Stochastic individual-based models with power law mutation rate on a general finite trait space

Loren Coquille, Anna Kraut, Charline Smadi

Stochastic individual-based models with power law mutation rate on a general finite trait space[*]

Loren Coquille[†]   Anna Kraut[‡]   Charline Smadi[§]

Abstract

We consider a stochastic individual-based model for the evolution of a haploid, asexually reproducing population. The space of possible traits is given by the vertices of a (possibly directed) finite graph $\mathcal{G} = (V, E)$. The evolution of the population is driven by births, deaths, competition, and mutations along the edges of $\mathcal{G}$. We are interested in the large population limit under a mutation rate μ_K given by a negative power of the carrying capacity K of the system: $\mu_K = K^{-1/\alpha}, \alpha > 0$. This results in several mutant traits being present at the same time and competing for invading the resident population. We describe the time evolution of the orders of magnitude of each sub-population on the $\log K$ time scale, as K tends to infinity. Using techniques developed in [8], we show that these are piecewise affine continuous functions, whose slopes are given by an algorithm describing the changes in the fitness landscape due to the succession of new resident or emergent types. This work generalises [25] to the stochastic setting, and Theorem 3.2 of [6] to any finite mutation graph. We illustrate our theorem by a series of examples describing surprising phenomena arising from the geometry of the graph and/or the rate of mutations.

Keywords: eco-evolution; adaptive dynamics; birth and death processes; finite graph; selective sweep; coupling; competitive Lotka-Volterra systems.
MSC2020 subject classifications: 92D25; 60J80; 92D15; 37N25.
Submitted to EJP on March 10, 2020, final version accepted on August 27, 2021.
Supersedes arXiv:2003.03452.

[†]Univ. Grenoble Alpes, CNRS, IF, 38000 Grenoble, France. E-mail: Loren.Coquille@univ-grenoble-alpes.fr
[‡]Rheinische Friedrich-Wilhelms-Universität Bonn, Germany. E-mail: kraut@iam.uni-bonn.de
[§]Univ. Grenoble Alpes, INRAE, LESSEM, F-38402 St-Martin-d'Hères, France and Univ. Grenoble Alpes, CNRS, IF, 38000 Grenoble, France. E-mail: charline.smadi@inrae.fr

1 Introduction

Adaptive dynamics is a biological theory that was developed to study the interplay between ecology and evolution. It involves the three mechanisms of heredity, mutations, and natural selection. It was first introduced in the 1990ies by Metz, Geritz, Bolker, Pacala, Dieckmann, Law, and coauthors [30, 18, 23, 4, 5, 17], who mostly considered a deterministic setting but also heuristically mentioned first stochastic versions. A paradigm of adaptive dynamics is the separation of the slow evolutionary and the fast ecological time scales, which is a result of reproduction with rare mutations. Invasion, fixation or extinction of a mutant population is determined by its invasion fitness, that describes the exponential growth rate of a single mutant in the current (coexisting) population(s) at equilibrium.

Stochastic individual-based models of adaptive dynamics have been rigorously constructed and first studied in the seminal work of Fournier and Méléard [22], and there is now a growing literature on these models. The population consists of a collection of individuals who reproduce, with or without mutation, or die after random exponential times depending on the current state of the whole population. The population size is controlled by a *carrying capacity K* which represents the amount of available resources. This class of models has first been studied in the original context of separation between evolutionary and ecological time scales. That is in the joint limit of large populations and rare mutations such that a mutant either dies out or fixates before the next mutation occurs. Mathematically this amounts to considering a probability of mutation satisfying in particular

$$\mu_K \ll 1/K \log K \quad \text{as } K \to \infty. \tag{1.1}$$

We will call this regime 'rare mutation regime' in the sequel. The description of the succession of mutant invasions, on the mutation time scale $1/K\mu_K$, in a monomorphic [7] or polymorphic [10, 2] asexual population gives rise respectively to the so-called *Trait Substitution Sequence* or *Polymorphic Evolution Sequence*. Extensions of the question to sexual populations were then studied, both in the haploid [39, 13] and the diploid [12, 33] cases.

It is natural to consider the effect of a higher mutation rates, where mutation events are no longer separated, if we want to describe several mutant traits being present microscopically at the same time and competing for invading the resident population. The mutation rate given by

$$\mu_K = K^{-\frac{1}{\alpha}}, \quad \text{for } \alpha > 0 \tag{1.2}$$

was considered in different contexts [20, 40, 6, 8] and will be the concern of the present paper. Notice that another mutation scale has been considered in [2, 3] to model the interaction of few mutants in the case without recurrent mutations, namely μ_K of order $1/K \log K$.

Another approach to adaptive dynamics has been introduced by Maynard Smith [29] under the name of *adaptive walks*. This was further developed by Kauffman and Levin [24] and many others, as mentioned below. Here, a given finite graph represents the possible types of individuals (vertices) together with their possibilities of mutation (edges). A fixed, but possibly random, fitness landscape assigns real numbers to the vertices of the graph. The evolution of the population is modelled as a random walk on the graph that moves towards higher fitnesses. This can be interpreted as the adaptation of the population to its environment. In contrast to the adaptive dynamics context, this fitness landscape is not dependent on the current state of the population. Adaptive walks move along edges towards neighbours of increasing fitness, according to some transition law, towards a local or global maximum. In particular, in such models it is not possible for a population to cross a fitness valley. This is partially solved by a variation of this

model, called *adaptive flight* [32]. It consists in a walk jumping between local fitness maxima, before eventually reaching a global maximum. The questions of the distribution of maxima [34], the typical length of a walk [35], or the typical accessibility properties of the fitness landscape [26, 38, 1] have been studied under different assumptions on the graph structure, the fitness law, or the transition law of the walk. Moreover, comparisons of these models with actual empirical fitness landscapes have been performed in [41]. As Kraut and Bovier showed [25], adaptive walks and flights arise as the limit of individual-based models of adaptive dynamics, when the large population followed by the rare mutations limit is taken. They also conjecture, and this will be proved in the present article, that similar results hold in the stochastic setting under the mutation rate (1.2), as we detail below.

In this paper, we consider an individual-based Markov process that models the evolution of a haploid, asexually reproducing population. The space of possible traits is given by the vertices of a (possibly directed) finite graph $\mathcal{G} = (V, E)$. The evolution of the population is driven by birth, death, and competition rates, which are fixed and depend on the traits, as well as mutations towards nearest neighbours in the graph $\mathcal{G}$. We start with a macroscopic initial condition (that is to say of order K, see Definition 2.2) and we are interested in the stochastic process given by the large population limit under the mutation rate (1.2). We describe the time evolution of the orders of magnitude of each sub-population on the $\log K$ time scale, as K tends to infinity. We show that the limiting process is deterministic, given by piecewise affine continuous functions, which are determined by an algorithm describing the changes in the fitness landscape due to the succession of new resident or emergent types.

This work constitutes an extension of the paper by Kraut and Bovier [25] to the stochastic setting. They consider the deterministic system resulting from the large population limit of the individual-based model ($K \to \infty$), and let the mutation probability μ tend to zero. By rescaling the time by $\log(1/\mu)$, they prove that the limiting process is a deterministic adaptive walk that jumps between different equilibria of coexisting traits. A corollary of our results gives the same behaviour, on the $\log K$ time scale, for the stochastic process under the scaling (1.2) for α larger than the diameter of the graph $\mathcal{G}$. Kraut and Bovier also study a variation of the model, where they modify the deterministic system such that the subpopulations can only reproduce when their size lies above a certain threshold μ^{α}. This limits the radius in which a resident population can foster mutants, and mimics the scaling (1.2) that we consider. The resulting limiting processes are adaptive flights (which are not restricted to jumping to nearest neighbours), and thus can cross valleys in the fitness landscape and reach a global fitness maximum. We obtain the same behaviour, on the $\log K$ time scale, for the stochastic process under the scaling (1.2) without any restriction on α.

The results of the present paper can also be seen as a generalisation of Theorem 3.2 in [6] by Bovier, Coquille and Smadi to any finite trait space. Indeed, they consider the graph with vertices $V = \{0, \ldots, L\}$ embedded in $\mathbb{N}$ and 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. Several speeds of the mutation rate are considered, and in particular, when $\alpha > L$, the exit time of the valley is computed on the $\log K$ time scale. This becomes a corollary of our results, and we can give an algorithmic description of the rescaled process for more general graphs endowed with a fitness valley, as we discuss in several examples in Section 3.

Our proof heavily relies on couplings of the original process with logistic birth and death processes with non-constant immigration, and the analysis of the latter simpler processes on the $\log K$ time scale. This approach was developed by Champagnat, Méléard, and Tran in [8]. They consider an individual-based model for the evolution

of a discrete population performing horizontal gene transfer and mutations on $V = [0,4] \cap \delta \mathbb{N}, \delta > 0$. Their goal is to analyze the trade-off between natural selection, which drives the population to higher birth-rates, and transfer, which drives the population to lower ones. Under the mutation rate (1.2), they exhibit parameter regimes where different evolutionary outcomes appear, in particular evolutionary suicide and emergence of a cyclic behaviour. As in the present paper, their results characterize the time evolution of the orders of magnitude of each sub-population on the $\log K$ time scale, which are shown to be piecewise affine continuous functions whose slopes are given by an algorithm describing the succession of phases when a given type is dominant or resident. Their proofs provide us with the main ingredients needed for our results. However, the graph structure they choose simplifies the inductions and we have to generalise their approach to treat the case of more general graphs, in the proof spirit of Kraut and Bovier [25].

Our results are general, and could be applied to have a better understanding of evolutionary trajectories in complex fitness landscapes. There are now more and more empirical studies of fitness landscapes (see [15] for a comprehensive review of data and tools up to 2014 for instance), and the probability and effect of specific mutations in given landscapes are better and better understood. For instance oriented mutation graphs can stem from mutation bias, through codon usage bias or similar molecular phenomena which make some mutations more probable than others [36].

We present a series of specific examples where surprising phenomena arise from the geometry of the graph $\mathcal{G}$ and/or the rate of mutations (1.2). Most of them could not happen under a different scaling of mutation rates.

- In Example 3.2, we describe a scenario where the ancestry of the resident population consists, with high probability, of back mutations towards a previously extinct trait, although the mutations that happen in between are not deleterious. In other words, the final resident individuals, say of trait v, although they can be produced from a wild type directly, come with high probability from a sequence of *non deleterious* mutations which went back to the wild type before mutating to v. This phenomenon can also happen in the rare mutation regime (1.1), that is for $\alpha \in (0,1)$, on the mutation time scale $(1/K\mu_K \gg \log K)$, where invading mutants fully replace the resident population before a new mutant arises. We show that it can still occur for higher mutation rates of the form (1.2), on a $\log K$ time scale, when parameters are chosen such that temporary extinction of the original trait is likely. Such mutational reversions have been observed (see [16] for instance).

- If evolution and mutation time scales are separated (i.e. in the rare mutation regime (1.1)), mutations occur one at a time, and the number of successive resident traits from the wild type to the type gathering k successively beneficial mutations is k. This is not the case if mutations are faster, in which case it is possible to observe either more or less successive resident traits. We will show this in Examples 3.3 and 3.4.

- In Example 3.5, we show that adding a new possible mutation path towards a fit trait can increase the time until it appears macroscopically. This is in the spirit of the paradox called *price of anarchy* in game theory or more specifically *Braess paradox* in the study of traffic networks congestion. Motter showed that this paradox may often occur in biological and ecological systems [31]. He studies the removal of part of a metabolic network to ensure its long term persistence, with applications to cancer, antibiotics and metabolic diseases. Another field of application is the food webs management, where selective removal of some species from the network can potentially have a positive outcome of preventing a series of further extinctions [37].

- Another counter-intuitive phenomenon arising from the mutation rate (1.2), presented in Example 3.6, is the possibility to observe, for a cyclic clockwise oriented mutation graph, successive counter-clockwise resident populations. This means that the macroscopic succession of resident traits is not necessarily representative of the mutation graph. In particular, this may call into question the interpretation in terms of mutation graphs of some experiments in experimental evolution (see [28] for instance).

- In Examples 3.7 and 3.8, we show that the mutation rate (1.2) does not restrict the range of the corresponding adaptive flights on the trait space (i.e. the distance that the limiting process can jump) to $\lfloor \alpha \rfloor$.

- We finally study the framework of fitness valley crossings. Combining our results with Theorem 3.3 of [6], we construct Examples 3.9 and 3.11, where effective random walks on the trait space appear on the time scale K^β, for some positive β. Those limiting adaptive flights arise as a result of a "fast" equilibration on the $\log K$ time scale followed by exponential waiting times until fitness valleys get crossed. This makes sense biologically, since there may be traits with positive invasion fitness that can be reached through several consecutive mutation steps [27, 14].

The remainder of this paper is organised as follows. In Section 2 we define the model and present our results. In Section 3 we illustrate our results by a series of examples describing surprising phenomena arising from the geometry and/or the rate of mutations. Section 4 is devoted to the proofs. In the Appendix, we present and extend some technical results.

2 Convergence on the $\log K$-time scale

2.1 Model

We consider an individual-based Markov process that models the evolution of a haploid, asexually reproducing population. The space of possible traits is given by the vertices of a (possibly directed) finite graph $\mathcal{G} = (V, E)$. Let $\mathcal{M}_p(V)$ be the set of probability measures on V.

For all traits $v, w \in V$ and every $K \in \mathbb{N}$, we introduce the following parameters:

- $b_v \in \mathbb{R}_+$, the birth rate of an individual of trait v,

- $d_v \in \mathbb{R}_+$, the (natural) death rate of an individual of trait v,

- $c_{v,w}^K = c_{v,w}/K \in \mathbb{R}_+$, the competition imposed by an individual of trait w onto an individual of trait v,

- $\mu_K \in [0, 1]$, the probability of mutation at a birth event,

- $m(v, \cdot) \in \mathcal{M}_p(V)$, the law of the trait of a mutant offspring produced by an individual of trait v.

The process N^K with values in $\mathbb{D}(\mathbb{R}_+, \mathbb{N}^V)$ describes the state of the population, where $N_v^K(t)$ denotes the number of individuals of trait $v \in V$ alive at time $t \geq 0$. We assume that edges in E mark the possibility of mutation and hence $m(v, w) > 0$ if and only if $(v, w) \in E$.

Remark 2.1. We could also allow for μ_K to depend on $v \in V$ as long as $\mu_K(v) = \mu_K h(v)$ for some strictly positive function h that is independent of K. However, this would not change the characterisation of the limit, and hence we assume a constant μ_K to simplify the notation.

Moreover, we assume that, for every $v \in V$, $c_{v,v} > 0$, which ensures the finiteness of the population size (see Corollary A.7). The parameter K is scaling the competitive pressure and, through this self-competition, fixes the equilibrium size of the population to the order of K.

As a consequence of our parameter definitions, the process N^K is characterised by its infinitesimal generator:

$$
\mathcal{L}^K \phi(N) = \sum_{v \in V} (\phi(N + \delta_v) - \phi(N)) \left(N_v b_v (1 - \mu_K) + \sum_{w \in V} N_w b_w \mu_K m(w, v) \right)
$$
$$
+ \sum_{v \in V} (\phi(N - \delta_v) - \phi(N)) N_v \left(d_v + \sum_{w \in V} c_{v,w}^K N_w \right), \tag{2.1}
$$

where $\phi : \mathbb{N}^V \to \mathbb{R}$ is measurable and bounded. Such processes have been explicitly constructed in terms of Poisson random measures in [22].

Due to the scaling of the competition c^K, the population admits a quasi-stationary distribution with a total population size of order K, when at least one type has a birth rate $b_.$ strictly larger than its death rate $d_.$ (see [7] or [11] for details). Since the mutation probability μ_K tends to zero as $K \to \infty$, the process N^K/K converges (on finite time intervals) to the mutation-free Lotka-Volterra system (2.3) involving all initial coexisting resident traits. We are interested in the long-term evolution of the population and want to study successive invasions by new mutant populations. Given the fact that a mutant population that is initially of order K^γ, $\gamma < 1$, needs a time of order $\log K$ to grow exponentially to the order of K, we have to rescale the time by $\log K$ to obtain a non trivial limit.

It is convenient to describe the population size of a certain trait $v \in V$ by its K-exponent at time t, $\beta_v^K(t)$, defined by

$$
\beta_v^K(t) := \frac{\log(1 + N_v^K(t \log K))}{\log K}, \quad t \geq 0, \tag{2.2}
$$

which is equivalent to $N_v^K(t \log K) = K^{\beta_v^K(t)} - 1$. Since the population size is restricted to order K by the competition, β_v^K ranges between 0 and 1, as $K \to \infty$ (see Corollary A.7 for a rigorous statement).

For the sake of readability, we now introduce the terminology we will use in the sequel.

