

The polymorphic evolution sequence for populations with phenotypic plasticity*

Martina Baar[†] Anton Bovier[‡]

Abstract

In this paper we study a class of stochastic individual-based models that describe the evolution of haploid populations where each individual is characterised by a phenotype and a genotype. The phenotype of an individual determines its natural birth- and death rates as well as the competition kernel, $c(x, y)$ which describes the induced death rate that an individual of type x experiences due to the presence of an individual of type y . When a new individual is born, with a small probability a mutation occurs, i.e. the offspring has different genotype as the parent. The novel aspect of the models we study is that an individual with a given genotype may express a certain set of different phenotypes, and during its lifetime it may switch between different phenotypes, with rates that are much larger than the mutation rates and that, moreover, may depend on the state of the entire population. The evolution of the population is described by a continuous-time, measure-valued Markov process. In [4], such a model was proposed to describe tumor evolution under immunotherapy. In the present paper we consider a large class of models which comprises the example studied in [4] and analyse their scaling limits as the population size tends to infinity and the mutation rate tends to zero. Under suitable assumptions, we prove convergence to a Markov jump process that is a generalisation of the polymorphic evolution sequence (PES) as analysed in [9, 11].

Keywords: adaptive dynamics; canonical equation; large population limit; mutation-selection individual-based model.

AMS MSC 2010: 60K35; 60J85; 92D25.

Submitted to EJP on August 5, 2017, final version accepted on July 3, 2018.

*M. B. is supported by the German Research Foundation through the Priority Programme 1590 “Probabilistic Structures in Evolution”. A.B. is partially supported by the German Research Foundation in the Collaborative Research Center 1060 “The Mathematics of Emergent Effects”, the Priority Programme 1590 “Probabilistic Structures in Evolution”, the Hausdorff Center for Mathematics (HCM), and the Cluster of Excellence “ImmunoSensation” at Bonn University.

[†]Institut für Angewandte Mathematik, Rheinische Friedrich-Wilhelms-Universität, Endenicher Allee 60, 53115 Bonn, Germany

E-mail: mbaar@uni-bonn.de

[‡]Institut für Angewandte Mathematik, Rheinische Friedrich-Wilhelms-Universität, Endenicher Allee 60, 53115 Bonn, Germany

E-mail: bovier@uni-bonn.de <http://wt.iam.uni-bonn.de/bovier>

1 Introduction

Over the last decade there has been increasing interest in the mathematical analysis of so-called *stochastic individual based models of adaptive dynamics*. These models were introduced in a series of papers by Bolker, Pacala, Dieckmann, and Law [7, 8, 14]. They describe the evolution of a population of individuals characterised by their phenotypes under the influence of the evolutionary mechanisms of birth, death, mutation, and ecological competition in an inhomogeneous "fitness landscape" as a measure valued Markov process. Note that individual based models have a much longer tradition in ecology (see the book by Grimm and Railsback [20] or the more recent review by Grimm and DeAngelis [13]). In these models there appear two natural scaling parameters. The *carrying capacity*, K , which regulates the size of the population and that can reasonably be considered as a large parameter, and the mutation rate (of advantageous mutations), u , that in many biological situations can be taken as a small parameter. In a series of remarkable papers, Champagnat and Méléard [9, 11] (and others) have analysed the limiting processes that arise in the limit when K is taken to infinity while at the same time $u = u_K$ tends to zero. Under conditions that ensure the separation of the *ecological* and *evolutionary* time scales. This means that the mutation rates are so small that the system has time to equilibrate (ecological time scale) between two mutational events. On the time scale where mutations occur (evolutionary time scale), the evolution of the population can then be described as a Markov jump process along a sequence of equilibria of, in general, polymorphic populations. An important (and in some sense generic) special case occurs when the mutant population fixates while the resident population dies out in each step. The corresponding jump process is called the *Trait Substitution Sequence* (TSS) in adaptive dynamics. Champagnat [9] derived criteria in the context of individual-based models under which convergence to the TSS can be proven. The general process is called the *Polymorphic Evolution Sequence* (PES) [11]. Here the limit is describes as a jump process between possibly polymorphic equilibria of systems of Lotka-Volterra equations of increasing dimension.

In the present paper we extend this analysis to models where an additional biological phenomenon is present, the so-called *phenotypic plasticity*[19] or *bet hedging* [6]. By this we mean the following. Individuals are no longer described by their phenotype, but by both their *genotype* and their phenotype. Moreover, an individual of a given genotype can express several phenotypes and it can change its phenotype during the course of its lifetime. The term phenotypic plasticity is often used to describe a situation where the expression of the phenotype depends on the environment in a deterministic way, while bet hedging describes stochastic switches that "hedge" against future changes in the environment. In the model we discuss, both aspects, stochastic switching and environmental dependence are present.

Our original motivation comes from applications to cancer therapy, where it is well-known that phenotypic switching is of utmost importance and in fact a major obstacle to successful therapies (see, e.g. [21] and references therein). Specifically, Landsberg et al [25] showed in a mouse model of melanoma that phenotypic switches in the expression of surface markers of tumor cells are responsible for evasion of immunotherapy with adoptive cell transfer. Moreover, they showed that switch rates are affected by the presence of cytokines, in particular $\text{TNF-}\alpha$ that is produced in an inflammatory environment. In [4] we reported on first attempts to model these experiments in the context of the process discussed in the present paper.

Phenotypic switches without mutations are certainly relevant in many if not most biological systems (see e.g. [6, 19] and references therein).

Here we take a broader look at a large class of models. By expanding the techniques of [11] we prove that the microscopic process converges on the evolutionary time scale

to a generalisation of the Polymorphic Evolution Sequences (PES) (cf. Thm. 3.6). The main difference in the proof is that we have to couple the process with multi-type branching processes instead of normal branching processes, which leads also to a different definition of invasion fitness in this setting. Note that we gave in [4] already heuristic arguments why the process should converge to a Markov jump process. The aim of this paper is to give the rigorous statement and its proof.

The remainder of this paper is organised as follows. In Section 2 we define the model, give a pathwise description of the Markov process we are studying and state the convergence towards a quadratic system of ODEs in the large population limit. In Sections 3 we consider the case of rare mutations and fast switches. More precisely, we state the convergence to the Polymorphic Evolution Sequences with phenotypic Plasticity (PESP) in Subsection 3.2 and prove it in Subsection 3.3.

2 The microscopic model

In this section we introduce the stochastic individual-based model we analyse (cf. [4, 17, 9, 11, 5]). The evolutionary process changes populations on a macroscopic level, but the basic mechanisms of evolution, heredity, variation (in our context caused by mutation and phenotypical switching), and selection, act on the microscopic level of the individuals. We describe the evolving population as a stochastic system of interacting individuals, where each individual is characterised by its phenotype and its genotype.

Let $l \geq 1$ and \mathcal{X} a finite set of the form $\mathcal{X} = \mathcal{G} \times \mathcal{P}$, where \mathcal{G} is the set of genotypes and \mathcal{P} is the set of phenotypes. We call \mathcal{X} the *trait space* of the population. As usual, we introduce a parameter $K \in \mathbb{N}$, called the *carrying capacity*. This parameter allows to scale the population size and can be interpreted as the size of available space or the amount of available resources. Let $\mathcal{M}(\mathcal{X})$ be the set of finite, non-negative measures on \mathcal{X} , equipped with the topology of weak convergence, and let $\mathcal{M}^K(\mathcal{X}) \subset \mathcal{M}(\mathcal{X})$ be the set of finite point measures on \mathcal{X} rescaled by K , i.e.

$$\mathcal{M}^K(\mathcal{X}) \equiv \left\{ \frac{1}{K} \sum_{i=1}^n \delta_{x_i} : n \in \mathbb{N}_0, x_1, \dots, x_n \in \mathcal{X} \right\}, \quad (2.1)$$

where δ_x denotes the Dirac mass at $x \in \mathcal{X}$. We model the time evolution of a population as an $\mathcal{M}^K(\mathcal{X})$ -valued, continuous time Markov process $(\nu_t^K)_{t \geq 0}$. To account for the process basic mechanisms of evolution and the phenotypic plasticity, we introduce the following parameters:

- (i) $b(p) \in \mathbb{R}_+$ is the *rate of birth* of an individual with phenotype $p \in \mathcal{P}$.
- (ii) $d(p) \in \mathbb{R}_+$ is the *rate of natural death* of an individual with with phenotype $p \in \mathcal{P}$.
- (iii) $c(p, \tilde{p})K^{-1} \in \mathbb{R}_+$ is the *competition kernel*, which models the competitive pressure an individual with phenotype $p \in \mathcal{P}$ feels from an individual with phenotype $\tilde{p} \in \mathcal{P}$ and is inversely proportional to the carrying capacity K .
- (iv) $s_{\text{nat.}}^g(p, \tilde{p}) \in \mathbb{R}_+$ is the *natural switch kernel*, which models the natural switching from phenotype p to \tilde{p} of individuals with genotype g .
- (v) $s_{\text{ind.}}^g(p, \tilde{p})(\hat{p})K^{-1} \in \mathbb{R}_+$ is the *induced switch kernel*, which models the switching from phenotype p to \tilde{p} of individuals with genotype g induced by an individual with phenotype \hat{p} . (Compare with the cytokine-induced switch of [4], especially the one of TNF- α (Tumour Necrosis Factor).)
- (vi) $u_K m(g)$ with $u_K, m(g) \in [0, 1]$ is the *probability that a mutation occurs at birth* from an individual with genotype $g \in \mathcal{G}$, where u_K is a scaling parameter.
- (vii) $M((g, p), (\tilde{g}, \tilde{p}))$ is the *mutation law*, i.e. if a mutant is born from an individual with trait (g, p) , then the mutant's trait is (\tilde{g}, \tilde{p}) with probability $M((g, p), (\tilde{g}, \tilde{p}))$.

Note that most of the parameters depend on the phenotype only and that we explicitly allow that individuals with different genotypes can express the same phenotype and, conversely, that individuals with the same genotype can express different phenotypes.

We assume that phenotypic switches can be enhanced by the environment, i.e. induced switching can only strengthen natural switching and that phenotypic switches happen with rate of order one, i.e. they occur with positive probability during the lifetime of an individual (unless they are set to zero). One might also be interested in cases where these rates are much smaller, i.e. tending to zero with K . This would require a different analysis and is beyond the scope of this paper.

Assumption 2.1. *For simplicity, we assume that $s_{ind}^g(p, \tilde{p})(\hat{p})K^{-1} = 0$, for all $\hat{p} \in \mathcal{X}$, whenever $s_{nat}^g(p, \tilde{p}) = 0$, i.e. depending on the environment the total switching rate can be larger or smaller but not zero or non-zero.*

At any time $t \geq 0$, we consider a finite population which consist of N_t individuals and each individual is characterised its trait $x_i(t) \in \mathcal{X}$. The state of a population at time t is the measure

$$\nu_t^K = \frac{1}{K} \sum_{i=1}^{N_t} \delta_{x_i(t)}. \tag{2.2}$$

The population process ν^K is a $\mathcal{M}^K(\mathcal{X})$ -valued Markov process with infinitesimal generator \mathcal{L}^K , defined, for any bounded measurable function $\phi : \mathcal{M}^K(\mathcal{X}) \rightarrow \mathbb{R}$ and for all $\mu^K \in \mathcal{M}^K(\mathcal{X})$, by

$$\begin{aligned} & (\mathcal{L}^K \phi)(\mu^K) \tag{2.3} \\ &= \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \left(\phi \left(\mu^K + \frac{\delta_{(g,p)}}{K} \right) - \phi(\mu^K) \right) (1 - u_K m(g)) b(p) K \mu^K(g, p) \\ &+ \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{G} \times \mathcal{P}} \left(\phi \left(\mu^K + \frac{\delta_{(\tilde{g}, \tilde{p})}}{K} \right) - \phi(\mu^K) \right) u_K m(g) M((g, p), (\tilde{g}, \tilde{p})) b(p) K \mu^K(g, p) \\ &+ \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \left(\phi \left(\mu^K - \frac{\delta_{(g,p)}}{K} \right) - \phi(\mu^K) \right) \left(d(p) + \sum_{\tilde{p} \in \mathcal{P}} c(p, \tilde{p}) \mu^K(\tilde{p}) \right) K \mu^K(g, p) \\ &+ \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \sum_{\tilde{p} \in \mathcal{P}} \left(\phi \left(\mu^K + \frac{\delta_{(g,\tilde{p})}}{K} - \frac{\delta_{(g,p)}}{K} \right) - \phi(\mu^K) \right) \\ &\quad \times \left(s_{nat}^g(p, \tilde{p}) + \sum_{\hat{p} \in \mathcal{P}} s_{ind}^g(p, \tilde{p})(\hat{p}) \mu^K(\hat{p}) \right) K \mu^K(g, p). \end{aligned}$$

The first and second terms describe the births (without and with mutation), the third term describes the deaths due to age or competition, and the last term describes the phenotypic plasticity. Observe that the first and second terms are linear (in μ^K), but the third and fourth terms are non-linear. The only difference to the standard model is the presence of the fourth term, that corresponds to the phenotypic switches. However, this term changes the dynamics substantially. In particular, the system of differential equations which arises in the large population limit without mutation ($u_K = 0$) is not a generalised Lotka-Volterra system anymore, i.e. does not have the form $\dot{n} = n f(n)$, where f is linear in n (cf. Thm. 2.4 and Def. 3.2).

Remark 2.2. (i) Since \mathcal{X} is finite, we could also represent the population state as an $|\mathcal{X}|$ -dimensional vector. More precisely, let E be a subset of $\mathbb{R}^{|\mathcal{X}|}$ and $E^K \equiv E \cap \{n/K : n \in \mathbb{N}_0^{|\mathcal{X}|}\}$ (we denote by \mathbb{N}_0 the non-negative integers), then, for fixed $K \geq 1$, the population process can be constructed as Markov process with state space E^K by using independent standard Poisson processes (cf. [16] Chap. 11).

- (ii) For an extension to a non-finite trait space, e.g. if \mathcal{G} and \mathcal{P} are compact subsets of \mathbb{R}^k , for some $k \geq 1$, the modelling of switching the phenotype has to be changed in the following way: Each individual with trait $(g, p) \in \mathcal{G} \times \mathcal{P}$ has instead of the natural switch kernel $s_{\text{nat}}^g(p, \tilde{p})$ a natural switch rate $s_{\text{nat}}(g, p)$ combined with a probability measure $S_{\text{nat}}^{(g,p)}(d\tilde{p})$ on \mathcal{P} and instead of the induced switch kernel $s_{\text{ind}}^g(p, \tilde{p})(\hat{p})K^{-1}$ a induced switch kernel $s_{\text{ind}}((g, p), \hat{p})K^{-1}$ combined with a family of probability measure $\{S_{\text{ind}}^{((g,p),\hat{p})}(d\tilde{p})\}$ on \mathcal{P} .

