \mapsto f_{i0}$, *red curve*: $i \mapsto f_{iL}$.

without these 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 imposes 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* $(K, \mu) \to (\infty, \mu)$, *then* $\mu \to 0$. The first regime we are interested in is the case when μ 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* 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 modeled by a Moran model with immigration, and back mutations are not considered.

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

$$x^{\mu}(0) = (\bar{x}_0, 0, \dots, 0). \tag{3.5}$$

Then, for $i \in [\![0, L]\!]$, *as* $\mu \to 0$, *uniformly on bounded time intervals*,

$$\frac{\log[x_i^{\mu}(t \cdot \log(1/\mu))]}{\log(1/\mu)} \to x_i(t), \tag{3.6}$$

where $x_i(t)$ *is piecewise linear. More precisely*:

1. *In the case of* 1*-sided mutations*, $m_{ij} = m_{ij}^{(1)}$, *for* $i \in [\![0, L-1]\!]$,

$$x_i(t) = \begin{cases} -i & \textit{for } 0 \leq t < L/f_{L0}, \\ -i - (t - L/f_{L0}) \min\limits_{k \in [\![0,i]\!]} |f_{kL}| & \textit{for } t > L/f_{L0}, \end{cases} \tag{3.7}$$

and

$$x_L(t) = \begin{cases} -L + f_{L0}t & \textit{for } 0 \leq t < L/f_{L0}, \\ 0 & \textit{for } t > L/f_{L0}. \end{cases} \tag{3.8}$$

2. *In the case of* 2*-sided mutations*, $m_{ij} = m_{ij}^{(2)}$: *consider the sequence* $\{i_1, \dots, i_r\}$ *of "fitness records", defined recursively by* $i_1 = 0$, $i_k = \min\{i \in [\![0, L-1]\!] : f_{iL} < f_{i_{k-1}L}\}$,

$$x_i(t) = \begin{cases} -i \vee \bigl(-L - (L-i) + f_{L0}t\bigr) \\ \qquad \textit{for } 0 \leq t < L/f_{L0}, \\ -(L-i) \vee \max\limits_{k \in [\![0,i]\!]} \{-i - |f_{kL}|(t - L/f_{L0})\} \\ \quad \vee \max\limits_{k \in [\![1,r]\!]} \{-i_k - |i - i_k| - |f_{i_k L}|(t - L/f_{L0})\} \\ \qquad \textit{for } t > L/f_{L0}. \end{cases} \tag{3.9}$$

Moreover,

$$\bigl(x_0^\mu(t \log(1/\mu)), x_L^\mu(t \log(1/\mu))\bigr) \to \begin{cases} (\bar{x}_0, 0) & \textit{for } 0 \leq t < L/f_{L0}, \\ (0, \bar{x}_L) & \textit{for } t > L/f_{L0}. \end{cases} \tag{3.10}$$

The shape of $x(t) := (x_0(t), \dots, x_L(t))$ can be seen on Figures 2 and 3 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 $O(\mu^i)$ in a time of order $o(1)$, then the L-population, starting from a size $O(\mu^L)$, grows exponentially with rate f_{L0} until reaching order one (which takes a time L/f_{L0}) 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 $O(\mu^i)$ with a rate given by the lowest (negative) fitness of its left neighbours, $(\min_{j \in [\![0,i]\!]} |f_{jL}|)$ 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

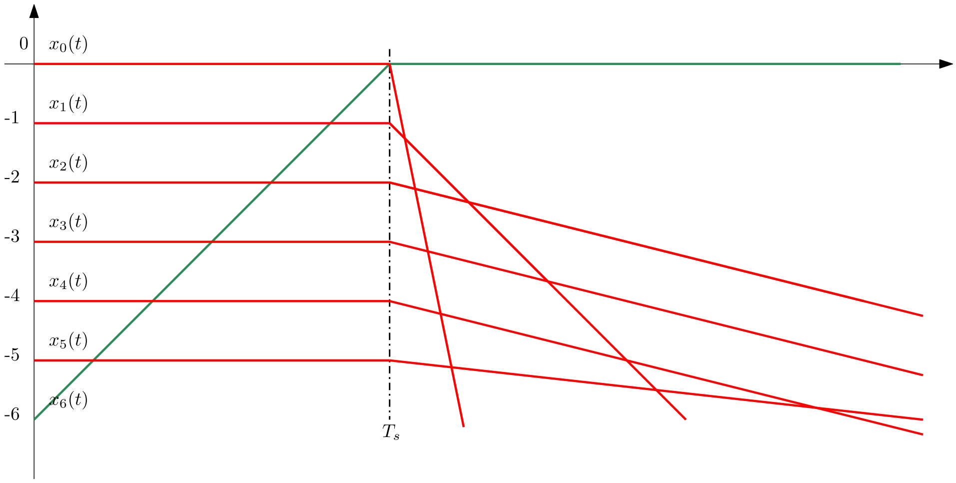

FIG. 2. *Graph of* $x(t)$ *in the* 1-*sided case* $m_{ij} = m_{ij}^{(1)}$ *for* $L = 6$ *and* $f_{60} = 1$, $(f_{06}, f_{16}, f_{26}, f_{36}, f_{46}, f_{56}) = (-5, -1, -0.25, -1.5, -2, -0.05)$, *which is the fitness landscape depicted in Figure* 1.

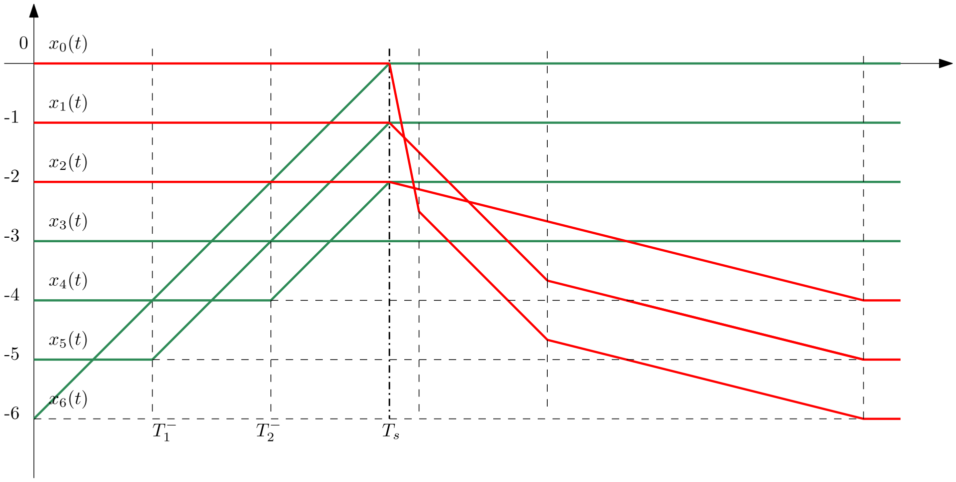

FIG. 3. *Graph of* $x(t)$ *in the* 2-*sided case* $m_{ij} = m_{ij}^{(2)}$ *for* $L = 6$ *and* $f_{60} = 1$, $(f_{06}, f_{16}, f_{26}) = (-5, -1, -0.25)$, *which is* (*compatible with*) *the fitness landscape depicted in Figure* 1.

respect to the competition exerted by the L-population, which has a size of order 1. As a consequence, x_0^μ has a dynamics close to this of the solution to

$$\dot{\tilde{x}}_0(t) = f_{0L}\tilde{x}_0(t), \tag{3.11}$$

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

$$\dot{\tilde{x}}_1(t) = f_{1L}\tilde{x}_1(t) + \mu\tilde{x}_0(t) = f_{1L}\tilde{x}_1(t) + \mu\tilde{x}_0(0)\mathrm{e}^{f_{0L}t}, \tag{3.12}$$

that is to say

$$x_1(t) \approx x_1(0)\mathrm{e}^{f_{1L}t} + \mu\frac{x_0(0)}{f_{0L} - f_{1L}}\left(\mathrm{e}^{f_{0L}t} - \mathrm{e}^{f_{1L}t}\right). \tag{3.13}$$

From these heuristics, we get that

$$\begin{aligned} x_1^\mu\big(t\log(1/\mu)\big) &\approx x_1^\mu(0)\mu^{|f_{1L}|t} + \mu\frac{x_0^\mu(0)}{|f_{1L}| - |f_{0L}|}\left(\mu^{|f_{0L}|t} - \mu^{|f_{1L}|t}\right) \\ &= \mu\left(C_1\mu^{|f_{1L}|t} + \frac{C_0}{|f_{1L}| - |f_{0L}|}\left(\mu^{|f_{0L}|t} - \mu^{|f_{1L}|t}\right)\right), \end{aligned} \tag{3.14}$$

where C_0 and C_1 are of order 1. We thus see that the leading order is $\mu^{1+\inf\{|f_{0L}|,|f_{1L}|\}t}$. 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 $\mu^{i+\inf\{|f_{0L}|,|f_{1L}|,\dots,|f_{iL}|\}t}$.

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 to the heuristics we have just described, except that mutants from the i-population ($i \in [\![1, L]\!]$) might also have an impact on the decrease rate of the $(i-1)$-population. This is the case if $x_i^\mu / x_{i-1}^\mu \geq C/\mu$, for a positive constant C. Under this condition, the number of type-$(i-1)$ individuals produced by mutations of type i-individuals has the same order as the type $(i-1)$ population size.

3.2. *Stochastic limit* $(K, \mu) \to (\infty, 0)$. When the mutation probability is small, the dynamics and time scale of the invasion process depends on the scaling of the mutation probability per reproductive event, μ, with respect to the carrying capacity K. We consider in this section mutation probabilities with two possible forms. Either,

$$\mu = f(K)K^{-1/\alpha} \quad \text{with } \alpha \geq 1 \text{ and } \left|\ln f(K)\right| = o(\ln K), \tag{3.15}$$

or

$$\mu = o(1/K). \tag{3.16}$$

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

For $v \geq 0$ and $0 \leq i \leq L$, let $T_v^{(K,i)}$ denote the first time the i-population reaches the size $\lfloor vK \rfloor$,

$$T_v^{(K,i)} := \inf\{t \geq 0, X_i(t) = \lfloor vK \rfloor\}. \tag{3.17}$$

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

$$t(L,\alpha) := \frac{L}{\alpha}\frac{1}{f_{L0}} + \sup\left\{\left(1-\frac{i}{\alpha}\right)\frac{1}{|f_{iL}|}, 0 \le i \le L-1\right\}, \tag{3.18}$$

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

$$T_0^{(K,\Sigma)} := \inf\left\{t \ge 0, \sum_{0\le i\le L-1} X_i(t) = 0\right\}. \tag{3.19}$$

With this notation, we have the following asymptotic result.

THEOREM 3.2. *Assume that* (3.15) *holds and that* $L < \alpha < \infty$. *Then there exist two positive constants* ε_0 *and* c *such that, for every* $0 < \varepsilon \le \varepsilon_0$,

$$\liminf_{K\to\infty} \mathbb{P}\left((1-c\varepsilon)\frac{1}{\alpha}\frac{L}{f_{L0}} < \frac{T_\varepsilon^{(K,L)}}{\log K} < \frac{T_{\bar{x}_L-\varepsilon}^{(K,L)}}{\log K} < (1+c\varepsilon)\frac{1}{\alpha}\frac{L}{f_{L0}}\right) \ge 1-c\varepsilon. \tag{3.20}$$

Moreover,

$$\frac{T_0^{(K,\Sigma)}}{\log K} \to t(L,\alpha) \quad \textit{in probability}, (K\to\infty) \tag{3.21}$$

and there exists a positive constant V *such that*

$$\limsup_{K\to\infty} \mathbb{P}\Big(\sup_{t\le e^{KV}} \big|X_L\big(T_{\bar{x}_L-\varepsilon}^{(K,L)}+t\big) - \bar{x}_L K\big| > c\varepsilon K\Big) \le c\varepsilon. \tag{3.22}$$

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 prove that the system performs small fluctuations around the deterministic evolution studied above: the i-population first stabilises around $O(K\mu^i)$ in a time of order one, then the L-population grows exponentially with rate f_{L0} until reaching order K (supercritical branching process, needs a time close to $L\log K/(\alpha f_{L0})$) while the other types stay stable, the swap between populations 0 and L then takes a time of order one, and finally, for $i \ne L$, the i-population decays exponentially from $O(K\mu^i)$ to extinction with a rate given by the lowest (negative) fitness of its left neighbours (subcritical branching process, needs a time close to $(\sup_{j\in[\![0,i]\!]}(1-j/\alpha)/|f_{jL}|)\log K)$. Thus, the time until extinction of all non-L populations is close to the constant (3.18) times $\log K$.

Note that Theorem 3.2 is close in spirit to the results of Durrett and Mayberry in [26], and some of our techniques of proof are similar. However, the processes they consider differ from ours at many levels. More precisely, they consider a Moran model with either fixed or growing population size (with a growth independent of the composition in traits of the population), and mutants with increasing fitnesses, while we work with a model with varying population size (where variations depend on the population composition via trait dependent competitive interactions) and allow negative fitnesses. Moreover in [26], all mutations have the same effect and back mutations are not considered, whereas it leads to interesting behaviour and more technicalities in our case. Finally, the way mutations are encoded in Moran-like models do not allow to distinguish between effects due to birth rate, death rate and competition. The class of models we consider allow a much wider variety of mutations (see Section 3 in [11] for a detailed discussion on these aspects).

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 longer 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 $0 < \rho < 1$,

$$\lambda(\rho) := \sum_{k=1}^{\infty} \frac{(2k)!}{(k-1)!(k+1)!}\rho^k(1-\rho)^{k+1}, \tag{3.23}$$

and, for $\lfloor\alpha\rfloor + 1 \le i \le L-1$, set $\rho_i := b_i/(b_i + d_i + c_{i0}\bar{x}_0)$.

THEOREM 3.3.

- *Assume that* (3.15) *holds*, $\alpha \notin \mathbb{N}$ *and* $1 < \alpha < L$. *Then there exist two positive constants* ε_0 *and* c, *and two exponential random variables* E_- *and* E_+ *with parameters*

$$\begin{aligned} &(1+c\varepsilon)\frac{\bar{x}_0 b_0\cdots b_{\lfloor\alpha\rfloor-1}}{|f_{10}|\cdots|f_{\lfloor\alpha\rfloor 0}|}\frac{f_{L0}}{b_L}\prod_{i=\lfloor\alpha\rfloor+1}^{L-1}\lambda(\rho_i) \quad \text{and} \\ &(1-c\varepsilon)\frac{\bar{x}_0 b_0\cdots b_{\lfloor\alpha\rfloor-1}}{|f_{10}|\cdots|f_{\lfloor\alpha\rfloor 0}|}\frac{f_{L0}}{b_L}\prod_{i=\lfloor\alpha\rfloor+1}^{L-1}\lambda(\rho_i), \end{aligned} \tag{3.24}$$

such that, for every $\varepsilon \le \varepsilon_0$,

$$\liminf_{K\to\infty}\mathbb{P}\Big(E_- \le \big(T^{(K,L)}_{\bar{x}_L-\varepsilon} \vee T^{(K,\Sigma)}_0\big)K\mu^L \le E_+\Big) \ge 1 - c\varepsilon. \tag{3.25}$$

- *There exists a positive constant V such that if μ satisfies*

$$K\mu \ll 1 \quad \text{and} \quad e^{VK} \gg 1/K\mu^L, \tag{3.26}$$

then the same conclusion holds, with the corresponding parameters, for E_- and E_+:

$$(1+c\varepsilon)\bar{x}_0 \frac{f_{L0}}{b_L} \prod_{i=1}^{L-1} \lambda(\rho_i) \quad \text{and} \quad (1-c\varepsilon)\bar{x}_0 \frac{f_{L0}}{b_L} \prod_{i=1}^{L-1} \lambda(\rho_i). \tag{3.27}$$

Moreover, under both assumptions, there exists a positive constant V such that

$$\limsup_{K\to\infty} \mathbb{P}\Big(\sup_{t \le \mathrm{e}^{KV}} \big| X_L\big(T^{(K,L)}_{\bar{x}_L-\varepsilon} + t\big) - \bar{x}_L K \big| > c\varepsilon K \Big) \le c\varepsilon.$$

In the first case, the typical trajectories of the process are as follows: mutant populations of type i, for $1 \le i \le \lfloor \alpha \rfloor$, reach a size of order $K\mu^i \gg 1$ in a time of order $b_{i-1} \log K/f_{i0}$ (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 $\lfloor \alpha \rfloor + 1 \le i \le L$, describe a.s. finite excursions, whose a proportion of order μ 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 $d_i + c_{i0}\bar{x}_0$ excepting the first individual. Hence, $\mu\lambda(\rho_i)$ is the approximated probability for a type i-population ($\lfloor \alpha \rfloor + 1 \le i \le L-1$) 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 μ 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 crosses the fitness valley with 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 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 are 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 μ is too small, the population gets 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 provide some bounds in the next results.