Definition 2.2.   *1. A trait $v \in V$ with exponent β_v^K is called macroscopic if, for every $\varepsilon > 0$, there exists K_ε such that, for every $K \geq K_\varepsilon$, $\beta_v^K > 1 - \varepsilon$.*

   2. A trait that is not macroscopic is called microscopic.

   3. The set of living traits is the set $\{v \in V : \beta_v^K > 0\}$.

When K is large enough, the macroscopic traits interact on any finite time interval according to the corresponding mutation-free Lotka-Volterra system (see Chapter 11, Theorem 2.1 in [21] for the proof of this law of large numbers): Let $\mathbf{v} \subset V$, then the mutation-free Lotka-Volterra system associated to $\mathbf{v}$ is

$$
\dot{n}_w(t) = \left(b_w - d_w - \sum_{v \in \mathbf{v}} c_{w,v} n_v(t) \right) n_w(t), \quad w \in \mathbf{v}, \ t \geq 0. \tag{2.3}
$$

For a subset $\mathbf{v} \subset V$ of traits, we denote by $\bar{n}(\mathbf{v}) \in \mathbb{R}_+^V$ the unique stable equilibrium of the Lotka-Volterra system (2.3), when it exists, and where to simplify notations, we

extend it by $\bar{n}_w(\mathbf{v}) = 0$ for $w \notin \mathbf{v}$. In the case where $\mathbf{v} = \{v\}$, we obtain from classical results on Lotka-Volterra models (see [7] for instance)

$$\bar{n}_v(v) = (b_v - d_v)/c_{v,v} \vee 0.$$

If $\mathbf{v}$ denotes the set of macroscopic traits, we call the traits $v \in \mathbf{v}$ such that $\bar{n}_v(\mathbf{v}) > 0$ *resident* and denote them by $\tilde{\mathbf{v}}$.

The approximate rate at which a mutant of trait w grows in a population of coexisting resident traits $\tilde{\mathbf{v}} \subset \mathbf{v}$ is called *invasion fitness* and is denoted by $f_{w,\mathbf{v}}$, where

$$f_{w,\mathbf{v}} := b_w - d_w - \sum_{v \in \mathbf{v}} c_{w,v} \bar{n}_v(\mathbf{v}).$$

If $f_{w,\mathbf{v}} > 0$, the trait w is called *fit*. If $f_{w,\mathbf{v}} < 0$, the trait w is called *unfit*. Note that for all $w \in \mathbf{v}$, $f_{w,\mathbf{v}} \leq 0$, else the equilibrium would not be stable. The case of $f_{w,\mathbf{v}} = 0$, for $w \in V \backslash \mathbf{v}$ is excluded (see Remark 2.5).

Mutants can be produced along (directed) edges of the graph. We denote by $d(v, w)$ the graph distance, i.e. the length of the shortest (directed) path from v to w in $\mathcal{G} = (V, E)$. For a subset $\mathbf{v} \subset V$ we define

$$d(\mathbf{v}, w) := \min_{v \in \mathbf{v}} d(v, w) \quad \text{and} \quad d(w, \mathbf{v}) := \min_{v \in \mathbf{v}} d(w, v).$$

2.2 Results

Let a finite graph $\mathcal{G} = (V, E)$ be given. We will always make two assumptions in the sequel, that we present now. The first one ensures that there is a competition between individuals of the same type, which prevents the population size from going to infinity.

Assumption 2.3.

$$c_{v,v} > 0 \quad \forall v \in V.$$

The second assumption is more technical. It excludes that mutants are produced at distance $\alpha \in \mathbb{N}$ from the resident traits or that non-resident traits have fitness $f_{w,\mathbf{v}} = 0$. In both cases, the respective subpopulations can neither be approximated by sub- nor super-critical branching processes. In the first case, the fixation of the population would occur randomly, and in the second, the population could both grow or shrink due to fluctuations.

Assumption 2.4.

$$\alpha \in \mathbb{R}_+^* \backslash \mathbb{N} \quad \text{and} \quad f_{w,\mathbf{v}} \neq 0 \quad \forall \, \mathbf{v} \subset V, \, w \notin \{v \in \mathbf{v} : \bar{n}_v(\mathbf{v}) > 0\}.$$

The trajectories $(\beta_w(t), w \in V)$, which are limits of $(\beta_w^K(t), w \in V)$ (recall (2.2)), are defined up to a stopping time T_0 by the following inductive procedure:

Let $\mathbf{v}_0 \subset V$ be the initial set of macroscopic (not necessarily resident) traits. For simplicity, we assume that the initial orders of population sizes converge to $\beta_w(0) := \left(1 - \frac{d(\mathbf{v}_0, w)}{\alpha}\right)_+$. More general initial conditions are discussed in Remark 2.9.

The increasing sequence of invasion times is denoted by $(s_k)_{k \geq 0}$, where $s_0 := 0$ and, for $k \geq 1$,

$$s_k := \inf\{t > s_{k-1} : \exists \, w \in V \backslash \mathbf{v}_{k-1} : \beta_w(t) = 1\}.$$

Here, $\mathbf{v}_k$ denotes the set of previously coexisting resident traits $\tilde{\mathbf{v}}_{k-1}$ together with the trait $w \in V \backslash \mathbf{v}_{k-1}$ that satisfies $\beta_w(s_k) = 1$.

For $s_{k-1} \leq t \leq s_k$, for any $w \in V$, $\beta_w(t)$ is defined by

$$\beta_w(t) := \max_{u \in V} \left[\beta_u(s_{k-1}) + (t - t_{u,k} \wedge t) f_{u,\mathbf{v}_{k-1}} - \frac{d(u, w)}{\alpha}\right] \vee 0, \tag{2.4}$$

where, for any $w \in V$,

$$t_{w,k} := \begin{cases} \inf\{t \geq s_{k-1} : \exists\, u \in V : d(u,w) = 1, \beta_u(t) = \frac{1}{\alpha}\} & \text{if } \beta_w(s_{k-1}) = 0 \\ s_{k-1} & \text{else} \end{cases} \qquad (2.5)$$

is the first time in $[s_{k-1}, s_k]$ when this trait arises.

The stopping time T_0, that terminates the inductive construction of the limiting trajectories, is set to s_k if

(a) there is more than one $w \in V \backslash \mathbf{v}_{k-1}$ such that $\beta_w(s_k) = 1$;

(b) the mutation-free Lotka-Volterra system associated to $\mathbf{v}_k$ does not have a unique globally attractive stable equilibrium (in particular, if such an equilibrium does not exist for $\mathbf{v}_0$, T_0 is set to 0);

(c) there exists $w \in V \backslash \tilde{\mathbf{v}}_{k-1}$ such that $\beta_w(s_k) = 0$ and $\beta_w(s_k - \varepsilon) > 0$ for all $\varepsilon > 0$ small enough.

(d) there exists $w \in V \backslash \mathbf{v}_{k-1}$ such that $s_k = t_{w,k}$.

Remark 2.5. Notice that conditions (a), (c), and (d) in the definition of T_0 are here to exclude very specific and non generic cases where one coordinate reaches 1 while another reaches 1 or reaches 0 from above, or a new trait arises at the exact same time. They are difficult to handle for technical reasons.

Condition (b) is necessary since we can only describe the system as long as the Lotka-Volterra phases are governed by the convergence to a unique equilibrium. The existence of such an equilibrium is not always guaranteed and not yet fully characterized in arbitrary dimensions (see [42] for the case of dimension 3 for instance).

Remark 2.6. Note that $\beta_w(t)$ in (2.4) can never exceed 1 since $f_{u,\mathbf{v}_{k-1}} \leq 0$, for all $u \in \mathbf{v}_{k-1}$, and the stopping time s_k is triggered whenever a trait $u \in V \backslash \mathbf{v}_{k-1}$ reaches $\beta_u(t) = 1$ (which implies $\beta_u(s_{k-1}) + (t - t_{u,k} \wedge t) f_{u,\mathbf{v}_{k-1}} \leq 1$).

The $t_{w,k}$'s do not keep track of traits that die out in $[s_{k-1}, s_k]$ and then reappear. However, since the fitnesses do not change between invasions, such a trait would have a negative invasion fitness (else it would not die out). Hence, it would not start growing on its own if it reappears, but only follow along another trait due to mutants. It would therefore not contribute to the maximum over $u \in V$ in (2.4).

With these definitions at hand we can now formulate the main results of this paper. The first one describes the convergence of the orders of the population sizes.

Theorem 2.7. *Let $\mathcal{G} = (V, E)$ be a finite graph. Suppose that Assumption 2.3 and 2.4 hold and consider the model defined by (2.1) with $\mu_K = K^{-1/\alpha}$. Let $\mathbf{v}_0 \subset V$ and assume that, for every $w \in V$,*

$$\beta_w^K(0) \to \left(1 - \frac{d(\mathbf{v}_0, w)}{\alpha}\right)_+, \quad (K \to \infty) \quad \text{in probability.} \qquad (2.6)$$

Then, for all $T > 0$, as $K \to \infty$, the sequence $((\beta_w^K(t), w \in V), t \in [0, T \wedge T_0])$ converges in probability in $\mathbb{D}([0, T \wedge T_0], \mathbb{R}_+^V)$ to the deterministic, piecewise affine, continuous function $((\beta_w(t), w \in V), t \in [0, T \wedge T_0])$, which is defined in (2.4).

The second result describes the convergence of the rescaled population sizes and precisely quantifies the times of invasions.

Proposition 2.8. *Let $\mathcal{G} = (V, E)$ be a finite graph. Suppose that Assumption 2.3 and 2.4 hold and consider the model defined by (2.1) with $\mu_K = K^{-1/\alpha}$. With the same notations as in Theorem 2.7, for all $T > 0$, as $K \to \infty$, the sequence $((N_w^K(t \log K)/K, w \in V), t \in [0, T \wedge T_0])$ converges in the sense of the finite dimensional distributions to a deterministic jump process $((N_w(t), w \in V), t \in [0, T \wedge T_0])$, which is defined as follows:*

(i) *For $t \in [0, T_0]$, $N(t)$ jumps between different Lotka-Volterra equilibria according to*

$$N_w(t) := \sum_{k \in \mathbb{N}: s_{k+1} \leq T_0} \mathbf{1}_{s_k \leq t < s_{k+1}} \mathbf{1}_{w \in \mathbf{v}_k} \bar{n}_w(\mathbf{v}_k).$$

(ii) *The invasion times s_k and the times $t_{w,k}$ when new mutants arise can be calculated as follows. We define the increasing sequence $(\tau_\ell, \ell \geq 0) = \{s_k, k \geq 0\} \cup \{t_{w,k}, w \in V, k \geq 0\}$ of invasion times or appearance times of new mutants, and $(M_\ell, \ell \geq 0)$ the sets of living traits in the time interval $(\tau_\ell, \tau_{\ell+1}]$. Initially, $\tau_0 = s_0 = 0$ and, according to (2.6), $M_0 = \{w \in V : d(\mathbf{v}_0, w) < \alpha\} = \{w \in V : \beta_w(0) > 0\}$. For $s_{k-1} \leq \tau_{\ell-1} < s_k$, τ_ℓ is defined as*

$$\tau_\ell := s_k \wedge \min\{t_{w,k} : w \in V, t_{w,k} > \tau_{l-1}\}.$$

Given τ_ℓ and $M_{\ell-1}$, we set $M_\ell := (M_{\ell-1} \backslash \{w \in V : \beta_w(\tau_\ell) = 0\}) \cup \{w \in V : \tau_\ell = t_{w,k}\}$. τ_ℓ is then given by

$$\tau_\ell - \tau_{\ell-1} = \min_{\substack{w \in M_{\ell-1}: \\ f_{w,\mathbf{v}_{\ell-1}} > 0}} \frac{\left(1 \wedge \frac{d(w, V \backslash M_{\ell-1})}{\alpha}\right) - \beta_w(\tau_{\ell-1})}{f_{w,\mathbf{v}_{\ell-1}}}. \qquad (2.7)$$

Remark 2.9. We could allow for more general initial conditions of the form

$$\beta_w^K(0) \to \tilde{\beta}_w \in [0, 1],$$

with $\tilde{\beta}_w$, $w \in V$, deterministic and $\mathbf{v}_0 := \{w \in V : \tilde{\beta}_w = 1\} \neq \emptyset$. An inductive application of Corollary A.7, similar to the induction proving (4.9), implies that within a time of order 1, for all $w \in V$, $\beta_w^K \cong \max_{u \in V}[\tilde{\beta}_u - d(u, w)/\alpha]_+$. This is due to incoming mutations from larger neighbouring populations. We therefore set $\beta_w(0) := \max_{u \in V}[\tilde{\beta}_u - d(u, w)/\alpha]_+$ in Theorem 2.7 and $M_0 := \{w \in V : \beta_w(0) > 0\}$ in Proposition 2.8. The rest of the results remains unchanged.

Note that the specific initial conditions that we choose in the results above are quite natural as they depict the situation of a resident (wildtype) population spreading new mutants. They are simply the $\beta_w(0)$'s that arise from $\tilde{\beta}_w = 1$ for $w \in \mathbf{v}_0$ and 0 else.

Remark 2.10. The limiting jump process $N(t)$ resembles an adaptive walk or flight, as studied in [35, 32, 38, 34, 1]. For a constant competition kernel $c_{v,w} \equiv c > 0$, we consider the fixed fitness landscape given by $r_v = b_v - d_v$. Since in this case $f_{w,v} = r_w - r_v$, the process jumps along edges towards traits of increasing fitness r.

The above results are in the vein of Theorem 2.1 and Corollary 2.3 in [8]. There are however many differences between the setting considered in [8] and our setting.

Due to the horizontal transfer between individuals, Champagnat and coauthors obtained trajectories where a "dominant" population, i.e. with the size of highest order, could be non resident, i.e. of order negligible with respect to K. They could also witness extinction on a $\log K$ time scale as well as evolutionary suicide. The absence of horizontal transfer in our case prevents such behaviours.

We consider a general finite graph of mutations with possible back mutations, whereas their graph was embedded in $\mathbb{Z}$ and did not allow for back mutations. We also allow for the coexistence of several resident traits in the population at equilibrium. The two main difficulties in the proofs compared to [8] are thus to handle the generality of the graph of mutations, and to extend some approximation results to the multidimensional case.

3 Surprising phenomena arising from geometry and mutation rate

In this section, we present some non intuitive behaviours of the population process, which stem from the mutation scale or the generality of the mutational graph that we allow for. They are direct applications of Theorem 2.7 and Proposition 2.8, and provide explicit computations of exponents (2.4) and time intervals (2.7).

Several examples are built on directed graphs. Although this is not a necessary condition to obtain the desired phenomena, it allows a simplified study (especially of the decay phases).

The examples are supported by figures that display the respective exponents β over time. Note that in the formulation of the examples we always assume monomorphic initial conditions with one subpopulation (i.e. the wildtype) at its equilibrium and the other traits at a population size of zero. These initial conditions are formulated in terms of the limiting process, i.e. with the equilibrium size of a subpopulation being $\bar{n}$ and not $\bar{n}K$. The monomorphic initial conditions allow a comparison between different graphs, particularly in Example 4. The figures however display initial values for β's that are of the form of Theorem 2.7 and Proposition 2.8, where mutations of the wildtype have spread in a radius of α. These are attained after a time of order 1, much in the spirit of Remark 2.9.

We first introduce some notations for the sake of readability. Recall that the following symbols are not transitive, which explains why we may witness such suprising dynamics.

Definition 3.1. *Let $w, v \in V$ and $\mathbf{v} \subset V$. We write*

1. *with high probability to mean "with a probability converging to 1 as $K \to \infty$",*
2. *$w > \mathbf{v}$ if and only if $f_{w,\mathbf{v}} > 0$, that is if w can invade in $\mathbf{v}$,*
3. *$w < \mathbf{v}$ if and only if $f_{w,\mathbf{v}} < 0$, that is if w cannot invade in $\mathbf{v}$,*
4. *$w \gg v$ (or $v \ll w$) if and only if $f_{w,v} > 0$ and $f_{v,w} < 0$, that is if w can invade in v and fixate,*
5. *$w \equiv v$ if and only if $f_{w,v} > 0$ and $f_{v,w} > 0$, that is if w and v can coexist,*
6. *$w \frown v$ if and only if $f_{w,v} < 0$ and $f_{v,w} < 0$, that is if w and v can neither invade in each other.*

3.1 Back mutations before adaptation

In the following, we build an example where the ancestry of the resident population comes from back mutations from an ancestral trait, even if the mutations happening in between are not deleterious.

Example 3.2. Let us consider the graph $\mathcal{G}$ depicted on Figure 1 where $V = \{0, 1, 2, 3\}$ and $E = \{[0,1], [1,2], [2,0], [0,3]\}$. Let $\alpha > 2$, an initial condition given by $(\bar{n}(0), 0, 0, 0)$ and a fitness landscape given by

$$0 \ll 1 \ll 2, \quad 3 \frown 0, \quad 3 \frown 1$$

$$0 \equiv 2, \quad 3 > \{0, 2\}, \quad 2 < 3 \tag{3.1}$$

$$f_{2,0} < 2f_{1,0} \tag{3.2}$$

$$f_{0,1} \geq f_{3,1}, \quad f_{2,0} \leq f_{1,0} \tag{3.3}$$

$$i_1 := \frac{1 - 1/\alpha}{f_{0,1}} < \frac{-(1 - 4/\alpha)}{f_{2,1}} =: i_2 \tag{3.4}$$

In this case, Proposition 2.8 implies that on the $\log K$ time scale, the rescaled macroscopic population then jumps from traits $0 - 1 - 2$ then to coexistence between 0 and
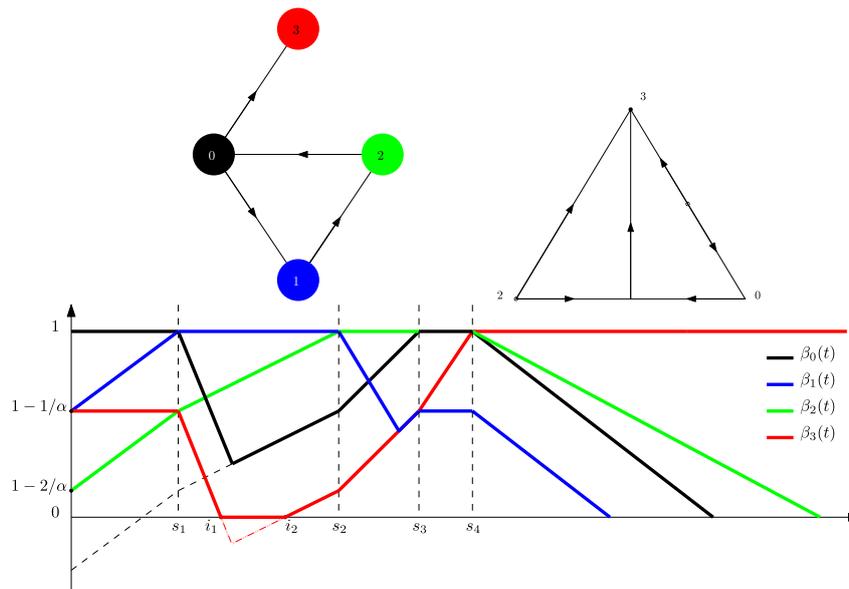
Figure 1: Graph $\mathcal{G}$, phase portrait of traits 0-2-3, and exponents $\beta(t)$ of Example 3.2.