2.1 Explicit construction of the population process with phenotypic plasticity

It is useful to give a pathwise description of ν^K in terms of Poisson point measures (cf. [17]). Let us recall this construction. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an abstract probability space. On this space, we define the following independent random elements:

- (i) a convergent sequence $(\nu_0^K)_{K \geq 1}$ of $\mathcal{M}^K(\mathcal{X})$ -valued random measures (the random initial population),
- (ii) $|\mathcal{X}|$ independent Poisson point measures $(N_{(g,p)}^{\text{birth}}(ds, di, d\theta))_{(g,p) \in \mathcal{X}}$ on $[0, \infty) \times \mathbb{N} \times \mathbb{R}_+$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) d\theta$,
- (iii) $|\mathcal{X}|$ independent Poisson point measures $(N_{(g,p)}^{\text{mut.}}(ds, di, d\theta, dx))_{(g,p) \in \mathcal{X}}$ on $[0, \infty) \times \mathbb{N} \times \mathbb{R}_+ \times \mathcal{X}$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) d\theta \sum_{\tilde{x} \in \mathcal{X}} \delta_{\tilde{x}}(dx)$.
- (iv) $|\mathcal{X}|$ independent Poisson point measures $(N_{(g,p)}^{\text{death}}(ds, di, d\theta))_{(g,p) \in \mathcal{X}}$ on $[0, \infty) \times \mathbb{N} \times \mathbb{R}_+$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) d\theta$,
- (v) $|\mathcal{X}|$ independent Poisson point measures $(N_{(g,p)}^{\text{switch}}(ds, di, d\theta, dp))_{(g,p) \in \mathcal{X}}$ on $[0, \infty) \times \mathbb{N} \times \mathbb{R}_+ \times \mathcal{P}$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) d\theta \sum_{\tilde{p} \in \mathcal{P}} \delta_{\tilde{p}}(dp)$,

Then, ν^K is given by the following equation

$$\begin{aligned}
 \nu_t^K &= \nu_0^K + \sum_{(g,p) \in \mathcal{X}} \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\{i \leq K \nu_{s-}^K(g,p), \theta \leq b(p)(1-u_K m(g))\}} \frac{1}{K} \delta_{(g,p)} N_{(g,p)}^{\text{birth}}(ds, di, d\theta) \\
 &+ \sum_{(g,p) \in \mathcal{X}} \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{\mathcal{X}} \mathbb{1}_{\{i \leq K \nu_{s-}^K(g,p), \theta \leq b(p)u_K m(g)M((g,p),x)\}} \frac{1}{K} \delta_x N_{(g,p)}^{\text{mut.}}(ds, di, d\theta, dx) \\
 &- \sum_{(g,p) \in \mathcal{X}} \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\{i \leq K \nu_{s-}^K(g,p), \theta \leq d(p) + \sum_{\tilde{p} \in \mathcal{P}} c(p,\tilde{p})\nu_{s-}^K(\tilde{p})\}} \frac{1}{K} \delta_{(g,p)} N_{(g,p)}^{\text{death}}(ds, di, d\theta) \\
 &+ \sum_{(g,p) \in \mathcal{X}} \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{\mathcal{P}} \mathbb{1}_{\{i \leq K \nu_{s-}^K(g,p), \theta \leq s_{\text{nat}}^g(p,\tilde{p}) + \sum_{\tilde{p} \in \mathcal{P}} s_{\text{ind}}^g(p,\tilde{p})(\hat{p})\nu_{s-}^K(\hat{p})\}} \\
 &\quad \times \frac{1}{K} (\delta_{(g,\tilde{p})} - \delta_{(g,p)}) N_{(g,p)}^{\text{switch}}(ds, di, d\theta, d\tilde{p}).
 \end{aligned}
 \tag{2.4}$$

Remark 2.3. This construction uses that \mathcal{X} is a discrete set and is in some sense closer to the definition given in [16] (p. 455). For non-discrete trait spaces the process can be constructed as in [17].

2.2 The law of large numbers

If the mutation rate is independent of K and the initial conditions converge to a deterministic limit, then the sequence of rescaled processes, $(\nu^K)_{K \geq 1}$, converges in probability as $K \uparrow \infty$ to the solution of a system of ODEs. This follows directly from the law of large numbers for density depending processes, see, e.g. Ethier and Kurtz [16], Chap. 11 or [15] Chap. 10. The following theorem gives a precise statement.

Theorem 2.4. Let $u_K \equiv 1$. Suppose that the initial conditions converge in probability to a deterministic limit, i.e. $\nu_0^K \xrightarrow{P} \nu_0$ as $K \uparrow \infty$, where ν_0 is a finite measure on \mathcal{X} . Then, for every $T > 0$, there exists a deterministic function $\xi \in C([0, T], \mathcal{M}_F(\mathcal{X}))$ such that, for all $\eta > 0$,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} \|\nu_t^K - \xi_t\|_{TV} > \eta \right] = 0, \tag{2.5}$$

where $\|\cdot\|_{TV}$ is the total variation norm. Moreover, let \mathbf{n} be the unique solution to the dynamical system

$$\begin{aligned} \dot{\mathbf{n}}_{(g,p)}(t) = & \mathbf{n}_{(g,p)}(t) \left((1 - m(g))b(p) - d(p) - \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{G} \times \mathcal{P}} c(p, \tilde{p}) \mathbf{n}_{(\tilde{g}, \tilde{p})}(t) \right. \\ & \left. - \sum_{\tilde{p} \in \mathcal{P}} \left(s_{\text{nat.}}^g(p, \tilde{p}) + \sum_{(\hat{g}, \hat{p}) \in \mathcal{G} \times \mathcal{P}} s_{\text{ind.}}^g(p, \tilde{p})(\hat{p}) \mathbf{n}_{(\hat{g}, \hat{p})}(t) \right) \right) \\ & + \sum_{\tilde{p} \in \mathcal{P}} \mathbf{n}_{(g, \tilde{p})}(t) \left(s_{\text{nat.}}^g(\tilde{p}, p) + \sum_{(\hat{g}, \hat{p}) \in \mathcal{G} \times \mathcal{P}} s_{\text{ind.}}^g(\tilde{p}, p)(\hat{p}) \mathbf{n}_{(\hat{g}, \hat{p})}(t) \right) \\ & + \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{G} \times \mathcal{P}} \mathbf{n}_{(\tilde{g}, \tilde{p})}(t) m(\tilde{g})b(\tilde{p})M((\tilde{g}, \tilde{p}), (g, p)), \quad (g, p) \in \mathcal{G} \times \mathcal{P}, \end{aligned} \tag{2.6}$$

with initial condition $\mathbf{n}_x(0) = \nu_0(x)$, for all $x \in \mathcal{X}$.

Then ξ is given as $\xi_t = \sum_{x \in \mathcal{X}} \mathbf{n}_x(t) \delta_x$.

Proof. This result follows from Theorem 2.1 in Chapter 11 of [16], since we can construct the processes as described in Remark 2.2 (i). Since these processes have the same laws, converge almost surely and the limit is deterministic, our processes converge in probability to the same limit. Alternatively, the theorem can be proven by using a large deviation principle on $[0, T]$ (cf. [15] Chap. 10 and Theorem 3.8) \square

Remark 2.5. If the trait spaces is not finite, one can obtain a similar result, cf. [17].

3 The interplay between rare mutations and fast switches

In this section we state our main results. As in previous work, we place ourselves under the assumptions

$$\forall V > 0, \quad \exp(-VK) \ll u_K \ll \frac{1}{K \ln K}, \quad \text{as } K \uparrow \infty, \tag{3.1}$$

which ensure that a population reaches equilibrium before a new mutant appears and that mutations occur before the population dies out. Under these assumptions, we prove that the individual-based process with phenotypic plasticity converges to a generalisation of the PES. Let us start with describing the techniques used in [11].

The key element in the proof of the convergence to the PES used by Champagnat and Méléard [11] is a precise analysis of how a mutant population fixates. A crucial assumption in [11] is that the competitive Lotka-Volterra systems that describes the large population limit always have a unique stable fixed point $\bar{\mathbf{n}}$. Thus, the main task is to study the invasion of a mutant that has just appeared in a population close to equilibrium. The invasion can be divided into three steps: First, as long as the mutant population size is smaller than $K\epsilon$, for a fixed small $\epsilon > 0$, the resident population stays close to its equilibrium. Therefore, the mutant population can be approximated by a branching process. Second, once the mutant population reaches the level $K\epsilon$, the whole system is close to the solution of the corresponding deterministic system and reaches

an ϵ -neighbourhood of \bar{n} in finite time. Third, the subpopulations which have a zero coordinate in \bar{n} can be approximated by subcritical branching processes until they die out.

The first and third steps require a time of order $\ln(K)$, whereas the second step requires only a time of order one, independent of K . Since the expected time between two mutations is of order $1/(u_K K)$, the upper bound on u_K in (3.1) guarantees that, with high probability, the three steps of an invasion are completed before a new mutation occurs.

In the first invasion step the invasion fitness of a mutant plays a crucial role. Given a population in a stable equilibrium that populates a certain set of traits, say $M \subset \mathcal{X}$, the invasion fitness $f(x, M)$ is the growth rate of a population consisting of a single individual with trait $x \notin M$ in the presence of the equilibrium population \bar{n} on M . In the case of the standard model, it is given by

$$f(x, M) = b(x) - d(x) - \sum_{y \in M} c(x, y) \bar{n}_y. \tag{3.2}$$

Positive $f(x, M)$ implies that a mutant appearing with trait x from the equilibrium population on M has a positive probability (uniformly in K) to grow to a population of size of order K ; negative invasion fitness implies that such a mutant population will die out with probability tending to one (as $K \uparrow \infty$) before it can reach a size of order K . The reason for this is that the branching process (birth-death process) which approximates the mutant population is supercritical if $f(x, M)$ is positive and subcritical if $f(x, M)$ is negative.

In order to describe the dynamics of a phenotypically heterogeneous population on the evolutionary time scale, we have to adapt the notion of invasion fitness to the case where fast phenotypic switches are present. Since switches between phenotypes associated to the same genotype happen at times of order one, the growth rate of the initial mutant phenotype does not determine the probability of fixation. See [12] for a similar issue in a model with sexual reproduction. In the proof of Theorem 3.6, we approximate the dynamics of the mutant population by a *multi-type branching process* until it reaches a size $K\epsilon$ (or dies out). A continuous-time multi-type branching process is supercritical if and only if the largest eigenvalue of its infinitesimal generator is strictly positive (cf. [2, 27]). Therefore, this eigenvalue will provide an appropriate generalisation of the invasion fitness.

3.1 The competitive Lotka-Volterra system with phenotypic plasticity

We first consider the large population limit without mutation ($u_K \equiv 0$). Assume that the initial condition is supported on d traits, $(\mathbf{g}, \mathbf{p}) = ((g_1, p_1), \dots, (g_d, p_d)) \in (\mathcal{G} \times \mathcal{P})^d$, and that the sequence of the initial conditions converges in distribution

$$\lim_{K \uparrow \infty} \nu_0^K = \sum_{i=1}^d n_i(0) \delta_{(g_i, p_i)}, \quad \text{in law,} \quad \text{where } n_i(0) > 0, \text{ for all } i \in \{1 \dots d\}. \tag{3.3}$$

By Theorem 2.4, for every $T > 0$, the sequences of processes $\nu^K \in \mathbb{D}([0, T], \mathcal{M}^K(\mathcal{X}))$ generated by \mathcal{L}^K with initial state ν_0^K converges in distribution, as $K \uparrow \infty$, to a deterministic function $\xi \in C([0, T], \mathcal{M}(\mathcal{X}))$. Since $u_K \equiv 0$, no new genotype can appear in the population process ν^K . Moreover, not every genotype can express every phenotype. Let us describe the support of ν_t^K more precisely.

For all $g \in \mathcal{G}$, let X^g be a stationary discrete-time Markov chain with state space \mathcal{P} and transition probabilities

$$\mathbb{P}[X_i^g = \tilde{p} \mid X_{i-1}^g = p] = \frac{s^g(p, \tilde{p})}{\sum_{\hat{p} \in \mathcal{P}} s^g(p, \hat{p})}, \quad \text{if } \sum_{\hat{p} \in \mathcal{P}} s^g(p, \hat{p}) > 0, \quad (3.4)$$

and

$$\mathbb{P}[X_i^g = p \mid X_{i-1}^g = p] = 1, \quad \text{if } \sum_{\hat{p} \in \mathcal{P}} s^g(p, \hat{p}) = 0. \quad (3.5)$$

The Markov chains $\{X^g, g \in \mathcal{G}\}$ contain only partial information on the switching behaviour of the process ν^K , but we see that this is the key information needed later.

In the sequel we work under the following simplifying assumption:

Assumption 3.1. For all $g \in \mathcal{G}$, all communicating classes of X^g are recurrent.

We denote the communicating class associated with $(g, p) \in \mathcal{G} \times \mathcal{P}$ by $[p]_g$. This is the communicating class of X^g which contains p , i.e. p can be seen as a representative of the class, which has an equivalence relation depending on g . By Assumption 2.1, this ensures

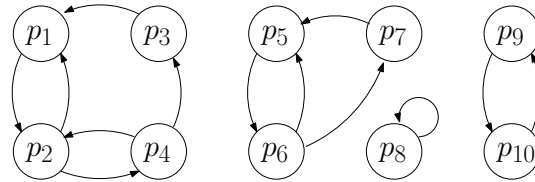


Figure 1: Example of a Markov chain X^g . Here, $\mathcal{P} = \{p_1, \dots, p_{10}\}$ and X^g has four communicating classes: $\{p_1, p_2, p_3, p_4\}$, $\{p_5, p_6, p_7\}$, $\{p_8\}$, $\{p_9, p_{10}\}$. The class $\{p_8\}$ has only one element, i.e. $\sum_{i=1}^{10} s^g(p_8, p_i) = 0$ in this example.

that if we start with a large enough population consisting only of individuals carrying the same trait (g, p) , then, after a time of order 1, all phenotypes in the class $[p]_g$ will be present in the population, but none of the other classes. Observe that these Markov chains do not describe the dynamics of the whole process. If we allowed transient states this would not imply that the trait would get extinct, since its growth rate could be larger than the switching rate.

Thus, $[p]_g$ is the set of phenotypes which are reachable in the Markov chain X^g with $X_0^g = p$ and the set of traits which can appear in the population process ν^K is given by

$$\mathcal{X}_{(\mathbf{g}, \mathbf{p})} \equiv \bigcup_{i=1}^d \{g_i\} \times [p_i]_{g_i}. \quad (3.6)$$

With this notation, ξ is given by $\xi(t) = \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathbf{n}_x(t) \delta_x$, where \mathbf{n} is the solution of the competitive Lotka-Volterra system with phenotypic plasticity defined below.