The previous theorem (Theorem 3.3) provided a wide range of mutation probabilities μ 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,

$$\rho_0(K) := \sqrt{K}\exp\bigl(-K\bigl(b_0 - d_0 + d_0 \ln(d_0/b_0)\bigr)\bigr). \tag{3.28}$$

More precisely, it is stated in [18] that $\mathbb{E}_\nu[T_0^{(\mathrm{mono})}] = 1/\rho_0(K)$, where ν is the stationary distribution of a monomorphic 0-population, and $T_0^{(\mathrm{mono})}$ its extinction time. We also need to introduce the two stopping times

$$\begin{aligned} &T_0 := \inf\{t \geq 0, X_i(t) = 0, \forall\, 0 \leq i \leq L\} \quad \text{and} \\ &B_L := \inf\{t \geq 0, X_L(t) > 0\}, \end{aligned} \tag{3.29}$$

as well as the following assumption.

ASSUMPTION 2. The birth and death rates satisfy the conditions

$$b_i < d_i, \quad 1 \leq i \leq L-1. \tag{3.30}$$

Then we have the following result.

THEOREM 3.4. *Suppose that Assumption 1 holds.*

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

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, that is, if $f_{L0} > 0$ 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 nontrivial 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 $i = 1, \dots, L-1$. In the 1-sided case, those stay of order $K\mu^i$, while in the 2-sided case, they stay of order $K\mu^{\min\{i,L-i\}}$. There is no complicated decay phase as in Section 6.3.2, and its stochastic analog.
- If the mutation probability μ 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 $L!$ in the small mutation regime (Theorem 3.3) and will stay the same in the large mutation regime (Theorem 3.2).

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 [38, 43] and in viruses [30, 41], 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 [40, 49, 52], 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 Gillepsie [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, 32, 33, 50, 51] 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

$$b_i = b \quad \text{and} \quad |f_{ij}| = |f_{ji}| \quad \forall\, 0 \leq i, j \leq L, \tag{5.1}$$

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 [2, 34, 42, 44, 45] 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 multitype branching processes. 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* $\zeta \in \{1, 2\}$, $(b_0, \dots, b_L) \in (\mathbb{R}^+)^{L+1}$, $(\ell_0, \dots, \ell_L) \in (\mathbb{R}^+)^{L+1}$, $(p_0, \dots, p_L) \in (\mathbb{R}^+)^{L+1}$ *and* $(f_0, \dots, f_L) \in \mathbb{R}^{L+1}$ *such that*

$$f_i \neq f_j \quad \textit{for all } i \neq j. \tag{6.1}$$

Let

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

Then the solution to the linear system

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

with initial condition

$$y(0) = (\ell_0\mu^{p_0}, \dots, \ell_L\mu^{p_L}), \tag{6.4}$$

satisfies

$$\lim_{\mu\to 0} \frac{\log(y_i(t\log(1/\mu)))}{\log(1/\mu)} = -m_i(t) := -\min_{\substack{\gamma,\alpha\in[\![0,L]\!]:\\ \ell_\gamma\neq 0, \gamma\le\alpha\le i}} \{i - \gamma + p_\gamma - tf_\alpha\}, \tag{6.5}$$

with the convention $p/0 = \infty$, *for* $p \ge 0$.

Note that Assumption 6.1 intuitively ensures that contributions coming from mutants of different traits are different (when computing the growth or decrease rate of a given trait). It is then clear which one wins in equation (6.5). If this assumption does not hold, it could happen that prefactors (in front of powers of μ) matter, and we do not want to enter into such an analysis. Mathematically, it ensures the matrix M_ζ in (6.2) to be diagonalisable for μ small enough, and thus to obtain explicit expressions for change of basis matrices in the proof below.

PROOF OF LEMMA 6.1. Under assumption (6.1), the matrix M_ζ in (6.2) is diagonalisable for μ small enough: it can be checked that $M_\zeta = SDS^{-1}$ with

$$D = ((f_i - b_i\mu)\delta_{ij})_{0\le i,j\le L}, \tag{6.6}$$

$$\begin{aligned} S &= \left(\left(\frac{\zeta}{\mu}\right)^{L-i} \frac{\prod_{k=i+1}^{L}(f_j - f_k) + \mu(b_k - b_j)}{\prod_{\ell=i}^{L-1} b_\ell} \mathbf{1}_{[i\ge j]}\right)_{0\le i,j\le L} \\ &=: (\mu^{i-L}C_{ij}\mathbf{1}_{[i\ge j]})_{0\le i,j\le L}, \end{aligned} \tag{6.7}$$

$$\begin{aligned} S^{-1} &= \left(\left(\frac{\mu}{\zeta}\right)^{L-j} \frac{\prod_{\ell=j}^{L-1} b_\ell}{\prod_{k=j,k\neq i}^{L}(f_i - f_k) + \mu(b_k - b_i)} \mathbf{1}_{[i\ge j]}\right)_{0\le i,j\le L} \\ &=: (\mu^{L-j}C'_{ij}\mathbf{1}_{[i\ge j]})_{0\le i,j\le L}. \end{aligned} \tag{6.8}$$

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

$$y(t) = \exp(tM_\zeta)y(0) = S\exp(tD)S^{-1}y(0), \tag{6.9}$$

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

$$\begin{aligned} y_i(t) &= \sum_{\alpha,\gamma=0}^{L} S_{i\alpha}\mathrm{e}^{t(f_\alpha - b_\alpha\mu)} S^{-1}_{\alpha\gamma}\ell_\gamma \mu^{p_\gamma} \\ &= \sum_{\gamma:\ell_\gamma\neq 0}\sum_{\gamma\le\alpha\le i} \mu^{i-\gamma+p_\gamma}\mathrm{e}^{t(f_\alpha-b_\alpha\mu)} \cdot C_{i\alpha}C'_{\alpha\gamma}\ell_\gamma. \end{aligned} \tag{6.10}$$

Thus,

$$y_i\big(t\log(1/\mu)\big) = \sum_{\gamma:\ell_\gamma\neq 0}\sum_{\gamma\le\alpha\le i} \mu^{i-\gamma+p_\gamma-t(f_\alpha-b_\alpha\mu)} \cdot \big(C_{i\alpha}C'_{\alpha\gamma}\ell_\gamma\big). \tag{6.11}$$

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

$$y_i\big(t\log(1/\mu)\big) = C\mu^{m_i(t)}\big(1 + o(1)\big), \tag{6.12}$$

which implies the assertion of (6.5) and concludes the proof of the lemma. □

6.1. *Before the swap.*

6.1.1. *Time interval* $0 \le t \le T^-_{L-1}$. If $m_{ij} = m^{(2)}_{ij}$, let $\zeta = 2$ and

$$\begin{aligned} \tau^-_{L-1}(\varepsilon,\mu) = {}& \inf\{t : \exists\, i \in [\![0, L]\!] \text{ s.t. } x^\mu_i(t) > \mu^{i-\varepsilon}\} \\ & \wedge \inf\{t : |x^\mu_0(t) - \bar{x}_0| > \varepsilon\} \wedge \inf\{t : x^\mu_L(t) > \mu^{L-2+\varepsilon}\}, \end{aligned} \tag{6.13}$$

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

$$\begin{aligned} \tau^-_{L-1}(\varepsilon,\mu) = {}& \inf\{t : \exists\, i \in [\![0, L]\!] \text{ s.t. } x^\mu_i(t) > \mu^{i-\varepsilon}\} \\ & \wedge \inf\{t : |x^\mu_0(t) - \bar{x}_0| > \varepsilon\} \wedge \inf\{t : x^\mu_L(t) > \varepsilon\} \end{aligned} \tag{6.14}$$

and define

$$T^-_{L-1} := \lim_{\varepsilon\to 0}\lim_{\mu\to 0} \frac{\tau^-_{L-1}(\varepsilon,\mu)}{\log(1/\mu)}. \tag{6.15}$$

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

$$\frac{dx^\mu_i}{dt} \ge (f_{i0} - C\varepsilon)x^\mu_i + \mu\Big(\frac{b_{i-1}}{\zeta}x^\mu_{i-1} - b_i x^\mu_i\Big). \tag{6.16}$$

Hence, by the Gronwall lemma, x^μ is bigger than the solution to $\frac{dy}{dt} = M_\zeta y$ with $f_i = f_{i0} - C\varepsilon$. Applying Lemma 6.1 with $y(0) = (\bar{x}_0, 0, \dots, 0)$, and thus $\ell_0 = \bar{x}_0$, $p_0 = 0$, $\ell_i = 0$ for $i \neq 0$, we get, using (3.1), for ε small enough and $t > 0$,

$$\lim_{\mu\to 0} \frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} \geq \begin{cases} -i - C\varepsilon t & \text{for } i = 0, \dots, L-1, \\ -L + t(f_{L0} - C\varepsilon) & \text{for } i = L. \end{cases} \tag{6.17}$$

On the other hand, on the same time interval, we have, for some positive C, the upper bound

$$\frac{dx_i^\mu}{dt} \leq (f_{i0} + C\varepsilon)x_i^\mu + \mu\left(\frac{b_{i-1}}{\zeta}x_{i-1}^\mu - b_i x_i^\mu\right) + E_i, \tag{6.18}$$

where, until τ_{L-1}^-, with $\kappa := \sup b_i/2$,

$$(E,\zeta) = \begin{cases} (\kappa\mu\cdot(\mu^{1-\varepsilon}, \mu^{2-\varepsilon}, \dots, \mu^{L-1-\varepsilon}, \mu^{L-2+\varepsilon}, 0), 2) \\ \quad \text{if } m_{ij} = m_{ij}^{(2)}, \\ ((0, 0, \dots, 0), 1) \\ \quad \text{if } m_{ij} = m_{ij}^{(1)}. \end{cases} \tag{6.19}$$

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

$$\begin{aligned} y(t) &= \mathrm{e}^{tM_\zeta}\left(y(0) + \left(\int_0^t \mathrm{e}^{-sM_\zeta}\,ds\right)E\right) \\ &= \mathrm{e}^{tM_\zeta}y(0) + S\left(\int_0^t \mathrm{e}^{(t-s)D}\,ds\right)S^{-1}E \\ &= \mathrm{e}^{tM_\zeta}y(0) + S\left(\frac{\mathrm{e}^{(f_i - b_i\mu)t}}{f_i - b_i\mu}\delta_{ij}\right)S^{-1}E - S\left(\frac{1}{f_i - b_i\mu}\delta_{ij}\right)S^{-1}E. \end{aligned} \tag{6.20}$$

Now we compute the order of magnitude of each term as in (6.11) in the proof of Lemma 6.1 and show that the two terms in (6.20) involving E are negligible with respect to the main term. Set

$$e_1(t) := S(\mathrm{e}^{(f_i - b_i\mu)t}\delta_{ij})S^{-1}E, \qquad e_2 := S(\delta_{ij})S^{-1}E. \tag{6.21}$$

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

$$\begin{aligned} &(e_1(t\log(1/\mu)) \vee e_2)_i \\ &\quad = O\big(\mu^{\min_{\gamma\in[\![0,L-1]\!],\gamma\leq i}\{i-\gamma+(2+\gamma-\varepsilon)\mathbf{1}_{[\gamma<L-1]}+\{i-\gamma+(\gamma+\varepsilon)\}\mathbf{1}_{[\gamma=L-1]}\}}\big) \\ &\quad = O\big(\mu^{(i+\varepsilon)\mathbf{1}_{[i=L-1]}+(i+2-\varepsilon)\mathbf{1}_{[i<L-1]}}\big) = o(\mu^i), \end{aligned} \tag{6.22}$$

and if $i = L$ we get

$$(e_1(t\log(1/\mu)) \vee e_2)_L = O(\mu^{L-t(f_{L0}+C\varepsilon)+\varepsilon}) = o(\mu^{L-t(f_{L0}+C\varepsilon)}). \tag{6.23}$$

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

$$\begin{aligned} \lim_{\mu\to 0} \frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} &\leq \lim_{\mu\to 0} \frac{\log(y_i(t\log(1/\mu)))}{\log(1/\mu)} \\ &= \begin{cases} -i+C\varepsilon t & \text{for } i = 0, \dots, L-1, \\ -L + t(f_{L0} + C\varepsilon) & \text{for } i = L. \end{cases} \end{aligned} \tag{6.24}$$

Finally, observe that, as the only growing population is the one with trait L,

$$T_{L-1}^- = \lim_{\varepsilon\to 0} \lim_{\mu\to 0} \frac{\tau_{L-1}^-(\varepsilon,\mu)}{\log(1/\mu)} = \begin{cases} 2/f_{L0} & \text{for } m_{ij} = m_{ij}^{(2)}, \\ L/f_{L0} & \text{for } m_{ij} = m_{ij}^{(1)}. \end{cases} \tag{6.25}$$

In the case $m_{ij} = m_{ij}^{(1)}$, the proof continues directly with Section 6.2.