2, followed by the invasion and fixation of 3 which is produced with high probability, due to Condition (3.4), by individuals of type 0 which have the sequence $0 - 1 - 2$ as ancestry. In other words, the final resident individuals of trait 3, although they can be produced by individuals of trait 0 directly, come from a sequence of mutations which went around the loop $0 - 1 - 2$ of $\mathcal{G}$. Conditions (3.1), summarized in Figure 1, imply phase portrait number 8 in the classification of Zeeman [42]. Condition (3.2) ensures that trait 1 becomes resident before 2. Condition (3.3) is not necessary but allows to simplify the setting. The exponents are drawn in Figure 1.

Note that this scenario can also happen in the rare mutation regime considered in [7] (for example $\alpha \in (0, 1)$): the average waiting time until a mutant of type 1 appears is then of order $O(1/K\mu_K) = O(K^{-1+1/\alpha}) \gg \log K$. Once it has appeared, it survives with positive probability and the succession of invasions and fixations above takes place on the $\log K$ time scale, separated by mutation events on the $K^{-1+1/\alpha}$ time scale. What is new in our case is that such a scenario can still take place for higher mutation rates than the ones considered in [7], and on a $\log K$ time scale.

3.2 Non-intuitive mutational pathways in the high mutation framework

3.2.1 Longer or shorter path than expected

If evolution and mutation time scales are separated (i.e. in the rare mutation regime), mutations occur one at a time, and the number of successive resident traits from the wild type to the type gathering k successively beneficial mutations is k. This is not the case if mutations are faster, in which case it is possible to observe either more or less resident traits, as the following examples show.

Example 3.3. Let us consider the directed graph $\mathcal{G}$ depicted on Figure 2, where $V = \{00, 01, 10, 11\}$ and $E = \{[00, 01], [00, 10], [01, 11], [10, 11]\}$. Let $\alpha > 2$, an initial condition
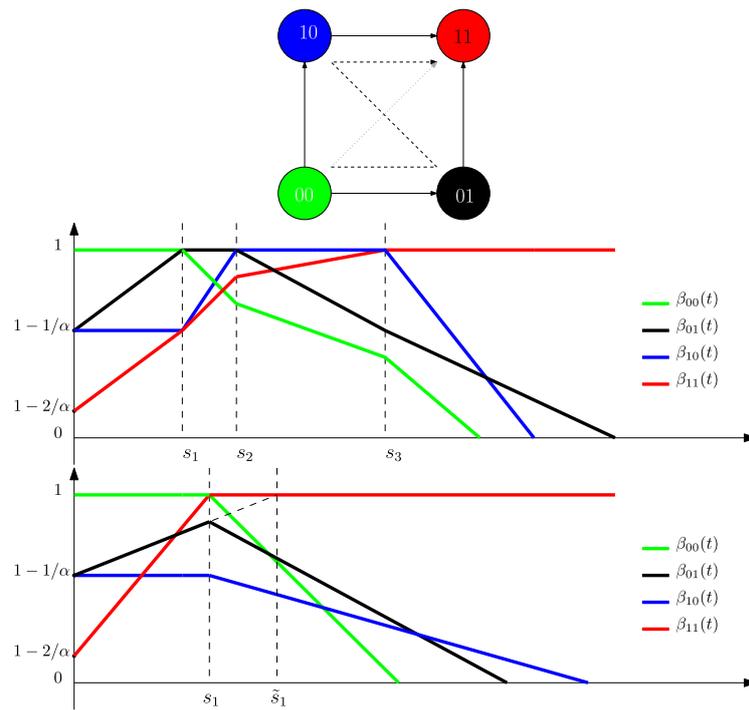
Figure 2: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Examples 3.3 and 3.4.

given by $(\bar{n}(00), 0, 0, 0)$ and a fitness landscape given by

$$00 \ll 01 \ll 10 \ll 11, \quad and \quad 01 \ll 11$$

$$10, 11 \frown 00$$

$$f_{11,00} < f_{01,00} \tag{3.5}$$

$$f_{10,01} > f_{11,01} \tag{3.6}$$

In this case, in the rare mutation regime, the rescaled macroscopic population jumps along $00 - 01 - 11$.

In the regime of Theorem 2.7, Proposition 2.8 implies that the rescaled macroscopic population jumps along $00 - 01 - 10 - 11$ on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 2. Note that Condition (3.5) ensures that 11 does not invade before 01, it is not necessary but allows to simplify the setting. Condition (3.6) ensures that 11 does not invade before 10.

Example 3.4. Let us consider the directed graph $\mathcal{G}$ depicted on Figure 2, where $V = \{00, 01, 10, 11\}$ and $E = \{[00, 01], [00, 10], [01, 11], [10, 11]\}$. Let $\alpha > 2$, an initial condition given by $(\bar{n}(00), 0, 0, 0)$ and a fitness landscape given by

$$01 > 00, \quad 11 \gg 00$$

$$10 < 00 \tag{3.7}$$

$$01, 10 < 11$$

$$\frac{2}{f_{11,00}} < \frac{1}{f_{01,00}} \tag{3.8}$$

In this case, in the rare mutation regime, the rescaled macroscopic population still jumps along $00 - 01 - 11$, under the additional assumption that $f_{11,01} > 0$.

In the regime of Theorem 2.7, Proposition 2.8 implies that the rescaled macroscopic population directly jumps from 00 to 11 on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 2. Condition (3.8) ensures that 11 fixates before 01. Condition (3.7) is not necessary but allows to simplify the setting. Note that equation (2.7) implies that $s_1 = 2/f_{11,00}$ and $\tilde{s}_1 = 1/f_{01,00}$.

3.2.2 Price of anarchy

We build an example where adding a new possible mutation path to a fit trait increases the time until it appears macroscopically.

Example 3.5. Let us consider the graph $\mathcal{G}$ depicted on Figure 3, where $V = \{1, 2a, 2b, 3\}$ and the edge set is either $E_1 = \{[1, 2a], [2a, 3], [2b, 3], [3, 2b]\}$ or $E_2 = \{[1, 2a], [2a, 3], [2b, 3], [2a, 2b], [3, 2b]\}$. Let $\alpha > 3$, an initial condition given by $(\bar{n}(1), 0, 0, 0)$ and a fitness landscape given by

$$1 \ll 2a \ll 3, \quad and \quad 2a \ll 2b \tag{3.9}$$
$$1 < 2b, \quad and, \quad 1, 2b < 3$$

$$f_{2a,1} \geq f_{3,1}, f_{2b,1}, \quad and \quad \frac{1}{f_{2b,2a}} < \frac{1}{f_{3,2a}} < \frac{2}{f_{2b,2a}} \tag{3.10}$$

$$0 < f_{3,2b} < f_{3,2a}. \tag{3.11}$$

In this case, if the edge set is E_1, Proposition 2.8 implies that the rescaled macroscopic population jumps along traits $1 - 2a - 3$ in a time T on the $\log K$ time scale. But if the edge set is E_2, the population jumps along $1 - 2a - 2b - 3$ and the time to reach 3 is $\tilde{T} > T$ (see Figure 3). Condition (3.10) ensures that $2b$ invades first when the edge set is E_2 but not when it is E_1, in other words β_{2b} reaches 1 before β_3 if started at $1 - 1/\alpha$ but not at $1 - 2/\alpha$. And Condition (3.11) enlarges the time of fixation of 3. Note that the first inequality in Condition (3.10) is not necessary but allows to simplify the second one. Moreover, observe that equation (2.7) implies $\tilde{s}_2 - \tilde{s}_1 = 1/f_{2b,2a}$ and $s_2 - s_1 = 1/f_{3,2a}$. Note that in the rare mutation regime we can observe this phenomenon on the mutation time scale, but only with probability strictly smaller than 1, since both $2b$ and 3 are fit with respect to $2a$ and can both invade with positive probability once they are produced.

3.2.3 Counter cycle

Example 3.6. Let us consider the graph $\mathcal{G}$ depicted on Figure 4, where $V = \{1, 2, 3\}$ and the edge set is $E = \{[1, 2], [2, 3], [3, 1]\}$. Let $\alpha > 2$, an initial condition given by $(\bar{n}(1), 0, 0)$ and a fitness landscape given by

$$1 \gg 2, \quad 2 \gg 3, \quad 3 \gg 1.$$

In this case, Proposition 2.8 implies that the rescaled macroscopic population jumps along traits $1 - 3 - 2$ (in the clockwise sense) although the mutations are directed counterclockwise. More precisely, the exponents are drawn in Figure 4. Moreover, if Conditions (3.12) below are fulfilled the period is shorter and shorter, and acceleration takes place, as it is depicted in Figure 4.

$$f_{2,3} > -f_{1,3}$$
$$f_{1,2} > -f_{3,2}$$
$$f_{3,1} > -f_{2,1} \tag{3.12}$$

Note that in the rare mutation regime, with the chosen parameters, there would be no evolution since $2 < 1$. Moreover, there are no parameters such that counter cyclic or accelerating behaviour could arise.
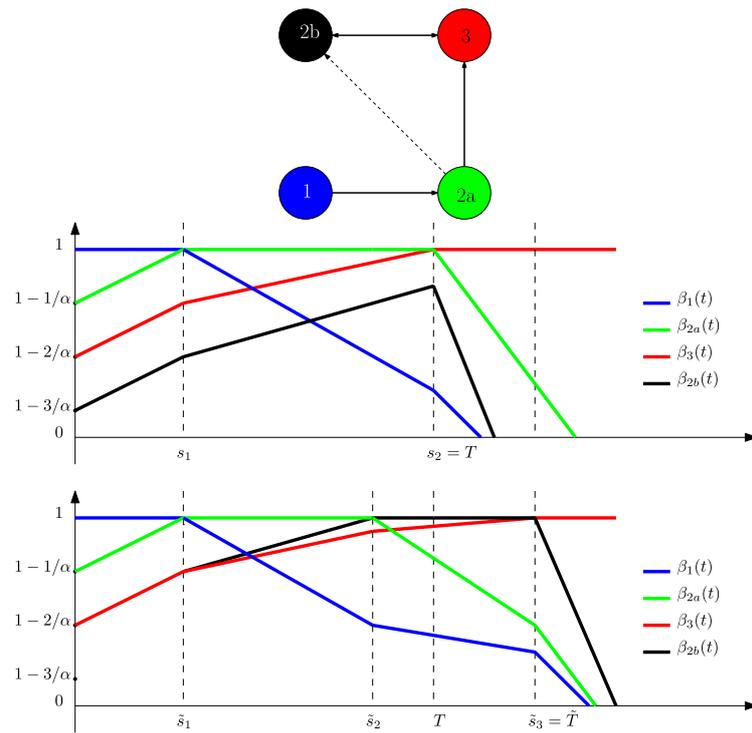
Figure 3: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.5, with edge set E_1 (above) and E_2 (below).
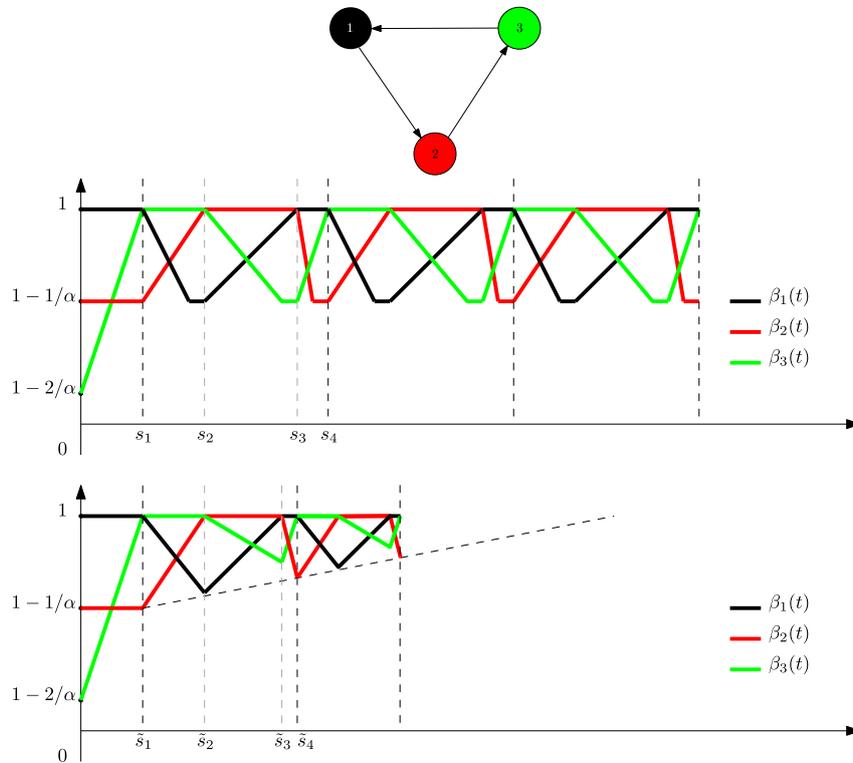
Figure 4: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.6, without Assumption 3.12 (above) and with Assumption 3.12 (below).

3.3 Arbitrary large jumps on the $\log K$-time scale

A natural question to ask is if the "cut-off" α restricts the range of the jumps, on the $\log K$ time-scale, to traits which are at a distance less than α. The answer is no, as the following example shows.

Example 3.7. Let us consider the graph $\mathcal{G}$ depicted on Figure 5, where $V = \{0, 1, 2, 3, 4\}$ and $E = \{[0, 1], [1, 2], [2, 3], [3, 4]\}$. Let $\alpha \in (3, 4)$, an initial condition given by $(\bar{n}(0), 0, \ldots, 0)$ and a fitness landscape given by

$$1, 2 < 0, \quad 3, 4 > 0, \quad 0, 1, 2, 3 < 4$$
$$\frac{1}{f_{4,0}} + \frac{-1 + 4/\alpha}{f_{3,0}} < \frac{3/\alpha}{f_{3,0}} \tag{3.13}$$

In this case, the cut-off is in between traits 3 and 4 (meaning that $K\mu_K^i \to 0$ for $i > 3$) and thus the population of trait 4 vanishes at time 0. However, Proposition 2.8 implies that the rescaled macroscopic population jumps from trait 0 to trait 4 in a time

$$s_1 = \frac{-1 + 4/\alpha}{f_{30}} + \frac{1}{f_{40}}$$

on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 5. Condition (3.13) ensures that trait 4 fixates before trait 3.

It is easy to generalize this example to construct jumps to any distance L larger than α, by taking larger and larger fitnesses after the negative fitness region. The condition implying emergence of trait L is then a little more technical to write, since one has to compute the time for the piecewise affine function $\beta_L(t)$ (with multiple slope-breaks) to reach 1 before the other traits. Example 3.7 constitutes the simplest non-trivial example of this phenomenon. Example 3.8 is a further case where a more distant trait fixates, and two intermediate times $t_{4,1}$ and $t_{5,1}$ occur (recall the definition in (2.5)).

Example 3.8. Let us consider the graph $\mathcal{G}$ depicted on Figure 6, where $V = \{0, 1, 2, 3, 4, 5\}$ and $E = \{[0, 1], [1, 2], [2, 3], [3, 4], [4, 5]\}$. Let $\alpha \in (3, 4)$, an initial condition given by $(\bar{n}(0), 0, \ldots, 0)$ and a fitness landscape given by

$$1, 2 < 0, \quad 3, 4, 5 > 0, \quad 0, 1, 2, 3, 4 < 5$$
$$f_{3,0} < f_{4,0} < f_{5,0} \quad \text{and} \quad \frac{-1 + 4/\alpha}{f_{3,0}} + \frac{-4/\alpha + 5/\alpha}{f_{4,0}} + \frac{1}{f_{5,0}} < \frac{3/\alpha}{f_{3,0}}. \tag{3.14}$$

In this case, the cut-off is in between traits 3 and 4 (meaning that $K\mu_K^i \to 0$ for $i > 3$) thus population of trait 4 and 5 vanishes at time 0. However, Proposition 2.8 implies that the rescaled macroscopic population jumps from trait 0 to trait 5 in a time

$$s_1 = \frac{-1 + 4/\alpha}{f_{3,0}} + \frac{-4/\alpha + 5/\alpha}{f_{4,0}} + \frac{1}{f_{5,0}}$$

on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 6. Condition (3.14) ensures that trait 5 fixates before traits 3 and 4. The first inequality is not needed but allows to simplify the second one. The dotted lines in the figures allow to construct the points where some exponents become positive.