Definition 3.2. For any $(\mathbf{g}, \mathbf{p}) \in (\mathcal{G} \times \mathcal{P})^d$, we denote by $LVS(d, (\mathbf{g}, \mathbf{p}))$ the competitive Lotka-Volterra system with phenotypic plasticity. This is an $|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|$ -dimensional system of ODEs given by

$$\begin{aligned} \dot{\mathbf{n}}_{(g,p)} = & \mathbf{n}_{(g,p)} \left(b(p) - d(p) - \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(p, \tilde{p}) \mathbf{n}_{(\tilde{g}, \tilde{p})} \right. \\ & \left. - \sum_{\tilde{p} \in [p]_g} \left(s_{\text{nat.}}^g(p, \tilde{p}) + \sum_{(\hat{g}, \hat{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind.}}^g(p, \tilde{p})(\hat{p}) \mathbf{n}_{(\hat{g}, \hat{p})}(t) \right) \right) \\ & + \sum_{\tilde{p} \in [p]_g} \mathbf{n}_{(g, \tilde{p})} \left(s_{\text{nat.}}^g(\tilde{p}, p) + \sum_{(\hat{g}, \hat{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind.}}^g(\tilde{p}, p)(\hat{p}) \mathbf{n}_{(\hat{g}, \hat{p})}(t) \right), \quad (g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}. \quad (3.7) \end{aligned}$$

We choose the (possibly misleading) name *competitive Lotka-Volterra system with phenotypic plasticity* to emphasise that we add phenotypic plasticity (induced by switching rates) in the usual competitive Lotka-Volterra system. Note, however, that the system *LVS* is *not* a system of Lotka-Volterra equations.

We now introduce the notation of coexisting traits in this context (cf. [11]).

Definition 3.3. For any $d \geq 2$, we say that the distinct traits $(g_1, p_1), \dots, (g_d, p_d)$ coexist if the system $LVS(d, (\mathbf{g}, \mathbf{p}))$ has a unique non-trivial equilibrium $\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p}) \in (0, \infty)^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$ which is locally strictly stable, meaning that all eigenvalues of the Jacobian matrix of the system $LVS(d, (\mathbf{g}, \mathbf{p}))$ at $\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})$ have strictly negative real parts.

Note that if $(g_1, p_1), \dots, (g_d, p_d)$ coexist, then all traits of $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ coexist and the equilibrium $\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})$ is asymptotically stable. We will prove later that if the traits $(g_1, p_1), \dots, (g_d, p_d)$ coexist, then the *invasion probability* of a mutant trait (\tilde{g}, \tilde{p}) which appears in the resident population $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ close to $\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})$ is given by the function

$$1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}), \tag{3.8}$$

where $q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p})$ is given as follows: Let us denote the elements of $[\tilde{p}]_{\tilde{g}}$ by $\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_{|\tilde{p}]_{\tilde{g}}|}$ and assume without loss of generality that $\tilde{p} = \tilde{p}_1$. Then, $q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p})$ is the first component of the smallest nonnegative solution of

$$\mathbf{u}(\mathbf{y}) = \mathbf{0}, \tag{3.9}$$

where \mathbf{u} is a map from $\mathbb{R}^{|\tilde{p}]_{\tilde{g}}|}$ to $\mathbb{R}^{|\tilde{p}]_{\tilde{g}}|}$ defined, for all $i \in \{1, \dots, |\tilde{p}]_{\tilde{g}}|\}$, by

$$\begin{aligned} u_i(\mathbf{y}) &\equiv b(\tilde{p}_i) y_i^2 \\ &+ \sum_{j=1}^{|\tilde{p}]_{\tilde{g}}|} \left(s_{\text{nat.}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind.}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j)(p) \bar{\mathbf{n}}_{(g,p)} \right) y_j + d(\tilde{p}_i) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_i, p) \bar{\mathbf{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) \\ &- \left(b(\tilde{p}_i) + \sum_{j=1}^{|\tilde{p}]_{\tilde{g}}|} \left(s_{\text{nat.}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind.}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j)(p) \bar{\mathbf{n}}_{(g,p)} \right) + d(\tilde{p}_i) \right. \\ &\left. + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_i, p) \bar{\mathbf{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) \right) y_i. \end{aligned} \tag{3.10}$$

In fact, $(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}))$ is the (asymptotic as $K \uparrow \infty$) probability that a single mutant reaches a population size of order K in a resident population with traits $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$. We obtain this by approximating the mutant population with multi-type branching processes (cf. proof of Thm. 3.11). The function $(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}))$ plays the same role as the function $[f(y; \mathbf{x})]_+ / b(y)$ in the standard case (cf. [11]).

To obtain that the process jumps on the evolutionary time scale from one equilibrium to the next, we need an assumption to prevent cycles, unstable equilibria or chaotic dynamics in the deterministic system (cf. [11] Ass. B).

Assumption 3.4. For any given traits $(g_1, p_1), \dots, (g_d, p_d) \in \mathcal{G} \times \mathcal{P}$ that coexist, and for any mutant trait $(\tilde{g}, \tilde{p}) \in \mathcal{X} \setminus \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ such that $1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) > 0$, there exists a neighbourhood $U \subset \mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}| + |\tilde{p}]_{\tilde{g}}|}$ of $(\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p}), 0, \dots, 0)$, such that all solutions of $LVS(d + 1, ((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})))$ with initial condition in $U \cap (0, \infty)^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}| + |\tilde{p}]_{\tilde{g}}|}$ converge as $t \uparrow \infty$ to a unique locally strictly stable equilibrium in $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}| + |\tilde{p}]_{\tilde{g}}|}$ denoted by $\mathbf{n}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$.

We write \mathbf{n}^* and not $\bar{\mathbf{n}}$ to emphasise that some components of \mathbf{n}^* can be zero. We use the shorthand notation $((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$, for $((g_1, p_1), \dots, (g_d, p_d), (\tilde{g}, \tilde{p}))$. Assumption 3.4 does not have to hold for all traits in $\mathcal{X} \setminus \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$, but only for those traits (\tilde{g}, \tilde{p}) which can appear in the resident population by mutation, i.e. only if $\sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} m(g) M((g, p), (\tilde{g}, \tilde{p}))$ is positive.

Remark 3.5. It is possible to extend the definitions and assumptions for the study of rare mutations and fast switches in populations with *non-discrete trait* space if one assumes that an individual can change its phenotype only to finitely many other phenotypes. This must be encoded in the switching kernels. More precisely, for all $(g, p) \in \mathcal{G} \times \mathcal{P}$ the communicating class $[p]_g$ should contain finitely many elements.

3.2 Convergence to the generalised polymorphic evolution sequence

In this subsection we state the main theorem of this paper and give the general idea of the proof illustrated by an example.

Theorem 3.6. *Suppose that the Assumptions 2.1, 3.1 and 3.4 hold. Fix $(g_1, p_1), \dots, (g_d, p_d) \in \mathcal{G} \times \mathcal{P}$ coexisting traits and assume that the initial conditions have support $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ and converge almost surely to $\bar{n}(\mathbf{g}, \mathbf{p})$, i.e. $\lim_{K \uparrow \infty} \nu_0^K = \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_x(\mathbf{g}, \mathbf{p}) \delta_x$ a.s.. Furthermore, assume that*

$$\forall V > 0, \quad \exp(-VK) \ll u_K \ll \frac{1}{K \ln(K)}, \quad \text{as } K \uparrow \infty. \tag{3.11}$$

Then, the sequence of the rescaled processes $(\nu_{t/Ku_K}^K)_{t \geq 0}$, generated by \mathcal{L}^K with initial state ν_0^K , converges in the sense of finite dimensional distributions to the measure-valued pure jump process Λ defined as follows: $\Lambda_0 = \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) \delta_{(g,p)}$ and the process Λ jumps, for all $(\hat{g}, \hat{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$, from

$$\sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) \delta_{(g,p)} \quad \text{to} \quad \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p}), (\hat{g}, \hat{p})}} \mathbf{n}_{(g,p)}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) \delta_{(g,p)} \tag{3.12}$$

with infinitesimal rate

$$m(\hat{g})b(\hat{p})\bar{n}_{(\hat{g}, \hat{p})}(\mathbf{g}, \mathbf{p})(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}))M((\hat{g}, \hat{p}), (\tilde{g}, \tilde{p})). \tag{3.13}$$

- Remark 3.7.** (i) The convergence cannot hold in law with respect to the Skorokhod topology (cf. [9]). It holds only in the sense of finite dimensional distributions on $\mathcal{M}_F(\mathcal{X})$, the set of finite positive measures on \mathcal{X} equipped with the topology of the total variation norm.
- (ii) The process Λ is a generalised version of the usual PES. Therefore, we call Λ Polymorphic Evolution Sequence with phenotypic Plasticity (PESP).
- (iii) Assumption 3.4 is essential for this statement. In the case when the dynamical system has multiple attractors and different points near the initial state lie in different basins of attraction, it is not clear and may be random which attractor the system approaches. The characterisation of the asymptotic behaviour of the dynamical system is needed to describe the final state of the stochastic process. This is in general a difficult and complex problem, which is not doable analytically and requires numerical analysis. Thus, we restrict ourselves to the Assumption 3.4.

We describe in the following the general idea of the proof, which is quite similar to the one given in [11]. The population is either in a stable phase or in an invasion phase. Until the first mutant appears, the population is in a stable phase, i.e. the population stays close to a given equilibrium. From the first mutational event until the population reaches again a stable state, the population is in an invasion phase. In fact, the mutant either survives and the population reaches fast a new stable state (where the mutant trait is present), or the mutant goes extinct and the population is again in the old stable state. After this, the population is again in a stable phase until the next mutation, etc..

Note that we prove in the following that the invasion phases are relatively short ($O(\ln(K))$) compared to the stable phase ($O(1/u_K K)$). Since we study the process on

the time scale $1/Ku_K$, the limit process proceeds as a pure jump process which jumps from one stable state to another.

The stable phase: Fix $\epsilon > 0$. Let $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ be the support of the initial conditions. For large K , the population process ν^K is, with high probability, still in a small neighbourhood of the equilibrium $\bar{n}(\mathbf{g}, \mathbf{p})$ when the first mutant appears. In fact, using large deviation results on the problem of exit from a domain (cf. [18]), we obtain that there exists a constant $M > 0$ such that the first time ν^K leave the $M\epsilon$ -neighbourhood of $\bar{n}(\mathbf{g}, \mathbf{p})$ is bigger than $\exp(VK)$, for some $V > 0$, with high probability. Thus, until this stopping time, mutations born from individuals with trait $x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ appear with a rate which is close to

$$u_K m(x) b(x) K \bar{n}_x(\mathbf{g}, \mathbf{p}).$$

The condition $1/(Ku_K) \ll \exp(VK)$, for all $V > 0$, in (3.11) ensures that the first mutation appears before this exit time.

The invasion phase: We divide the invasion of a given mutant trait (\tilde{g}, \tilde{p}) into *three steps*, as in [9] and [11] (cf. Fig. 2). In the first step, from a mutational event until the mutant population goes extinct or the mutant density reaches the value ϵ , the number of mutant individuals is small (cf. Fig. 2, $[0, t_1]$). Thus, applying a perturbed version of the large deviation result we used in the first phase, we obtain that the resident population stays close to its equilibrium density $\bar{n}(\mathbf{g}, \mathbf{p})$ during this step. Using similar arguments as Champagnat et al. [9, 11], we prove that the mutant population is well approximated by a $|\tilde{p}_{\tilde{g}}|$ -type branching process Z , as long as the mutant population has less than ϵK individuals. More precisely, let us denote the elements of $[\tilde{p}]_{\tilde{g}}$ by $\tilde{p}_1, \dots, \tilde{p}_{|\tilde{p}_{\tilde{g}}|}$, then, for each $1 \leq i \leq |\tilde{p}_{\tilde{g}}|$, each individual in Z (carrying trait (\tilde{g}, \tilde{p}_i)) undergoes

- (i) birth (without mutation) with rate $b(\tilde{p}_i)$,
- (ii) death with rate $d(\tilde{p}_i) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_i, p) \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p})$ and
- (iii) switch to \tilde{p}_j with rate $s_{\text{nat}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j)(p) \bar{n}_{(g,p)}$, for all $1 \leq j \leq |\tilde{p}_{\tilde{g}}|$.

This continuous-time multi-type branching process is supercritical if and only if the largest eigenvalue of its infinitesimal generator, which we denote by λ_{\max} , is larger than zero. Hence, the mutant invades with positive probability if and only if $\lambda_{\max} > 0$. Moreover, the probability that the density of the mutant's genotype, $\nu^K(\tilde{g})$, reaches ϵ at some time t_1 is close to the probability that the multi-type branching process reaches the total mass ϵK , which converges as $K \uparrow \infty$ to $(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}))$.

In the second step, we obtain as a consequence of Theorem 2.4 that once the mutant density has reached ϵ , for large K , the stochastic process ν^K can be approximated on any finite time interval by the solution of $LVS(d + 1, ((g_1, p_1), \dots, (g_d, p_d), (\tilde{g}, \tilde{p})))$ with a given initial state. By Assumption 3.4, this solution reaches the ϵ -neighbourhood of its new equilibrium $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ in finite time. Therefore, for large K , the stochastic process ν^K also reaches with high probability the ϵ -neighbourhood of $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ at some finite (K independent) time t_2 .

In the third step, we use similar arguments as in the first step. Since $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ is a strongly locally stable equilibrium (Ass. 3.4), the stochastic process ν_t^K stays close $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ and we can approximate the densities of the traits $(g, p) \in \mathcal{X}_{((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))}$ with $n_{(g,p)}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) = 0$ by $|\tilde{p}_{\tilde{g}}|$ -type branching processes which are subcritical and therefore become extinct, a.s..

The duration of the first and third step are proportional to $\ln(K)$, whereas the time of the second step is bounded. Thus, the second inequality in (3.11) guarantees that, with high probability, the three steps of invasion are completed before a new mutation

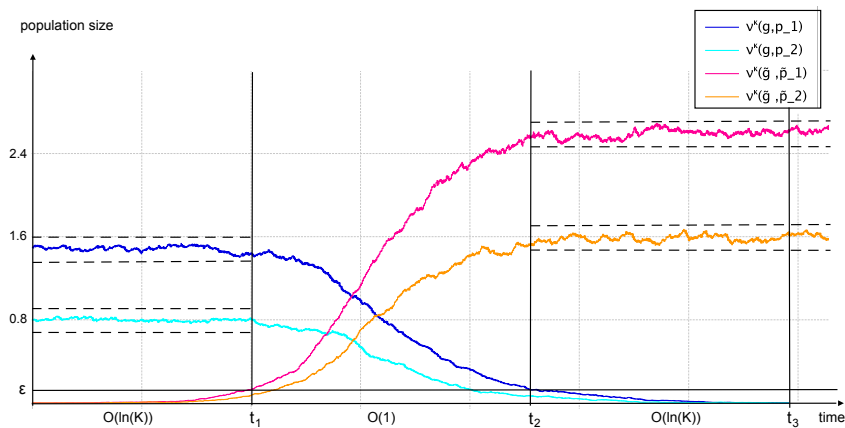


Figure 2: The three steps of one invasion phase.

occurs. After the last step the process is again back in a stable phase, but with a possibly different resident population, until the next mutation happens.