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

$$\begin{aligned} \tau_{L-2}^-(\varepsilon,\mu) = \inf\{t : \exists\, i \in [\![0, L-1]\!] \text{ s.t. } x_i^\mu(t) > \mu^{i-\varepsilon}\} \\ \wedge \inf\{t : |x_0^\mu(t) - \bar{x}_0| > \varepsilon\} \wedge \inf\{t : x_{L-1}^\mu(t) > \mu^{L-3+\varepsilon}\} \\ \wedge \inf\{t : x_L^\mu(t) > \mu^{L-4+\varepsilon}\} \end{aligned} \tag{6.26}$$

and define

$$T_{L-2}^- := \lim_{\varepsilon\to 0} \lim_{\mu\to 0} \frac{\tau_{L-2}^-(\varepsilon,\mu)}{\log(1/\mu)}. \tag{6.27}$$

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

$$\begin{aligned} \frac{dx_i^\mu}{dt} &\geq (f_{i0} - C\varepsilon)x_i^\mu \\ &\quad + \mu\left(\frac{b_{i-1}}{2} x_{i-1}^\mu \mathbf{1}_{[i<L-1]} + \frac{b_{i+1}}{2} x_{i+1}^\mu \mathbf{1}_{[i=L-1]} - b_i x_i^\mu\right). \end{aligned} \tag{6.28}$$

Hence, by the Gronwall lemma, and notation 6.3, x^μ is bigger than the solution to

$$\frac{dy}{dt} = \begin{pmatrix} M_{\text{left}}(L-2) & 0 \\ 0 & M_{\text{right}}(1) \end{pmatrix} y =: M'(L-2,1)y, \tag{6.29}$$

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

$$M_{\text{right}} = \begin{pmatrix} f_{L-1,0} - C\varepsilon & \frac{\mu}{2} b_L \\ 0 & f_{L0} - C\varepsilon \end{pmatrix}. \tag{6.30}$$

Applying then twice Lemma 6.1, once with $M_{\mathrm{left}}(L-2)$ and $y_{\mathrm{left}} = (y_0, \dots, y_{L-2})$ and once with $M_{\mathrm{right}}(1)$ (treated as $M(1)$ with "reversed indices", that is, f_i, b_i replaced by f_{L-i}, b_{L-i}) and $y_{\mathrm{right}} = (y_{L-1}, y_L)$, with

$$y(0) = (\bar{x}_0, \mu, \mu^2, \dots, \mu^{L-1}, \mu^{L-2}), \tag{6.31}$$

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

$$\begin{aligned} &\lim_{\mu\to 0} \frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} \\ &\quad \geq \begin{cases} -i - C\varepsilon t & \text{for } i = 0, \dots, L-2, \\ -(L-1) + t(f_{L0} - C\varepsilon) & \text{for } i = L-1, \\ -(L-2) + t(f_{L0} - C\varepsilon) & \text{for } i = L. \end{cases} \end{aligned} \tag{6.32}$$

On the other hand, we have the upper bound

$$\begin{aligned} \frac{dx_i^\mu}{dt} &\leq (f_{i0} + C\varepsilon)x_i^\mu \\ &\quad + \mu\left(\frac{b_{i-1}}{2}x_{i-1}^\mu \mathbf{1}_{[i<L-1]} + \frac{b_{i+1}}{2}x_{i+1}^\mu \mathbf{1}_{[i=L-1]} - b_i x_i^\mu\right) + E_i, \end{aligned} \tag{6.33}$$

where until τ_{L-2}^- we have

$$E = \mu \cdot (\mu^{1-\varepsilon}, \mu^{2-\varepsilon}, \dots, \mu^{L-2-\varepsilon}, \mu^{L-3+\varepsilon}, \mu^{L-2-\varepsilon}, \mu^{L-3+\varepsilon}). \tag{6.34}$$

By the Gronwall lemma, x^μ 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 (6.32) also as an upper bound, with $f_{L0} - C\varepsilon$ replaced by $f_{L0} + C\varepsilon$, and $-i - C\varepsilon t$ replaced by $-i + C\varepsilon t$. Finally, observe that

$$T_{L-2}^- = \lim_{\varepsilon\to 0}\lim_{\mu\to 0} \frac{\tau_{L-2}^-(\varepsilon, \mu)}{\log(1/\mu)} = \frac{4}{f_{L0}}. \tag{6.35}$$

6.1.3. *Induction until $T_{L/2}^-$.* In this section, if L is odd, then $L/2$ has to be replaced by $\lfloor L/2 \rfloor$. For $k \in \{3, \dots, L/2\}$, we treat the time interval $T_{L-k}^- \leq t \leq T_{L-(k+1)}^-$. Let $m_{ij} = m_{ij}^{(2)}$ and

$$\begin{aligned} \tau_{L-k}^-(\varepsilon, \mu) = {} & \inf\{t : \exists\, i \in [\![0, L-k+1]\!] \text{ s.t. } x_i^\mu(t) > \mu^{i-\varepsilon}\} \\ & \wedge \inf\{t : |x_0^\mu(t) - \bar{x}_0| > \varepsilon\} \\ & \wedge \inf\{t : \exists\, j \in [\![1, k]\!] \text{ s.t. } x_{L-k+j}^\mu(t) > \mu^{(L-k+j)-2j+\varepsilon}\} \end{aligned} \tag{6.36}$$

and define

$$T^-_{L-k} := \lim_{\varepsilon\to 0}\lim_{\mu\to 0}\frac{\tau^-_{L-k}(\varepsilon,\mu)}{\log(1/\mu)}. \tag{6.37}$$

For $t\in[\tau^-_{L-k}(\varepsilon,\mu),\tau^-_{L-k-1}(\varepsilon,\mu)]$, we have the lower bound

$$\begin{aligned}\frac{dx_i^\mu}{dt} &\ge (f_{i0}-C\varepsilon)x_i^\mu \\ &\quad + \mu\biggl(\frac{b_{i-1}}{2}x_{i-1}^\mu \mathbf{1}_{[i<L-k+1]}+\frac{b_{i+1}}{2}x_{i+1}^\mu \mathbf{1}_{[i\ge L-k+1]}-b_i x_i^\mu\biggr).\end{aligned} \tag{6.38}$$

Hence, by the Gronwall lemma, x^μ is bigger than the solution to

$$\frac{dy}{dt}=\begin{pmatrix} M_{\mathrm{left}}(L-k) & 0 \\ 0 & M_{\mathrm{right}}(k-1)\end{pmatrix} y =: M'(L-k,k-1)y, \tag{6.39}$$

where $M_{\mathrm{left}}(L-k)=M_2(L-k)$ with $f_i=f_{i0}-C\varepsilon$ and

$$M_{\mathrm{right}}(k)=\begin{pmatrix} f_{L-k,0}-C\varepsilon & \frac{\mu}{2}b_{L-k+1} & & 0 \\ & \ddots & \ddots & \\ & & f_{L-1,0}-C\varepsilon & \frac{\mu}{2}b_L \\ 0 & & & f_{L0}-C\varepsilon \end{pmatrix}. \tag{6.40}$$

Applying twice Lemma 6.1, once with M_{left} and $y_{\mathrm{left}}=(y_0,\dots,y_{L-k-1})$ and once with M_{right} (treated as $M(k)$ with "reversed indices", f_i, b_i replaced by f_{L-i}, b_{L-i}) and $y_{\mathrm{right}}=(y_{L-k},\dots,y_L)$, with

$$y(0)=\bigl(\bar{x}_0,\mu,\mu^2,\dots,\mu^{L-k},\mu^{L-k+1},\mu^{L-k},\mu^{L-k-1},\dots,\mu^{L-2k}\bigr), \tag{6.41}$$

up to $o_\varepsilon(1)$ terms in the powers of μ due to the range of possible initial conditions coming from the previous phase, we get

$$\begin{aligned}&\lim_{\mu\to 0}\frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} \\ &\quad\ge \begin{cases} -i-C\varepsilon t & \text{for } i=0,\dots,L-k, \\ -i+j-1+t(f_{L0}-C\varepsilon) & \text{for } i=L-k+j,\ j=1,\dots,k.\end{cases}\end{aligned} \tag{6.42}$$

On the other hand, we have the upper bound

$$\begin{aligned}\frac{dx_i^\mu}{dt} &\le (f_{i0}+C\varepsilon)x_i^\mu \\ &\quad + \mu\biggl(\frac{b_{i-1}}{2}x_{i-1}^\mu \mathbf{1}_{[i<L-k+1]}+\frac{b_{i+1}}{2}x_{i+1}^\mu \mathbf{1}_{[i\ge L-k+1]}-b_i x_i^\mu\biggr)+E_i,\end{aligned} \tag{6.43}$$

where on the time interval $[\tau^-_{L-k}(\varepsilon,\mu),\tau^-_{L-k-1}(\varepsilon,\mu)]$ we have

$$\begin{aligned}E=\mu\cdot\big(&\mu^{1-\varepsilon},\mu^{2-\varepsilon},\dots,\mu^{L-k-\varepsilon},\mu^{L-k-1+\varepsilon},\mu^{L-k-\varepsilon},\\&\mu^{L-k-1+\varepsilon},\mu^{L-k-2+\varepsilon},\dots,\mu^{L-1-2(k-1)+\varepsilon}\big).\end{aligned}\tag{6.44}$$

By the Gronwall lemma, x^μ is thus 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 the constant in the two blocks), we get (6.42) also as an upper bound, with $f_{L0}-C\varepsilon$ replaced by $f_{L0}+C\varepsilon$, and $-i-C\varepsilon t$ replaced by $-i+C\varepsilon t$. Finally, observe that

$$T^-_{L-k}=\lim_{\varepsilon\to0}\lim_{\mu\to0}\frac{\tau^-_{L-k}(\varepsilon,\mu)}{\log(1/\mu)}=\frac{2k}{f_{L0}}.\tag{6.45}$$

6.2. *The swap.* Let $m_{ij}=m^{(2)}_{ij}$ and

$$\begin{aligned}\tau^s(\varepsilon,\mu)=&\inf\{t:\exists\, i\in[\![0,L/2]\!]\text{ s.t. }x_i^\mu(t)>\mu^{i-\varepsilon}\}\wedge\inf\{t:x_0^\mu(t)<\varepsilon\}\\&\wedge\inf\{t:\exists\, i\in[\![L/2,L]\!]\text{ s.t. }x_i^\mu(t)>\mu^{L-i-\varepsilon}\}\\&\wedge\inf\{t:x_L^\mu(t)>\bar{x}_L-\varepsilon\},\end{aligned}\tag{6.46}$$

or $m_{ij}=m^{(1)}_{ij}$ and

$$\begin{aligned}\tau^s(\varepsilon,\mu)=&\inf\{t:\exists\, i\in[\![0,L-1]\!]\text{ s.t. }x_i^\mu(t)>\mu^{i-\varepsilon}\}\\&\wedge\inf\{t:x_0^\mu(t)<\varepsilon\}\wedge\inf\{t:x_L^\mu(t)>\bar{x}_L-\varepsilon\}.\end{aligned}\tag{6.47}$$

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

$$\frac{dx^\mu_\chi}{dt}\ge(b_\chi-d_\chi-c_{\chi0}x_0^\mu-c_{\chi L}x_L^\mu-C\mu)x^\mu_\chi-C'\mu^{1-\varepsilon}\tag{6.48}$$

and the upper bounds

$$\frac{dx^\mu_\chi}{dt}\le(b_\chi-d_\chi-c_{\chi0}x_0^\mu-c_{\chi L}x_L^\mu)x^\mu_\chi+C'\mu^{1-\varepsilon}.\tag{6.49}$$

Let $(\tilde{x}_0,\tilde{x}_L)$ denote the solution of the unperturbed system, that is, of

$$\frac{dx_\chi}{dt}=(b_\chi-d_\chi-c_{\chi0}x_0-c_{\chi L}x_L)x_\chi.\tag{6.50}$$

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 ε-neighbourhood of this equilibrium from initial conditions $(\bar{x}_0-\varepsilon,\varepsilon)$ is of order $O(1)$. Applying the Gronwall lemma to the function $|x_0^\mu(t)-\tilde{x}_0(t)|^2+|x_L^\mu(t)-\tilde{x}_L(t)|^2$, (6.48) and (6.49) imply that on any compact time interval $(x_0^\mu,x_L^\mu)\to(\tilde{x}_0,\tilde{x}_L)$ as $\mu\downarrow0$. Moreover, for all

μ small enough, the system (x_0^μ, x_L^μ) has a stable equilibrium that converges to $(0, \bar{x}_L)$, as $\mu \downarrow 0$.

For the populations $(x_1^\mu, \dots, x_{L-1}^\mu)$, we have, for $t \in [\tau_{L/2}^-(\varepsilon, \mu), \tau^s(\varepsilon, \mu)]$, the lower bounds

$$\begin{aligned}\frac{dx_i^\mu}{dt} &\geq (b_i - d_i - c_{i0}\bar{x}_0 - c_{iL}\bar{x}_L - C\varepsilon)x_i^\mu \\ &\quad + \mu\left(\frac{b_{i-1}}{2}x_{i-1}^\mu \mathbf{1}_{[i<L/2]} + \frac{b_{i+1}}{2}x_{i+1}^\mu \mathbf{1}_{[i\geq L/2]} - b_i x_i^\mu\right),\end{aligned} \tag{6.51}$$

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

$$\frac{dx_i^\mu}{dt} \geq (b_i - d_i - c_{i0}\bar{x}_0 - c_{iL}\bar{x}_L - C\varepsilon)x_i^\mu + \mu\big(b_{i-1}x_{i-1}^\mu - b_i x_i^\mu\big), \tag{6.52}$$

in the case $m_{ij} = m_{ij}^{(1)}$. We have decoupled traits 0 and L from traits $1, \dots, L-1$. 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 hold:

1. In the case $m_{ij} = m_{ij}^{(1)}$, the solution $(x_1^\mu, \dots, x_{L-1}^\mu)$ is smaller than the solution to

$$\frac{dy}{dt} = M_1(L-2)y \tag{6.53}$$

with $f_i = b_i - d_i - c_{i0}\bar{x}_0 - c_{iL}\bar{x}_L - C\varepsilon < 0$ and initial conditions:

$$y(0) = \big(\mu, \mu^2, \dots, \mu^{L-1}\big) \tag{6.54}$$

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

$$\lim_{\mu\to 0} \frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} \geq -i + t \sup_{1\leq\alpha\leq i} f_\alpha \quad \text{for } i = 1, \dots, L-1. \tag{6.55}$$

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

$$\lim_{t\to 0}\lim_{\mu\to 0} \frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} \geq -i \quad \text{for } i = 1, \dots, L-1. \tag{6.56}$$

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

$$\frac{dy}{dt} = \begin{pmatrix} M_{\text{left}}(L/2-2) & 0 \\ 0 & M_{\text{right}}(L/2-1) \end{pmatrix} y =: M'(L/2-2, L/2-1)y \tag{6.57}$$

with $f_i = b_i - d_i - c_{i0}\bar{x}_0 - c_{iL}\bar{x}_L - C\varepsilon < 0$ and initial conditions

$$y(0) = \big(\mu, \mu^2, \dots, \mu^{L/2+1}, \mu^{L/2}, \mu^{L/2-1} \dots, \mu\big) \tag{6.58}$$

up to $o_\varepsilon(1)$ terms in the powers of μ 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

$$y(0) = (\mu, \mu^2, \dots, \mu^{\lfloor L/2 \rfloor}, \mu^{\lfloor L/2 \rfloor}, \mu^{\lfloor L/2 \rfloor - 1} \dots, \mu) \tag{6.59}$$

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

$$\lim_{t\to 0} \lim_{\mu\to 0} \frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} \ge -\min\{i, L-i\} \quad \text{for } i = 1, \dots, L-1. \tag{6.60}$$

On the other hand, we have the upper bound

$$\frac{dx_i^\mu}{dt} \le F x_i^\mu + \mu^{1-\varepsilon}, \tag{6.61}$$

with some $F > 0$. Thus, by the Gronwall lemma,

$$x_i^\mu(t) \le \mu^{i-\varepsilon} \quad \text{for } t < \frac{\varepsilon}{F}\log(1/\mu) \text{ and for } i = 1, \dots L/2 \tag{6.62}$$

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

$$\lim_{\mu\to 0} \frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} \le \begin{cases} -\min\{i, L-i\} & \text{if } m_{ij} = m_{ij}^{(2)}, \\ -i & \text{if } m_{ij} = m_{ij}^{(1)}. \end{cases} \tag{6.63}$$

The duration of the swap vanishes (on the time scale $\log(1/\mu)$) in the limit $\mu \to 0$. We thus have $T^s = T^-_{L/2}$.