3.4 Effective random walk across fitness valleys

The following two examples consider cases where there is no fitter trait within the α-radius of the initial resident trait. The population hence has to cross a valley in the
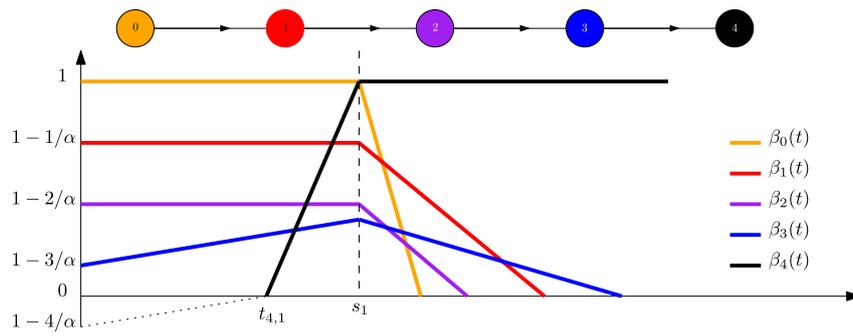
Figure 5: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.7.
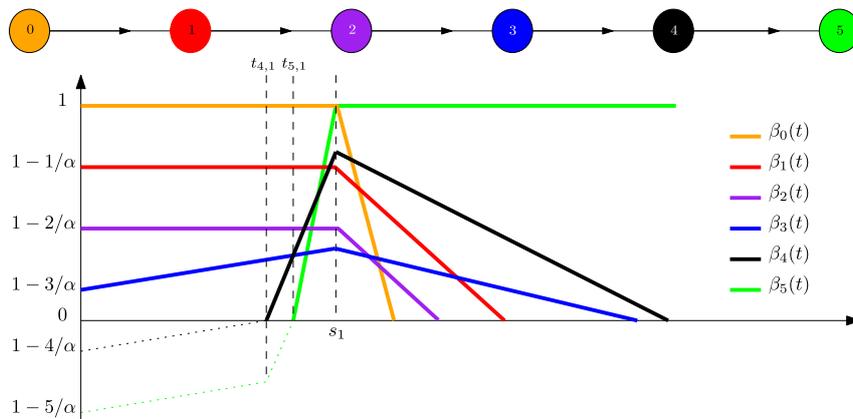
Figure 6: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.8.

fitness landscape, with mutations that occur much more rarely than within a time of order 1 or $\log K$, resulting in evolutionary adaptation only being visible on a time scale longer than $\log K$. The examples combine a result from [6], which treats the crossing of a fitness valley of width $L > \alpha$ on the time scale $1/K\mu_K^L$, with the main results of this paper, which describe the faster dynamics on the $\log K$ time scale once the valley is crossed.

For the convenience of the reader, we briefly recall the key heuristics of the result in [6] for our simple situation of a fitness valley of width 2. Think of the trait space of $\{0, i, 1\}$, where 0 is the resident trait, i an intermediate unfit trait and 1 a fit trait (with respect to 0). Mutation is possible from 0 to i and from i to 1 and we assume $1 < \alpha < 2$. Then, within a time of order 1, trait 0 attains a population size close to $\bar{n}_0 K$. It produces mutants of type i at rate $\bar{n}_0 K b_0 \mu_K$.

This mutant population has a growth rate of $f_{i,0} < 0$ and would therefore shrink and go extinct without the incoming mutations from trait 0. The equilibrium size of trait i, where the negative growth rate is exactly balanced by the incoming mutations is $\bar{n}_0 b_0 K \mu_K / |f_{i,0}|$. Consequently, mutants of trait 1 are produced by trait i at rate $\bar{n}_0 b_0 b_i K \mu_K^2 / |f_{i,0}| \ll 1$. These mutations determine the time scale of $1/K\mu_K^2 = K^{-1+2/\alpha} \gg \log K$ on which the crossing of the fitness valley takes place.

Finally, the population size of mutants of trait 1, stemming from one of these rare mutation events, can be approximated by a birth and death process with birth rate b_1 and death rate $d_1 - c_{1,0}\bar{n}_0$. The probability of invasion in the population, i.e. non-extinction, is therefore $(b_1 - d_1 - c_{1,0}\bar{n}_0)/b_1 = f_{1,0}/b_1$, which is a classical result (see e.g. [7]). Note

that, on the time scale $1/K\mu_K^2$, only the jump from 0 to 1 is visible and it occurs at a random exponential time with parameter $\bar{n}_0 b_0 b_i/|f_{i,0}| \cdot f_{1,0}/b_1$.

We now turn to the concrete examples.

3.4.1 2 effective sites

Example 3.9. Let us consider the graph $\mathcal{G}$ depicted on Figure 7, where $V = \{0, 1a, 1b, i\}$ and $E = \{[0, i], [i, 1a], [1a, 1b], [i, 1b]\}$. We suppose that whenever there are several outgoing edges from a vertex v, the mutation kernel is uniform among the nearest neighbouring vertices. Let $\alpha \in (1, 2)$, an initial condition given by $(\bar{n}(0), 0, \dots, 0)$ and a fitness landscape given by

$$1a \gg 0 \gg 1b \gg 1a$$
$$i < 0, \quad i < 1a, \quad i < 1b.$$

In this case, according to [6], the time to cross the fitness valley is of order $O(1/K\mu_K^2) = O(K^{-1+2/\alpha}) \gg \log K$, thus the first mutant of type $1a$ will appear on this time scale, and will invade with positive probability. Then, once $1a$ has become the resident trait and the fitness landscape is changed, type $1b$ fixates deterministically in a time of order $O(\log K)$ and one has to wait again a time of order $O(K^{-1+2/\alpha})$ until the appearance of the next mutant of type 0. Thus, on the time scale $O(K^{-1+2/\alpha})$, the population process converges to a jump process between the two states 0 and $1b$ with positive jump rates although the fitness $f_{1b,0}$ is negative.

Theorem 3.10. *As $K \to \infty$, the following convergence holds*

$$(N_0^K, N_{1b}^K)(tK^{-1+2/\alpha})/K \Rightarrow \bar{n}(X_t)\delta_{X_t}$$

for finite dimensional distributions, where X_t is a continuous time Markov chain on $\{0, 1b\}$ with transition rates

$$r_{0 \to 1b} = \frac{\bar{n}(0)b_0 b_i}{2|f_{i,0}|} \frac{f_{1a,0}}{b_{1a}},$$
$$r_{1b \to 0} = \frac{\bar{n}(1b)b_{1b} b_i}{2|f_{i,1b}|} \frac{f_{0,1b}}{b_0}.$$

Note that the additional $1/2$ in the rates, when compared to the heuristics above, stems from the fact that mutants of trait i can be either of type 0 or 1 with probability $1/2$ each and only one of those contributes to the respective jump rate.

3.4.2 3 effective sites

Example 3.11. Let us consider the graph $\mathcal{G}$ depicted on Figure 7. We suppose that whenever there are several outgoing edges from a vertex v, the mutation kernel is uniform among the nearest neighbouring vertices. Let $\alpha \in (1, 2)$, an initial condition given by $(\bar{n}(0), 0, \dots, 0)$ and a fitness landscape given by

$$
\begin{array}{ccc}
1a \gg 0 & 2a \gg 0 & 2a \gg 1b \\
1b \gg 1a & 2b \gg 2a & 1a \gg 2b \\
0 \gg 1b & 0 \gg 2b & 1b \gg 2b \\
i < 0 & j < 0 & k < 1a,\, k < 1b \\
i < 1a,\, i < 1b & j < 2a,\, j < 2b & k < 2a,\, k < 2b
\end{array}
$$
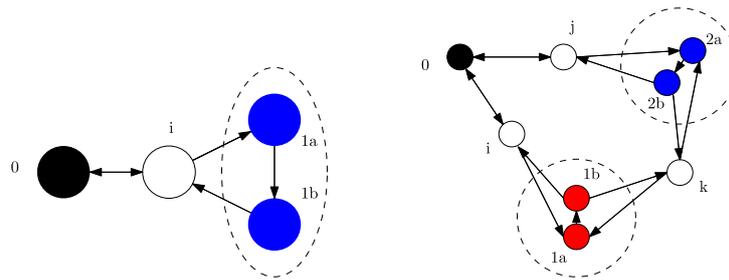
Figure 7: Graph $\mathcal{G}$ of Examples 3.9 and 3.11.

Thus, following [6] and the arguments of the previous example, on the time scale $O(K^{-1+2/\alpha})$, the population process converges to a jump process between the three states $\{0, 1b, 2b\}$ with positive jump rates. More precisely,

Theorem 3.12. *As $K \to \infty$, the following convergence holds*

$$(N_0^K, N_{1b}^K, N_{2b}^K)(tK^{-1+2/\alpha})/K \Rightarrow \bar{n}(X_t)\delta_{X_t}$$

for finite dimensional distributions, where X_t is a continuous time Markov chain on $\{0, 1b, 2b\}$ with transition rates:

$$\begin{aligned}
r_{0\to 1b} &= \frac{\bar{n}(0)b_0 b_i}{4|f_{i,0}|}\frac{f_{1a,0}}{b_{1a}}, & r_{1b\to 0} &= \frac{\bar{n}(1b)b_{1b}b_i}{4|f_{i,1b}|}\frac{f_{0,1b}}{b_0}\\
r_{0\to 2b} &= \frac{\bar{n}(0)b_0 b_j}{4|f_{j,0}|}\frac{f_{2a,0}}{b_{2a}}, & r_{2b\to 0} &= \frac{\bar{n}(2b)b_{2b}b_j}{4|f_{j,2b}|}\frac{f_{0,2b}}{b_0}\\
r_{1b\to 2b} &= \frac{\bar{n}(1b)b_{1b}b_k}{4|f_{k,1b}|}\frac{f_{2a,1b}}{b_{2a}}, & r_{2b\to 1b} &= \frac{\bar{n}(2b)b_{2b}b_k}{4|f_{k,2b}|}\frac{f_{1a,2b}}{b_{1a}}.
\end{aligned}$$

As above, the $1/4$ is the result of multiple choices for mutant traits at mutation events.

4 Proof of Theorem 2.7 and Proposition 2.8

This section is dedicated to the proofs of our main results. As they are technical and involve many stopping times, we begin with a rough outline of the strategy of the proof.

Throughout the proof, we define several stopping times to divide the times between invasions into sub-steps. Heuristically they correspond to the following events:

- σ_k^K, the time when the kth invasion has taken place and a new equilibrium is reached.
- $\theta_{k,m,C}^K$, the first time after σ_{k-1}^K when either the macroscopic traits stray too far from their equilibrium or at least one of the (formerly) microscopic traits becomes macroscopic (recall Definition 2.2)
- s_k^K, the first time after σ_{k-1}^K when a microscopic trait becomes almost macroscopic, i.e. reaches an order of $K^{1-\varepsilon_k}$.
- $t_{w,k}^K$, the first time after σ_{k-1}^K when trait w has a positive population size. ($t_{w,k}^K = \sigma_{k-1}^K$ for all traits that are alive at this time.)

The conditions satisfied by $m > 0$, $C > 0$ and ε_k will be precised later on. m and ε_k are typically small (see (4.27) and point (3) on page 19). The conditions satisfied by C will be specified in Section 4.4.

As in Proposition 2.8, $(\tau_\ell^K, \ell \geq 0)$ is the collection of both $(s_k^K, k \geq 0)$ and $(t_{w,k}^K, k \geq 0, w \in V)$. Figure 8 visualises the different stopping times for the case of one macroscopic and two microscopic traits.

The proof consists of five parts, corresponding respectively to Sections 4.2 to 4.6 below:
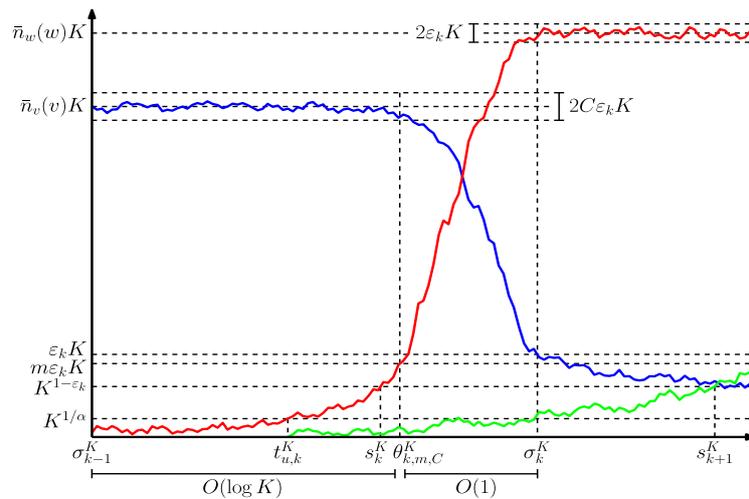
Figure 8: Schematic evolution of macroscopic trait v (blue) and microscopic traits $w = l_k^K$ (red) and u (green), where $d(w, u) = 1$, during the kth invasion step.

1. In the longest and most involved part of the proof, we study the growth dynamics of the different subpopulations in the time interval $[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$, making use of several results from [8], which are restated in the Appendix, and generalised when needed. Similar to [25], we prove lower and upper bounds for $\beta_w^K(t)$ via an induction, successively taking into account incoming mutants originating from traits of increasing distance to w. We prove that $\beta_w^K(t)$ follows the characterisation of $\beta_w(t)$ in Theorem 2.7 up to an error of order ε_k for large K. The proof shows in addition that rescaled macroscopic populations remain arbitrarily close to their Lotka-Volterra equilibrum provided K is taken large enough, which implies part (i) of Proposition 2.8.

2. We construct the sets M_ℓ^K and calculate the value of $\tau_\ell^K - \tau_{\ell-1}^K$, proving part (ii) of Proposition 2.8.

3. We prove that s_k^K and $\theta_{k,m,C}^K$ are equal up to an error η_k that goes to zero as $\varepsilon_k \to 0$ and conclude that s_k^K converges to s_k when $K \to \infty$.

4. We prove that the stopping time $\theta_{k,m,C}^K$ is triggered by a (formerly) microscopic trait reaching order K, and not by the macroscopic traits deviating from their equilibrium.

5. Knowing that we have non-vanishing population sizes at $\theta_{k,m,C}^K$, we finally consider the Lotka-Volterra phase involving $\mathbf{v}_{k-1}$ and the trait l_k^K that has newly reached order K, proving that the initial conditions for the next step, characterised in the definition of σ_k^K, are satisfied after a time of order 1. This concludes the proof of Theorem 2.7.

Recall the definitions provided in Theorem 2.7 and Proposition 2.8, and the definition of $\tilde{\mathbf{v}}$, the support of the mutation free Lotka-Volterra equilibrium associated to $\mathbf{v}$, on page 7.

Similarly as in [8], the strategy of the proof consists in performing an induction on successive phases k, during which the population sizes of the set of traits $\tilde{\mathbf{v}}_k$ are close to their equilibrium value and the population sizes of the set of traits $V \setminus \tilde{\mathbf{v}}_k$ are small with respect to K. To be more precise, we will introduce a sequence of stopping times $(\sigma_k^K \log K, k \in \mathbb{N})$ (see definition in (4.30)) satisfying the following conditions, as soon as $s_k < T$:

Assumption 4.1.

1. $\sigma_k^K \to s_k$ *in probability when K goes to infinity*

2. *For any $0 < \varepsilon_k < 1 \wedge \inf_{w \in \tilde{\mathbf{v}}_k} \bar{n}_w(\mathbf{v}_k)$, with high probability*

   (a) *For every $w \in \tilde{\mathbf{v}}_k$,*
   $$\frac{N_w^K(\sigma_k^K \log K)}{K} \in [\bar{n}_w(\mathbf{v}_k) - \varepsilon_k, \bar{n}_w(\mathbf{v}_k) + \varepsilon_k].$$

   (b) *For every $w \in \mathbf{v}_k \setminus \tilde{\mathbf{v}}_k$,*
   $$K^{1-\varepsilon_k} \leq N_w^K(\sigma_k^K \log K) \leq \varepsilon_k K.$$

   (c) *There exists $\bar{c}_k < \infty$ such that for every $w \notin \mathbf{v}_k$, either $N_w^K(\sigma_k^K \log K) = 0$ if $\beta_w(s_k) = 0$ or*
   $$0 < \beta_w(s_k) - \bar{c}_k\varepsilon_k < \frac{\log\left(1 + N_w^K(\sigma_k^K \log K)\right)}{\log K} = \beta_w^K(\sigma_k^K) < \beta_w(s_k) + \bar{c}_k\varepsilon_k < 1.$$

To be more precise, for $k \geq 1$, the time interval $[\sigma_{k-1}^K \log K, \sigma_k^K \log K]$ will be divided into two parts:

- a 'stochastic phase' $[\sigma_{k-1}^K \log K, \theta_{k,m,C}^K \log K]$ needed for the trait
$$l_k^K := \mathbf{v}_k^K \setminus \mathbf{v}_{k-1}^K$$
to reach a size of order K,

- a 'deterministic phase' $[\theta_{k,m,C}^K \log K, \sigma_k^K \log K]$ needed for the mutation free Lotka-Volterra system associated to $\tilde{\mathbf{v}}_{k-1}^K \cup l_k^K$ to reach a neighbourhood of its equilibrium.