An example: Figure 2 shows the invasion phase of a single mutant with trait (\tilde{g}, \tilde{p}_1) , which appeared (at time 0) in a population close to $\bar{n}(g, p)$ (indicated by the dashed lines). In this example the resident population consists of two coexisting traits (g, p_1) and (g, p_2) and the mutant individuals can switch to one other phenotype only, i.e. $[\tilde{p}_1]_{\tilde{g}} = \{\tilde{p}_1, \tilde{p}_2\}$. The parameters of the simulation of Figure 2 are given in Table 1. The stable fixed point of

Table 1: Parameters of Figure 2

$b(p_1) = 3$	$d(p_1) = 1$	$c(p_1, p_1) = 1$	$c(p_1, p_2) = 0.7$	$c(p_1, \tilde{p}_1) = 0.7$	$c(p_1, \tilde{p}_2) = 0.7$	$s_{\text{nat}}^g(p_1, p_2) = 1$
$b(p_2) = 3$	$d(p_2) = 1$	$c(p_2, p_1) = 0.7$	$c(p_2, p_2) = 1$	$c(p_2, \tilde{p}_1) = 0.7$	$c(p_2, \tilde{p}_2) = 0.7$	$s_{\text{nat}}^g(p_2, p_1) = 2$
$b(\tilde{p}_1) = 5$	$d(\tilde{p}_1) = 1$	$c(\tilde{p}_1, p_1) = 0.7$	$c(\tilde{p}_1, p_2) = 0.7$	$c(\tilde{p}_1, \tilde{p}_1) = 1$	$c(\tilde{p}_1, \tilde{p}_2) = 0.7$	$s_{\text{nat}}^g(\tilde{p}_1, \tilde{p}_2) = 1.5$
$b(\tilde{p}_2) = 4$	$d(\tilde{p}_2) = 1$	$c(\tilde{p}_2, p_1) = 0.7$	$c(\tilde{p}_2, p_2) = 0.7$	$c(\tilde{p}_2, \tilde{p}_1) = 0.7$	$c(\tilde{p}_2, \tilde{p}_2) = 1$	$s_{\text{nat}}^g(\tilde{p}_2, \tilde{p}_1) = 2$
$K = 2000$	$u_K = 0$	$u_0^K(g, p_1) = 1.5$	$u_0^K(g, p_2) = 0.8$	$u_0^K(\tilde{g}, \tilde{p}_1) = 1/K$	$u_0^K(\tilde{g}, \tilde{p}_2) = 0$	$s_{\text{ind}}(\dots)(\cdot) \equiv 0$

the system $LVS(2, ((g, p_1), (g, p_2)))$ is $\bar{n}((g, p_1), (g, p_2)) \approx (1.507, 0.809)$. The infinitesimal generator of the multi-type branching process that approximates the mutant population in the first step is approximately

$$\begin{pmatrix} 0.879 & 1.5 \\ 2 & -0.621 \end{pmatrix}. \tag{3.14}$$

Since the largest eigenvalue of this matrix is positive (≈ 2.016), the mutant population reaches with positive probability the second invasion step (cf. Fig. 2).

Moreover, $n^* \approx (0, 0, 2.608, 1.608)$ is the unique locally strictly stable fixed point of the dynamical system $LVS(4, ((g, p_1), (g, p_2), (\tilde{g}, \tilde{p}_1), (\tilde{g}, \tilde{p}_2)))$. The dynamical system and hence also the stochastic process reach in finite time the ϵ -neighbourhood of this value. The infinitesimal generator of the multi-type branching process that approximates the resident population in the third step is approximately

$$\begin{pmatrix} -1.951 & 2 \\ 1 & -2.951 \end{pmatrix}. \tag{3.15}$$

The largest eigenvalue of this matrix is negative (≈ -0.951), meaning that the process is subcritical and goes extinct, a.s.. Therefore, there exists a time t_3 such that all individuals which carry trait (g, p_1) or (g, p_2) are a.s. dead at time t_3 .

3.3 Proof of Theorem 3.6

In this paragraph we prove the convergence to the PES. (The proof uses the same arguments and techniques as [11], which were developed in [9]. However, some extensions are necessary if fast phenotypic switches are included in the process, which we state and prove in this subsection.) We start with an analog of Theorem 3 of [9]. Part (i) of the following theorem strengthens Theorem 2.4, and part (ii) provides control of exit from an attractive domain in the polymorphic case with phenotypic plasticity.

Theorem 3.8. (i) Assume that the initial conditions have support $\{(g_1, p_1), \dots, (g_d, p_d)\}$ and are uniformly bounded, i.e., for all $1 \leq i \leq d$, $\nu_0^K(g_i, p_i) \in A$, where A is a compact subset of $\mathbb{R}_{>0}$ ($\mathbb{R}_{>0}$ denotes the set of strictly positive real numbers). Then, for all $T > 0$ and for all $\eta > 0$,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathbf{n}_x(t, \nu_0^K) \delta_x \right\|_{TV} > \eta \right] = 0, \tag{3.16}$$

where $\mathbf{n}(t, \nu_0^K) \in \mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$ denotes the value of the solution of $LVS(d, (\mathbf{g}, \mathbf{p}))$ at time t with initial condition $\mathbf{n}_x(0, \nu_0^K) = \nu_0^K(x)$, for all $x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$. Note that the measure $\sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathbf{n}_x(t, \nu_0^K) \delta_x$ depends on K , since the initial condition and hence the solution of $LVS(d, (\mathbf{g}, \mathbf{p}))$ depends on K .

(ii) Let $(g_1, p_1), \dots, (g_d, p_d) \in \mathcal{X}$ coexist. Assume that, for any $K \geq 1$, $\text{Supp}(\nu_0^K) = \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$. Let $\tau_{mut.}$ be the first mutation time. Define the first exit time from the ξ -neighbourhood of $\bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p})$ by

$$\theta_{exit}^{K, \xi} \equiv \inf \{ t \geq 0 : \exists x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})} : |\nu_t^K(x) - \bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p})| > \xi \}. \tag{3.17}$$

Then there exist $\epsilon_0 > 0$ and $M > 0$ such that, for all $\epsilon < \epsilon_0$, there exists $V > 0$ such that if the initial state of ν^K lies in the ϵ -neighbourhood of $\bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p})$, the probability that $\theta_{exit}^{K, M\epsilon}$ is larger than $e^{KV} \wedge \tau_{mut.}$ converges to one, i.e.

$$\lim_{K \uparrow \infty} \sup_{\mathbf{n}^K \in (\mathbb{N}/K)^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|} \cap B_\epsilon(\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p}))} \mathbb{P} \left[\theta_{exit}^{K, M\epsilon} < e^{KV} \wedge \tau_{mut.} \mid \nu_0^K(x) = n_x^K, \right. \\ \left. \text{for all } x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})} \right] = 0, \tag{3.18}$$

where $\mathbf{n}^K \equiv (n_x^K)_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}}$ and $B_\epsilon(\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p}))$ denotes the ϵ -neighbourhood of $\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})$.

Moreover, (3.18) also holds if, for all $(g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$, the total death rate of an individual with trait (g, p) ,

$$d(p) + \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(p, \tilde{p}) \nu_t^K(\tilde{g}, \tilde{p}), \tag{3.19}$$

and the total switch rates of an individual with trait (g, p) ,

$$s_{nat.}^g(p, p_i) + \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{ind.}^g(p, p_i)(\tilde{p}) \nu_t^K(\tilde{g}, \tilde{p}) \text{ for all } p_i \in [p]_g, \tag{3.20}$$

are perturbed by additional random processes that are uniformly bounded by $\bar{c}\epsilon$ respectively $\bar{s}_{ind.}\epsilon$, where \bar{c} and $\bar{s}_{ind.}$ are upper bounds for the parameters of competition and induced switch.

Remark 3.9. (i) One consequence of the second part of (ii) is that, with high probability, the process stays in the $M\epsilon$ -neighbourhood of $\bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p})$ until the first time that a mutant's density reaches the value ϵ . In other words, let $\theta_{Invasion}^K$ denote the

first time that a mutant's density reaches the value ϵ , i.e

$$\theta_{\text{Invasion}}^K \equiv \left\{ t \geq 0 : \exists (g, p) \notin \mathcal{X}_{(\mathbf{g}, \mathbf{p})} : \sum_{\tilde{p} \in [p]_g} \nu_t^K(g, \tilde{p}) \geq \epsilon \right\}. \quad (3.21)$$

Then, the probability that $\theta_{\text{exit}}^{K, M\epsilon}$ is larger than $e^{KV} \wedge \theta_{\text{Invasion}}^K$ converges to one. We use this result also for the third invasion step.

- (ii) Since $\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})$ is a locally strictly stable fixed point of the system $LVS(d, (\mathbf{g}, \mathbf{p}))$, there exists a constant $M > 0$ such that, for all $\epsilon > 0$ small enough, for all trajectories $\mathbf{n}(t)$ with $\|\mathbf{n}(0) - \bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})\| < \epsilon$, it holds that $\sup_{t \geq 0} \|\mathbf{n}(t) - \bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})\| < M\epsilon$.

Proof. The main task to prove (i) is to show that a large deviation principle on $[0, T]$ holds for a slightly modify process and that the ν^K has the same law on the random time interval we need to control it. In fact, Theorem 10.2.6 of [15] can be applied to obtain the large deviation principle. The main task to prove (ii) is to show that the classical estimates for exit times from a domain (cf. [18]) for the jump process ν^K can be used. Note that Freidlin and Wentzell study in [18] mainly small white noise perturbations of dynamical systems. However, there also are some comments on the generalisation to dynamical systems with small jump-like perturbations (cf. [18], Sec. 5.4) \square

The following Lemma describes the asymptotic behaviour of τ_{mut} and can be seen as an extension of Lemma 2 of [9] or Lemma A.3 of [11].

Lemma 3.10. *Let $(g_1, p_1), \dots, (g_d, p_d) \in \mathcal{X}$ coexist. Assume that, for any $K \geq 1$, $\text{Supp}(\nu_0^K) = \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$. Let τ_{mut} denote the first mutation time. Then, there exists $\epsilon_0 > 0$ such that if the initial states of ν^K belong to the ϵ_0 -neighbourhood of $\bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p})$, then, for all $\epsilon \in (0, \epsilon_0)$,*

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\tau_{\text{mut}} > \ln(K), \sup_{t \in [\ln(K), \tau_{\text{mut}}]} \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p}) \delta_x \right\|_{TV} < \epsilon \right] = 1, \quad (3.22)$$

Moreover, $(\tau_{\text{mut}} u_K K)_{K \geq 1}$ converges in law to an exponential distributed random variable with parameter $\sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} m(g)b(p)\bar{\mathbf{n}}_{(g,p)}(\mathbf{g}, \mathbf{p})$ and the probability that the mutant, which appears at time τ_{mut} , is born from an individual with trait $(g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ converges to

$$\frac{m(g)b(p)\bar{\mathbf{n}}_{(g,p)}(\mathbf{g}, \mathbf{p})}{\sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} m(\tilde{g})b(\tilde{p})\bar{\mathbf{n}}_{(\tilde{g}, \tilde{p})}(\mathbf{g}, \mathbf{p})} \quad (3.23)$$

as $K \uparrow \infty$.

Proof. There exist constants $C > 0$ and $V > 0$, such that on the time interval $[0, \exp(KV)]$, with probability tending to one, the total mass of the population, $\nu_t^K(\mathcal{X})$, is bounded from above by C . Therefore, we can construct an exponential random variable A with parameter $C'Ku_K$, where $C' = C \max_{g \in \mathcal{G}, p \in \mathcal{P}} m(g)b(p)$, such that

$$A \leq \tau_{\text{mut}} \quad \text{on the event} \quad \{\tau_{\text{mut}} < \exp(KV)\}. \quad (3.24)$$

Thus, $\mathbb{P}[\tau_{\text{mut}} > \ln(K)] \geq \mathbb{P}[A > \ln(K)] = e^{-C' \ln(K)Ku_K}$. Since (3.11) implies that $\ln(K)Ku_K$ converges to zero as $K \uparrow \infty$, we get $\lim_{K \uparrow \infty} \mathbb{P}[\tau_{\text{mut}} > \ln(K)] = 1$.

The fixed point $\bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})$ is asymptotic stable. Thus, $\exists \epsilon_0 > 0 : \forall \tilde{\epsilon} \in (0, \epsilon_0) \exists T(\tilde{\epsilon})$:

$$\|\mathbf{n}(\mathbf{g}, \mathbf{p})(0) - \bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})\| < \epsilon_0, \quad \text{implies} \quad \sup_{t \geq T(\tilde{\epsilon})} \|\mathbf{n}(\mathbf{g}, \mathbf{p})(t) - \bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})\| < \tilde{\epsilon}/2. \quad (3.25)$$

In words, there exists a finite time $T(\tilde{\epsilon})$ such that all trajectories, which start in the ϵ_0 neighbourhood of the fixed point, stay after $T(\tilde{\epsilon})$ in the $\tilde{\epsilon}/2$ -neighbourhood of the fixed point.

Next, we apply the last theorem: By (i), for all $\tilde{\epsilon} \in (0, \epsilon_0) \exists T(\tilde{\epsilon})$ such that, for K large enough,

$$\left\| \nu_{T(\tilde{\epsilon})}^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_x(\mathbf{g}, \mathbf{p}) \delta_x \right\|_{TV} < \tilde{\epsilon} \quad a.s.. \tag{3.26}$$

Then, by (ii), there exist $\epsilon_0 > 0$ and $M > 0$: for all $\tilde{\epsilon} \in (0, \epsilon_0)$, there exists $V > 0$ such that

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\sup_{t \in [T(\tilde{\epsilon}), e^{KV} \wedge \tau_{mut.})} \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_x(\mathbf{g}, \mathbf{p}) \delta_x \right\|_{TV} < M\tilde{\epsilon} \right] = 1. \tag{3.27}$$

Moreover, for all $\tilde{\epsilon} \in (0, \epsilon_0)$ there exists $K_0 \in \mathbb{N}$ such that $T(\tilde{\epsilon}) < \ln(K)$, for all $K \geq K_0$. Thus, setting $\epsilon = M\tilde{\epsilon}$, ends the proof of (3.22), provided that $\lim_{K \uparrow \infty} \mathbb{P}[\tau_{mut.} < e^{KV}] = 1$.