6.3. *After the swap.*

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

$$\begin{aligned} \tau^+(\varepsilon,\mu) = \inf\{t : \exists\, i \in [\![1, L-1]\!] \text{ s.t. } x_i^\mu(t) > \mu^{i-2\varepsilon}\} \\ \wedge \inf\{t : |x_L^\mu(t) - \bar{x}_L| > \varepsilon\}. \end{aligned} \tag{6.64}$$

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

$$\frac{dx_i^\mu}{dt} \ge (f_{iL} - C\varepsilon)x_i^\mu + \mu(b_{i-1}x_{i-1}^\mu - b_i x_i^\mu). \tag{6.65}$$

Hence, by the Gronwall lemma, and notation (6.3), x^μ is bigger than the solution to $\frac{dy}{dt} = M_1 y$ with $f_i = f_{iL} - C\varepsilon$. Applying Lemma 6.1 with

$$y(0) = (\varepsilon, \mu, \dots, \mu^{L-1}, \bar{x}_L - \varepsilon), \tag{6.66}$$

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

$$\begin{aligned}\lim_{\mu\to 0}\frac{\log(x_i^\mu(t\log(1/\mu)))}{\log(1/\mu)} &\geq -\min_{\alpha\leq i}\{i - L\delta_{i,L} - t(f_{\alpha L} - C\varepsilon)\}\\ &= -i + L\delta_{i,L} - t\min_{\alpha\in[\![0,i]\!]}|f_{\alpha L}| + tC\varepsilon.\end{aligned}\tag{6.67}$$

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

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

$$\begin{aligned}\tau^+(\varepsilon,\mu) = \inf\{t > T^s : \exists\, i\in[\![0, L-1]\!] : x_i^\mu(t) > \varepsilon\}\\ \wedge \inf\{t : |x_L^\mu(t) - \bar{x}_L| > \varepsilon\}.\end{aligned}\tag{6.68}$$

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

$$\frac{dx_i^\mu}{dt} \geq (f_{iL} - C\varepsilon)x_i^\mu + \mu\left(\frac{b_{i-1}}{2}x_{i-1}^\mu + \frac{b_{i+1}}{2}x_{i+1}^\mu - b_i x_i^\mu\right)\tag{6.69}$$

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

$$\tilde{x}_i^\mu := \frac{\log[x_i^\mu(t\cdot\log(\frac{1}{\mu}))]}{\log(\frac{1}{\mu})}.\tag{6.70}$$

We thus have

$$\frac{d\tilde{x}_i^\mu}{dt} \geq f_{iL} - C\varepsilon - \mu + \frac{b_{i-1}}{2}\mu^{1+\tilde{x}_i^\mu - \tilde{x}_{i-1}^\mu} + \frac{b_{i+1}}{2}\mu^{1+\tilde{x}_i^\mu - \tilde{x}_{i+1}^\mu}\tag{6.71}$$

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

$$\begin{aligned}\tilde{x}^\mu(0) = \biggl(&\frac{\log\varepsilon}{\log(1/\mu)}, -1, -2, \ldots,\\ &-L/2, -L/2+1, \ldots, -1, \frac{\log(\bar{x}_L - \varepsilon)}{\log(1/\mu)}\biggr)\end{aligned}\tag{6.72}$$

up to $o_\varepsilon(1)$ 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

$$\begin{aligned}\tilde{x}^\mu(0) = \biggl(&\frac{\log\varepsilon}{\log(1/\mu)}, -1, -2, \ldots,\\ &-\lfloor L/2\rfloor, -\lfloor L/2\rfloor, \ldots, -1, \frac{\log(\bar{x}_L - \varepsilon)}{\log(1/\mu)}\biggr),\end{aligned}\tag{6.73}$$

but the proof idea stays the same. Let $\delta > 0$ and $T_i^{\text{in}}(\delta,\mu) := \inf\{t > 0 : \tilde{x}_i^\mu > \tilde{x}_{i-1}^\mu - (1-\delta) \text{ or } \tilde{x}_i^\mu > \tilde{x}_{i+1}^\mu - (1-\delta)\}$. Then, for $t \in [0, T_i^{\text{in}}]$, that is, when $\tilde{x}_i^\mu$ is above one of its neighbours minus $1-\delta$, then, for μ and δ small enough, the slope of $\tilde{x}_i^\mu$ is prescribed by the fitness of trait i with respect to trait L (up to a multiple of ε). Indeed, by (6.71),

$$\begin{aligned} d\tilde{x}_i^\mu/dt &\geq f_{iL} - C\varepsilon - \mu \geq f_{iL} - 2C\varepsilon, \\ d\tilde{x}_i^\mu/dt &\leq f_{iL} + C\varepsilon - \mu + \frac{b_{i-1}}{2}\mu^\delta + \frac{b_{i+1}}{2}\mu^\delta \leq f_{iL} + 2C\varepsilon. \end{aligned} \tag{6.74}$$

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

Assume by contradiction that $\inf\{T_i^{\text{out}}, i \in \{0,\dots,L\}\} < \infty$.

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

$$1 + \tilde{x}_i^\mu - \tilde{x}_{i-1}^\mu = -\delta \quad \text{and} \quad 1 + \tilde{x}_i^\mu - \tilde{x}_{i+1}^\mu > -\delta, \tag{6.75}$$

the two other possibilities (right neighbour or both) are similar. By (6.71), for μ small enough,

$$\frac{d\tilde{x}_i^\mu}{dt}(T_i^{\text{out}}) \geq f_{iL} - C\varepsilon - \mu + \min\left\{\frac{b_{i-1}}{2}, \frac{b_{i+1}}{2}\right\}\mu^{-\delta} = O(\mu^{-\delta}), \tag{6.76}$$

$$\begin{aligned} \frac{d\tilde{x}_{i-1}^\mu}{dt}(T_i^{\text{out}}) &= f_{i-1,L} - C\varepsilon - \mu + \frac{b_{i-2}}{2}\mu^{1+\tilde{x}_{i-1}^\mu - \tilde{x}_{i-2}^\mu} + \frac{b_i}{2}\mu^{2+\delta} \\ &\leq \frac{b_{i-2}}{2}o(\mu^{-\delta}) + \frac{b_i}{2}\mu^{2+\delta} = o(\mu^{-\delta}), \end{aligned} \tag{6.77}$$

where the upper bounds in (6.77) come from the assumption that $\tilde{x}_i^\mu$ is the highest population among these exiting their security region. Indeed, if in (6.77) we had $1 + \tilde{x}_{i-1}^\mu - \tilde{x}_{i-2}^\mu = -\delta$ then, by definition, $\tilde{x}_{i-1}^\mu$ would exit its security region, thus we would have $\tilde{x}_{i-1}^\mu \leq \tilde{x}_i^\mu$, which contradicts (6.75). Thus, $1 + \tilde{x}_{i-1}^\mu - \tilde{x}_{i-2}^\mu > -\delta$. The equations (6.76) and (6.77) imply that the derivative $\frac{d}{dt}(\tilde{x}_i^\mu - \tilde{x}_{i-1}^\mu)(T_i^{\text{out}})$ is as large as needed. Thus, again by the continuity of the solutions, this implies the existence of some $t' < t$ such that $\tilde{x}_i^\mu(t') < \tilde{x}_{i-1}^\mu(t') - (1+\delta)$. Hence, $t' < T_i^{\text{out}}$, which is a contradiction. This implies that $T_i^{\text{out}} = \infty$, for all $i \in \{0,\dots,L\}$.

This allows us to describe the limit of $\tilde{x}^\mu$ as $\mu \to 0$. A helpful example is given in Figure 3. First, as $f_{LL} = 0$, equation (6.74) and the previous reasoning imply that until $\tau^+(\varepsilon, \mu)$,

$$\frac{d\tilde{x}_L^\mu}{dt} \geq -C\varepsilon \quad \text{and} \quad \frac{d\tilde{x}_L^\mu}{dt} \leq +C\varepsilon, \tag{6.78}$$

which implies that $\tilde{x}_L^\mu \to 0$ (take the limits $\varepsilon \to 0$ after $\mu \to 0$). Now the initial condition (6.72) and Assumption 1 imply that $\tilde{x}_i^\mu \to -(L-i)$, for $i = L/2, \dots, L-1$. Indeed, $\tilde{x}_L^\mu$ is close to 0, and $\tilde{x}_i^\mu(0) = -(L-i)$ 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 $\tilde{x}_i^\mu$ behaves piecewise linearly in the limit $\mu \to 0$ and given the sequence $(f_{0L}, \dots, f_{L-1,L})$ one can construct the successive slopes by following the rule "$\tilde{x}_i^\mu$ tries to decay with slope f_{iL} while being at distance at most 1 of $\tilde{x}_{i-1}^\mu$ and $\tilde{x}_{i+1}^\mu$; if it is not possible, then it stays parallel to the largest of its neighbours, either $\tilde{x}_{i-1}^\mu$ or $\tilde{x}_{i+1}^\mu$".

More precisely, consider the sequence $\{i_1, \dots, i_r\}$ 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 $(\tilde{x}^\mu(t))_{t>0}$, starting with initial condition (6.72), stays in an ε-neighbourhood of the deterministic process $x(t)$ given by

$$x_i(t) = -(L-i) \vee \max_{k \in [\![0,i]\!]} \{-i - |f_{kL}|t\} \vee \max_{k \in [\![1,r]\!]} \{-i_k - |i - i_k| - |f_{i_kL}|t\}. \tag{6.79}$$

Once again, Figure 3 provides a helpful example to compute the formula.

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 slowly varying function (recall (3.15)).

7.1. *Poisson representation.*

In the vein of Fournier and Méléard [28], we represent the population process in terms of Poisson measures. Let $(Q_k^{(b)}, Q_k^{(m)}, Q_k^{(d)}, 0 \leq k \leq L)$ be independent Poisson random measures on $\mathbb{R}_+^2$ with intensity $ds\, d\theta$, 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

$$d_i^K(x) = D_i^K(x)x_i = \left(d_i + \sum_{j=0}^{L} \frac{c_{ij}}{K} x_j\right) x_i, \tag{7.1}$$

for the total death rate of the subpopulation 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,

$$
\begin{aligned}
h(X(t)) = h(X(0)) &+ \sum_{k=0}^{L} \int_0^t \int_{\mathbb{R}_+} \big(h(X(s^-)+e_k) - h(X(s^-))\big) \\
&\times \mathbf{1}_{\theta \le (1-\mu) b_k X_k(s^-)} Q_k^{(b)}(ds, d\theta) \\
&+ \sum_{k=1}^{L} \int_0^t \int_{\mathbb{R}_+} \big(h(X(s^-)+e_k) - h(X(s^-))\big) \\
&\times \mathbf{1}_{\theta \le \mu b_{k-1} X_{k-1}(s^-)} Q_k^{(m)}(ds, d\theta) \\
&+ \sum_{k=0}^{L} \int_0^t \int_{\mathbb{R}_+} \big(h(X(s^-)-e_k) - h(X(s^-))\big) \\
&\times \mathbf{1}_{\theta \le D_k^K(X(s^-)) X_k(s^-)} Q_k^{(d)}(ds, d\theta).
\end{aligned}
\tag{7.2}
$$

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

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

and the stopping times T_ε^K and S_ε^K, which denote respectively the hitting time of $\lfloor \varepsilon K \rfloor$ by the total mutant population $(X_1 + \cdots + X_L)$ and the exit time of I_ε^K by the resident 0-population,

$$
\begin{aligned}
T_\varepsilon^K &:= \inf\Big\{t \ge 0, \sum_{1 \le i \le L} X_i(t) = \lfloor \varepsilon K \rfloor\Big\}, \\
S_\varepsilon^K &:= \inf\{t \ge 0, X_0(t) \notin I_\varepsilon^K\}.
\end{aligned}
\tag{7.4}
$$

As shown in [14], we know that as long as the total mutant population size is smaller than εK, 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, $c_{i0}\bar{x}_0$. This allows us to couple i-population sizes ($1 \le i \le L-1$) 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 εK, whereas the L-mutant population size exceeds the size εK, the 0 and L populations 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 is close to $\log K/|f_{0L}|$ (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 reach the size $\lfloor \varepsilon K \rfloor$. 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 is of order $\log K$, and we 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 excursion 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 is 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 depends strongly on the mutation probability per reproductive event, μ.

To study the case when $K\mu^L \gg 1$, we 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 $t \leq T_\varepsilon^K \wedge S_\varepsilon^K$,

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

By definition of the population process in (7.2) and of the stopping times T_ε^K and S_ε^K in (7.4), the following processes satisfy (7.5):

$$X_0^{(\pm)}(t) = K\left(\bar{x}_0 \pm 2\varepsilon \frac{\sup_{1\leq i\leq L} c_{0i}}{c_{00}}\right) =: x_0^{(\pm)} K, \tag{7.6}$$

and, for $1 \leq i \leq L-1$ and $* \in \{-, +\}$,

$$\begin{aligned} X_i^{(*)}(t) = &\int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq (1-\mu) b_i X_i^{(*)}(s^-)} Q_k^{(b)}(ds, d\theta) \\ &+ \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq \mu b_{i-1} X_{i-1}^{(*)}(s^-)} Q_k^{(m)}(ds, d\theta) \\ &- \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \leq (d_i + c_{i0} x_0^{(\bar{*})} + \mathbf{1}_{\{*=-\}} \varepsilon \sup_{1\leq j\leq L} c_{ij}) X_i^{(*)}(s^-)} Q_k^{(d)}(ds, d\theta), \end{aligned} \tag{7.7}$$

where $\bar{*} = \{-,+\} \setminus *$ and we used the same Poisson measures as in (7.2). Note that from this representation, we get directly the classical semimartingale decomposition for $X_i^{(-)}$ and $X_i^{(+)}$: for $* \in \{-,+\}$,

$$X_i^{(*)}(t) = M_i^{(*)}(t) + A_i^{(*)}(t), \tag{7.8}$$

where $M_i^{(*)}$ is a square integrable martingale and $A_i^{(*)}$ is a finite variation process, namely

$$\begin{aligned} M_i^{(-)}(t) = {} & \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le (1-\mu) b_i X_i^{(-)}(s^-)} \bigl(Q_k^{(b)}(ds, d\theta) - ds\, d\theta\bigr) \\ & + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le \mu b_{i-1} X_{i-1}^{(-)}(s^-)} \bigl(Q_k^{(m)}(ds, d\theta) - ds\, d\theta\bigr) \\ & - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le (d_i + c_{i0} x_0^{(+)} + \varepsilon \sup_{1 \le j \le L} c_{ij}) X_i^{(-)}(s^-)} \\ & \times \bigl(Q_k^{(d)}(ds, d\theta) - ds\, d\theta\bigr), \end{aligned} \tag{7.9}$$

$$\begin{aligned} A_i^{(-)}(t) = {} & \mu b_{i-1} \int_0^t X_{i-1}^{(-)}(s)\, ds \\ & + \Bigl((1-\mu) b_i - d_i - c_{i0} x_0^{(+)} - \varepsilon \sup_{1 \le j \le L} c_{ij}\Bigr) \int_0^t X_i^{(-)}(s)\, ds, \end{aligned} \tag{7.10}$$

and the same expression for $M_i^{(+)}$ and $A_i^{(+)}$ by replacing the (−)'s by (+)'s and the terms

$$d_i + c_{i0} x_0^{(+)} + \varepsilon \sup_{1 \le j \le L} c_{ij} \tag{7.11}$$

by

$$d_i + c_{i0} x_0^{(-)}. \tag{7.12}$$

Finally, we recall the expression of the quadratic variation of $M_i^{(-)}$,

$$\begin{aligned} \langle M_i^{(-)} \rangle_t = {} & \mu b_{i-1} \int_0^t X_{i-1}^{(-)}(s)\, ds \\ & + \Bigl((1-\mu) b_i + d_i + c_{i0} x_0^{(+)} + \varepsilon \sup_{1 \le j \le L} c_{ij}\Bigr) \int_0^t X_i^{(-)}(s)\, ds, \end{aligned} \tag{7.13}$$

and the one of $M_i^{(+)}$ is obtained by similar modifications as before.

Let us now introduce, for $1 \le k \le L-1$, the following notation:

$$\begin{aligned} -s_{k0}^{(+)} &:= (1-\mu) b_k - d_k - c_{k0} x_0^{(-)} \quad \text{and} \\ -s_{k0}^{(-)} &:= (1-\mu) b_k - d_k - c_{k0} x_0^{(+)} - \varepsilon \sup_{1 \le j \le L} c_{kj}, \end{aligned} \tag{7.14}$$

as well as for $* \in \{-,+\}$:

$$x_k^{(*)} := (1 * \varepsilon)^k \frac{b_0 \cdots b_{k-1} x_0^{(*)} \mu^k}{s_{10}^{(*)} \cdots s_{k0}^{(*)}} \quad \text{and} \quad t_\varepsilon^{(k)} := \frac{|\ln \varepsilon|}{s_{k0}^{(-)}}. \tag{7.15}$$

Notice that $s_{k0}^{(+)} \le s_{k0}^{(-)}$ and that $s_{k0}^{(+)}$ and $s_{k0}^{(-)}$ are positive, for ε small enough, by Assumption (A2).

LEMMA 7.1. *For every* $0 \le i \le L-1$,

$$x_i^{(-)} K \le \mathbb{E}[X_i^{(-)}(s)], \quad s \ge t_\varepsilon^{(1)} + \cdots + t_\varepsilon^{(i)}$$

and

$$\mathbb{E}[X_i^{(+)}(s)] \le x_i^{(+)} K, \quad s \ge 0.$$

PROOF. We prove this lemma by induction. The property is true for $i = 0$. Recall (7.7), (7.14) and (7.15). Then we get, for $1 \le i \le L-1$,

$$\frac{d}{dt}\mathbb{E}[X_i^{(*)}(t)] = b_{i-1}\mu\mathbb{E}[X_{i-1}^{(*)}(t)] - s_{i0}^{(*)}\mathbb{E}[X_i^{(*)}(t)], \quad \mathbb{E}[X_i^{(*)}(0)] = 0.$$

By the induction hypothesis, this yields, for every $t \ge 0$,

$$\frac{d}{dt}\mathbb{E}[X_i^{(+)}(t)] \le \mu b_{i-1} x_{i-1}^{(+)} K - s_{i0}^{(+)}\mathbb{E}[X_i^{(+)}(t)], \quad \mathbb{E}[X_i^{(*)}(0)] = 0$$

and, for every $t \ge t_\varepsilon^{(1)} + \cdots + t_\varepsilon^{(i)}$,

$$\frac{d}{dt}\mathbb{E}[X_i^{(-)}(t)] \ge \mu b_{i-1} x_{i-1}^{(-)} K - s_{i0}^{(-)}\mathbb{E}[X_i^{(-)}(t)], \quad \mathbb{E}[X_i^{(*)}(0)] = 0,$$

which completes the proof. □

LEMMA 7.2. *Let* $0 \le i < L$ *such that* $\lim_{K\to\infty} K\mu^i = \infty$, *and introduce the two counting processes*:

$$R_i^{(\pm)}(t) := \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\theta \le \mu b_i X_i^{(\pm)}(s^-)} Q_{i+1}^{(m)}(ds, d\theta) \tag{7.16}$$

and

$$\bar{R}_i^{(\pm)}(t) := \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\theta \le \mu b_i \mathbb{E}[X_i^{(\pm)}(s^-)]} Q_{i+1}^{(m)}(ds, d\theta), \tag{7.17}$$

where we use the same Poisson point measure as in (7.2). *Then* $M_i^{(\pm)} := R_i^{(\pm)} - \bar{R}_i^{(\pm)}$ *is a martingale and*

$$\mathbb{E}[(M_i^{(\pm)}(t))^2] \le 2\mu b_i x_i^{(\pm)} K t.$$

PROOF. We have

$$\begin{aligned}M_i^{(\pm)}(t) &= \int_0^t \int_{\mathbb{R}^+} (\mathbf{1}_{\theta \le \mu b_i X_i^{(\pm)}(s-)} - \mathbf{1}_{\theta \le \mu b_i \mathbb{E}[X_i^{(\pm)}(s-)]}) Q_{i+1}^{(m)}(ds, d\theta) \\ &= \int_0^t \int_{\mathbb{R}^+} (\mathbf{1}_{\theta \le \mu b_i X_i^{(\pm)}(s-)} - \mathbf{1}_{\theta \le \mu b_i \mathbb{E}[X_i^{(\pm)}(s-)]}) \tilde{Q}_{i+1}^{(m)}(ds, d\theta) \\ &\quad + \mu b_i \int_0^t (X_i^{(\pm)}(s) - \mathbb{E}[X_i^{(\pm)}(s)])\, ds.\end{aligned}$$

Hence, $M_i^{(\pm)}$ is a martingale. We can compute its quadratic variation via

$$\begin{aligned}\langle M_i^{(\pm)} \rangle_t &= \int_0^t \int_{\mathbb{R}^+} (\mathbf{1}_{\theta \le \mu b_i X_i^{(\pm)}(s)} - \mathbf{1}_{\theta \le \mu b_i \mathbb{E}[X_i^{(\pm)}(s)]})^2 \, ds \, d\theta \\ &= \mu b_i \int_0^t (X_i^{(\pm)}(s) + \mathbb{E}[X_i^{(\pm)}(s)] - 2(X_i^{(\pm)}(s) \wedge \mathbb{E}[X_i^{(\pm)}(s)]))\, ds \\ &= \mu b_i \int_0^t |X_i^{(\pm)}(s) - \mathbb{E}[X_i^{(\pm)}(s)]|\, ds\end{aligned}$$

As a consequence,

$$\begin{aligned}\mathbb{E}[(M_i^{(\pm)}(t))^2] = \mathbb{E}[\langle M_i^{(\pm)} \rangle_t] &\le \mu b_i \int_0^t \mathbb{E}[X_i^{(\pm)}(s) + \mathbb{E}[X_i^{(\pm)}(s)]]\, ds \\ &= 2\mu b_i \int_0^t \mathbb{E}[X_i^{(\pm)}(s)]\, ds,\end{aligned}$$

and we complete the proof applying Lemma 7.1. □

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 ε small enough (not scaling with K). The proof of the asymptotics

$$\liminf_{K \to \infty} \mathbb{P}\left((1 - c\varepsilon) \frac{1}{\alpha} \frac{L}{f_{L0}} < \frac{T_\varepsilon^{(K,L)}}{\log K} < (1 + c\varepsilon) \frac{1}{\alpha} \frac{L}{f_{L0}} \right) \ge 1 - c\varepsilon$$

follows this of Lemma 6.1 in [46]. To end the proof of Theorem 3.2, two more steps are needed. The first one is the study of the swap between 0 and L-populations, which leads to the first statement (3.20) of Theorem 3.2, and the second one is the study of the extinction phase of the unfit mutants, which leads to the second and third statements (3.21) and (3.22) of Theorem 3.2.

First, we need to show that once the L-population size has reached the value εK, the rescaled populations X_0^K and X_L^K 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

$$x_L(0) = \varepsilon \quad \text{and} \quad |x_0(0) - \bar{x}_0| \le 2\varepsilon \frac{\sup_{1 \le i \le L} c_{0i}}{c_{00}}. \tag{7.18}$$

This stays true until a time when X_L^K is close to its monomorphic equilibrium size $\bar{x}_L$ and X_0^K is smaller than ε^2. During this time interval, the i-population sizes, for $1 \le i \le L - 1$, do not evolve a lot. More precisely, there exists a constant ε_0 such that, for $\varepsilon \le \varepsilon_0$ and $1 \le i \le L - 1$, with a probability close to one, $\mu^{i+\varepsilon} \le X_i^K(t) \le \mu^{i-\varepsilon}$, where t describes an interval with a duration of order 1, which is the time needed for the rescaled population sizes (X_0^K, X_L^K) 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 A such that $-A\mathbb{E}[X_i^K(t)] \le d\mathbb{E}[X_i^K(t)]/dt \le A\mathbb{E}[X_i^K(t)]$ (for rigorous arguments, see the proof of Lemma 10 in [5]). This leads to the following rigorous statement: there exist a positive ε_0 and a function $f : x \mapsto f(x) \in (0, x^2)$ such that, for $\varepsilon \le \varepsilon_0$, there exist a stopping time U_ε^K and an event $\mathcal{E}$ such that

$$\frac{U_\varepsilon^K f_{L0}}{\log K} \underset{K\to\infty}{\to} 1 \quad \text{in probability}, \qquad \mathbb{P}(\mathcal{E}) \ge 1 - \varepsilon, \tag{7.19}$$

and almost surely on $\mathcal{E}$,

$$\begin{aligned} & f(\varepsilon) < X_0^K(U_\varepsilon^K) < \varepsilon^2, \qquad |X_L^K(U_\varepsilon^K) - \bar{x}_L| \le \varepsilon, \\ & K\mu^\varepsilon < X_i^K(U_\varepsilon^K)/\mu^i < K\mu^{-\varepsilon}, \quad 1 \le i \le L - 1. \end{aligned} \tag{7.20}$$

This proves part (3.20) of Theorem 3.2.

Second, we need to approximate the time for the i-populations $(0 \le i \le L - 1)$ to get extinct after the time U_ε^K. Let us define two stopping times:

$$V_\varepsilon^K := \inf\{t \ge U_\varepsilon^K, |X_L^K(t) - \bar{x}_L| > 2\varepsilon\}$$

and

$$W_\varepsilon^K := \inf\left\{t \ge U_\varepsilon^K, \sum_{0 \le i \le L-1} X_i^K(t) > \varepsilon\right\}.$$

We will prove the following property: there exist $\varepsilon_0, C, V > 0$ such that, for $\varepsilon \le \varepsilon_0$,

$$\liminf_{K\to\infty} \mathbb{P}(\mathrm{e}^{KV} < (W_\varepsilon^K \wedge V_\varepsilon^K)) \ge 1 - o_\varepsilon(1), \tag{7.21}$$

where $o_\varepsilon(1)$ is a function of ε which goes to 0 as ε goes to 0. This allows us to couple the i-population sizes $(0 \le i \le L - 1)$ with subcritical birth and death processes with inhomogeneous immigration in order to approximate their time to extinction.

To prove (7.21), 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 counter-selected, and whose initial size is smaller than $O(\varepsilon^2 K)$; second, the L-population size. Let us show that with a probability converging to 1 as $K \to \infty$, $W_\varepsilon^K < V_\varepsilon^K$. To this aim, notice that on the time interval $[0, V_\varepsilon^K]$, the death rate of the i-population ($0 \leq i \leq L-1$) satisfies

$$d_i + \sum_{j=0}^{L} \frac{c_{ij}}{K} X_j \geq d_i + c_{iL}(\bar{x}_L - 2\varepsilon). \tag{7.22}$$

Moreover, we know that almost surely on the event $\mathcal{E}$, we have $X_0(U_\varepsilon^K) \leq \varepsilon^2 K$. Hence, if we introduce, for $0 \leq i \leq L-1$ and $k \in \mathbb{N}$, the notation

$$T_k^{(X_i)} := \inf\{t \geq U_\varepsilon^K, X_i(t) = k\}, \tag{7.23}$$

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

$$\begin{aligned}\mathbb{P}(T_{\varepsilon K}^{(X_0)} < V_\varepsilon^K | \mathcal{E}) &\leq \frac{((d_0 + c_{0L}(\bar{x}_L - 2\varepsilon))/b_0)^{\varepsilon^2 K} - 1}{((d_0 + c_{0L}(\bar{x}_L - 2\varepsilon))/b_0)^{\varepsilon K} - 1} \\ &\leq \left(\frac{b_0}{d_0 + c_{0L}(\bar{x}_L - 2\varepsilon)}\right)^{\varepsilon K(1-\varepsilon)} \leq C\varepsilon,\end{aligned} \tag{7.24}$$

for any constant C, K large enough and ε small enough.

Let us denote by $\mathcal{M}_{01}$ the number of type 1 mutants produced by type 0-individuals during the time interval $[U_\varepsilon^K, V_\varepsilon^K]$. From (7.2), we have

$$\mathcal{M}_{01} = \int_{U_\varepsilon^K}^{V_\varepsilon^K \wedge T_0^{(X_0)}} \mathbf{1}_{\theta \leq \mu b_0 X_0(s^-)} Q_0^{(m)}(ds, d\theta). \tag{7.25}$$

Moreover, considering all the possible orderings of $T_{\varepsilon K}^{(X_0)}$, V_ε^K, $T_0^{(X_0)}$ and $\ln K/\sqrt{\varepsilon}$, we get

$$\begin{aligned}\mathbb{P}\biggl(\{T_{\varepsilon K}^{(X_0)} < V_\varepsilon^K \wedge T_0^{(X_0)}\} \cup \left\{V_\varepsilon^K \wedge T_0^{(X_0)} < T_{\varepsilon K}^{(X_0)} \wedge \frac{\ln K}{\sqrt{\varepsilon}}\right\} \\ \cup \left\{\frac{\ln K}{\sqrt{\varepsilon}} < V_\varepsilon^K \wedge T_0^{(X_0)} < T_{\varepsilon K}^{(X_0)}\right\}\biggr) = 1.\end{aligned}$$

Hence, using the Markov inequality, (7.24), 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

$$\begin{aligned}
&\mathbb{P}(\mathcal{M}_{01} > K\mu \ln K | \mathcal{E}) \\
&\quad \leq \frac{(\varepsilon K)(b_0\mu)(\ln K/\sqrt{\varepsilon})}{K\mu \ln K} + \mathbb{P}\big(T^{(X_0)}_{\varepsilon K} < V^K_\varepsilon | \mathcal{E}\big) \\
&\qquad + \mathbb{P}\left(\frac{\ln K}{\sqrt{\varepsilon}} < V^K_\varepsilon \wedge T^{(X_0)}_0 < T^{(X_0)}_{\varepsilon K} \middle| \mathcal{E}\right) \\
&\quad = o_\varepsilon(1).
\end{aligned} \tag{7.26}$$

Applying again (A.5), we find that each mutant of type 1 that is produced by a type 0 individual generates a type 1 population whose size has a probability to reach $\varepsilon/\mu \ln K$ that is bounded by

$$\left(\frac{b_1}{d_1 + c_{1L}(\bar{x}_L - 2\varepsilon)}\right)^{\varepsilon/\mu \ln K - 1}.$$

We deduce that

$$\begin{aligned}
&\mathbb{P}\big(T^{(X_1)}_{\varepsilon K} < V^K_\varepsilon | \mathcal{E}\big) \\
&\quad \leq \mathbb{P}(\mathcal{M}_{01} > K\mu \ln K | \mathcal{E}) + K\mu \ln K \left(\frac{b_1}{d_1 + c_{1L}(\bar{x}_L - 2\varepsilon)}\right)^{\varepsilon/\mu \ln K - 1} = o_\varepsilon(1).
\end{aligned}$$

We reiterate the reasoning for the other counter-selected mutant populations (i-populations with $2 \leq i \leq L-1$) to conclude

$$\mathbb{P}\big(W^K_\varepsilon < V^K_\varepsilon | \mathcal{E}\big) = o_\varepsilon(1). \tag{7.27}$$

By a direct application of Lemma A.1, we get the existence of a positive constant V such that

$$\liminf_{K\to\infty} \mathbb{P}\big(\mathrm{e}^{KV} < V^K_\varepsilon | V^K_\varepsilon \leq W^K_\varepsilon, \mathcal{E}\big) = 1. \tag{7.28}$$

Using (7.27) and (7.28), we get

$$\begin{aligned}
&\liminf_{K\to\infty} \mathbb{P}\big(\mathrm{e}^{KV} < (W^K_\varepsilon \wedge V^K_\varepsilon) | \mathcal{E}\big) \\
&\quad \geq \liminf_{K\to\infty} \mathbb{P}\big(\mathrm{e}^{KV} < V^K_\varepsilon \leq W^K_\varepsilon | \mathcal{E}\big) \\
&\quad = \liminf_{K\to\infty} \mathbb{P}\big(\mathrm{e}^{KV} < V^K_\varepsilon | V^K_\varepsilon \leq W^K_\varepsilon, \mathcal{E}\big)\mathbb{P}\big(V^K_\varepsilon \leq W^K_\varepsilon | \mathcal{E}\big) = 1 - o_\varepsilon(1).
\end{aligned} \tag{7.29}$$

This proves (7.21), and thus statement (3.22) of Theorem 3.2, as we recall that $\mathbb{P}(\mathcal{E}) \geq 1 - \varepsilon$.

We may now approximate the growth rates of the i-population sizes ($0 \leq i \leq L-1$) during the time interval $[U^K_\varepsilon, V^K_\varepsilon \wedge W^K_\varepsilon]$. For $0 \leq i \leq L-1$, let us introduce, for $* \in \{-,+\}$,

$$-\sigma_i^{(*)} := b_i(1-\mu) - d_i - c_{iL}(\bar{x}_L \bar{*} 2\varepsilon) - \mathbf{1}_{\{*=-\}} \sup_{0\leq k\leq L-1} c_{ik}\varepsilon, \tag{7.30}$$

where $\bar{*} = \{-,+\} \setminus *$. Notice that, for ε small enough the $(\sigma_i^{(*)})_{0 \le i \le L-1}$ are pairwise distinct by the fourth point of Assumption 1. We consider such an ε throughout the remainder of the proof to make sure that we do not divide by 0. Notice also that equation (7.30) ensures that there exists $C > 0$ such that, for ε small enough,

$$0 < |f_{iL}| - C\varepsilon < \sigma_i^{(+)} < |f_{iL}| < \sigma_i^{(-)} < |f_{iL}| + C\varepsilon.$$

From the definition of the process X in (7.2) and from (7.20), we get that almost surely on the event $\mathcal{E}$ and for $0 \le i \le L-1$,

$$P_i^{(-)}(t) \le X_i(U_\varepsilon^K + t) \le P_i^{(+)}(t) \quad \forall\, U_\varepsilon^K \le U_\varepsilon^K + t \le V_\varepsilon^K \wedge W_\varepsilon^K \tag{7.31}$$

where, for $t \ge 0$ and $* \in \{-,+\}$,

$$\begin{aligned} P_i^{(*)}(t) = X_i(U_\varepsilon^K) &+ \int_{U_\varepsilon^K}^{U_\varepsilon^K + t} \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le (1-\mu) b_i P_i^{(*)}(s^-)} Q_i^{(b)}(ds, d\theta) \\ &+ \int_{U_\varepsilon^K}^{U_\varepsilon^K + t} \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le \mu b_{i-1} P_{i-1}^{(*)}(s^-)} Q_i^{(m)}(ds, d\theta) \\ &- \int_{U_\varepsilon^K}^{U_\varepsilon^K + t} \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le ((1-\mu) b_i + \sigma_i^{(*)}) P_i^{(*)}(s^-)} Q_i^{(d)}(ds, d\theta), \end{aligned} \tag{7.32}$$

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

$$\beta_L := \left\{ k \in [\![0, L-1]\!] \text{ such that } \frac{|f_{kL}|}{1 - k/\alpha} = \inf_{0 \le j \le L-1} \left\{ \frac{|f_{jL}|}{1 - j/\alpha} \right\} \right\}. \tag{7.33}$$

We will see that the β_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 (7.20), we know that on the event $\mathcal{E}$ the size at time U_ε^K of the β_L-population is

$$C(\varepsilon, K) K \mu^{\beta_L} \quad \text{with } \mu^\varepsilon \le C(\varepsilon, K) \le \mu^{-\varepsilon}.$$

From (7.31) and (7.32), we see that almost surely on $\mathcal{E}$ and on the time interval $[U_\varepsilon^K, V_\varepsilon^K \wedge W_\varepsilon^K]$, the β_L-population size is larger than a subcritical birth and death process with initial state $C(\varepsilon, K) K \mu^{\beta_L}$, individual birth rate $b_{\beta_L}(1-\mu)$, and individual death rate $b_{\beta_L}(1-\mu) + \sigma_{\beta_L}^{(-)}$. Applying equation (A.3), we deduce that

$$\begin{aligned} \liminf_{K \to \infty} \mathbb{P}\biggl(\inf\{t \ge 0, X_{\beta_L}(U_\varepsilon^K + t) = 0\} \ge \left(1 - \frac{\beta_L}{\alpha}\right) \frac{(1-\varepsilon)}{\sigma_{\beta_L}^{(-)}} \ln K \Big| \mathcal{E} \biggr) \\ \ge 1 - \varepsilon. \end{aligned} \tag{7.34}$$

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 takes a time of order at most

$(1 + l\varepsilon)(1 - \beta_L/\alpha)\ln K/\sigma_{\beta_L}^{+}$, for some positive l (to be made precise later, see (7.39)) to get extinct. To simplify notation, let us introduce, for $0 \le i \le L - 1$ and $* \in \{-, +\}$,

$$\mathfrak{f}_i^{(*)} := \inf\{\sigma_j^{(*)}, 0 \le j \le i\}. \tag{7.35}$$

We will see that the mutant population whose size decreases the slowest provides the leading term and scale the time needed for all but the L populations to get extinct. To prove that, we now show by induction that there exists $\varepsilon_0 > 0$ and a sequence of positive functions, $(g_0 : x \mapsto x^2, g_1, \dots, g_{L-1})$, such that, for every $0 \le i \le L - 1$, $\varepsilon \le \varepsilon_0$ and $t \ge 0$,

$$\mathbb{E}[P_i^{(+)}(t)] \le g_i(\varepsilon)K\mu^{i-\varepsilon}\mathrm{e}^{-\mathfrak{f}_i^{(+)}t.} \tag{7.36}$$

For $i = 0$, from definitions (7.30), (7.32) and property (7.20), we get

$$\mathbb{E}[P_0^{(+)}(t)] \le \varepsilon^2 K\mathrm{e}^{-\sigma_0^{(+)}t} = \varepsilon^2 K\mathrm{e}^{-\mathfrak{f}_0^{(+)}t}.$$

Let us assume that (7.36) holds for every i such that $0 \le i \le i_0 < L-1$. Then from (7.30), (7.32) and the induction hypothesis, for $t \ge 0$,

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[P_{i_0+1}^{(+)}(t)] &\le -\sigma_{i_0+1}^{(+)}\mathbb{E}[P_{i_0+1}^{(+)}(t)] + \mu b_{i_0}\mathbb{E}[P_{i_0}^{(+)}(t)]\\ &\le -\sigma_{i_0+1}^{(+)}\mathbb{E}[P_{i_0+1}^{(+)}(t)] + \mu b_{i_0}g_{i_0}(\varepsilon)K\mu^{i_0-\varepsilon}\mathrm{e}^{-\mathfrak{f}_{i_0}^{(+)}t}\\ &= -\sigma_{i_0+1}^{(+)}\mathbb{E}[P_{i_0+1}^{(+)}(t)] + b_{i_0}g_{i_0}(\varepsilon)K\mu^{i_0+1-\varepsilon}\mathrm{e}^{-\mathfrak{f}_{i_0}^{(+)}t}.\end{aligned} \tag{7.37}$$

Applying the method of variation of parameters, we get, for every $t \ge 0$,

$$\begin{aligned}\mathbb{E}[P_{i_0+1}^{(+)}(t)] &\le \mathbb{E}[P_{i_0+1}^{(+)}(0)]\mathrm{e}^{-\sigma_{i_0+1}^{(+)}t}\\ &\quad + \frac{b_{i_0}g_{i_0}(\varepsilon)K\mu^{i_0+1-\varepsilon}}{\sigma_{i_0+1}^{(+)} - \mathfrak{f}_{i_0}^{(+)}}(\mathrm{e}^{-\mathcal{S}_{i_0}^{(+)}t} - \mathrm{e}^{-\sigma_{i_0+1}^{(+)}t})\\ &\le K\mu^{i_0+1-\varepsilon}\Big(\mathrm{e}^{-\sigma_{i_0+1}^{(+)}t} + \frac{b_{i_0}g_{i_0}(\varepsilon)}{\sigma_{i_0+1}^{(+)} - \mathfrak{f}_{i_0}^{(+)}}(\mathrm{e}^{-\mathfrak{f}_{i_0}^{(+)}t} - \mathrm{e}^{-\sigma_{i_0+1}^{(+)}t})\Big),\end{aligned} \tag{7.38}$$

where the last inequality is a consequence of (7.20). Hence, the $i_0 + 1$-population satisfies (7.36), with $\mathfrak{f}_{i_0+1}^{(+)} = \mathfrak{f}_{i_0}^{(+)} \wedge \sigma_{i_0+1}^{(+)}$, according to the definition (7.35), and

$$g_{i_0+1}(\varepsilon) = 1 + \frac{2b_{i_0}g_{i_0}(\varepsilon)}{|\sigma_{i_0+1}^{(+)} - \mathfrak{f}_{i_0}^{(+)}|}.$$

Moreover, let us introduce $l > 0$ such that, for ε small enough and for $0 \le i \le L - 1$,

$$\frac{1 - i/\alpha + \varepsilon/\alpha}{\mathfrak{f}_i^{(+)}} < (1 + l\varepsilon)\frac{1 - \beta_L/\alpha}{\sigma_{\beta_L}^{(+)}}, \tag{7.39}$$

(which is possible according to the definitions (7.30) and (7.33)) and define

$$s_K := \frac{(1+l\varepsilon)}{\sigma_{\beta_L}^{(+)}} \ln\bigl(K\mu^{\beta_L}\bigr).$$

Then, applying (7.31), (7.36) and the Markov inequality, we get

$$\begin{aligned}
&\mathbb{P}(\exists\, i \in [\![0, L-1]\!], X_i(s_K) \geq 1 | \mathcal{E}) \\
&\quad\leq \sum_{0\leq i\leq L-1} \mathbb{P}\bigl(X_i(s_K) \geq 1 | \mathcal{E}\bigr) \leq \sum_{0\leq i\leq L-1} \mathbb{P}\bigl(P_i^{(+)}(s_K) \geq 1 | \mathcal{E}\bigr) \\
&\quad\leq \sum_{0\leq i\leq L-1} \mathbb{E}\bigl(P_i^{(+)}(s_K) | \mathcal{E}\bigr) \leq \sum_{0\leq i\leq L-1} g_i(\varepsilon) K\mu^{i-\varepsilon} \mathrm{e}^{-\mathfrak{f}_i^{(+)} s_K} \\
&\quad= \sum_{0\leq i\leq L-1} g_i(\varepsilon) \exp\biggl[\mathfrak{f}_i^{(+)}\biggl(\frac{1}{\mathfrak{f}_i^{(+)}} \ln\bigl(K\mu^{i-\varepsilon}\bigr) - \frac{(1+l\varepsilon)}{\sigma_{\beta_L}^{(+)}} \ln\bigl(K\mu^{\beta_L}\bigr)\biggr)\biggr] \qquad (7.40)\\
&\quad= \sum_{0\leq i\leq L-1} g_i(\varepsilon) \exp\biggl[\mathfrak{f}_i^{(+)}\biggl(\frac{1 - i/\alpha + \varepsilon/\alpha}{\mathfrak{f}_i^{(+)}} - (1+l\varepsilon)\frac{1-\beta_L/\alpha}{\sigma_{\beta_L}^{(+)}}\biggr) \\
&\qquad \times \ln K\bigl(1+o(1)\bigr)\biggr],
\end{aligned}$$

where we used (3.15) in the last line. According to the definition of l, the last term goes to 0 as K goes to infinity.

Combining (7.19), (7.21), (7.34) and (A.1) proves statement (3.21) of Theorem 3.2, and thus completes the proof of this theorem. □

7.3. *Proof of Theorem* 3.3. Assume first that (3.15) holds and that $\alpha \notin \mathbb{N}$. Theorem 3.3 addresses the case where $K\mu^L$ is small. Only the $\lfloor\alpha\rfloor$ first mutant populations has a large size as

$$\begin{aligned}
K\mu^{\lfloor\alpha\rfloor} &= f^{\lfloor\alpha\rfloor}(K) K^{1-\lfloor\alpha\rfloor/\alpha} \to \infty, \quad K \to \infty, \\
K\mu^{\lfloor\alpha\rfloor+1} &= f^{\lfloor\alpha\rfloor+1}(K) K^{1-(\lfloor\alpha\rfloor+1)/\alpha} \to 0, \quad K \to \infty.
\end{aligned}$$

For $\lfloor\alpha\rfloor + 1 \leq i \leq L-1$, the i-mutant population sizes perform excursions until a successful L-individual is created. By successful L-individual, we mean a mutant L which generates a population outcompeting the other populations. Here again, the key tools are couplings with birth and death processes without competition.

Let us denote by $T^{(i)}$ (see definition in (7.41)) the birth time of the ith mutant of type $(\lfloor\alpha\rfloor + 1)$ descended from an individual of type $\lfloor\alpha\rfloor$ and by $X_0^{(i)}$ the type $(\lfloor\alpha\rfloor + 1)$-population generated by this individual. Then we use the lexicographic order to number the k-mutant populations, with $\lfloor\alpha\rfloor + 2 \leq k \leq L$ (see Figure 4 for an illustration). More precisely:

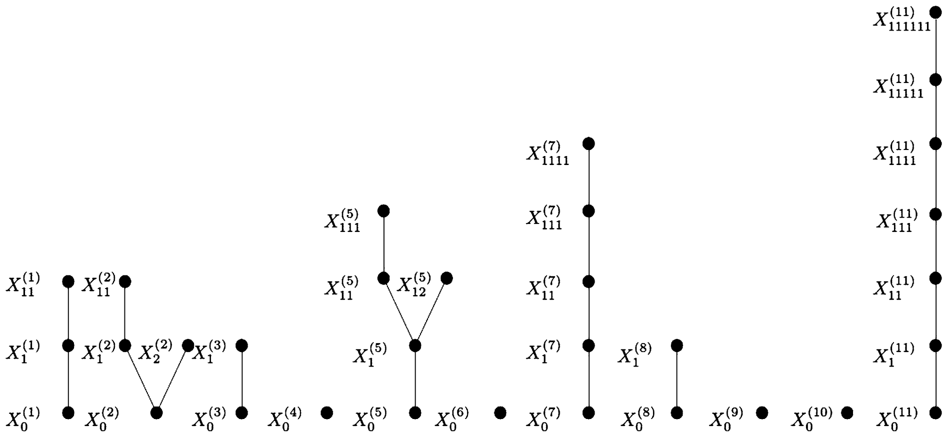

FIG. 4. *Numbering of the successive subpopulations: in this example, the first* $(\lfloor\alpha\rfloor+1)$*-mutant generates the* $(\lfloor\alpha\rfloor+1)$*-population* $X_0^{(1)}$. *One individual of the population* $X_0^{(1)}$ *gives birth to an* $(\lfloor\alpha\rfloor+2)$*-mutant, which generates the* $(\lfloor\alpha\rfloor+2)$*-population* $X_1^{(1)}$. *One individual of the population* $X_1^{(1)}$ *gives birth to an* $(\lfloor\alpha\rfloor+3)$*-mutant, which generates the* $(\lfloor\alpha\rfloor+3)$*-population* $X_{11}^{(1)}$. *The population* $X_{11}^{(1)}$ *gets extinct without giving birth to any* $(\lfloor\alpha\rfloor+4)$*-individual. The second* $(\lfloor\alpha\rfloor+1)$*-mutant produced by an* $\lfloor\alpha\rfloor$*-individual generates the* $(\lfloor\alpha\rfloor+1)$*-population* $X_0^{(2)}$. *Two individuals of the population* $X_0^{(2)}$ *give birth to an* $(\lfloor\alpha\rfloor+2)$*-mutant. These mutants generate the* $(\lfloor\alpha\rfloor+2)$*-populations* $X_1^{(2)}$ *and* $X_2^{(2)}$, *respectively. And so on with the notation previously introduced.*

- For $j\geq 1$, $X_j^{(i)}$ is the $(\lfloor\alpha\rfloor+2)$-population generated by the jth $(\lfloor\alpha\rfloor+2)$-mutant produced by an individual of type $(\lfloor\alpha\rfloor+1)$ belonging to the population $X_0^{(i)}$.
- For $j,k\geq 1$, $X_{jk}^{(i)}$ is the $(\lfloor\alpha\rfloor+3)$-population generated by the kth $(\lfloor\alpha\rfloor+3)$-mutant produced by an individual of type $(\lfloor\alpha\rfloor+2)$ belonging to the population $X_j^{(i)}\ldots.$

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 μ, and more than one $(i+1)$-mutant with a probability of order μ^2. The law of all trees can be approximated by the law of a subcritical Galton–Watson process, and trees are approximately independent. Hence, we will be able to approximate the probability for the $X_0^{(i)}$ 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 is close to an exponential random variable with mean one divided by this probability.

Recall the definition of the process $R_{\lfloor\alpha\rfloor}$ in (7.16). Then the stopping time $T^{(i)}$ which is the birth time of the ith $(\lfloor\alpha\rfloor+1)$-mutant produced by an $\lfloor\alpha\rfloor$-individual can be expressed as

$$T^{(i)}:=\inf\{t\geq 0,\ R_{\lfloor\alpha\rfloor}(t)\geq i\}. \tag{7.41}$$

In particular, from (7.5), we get, for every $t \le T_\varepsilon^K \wedge S_\varepsilon^K$,

$$R^{(-)}_{\lfloor\alpha\rfloor}(t) \le R_{\lfloor\alpha\rfloor}(t) \le R^{(+)}_{\lfloor\alpha\rfloor}(t) \quad \text{a.s.}, \tag{7.42}$$

where processes $R^{(\pm)}_{\lfloor\alpha\rfloor}$ have been defined in (7.16). Let us first, for the sake of simplicity, replace the processes $R^{(\pm)}_{\lfloor\alpha\rfloor}$ by the processes $\bar{R}^{(\pm)}_{\lfloor\alpha\rfloor}$, defined in (7.17), and introduce

$$T^{(i,\pm)} := \inf\{t \ge 0, \bar{R}^{(\pm)}_{\lfloor\alpha\rfloor}(t) \ge i\}.$$

We will prove later on that this does not modify the result.

Let u_K be a sequence such that

$$u_K \gg \ln K \quad \text{and} \quad \mu^{\lfloor\alpha\rfloor+1} K u_K \underset{K\to\infty}{\to} 0.$$

Using the Markov inequality and Lemma 7.1, we get

$$\begin{aligned}\mathbb{P}\bigl(T^{(1,-)} \le u_K\bigr) &\le \mathbb{P}\bigl(R^{(-)}(u_K) \ge 1\bigr) \le \mu b_{\lfloor\alpha\rfloor} u_K \bar{x}^{(+)}_{\lfloor\alpha\rfloor} \\ &= (1+\varepsilon)^{\lfloor\alpha\rfloor} \frac{b_0 \cdots b_{\lfloor\alpha\rfloor} x_0^{(+)}}{s_{10}^{(+)} \cdots s_{\lfloor\alpha\rfloor 0}^{(+)}} \mu^{\lfloor\alpha\rfloor+1} K u_K \to 0, \quad (K \to \infty).\end{aligned}$$

Following the ideas developed in Section 7.2, we may couple each k-mutant population ($\lfloor\alpha\rfloor + 1 \le k \le L-1$) 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 notation are tedious, but we nevertheless state rigorously the resulting properties. Let us denote by $(T_{\mathfrak{j}}^{(i)}, \mathfrak{j} \in \bigcup_{n\in\mathbb{N}} \mathbb{N}^n, i \in \mathbb{N})$ the time of appearance of the populations $(X_{\mathfrak{j}}^{(i)}, \mathfrak{j} \in \bigcup_{n\in\mathbb{N}} \mathbb{N}^n, i \in \mathbb{N})$. For all $\mathfrak{j} \in \bigcup_{n\in\mathbb{N}} \mathbb{N}^n$, $i \in \mathbb{N}$, $T_{\mathfrak{j}}^{(i)} := \inf\{t \ge 0, X_{\mathfrak{j}}^{(i)}(t) \ge 1\}$. Then we introduce birth and death processes $(X_{\mathfrak{j}}^{(i,*)}, \mathfrak{j} \in \bigcup_{n\in\mathbb{N}} \mathbb{N}^n, i \in \mathbb{N}, * \in \{-,+\})$ with birth and death rates

$$\Bigl(\bigl(b_{\mathfrak{t}(\mathfrak{j})}, (1-\mu) b_{\mathfrak{t}(\mathfrak{j})} + \sigma^{(*)}_{\mathfrak{t}(\mathfrak{j})}\bigr), \mathfrak{j} \in \bigcup_{n\in\mathbb{N}} \mathbb{N}^n, i \in \mathbb{N}, * \in \{-,+\}\Bigr),$$

where the $\sigma^{(*)}$'s have been defined in (7.30),

$$\mathfrak{t}(\mathfrak{j}) := \lfloor\alpha\rfloor + |\mathfrak{j}| + 1,$$

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

These processes can be constructed in such a way that, for all $\mathfrak{j} \in \bigcup_{n\in\mathbb{N}} \mathbb{N}^n$, $i \in \mathbb{N}$,

$$X_{\mathfrak{j}}^{(i,-)}(t) \le X_{\mathfrak{j}}^{(i)}(t) + \mathfrak{N}_{\mathfrak{j}}^{(i)}(t) \le X_{\mathfrak{j}}^{(i,+)}(t), \quad t \le T_\varepsilon^K \wedge S_\varepsilon^K, \tag{7.43}$$

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

Moreover, conditionally on $(T_{\mathrm{j}}^{(i)}, \mathrm{j}\in\bigcup_{n\in\mathbb{N}}\mathbb{N}^n, i\in\mathbb{N})$, the pairs of processes $((X_{\mathrm{j}}^{(i,-)}, X_{\mathrm{j}}^{(i,+)}), \mathrm{j}\in\bigcup_{n\in\mathbb{N}}\mathbb{N}^n, i\in\mathbb{N})$ are independent. This last point allows us to approximate the law of the random trees

$$\mathcal{T}^{(i)} := X_0^{(i)} \bigcup_{n\in\mathbb{N}} X_{\mathbb{N}^n}^{(i)}, \quad i\in\mathbb{N}$$

(an example is depicted in Figure 4) by the same law, and independently. Indeed from equation (7.43), we get that $(\mathcal{T}^{(i)})_{i\in\mathbb{N}}$, 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 $\mathcal{T}^{(i)}$ ($i\in\mathbb{N}$) with root $X_0^{(i)}$ independently, and approximate its probability to end with a L-mutant subpopulation. For simplicity, we write $|0|=0$.

Consider a vertex $X_{\mathrm{j}}^{(i)}$, $\mathrm{j}\in\{0\}\cup\mathbb{N}^{\mathbb{N}}$ of the tree $\mathcal{T}^{(i)}$. Due to the coupling (7.43), we are able to approximate the probability for this vertex to have no child, one child or more than one child. Before the time $T_\varepsilon^K \wedge S_\varepsilon^K$, the law of the number of offsprings produced by the $X_{\mathrm{j}}^{(i)}$ population is smaller (resp., larger) than the number of offspring produced by a population initiated by one individual, with individual birth rate $b_{\mathrm{t(j)}}$ and individual death rate $(1-\mu)b_{\mathrm{t(j)}}+\sigma_{\mathrm{t(j)}}^{(+)}$ (resp., $(1-\mu)b_{\mathrm{t(j)}}+\sigma_{\mathrm{t(j)}}^{(-)}$). Moreover, each offspring is a mutant of type $(\mathrm{t(j)}+1)$ with probability μ, and is a clone with probability $(1-\mu)$. Hence,

$$\begin{aligned}\mathbb{P}\big(X_{\mathrm{j}}^{(i)}\text{-pop produces 1 mutant}\big) &\le \sum_{n=0}^{\infty} p^{(b_{\mathrm{t(j)}},(1-\mu)b_{\mathrm{t(j)}}+\sigma_{\mathrm{t(j)}}^{(+)})}(n)n\mu(1-\mu)^{n-1}\\ &\le \mu \mathrm{e}^{(b_{\mathrm{t(j)}},b_{\mathrm{t(j)}}+\sigma_{\mathrm{t(j)}}^{(+)})},\end{aligned}$$

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

$$\begin{aligned}&\mathbb{P}\big(X_{\mathrm{j}}^{(i)}\text{-pop produces 1 mutant}\big)\\ &\quad\ge \sum_{n=0}^{\infty} p^{(b_{\mathrm{t(j)}},(1-\mu)b_{\mathrm{t(j)}}+\sigma_{\mathrm{t(j)}}^{(-)})}(n)n\mu(1-\mu)^{n-1}\\ &\quad\ge \mu \sum_{n=0}^{\mu^{-1/2}+1} p^{(b_{\mathrm{t(j)}},(1-\mu)b_{\mathrm{t(j)}}+\sigma_{\mathrm{t(j)}}^{(-)})}(n)n(1-\mu)^{\mu^{-1/2}} \qquad (7.44)\\ &\quad\ge \mu \sum_{n=0}^{\mu^{-1/2}+1} p^{(b_{\mathrm{t(j)}},(1-\mu)b_{\mathrm{t(j)}}+\sigma_{\mathrm{t(j)}}^{(-)})}(n)n(1-2\sqrt{\mu})\end{aligned}$$

$$= \mu(1 - 2\sqrt{\mu})\big(e^{(b_{\mathrm{t(j)}},(1-\mu)b_{\mathrm{t(j)}}+\sigma^{(-)}_{\mathrm{t(j)}})} - \mathbb{E}[\mathbf{1}_{\{|X^{(i)}_{\mathrm{j}}| \geq \mu^{-1/2}+1\}}|X^{(i)}_{\mathrm{j}}|]\big),$$

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

$$\begin{aligned}\mathbb{E}^2[\mathbf{1}_{\{|X^{(i)}_{\mathrm{j}}| \geq \mu^{-1/2}+1\}}|X^{(i)}_{\mathrm{j}}|] &\leq \mathbb{P}(|X^{(i)}_{\mathrm{j}}| \geq \mu^{-1/2}+1)\mathbb{E}[|X^{(i)}_{\mathrm{j}}|^2] \\ &\leq \sqrt{\mu}\mathbb{E}[|X^{(i)}_{\mathrm{j}}|]\mathbb{E}[|X^{(i)}_{\mathrm{j}}|^2] = O(\sqrt{\mu}),\end{aligned}$$

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

Adding (A.8), we may conclude that, as K goes to infinity,

$$\mathbb{P}(X^{(i)}_{\mathrm{j}}\text{-pop produces 1 mutant}) = e^{(b_{\mathrm{t(j)}},b_{\mathrm{t(j)}}+|f_{\mathrm{t(j)}0}|)}\mu\big(1+O(\varepsilon)\big).$$

Using again coupling (7.43) and (A.6), we get that

$$\begin{aligned}&\mathbb{P}(X^{(i)}_{\mathrm{j}}\text{-pop produces at least 2 mutants}) \\ &\quad\leq \sum_{n=0}^{\infty} p^{(b_{\mathrm{t(j)}},b_{\mathrm{t(j)}}+\sigma^{(+)}_{\mathrm{t(j)}})}(n)\frac{n(n-1)}{2}\mu^2 = O(\mu^2).\end{aligned}$$

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

$$\mu^{L-1-\lfloor\alpha\rfloor}\left(\prod_{k=\lfloor\alpha\rfloor+1}^{L-1} e^{(b_k,b_k+|f_{k0}|)}\right)\big(1+O(\varepsilon)\big). \tag{7.45}$$

Indeed, the probability for each vertex to produce one child is of order μ, and the probability to produce at least two children is of order μ^2. Since there is only a finite number of possible mutations, independent of μ, this implies that the probability for the tree $\mathcal{T}^{(i)}$ to have at least one vertex with two children and end with a L individual is of order $\mu^{L-\lfloor\alpha\rfloor}$, which is negligible compared to $\mu^{L-1-\lfloor\alpha\rfloor}$. Moreover, we know that each L-mutant has a probability close to f_{L0}/b_L to generate a population whose size hits the value εK, 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 L-individual to be born. If the times of appearance of the trees $\mathcal{T}^{(i)}$ had the law of a Poisson process with inhomogeneous parameter close to $\mu b_{\lfloor\alpha\rfloor}x_{\lfloor\alpha\rfloor}K$ (that is to say if we could consider $\bar{R}^{(\pm)}_{\lfloor\alpha\rfloor}$ instead of $R^{(\pm)}_{\lfloor\alpha\rfloor}$), this would complete the proof of the first point of Theorem 3.3. We now need to justify that the result stays true when considering $R^{(\pm)}_{\lfloor\alpha\rfloor}$. To achieve this goal, it is enough to prove the existence of two sequences $N_1(K)$ and $N_2(K)$ satisfying

$$N_1(K) \gg (K\mu^L)^{-1} \quad \text{and} \quad N_2(K) \ll (\mu^{L-1-\lfloor\alpha\rfloor})^{-1} \tag{7.46}$$

such that

$$\lim_{K\to\infty}\mathbb{P}\Big(\sup_{s\le N_1(K)}\big|R^{(\pm)}_{\lfloor\alpha\rfloor}(s)-\bar{R}^{(\pm)}_{\lfloor\alpha\rfloor}(s)\big|>N_2(K)\Big)=0. \tag{7.47}$$

Indeed, this implies that during the time interval under consideration (of order $(K\mu^L)^{-1}$), the difference between the number of 'trees' generated by the processes $R^{(\pm)}_{\lfloor\alpha\rfloor}$ and $\bar{R}^{(\pm)}_{\lfloor\alpha\rfloor}$ is much smaller than $(\mu^{L-1-\lfloor\alpha\rfloor})^{-1}$, and as each tree has a probability of order $(\mu^{L-1-\lfloor\alpha\rfloor})$ to generate a successful mutant, the same tree is at the origin of the successful mutant under the two counting processes under consideration with a probability close to 1.

To prove (7.47), we apply Doob's martingale inequality to $M_{\lfloor\alpha\rfloor}$. This yields

$$\begin{aligned}&\mathbb{P}\Big(\sup_{s\le N_1(K)}\big|R^{(\pm)}_{\lfloor\alpha\rfloor}(s)-\bar{R}^{(\pm)}_{\lfloor\alpha\rfloor}(s)\big|>N_2(K)\Big)\\&\quad\le\frac{\mathbb{E}[(M^{(\pm)}_{\lfloor\alpha\rfloor}(N_1(K)))^2]}{N_2^2(K)}\le\frac{2\mu b_{\lfloor\alpha\rfloor}\bar{x}^{(\pm)}_{\lfloor\alpha\rfloor}N_1(K)}{N_2^2(K)}\\&\quad\le CK\mu^{\lfloor\alpha\rfloor+1}\frac{N_1(K)}{N_2^2(K)}=C\frac{K\mu^L}{(\mu^{L-1-\lfloor\alpha\rfloor})^2}\frac{N_1(K)}{N_2^2(K)}\mu^{L-1-\lfloor\alpha\rfloor},\end{aligned}$$

where C is a finite constant. As $\mu^{L-1-\lfloor\alpha\rfloor}$ goes to 0 when K tends to ∞, the sequences $N_1(K)$ and $N_2(K)$ can be chosen in such a way that the last term in the previous series of inequalities goes to 0 when K tends to ∞, which completes the proof of (7.47).

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 ε 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 εK from its equilibrium size $\bar{n}_0K$ during a time larger than e^{KV}. As a consequence, if $K\mu\gg e^{-VK}$, the 0-population produces a large number of 1-mutants during the time interval $[0,e^{VK}]$, with a rate very close to $b_0\bar{n}_0K\mu$. Hence, the proof is very similar to the previous proof, where the $\lfloor\alpha\rfloor$-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:

$$\begin{aligned}X_0(t)=\lfloor\bar{x}_0K\rfloor&+\int_0^t\int_{\mathbb{R}_+}\mathbf{1}_{\theta\le(1-\mu)b_0X_0(s^-)}Q_0^{(b)}(ds,d\theta)\\&-\int_0^t\int_{\mathbb{R}_+}\mathbf{1}_{\theta\le D_0^K(X(s^-))X_0(s^-)}Q_0^{(d)}(ds,d\theta),\end{aligned}\tag{8.1}$$

where $D_0^K(X)$ is defined in (7.1). Thus, if we introduce the process Y_0 via

$$\begin{aligned} Y_0(t) = \lfloor \bar{x}_0 K \rfloor &+ \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le b_0 Y_0(s^-)} Q_0^{(b)}(ds, d\theta) \\ &- \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\theta \le (d_0 + c_{00} Y_0(s^-)/K) Y_0(s^-)} Q_0^{(d)}(ds, d\theta), \end{aligned}$$

we get that, almost surely, $X_0(t) \le Y_0(t)$, for all $t \ge 0$. Now consider a time v_K such that

$$\frac{1}{\rho_0(K)} \ll v_K \ll \frac{1}{K\mu^L}, \quad K \to \infty,$$

where $\rho_0(K)$ was defined in (3.28). If we apply inequality (3.7) of [18] to the process Y_0, we get

$$d_{\mathrm{TV}}\big(\mathbb{P}\big(Y_0(v_K) \in \cdot\big), \delta_0(\cdot)\big) \underset{K\to\infty}{\to} 0,$$

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

$$\mathbb{P}\big(X_0(v_K) > 0\big) \underset{K\to\infty}{\to} 0. \tag{8.2}$$

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

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[X_0(t)] &\le \mathbb{E}\Big[\Big(b_0 - d_0 - \frac{c_{00}}{K} X_0(t)\Big) X_0(t)\Big] \\ &\le (b_0 - d_0)\mathbb{E}[X_0(t)] - \frac{c_{00}}{K}\mathbb{E}^2[X_0(t)]. \end{aligned}$$

Thus, for all $t \ge 0$, we have $\mathbb{E}[X_0(t)] \le \bar{x}_0 K$. Next we bound the expectation of the total number Ξ_1 of type 1 individuals generated by type 0 individuals by mutations before the time v_K:

$$\mathbb{E}[\Xi_1] \le \int_0^{v_K} b_0 \mu \mathbb{E}[X_0(s)]\, ds \le b_0 \bar{x}_0 K \mu v_K. \tag{8.3}$$

We want to bound the probability that at least one type 1 individual born from a type 0 individual before time v_K has a line of descent containing a type L individual. Denote by ξ_i the event that the ith 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

$$\mathbb{P}(B_L < v_K) = \mathbb{P}\Big(\bigcup_{i \le \Xi_1} \xi_i\Big) = \mathbb{E}\Big[\mathbb{P}\Big(\bigcup_{i \le \Xi_1} \xi_i \Big| \Xi_1\Big)\Big] \le \mathbb{E}\Big[\sum_{i \le \Xi_1} \mathbb{P}(\xi_i | \Xi_1)\Big].$$

But recall that by Assumption 2, for $1 \le i \le L-1$, $b_i < d_i$. Hence, using (7.45), we see that the probability of the events $(\xi_i)_{1 \le i \le \Xi_1}$ can be bounded independently of Ξ_1 by

$$2\Big(\prod_{1\le i\le L-1} \mathrm{e}^{(b_i,d_i)}\Big)\mu^{L-1}.$$

This yields

$$\mathbb{P}(B_L < v_K) \le b_0 \bar{x}_0 v_K \Big(\prod_{1\le i\le L-1} \mathrm{e}^{(b_i,d_i)}\Big)\mu^L \underset{K\to\infty}{\to} 0.$$

Adding (8.2) completes the proof.

8.2. *Proof of Theorem* 3.4 *point* 1. We introduce v_K such that $\frac{1}{\rho_0(K)} \ll v_K \ll \frac{1}{K\mu}$. 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 completes the proof.

APPENDIX. 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 $(W_t)_{t\ge 0}$ 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*

$$\mathcal{S}_K = \inf\Big\{t\ge 0 : W_t \notin \Big[\Big(\frac{b-d}{c}-\eta_1\Big)K, \Big(\frac{b-d}{c}+\eta_2\Big)K\Big]\Big\}.$$

Then there exists $V>0$ such that, for any compact subset C of $](b-d)/c-\eta_1, (b-d)/c+\eta_2[$,

$$\lim_{K\to\infty} \sup_{k/K\in C} \mathbb{P}_k\big(\mathcal{S}_K < \mathrm{e}^{KV}\big) = 0. \tag{A.1}$$

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 $Z=(Z_t)_{t\ge 0}$ be a birth and death process with individual birth and death rates b and d. For $i\in\mathbb{Z}_+$, $T_i=\inf\{t\ge 0, Z_t=i\}$ 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* $t \geq 0$,

$$\mathbb{P}_i(T_0 \leq t) = \left(\frac{d(1-\mathrm{e}^{(d-b)t})}{b - d\mathrm{e}^{(d-b)t}}\right)^i. \tag{A.2}$$

- *If* $0 < b < d$ *and* $Z_0 = N$, *the following convergence holds*:

$$T_0/\log N \underset{N\to\infty}{\to} (d-b)^{-1} \quad \textit{in probability}. \tag{A.3}$$

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

$$T_N/\log N \underset{N\to\infty}{\to} (b-d)^{-1} \quad \textit{a.s.} \tag{A.4}$$

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

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

The last result of this Appendix concerns the size distribution of the total number of individuals in a subcritical 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* $k \geq 0$,

$$p^{(b,d)}(k) := \mathbb{P}(Z = k) = \frac{(2k)!}{k!(k+1)!}\left(\frac{b}{d+b}\right)^k\left(\frac{d}{d+b}\right)^{k+1}. \tag{A.6}$$

In particular,

$$\mathrm{e}^{(b,d)} := \mathbb{E}[Z] = \sum_{k=1}^{\infty} \frac{(2k)!}{(k-1)!(k+1)!}\left(\frac{b}{d+b}\right)^k\left(\frac{d}{d+b}\right)^{k+1}. \tag{A.7}$$

Moreover, there exist two positive constants c *and* ε_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*

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

REFERENCES

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

[2] ALEXANDER, H. K. (2013). Conditional distributions and waiting times in multitype branching processes. *Adv. in Appl. Probab.* **45** 692–718. MR3102468

[3] BAAR, M., BOVIER, A. and CHAMPAGNAT, N. (2017). From stochastic, individual-based models to the canonical equation of adaptive dynamics in one step. *Ann. Appl. Probab.* **27** 1093–1170. MR3655862

[4] BAAR, M., COQUILLE, L., MAYER, H., HÖLZEL, M., ROGAVA, M., TÜTING, T. and BOVIER, A. (2016). A stochastic model for immunotherapy of cancer. *Sci. Rep.* **6** 24169.

[5] BILLIARD, S. and SMADI, C. (2017). The interplay of two mutations in a population of varying size: A stochastic eco-evolutionary model for clonal interference. *Stochastic Process. Appl.* **127** 701–748. MR3605709

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

[7] BOLKER, B. M. and PACALA, S. W. (1999). Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. *Amer. Nat.* **153** 575–602.

[8] BOVIER, A., COQUILLE, L. and NEUKIRCH, R. (2018). The recovery of a recessive allele in a Mendelian diploid model. *J. Math. Biol.* **77** 971–1033. MR3856865

[9] BOVIER, A. and DEN HOLLANDER, F. (2015). *Metastability: A Potential-Theoretic Approach. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **351**. Springer, Cham. MR3445787

[10] BOVIER, A. and WANG, S.-D. (2013). Trait substitution trees on two time scales analysis. *Markov Process. Related Fields* **19** 607–642. MR3185194

[11] BRINK-SPALINK, R. and SMADI, C. (2017). Genealogies of two linked neutral loci after a selective sweep in a large population of stochastically varying size. *Adv. in Appl. Probab.* **49** 279–326. MR3631225

[12] BRITTON, T. and PARDOUX, E. (2017). Stochastic epidemics in a homogeneous community. In preparation.

[13] CARTER, A. J. and WAGNER, G. P. (2002). Evolution of functionally conserved enhancers can be accelerated in large populations: A population-genetic model. *Proc. R. Soc. Lond., B Biol. Sci.* **269** 953–960.

[14] CHAMPAGNAT, N. (2006). A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stochastic Process. Appl.* **116** 1127–1160. MR2250806

[15] CHAMPAGNAT, N., FERRIÈRE, R. and BEN AROUS, G. (2001). The canonical equation of adaptive dynamics: A mathematical view. *Selection* **2** 73–83.

[16] CHAMPAGNAT, N., FERRIÈRE, R. and MÉLÉARD, S. (2008). From individual stochastic processes to macroscopic models in adaptive evolution. *Stoch. Models* **24** 2–44. MR2466448

[17] CHAMPAGNAT, N. and MÉLÉARD, S. (2011). Polymorphic evolution sequence and evolutionary branching. *Probab. Theory Related Fields* **151** 45–94. MR2834712

[18] CHAZOTTES, J.-R., COLLET, P. and MÉLÉARD, S. (2016). Sharp asymptotics for the quasi-stationary distribution of birth-and-death processes. *Probab. Theory Related Fields* **164** 285–332. MR3449391

[19] CHAZOTTES, J.-R., COLLET, P. and MÉLÉARD, S. (2017). On time scales and quasi-stationary distributions for multitype birth-and-death processes. arXiv preprint arXiv:1702.05369.

[20] CORON, C., COSTA, M., LEMAN, H. and SMADI, C. (2018). A stochastic model for speciation by mating preferences. *J. Math. Biol.* **76** 1421–1463. MR3771426

[21] CORON, C., MÉLÉARD, S., PORCHER, E. and ROBERT, A. (2013). Quantifying the mutational meltdown in diploid populations. *Amer. Nat.* **181** 623–636.

[22] COWPERTHWAITE, M. C., BULL, J. J. and MEYERS, L. A. (2006). From bad to good: Fitness reversals and the ascent of deleterious mutations. *PLoS Comput. Biol.* **2** e141.

[23] DEPRISTO, M. A., HARTL, D. L. and WEINREICH, D. M. (2007). Mutational reversions during adaptive protein evolution. *Mol. Biol. Evol.* **24** 1608–1610.

[24] DIECKMANN, U. and LAW, R. (1996). The dynamical theory of coevolution: A derivation from stochastic ecological processes. *J. Math. Biol.* **34** 579–612. MR1393842

[25] DIECKMANN, U. and LAW, R. (2000). Moment approximations of individual-based models. In *The Geometry of Ecological Interactions*: *Simplifying Spatial Complexity* (U. Dieckmann, R. Law and J. A. J. Metz, eds.) 252–270. Cambridge University Press, Cambridge.

[26] DURRETT, R. and MAYBERRY, J. (2011). Traveling waves of selective sweeps. *Ann. Appl. Probab.* **21** 699–744. MR2807971

[27] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0838085

[28] FOURNIER, N. and MÉLÉARD, S. (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.* **14** 1880–1919. MR2099656

[29] GERITZ, S. A., METZ, J. A., KISDI, É. and MESZÉNA, G. (1997). Dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.* **78** 2024.

[30] GIACHETTI, C. and HOLLAND, J. J. (1988). Altered replicase specificity is responsible for resistance to defective interfering particle interference of an Sdi-mutant of vesicular stomatitis virus. *J. Virol.* **62** 3614–3621.

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

[32] GOKHALE, C. S., IWASA, Y., NOWAK, M. A. and TRAULSEN, A. (2009). The pace of evolution across fitness valleys. *J. Theoret. Biol.* **259** 613–620. MR2973179

[33] HAENO, H., MARUVKA, Y. E., IWASA, Y. and MICHOR, F. (2013). Stochastic tunneling of two mutations in a population of cancer cells. *PLoS ONE* **8** e65724.

[34] IWASA, Y., MICHOR, F. and NOWAK, M. A. (2004). Evolutionary dynamics of invasion and escape. *J. Theoret. Biol.* **226** 205–214. MR2069303

[35] IWASA, Y., MICHOR, F. and NOWAK, M. A. (2004). Stochastic tunnels in evolutionary dynamics. *Genetics* **166** 1571–1579.

[36] LEMAN, H. (2016). Convergence of an infinite dimensional stochastic process to a spatially structured trait substitution sequence. *Stoch. Partial Differ. Equ. Anal. Comput.* **4** 791–826. MR3554431

[37] LENSKI, R. E., OFRIA, C., PENNOCK, R. T. and ADAMI, C. (2003). The evolutionary origin of complex features. *Nature* **423** 139–144.

[38] MAISNIER-PATIN, S., BERG, O. G., LILJAS, L. and ANDERSSON, D. I. (2002). Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. *Mol. Microbiol.* **46** 355–366.

[39] METZ, J. A., GERITZ, S. A., MESZÉNA, G., JACOBS, F. J. and VAN HEERWAARDEN, J. S. (1995). Adaptive dynamics: A geometrical study of the consequences of nearly faithful reproduction. WP-95-099.

[40] MOORE, F. B.-G. and TONSOR, S. J. (1994). A simulation of Wright's shifting-balance process: Migration and the three phases. *Evolution* **48** 69–80.

[41] O'HARA, P. J., NICHOL, S. T., HORODYSKI, F. M. and HOLLAND, J. J. (1984). Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. *Cell* **36** 915–924.

[42] SAGITOV, S. and SERRA, M. C. (2009). Multitype Bienaymé–Galton–Watson processes escaping extinction. *Adv. in Appl. Probab.* **41** 225–246. MR2514952

[43] SCHRAG, S. J., PERROT, V. and LEVIN, B. R. (1997). Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. *Proc. R. Soc. Lond., B Biol. Sci.* **264** 1287–1291.

[44] SERRA, M. C. (2006). On the waiting time to escape. *J. Appl. Probab.* **43** 296–302. MR2225069

[45] SERRA, M. C. and HACCOU, P. (2007). Dynamics of escape mutants. *Theor. Popul. Biol.* **72** 167–178.

[46] SMADI, C. (2017). The effect of recurrent mutations on genetic diversity in a large population of varying size. *Acta Appl. Math.* **149** 11–51. MR3647031

[47] TRAN, V. C. (2008). Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM Probab. Stat.* **12** 345–386. MR2404035

[48] VAN DER HOFSTAD, R. (2016). *Random Graphs and Complex Networks*. Cambridge Univ. Press, Cambridge.

[49] WADE, M. J. and GOODNIGHT, C. J. (1991). Wright's shifting balance theory: An experimental study. *Science* **253** 1015–1018.

[50] WEINREICH, D. M. and CHAO, L. (2005). Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution* **59** 1175–1182.

[51] WEISSMAN, D. B., DESAI, M. M., FISHER, D. S. and FELDMAN, M. W. (2009). The rate at which asexual populations cross fitness valleys. *Theor. Popul. Biol.* **75** 286–300.

[52] WRIGHT, S. (1965). Factor interaction and linkage in evolution. *Proc. R. Soc. Lond., B Biol. Sci.* **162** 80–104.

A. BOVIER
INSTITUT FÜR ANGEWANDTE MATHEMATIK
RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN
ENDENICHER ALLEE 60
53115 BONN
GERMANY
E-MAIL: bovier@uni-bonn.de

L. COQUILLE
CNRS INSTITUT FOURIER
UNIVERSITÉ GRENOBLE ALPES
F-38000 GRENOBLE
FRANCE
E-MAIL: loren.coquille@univ-grenoble-alpes.fr

C. SMADI
UR LISC
LABORATOIRE D'INGÉNIERIE
POUR LES SYSTÈMES COMPLEXES
IRSTEA
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: charline.smadi@irstea.fr