Initialisation of the induction.

- $\underline{\sigma_0^K}$: By assumption,
$$\beta_w^K(0) \stackrel{K\to\infty}{\to} \left(1 - \frac{d(\mathbf{v}_0, w)}{\alpha}\right)_+.$$

Let us choose a small $\varepsilon_0 > 0$. Then from point (ii) of Lemma A.6, there exists a deterministic $T(\varepsilon_0) < \infty$ such that
$$\lim_{K\to\infty} \mathbb{P}\left(\|N^K(T(\varepsilon_0))/K - \bar{n}(\mathbf{v}_0)\|_\infty \leq \varepsilon_0\right) = 1.$$

Define $\sigma_0^K := T(\varepsilon_0)/\log K$. We can check that σ_0^K is a stopping time converging in probability to $s_0 = 0$ and satisfying Assumption 4.1. Moreover we know that the processes $(\beta_w^K, w \in V)$ vary on a time scale of order $\log K$ (see [7, 8] for instance). In particular, they do not vary during the time $T(\varepsilon_0)$ in the large K limit. This entails that σ_0^K satisfies Assumption 4.1.

- $\underline{\sigma_k^K, k \geq 1}$:
   Assume that $s_{k-1} < T_0$ and that $\sigma_{k-1}^K \log K$ is a stopping time satisfying Assumption 4.1. We will now construct σ_k^K.

4.1 Definitions and first properties

Let us introduce a small $\varepsilon_k > 0$ as well as a stopping time $\theta^K_{k,m,C} \log K$ via

$$\theta^K_{k,m,C} := \inf \left\{ t \geq \sigma^K_{k-1}, \exists w \in \mathbf{v}_{k-1}, \left| \frac{N^K_w(t \log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1}) \right| \geq C\varepsilon_k \right.$$

$$\left. \text{or} \sum_{w \notin \mathbf{v}_{k-1}} N^K_w(t \log K) \geq m\varepsilon_k K \right\}. \quad (4.1)$$

Recall that the conditions satisfied by $m > 0$ and $C > 0$ will be precised later on.

We will now finely study the population dynamics on the time interval $[\sigma^K_{k-1} \log K, (\theta^K_{k,m,C} \wedge T) \log K]$. To this aim, we will couple the subpopulations of individuals with a given trait with branching processes with immigration and use results on these processes derived in [8] and recalled (and generalized when needed) in the Appendices. The main difficulty of this step comes from the fact that as we allow for any finite graph of mutations, the immigration rate for a particular subpopulation may vary a lot on the time interval $[\sigma^K_{k-1} \log K, (\theta^K_{k,m,C} \wedge T) \log K]$. This is why we introduced in Proposition 2.8 the sequence of times $(\tau_\ell, \ell \in \mathbb{N})$, which corresponds to the times when mutants of a new type arise or a formerly microscopic trait becomes of order K.

Notice that although we make extensive use of the techniques and results developed in [8], the authors of this paper considered a specific graph embedded in $\mathbb{Z}$, and their proof structure, in particular inductions, relies on their graph structure. The current inductions are more involved and more in the proof spirit of [25].

To begin with, let us recall the rates of the different events for the population N^K_w, with $w \in V$, at time t:

- Reproductions without mutation:

$$\mathfrak{b}_w(t) := b_w(1 - K^{-1/\alpha})N^K_w(t). \quad (4.2)$$

- Death:

$$\mathfrak{d}_w(t) := \left(d_w + \sum_{x \in V} \frac{c_{w,x}}{K} N^K_x(t) \right) N^K_w(t). \quad (4.3)$$

- Reproductions with mutations towards the trait w:

$$\mathfrak{bm}_w(t) := K^{-1/\alpha} \sum_{x \in V, d(x,w)=1} b_x m(x,w) N^K_x(t). \quad (4.4)$$

Notice that for K large enough, as σ^K_{k-1} satisfies Assumption 4.1 and by definition of $\theta^K_{k,m,C}$, on the time interval $[\sigma^K_{k-1} \log K, (\theta^K_{k,m,C} \wedge T) \log K]$, we have

$$b(w,k,-)N^K_w(t) \leq \mathfrak{b}_w(t) \leq b(w,k,+)N^K_w(t), \quad (4.5)$$

$$d(w,\tilde{\mathbf{v}}_{k-1},k,+)N^K_w(t) \leq \mathfrak{d}_w(t) \leq d(w,\tilde{\mathbf{v}}_{k-1},k,-)N^K_w(t), \quad (4.6)$$

$$f(w,\tilde{\mathbf{v}}_{k-1},k,-) \leq f_{w,\tilde{\mathbf{v}}_{k-1}} \leq f(w,\tilde{\mathbf{v}}_{k-1},k,+), \quad (4.7)$$

where we have introduced the following notations, for any $w \in V$ and $* \in \{-,+\}$,

$$b(w,k,-) := (1 - \varepsilon_k)b_w, \quad b(w,k,+) := b_w,$$

$$d(w,\tilde{\mathbf{v}}_{k-1},k,-) := d_w + \sum_{x \in \tilde{\mathbf{v}}_{k-1}} c_{w,x} \bar{n}_x(\mathbf{v}_{k-1}) + \left(\sum_{x \in V} c_{w,x} \right)(m+C)\varepsilon_k,$$

$$d(w,\tilde{\mathbf{v}}_{k-1},k,+) := d_w + \sum_{x \in \tilde{\mathbf{v}}_{k-1}} c_{w,x} \bar{n}_x(\mathbf{v}_{k-1}) - \left(\sum_{x \in V} c_{w,x} \right) C\varepsilon_k,$$

$$f(w,\tilde{\mathbf{v}}_{k-1},k,*) := b(w,k,*) - d(w,\tilde{\mathbf{v}}_{k-1},k,*). \quad (4.8)$$

Hence the rate of reproduction without mutation, as well as the death rate do not vary significantly during the time interval $[\sigma_{k-1}^K \log K, (\theta_{k,m,C}^K \wedge T) \log K]$. The difficulty comes from the rate of mutations towards a given trait, which depends on the population sizes of its neighbours in the graph $\mathcal{G}$, which themselves depend on the population sizes of their neighbours and so on.

Let us introduce the times τ_ℓ^K and the sets M_ℓ^K, which correspond respectively to the times of invasion or appearance of new mutants (and will be the time steps of the algorithm to be described shortly later) and to the sets of living traits in the time interval $(\tau_\ell^K, \tau_{\ell+1}^K]$. To be more precise,

Definition 4.2. *Let $s_k^K := \inf\{t \geq \sigma_{k-1}^K : \exists w \in V \setminus \mathbf{v}_{k-1}^K, \beta_w^K(t) > 1 - \varepsilon_k\}$, and*

$$
t_{w,k}^K := \begin{cases} \inf\{t \geq \sigma_{k-1}^K : \exists\, u \in V : d(u,w) = 1, \beta_u^K(t) = \frac{1}{\alpha}\} & \text{if } \beta_w^K(\sigma_{k-1}^K) = 0 \\ \sigma_{k-1}^K & \text{else}, \end{cases}
$$

The sequences $(\tau_\ell^K, \ell \geq 0)$ and $(M_\ell^K, \ell \geq 0)$ are defined as follows:
$\tau_0^K = \sigma_0^K$ and, for $\sigma_{k-1}^K \leq \tau_{\ell-1}^K < s_k^K$,

$$
\tau_\ell^K = s_k^K \wedge \min\{t_{w,k}^K : w \in V, t_{w,k}^K > \tau_{\ell-1}^K\},
$$

that is to say the minimum between the time when a previously microscopic population becomes (almost) macroscopic, and the time of appearance of a new mutant. From the definition of the sequence $(\tau_\ell^K, \ell \geq 0)$ we can now define the sequence of sets of living traits $(M_\ell^K, \ell \geq 0)$ via

$$
\begin{aligned}
M_\ell^K &= \{w \in V : \beta_w^K(\tau_\ell^K) > 0 \text{ or } \tau_\ell^K = t_{w,k}^K\} \\
&= \left(M_{\ell-1}^K \setminus \{w \in V : \beta_w^K(\tau_\ell^K) = 0\}\right) \cup \{w \in V : \tau_\ell^K = t_{w,k}^K\}.
\end{aligned}
$$

4.2 Dynamics of the process on $[\tau_{\ell-1}^K \log K, \tau_\ell^K \log K]$

In this section, we study the growth dynamics of the different subpopulations in the time interval $[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$. We prove that $\beta_w^K(t)$ follows the characterisation of $\beta_w(t)$ in Theorem 2.7 up to an error of order ε_k for large K.

We will first prove that there exists a finite and positive constant C_ℓ such that with high probability, for every $w \in M_{\ell-1}^K$ and $t \in [\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$,

$$
\max_{u \in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+
$$

$$
\leq \beta_w^K(t) \leq
$$

$$
\max_{u \in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + C_\ell \varepsilon_k + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, +)\right]_+ \quad (4.9)
$$

Let us thus take t in $[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$. To obtain the lower bound in (4.9), we show by induction that, for any $n \geq 0$ and with high probability,

$$
\beta_w^K(t) \geq \max_{\substack{u \in M_{\ell-1}^K : \\ d(u,w) \leq n}} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+ \quad (4.10)
$$

Induction lower bound: • $\underline{n = 0:}$ let $w \in M_{\ell-1}^K$. From (4.5) and (4.6), we see that we can couple N_w^K with a process Z^K with law $BP_K\left(b(w,k,-), d(w, \tilde{\mathbf{v}}_{k-1}, k, -), \beta_w^K(\tau_{\ell-1}^K)\right)$ (see the definition of BP_K in Section A.1) in such a way that

$$
N_w^K(t \log K) \geq Z^K((t - \tau_{\ell-1}^K) \log K).
$$

Hence, from Corollary A.4, we obtain that with high probability,

$$\beta_w^K(t) \geq \left[\beta_w^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K) f(w, \tilde{\mathbf{v}}_{k-1}, k, -) \right]_+ .$$

Remark 4.3. Notice that the application of Lemma A.1 (which has been derived in [8]) would require $\beta_w^K(\tau_{\ell-1}^K) > 0$ and that this condition may not be satisfied for one of the $w \in \tilde{\mathbf{v}}_{k-1}$ (the trait which becomes macroscopic at time $\tau_{\ell-1}^K \log K$). However, the population of individuals w grows exponentially due to the mutations coming from another trait and there exists a finite c such that, for small $\delta > 0$, $N_w^K((\tau_{\ell-1}^K + \delta) \log K) \geq K^{c\delta}$. We could thus apply Lemma A.1 at this time, and later on let δ go to 0 to get the result. This is in words the statement of Corollary A.4.

$\bullet \underline{n \to n+1}$: Let $w, u', u \in M_{\ell-1}^K$ such that $d(u', w) = 1$ and $d(u, u') \leq n$. From now on, we will use the notation BPI_K, which is defined in Section A.2. From (4.5), (4.6), and (4.7), by looking only at the immigration coming from u', we see that we can couple N_w^K with a process Z^K with law

$$BPI_K \left(b(w, k, -), d(w, \tilde{\mathbf{v}}_{k-1}, k, -), f(u', \tilde{\mathbf{v}}_{k-1}, k, -), \beta_{u'}^K(\tau_{\ell-1}^K) - \frac{1}{\alpha}, \beta_w^K(\tau_{\ell-1}^K) \right)$$

in such a way that

$$N_w^K(t \log K) \geq Z^K((t - \tau_{\ell-1}^K) \log K).$$

By the induction hypothesis, with high probability,

$$\beta_{u'}^K(t) \geq \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u, u')}{\alpha} + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, -) \right]_+ , \qquad (4.11)$$

which implies that we can couple Z^K with a process Y^K with law

$$BPI_K \left(b(w, k, -), d(w, \tilde{\mathbf{v}}_{k-1}, k, -), f(u, \tilde{\mathbf{v}}_{k-1}, k, -), \beta_u^K(\tau_{\ell-1}^K) - \frac{d(u, u') + 1}{\alpha}, \beta_w^K(\tau_{\ell-1}^K) \right)$$

in such a way that

$$Z^K((t - \tau_{\ell-1}^K) \log K) \geq Y^K((t - \tau_{\ell-1}^K) \log K).$$

Hence, from Corollary A.4, even if we have to work in a time interval $[\tau_{\ell-1}^K + \delta, T]$, for a small positive δ, in the spirit of Remark 4.3, as $w \in M_{\ell-1}^K$ we obtain that with high probability,

$$\begin{aligned}
\beta_w^K(t) &\geq \left[\beta_w^K(\tau_{\ell-1}^K) \vee \left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u, u') + 1}{\alpha} \right) + (t - \tau_{\ell-1}^K) f(w, \tilde{\mathbf{v}}_{k-1}, k, -) \right]_+ \\
&\qquad \vee \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u, u') + 1}{\alpha} + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, -) \right]_+ \\
&\geq \left[\beta_w^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K) f(w, \tilde{\mathbf{v}}_{k-1}, k, -) \right]_+ \\
&\qquad \vee \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u, u') + 1}{\alpha} + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, -) \right]_+ .
\end{aligned}$$

As this is true for any u' such that $d(u', w) = 1$ and as the above bound is a decreasing function of $d(u, u')$, by taking the supremum over such u' we obtain

$$\beta_w^K(t) \geq \left[\beta_w^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K)f(w, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+$$
$$\vee \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K)f(u, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+.$$

Thus, with high probability,

$$\beta_w^K(t) \geq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n+1}} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K)f(u, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+,$$

which ends the induction for the lower bound.

Let us now proceed to the induction for the upper bound. We again take t in $[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$ and we will show that for any $n \in \mathbb{N}$ there exists a finite constant $C_{n,\ell}$ such that with high probability,

$$\beta_w^K(t) \leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n}} \left[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + C_{n,\ell}\varepsilon_k\right) \vee \left(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k\right)\right.$$
$$\left. + (t - \tau_{\ell-1}^K)f(u, \tilde{v}_{k-1}, k, +)\right]_+ \vee \left(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k\right).$$

Induction upper bound:

Throughout the induction for the upper bound, we will several times make use of the fact that we can bound the total immigration to one trait (which is the sum of the mutants coming from its neighbours) by the number of neighbours times the largest incoming mutation. More precisely, if I_w is the number of incoming neighbours of w,

$$\sum_{\substack{u \in V: \\ d(u,w)=1}} K^{\beta_u^K(t)} \leq I_w \max_{\substack{u \in V: \\ d(u,w)=1}} K^{\beta_u^K(t)} = \max_{\substack{u \in V: \\ d(u,w)=1}} K^{(\log I_w / \log K) + \beta_u^K(t)}. \tag{4.12}$$

Since the trait space is finite, for K large enough, we can assume that $\max_{w \in V} \log I_w / \log K \leq \varepsilon_k$.

• $\underline{n = 0}$: We observe that for K large enough $\beta_u^K(t) \leq 1 + \varepsilon_k$ for every $u \in M_{\ell-1}^K$ (see Corollary A.7).