Again, we can construct, for all $\epsilon > 0$, two exponential random variables $A^{1,K,\epsilon}$ and $A^{2,K,\epsilon}$ with parameters

$$a_1 u_K K \equiv \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} u_K m(g) b(p) (\bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) + \epsilon) K \tag{3.28}$$

and

$$a_2 u_K K \equiv \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} u_K m(g) b(p) (\bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) - \epsilon) K \tag{3.29}$$

such that

$$A^{1,K,\epsilon} \leq \tau_{mut.} \leq A^{2,K,\epsilon} \quad \text{on the event } \{T(\tilde{\epsilon}) < \tau_{mut.} < e^{KV}\}, \tag{3.30}$$

where $T(\tilde{\epsilon})$ is the time defined in equation (3.26) and $\tilde{\epsilon} = \epsilon/M$. Moreover, we have

$$\lim_{K \uparrow \infty} \mathbb{P}[\tau_{mut.} < \ln(K)] = 0 \quad \text{and} \quad \lim_{K \uparrow \infty} \mathbb{P}[A^{2,K,\epsilon} > e^{KV}] = 0, \tag{3.31}$$

because $u_K K e^{KV} \uparrow \infty$ as $K \uparrow \infty$. Therefore, for all $\epsilon > 0$, the probability of the event $\{T(\tilde{\epsilon}) < \tau_{mut.} < e^{KV}\}$ converges to one as K goes to infinity. Moreover, the random variables $A^{1,K,\epsilon} u_K K$ and $A^{2,K,\epsilon} u_K K$ converge both in law to the same exponential distributed random variable with parameter

$$\sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} m(g) b(p) \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) \tag{3.32}$$

as first $K \uparrow \infty$ and then $\epsilon \rightarrow 0$. The random variables $A, A^{1,K,\epsilon}$ and $A^{2,K,\epsilon}$ can easily be constructed by using the pathwise description of ν^K (cf. [3] or [10]). \square

Theorem 3.11 (The three steps of invasion). *Let $(g_1, p_1), \dots, (g_d, p_d) \in \mathcal{X}$ coexist. Assume that, for any $K \geq 1$, $\text{Supp}(\nu_0^K) = \mathcal{X}_{(\mathbf{g}, \mathbf{p})} \cup \{(\tilde{g}, \tilde{p})\}$. Let $\tau_{mut.}$ denote the next mutation time (after time zero) and define*

$$\theta_{NoJump}^{K,\xi} \equiv \inf \left\{ t \geq 0 : \nu_t^K(\tilde{g}) = 0 \text{ and } \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_x(\mathbf{g}, \mathbf{p}) \delta_x \right\|_{TV} < \xi \right\} \tag{3.33}$$

$$\theta_{Jump}^{K,\xi} \equiv \inf \left\{ t \geq 0 : \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))}} \mathbf{n}_x^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) \delta_x \right\|_{TV} < \xi \right\} \tag{3.34}$$

$$\text{and } \forall \hat{x} \notin \{x \in \mathcal{X} : \mathbf{n}_x^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) > 0\} : \nu_t^K(\hat{x}) = 0 \}.$$

Assume that we have a single initial mutant, i.e. $\nu_0^K(\tilde{g}, \tilde{p}) = 1/K$. Then, there exist $\epsilon_0 > 0, C > 0$, and $M > 0$ such that for all $\epsilon \in (0, \epsilon_0)$ if $\left\| \nu_0^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_x(\mathbf{g}, \mathbf{p}) \delta_x \right\|_{TV} < \epsilon$,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\theta_{NoJump}^{K,M\epsilon} < \theta_{Jump}^{K,M\epsilon} \right] \geq q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) - C\epsilon, \tag{3.35}$$

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\theta_{Jump}^{K,M\epsilon} < \theta_{NoJump}^{K,M\epsilon} \right] \geq 1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) - C\epsilon, \tag{3.36}$$

where $1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p})$ is the invasion probability defined in (3.8) and

$$\forall \eta > 0, \quad \lim_{K \uparrow \infty} \mathbb{P} \left[\theta_{Jump}^{K, M\epsilon} \wedge \theta_{NoJump}^{K, M\epsilon} \geq \frac{\eta}{u_K K} \wedge \tau_{mut.} \right] \leq C\epsilon. \quad (3.37)$$

The structure of the proof is similar to the one of Lemma 3 in [9] (cf. also Lem. A.4. of [11]). However, we have to extend the theory to multi-type branching processes. Thus, the proof is not a simple copy the arguments in [9]. Before proving the theorem, let us collect *some properties about multi-type continuous-time branching processes*. Most of these can be found in [2] or [27]. The limit theorems we need in the sequel were first obtained by Kesten and Stigum [23, 22, 24] in the discrete-time case and by Athreya [1] in the continuous-time case.

Let $Z(t)$ be a k -dimensional continuous-time branching process. Assume that $Z(t)$ is non-singular and that the first moments exist. (Note that a process is *singular* if and only if each individual has exactly one offspring and that the existence of the first moments is sufficient for the non-extinction hypothesis.) Then, the so-called *mean matrix* $M(t)$ of $Z(t)$ is the $k \times k$ matrix with elements

$$m_{ij}(t) \equiv \mathbb{E}[Z_j(t) | Z(0) = \mathbf{e}_i], \quad 1 \leq i, j \leq k, \quad (3.38)$$

and \mathbf{e}_i is the i -th unit vector in \mathbb{R}^k . It is well known (cf. [2] p. 202) that there exists a matrix \mathbf{A} , called the infinitesimal generator of the semigroup $\{\mathbf{M}(t), t \geq 0\}$, such that

$$\mathbf{M}(t) \equiv \exp(\mathbf{A}t) = \sum_{n=0}^{\infty} \frac{t^n (\mathbf{A})^n}{n!}. \quad (3.39)$$

Furthermore, let $\mathbf{r} = (r_1, \dots, r_k)$ be the vector of the branching rates, meaning that every individual of type i has an exponentially distributed lifetime of parameter r_i and let \mathbf{M} be the mean matrix of the corresponding discrete-time process, i.e. $\mathbf{M} \equiv \{m_{ij}, i, j = 1, \dots, k\}$, where m_{ij} is the expected number of type j offspring of a single type- i -particle in one generation. Then, we can identify the infinitesimal generator \mathbf{A} as

$$\mathbf{A} = \mathbf{R}(\mathbf{M} - \mathbf{I}), \quad (3.40)$$

where $\mathbf{R} = \text{diag}(r_1, \dots, r_k)$, i.e. $r_{ij} = r_i \delta_i(j)$ and \mathbf{I} is the identity matrix of size k .

Under the basic assumption of *positive regularity*, i.e. that there exists a time t_0 such that $\mathbf{M}(t_0)$ has strictly positive entries, the Perron-Frobenius theory asserts that

- (i) the largest eigenvalue of $\mathbf{M}(t_0)$ is real-valued and strictly positive,
- (ii) the algebraic and geometric multiplicities of this eigenvalue are both one, and
- (iii) the corresponding eigenvector has strictly positive components.

By (3.39), the eigenvalues of $\mathbf{M}(t)$ are given by $\exp(\lambda_i t)$, where $\{\lambda_i; i = 1, \dots, k\}$ are the eigenvalues of \mathbf{A} , and both matrices have the same eigenvectors, which implies that the left and right eigenvectors \mathbf{u} and \mathbf{v} of $\lambda_{\max}(\mathbf{A})$ can be chosen with strictly positive components and satisfying

$$\sum_{i=1}^k v_i u_i = 1 \quad \text{and} \quad \sum_{i=1}^k u_i = 1. \quad (3.41)$$

The process Z is called supercritical, critical, or subcritical according as $\lambda_{\max}(\mathbf{A})$ is larger, equal, or smaller than zero.

Observe that the following properties are equivalent (cf. [27] p. 95-99 and [26]):

Z is irreducible $\Leftrightarrow \mathbf{M}$ is irreducible $\Leftrightarrow \mathbf{A}$ is irreducible $\Leftrightarrow \mathbf{M}(t)$ is irreducible for all $t > 0$ $\Leftrightarrow \mathbf{M}(t) > 0$ for all $t > 0$.

In particular, irreducible implies positive regular. Note that a matrix is irreducible if it is not similar via a permutation to a block upper triangular matrix and that a Markov chain is irreducible if and only if the mean matrix is irreducible.

The following lemma is an extension of Theorem 4 of [9] for multi-type branching processes.

Lemma 3.12. *Let $(Z(t))_{t \geq 0}$ be a non-singular and irreducible k -dimensional continuous-time Markov branching process and \mathbf{q} the extinction vector of Z , i.e.*

$$q_i \equiv \mathbb{P}[Z(t) = 0, \text{ for some } t \geq 0 | Z(0) = \mathbf{e}_i], \quad \text{for } 1 \leq i \leq k. \quad (3.42)$$

Furthermore, let $(t_K)_{K \geq 1}$ be a sequence of positive numbers such that $\ln(K) \ll t_K$, define $T_\rho \equiv \inf\{t \geq 0 : \sum_{i=1}^k Z_i(t) = \rho\}$ and assume that, for all $i, j \in \{1, \dots, k\}$ and $t \in [0, \infty)$,

$$\mathbb{E}[Z_j(t) \ln(Z_j(t)) | Z(0) = \mathbf{e}_i] < \infty. \quad (3.43)$$

(i) *If Z is subcritical, i.e. $\lambda_{\max}(\mathbf{A}) < 0$, then, for any $\epsilon > 0$,*

$$\lim_{K \uparrow \infty} \mathbb{P} \left[T_0 \leq t_K \wedge T_{\lceil \epsilon K \rceil} \mid Z(0) = \mathbf{e}_i \right] = 1, \quad \text{for all } i \in \{1, \dots, k\}, \quad (3.44)$$

and

$$\lim_{K \uparrow \infty} \inf_{\mathbf{x} \in \partial B_{\epsilon K}} \mathbb{P} [T_0 \leq t_K \mid Z(0) = \mathbf{x}] = 1, \quad \text{where } \partial B_{\epsilon K} \equiv \{\mathbf{x} \in \mathbb{N}_0^k : \sum_{i=1}^k x_i = \lceil \epsilon K \rceil\}. \quad (3.45)$$

Moreover, for $\bar{u} = \frac{\max_{1 \leq i \leq k} u_i}{\min_{1 \leq j \leq k} u_j}$ and, for any $\epsilon > 0$,

$$\lim_{K \uparrow \infty} \sup_{\mathbf{x} \in B_{\epsilon^2 K}} \mathbb{P} \left[T_{\lceil \epsilon K \rceil} \leq T_0 \mid Z(0) = \mathbf{x} \right] \leq \bar{u} \epsilon, \quad \text{where } B_{\epsilon^2 K} \equiv \{\mathbf{x} \in \mathbb{N}_0^k : \sum_{i=1}^k x_i \leq \lceil \epsilon^2 K \rceil\}. \quad (3.46)$$

(ii) *If Z is supercritical, i.e. $\lambda_{\max}(\mathbf{A}) > 0$, then, for any $\epsilon > 0$ (small enough),*

$$\lim_{K \uparrow \infty} \mathbb{P} \left[T_0 \leq t_K \wedge T_{\lceil \epsilon K \rceil} \mid Z(0) = \mathbf{e}_i \right] = q_i, \quad \text{for all } i \in \{1, \dots, k\} \quad (3.47)$$

and

$$\lim_{K \uparrow \infty} \mathbb{P} \left[T_{\lceil \epsilon K \rceil} \leq t_K \mid Z(0) = \mathbf{e}_i \right] = 1 - q_i, \quad \text{for all } i \in \{1, \dots, k\}. \quad (3.48)$$

Moreover, conditionally on survival, the proportions of the different types present in the population converge almost surely, as $t \uparrow \infty$, to the corresponding ratios of the components of the eigenvector: for all $i = 1, \dots, k$,

$$\lim_{t \uparrow \infty} \frac{Z_i(t)}{\sum_{j=1}^k Z_j(t)} = \frac{v_i}{\sum_{j=1}^k v_j}, \quad \text{a.s. on } \{T_0 = \infty\}. \quad (3.49)$$

Remark 3.13. In the critical case, when $\lambda_{\max}(\mathbf{A}) = 0$, (3.44) still holds, i.e. a mutant will not fixate. But Eq. (3.45) does not hold, since the time to extinction is now much larger. This leads to complications that would obfuscate the picture. Assumption 3 on the existence of a unique strictly stable equilibrium excludes the occurrence of these critical situations.

Proof. We start with the proof of (i). Since $Z(t)$ is in this case a subcritical irreducible continuous-time branching process and $\mathbb{E}[Z_j(t) \ln(Z_j(t)) | Z(0) = \mathbf{e}_i] < \infty$, we obtain by

applying Satz 6.2.7 of [27] the existence of a constant $C > 0$ such that

$$\lim_{t \uparrow \infty} \frac{1 - q_i(t)}{e^{\lambda_{\max}(\mathbf{A})t}} = Cu_i, \tag{3.50}$$

where $q_i(t) \equiv \mathbb{P}[Z(t) = 0 \mid Z(0) = \mathbf{e}_i]$. Moreover, we have a non-explosion condition. Thus, for all $\epsilon > 0$, either $T_{\lceil \epsilon K \rceil}$ equals infinity or it converges to infinity as $K \uparrow \infty$. Putting both together, there exists a sequence s_K with $\lim_{K \uparrow \infty} s_K = +\infty$ such that

$$\lim_{K \uparrow \infty} \mathbb{P} \left[T_0 \leq t_K \wedge T_{\lceil \epsilon K \rceil} \mid Z(0) = \mathbf{e}_i \right] \geq \lim_{K \uparrow \infty} \mathbb{P} [T_0 \leq s_K \mid Z(0) = \mathbf{e}_i] = \lim_{K \uparrow \infty} q_i(s_K) = 1. \tag{3.51}$$

The branching property implies that, for all $\mathbf{x} \in \mathbb{N}^k$, $\mathbb{P}[Z(t) = 0 \mid Z(0) = \mathbf{x}] = \prod_{i=1}^k (q_i(t))^{x_i}$ (cf. [26] p. 25). So, we get

$$\inf_{\mathbf{x} \in \partial B_{\epsilon K}} \mathbb{P} [T_0 \leq t_K \mid Z(0) = \mathbf{x}] = \inf_{\mathbf{x} \in \partial B_{\epsilon K}} \mathbb{P} [Z(t_K) = 0 \mid Z(0) = \mathbf{x}] = \inf_{\mathbf{x} \in \partial B_{\epsilon K}} \prod_{i=1}^k (q_i(t_K))^{x_i}. \tag{3.52}$$

For all $i \in \{1, \dots, k\}$, $1 \geq (q_i(t_K))^{x_i} \geq (q_i(t_K))^{\lceil \epsilon K \rceil}$ and by (3.50) we have $1 - q_i(t_K) = O(e^{\lambda_{\max}(\mathbf{A})t_K})$. Moreover, for any sequence $(w_K)_{K \geq 1}$ such that $\lim_{K \uparrow \infty} w_K = 0$,

$$\lim_{K \uparrow \infty} \left(1 + \frac{w_K}{K} \right)^K = 1. \tag{3.53}$$

This implies that, for all t_K with $t_K \gg \ln(K)$ and $C > 0$, since $\lim_{K \uparrow \infty} C e^{\lambda_{\max}(\mathbf{A})t_K} \lceil \epsilon K \rceil = 0$,

$$\lim_{K \uparrow \infty} (1 - C e^{\lambda_{\max}(\mathbf{A})t_K})^{\lceil \epsilon K \rceil} = 1. \tag{3.54}$$

Thus, taking the limit $K \uparrow \infty$ in (3.52), we obtain the desired equation (3.45). To prove the inequality (3.46) we use the fact that $(\sum_{i=1}^k u_i Z_i(t)) e^{-\lambda_{\max} t}$ is a martingale (cf. [1], Prop. 2). By applying Doob's stopping theorem to the stopping time $T_{\lceil \epsilon K \rceil} \wedge T_0$ we obtain, for all $\mathbf{x} \in B_{\epsilon^2 K}$, that

$$\mathbb{E} \left[\left(\sum_{i=1}^k u_i Z_i(T_{\lceil \epsilon K \rceil}) \right) e^{-\lambda_{\max}(\mathbf{A})T_{\lceil \epsilon K \rceil}} \mathbf{1}_{\{T_{\lceil \epsilon K \rceil} < T_0\}} \mid Z(0) = \mathbf{x} \right] = \sum_{i=1}^k u_i x_i. \tag{3.55}$$

Therefore, since $\lambda_{\max}(\mathbf{A}) < 0$ in the subcritical case,

$$\mathbb{E} \left[\min_{1 \leq i \leq k} u_i \lceil \epsilon K \rceil \mathbf{1}_{\{T_{\lceil \epsilon K \rceil} < T_0\}} \mid Z(0) = \mathbf{x} \right] \leq \max_{1 \leq i \leq k} u_i \lceil \epsilon^2 K \rceil, \quad \text{for all } \mathbf{x} \in B_{\epsilon^2 K}, \tag{3.56}$$

which implies (3.46).

Let us continue by proving (ii). Since $Z(t)$ is supercritical in this case, applying Theorem 5.7.2 of [2] yields that

$$\lim_{t \uparrow \infty} Z(t)(\omega) e^{-\lambda_{\max}(\mathbf{A})t} = W(\omega) \mathbf{v}, \text{ a.s.}, \tag{3.57}$$

where W is a nonnegative random variable. Since we assume that, for all $i \in \{1, \dots, k\}$, $\mathbb{E}[Z_j(t) \ln(Z_j(t)) \mid Z(0) = \mathbf{e}_i] < \infty$, we get that

$$\mathbb{P}[W = 0 \mid Z(0) = \mathbf{e}_i] = q_i, \quad \mathbb{E}[W \mid Z(0) = \mathbf{e}_i] = u_i, \tag{3.58}$$

and W has an absolutely continuous distribution on $(0, \infty)$. All components of \mathbf{v} are strictly positive and $W > 0$, a.s., on the event $\{\omega : T_0(\omega) = \infty\}$. Hence, we have

$$Z(t) = O \left(e^{\lambda_{\max}(\mathbf{A})t} \right) \text{ a.s. on } \{T_0 = \infty\}. \tag{3.59}$$

This implies that $\lim_{K \uparrow \infty} \mathbb{P}[Z(t_K) < \lceil \epsilon K \rceil, T_0 = \infty] = 0$ and thus

$$\lim_{K \uparrow \infty} \mathbb{P}[T_0 = \infty, T_{\lceil \epsilon K \rceil} \geq t_K] = 0. \tag{3.60}$$

Note that we used that $t_K \gg \ln(K)$. Since $\mathbb{P}[T_0 = \infty | Z(0) = \mathbf{e}_i] = 1 - q_i$, we deduce (3.48). On the other hand, there exist two sequences s_K^1 and s_K^2 , which converge to infinity as $K \uparrow \infty$, such that, $\lim_{K \uparrow \infty} \mathbb{P}[s_K^1 \leq t_K \wedge T_{\lceil \epsilon K \rceil} \leq s_K^2] = 1$. This yields (3.47), because, for all $i \in \{1, \dots, k\}$ and $l = 1, 2$, we have that $\lim_{K \uparrow \infty} \mathbb{P}[T_0 < s_K^l | Z_0 = \mathbf{e}_i] = q_i$. Note that equation (3.49) is a simple consequence of (3.57). \square

Using these properties about multi-type branching processes we can now prove the theorem about the three steps of invasion.

Proof of Theorem 3.11. The first invasion step. Let us introduce the following stopping times

$$\theta_{\text{exit}}^{K, M\epsilon} = \inf \left\{ t \geq 0 : \|\nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_x(\mathbf{g}, \mathbf{p}) \delta_x\|_{TV} > M\epsilon \right\} \tag{3.61}$$

$$\tilde{\theta}_\epsilon^K = \inf \left\{ t \geq 0 : \nu_t^K(\tilde{g}) \geq \epsilon \right\} \tag{3.62}$$

$$\tilde{\theta}_0^K = \inf \left\{ t \geq 0 : \nu_t^K(\tilde{g}) = 0 \right\} \tag{3.63}$$

Until $\tilde{\theta}_\epsilon^K$ the mutant population $\nu_t^K(\tilde{g})$ influences only the death and switching rates of the resident population and this perturbation is uniformly bounded by $(\bar{c} + \bar{s}_{\text{ind}})\epsilon$. Thus, by applying Theorem 3.8 (ii), we obtain

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\theta_{\text{exit}}^{K, M\epsilon} < e^{KV} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_\epsilon^K \right] = 0. \tag{3.64}$$

On the time interval $[0, \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_\epsilon^K]$, the resident population can be approximated by $\sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{n}_x(\mathbf{g}, \mathbf{p}) \delta_x$ and no further mutant appears. This allows us to approximate $\nu_t^K(\tilde{g})$ by multi-type branching processes.

Let $k \equiv \lceil [\tilde{p}]_{\tilde{g}} \rceil$. We construct two $(\mathbb{N}_0)^k$ -valued processes $X^{1, \epsilon}(t)$ and $X^{2, \epsilon}(t)$, using the pathwise definition in terms of Poisson point measures of ν_t^K , which control the mutant population $\nu_t^K(\tilde{g})$. To this aim let us denote the elements of $[\tilde{p}]_{\tilde{g}}$ by $\tilde{p}_1, \dots, \tilde{p}_k$ (w.l.o.g. $\tilde{p} \equiv \tilde{p}_1$). Then, we define $X^{1, \epsilon}$ by

$$X^{1, \epsilon}(t) \equiv X^{1, \epsilon}(0) \tag{3.65}$$

$$\begin{aligned} &+ \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{ i \leq X_j^{1, \epsilon}(s-), \theta \leq b(\tilde{p}_i) - \epsilon \right\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{birth}}(ds, di, d\theta) \\ &- \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{ i \leq X_j^{1, \epsilon}(s-), \theta \leq d(\tilde{p}_j) + \sum_{(g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_j, p) \bar{n}_{(g, p)}(\mathbf{g}, \mathbf{p}) + \bar{c}M\epsilon \right\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{death}}(ds, di, d\theta) \\ &+ \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{[\tilde{p}]_{\tilde{g}}} \mathbb{1}_{\left\{ i \leq X_j^{1, \epsilon}(s-), i \neq j \right\}} \\ &\quad \times \left(\mathbb{1}_{\left\{ \theta \leq s_{\text{nat}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_i) + \sum_{(g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_i)(p) \bar{n}_{(g, p)}(\mathbf{g}, \mathbf{p}) - \bar{s}_{\text{ind}}M\epsilon \right\}} \mathbf{e}_l \right. \\ &\quad \left. - \mathbb{1}_{\left\{ \theta \leq s_{\text{nat}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l) + \sum_{(g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l)(p) \bar{n}_{(g, p)}(\mathbf{g}, \mathbf{p}) + \bar{s}_{\text{ind}}M\epsilon \right\}} \mathbf{e}_j \right) N_{(\tilde{g}, \tilde{p}_j)}^{\text{switch}}(ds, di, d\theta, d\tilde{p}_i), \end{aligned}$$

and similar $X^{2, \epsilon}$ by

$$\begin{aligned}
 X^{2,\epsilon}(t) &\equiv X^{2,\epsilon}(0) & (3.66) \\
 &+ \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\{i \leq X_j^{2,\epsilon}(s-), \theta \leq b(\tilde{p}_i) + \epsilon\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{birth}}(ds, di, d\theta) \\
 &- \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\{i \leq X_j^{2,\epsilon}(s-), \theta \leq d(\tilde{p}_j) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_j, p) \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) - \bar{c}M\epsilon\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{death}}(ds, di, d\theta) \\
 &+ \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{[\tilde{p}]_{\tilde{g}}} \mathbb{1}_{\{i \leq X_j^{1,\epsilon}(s-), i \neq j\}} \\
 &\quad \times \left(\mathbb{1}_{\{\theta \leq s_{\text{nat}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l)(p) \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) + \bar{s}_{\text{ind}}M\epsilon\}} \mathbf{e}_l \right. \\
 &\quad \left. - \mathbb{1}_{\{\theta \leq s_{\text{nat}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind}}^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l)(p) \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p}) - \bar{s}_{\text{ind}}M\epsilon\}} \mathbf{e}_j \right) N_{(\tilde{g}, \tilde{p}_j)}^{\text{switch}}(ds, di, d\theta, d\tilde{p}_l),
 \end{aligned}$$

where \mathbf{e}_j is the j -th unit vector in \mathbb{R}^k and N^{birth} , N^{death} , and N^{switch} are the collections of Poisson point measures defined in Subsection 2.1. Note that $X^{1,\epsilon}(t)$ and $X^{2,\epsilon}(t)$ are k -type branching processes with the following dynamics: For each $1 \leq i \leq k$, each individual in $X^{1,\epsilon}(t)$, respectively $X^{2,\epsilon}(t)$, with trait (\tilde{g}, \tilde{p}_i) undergoes

- (i) birth (without mutation) with rate $b(\tilde{p}_i) - \epsilon$, respectively $b(\tilde{p}_i) + \epsilon + 2(k-1)\bar{s}_{\text{ind}}M\epsilon$,
 - death with rate $D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) + \bar{c}M\epsilon + 2(k-1)\bar{s}_{\text{ind}}M\epsilon$, respectively $D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) - \bar{c}M\epsilon$, where $D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) \equiv d(\tilde{p}_i) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_i, p) \bar{n}_{(g,p)}(\mathbf{g}, \mathbf{p})$,
- (ii) switch to \tilde{p}_j with rate $S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i, \tilde{p}_j) - \bar{s}_{\text{ind}}M\epsilon$, for all $j \neq i$ (for both processes), where $S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i, \tilde{p}_j) \equiv s_{\text{nat}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} s_{\text{ind}}^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j)(p) \bar{n}_{(g,p)}$.

Moreover, the processes $X^{1,\epsilon}(t)$ and $X^{2,\epsilon}(t)$ have the following property: There exists a $K_0 > 1$ such that, for all $\tilde{p}_i \in [\tilde{p}]_{\tilde{g}}$ and for all $K \geq K_0$,

$$\forall 0 \leq t \leq \theta_{\text{exit}}^{K,\epsilon} \wedge \tau_{\text{mut}} \wedge \tilde{\theta}_\epsilon^K : \quad X_i^{1,\epsilon}(t) \leq \nu_t^K(\tilde{g}, \tilde{p}_i)K \leq X_i^{2,\epsilon}(t). \quad (3.67)$$

Hence, if $\tilde{\theta}_\epsilon^K \leq \theta_{\text{exit}}^{K,\epsilon} \wedge \tau_{\text{mut}}$, then

$$\inf \{t \geq 0 : X^{2,\epsilon}(t) = \lceil \epsilon K \rceil\} \leq \tilde{\theta}_\epsilon^K \leq \inf \{t \geq 0 : X^{1,\epsilon}(t) = \lceil \epsilon K \rceil\}. \quad (3.68)$$

On the other hand, if $\inf \{t \geq 0 : X^{2,\epsilon}(t) = 0\} \leq \inf \{t \geq 0 : X^{2,\epsilon}(t) = \lceil \epsilon K \rceil\} \wedge \theta_{\text{exit}}^{K,\epsilon} \wedge \tau_{\text{mut}}$, then

$$\tilde{\theta}_0^K \leq \inf \{t \geq 0 : X^{2,\epsilon}(t) = 0\}. \quad (3.69)$$

Next, let us identify the infinitesimal generator of the control processes $X^{1,\epsilon}$ and $X^{2,\epsilon}$. Therefore, define, for $i = 1, \dots, k$,

$$f_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}_i) \equiv b(\tilde{p}_i) - D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) - \sum_{j \neq i} S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i, \tilde{p}_j). \quad (3.70)$$

$(f_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}_i))$ would be the invasion fitness of phenotype \tilde{p}_i if there was no switch back from the other phenotypes to \tilde{p}_i . Then, by Equation (3.40), the infinitesimal generators are given by the following matrixes

$$\mathbf{A}(X^{l,\epsilon}) = \begin{pmatrix} f_{(\mathbf{g}, \mathbf{p})}^{l,\epsilon}(\tilde{g}, \tilde{p}_1) & S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_1, \tilde{p}_2) - \bar{s}_{\text{ind}}M\epsilon & \dots & S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_1, \tilde{p}_k) - \bar{s}_{\text{ind}}M\epsilon \\ S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_2, \tilde{p}_1) - \bar{s}_{\text{ind}}M\epsilon & f_{(\mathbf{g}, \mathbf{p})}^{l,\epsilon}(\tilde{g}, \tilde{p}_2) & & \\ \vdots & & \ddots & \vdots \\ S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_k, \tilde{p}_1) - \bar{s}_{\text{ind}}M\epsilon & & \dots & f_{(\mathbf{g}, \mathbf{p})}^{l,\epsilon}(\tilde{g}, \tilde{p}_k) \end{pmatrix} \quad (3.71)$$

for $l \in \{1, 2\}$, where $f_{(\mathbf{g}, \mathbf{p})}^{1, \epsilon}(\tilde{g}, \tilde{p}_i) \equiv f_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}_i) - \epsilon(1 + \bar{c}M + (k-1)\bar{s}_{\text{ind}}M)$ and $f_{(\mathbf{g}, \mathbf{p})}^{2, \epsilon}(\tilde{g}, \tilde{p}_i) \equiv f_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}_i) + \epsilon(1 + \bar{c}M + 3(k-1)\bar{s}_{\text{ind}}M)$.

We prove in the following that the number of mutant individuals grow with positive probability to ϵK before dying out if and only if λ_{\max} of $\mathbf{A}_{(\tilde{g}, \tilde{p})} \equiv \lim_{\epsilon \rightarrow 0} \mathbf{A}(X^{1, \epsilon})$ is strictly positive. Thus, $\lambda_{\max}(\mathbf{A}_{(\tilde{g}, \tilde{p})})$ is an appropriate generalisation of the invasion fitness of the class $[\tilde{p}]_{\tilde{g}}$:

$$F_{[\tilde{p}]_{\tilde{g}}}(\mathbf{g}, \mathbf{p}) \equiv \lambda_{\max}(\mathbf{A}_{(\tilde{g}, \tilde{p})}). \tag{3.72}$$

Since the birth and death rates of $X^{1, \epsilon}$ and $X^{2, \epsilon}$ are positive and since Assumption 3.1 implies that $\mathbf{M}(X^{1, \epsilon})$ and $\mathbf{M}(X^{2, \epsilon})$ are irreducible, we obtain that the processes $X^{1, \epsilon}$ and $X^{2, \epsilon}$ are non-singular and irreducible. Thus, $X^{1, \epsilon}$ and $X^{2, \epsilon}$ satisfy the conditions of Lemma 3.12. For $l \in \{1, 2\}$, let $\mathbf{q}(X^{l, \epsilon})$ denote the extinction probability vector of $X^{l, \epsilon}$, i.e.

$$\mathbf{q}(X^{l, \epsilon}) \equiv (q_1(X^{l, \epsilon}), \dots, q_k(X^{l, \epsilon})),$$

where $q_i(X^{l, \epsilon}) \equiv \mathbb{P}[X^{l, \epsilon}(t) = 0, \text{ for some } t \mid X^{l, \epsilon}(0) = \mathbf{e}_i]$.

Observe that $\mathbf{q}(X^{l, \epsilon}) = (1, \dots, 1)$ if $X^{l, \epsilon}$ is not supercritical. To characterise $\mathbf{q}(X^{l, \epsilon})$ in the supercritical case, let us introduce the following functions

$$\mathbf{u}^l : [0, 1]^k \times (-\eta, \eta) \rightarrow \mathbb{R}^k, \quad \text{where } \eta \text{ is some small enough constant and } l \in \{1, 2\}, \tag{3.73}$$

defined, for all $1 \leq i \leq k$, by

$$\begin{aligned} \mathbf{u}_i^1(\mathbf{y}, \epsilon) \equiv & \left(b(\tilde{p}_i) - \epsilon \right) y_i^2 + \sum_{j \neq i} (S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i, \tilde{p}_j) - \bar{s}_{\text{ind}}M\epsilon) y_j \\ & + D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) + \bar{c}M\epsilon + 2(k-1)\bar{s}_{\text{ind}}M\epsilon \\ & - \left(b(\tilde{p}_i) + \sum_{j \neq i} S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i, \tilde{p}_j) + D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) + (1 - \bar{c}M + (k-1)\bar{s}_{\text{ind}}M)\epsilon \right) y_i. \end{aligned} \tag{3.74}$$

and

$$\begin{aligned} \mathbf{u}_i^2(\mathbf{y}, \epsilon) \equiv & \left(b(\tilde{p}_i) + \epsilon + 2(k-1)\bar{s}_{\text{ind}}M\epsilon \right) y_i^2 + \sum_{j \neq i} (S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i, \tilde{p}_j) - \bar{s}_{\text{ind}}M\epsilon) y_j \\ & + D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) - \bar{c}M\epsilon \\ & - \left(b(\tilde{p}_i) + \sum_{j \neq i} S_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i, \tilde{p}_j) + D_{(\mathbf{g}, \mathbf{p})}(\tilde{p}_i) + (1 - \bar{c}M + (k-1)\bar{s}_{\text{ind}}M)\epsilon \right) y_i. \end{aligned} \tag{3.75}$$

Observe that $\mathbf{u}^1(\mathbf{y}, \epsilon)$ and $\mathbf{u}^2(\mathbf{y}, \epsilon)$ are the infinitesimal generating functions of $X^{1, \epsilon}$ and $X^{2, \epsilon}$ and that $\mathbf{u}^1(\mathbf{y}, 0) = \mathbf{u}^2(\mathbf{y}, 0)$. Moreover, the extinction vector of a multi-type branching process is given as the unique root of the generating function in the unit cube (cf. [2] p. 205 or [27] Chap. 5). Thus, in the supercritical case $\mathbf{q}(X^{1, \epsilon})$ is the unique solution of

$$\mathbf{u}^1(\mathbf{y}, \epsilon) = 0 \quad \text{for } \mathbf{y} \in [0, 1]^k \tag{3.76}$$

and $\mathbf{q}(X^{2, \epsilon})$ is the unique solution of

$$\mathbf{u}^2(\mathbf{y}, \epsilon) = 0 \quad \text{for } \mathbf{y} \in [0, 1]^k. \tag{3.77}$$

These solutions are in general not analytic. Applying Lemma 3.12 to $X^{1, \epsilon}$ and $X^{2, \epsilon}$ we obtain that there exists $C_1 > 0$ such that, for all $\eta > 0$, $\epsilon > 0$ sufficiently small and K

large enough,

$$\begin{aligned} \mathbb{P} \left[\theta_{\text{No Jump}}^{K, M\epsilon} < \frac{\eta}{K u_K} \wedge \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_\epsilon^K \right] &\geq \mathbb{P} \left[\inf\{t \geq 0 : X^{2,\epsilon}(t) = 0\} < \frac{\eta}{K u_K} \right] \\ &\geq q_1(X^{2,\epsilon}) - C_1\epsilon \end{aligned} \tag{3.78}$$

and

$$\begin{aligned} \mathbb{P} \left[\tilde{\theta}_\epsilon^K < \frac{\eta}{K u_K} \wedge \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_0^K \right] &\geq \mathbb{P} \left[\inf\{t \geq 0 : X^{1,\epsilon}(t) = 0\} < \frac{\eta}{K u_K} \right] \\ &\geq 1 - q_1(X^{1,\epsilon}) - C_1\epsilon. \end{aligned} \tag{3.79}$$

If $X^{2,\epsilon}$ is sub- or critical for ϵ small enough, then $\lim_{\epsilon \downarrow 0} q_1(X^{2,\epsilon}) = \lim_{\epsilon \downarrow 0} q_1(X^{1,\epsilon}) = 1$. In the supercritical case, let $\mathbf{q} \in [0, 1]^k$ be the solution of $\mathbf{u}^1(\mathbf{y}, 0) = \mathbf{u}^2(\mathbf{y}, 0) = 0$. Then, by applying the implicit function theorem, there exist open sets $U^1 \subset \mathbb{R}$ and $U^2 \subset \mathbb{R}$ containing 0, open sets $V^1 \subset \mathbb{R}^k$ and $V^2 \subset \mathbb{R}^k$ containing \mathbf{q} , and two unique continuously differentiable functions $g^1 : U^1 \rightarrow V^1$ and $g^2 : U^2 \rightarrow V^2$ such that

$$\{(\epsilon, g^1(\epsilon)) | \epsilon \in U^1\} = \{(\epsilon, \mathbf{y}) \in U^1 \times V^1 | \mathbf{u}^1(\mathbf{y}, \epsilon) = 0\}. \tag{3.80}$$

and

$$\{(\epsilon, g^2(\epsilon)) | \epsilon \in U^2\} = \{(\epsilon, \mathbf{y}) \in U^2 \times V^2 | \mathbf{u}^2(\mathbf{y}, \epsilon) = 0\}. \tag{3.81}$$

By definition, $g^1(0) = g^2(0) = \mathbf{q}$ and $q_1 = q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p})$. We can linearise and obtain that there exists a constant $C_2 > 0$ such that

$$q_1(X^{1,\epsilon}) \leq q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) + C_2\epsilon \quad \text{and} \quad q_1(X^{2,\epsilon}) \geq q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) - C_2\epsilon \tag{3.82}$$

Therefore,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\theta_{\text{No Jump}}^{K, M\epsilon} \wedge \tilde{\theta}_\epsilon^K < \frac{\eta}{K u_K} \wedge \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}} \right] \geq 1 - 2(C_1 + C_2)\epsilon. \tag{3.83}$$

Conditionally on survival, the proportions of the different phenotypes in $X^{1,\epsilon}$ converge almost surely, as $t \uparrow \infty$, to the corresponding ratios of the components of the eigenvector, which are all strictly positive (cf. Lem. 3.12, Eq. (3.49)). Moreover, there exists a constant $C_3 > 0$ such that, for all ϵ small enough,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\left\{ \tilde{\theta}_\epsilon^K < \frac{\eta}{K u_K} \wedge \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}} \right\} \cap \left\{ \inf\{t \geq 0 : X^{1,\epsilon}(t) = 0\} < \infty \right\} \right] < C_3\epsilon \tag{3.84}$$

and $\tilde{\theta}_\epsilon^K$ converges to infinity as $K \uparrow \infty$. Thus, conditionally on $\{\tilde{\theta}_\epsilon^K < \frac{\eta}{K u_K} \wedge \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}}\}$, there exists a (small) constant $C_4 > 0$ such that the probability that the densities of the phenotypes $\{\tilde{p}_1, \dots, \tilde{p}_k\}$, are all larger than $C_4\epsilon$ at time $\tilde{\theta}_\epsilon^K$ converges to one as first $K \uparrow \infty$ and then $\epsilon \rightarrow 0$. More precisely, there exists constants $C_4 > 0$ and $C_5 > 0$ such that, for all ϵ small enough,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\tilde{\theta}_\epsilon^K < \frac{\eta}{K u_K} \wedge \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}}, \exists i \in \{1, \dots, k\} : \nu_{\tilde{\theta}_\epsilon^K}^K(\tilde{p}_i) \leq C_4\epsilon \right] \leq C_5\epsilon. \tag{3.85}$$

The second invasion step. By Assumption 3.4, any solution of $LVS(d+1, ((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})))$ with initial state in the compact set

$$A \equiv \left\{ x \in \mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|} : |x - \bar{\mathbf{n}}(\mathbf{g}, \mathbf{p})| \leq M\epsilon \right\} \times [C_4\epsilon, \epsilon]^k \tag{3.86}$$

converge, as $t \uparrow \infty$, to the unique locally strictly stable equilibrium $\mathbf{n}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$. Therefore, for all $\epsilon > 0$ there exists $T(\epsilon) \in \mathbb{R}$ such that any of these trajectories do not leave the set

$$\left\{ x \in \mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|+k} : |x - \mathbf{n}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))| \leq \epsilon^2/2 \right\} \tag{3.87}$$

after time $T(\epsilon)$. Back to the stochastic system, let us introduce on the event $\{\tilde{\theta}_\epsilon^K < \frac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}}\}$ the following stopping time

$$\theta_{\text{near } n^*}^{K,\epsilon} = \inf \left\{ t \geq \tilde{\theta}_\epsilon^K : \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))}} \mathbf{n}_x^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) \delta_x \right\|_{TV} < \epsilon^2 \right\}. \quad (3.88)$$

Then, we conclude by using the strong Markov property at $\tilde{\theta}_\epsilon^K$ and Theorem 3.8 (i) on $[0, T(\epsilon)]$ that there exists a constant $C_6 > 0$ such that, for all ϵ small enough,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\tilde{\theta}_\epsilon^K < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K} \text{ and } \sup_{s \in [\tilde{\theta}_\epsilon^K, \tilde{\theta}_\epsilon^K + T(\epsilon)]} \left\| \nu_s^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathbf{n}_x(s, \nu_0^K) \delta_x \right\|_{TV} \leq \epsilon^2 \right] \geq 1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) - C_6\epsilon, \quad (3.89)$$

which implies

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\tilde{\theta}_\epsilon^K < \theta_{\text{near } n^*}^{K,\epsilon} < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K} \right] \geq 1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) - C_6\epsilon. \quad (3.90)$$

We used that, at time $\tilde{\theta}_\epsilon^K$, the stochastic process ν^K (considered as element of $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})|+k}}$) lies in the compact set A , where A is defined in (3.86).

The third invasion step. After time $\theta_{\text{near } n^*}^{K,\epsilon}$ we use again comparisons with multi-type branching processes to show that all individuals carrying a trait which is not present in the new equilibrium n^* die out. To this aim let us define

$$\mathcal{X}_{\text{extinct}}^{n^*} = \{(g, p) \in \mathcal{X}_{((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))} : \mathbf{n}_{(g,p)}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) = 0\} \quad (3.91)$$

For proving that the populations with traits in $\mathcal{X}_{\text{extinct}}^{n^*}$ stay small after $\theta_{\text{near } n^*}^{K,\epsilon}$ and that the populations with traits not in $\mathcal{X}_{\text{extinct}}^{n^*}$ stay close to its equilibrium value after $\theta_{\text{near } n^*}^{K,\epsilon}$, let us define

$$\theta_{\text{not small}}^{K,\epsilon} = \inf \left\{ t \geq \theta_{\text{near } n^*}^{K,\epsilon} : \exists (g, p) \in \mathcal{X}_{\text{extinct}}^{n^*} \text{ such that } \nu_t^K(g, p) > \epsilon \right\} \quad (3.92)$$

and

$$\theta_{\text{exit } n^*}^{K,M\epsilon} \equiv \inf \left\{ t \geq \theta_{\text{near } n^*}^{K,\epsilon} : \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))}} \mathbf{n}_x^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) \delta_x \right\|_{TV} > M\epsilon \right\}. \quad (3.93)$$

By using first the strong Markov property at $\theta_{\text{near } n^*}^{K,\epsilon}$, we can apply Theorem 3.8 (ii) and obtain that there exist constants $M > 0$ and $C_7 > 0$ such that, for all ϵ small enough,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\tilde{\theta}_\epsilon^K < \theta_{\text{near } n^*}^{K,\epsilon} < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K} \text{ and } \theta_{\text{exit } n^*}^{K,M\epsilon} < e^{KV} \wedge \tau_{\text{mut.}} \wedge \theta_{\text{not small}}^{K,\epsilon} \right] < C_7\epsilon \quad (3.94)$$

This is obtained in a similar way as Equation (3.64) in the first step. Note that $(g, p) \in \mathcal{X}_{\text{extinct}}^{n^*}$ implies that $(g, p_i) \in \mathcal{X}_{\text{extinct}}^{n^*}$, for all $p_i \in [p]_g$, which is a consequence of Assumption 3.1.

Using the same arguments as in the first step, we can construct, for all $(g, p) \in \mathcal{X}_{\text{extinct}}^{n^*}$, a $[p]_g$ -type continuous-time branching process $Y^{\epsilon, (g,p)}(s)$ with initial condition

$$Y_i^{\epsilon, (g,p)}(0) = \nu_{\theta_{\text{near } n^*}^{K,\epsilon}}^{K, \epsilon}(g, p_i)K, \quad \text{for all } p_i \in [p]_g \quad (3.95)$$

such that, for all K large enough and, for all $t \in [\theta_{\text{near } n^*}^{K,\epsilon}, \theta_{\text{exit } n^*}^{K,M\epsilon} \wedge \theta_{\text{not small}}^{K,\epsilon} \wedge \tau_{\text{mut.}}]$,

$$\nu_t^K(g, p_i)K \leq Y_i^{\epsilon, (g,p)}(t - \theta_{\text{near } n^*}^{K,\epsilon}), \quad \text{for all } p_i \in [p]_g. \quad (3.96)$$

Moreover, $Y^{\epsilon, (g,p)}(t)$ is characterised as follows: For each $p_i \in [p]_g$, each individual in $Y^{\epsilon, (g,p)}(t)$ with trait (g, p_i) undergoes

- (i) birth (without mutation) with rate $b(p_i) + 2(|[p]_g| - 1)s_{\text{ind.}}(M + |\mathcal{X}_{\text{extinct}}^{\text{n}^*}|)\epsilon$
 - death with rate $d(p_i) + \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})}} c(p_i, \tilde{p}) \mathbf{n}_{(\tilde{g}, \tilde{p})}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) - \bar{c}(M + |\mathcal{X}_{\text{extinct}}^{\text{n}^*}|)\epsilon$
- (ii) for all $j \neq i$, switch to p_j with rate $s_{\text{nat.}}^g(p_i, p_j) + \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})}} s_{\text{ind.}}^g(p_i, p_j)(\tilde{p}) \mathbf{n}_{(\tilde{g}, \tilde{p})}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) - \bar{s}_{\text{ind.}}(M + |\mathcal{X}_{\text{extinct}}^{\text{n}^*}|)\epsilon$.

Let $\mathbf{A}(Y^{\epsilon, (g,p)})$ denote the infinitesimal generator of the process $Y^{\epsilon, (g,p)}$. Since the equilibrium $\mathbf{n}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ is locally strictly stable (cf. Ass. 3.4), the eigenvalues of the Jacobian matrix of the dynamical system at $\mathbf{n}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ are all strictly negative. If ϵ is small enough, this implies that all eigenvalues of $\{\mathbf{A}(Y^{\epsilon, (g,p)}), (g, p) \in \mathcal{X}_{\text{extinct}}^{\text{n}^*}\}$ are strictly negative. (There exists an order of the elements of $\mathcal{X}_{(\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})}$ such that the Jacobian matrix is an upper-block-triangular matrix and $\{\mathbf{A}(Y^{0, (g,p)}), (g, p) \in \mathcal{X}_{\text{extinct}}^{\text{n}^*}\}$ are on the diagonal.) Thus, for all ϵ small enough, the branching processes $\{Y^{\epsilon, (g,p)}, (g, p) \in \mathcal{X}_{\text{extinct}}^{\text{n}^*}\}$ are all subcritical. Moreover, we can apply Lemma 3.12 and get, for all ϵ small enough and $(g, p) \in \mathcal{X}_{\text{extinct}}^{\text{n}^*}$

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\inf\{t \geq 0 : Y^{\epsilon, (g,p)}(t) = 0\} \leq \frac{\eta}{Ku_K} \right] = 1, \tag{3.97}$$

and there exists a constant C_8 such that, for all ϵ small enough and $(g, p) \in \mathcal{X}_{\text{extinct}}^{\text{n}^*}$,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\inf\{t \geq 0 : Y^{\epsilon, (g,p)}(t) = \lceil \epsilon K \rceil\} \leq \inf\{t \geq 0 : Y^{\epsilon, (g,p)}(t) = 0\} \right] \leq C_8 \epsilon. \tag{3.98}$$

Hence, there exists a constant $M > 0$ and $C_9 > 0$ such that, for all $\eta > 0$ and ϵ small enough,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\tilde{\theta}_\epsilon^K < \theta_{\text{Jump}}^{K, M\epsilon} < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K} \wedge \theta_{\text{not small}}^{K, \epsilon} \right] \geq 1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) - C_9 \epsilon, \tag{3.99}$$

which finishes the proof of the theorem. □

Combining all the previous results, we can prove similar as in [9] that for, all $\epsilon > 0, t > 0$ and $\Gamma \subset \mathcal{X}$,

$$\lim_{K \uparrow \infty} \mathbb{P} \left[\text{Supp}(\nu_{t/Ku_K}^K) = \Gamma, \text{ all traits of } \Gamma \text{ coexist in } LVS(|\Gamma|, \Gamma), \right] \tag{3.100}$$

$$\text{and } \|\nu_{t/Ku_K}^K - \sum_{x \in \Gamma} \bar{n}_x(\Gamma) \delta_x\|_{TV} < \epsilon \Big] = \mathbb{P}[\text{Supp}(\Lambda_t) = \Gamma]$$

where Λ is the PES with phenotypic plasticity defined in Theorem 3.6. Finally, generalising this to any sequence of times $0 < t_1 < \dots < t_n$, implies that $(\nu_{t/Ku_K}^K)_{t \geq 0}$ converges in the sense of finite dimensional distributions to $(\Lambda_t)_{t \geq 0}$ (cf. [9], Cor. 1 and Lem. 1), which ends the proof of Theorem 3.6.

3.4 Examples

Figure 3 shows two examples where in a population consisting only of type (g, p) and being close to $\mathbf{n}(g, p)$ a mutation to genotype \tilde{g} occurs. In these example, \tilde{g} is associated with two possible phenotypes \tilde{p}_1 and \tilde{p}_2 .

In example (A), we start with a single mutant carrying trait (\tilde{g}, \tilde{p}_1) and which can switch to \tilde{p}_2 but the back-switch is relative weak (cf. Tab. 2). According to definition (3.70) we have $f_{(g,p)}(\tilde{g}, \tilde{p}_1) < 0$ and $f_{(g,p)}(\tilde{g}, \tilde{p}_2) > 0$. However, the global fitness of the genotype \tilde{g} is positive. More precisely, it is given by the largest eigenvalue of $\begin{pmatrix} -3 & 2 \\ 0.6 & 1 \end{pmatrix}$, which equals approximatively 1.280. Therefore, the multi-type branching process approximating the mutant population in the first step is supercritical. This does

PES with phenotypic plasticity

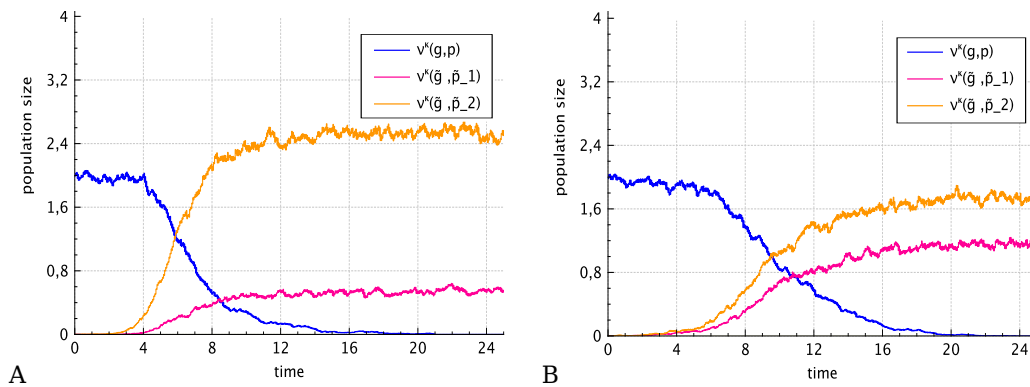


Figure 3: Simulations of the invasion phase with $K = 1000$. (A) The mutant phenotype \tilde{p}_1 has a negative initial growth rate but can switch to \tilde{p}_2 which has a positive one. The fitness of the genotype \tilde{g} is positive. (B) The fitness of the mutant genotype \tilde{g} is positive, although each phenotype has a negative initial growth rate. This is possible because an outgoing switch is a loss of a cell for a phenotype, but not for the whole genotype.

not depend on the phenotype of the first mutant, i.e. we would have the same if we had started with a single mutant carrying trait (\tilde{g}, \tilde{p}_2) . However, the probability of invasion depends on this. In this example, the invasion probability is given by the solution of

$$2y_1^2 + 2y_2 + 3 - 7y_1 = 0, \quad (3.101)$$

$$4y_2^2 + 0.6y_1 + 2.4 - 7y_2 = 0. \quad (3.102)$$

Thus, if we start with the trait (\tilde{g}, \tilde{p}_1) , the invasion probability is approximately 0.199. Whereas it is 0.338 if the first one has trait (\tilde{g}, \tilde{p}_2) . In Figure 3 (A), the mutant population with genotype \tilde{g} survives and the stochastic process is attracted to the new equilibrium $\mathbf{n}^*((g, p), (\tilde{g}, \tilde{p}_1), (\tilde{g}, \tilde{p}_2)) \approx (0, 0.543, 2.554)$, which is a strictly stable.

Table 2: Parameters of Figure 3 (A)

$b(p) = 3$	$d(p) = 1$	$c(p, p) = 1$	$c(p, \tilde{p}_1) = 1$	$c(p, \tilde{p}_2) = 0.7$	$s_{\text{ind}}(\cdot, \cdot)(\cdot) \equiv 0$	$\nu_0^K(g, p) = 2$
$b(\tilde{p}_1) = 2$	$d(\tilde{p}_1) = 1$	$c(\tilde{p}_1, p) = 1$	$c(\tilde{p}_1, \tilde{p}_1) = 1$	$c(\tilde{p}_1, \tilde{p}_2) = 0.5$	$s^{\tilde{g}}(\tilde{p}_1, \tilde{p}_2) = 2$	$\nu_0^K(\tilde{g}, \tilde{p}_1) = K^{-1}$
$b(\tilde{p}_2) = 4$	$d(\tilde{p}_2) = 1$	$c(\tilde{p}_2, p) = 0.7$	$c(\tilde{p}_2, \tilde{p}_1) = 0.5$	$c(\tilde{p}_2, \tilde{p}_2) = 1$	$s^{\tilde{g}}(\tilde{p}_2, \tilde{p}_1) = 0.6$	$\nu_0^K(\tilde{g}, \tilde{p}_2) = 0$

In example (B), $f_{(g,p)}(\tilde{g}, \tilde{p}_1)$ and $f_{(g,p)}(\tilde{g}, \tilde{p}_2)$ are both negative. Nevertheless, the fitness of the genotype is positive and thus the mutant invades with positive probability. (It is given by the largest eigenvalue of $\begin{pmatrix} -3 & 2 \\ 2 & -0.4 \end{pmatrix}$, which equals approximately 0.685.) However, the invasion probability is smaller in this example. It is approximately 0.127 if we start with the trait (\tilde{g}, \tilde{p}_1) and 0.207 else. In Figure 3 (B), the mutant population survives and the process is attracted to the stable fixed point $\mathbf{n}^*((g, p), (\tilde{g}, \tilde{p}_1), (\tilde{g}, \tilde{p}_2)) \approx (0, 1.153, 1.745)$. Hence, this examples illustrate that the usual definition of invasion fitness fails for populations with phenotypic plasticity.

Table 3: Parameters of Figure 3 (B)

$b(p) = 3$	$d(p) = 1$	$c(p, p) = 1$	$c(p, \tilde{p}_1) = 1$	$c(p, \tilde{p}_2) = 0.7$	$s_{\text{ind}}(\cdot, \cdot)(\cdot) \equiv 0$	$\nu_0^K(g, p) = 2$
$b(\tilde{p}_1) = 2$	$d(\tilde{p}_1) = 1$	$c(\tilde{p}_1, p) = 1$	$c(\tilde{p}_1, \tilde{p}_1) = 1$	$c(\tilde{p}_1, \tilde{p}_2) = 0.5$	$s^{\tilde{g}}(\tilde{p}_1, \tilde{p}_2) = 2$	$\nu_0^K(\tilde{g}, \tilde{p}_1) = 1/K$
$b(\tilde{p}_2) = 4$	$d(\tilde{p}_2) = 1$	$c(\tilde{p}_2, p) = 0.7$	$c(\tilde{p}_2, \tilde{p}_1) = 0.5$	$c(\tilde{p}_2, \tilde{p}_2) = 1$	$s^{\tilde{g}}(\tilde{p}_2, \tilde{p}_1) = 2$	$\nu_0^K(\tilde{g}, \tilde{p}_2) = 0$

References

- [1] K. B. Athreya. Some results on multitype continuous time Markov branching processes. *Ann. Math. Stat.*, 39:347–357, 1968. MR-0221600
- [2] K. B. Athreya and P. E. Ney. *Branching processes*. Die Grundlehren der mathematischen Wissenschaften, Vol. 196. Springer-Verlag Berlin Heidelberg, 1972. MR-0373040
- [3] M. Baar, A. Bovier, and N. Champagnat. From stochastic, individual-based models to the canonical equation of adaptive dynamics - in one step. *Ann. Appl. Probab.*, 27:1093–1170, 2017. MR-3655862
- [4] M. Baar, L. Coquille, H. Mayer, M. Hölzel, M. Rogava, T. Tüting, and A. Bovier. A stochastic model for immunotherapy of cancer. *Scientific Reports*, 6:24169, 2016.
- [5] V. Bansaye and S. Méléard. *Stochastic models for structured populations. Scaling limits and long time behavior*, volume 1 of *Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems*. Springer, Cham; MBI Mathematical Biosciences Institute, Ohio State University, Columbus, OH, 2015. MR-3380810
- [6] H. J. E. Beaumont, J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey. Experimental evolution of bet hedging. *Nature*, 462:90–93, 2009.
- [7] B. Bolker and S. W. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.*, 52(3):179 – 197, 1997.
- [8] B. M. Bolker and S. W. Pacala. Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. *Am. Nat.*, 153(6):575–602, 1999.
- [9] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.*, 116(8):1127–1160, 2006. MR-2250806
- [10] N. Champagnat, P-E. Jabin, and S. Méléard. Adaptation in a stochastic multi-resources chemostat model. *J. Math. Pures Appl.*, 101(6):755–788, 2014. MR-3205642
- [11] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Prob. Theory Rel.*, 151(1-2):45–94, 2011. MR-2834712
- [12] P. Collet, S. Méléard, and J. A. J. Metz. A rigorous model study of the adaptive dynamics of Mendelian diploids. *J. Math. Biol.*, 67(3):569–607, 2013. MR-3084364
- [13] D. L. DeAngelis and V. Grimm. Deangelis dl, grimm v. individual-based models in ecology after four decades. *F1000Prime Reports*, 6(39), 2014.
- [14] U. Dieckmann and R. Law. Moment approximations of individual-based models. In U. Dieckmann, R. Law, and J. A. J. Metz, editors, *The geometry of ecological interactions: simplifying spatial complexity*, pages 252–270. Cambridge University Press, 2000. MR-2265382
- [15] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1997. MR-1431744
- [16] S. N. Ethier and T. G. Kurtz. *Markov processes. Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. MR-0838085
- [17] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004. MR-2099656
- [18] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Heidelberg, 3rd edition, 2012. MR-2953753
- [19] G. Fusco and A. Minelli. Phenotypic plasticity in development and evolution: facts and concepts. *Phil. Trans. Royal Soc. London B: Biol. Sci.*, 365:547–556, 2010.
- [20] V. Grimm and S. F. Railsback. *Individual-based modeling and ecology*. Princeton Series in Theoretical and Computational Biology. Princeton University Press, Princeton, NJ, 2005. MR-2153370
- [21] M. Hölzel, A. Bovier, and T. Tüting. Plasticity of tumour and immune cells: a source of heterogeneity and a cause for therapy resistance? *Nat. Rev. Cancer*, 13(5):365–376, 2013.

- [22] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Stat.*, 37:1463–1481, 1966. MR-0200979
- [23] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Stat.*, 37:1211–1223, 1966. MR-0198552
- [24] H. Kesten and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.*, 17:309–338, 1967. MR-0205340
- [25] J. Landsberg, J. Kohlmeyer, M. Renn, T. Bald, M. Rogava, M. Cron, M. Fatho, V. Lennerz, T. Wölfel, M. Hölzel, and T. Tüting. Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. *Nature*, 490(7420):412–416, 10 2012.
- [26] S. Pénisson. *Conditional limit theorems for multitype branching processes and illustration in epidemiological risk analysis*. PhD thesis, Universität Potsdam, Potsdam (Germany), 2010.
- [27] B. A. Sewastjanow. *Verzweigungsprozesse*. R. Oldenbourg Verlag, Munich-Vienna, 1975. MR-0408019

Acknowledgments. We thank two anonymous referees for valuable comments that helped to improve the paper.

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 (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>