From (4.5), (4.6), and (4.7), we see that we can couple N_w^K with a process Z^K with law $BPI_K\left(b(w,k,+), d(w, \tilde{\mathbf{v}}_{k-1}, k, +), 0, 1 - \frac{1}{\alpha} + 2\varepsilon_k, \beta_w^K(\tau_{\ell-1}^K)\right)$ in such a way that

$$N_w^K(t \log K) \leq Z^K((t - \tau_{\ell-1}^K) \log K).$$

Hence from Corollary A.4, even if we have to work in a time interval $[\tau_{\ell-1}^K + \delta, T]$, for a small positive δ, in the spirit of Remark 4.3, as $w \in M_{\ell-1}^K$ we obtain that with high probability,

$$\beta_w^K(t) \leq \left[\beta_w^K(\tau_{\ell-1}^K) \vee \left(1 - \frac{1}{\alpha} + 2\varepsilon_k\right) + (t - \tau_{\ell-1}^K)f(w, \tilde{\mathbf{v}}_{k-1}, k, +)\right]_+ \vee \left(1 - \frac{1}{\alpha} + 2\varepsilon_k\right).$$

• $\underline{n \to n+1}$: For $w, u' \in M_{\ell-1}^K$ such that $d(u',w) = 1$, by the induction hypothesis we have the existence of a finite constant $C_{n,\ell}$ such that, with high probability,

$$\beta_{u'}^K(t) \leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,u') \leq n}} \left[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')}{\alpha} + C_{n,\ell}\varepsilon_k \right) \vee \left(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k \right) \right.$$

$$\left. + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, +) \right]_+ \vee \left(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k \right).$$

From (4.5), (4.6), and (4.7), by looking at the maximal immigration coming from a neighbouring u' and adding another ε_k in the spirit of (4.12), we thus see that we can couple N_w^K with multiple processes $Z^{K,u,u'}$ and Z^K with respective laws

$$BPI_K \Big(b(w,k,+), d(w, \tilde{\mathbf{v}}_{k-1}, k, +), f(u, \tilde{\mathbf{v}}_{k-1}, k, +),$$

$$\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right), \beta_w^K(\tau_{\ell-1}^K) \Big)$$

and

$$BPI_K \left(b(w,k,+), d(w, \tilde{\mathbf{v}}_{k-1}, k, +), 0, 1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k, \beta_w^K(\tau_{\ell-1}^K) \right)$$

in such a way that

$$N_w^K(t \log K) \leq \max_{\substack{u' \in M_{\ell-1}^K: \\ d(u',w)=1}} \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,u') \leq n}} Z^{K,u,u'}((t - \tau_{\ell-1}^K)\log K) \vee Z^K((t - \tau_{\ell-1}^K)\log K).$$

Hence from Corollary A.4, even if we have to work in a time interval $[\tau_{\ell-1}^K + \delta, T]$, for a small positive δ, in the spirit of Remark 4.3, as $w \in M_{\ell-1}^K$ we obtain that with high probability,

$$\beta_w^K(t) \leq \max_{\substack{u' \in M_{\ell-1}^K: \\ d(u',w)=1}} \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,u') \leq n}} \left\{ \left[\beta_w^K(\tau_{\ell-1}^K) \vee \left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) \right. \right.$$

(4.13)

$$+ (t - \tau_{\ell-1}^K) f(w, \tilde{\mathbf{v}}_{k-1}, k, +) \Big]_+$$

$$\vee \left[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}) f(u, \tilde{\mathbf{v}}_{k-1}, k, +) \right]_+$$

(4.14)

$$\vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) \Big\}$$

$$\leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n+1}} \left\{ \left[\beta_w^K(\tau_{\ell-1}^K) \vee \left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) \right. \right.$$

(4.15)

$$+ (t - \tau_{\ell-1}^K) f(w, \tilde{\mathbf{v}}_{k-1}, k, +) \Big]_+$$

$$\vee \left[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}) f(u, \tilde{\mathbf{v}}_{k-1}, k, +) \right]_+$$

(4.16)

$$\vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) \Big\}.$$

(4.17)

In order to simplify the right hand side of the previous inequality, we will show that for any $\ell \in \mathbb{N}$ there exists a finite and positive constant C_ℓ such that for any $(u,w) \in V^2$, with high probability

$$\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} \leq \beta_w^K(\tau_{\ell-1}^K) + C_\ell \varepsilon_k. \qquad (4.18)$$

Combining (4.17) and (4.18) yields that with high probability,

$$
\begin{aligned}
\beta_w^K(t) \leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n+1}} &\left\{ \left[\left(\beta_w^K(\tau_{\ell-1}^K) + (C_{n,\ell} + 1 + C_\ell)\varepsilon_k \right) \right. \right. \\
&\left. \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}^K)f(w, \tilde{\mathbf{v}}_{k-1}, k, +) \right]_+ \\
&\vee \left[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell} + 1)\varepsilon_k \right) \right. \\
&\left. \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1})f(u, \tilde{\mathbf{v}}_{k-1}, k, +) \right]_+ \\
&\left. \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) \right\} \\
\leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n+1}} &\left[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell} + 1 + C_\ell)\varepsilon_k \right) \right. \\
&\left. \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1})f(u, \tilde{\mathbf{v}}_{k-1}, k, +) \right]_+ \\
&\vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right),
\end{aligned}
$$

which ends the induction for the upper bound.

Let us now derive inequality (4.18). It is obtained by an induction on ℓ. If $\ell = 1$, by (2.6) and the triangle inequality,

$$
\begin{aligned}
\lim_{K \to \infty} \beta_u^K(0) - \frac{d(u,w)}{\alpha} &= \left[1 - \frac{d(\mathbf{v}_0, u)}{\alpha} \right]_+ - \frac{d(u,w)}{\alpha} \\
&\leq \left[1 - \frac{d(\mathbf{v}_0, u)}{\alpha} - \frac{d(u,w)}{\alpha} \right]_+ \leq \left[1 - \frac{d(\mathbf{v}_0, w)}{\alpha} \right]_+ = \lim_{K \to \infty} \beta_w^K(0).
\end{aligned}
$$

As the convergence is in probability, it means that for K large enough, there exists a finite $C_{u,w}$ such that with a probability larger than $1 - \varepsilon_k$,

$$
\beta_u^K(0) - \frac{d(u,w)}{\alpha} \leq \beta_w^K(0) + C_{u,w}\varepsilon_k. \tag{4.19}
$$

As there are only finitely many traits, $\sup_{u,w \in V} C_{u,w} < \infty$. Moreover, as ε_k can be chosen as small as we want and as we want to prove a convergence in probability, we may focus on the event where inequality (4.19) is satisfied. We will do that later on without mentioning it again for the sake of readability.

Now assume that (4.18) is true for $\ell - 1 \in \mathbb{N}$. Let us first prove that it still holds for ℓ.

From the previous step on the time interval $[\tau_{\ell-2}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K]$, we know that if $\tau_{\ell-1}^K \leq T \wedge \theta_{k,m,C}^K$, for any $w \in V$ and K large enough,

$$
\max_{u \in M_{\ell-2}^K} \left[\beta_u^K(\tau_{\ell-2}^K) - \frac{d(u,w)}{\alpha} + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(u, \tilde{\mathbf{v}}_{k-2}, k, -) \right]_+ \leq \beta_w^K(\tau_{\ell-1}^K).
$$

Now let us take $u \in V$. We also deduce from the previous step that for K large enough

$$
\beta_u^K(\tau_{\ell-1}^K) \leq \max_{u' \in M_{\ell-2}^K} \left[\beta_{u'}^K(\tau_{\ell-2}^K) - \frac{d(u',u)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(u', \tilde{\mathbf{v}}_{k-2}, k, +) \right]_+.
$$

In particular there exists $\tilde{u} \in V$ such that $d(\tilde{u}, u) \leq \lfloor \alpha \rfloor$ and for K large enough

$$
\beta_u^K(\tau_{\ell-1}^K) \leq \left[\beta_{\tilde{u}}^K(\tau_{\ell-2}^K) - \frac{d(\tilde{u}, u)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(\tilde{u}, \tilde{\mathbf{v}}_{k-2}, k, +) \right]_+.
$$

Thus, for K large enough,

$$
\begin{aligned}
\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} &\le \left[\beta_{\tilde{u}}^K(\tau_{\ell-2}^K) - \frac{d(\tilde{u},u)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(\tilde{u},\tilde{\mathbf{v}}_{k-2},k,+) \right]_+ - \frac{d(u,w)}{\alpha} \\
&\le \left[\beta_{\tilde{u}}^K(\tau_{\ell-2}^K) - \frac{d(\tilde{u},u)+d(u,w)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(\tilde{u},\tilde{\mathbf{v}}_{k-2},k,+) \right]_+ \\
&\le \left[\beta_{\tilde{u}}^K(\tau_{\ell-2}^K) - \frac{d(\tilde{u},w)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(\tilde{u},\tilde{\mathbf{v}}_{k-2},k,+) \right]_+ \\
&\le \left[\beta_{\tilde{u}}^K(\tau_{\ell-2}^K) - \frac{d(\tilde{u},w)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(\tilde{u},\tilde{\mathbf{v}}_{k-2},k,-) \right]_+ + C\varepsilon_k \\
&\le \max_{\tilde{u}\in M_{\ell-2}^K} \left[\beta_{\tilde{u}}^K(\tau_{\ell-2}^K) - \frac{d(\tilde{u},w)}{\alpha} + (\tau_{\ell-1}^K - \tau_{\ell-2}^K)f(\tilde{u},\tilde{\mathbf{v}}_{k-2},k,-) \right]_+ + (C_{\ell-1}+C)\varepsilon_k \\
&\le \beta_w^K(\tau_{\ell-1}^K) + (C_{\ell-1}+C)\varepsilon_k,
\end{aligned}
$$

where we used (4.8), (4.10), the bound $\tau_{\ell-1}^K - \tau_{\ell-2}^K \le T$, and

$$
C := \max_{\tilde{u}\in M_{\ell-2}^K} \left(b_{\tilde{u}} + (m+6)\left(\sum_{x\in V} c_{\tilde{u},x} \right) \right) T.
$$

This entails (4.18).

To conclude the proof of (4.9), we just need to notice that for $n > \lfloor \alpha \rfloor$, if ε_k is small enough,

$$
1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k < 0.
$$

As $\tau_\ell^K - \tau_{\ell-1}^K \le T$, Equation (4.9) tells us that, with an error of order ε_k which is as small as we want, with high probability, the growth of traits $w \in M_{\ell-1}^K$ follows, for $t \in [\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$

$$
\beta_w^K(t) \cong \max_{u\in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K)f_{u,\mathbf{v}_{k-1}} \right]_+ .
$$

To avoid repetition, we will write $\cong$ in the sequel to indicate approximations with high probability, with an error of order ε_k.

4.3 Value of τ_ℓ^K and construction of M_ℓ^K

In this section, we construct the sets M_ℓ^K and calculate the value of $\tau_\ell^K - \tau_{\ell-1}^K$, proving part (ii) of Proposition 2.8.

Let us assume for the moment (it will be proven in Section 4.4) that the following holds with high probability:

$$
[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K] = [\tau_{\ell-1}^K, \tau_\ell^K]. \tag{4.20}
$$

Our aim now is to find the duration $\tau_\ell^K - \tau_{\ell-1}^K$ and to construct the set M_ℓ^K knowing the set $M_{\ell-1}^K$.

To reach τ_ℓ^K, two events are possible. Either one living non resident trait reaches a size of order K, or a new mutant appears.

Let us consider the first type of event. In fact, we have to be more precise on the time when a new trait has a size which reaches order K, this is why we defined s_k^K as the time when one trait has a size which reaches order $K^{1-\varepsilon_k}$. Notice that we may choose ε_k small enough to be sure that it corresponds to the trait whose exponent reaches 1 at

time s_k in the deterministic sequence $(s_j, j \in \mathbb{N})$ defined in Theorem 2.7. (if there exist two such traits, condition $(iv)(a)$ is fulfilled and T_0 is set to s_k). Notice that if $f_{u,\mathbf{v}_{k-1}} < 0$, for any $w \in V$,

$$t \mapsto \beta_u^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K)f_{u,\mathbf{v}_{k-1}} - \frac{d(u,w)}{\alpha} \tag{4.21}$$

is decreasing and thus will not reach $1 - \varepsilon_k$ if it is smaller than this value at time $\tau_{\ell-1}^K$. Hence if we denote by u_0 the element of $M_{\ell-1}^K$ such that $\beta_{u_0}^K(\tau_\ell^K) = 1 - \varepsilon_k$, we get

$$1 - \varepsilon_k = \beta_{u_0}^K(\tau_\ell^K) \cong \max_{\substack{u \in M_{\ell-1}^K \\ f_{u,\mathbf{v}_{k-1}} > 0}} \left[\beta_u^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K)f_{u,\mathbf{v}_{k-1}} - \frac{d(u,u_0)}{\alpha} \right].$$

Now assume by contradiction that there is $u_1 \neq u_0 \in M_{\ell-1}^K$ such that:

$$1 - \varepsilon_k = \beta_{u_0}^K(\tau_\ell^K) \cong \beta_{u_1}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f_{u_1,\mathbf{v}_{k-1}} - \frac{d(u_1,u_0)}{\alpha}.$$

This implies

$$\beta_{u_1}^K(\tau_\ell^K) \geq \beta_{u_1}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f_{u_1,\mathbf{v}_{k-1}} > 1,$$

as soon as $\varepsilon_k < 1/\alpha$, which yields a contradiction. This implies that if there exists $u_0 \in M_{\ell-1}^K$ such that $\beta_{u_0}^K(\tau_\ell^K) = 1 - \varepsilon_k$, then

$$\beta_{u_0}^K(\tau_\ell^K) \cong \beta_{u_0}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f_{u_0,\mathbf{v}_{k-1}}$$

and with high probability, the value of $\tau_\ell^K - \tau_{\ell-1}^K$ satisfies,

$$\tau_\ell^K - \tau_{\ell-1}^K \cong \min_{\substack{w \in M_{\ell-1}^K: \\ f_{w,\mathbf{v}_{k-1}} > 0}} \frac{1 - \beta_w^K(\tau_{\ell-1}^K)}{f_{w,\mathbf{v}_{k-1}}}. \tag{4.22}$$

Let us now consider the second type of event, that is to say that there exist $u_0 \notin M_{\ell-1}^K$ and $u_1 \in M_{\ell-1}^K$ such that $d(u_1,u_0) = 1$ and $\beta_{u_1}^K(\tau_\ell^K) = 1/\alpha$. Notice again that if $f_{u,\mathbf{v}_{k-1}} < 0$, the function defined in (4.21) is decreasing and thus will not reach $1/\alpha$ if it is smaller than this value at time $\tau_{\ell-1}^K$.

By definition we have

$$\frac{1}{\alpha} = \beta_{u_1}^K(\tau_\ell^K) \cong \max_{u \in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f_{u,\mathbf{v}_{k-1}} - \frac{d(u,u_1)}{\alpha} \right].$$

Denote by $u_2 \in M_{\ell-1}^K$ the trait realizing the maximum in the previous equation, that is to say

$$\frac{1}{\alpha} \cong \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f_{u_2,\mathbf{v}_{k-1}} - \frac{d(u_2,u_1)}{\alpha}.$$

This equality can be rewritten as

$$\beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f_{u_2,\mathbf{v}_{k-1}} \cong \frac{d(u_2,u_1) + 1}{\alpha}.$$

Let us now make a *reductio ad absurdum* to prove that $d(u_2,u_1) + 1 = d(u_2,u_0)$. Let us thus assume that

$$d(u_2,u_1) + 1 > d(u_2,u_0) \Leftrightarrow d(u_2,u_1) \geq d(u_2,u_0), \tag{4.23}$$

and take u_1' such that

$$d(u_2,u_1') + 1 = d(u_2,u_0).$$

Let us first assume (we will prove it later) that $u_1' \in M_{\ell-1}^K$. In this case, using the proof for the lower bound, we obtain that with high probability

$$
\begin{aligned}
\beta_{u_1'}^K(\tau_\ell^K) &\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f(u_2, \tilde{\mathbf{v}}_{k-1}, k, -) - \frac{d(u_2, u_1')}{\alpha} \\
&\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f(u_2, \tilde{\mathbf{v}}_{k-1}, k, -) - \frac{d(u_2, u_1) - 1}{\alpha} \cong \frac{2}{\alpha}.
\end{aligned}
$$

As $d(u_1', u_0) = 1$, this means that u_0 becomes a living trait before the time τ_ℓ^K, which is in contradiction with the definition of τ_ℓ^K.

Let us now assume that $u_1' \notin M_{l-1}^K$ and consider a sequence of vertices $v_0 = u_2, v_1, ..., v_{d(u_2, u_1')} = u_1'$ such that $d(u_2, v_i) = i$ and $d(v_i, u_1') = d(u_2, u_1') - i$. Let

$$
i_0 := \max\{0 \leq i \leq d(u_2, u_1') - 1, v_i \in M_{\ell-1}^K\}.
$$

Then

$$
d(u_2, v_{i_0}) \leq d(u_2, u_1') - 1 \leq d(u_2, u_1) - 2,
$$

and with high probability

$$
\begin{aligned}
\beta_{v_{i_0}}^K(\tau_\ell^K) &\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f(u_2, \tilde{\mathbf{v}}_{k-1}, k, -) - \frac{d(u_2, v_{i_0})}{\alpha} \\
&\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f(u_2, \tilde{\mathbf{v}}_{k-1}, k, -) - \frac{d(u_2, u_1) - 2}{\alpha} \cong \frac{3}{\alpha},
\end{aligned}
$$

and thus v_{i_0+1} becomes a living trait before the time τ_ℓ^K, which again is in contradiction with the definition of τ_ℓ^K. We thus obtain a contradiction and deduce that (4.23) is not satisfied. We conclude that

$$
\beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K)f_{u_2, \mathbf{v}_{k-1}} \cong \frac{d(u_2, u_1) + 1}{\alpha} = \frac{d(u_2, u_0)}{\alpha}.
$$

Hence, when τ_ℓ^K corresponds to the arrival of a new mutant,

$$
\tau_\ell^K - \tau_{\ell-1}^K \cong \min_{\substack{w \in M_{\ell-1}^K \\ f_{w, \mathbf{v}_{\ell-1}} > 0}} \frac{\frac{d(w, V \setminus M_{\ell-1}^K)}{\alpha} - \beta_w^K(\tau_{\ell-1}^K)}{f_{w, \mathbf{v}_{k-1}}}. \tag{4.24}
$$

Combining (4.22) and (4.24), we finally obtain:

$$
\tau_\ell^K - \tau_{\ell-1}^K \cong \min_{\substack{w \in M_{\ell-1}^K: \\ f_{w, \mathbf{v}_{k-1}} > 0}} \frac{\left(1 \wedge \frac{d(w, V \setminus M_{\ell-1}^K)}{\alpha}\right) - \beta_w^K(\tau_{\ell-1}^K)}{f_{w, \mathbf{v}_{k-1}}}.
$$

To obtain M_ℓ^K from $M_{\ell-1}^K$, we suppress the traits $w \in M_{\ell-1}^K$ such that $\beta_w^K(\tau_\ell^K) = 0$ (if condition $(iv)(c)$ is not satisfied, otherwise T_0 is set to s_k) and if $\tau_\ell \neq s_k$, we add the traits which are at distance 1 from the $w \in V$ satisfying

$$
w \in \arg\min_{\substack{w \in M_{\ell-1}^K: \\ f_{w, \mathbf{v}_{k-1}} > 0}} \frac{\left(1 \wedge \frac{d(w, V \setminus M_{\ell-1}^K)}{\alpha}\right) - \beta_w^K(\tau_{\ell-1}^K)}{f_{w, \mathbf{v}_{k-1}}}.
$$

4.4 Value of $\theta^K_{k,m,C}$ and convergence of s^K_k to s_k

In this section, we prove that s^K_k and $\theta^K_{k,m,C}$ are equal up to an error η_k that goes to zero as $\varepsilon_k \to 0$ and conclude that s^K_k converges to s_k when $K \to \infty$.

Recall the definition of $\theta^K_{k,m,C}$ in (4.1). We thus have constructed, on the time interval $[(\sigma^K_{k-1} \wedge T) \log K, (s^K_k \wedge \theta^K_{k,m,C} \wedge T) \log K]$, the times $(\tau^K_\ell, \ell \in \mathbb{N})$ and the sets $(M^K_\ell, \ell \in \mathbb{N})$ of living traits between times τ^K_ℓ and $\tau^K_{\ell+1}$. We will now study the dynamics of the process on the time interval $[(\sigma^K_{k-1} \wedge T) \log K, (\sigma^K_k \wedge \theta^K_{k,m,C} \wedge T) \log K]$ (σ^K_k to be defined later in order to satisfy Assumption 4.1). Recall that l^K_k is the trait $w \in V$ such that $\beta^K_w(s^K_k) = 1 - \varepsilon_k$ and introduce

$$\eta_k := 2\varepsilon_k / \left(f_{l^K_k, \mathbf{v}_{k-1}} - \left(b_{l^K_k} + \left(\sum_{x \in V} c_{l^K_k, x} \right) (C + m) \right) \varepsilon_k \right).$$

We will first prove that

$$\lim_{K \to \infty} \mathbb{P} \left(s^K_k \leq \theta^K_{k,m,C} \leq s^K_k + \eta_k \Big| s^K_k < T \right) = 1. \tag{4.25}$$

The first step consists in showing that

$$\lim_{K \to \infty} \mathbb{P} \left(\theta^K_{k,m,C} < s^K_k \Big| s^K_k < T \right) = 0. \tag{4.26}$$

By definition of s^K_k, we have

$$\sup_{w \in V \smallsetminus \mathbf{v}_{k-1}} \sup_{\sigma^K_{k-1} \leq t \leq s^K_k} \beta^K_w(t) \leq 1 - \varepsilon_k.$$

Moreover, applying Lemma A.6 to $\mathbf{v}_{k-1}$ we obtain that

$$\lim_{K \to \infty} \mathbb{P} \left(\forall t \in [\sigma^K_{k-1}, s^K_k], \sup_{w \in \mathbf{v}_{k-1}} \left| \frac{N^K_w(t \log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1}) \right| \leq C\varepsilon_k \Big| s^K_k < T \right) = 1.$$

As a consequence, (4.26) holds true. Notice that the value of C in the definition of $\theta^K_{k,m,C}$ in (4.1) is a consequence of the previous limit. The constant C is the one needed for Lemma A.6 to hold, and thus depends on the parameters of the process.

Now assume by contradiction that

$$s^K_k + \eta_k \leq \theta^K_{k,m,C} < T.$$

Then on the time interval $[s^K_k, s^K_k + \eta_k]$, by definition of $\theta^K_{k,m,C}$, the l^K_k population has a growth rate bounded from below by

$$f_{l^K_k, \mathbf{v}_{k-1}} - \left(b_{l^K_k} + \left(\sum_{x \in V} c_{l^K_k, x} \right) (C + m) \right) \varepsilon_k.$$

Hence by coupling, with high probability,

$$\beta^K_{l^K_k}(s^K_k + \eta_k) \geq 1 - \varepsilon_k + \left(f_{l^K_k, \mathbf{v}_{k-1}} - \left(b_{l^K_k} + \left(\sum_{x \in V} c_{l^K_k, x} \right) (C + m) \right) \varepsilon_k \right) \eta_k = 1 + \varepsilon_k,$$

which leads to a contradiction, as the total population size cannot be of order larger than K in the limit $K \to \infty$, see Corollary A.7.

This proves (4.25). In particular, this implies that s^K_k converges to s_k in probability when K goes to infinity, as soon as $T > s_k$.

4.5 Value of the process at time $\theta_{k,m,C}^K \log K$

In this section, we prove that the stopping time $\theta_{k,m,C}^K$ is triggered by a (formerly) microscopic trait reaching order K, and not by the macroscopic traits deviating from their equilibrium.

We are now interested in the value of the process at time $\theta_{k,m,C}^K \log K$. First notice that according to Proposition A.2 in [10] and (4.25),

$$\lim_{K\to\infty} \mathbb{P}\left(\forall t \in [\sigma_{k-1}^K, (s_k^K + \eta_k) \wedge \theta_{k,m,C}^K], w \in \mathbf{v}_{k-1}, \left|\frac{N_w^K(t\log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1})\right| < C\varepsilon_k\right)$$

$$= \lim_{K\to\infty} \mathbb{P}\left(\forall t \in [\sigma_{k-1}^K, \theta_{k,m,C}^K], w \in \mathbf{v}_{k-1}, \left|\frac{N_w^K(t\log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1})\right| < C\varepsilon_k\right) = 1. \quad (4.27)$$

Notice that m has to be chosen small enough for this limit to hold, and thus depends on the parameters of the Lotka-Volterra deterministic system associated to $\mathbf{v}_{k-1}$. To be more precise, m has to be chosen small enough for the assumption (A.5) in Lemma A.6 to hold true with ε_k in place of ε. We choose such an m in the definition of $\theta_{k,m,C}^K$ in (4.1).

Hence we obtain that with high probability,

$$\sum_{w\in V\smallsetminus\mathbf{v}_{k-1}} N_w^K(\theta_{k,m,C}^K \log K) \geq m\varepsilon_k K.$$

If condition $(iv)(a)$ of Theorem 2.7 is satisfied T_0 is set at s_k and the induction is stopped. Otherwise there exists $\gamma > 0$ such that if ε_k is small enough $\beta_w^K(s_k^K) < 1 - \gamma$ for every $w \in V \smallsetminus (\tilde{\mathbf{v}}_{k-1}^K \cup \{l_k^K\})$. Thus again by coupling, as the growth rates of the populations are limited and η_k may be as small as we want, with high probability,

$$\sum_{w\in V\smallsetminus\tilde{\mathbf{v}}_{k-1}^K, w\neq l_k^K} N_w^K(\theta_{k,m,C}^K \log K) \leq K^{1-\gamma/2}. \quad (4.28)$$

From the two last inequalities we deduce that with high probability,

$$N_{l_k^K}^K(\theta_{k,m,C}^K \log K) \geq m\varepsilon_k K/2. \quad (4.29)$$

4.6 Construction of σ_k^K and Assumption 4.1

In this section, knowing that we have non-vanishing population sizes at $\theta_{k,m,C}^K$, we finally consider the Lotka-Volterra phase involving $\tilde{\mathbf{v}}_{k-1}$ and the trait l_k^K that has newly reached order K, proving that the initial conditions for the next step, characterised in the definition of σ_k^K, are satisfied after a time of order 1. This will conclude the proof of Theorem 2.7.

Let us now introduce the stopping time σ_k^K, via:

$$\sigma_k^K := \inf\{t \geq \theta_{k,m,C}^K, \forall w \in \mathbf{v}_k^K, |N_w^K(t\log K)/K - \bar{n}_w(\mathbf{v}_k)| \leq \varepsilon_k\}. \quad (4.30)$$

The last step of the proof consists in showing that σ_k^K indeed satisfies Assumption 4.1. First $\sigma_k^K \log K$ is a stopping time. Second, from (4.27), (4.28), (4.29) and an application of Lemma A.6 there exists $T(\varepsilon_k) < \infty$ such that

$$\lim_{K\to\infty} \mathbb{P}\left(\left|N_w^K(\theta_{k,m,C}^K \log K + T(\varepsilon_k))/K - \bar{n}_w(\mathbf{v}_k)\right| \leq \varepsilon_k, \forall w \in \mathbf{v}_k\right) = 1.$$

Moreover, during a time of order one, the order of population sizes does not vary more than a constant times ε_k (result similar in spirit to Lemma B.9 in [8]). Adding that s_k^K converges to s_k in probability when K goes to infinity, as well as (4.25), we obtain that Assumption 4.1 holds. It ends the proof of Theorem 2.7 and Proposition 2.8.

A Couplings with branching processes and logistic processes with immigration

The aim of this section is to collect various couplings of the populations with simpler processes like branching processes and logistic processes with immigration, and to state some properties of these simpler processes. These results have been derived in [8] (note that we need to slightly generalise some of them), and we state them for the sake of readability. For simplicity we keep the notations of [8].

A.1 Branching process

In this subsection, we recall Lemma A.1 of [8], which describes the dynamics of a birth and death process on a $\log K$ time scale. For $b, d, \beta \geq 0$, let $BP_K(b, d, \beta)$ denote the law of a process $(Z^K(t), t \geq 0)$ with initial state $Z^K(0) = \lfloor K^\beta - 1 \rfloor$, individual birth rate b and individual death rate d.

Lemma A.1 (Lemma A.1 in [8])**.** *Let $(Z^K(t), t \geq 0)$ be a $BP_K(b, d, \beta)$ process such that $\beta > 0$. The process $(\log(1 + Z^K(t \log K))/\log K, t > 0)$ converges when K tends to infinity in probability in $L^\infty([0, T])$ for all $T > 0$ to the continuous deterministic function given by*

$$\bar{\beta} : t \mapsto \beta + (b - d)t \vee 0.$$

In addition, if $b < d$, for all $t > \beta/(d - b)$,

$$\lim_{K \to \infty} \mathbb{P}\left(Z^K_{t \log K} = 0\right) = 1.$$

A.2 Branching process with immigration

In this subsection, we recall Lemma B.4 and Theorem B.5 of [8], illustrated in Figure B.1 therein, which describe the dynamics of birth and death processes with immigration on a $\log K$ time scale. For $b, d, \beta \geq 0$, $a, c \in \mathbb{R}$, $BPI_K(b, d, a, c, \beta)$ denotes the law of a process $(Z^K(t), t \geq 0)$ with initial state $Z^K(0) = \lfloor K^\beta - 1 \rfloor$, individual birth rate b, individual death rate d, and immigration rate $K^c e^{as}$ at time $s \geq 0$.

Lemma A.2 (Lemma B.4 in [8])**.** *Assume that $\beta < c$. Then for all $\varepsilon > 0$ and all $\bar{a} > |b - d| \vee |a|$,*

$$\lim_{K \to \infty} \mathbb{P}\left(Z^K(\varepsilon \log K) \in \left[K^{c - \bar{a}\varepsilon}, K^{c + \bar{a}\varepsilon}\right]\right) = 1.$$

Lemma A.3 (Theorem B.5 in [8])**.** *Let $(Z^K(t), t \geq 0)$ be a $BPI_K(b, d, a, c, \beta)$ process with $c \leq \beta$ and assume that $\beta > 0$. The process $(\log(1 + Z^K(t \log K))/\log K, t > 0)$ converges when K tends to infinity in probability in $L^\infty([0, T])$ for all $T > 0$ to the continuous deterministic function $\bar{\beta}$ given by*

$$\bar{\beta} : t \mapsto (\beta + (b - d)t) \vee (c + at) \vee 0.$$

In addition, in the case where $c \neq 0$ or $a \neq 0$, for all compact intervals $I \subset \mathbb{R}_+$ which do not intersect the support of $\bar{\beta}$,

$$\lim_{K \to \infty} \mathbb{P}\left(Z^K(t \log K) = 0, \forall t \in I\right) = 1.$$

We will mostly use a corollary of those two lemmas, which is valid without the assumption $c \leq \beta$ but on a time interval $[\delta, T]$, for any $\delta > 0$. The idea of the proof has been explained in Remark 4.3.

Corollary A.4. *Let $(Z^K(t), t \geq 0)$ be a $BPI_K(b, d, a, c, \beta)$ process with $\beta \geq 0$, and either $c > 0$ or both $c = 0$ and $a > 0$. For any $\delta > 0$ and $T > 0$, the process*

$(\log(1 + Z^K(t \log K))/\log K, t \in [\delta, T])$ *converges when* K *tends to infinity in probability in* $L^\infty([\delta, T])$ *to the continuous deterministic function* $\bar{\beta}$ *given by*

$$\bar{\beta} : t \mapsto ((\beta \vee c) + (b - d)t) \vee (c + at) \vee 0.$$

A.3 Logistic birth and death process with immigration

We recall that for a subset $\mathbf{v} \subset V$ of traits that can coexist at a strictly positive equilibrium in the Lotka-Volterra system (2.3), $\bar{n}(\mathbf{v}) \in \mathbb{R}^{\mathbf{v}}_+$ denotes this equilibrium. The next result states that if all traits in $\mathbf{v}$ have an initial population of order K and the immigration of individuals with traits in $\mathbf{v}$ is small enough, the equilibrium $\bar{n}(\mathbf{v})K$ is reached in a time of order 1 and the populations of individuals whose traits belong to $\mathbf{v}$ will keep a size close to its equilibrium during a time of order larger than $\log K$

This result is a generalisation of Lemma C.1 in [8] to the multidimensional case and with (slightly) varying rates.

We thus consider a subset $\mathbf{v} \subset V$ of traits and denote by $(\mathbf{b_v}(t), t \geq 0) := ((b_w(t), w \in \mathbf{v}), t \geq 0)$, $(\mathbf{d_v}(t), t \geq 0) := ((d_w(t), w \in \mathbf{v}), t \geq 0)$, and $(\mathbf{c_v}(t), t \geq 0) := ((c_{w_1, w_2}(t), (w_1, w_2) \in \mathbf{v}^2), t \geq 0)$ its birth, natural death, and death by competition rates that we allow to vary in time, as well as $(\mathbf{g_v}(t), t \geq 0) := ((g_w(t), w \in \mathbf{v}), t \geq 0)$ a function with values in $\mathbb{R}^{\mathbf{v}}_+$. We denote by $LBDI_K(\mathbf{b_v}, \mathbf{d_v}, \mathbf{c_v}, \mathbf{g_v})$ the law of a logistic birth and death process with immigration $\mathbf{Z}^K := ((Z_w(t)^K, w \in \mathbf{v}), t \geq 0)$ where, at time t, an individual with a trait $w \in \mathbf{v}$ has a birth rate $b_w(t)$, a death rate $d_w(t) + \sum_{x \in \mathbf{v}} c_{w,x}(t) Z_x^K(t)/K$ and an immigration rate $g_w(t)$.

In order to state the next result, we introduce the following assumption:

Assumption A.5. *There exist* $\varepsilon, \eta, \mathcal{T} > 0$ *such that*

$$\sup_{w_1, w_2 \in \mathbf{v}} \left\{ \left| \mathbf{b}_{w_1}(t) - \bar{\mathbf{b}}_{w_1} \right|, \left| \mathbf{d}_{w_1}(t) - \bar{\mathbf{d}}_{w_1} \right|, \left| \mathbf{c}_{w_1, w_2}(t) - \bar{\mathbf{c}}_{w_1, w_2} \right| \right\} < \varepsilon$$

and $g_w(t) \leq K^{1-\eta}$, *for all* $t \leq \mathcal{T} \log K$.

Lemma A.6. *Let* $\mathbf{v} \subset V$ *and assume that the mutation-free Lotka-Volterra system* (2.3) *associated to* $\mathbf{v}$ *and with rates* $(\bar{\mathbf{b}}_{\mathbf{v}}, \bar{\mathbf{d}}_{\mathbf{v}}, \bar{\mathbf{c}}_{\mathbf{v}}) \in (\mathbb{R}^+_*)^{\mathbf{v}} \times (\mathbb{R}^+_*)^{\mathbf{v}} \times (\mathbb{R}^+_*)^{\mathbf{v}^2}$ *admits a unique positive globally attractive stable equilibrium* $\bar{n}(\mathbf{v})$. *Assume that* Z^K *follows the law* $LBDI_K(\mathbf{b_v}, \mathbf{d_v}, \mathbf{c_v}, \mathbf{g_v})$.

(i) *Suppose that Assumption A.5 holds for a triplet* $(\varepsilon, \eta, \mathcal{T})$. *Then there exist* $C, \varepsilon_0^{(1)} > 0$ *such that if* $\varepsilon \leq \varepsilon_0^{(1)}$ *and* $\|\mathbf{Z}^K(0)/K - \bar{n}(\mathbf{v})\|_\infty \leq \varepsilon$, *then*

$$\lim_{K \to \infty} \mathbb{P} \left(\forall t \in [0, \mathcal{T} \log K], \|\mathbf{Z}^K(t)/K - \bar{n}(\mathbf{v})\|_\infty \leq C\varepsilon \right) = 1.$$

(ii) *Suppose that Assumption A.5 holds for a triplet* $(\varepsilon, \eta, \mathcal{T})$. *Let* $\varepsilon_1, \varepsilon_2 > 0$. *Then there exist* $T(\varepsilon_1, \varepsilon_2) < \infty$ *and* $\varepsilon_0^{(2)} > 0$ *such that if* $\varepsilon < \varepsilon_0^{(2)}$,

$$\lim_{K \to \infty} \sup_{Z_v^K(0)/K \geq \varepsilon_1 \ \forall v \in \mathbf{v}} \mathbb{P} \left(\forall t \in [T(\varepsilon_1, \varepsilon_2), \mathcal{T} \log K], \|\mathbf{Z}^K(t)/K - \bar{n}(\mathbf{v})\|_\infty \leq \varepsilon_2 \right) = 1.$$

Proof. The case where the functions $\mathbf{b_v}$, $\mathbf{d_v}$, $\mathbf{c_v}$ are constant is a direct generalisation of Lemma C.1 in [8], whose proof follows arguments similar to the ones given in [7, 10] or in the Proposition 4.2 in [9] to handle the addition of (negligible) immigration. We do not provide it. Let us explain how we deal with varying rates for point (i). Let us choose $w_0 \in \mathbf{v}$, and introduce for $w_1, w_2 \in \mathbf{v}$:

$$\tilde{b}_{w_1} = \begin{cases} \bar{b}_{w_1} + \varepsilon & \text{if } w_1 \neq w_0 \\ \bar{b}_{w_1} - \varepsilon & \text{if } w_1 = w_0 \end{cases}$$

$$\tilde{d}_{w_1} = \begin{cases} \bar{d}_{w_1} - \varepsilon & \text{if } w_1 \neq w_0 \\ \bar{d}_{w_1} + \varepsilon & \text{if } w_1 = w_0 \end{cases}$$

$$\tilde{c}_{w_1,w_2} = \begin{cases} \bar{c}_{w_1,w_2} - \varepsilon & \text{if } w_1 \neq w_0 \\ \bar{c}_{w_1,w_2} + \varepsilon & \text{if } w_1 = w_0 \end{cases}$$

Then we can couple a process Z^K with the law $LBDI_K(\mathbf{b_v}, \mathbf{d_v}, \mathbf{c_v}, \mathbf{g_v})$ with a process $\tilde{Z}^K$ with the law $LBDI_K(\tilde{\mathbf{b}}_\mathbf{v}, \tilde{\mathbf{d}}_\mathbf{v}, \tilde{\mathbf{c}}_\mathbf{v}, \mathbf{g_v})$ such that for every $t \geq 0$, $\tilde{Z}^K_{w_0}(t) \leq Z^K_{w_0}(t)$ and $\tilde{Z}^K_w(t) \geq Z^K_w(t)$ for every $w \in \mathbf{v} \setminus w_0$. Moreover, as the equilibrium of a Lotka-Volterra system is continuous with respect to its coefficient, there is a positive $\tilde{C}$ such that for ε small enough, and if we denote by $\bar{n}^{(w_0)}(\mathbf{v})$ the equilibrium of the Lotka-Volterra system with the coefficients $\tilde{\mathbf{b}}_\mathbf{v}, \tilde{\mathbf{d}}_\mathbf{v}, \tilde{\mathbf{c}}_\mathbf{v}$ we have just introduced, $\|\bar{n}^{(w_0)}(\mathbf{v}) - \bar{n}(\mathbf{v})\| \leq \tilde{C}\varepsilon$. Hence applying the point (i) for the process $\tilde{Z}^K$, we obtain upper bounds for coordinates $w \neq w_0$ and a lower bound for the coordinate w_0, for the process Z^K. Doing the same and the reverse bounds for the other elements of $\mathbf{v}$ gives the result for some $C > \tilde{C}$ that takes into account the fluctuations around the varied equilibria.

Let us now explain how we get point (ii). First, we consider for every $w_0 \in \mathbf{v}$ the deterministic Lotka-Volterra process $(\tilde{z}_{w_0}(t), t \geq 0)$ with parameters $(\tilde{\mathbf{b}}_\mathbf{v}, \tilde{\mathbf{d}}_\mathbf{v}, \tilde{\mathbf{c}}_\mathbf{v}, \mathbf{g_v})$ introduced in the proof of point (i), and an initial state larger than ε_1. We know that there exists a time $T(\varepsilon_1, \varepsilon_2, w_0)$ such that if ε in Assumption A.5 is small enough this process satisfies:

$$|\tilde{z}_{w_0}(t) - \bar{n}(\mathbf{v})| \leq \frac{\varepsilon_2 \wedge \varepsilon_0^{(1)}}{3C}, \quad \forall t \geq T(\varepsilon_1, \varepsilon_2, w_0),$$

where C and $\varepsilon_0^{(1)}$ have been defined in point (i). From Chapter 11, Theorem 2.1 in [21], we also know that

$$\lim_{K \to \infty} \mathbb{P}\left(\sup_{t \leq T(\varepsilon_1, \varepsilon_2, w_0)} \|\tilde{Z}^K(t)/K - \tilde{z}_{w_0}(t)\| \leq \frac{\varepsilon_2 \wedge \varepsilon_0^{(1)}}{3C} \right) = 1.$$

Applying point (i) we thus obtain

$$\lim_{K \to \infty} \mathbb{P}\left(\forall t \in [T(\varepsilon_1, \varepsilon_2, w_0), \mathcal{T} \log K], \|\tilde{Z}^K(t)/K - \bar{n}(\mathbf{v})\|_\infty \leq \varepsilon_2 \right) = 1.$$

As this holds for any w_0 in the finite graph V this concludes the proof of point (ii). $\qquad \square$

We end this section with a result stating that the time needed for the total population size of a logistic birth and death process (with or without mutations) to reach (and stay smaller than or equal to) an order K is of order one for K large enough.

Corollary A.7. *Let us consider a subset $\mathbf{v} \subset V$ of traits, $(\mathbf{b_v}, \mathbf{d_v}, \mathbf{c_v})$ be in $(\mathbb{R}_*^+)^\mathbf{v} \times (\mathbb{R}_*^+)^\mathbf{v} \times (\mathbb{R}_*^+)^{\mathbf{v}^2}$ and let Z^K follow the law $LBDI_K(\mathbf{b_v}, \mathbf{d_v}, \mathbf{c_v}, 0)$, and $\mathcal{Z}^K$ denote the total population size of the process Z^K. For every $\varepsilon > 0$ there exist $T(\varepsilon), C < \infty$ such that*

$$\lim_{K \to \infty} \mathbb{P}\left(\sup_{T(\varepsilon) < t < e^{CK}} \frac{\log(1 + \mathcal{Z}^K(t))}{\log K} < 1 + \varepsilon \right) = 1.$$

Remark A.8. Notice that this result only treats mutation-free logistic birth and death processes. However, mutation within $\mathbf{v}$ does not affect the total population size and hence the result can be transferred to such cases. Considering $\mathbf{v} = V$, Corollary A.7 therefore implies the same asymptotic bound for the total population size of the process that we consider in Theorem 2.7 and Proposition 2.8, and hence also for each subpopulation.

Proof. The process $\mathcal{Z}^K$ increases by 1 at a rate

$$\sum_{w \in \mathbf{v}} b_w Z_w^K \leq (\sup_{w \in \mathbf{v}} b_w) \mathcal{Z}^K =: \mathcal{B} \mathcal{Z}^K$$

and decreases by 1 at a rate

$$\sum_{w \in \mathbf{v}} \left(d_w + \sum_{u \in \mathbf{v}} \frac{c_{w,u}}{K} Z_u^K \right) Z_w^K \geq \frac{1}{K} (\inf_{u \in \mathbf{v}} c_{u,u}) \sum_{w \in \mathbf{v}} \left(Z_w^K \right)^2$$
$$\geq \frac{1}{K} (\inf_{u \in \mathbf{v}} c_{u,u}) \frac{1}{Card(\mathbf{v})} \left(\mathcal{Z}^K \right)^2 =: \frac{\mathcal{C}}{K} \left(\mathcal{Z}^K \right)^2 .$$

Hence the process $\mathcal{Z}^K$ can be coupled with a logistic birth and death process $\mathcal{N}^K$ with individual birth rate $\mathcal{B}$ and individual death rate $\mathcal{C} \mathcal{N}^K / K$ in such a way that for every $t \geq 0$, if $\mathcal{Z}^K(0) = \mathcal{N}^K(0)$

$$\mathcal{Z}^K(t) \leq \mathcal{N}^K(t).$$

But from Chapter 11, Theorem 2.1 in [21], we know that on any finite time interval, the rescaled process $\mathcal{N}^K / K$ converges in probability to the solution to the logistic equation $\dot{\varkappa} = \varkappa(\mathcal{B} - \mathcal{C}\varkappa)$, $\varkappa(0) = \varkappa_0$ if $\mathcal{N}^K(0)/K$ converges in probability to $\varkappa_0$. The one dimensional logistic equation has an explicit solution, and in particular, we know that its equilibrium is $\mathcal{B}/\mathcal{C}$, that it comes down from infinity, and that it takes a time

$$\frac{1}{\mathcal{B}} \log \left(\frac{\bar{\varkappa}}{\bar{\varkappa} - \mathcal{B}/\mathcal{C}} \right)$$

to reach $\bar{\varkappa} > \mathcal{B}/\mathcal{C}$ from an infinite initial condition. As a consequence, $\mathcal{N}^K$ takes a time of order one to become smaller than $2\bar{\varkappa}K$, and as $\mathcal{B}/\mathcal{C}$ is a globally hyperbolic equilibrium for the function $\varkappa$, classical large deviation results (see [19] for instance) entail that $\mathcal{N}^K / K$ will stay an exponential (in K) time in any compact interval of R_+^* including $\mathcal{B}/\mathcal{C}$. This concludes the proof. $\square$

References

[1] Julien Berestycki, Eric Brunet, and Zhan Shi, *The number of accessible paths in the hypercube*, Bernoulli **22** (2016), no. 2, 653–680. MR-3449796

[2] Sylvain Billiard and Charline Smadi, *The interplay of two mutations in a population of varying size: a stochastic eco-evolutionary model for clonal interference*, Stoch. Process. their Appl. **127** (2017), no. 3, 701–748. MR-3605709

[3] Sylvain Billiard and Charline Smadi, *Stochastic dynamics of three competing clones: Conditions and times for invasion, coexistence, and fixation*, The American Naturalist **195** (2020), no. 3, 463–484.

[4] Benjamin M Bolker and Stephen W Pacala, *Using Moment Equations to Understand Stochastically Driven Spatial Pattern Formation in Ecological Systems*, Theor. Popul. Biol. **52** (1997), no. 3, 179–197.

[5] Benjamin M Bolker and Stephen W Pacala, *Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal*, Am Nat **153** (1999), no. 6, 575–602.

[6] Anton Bovier, Loren Coquille, and Charline Smadi, *Crossing a fitness valley as a metastable transition in a stochastic population model*, Ann. Appl. Probab. **29** (2019), no. 6, 3541–3589. MR-4047987

[7] Nicolas Champagnat, *A microscopic interpretation for adaptive dynamics trait substitution sequence models*, Stoch. Process. their Appl. **116** (2006), no. 8, 1127–1160. MR-2250806

[8] Nicolas Champagnat, Sylvie Méléard, and Viet Chi Tran, *Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer*, ArXiv Prepr. arXiv:1901.02385 (2019). MR-4312848

[9] Nicolas Champagnat, Pierre-Emmanuel Jabin, and Sylvie Méléard, *Adaptation in a stochastic multi-resources chemostat model*, J. Math. Pures Appl. **101** (2014), no. 6, 755–788. MR-3205642

[10] Nicolas Champagnat and Sylvie Méléard, *Polymorphic evolution sequence and evolutionary branching*, Probab. Theory Relat. Fields **151** (2011), no. 1-2, 45–94. MR-2834712

[11] Jean-René Chazottes, Pierre Collet, and Sylvie Méléard, *On time scales and quasi-stationary distributions for multitype birth-and-death processes*, (2017). MR-4029154

[12] Pierre Collet, Sylvie Méléard, and Johan A J Metz, *A rigorous model study of the adaptive dynamics of Mendelian diploids*, J. Math. Biol. **67** (2013), no. 3, 569–607. MR-3084364

[13] Camille Coron, Manon Costa, Hélène Leman, and Charline Smadi, *A stochastic model for speciation by mating preferences*, J. Math. Biol. **76** (2018), no. 6, 1421–1463. MR-3771426

[14] Matthew C Cowperthwaite, James J Bull, and Lauren A Meyers, *From Bad to Good: Fitness Reversals and the Ascent of Deleterious Mutations*, PLoS Comput. Biol. **2** (2006), no. 10, e141.

[15] J Arjan G M de Visser and Joachim Krug, *Empirical fitness landscapes and the predictability of evolution*, Nature Reviews Genetics **15** (2014), no. 7, 480–490.

[16] Mark A DePristo, Daniel L Hartl, and Daniel M Weinreich, *Mutational Reversions During Adaptive Protein Evolution*, Mol. Biol. Evol. **24** (2007), no. 8, 1608–1610.

[17] Ulf Dieckmann and Richard Law, *Moment approximations of individual-based models*, Geom. Ecol. Interact. Simpl. Spat. Complex. (U Dieckmann, R Law, and J A J Metz, eds.), Cambridge University Press, 2000, pp. 252–270. MR-2265382

[18] Ulf Dieckmann and Richard Law, *The dynamical theory of coevolution: a derivation from stochastic ecological processes*, J. Math. Biol. **34** (1996), no. 5-6, 579–612. MR-1393842

[19] Paul Dupuis and Richard S Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, USA, feb 1997. MR-1431744

[20] Rick Durrett and John Mayberry, *Traveling waves of selective sweeps*, Ann. Appl. Probab. **21** (2011), no. 2, 699–744. MR-2807971

[21] Stewart N Ethier and Thomas G Kurtz, *Markov Processes: Characterization and Convergence*, 1986. MR-0838085

[22] Nicolas Fournier and Sylvie Méléard, *A microscopic probabilistic description of a locally regulated population and macroscopic approximations*, Ann. Appl. Probab. **14** (2004), no. 4, 1880–1919. MR-2099656

[23] Stefan A H Geritz, Johan A J Metz, Éva Kisdi, and Géza Meszéna, *Dynamics of Adaptation and Evolutionary Branching*, Phys. Rev. Lett. **78** (1997), no. 10, 2024–2027.

[24] Stuart Kauffman and Simon Levin, *Towards a general theory of adaptive walks on rugged landscapes*, J. Theor. Biol. **128** (1987), no. 1, 11–45. MR-0907587

[25] Anna Kraut and Anton Bovier, *From adaptive dynamics to adaptive walks*, J. Math. Biol. **79** (2019), no. 5, 1699–1747. MR-4019937

[26] Joachim Krug, *Accessibility percolation in random fitness landscapes*, to Appear "Probabilistic Struct. Evol. ed. by E. Baake A. Wakolbinger (2019).

[27] Richard E Lenski, Charles Ofria, Robert T Pennock, and Christoph Adami, *The evolutionary origin of complex features*, Nature **423** (2003), no. 6936, 139–144.

[28] Rohan Maddamsetti, Richard E Lenski, and Jeffrey E Barrick, *Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with escherichia coli*, Genetics **200** (2015), no. 2, 619–631.

[29] John Maynard Smith, *Natural selection and the concept of a protein space*, Nature **225** (1970), no. 5232, 563–564.

[30] Johan A J Metz, Stefan A H Geritz, G Meszena, F J A Jacobs, and J S van Heerwaarden, *Adaptive Dynamics: A Geometrical Study of the Consequences of Nearly Faithful Reproduction*, Iiasa working paper, IIASA, Laxenburg, Austria, 1995. MR-1773114

[31] Adilson E Motter, *Improved network performance via antagonism: From synthetic rescues to multi-drug combinations*, BioEssays **32** (2010), no. 3, 236–245.

[32] Johannes Neidhart and Joachim Krug, *Adaptive walks and extreme value theory*, Phys. Rev. Lett. **107** (2011), no. 17.

[33] Rebecca Neukirch and Anton Bovier, *Survival of a recessive allele in a Mendelian diploid model*, J. Math. Biol. (2016), 1–54. MR-3665247

[34] Stefan Nowak and Joachim Krug, *Analysis of adaptive walks on NK fitness landscapes with different interaction schemes*, J. Stat. Mech. Theory Exp. **2015** (2015), no. 6. MR-3370889

[35] H Allen Orr, *A minimum on the mean number of steps taken in adaptive walks*, J. Theor. Biol. **220** (2003), no. 2, 241–247. MR-2043002

[36] Joshua B Plotkin and Grzegorz Kudla, *Synonymous but not the same: The causes and consequences of codon bias*, jan 2011, pp. 32–42.

[37] Sagar Sahasrabudhe and Adilson E Motter, *Rescuing ecosystems from extinction cascades through compensatory perturbations*, Nat. Commun. **2** (2011), no. 1.

[38] B Schmiegelt and J Krug, *Evolutionary Accessibility of Modular Fitness Landscapes*, J. Stat. Phys. **154** (2014), no. 1-2, 334–355. MR-3162544

[39] Charline Smadi, *An eco-evolutionary approach of adaptation and recombination in a large population of varying size*, Stoch. Process. their Appl. **125** (2015), no. 5, 2054–2095. MR-3315623

[40] Charline Smadi, *The Effect of Recurrent Mutations on Genetic Diversity in a Large Population of Varying Size*, Acta Appl. Math. **149** (2017), no. 1, 11–51. MR-3647031

[41] Ivan G Szendro, Martijn F Schenk, Jasper Franke, Joachim Krug, and J Arjan G M de Visser, *Quantitative analyses of empirical fitness landscapes*, J. Stat. Mech. Theory Exp. **2013** (2013), no. 1. MR-3036207

[42] M L Zeeman, *Hopf bifurcations in competitive three-dimensional Lotka-Volterra systems*, Dynam. Stab. Syst. **8** (1993), no. 3, 189–217. MR-1246002

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm