

## Interim Report

IR-13-034

### **A new proof for the convergence of an individual based model to the trait substitution sequence**

Ankit Gupta  
Hans J.A.J. Metz ([metz@iiasa.ac.at](mailto:metz@iiasa.ac.at))  
Viet Chi Tran

---

#### **Approved by**

Ulf Dieckmann  
Director, Evolution and Ecology Program

June 2015

# A new proof for the convergence of an individual based model to the Trait substitution sequence

Ankit Gupta\* and J.A.J. Metz<sup>†</sup> and Viet Chi Tran<sup>‡</sup>

March 14, 2012

## Abstract

We consider a continuous time stochastic individual based model for a population structured only by an inherited vector trait and with logistic interactions. We consider its limit in a context from adaptive dynamics: the population is large, the mutations are rare and we view the process in the timescale of mutations. Using averaging techniques due to Kurtz (1992), we give a new proof of the convergence of the individual based model to the trait substitution sequence of Metz et al. (1992) first worked out by Dieckman and Law (1996) and rigorously proved by Champagnat (2006): rigging the model such that “invasion implies substitution”, we obtain in the limit a process that jumps from one population equilibrium to another when mutations occur and invade the population.

Keywords: birth and death process; structured population; adaptive dynamics; individual based model; averaging technique; trait substitution sequence

Mathematical Subject Classification (2000): 92D15; 60J80; 60K35; 60F99

## 1 Introduction: The logistic birth and death model

We consider a stochastic individual based model (IBM) with trait structure and evolving as a result of births and deaths, that has been introduced by Dieckmann and Law [9] and Metz et al. [21] and in rigorous detail by Champagnat [2]. We study its limit in an evolutionary time scale when the population is large and the mutations are rare.

Champagnat [2] established the first rigorous proof of the convergence of a sequence of such IBMs to the trait substitution sequence process (TSS) introduced by Metz et al.[20] (with Metz et al. [21] as follow up) which, following Dieckmann and Law [9], can be explained as follows. In the limit the time scales of ecology and evolution are separated. Mutations are rare and before a mutant arises, the resident population stabilizes around an equilibrium. Under the “invasion implies substitution” Ansatz, there cannot be long term coexistence of two different traits. Evolution thus proceeds as a succession of monomorphic population equilibria. The fine structure of the transitions disappear on the time scale that is considered and, when a mutation occurs, invades, and fixates in the population by completely replacing the resident trait, the TSS jumps

---

\*CMAP, Ecole Polytechnique, UMR CNRS 7641, Route de Saclay, 91128 Palaiseau Cédex, France

<sup>†</sup>Mathematical Institute & Institute of Biology & NCB Naturalis, Leiden, Niels Bohrweg 1, 2333 CA, Leiden, Netherlands ; EEP, IIASA, Laxenburg, Austria

<sup>‡</sup>Equipe Probabilité Statistique, Laboratoire Paul Painlevé, UMR CNRS 8524, UFR de Mathématiques, Université des Sciences et Technologies Lille 1, Cité Scientifique, 59655 Villeneuve d’Ascq Cédex, France ; CMAP, Ecole Polytechnique

from one state to another. Champagnat’s proof is based on some fine estimates, including some fine large deviations results, to combine several approximations of the microscopic process. We separate the different time scales that are involved using averaging techniques due to Kurtz [15], and thus propose a new simplified proof of Champagnat’s results that skips many technicalities linked with fine approximations of the IBM. The aim of this paper is to exemplify the use of such averaging techniques in adaptive dynamics, which we hope will pave the way for generalizations of the TSS.

We consider a structured population where each individual is characterized by a trait  $x \in \mathbb{X}$ , a compact subset of  $\mathbb{R}^d$ . We are interested in large populations. We assume that the population’s initial size to be proportional to a parameter  $K \in \mathbb{N}^* = \{1, 2, \dots\}$ , to be interpreted as the area to which the population is confined, that we will let go to infinity while keeping the density constant by counting individuals with weight  $1/K$ . The population is assumed to be well mixed and its density is assumed to be limited by a fixed availability of resources per unit of area. The population at time  $t$  can be described by the following point measure on  $\mathbb{X}$

$$X^K(t) = \frac{1}{K} \sum_{i=1}^{N^K(t)} \delta_{x_t^i}, \quad (1.1)$$

where  $N^K(t)$  is the total number of individuals in the population at time  $t$  and where  $x_t^i \in \mathbb{X}$  denotes the trait of individual  $i$  living at time  $t$ , the latter being numbered in lexicographical order.

The population evolves by births and deaths. An individual with trait  $x \in \mathbb{X}$  gives birth to new individuals at rate  $b(x)$ , where  $b(x)$  is a continuous positive function on  $\mathbb{X}$ . With probability  $u_K p(x) \in [0, 1]$ , the daughter is a mutant with trait  $y$ , where  $y$  is drawn from the mutation kernel  $m(x, dy)$  supported on  $\mathbb{X}$ . Here  $u_K \in [0, 1]$  is a parameter depending on  $K$  that scales the probability of mutation. With probability  $1 - u_K p(x) \in [0, 1]$ , the daughter is a clone of her mother, with the same value of the trait  $x$ . In a population described by  $X \in \mathcal{M}_F(\mathbb{X})$ , the individual with trait  $x$  dies at rate  $d(x) + \int_{\mathbb{X}} \alpha(x, y) X(dy)$ , where the natural death rate  $d(x)$  and the competition kernel  $\alpha(x, y)$  are positive continuous functions.

**Assumption 1.1** *We assume that the functions  $b$ ,  $d$  and  $\alpha$  satisfy the following hypotheses:*

(A) *For all  $x \in \mathbb{X}$ ,  $b(x) - d(x) > 0$  and  $p(x) > 0$ .*

(B) *“Invasion implies substitution”: For all  $x$  and  $y$  in  $\mathbb{X}$ , we either have:*

$$(b(y) - d(y))\alpha(x, x) - (b(x) - d(x))\alpha(y, x) < 0 \quad (1.2)$$

$$\text{or} \quad \begin{cases} (b(y) - d(y))\alpha(x, x) - (b(x) - d(x))\alpha(y, x) > 0 \\ (b(x) - d(x))\alpha(y, y) - (b(y) - d(y))\alpha(x, y) < 0. \end{cases} \quad (1.3)$$

(C) *There exist  $\underline{\alpha}$  and  $\bar{\alpha} > 0$  such that for every  $x, y \in \mathbb{X}$ :*

$$0 < \underline{\alpha} \leq \alpha(x, y) \leq \bar{\alpha}. \quad (1.4)$$

Part (A) of Assumption 1.1 says that in the absence of competition, the population has a positive natural growth rate. Also the probability of a birth resulting in a mutation is positive. Part (B) corresponds to a condition known in adaptive dynamics as “invasion implies substitution”. It can be obtained from the analysis of the equilibria of the Lotka-Volterra system

that results from the ordinary large number limit of the logistic competition process without mutation. The condition has as consequences that when a mutant population manages to reach a sufficiently large size, it cannot coexist with the resident population: one of the two types has to become extinct. Hence, the population should be monomorphic away from the mutation events.

In this paper, we use the methods of Kurtz [16], based on martingale problems, for separating time scales in measure-valued processes. We show that on the one hand the populations stabilize around their equilibria on the fast ecological scale, while on the other hand, rare mutations at the evolutionary time scale may induce switches from one trait to another. This provides a new proof to the result of Champagnat [2], which we hope may pave the way to generalizations of the TSS. Our proof differs from the one in [2] in that we do not require comparisons with partial differential equations and large deviation results to exhibit the stabilization of the population around the equilibria determined by the resident trait.

In Section 2, we describe the IBM introduced in [2]. The model accounts only for a trait-structure and otherwise has very simple dynamics. Generalization to more general (possibly functional, as first introduced in [8, 22]) trait spaces are considered in [13] where the “invasion implies substitution” assumption is also relaxed (see also [21, 12, 5]). There exist several other possible generalizations of the TSS, including e.g. physiological structure [10, 18] or diploidy [6]. We consider the process that counts the new traits appearing due to mutations, and the occupation measure  $\Gamma^K$  of the process  $X^K$  under a changed time scale. The tightness of the couple of processes is studied in Section 3. The limiting values are shown to satisfy an equation that is considered in Section 4. This equation says that when a favorable mutant appears, then the distribution describing the population jumps to the equilibrium characterized by the new trait with a probability depending on the fitness of the mutant trait compared to the resident one. From the consideration of monomorphic and dimorphic populations and using couplings with branching processes, we prove the convergence in distribution of  $\{\Gamma^K\}$  to the occupation measure  $\Gamma$  of a pure jump Markov process that is called TSS.

**Notation:**

Let  $E$  be a Polish space and let  $\mathcal{B}(E)$  be its Borel sigma field. We denote by  $\mathcal{M}_F(E)$  (resp.  $\mathcal{M}_P(E)$ ) the set of nonnegative finite (resp. point) measures on  $E$ , endowed with the topology of weak convergence. If  $E$  is compact this topology coincides with the topology of vague convergence (see e.g. [14]). If  $E$  is compact, then for any  $M > 0$ , the set  $\{\mu \in \mathcal{M}_F(E) : \mu(E) \leq M\}$  is compact in this topology. For a measure  $\mu$ , we denote its support by  $\text{supp}(\mu)$ . If  $f$  is a bounded measurable function on  $E$  and  $\mu \in \mathcal{M}_F(E)$ , we use the notation:  $\langle \mu, f \rangle = \int_E f(x)\mu(dx)$ . With an abuse of notation,  $\langle \mu, x \rangle = \int_E x\mu(dx)$ . Convergence in distribution of a sequence of random variables (or processes) is denoted by ‘ $\Rightarrow$ ’. The minimum of any two numbers  $a, b \in \mathbb{R}$  is given by  $a \wedge b$  and for any  $a \in \mathbb{R}$ , its positive part is denoted by  $[a]^+$ . For any two  $\mathbb{N}^*$ -valued sequences  $\{a_K : K \in \mathbb{N}^*\}$  and  $\{b_K : K \in \mathbb{N}^*\}$  we say that  $a_K \ll b_K$  if  $a_K/b_K \rightarrow 0$  as  $K \rightarrow \infty$ .

Define a class of test functions on  $\mathcal{M}_F(\mathbb{X})$  by

$$\mathbb{F}_b^2 = \{F_f : F_f(\mu) = F(\langle \mu, f \rangle), f \in C_b(\mathbb{X}, \mathbb{R}) \text{ and } F \in C_b^2(\mathbb{R}, \mathbb{R}) \text{ with compact support}\}.$$

Here  $C_b(\mathbb{X}, \mathbb{R})$  is the set of all continuous and bounded real functions on  $\mathbb{X}$  and  $C_b^2(\mathbb{R}, \mathbb{R})$  is the set of bounded, twice continuously differentiable real-valued functions on  $\mathbb{R}$  with bounded first and second derivative. This class  $\mathbb{F}_b^2$  is separable and it is known (see for example [7]) that it characterizes the convergence in distribution on  $\mathcal{M}_F(\mathbb{X})$ .

The value at time  $t$  of a process  $X$  is denoted  $X(t)$  or sometimes  $X_t$  for notational convenience.

## 2 IBM in the evolutionary time-scale

The process  $X^K$  is characterized by its generator  $L^K$ , defined as follows. To simplify notation, we define the transition kernel:

$$M^K(x, dy) = u_K p(x) m(x, dy) + (1 - u_K p(x)) \delta_x(dy). \quad (2.5)$$

For any  $F_f \in \mathbb{F}_b^2$  let

$$\begin{aligned} L^K F_f(X) &= K \int_E b(x) \left[ \int_{\mathbb{X}} \left( F_f \left( X + \frac{1}{K} \delta_y \right) - F_f(X) \right) M^K(x, dy) \right] X(dx) \\ &+ K \int_E \left( d(x) + \langle X, \alpha(x, \cdot) \rangle \right) \left( F_f \left( X - \frac{1}{K} \delta_x \right) - F_f(X) \right) X(dx). \end{aligned} \quad (2.6)$$

Let  $K \in \mathbb{N}^*$  be fixed. The martingale problem for  $L^K$  has a unique solution for any initial condition  $X^K(0) \in \mathcal{M}_F(\mathbb{X})$ . It is possible to construct the solution of the martingale problem by considering a stochastic differential equation (SDE) driven by Poisson point processes and which corresponds to the IBM used for simulations (see [3, 4]). We need the following estimate to proceed, which is shown in [2, Lemma 1]

**Lemma 2.1** *Suppose that  $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X^K(0), 1 \rangle^2) < \infty$ , then*

$$\sup_{K \geq 1, t \geq 0} \mathbb{E}(\langle X^K(t), 1 \rangle^2) < +\infty.$$

In the sequel, we hence make the following assumptions about the initial condition.

**Assumption 2.2** *Suppose that the sequence of  $\mathcal{M}_F(\mathbb{X})$ -valued random variables  $\{X^K(0) : K \in \mathbb{N}^*\}$  satisfy the following conditions.*

(A) *There exists  $x_0 \in \mathbb{X}$  such that  $\text{supp}(X^K(0)) = \{x_0\}$  for all  $K \in \mathbb{N}^*$ .*

(B)  $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle X^K(0), 1 \rangle^2) < \infty$ .

(C)  $X^K(0) \Rightarrow X(0)$  as  $K \rightarrow \infty$  and  $\langle X(0), 1 \rangle > 0$  a.s.

From (2.6), we can see that the dynamics has two time scales. The slower time scale is of order  $Ku_K$  and it corresponds to the occurrence of new mutants while the faster time scale is of order 1 and it corresponds to the birth-death dynamics. We consider rare mutations and will therefore be lead to suppose that for any  $c > 0$

**Assumption 2.3**

$$\log K \ll \frac{1}{Ku_K} \ll e^{cK}.$$

Consider the process

$$Z^K(t) = X^K \left( \frac{t}{Ku_K} \right), \quad t \geq 0. \quad (2.7)$$

In what follows, we denote by  $\{\mathcal{F}_t^K : t \geq 0\}$  the canonical filtration associated with  $Z^K$ . Due to the change in time, the generator  $\mathbb{L}^K$  of  $Z^K$  is the generator  $L^K$  of  $X^K$  multiplied by  $(1/Ku_K)$ . Hence for any  $F_f \in \mathbb{F}_b^2$

$$\mathbb{L}^K F_f(Z) = \int_{\mathbb{X}} p(x) b(x) \left[ \int_{\mathbb{X}} \left( F_f \left( Z + \frac{1}{K} \delta_y \right) - F_f(Z) \right) m(x, dy) \right] Z(dx)$$

$$\begin{aligned}
& + \frac{1}{Ku_K} \left[ \int_{\mathbb{X}} b(x)(1 - u_K p(x)) K \left( F_f \left( Z + \frac{1}{K} \delta_x \right) - F_f(Z) \right) Z(dx) \right. \\
& \left. + \int_{\mathbb{X}} (d(x) + \langle Z, \alpha(x, \cdot) \rangle) K \left( F_f \left( Z - \frac{1}{K} \delta_x \right) - F_f(Z) \right) Z(dx) \right]. \quad (2.8)
\end{aligned}$$

In the process  $Z^K$  we have compressed time so that new mutants occur at rate of order 1. When we work at this time scale we can expect that between subsequent mutations, the fast birth-death dynamics will average (see e.g. [16]). Our aim is to exploit this separation in the time scales of ecology (which is related to the births and deaths of individuals) and of evolution (which is linked to mutations).

To study the averaging phenomenon, for the fast birth-death dynamics, we use the martingale techniques developed by Kurtz [16]. We introduce the occupation measure  $\Gamma^K$  defined for any  $t \geq 0$  and for any set  $A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X}))$  by

$$\Gamma^K([0, t] \times A) = \int_0^t \mathbb{1}_A(Z^K(s)) ds. \quad (2.9)$$

Kurtz's techniques have been used in the context of measure-valued processes in [19, 17] for different population dynamic problems, but an additional difficulty arises here due to the presence of non-linearities also at the fast time scale.

We introduce a  $\mathcal{M}_P(\mathbb{X})$ -valued process  $\{\chi^K(t) : t \geq 0\}$  which keeps track of the traits that have appeared in the population. That is, for each  $t \geq 0$ ,  $\chi^K(t)$  is a counting measure on  $\mathbb{X}$  that weights the traits that have appeared in the population until time  $t$ . The process  $\chi^K$  is a pure-jump Markov process that satisfies the following martingale problem. For any  $F_f \in \mathbb{F}_b^2$

$$M_t^{\chi^K} := F_f(\chi^K(t)) - F_f(\chi^K(0)) \quad (2.10)$$

$$\begin{aligned}
& - \int_0^t \int_{\mathbb{X}} p(x)b(x) \int_{\mathbb{X}} (F_f(\chi^K(s) + \delta_y) - F_f(\chi^K(s))) m(x, dy) Z^K(s, dx) ds \\
& = F_f(\chi^K(t)) - F_f(\chi^K(0)) \\
& - \int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \left[ \int_{\mathbb{X}} p(x)b(x) \int_{\mathbb{X}} (F_f(\chi^K(s) + \delta_y) - F_f(\chi^K(s))) m(x, dy) \mu(dx) \right] \Gamma^K(ds \times d\mu), \quad (2.11)
\end{aligned}$$

is a square integrable  $\{\mathcal{F}_t^K\}$ -martingale.

The main result of the paper proves the convergence of  $\{(\chi^K, \Gamma^K)\}$  to a limit  $(\chi, \Gamma)$ , where the slow component  $\chi$  is a jump Markov process and the fast component stabilizes in an equilibrium that depends on the value of the slow component.

**Theorem 2.4** *Suppose that Assumptions 1.1 and 2.2 hold.*

(A) *There exists a process  $\chi$  with paths in  $\mathbb{D}(\mathcal{M}_P(\mathbb{X}), [0, \infty))$  and a random measure  $\Gamma \in \mathcal{M}_F([0, \infty) \times \mathcal{M}_F(\mathbb{X}))$  such that  $(\chi^K, \Gamma^K) \Rightarrow (\chi, \Gamma)$  as  $K \rightarrow \infty$ , where  $(\chi, \Gamma)$  are characterized as follows. For all functions  $F_f \in \mathbb{F}_b^2$ ,*

$$\begin{aligned}
& F_f(\chi(t)) - F_f(\chi(0)) \\
& - \int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \left[ \int_{\mathbb{X}} b(x)p(x) \int_{\mathbb{X}} (F_f(\chi(s) + \delta_y) - F_f(\chi(s))) m(x, dy) \mu(dx) \right] \Gamma(ds \times d\mu) \quad (2.12)
\end{aligned}$$

is a square integrable martingale for the filtration

$$\mathcal{F}_t = \sigma \{ \chi(s), \Gamma([0, s] \times A) : s \in [0, t], A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X})) \} \quad (2.13)$$

and for any  $t \geq 0$

$$\int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu) \Gamma(ds \times d\mu) = 0 \text{ for all } F_f \in \mathbb{F}_b^2 \text{ a.s.}, \quad (2.14)$$

where the nonlinear operator  $\mathbb{B}$  is defined by

$$\mathbb{B}F_f(\mu) = F'(\langle \mu, f \rangle) \int_{\mathbb{X}} (b(x) - (d(x) + \langle \mu, \alpha(x, \cdot) \rangle)) f(x) \mu(dx). \quad (2.15)$$

(B) Moreover for any  $t > 0$  and  $A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X}))$  we have

$$\Gamma([0, t] \times A) = \int_0^t \mathbb{1}_A(\hat{n}_{\chi'(s)} \delta_{\chi'(s)}) ds, \quad (2.16)$$

where  $\{\chi'(t) : t \geq 0\}$  is a  $\mathbb{X}$ -valued Markov jump process with  $\chi'(0) = x_0$  and generator given by

$$\mathbb{C}f(x) = b(x)p(x)\hat{n}_x \int_{\mathbb{X}} \frac{[\text{Fit}(y, x)]^+}{b(y)} (f(y) - f(x)) m(x, dy), \quad (2.17)$$

for any  $f \in C_b(\mathbb{X}, \mathbb{R})$ . Here the population equilibrium  $\hat{n}_x$  and the fitness function  $\text{Fit}(y, x)$  are given by:

$$\hat{n}_x = (b(x) - d(x))/\alpha(x, x), \quad \text{and} \quad \text{Fit}(y, x) = b(y) - d(y) - \alpha(y, x)\hat{n}_x. \quad (2.18)$$

This fitness function provides information on the stability of the Lotka-Volterra system (see (4.27) below) and describes the growth rate of a negligible mutant population with trait  $y$  in an environment characterized by  $\hat{n}_x$ .

### 3 Tightness of $\{(\chi^K, \Gamma^K)\}$

To study the limit when  $K \rightarrow +\infty$ , we proceed by a tightness-uniqueness argument. First, we show the tightness of the distributions of  $\{(\chi^K, \Gamma^K) : K \in \mathbb{N}^*\}$  and derive certain properties of the limiting distribution. The limiting values of  $\{\Gamma^K\}$  satisfy an equation that characterizes the state of the population between two mutations, thanks to the ‘‘invasion implies substitution’’ assumption.

**Theorem 3.1** *Suppose that Assumption 1.1 is satisfied and  $\sup_{K \geq 1} \mathbb{E}(\langle X^K(0), 1 \rangle^2) < \infty$ . Then:*

(A) *The distributions of  $\{(\chi^K, \Gamma^K) : K \in \mathbb{N}^*\}$  are tight in the space*

$$\mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathbb{X})) \times \mathcal{M}_F(\mathbb{R}_+ \times \mathcal{M}_F(\mathbb{X}))).$$

(B) *Suppose that  $(\chi^K, \Gamma^K) \Rightarrow (\chi, \Gamma)$ , along some subsequence, as  $K \rightarrow \infty$ . Then  $\chi$  is characterized by the martingale problem given by (2.12) and  $\Gamma$  satisfies the equation (2.14).*

In fact, the Assumption 2.3 can be relaxed and the proof of Theorem 3.1 only requires that  $\lim_{K \rightarrow +\infty} Ku_K = 0$ .

**Proof.**[Proof of Th. 3.1] To prove the tightness of  $\{\chi^K : K \geq 1\}$ , we use a criterion from [11]. Let  $n^K(t) = \langle \chi^K(t), 1 \rangle$  for  $t \geq 0$ . This process counts the number of mutations in the population. For any  $T > 0$  we have by Lemma 2.1 that

$$\sup_{K \geq 1} \mathbb{E}(n^K(T)) \leq \|b\|_\infty T \sup_{K \geq 1, t \leq T} \mathbb{E}(\langle Z^K(t), 1 \rangle) \leq \|b\|_\infty T \sup_{K \geq 1, t \geq 0} \mathbb{E}(\langle X^K(t), 1 \rangle) < \infty. \quad (3.19)$$

From this estimate and the martingale problem (2.11), it can be checked by using Aldous and Rebolledo criteria that for every test function  $f \in C_b(\mathbb{X}, \mathbb{R})$ , the laws of  $\langle \chi^K, f \rangle$  are tight in  $\mathbb{D}([0, \infty), \mathbb{R})$  and that the compact containment condition is satisfied.

Let us now prove the tightness of  $\{\Gamma^K : K \in \mathbb{N}^*\}$ . Let  $\epsilon > 0$  be fixed. Using Lemma 2.1, there exists a  $N_\epsilon > 0$  such that

$$\sup_{K \geq 1, t \geq 0} \mathbb{P}(\langle Z^K(t), 1 \rangle > N_\epsilon) < \epsilon. \quad (3.20)$$

Since  $\mathbb{X}$  is compact, the set  $\mathcal{K}_\epsilon = \{\mu \in \mathcal{M}_F(\mathbb{X}), \langle \mu, 1 \rangle \leq N_\epsilon\}$  is compact. We deduce that for any  $T > 0$

$$\inf_{K \geq 1} \mathbb{E}(\Gamma^K([0, T] \times \mathcal{K}_\epsilon)) \geq (1 - \epsilon)T. \quad (3.21)$$

Indeed

$$\Gamma^K([0, T] \times \mathcal{K}_\epsilon) = \Gamma^K([0, T] \times \mathcal{M}_F(\mathbb{X})) - \Gamma^K([0, T] \times \mathcal{K}_\epsilon^c) = T - \int_0^T \mathbb{1}_{\mathcal{K}_\epsilon^c}(Z^K(t)) dt$$

and the result follows from the Fubini theorem and from (3.20). From Lemma 1.3 of [16],  $\{\Gamma^K : K \in \mathbb{N}^*\}$  is a tight family of random measures. The joint tightness of  $\{(\chi^K, \Gamma^K) : K \in \mathbb{N}^*\}$  is immediate from the tightness of  $\{\chi^K : K \in \mathbb{N}^*\}$  and  $\{\Gamma^K : K \in \mathbb{N}^*\}$ . This proves part (A).

We now prove part (B). Our proof is adapted from the proof of Theorem 2.1 in [16]. From part (A) we know that the distributions  $\{(\chi^K, \Gamma^K) : K \in \mathbb{N}^*\}$  are tight. Therefore there exists a subsequence  $\{\eta_K\}$  along which  $(\chi^K, \Gamma^K)$  converges in distribution to a limit  $(\chi, \Gamma)$ . We can take the limit in (2.11) along this subsequence. Except for a denumerable subset of times,  $M_t^{\chi, K}$  converges in distribution to the martingale given by (2.12).

Let us now show that the limiting value  $\Gamma$  satisfies (2.14). From (2.8) for any  $F_f \in \mathbb{F}_b^2$ , we get that

$$\begin{aligned} m_t^{F, f, K} &= F_f(Z^K(t)) - F_f(Z^K(0)) - \int_0^t \mathbb{L}^K F_f(Z^K(s)) ds \\ &= F_f(Z^K(t)) - F_f(Z^K(0)) - \left( \frac{1}{Ku_K} \right) \int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B} F_f(\mu) \Gamma^K(ds \times d\mu) - \frac{\delta^{F, f, K}(t)}{Ku_K} \end{aligned} \quad (3.22)$$

is a martingale. Here the operator  $\mathbb{B}$  is defined by (2.15) and

$$\delta^{F, f, K}(t) = \int_0^t (Ku_K \mathbb{L}^K F_f(Z^K(s)) - \mathbb{B} F_f(Z^K(s))) ds. \quad (3.23)$$

For any  $\mu \in \mathcal{M}_F(\mathbb{X})$  we have



$$\begin{aligned}
& Ku_K \mathbb{L}^K F_f(\mu) - \mathbb{B}F_f(\mu) \\
&= Ku_K \int_{\mathbb{X}} p(x)b(x) \left[ \int_{\mathbb{X}} \left( F_f \left( \mu + \frac{1}{K} \delta_y \right) - F_f(\mu) \right) m(x, dy) \right] \mu(dx) \\
&+ K \int_{\mathbb{X}} b(x) \left( F_f \left( \mu + \frac{1}{K} \delta_x \right) - F_f(\mu) - \frac{1}{K} F'(\langle \mu, f \rangle) f(x) \right) \mu(dx) \\
&+ K \int_{\mathbb{X}} (d(x) + \langle \mu, \alpha(x, \cdot) \rangle) \left( F_f \left( \mu - \frac{1}{K} \delta_x \right) - F_f(\mu) + \frac{1}{K} F'(\langle \mu, f \rangle) f(x) \right) \mu(dx) \\
&- Ku_K \int_{\mathbb{X}} p(x)b(x) \left( F_f \left( \mu + \frac{1}{K} \delta_x \right) - F_f(\mu) \right) \mu(dx).
\end{aligned}$$

For any  $x \in \mathbb{X}$  and  $\mu \in \mathcal{M}_F(\mathbb{X})$ , we have by Taylor expansion that for some  $\alpha_1, \alpha_2 \in (0, 1)$ :

$$\begin{aligned}
F_f \left( \mu \pm \frac{1}{K} \delta_x \right) - F_f(\mu) &= \pm F' \left( \langle \mu, f \rangle + \alpha_1 \frac{f(x)}{K} \right) \frac{f(x)}{K} \\
&= \pm F'(\langle \mu, f \rangle) \frac{f(x)}{K} + \frac{f(x)^2}{2K^2} F'' \left( \langle \mu, f \rangle + \alpha_2 \frac{f(x)}{K} \right).
\end{aligned}$$

Therefore we get

$$\left| F_f \left( \mu \pm \frac{1}{K} \delta_x \right) - F_f(\mu) \right| \leq \frac{\|F'\|_{\infty} \|f\|_{\infty}^2}{K}.$$

and

$$\left| K \left( F_f \left( \mu \pm \frac{1}{K} \delta_x \right) - F_f(\mu) \mp \frac{1}{K} F'(\langle \mu, f \rangle) f(x) \right) \right| \leq \frac{\|F''\|_{\infty} \|f\|_{\infty}^2}{2K}.$$

Using these estimates and Assumption 1.1,

$$\begin{aligned}
|Ku_K \mathbb{L}^K F_f(\mu) - \mathbb{B}F_f(\mu)| &\leq 2u_K \|b\|_{\infty} \|F'\|_{\infty} \|f\|_{\infty}^2 \langle \mu, 1 \rangle \\
&+ \frac{\|F''\|_{\infty} \|f\|_{\infty}^2}{2K} ((\|b\|_{\infty} + \|d\|_{\infty}) \langle \mu, 1 \rangle + \bar{\alpha} \langle \mu, 1 \rangle^2).
\end{aligned}$$

Pick any  $T > 0$ . This estimate along with Lemma 2.1 implies that as  $K \rightarrow +\infty$ ,  $\delta^{F, f, K}(t)$  (given by (3.23)) converges to 0 in  $L^1(d\mathbb{P})$ , uniformly in  $t \in [0, T]$ . Multiplying (3.22) by  $Ku_K$ , we get that along the subsequence  $\eta_K$ , the sequence of martingales  $\{Ku_K m^{F, f, K} : K \in \mathbb{N}^*\}$  converge in  $L^1(d\mathbb{P})$ , uniformly in  $t \in [0, T]$  to  $\int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu) \Gamma(ds \times d\mu)$ . The limit is itself a martingale. Since it is continuous and has paths of bounded variation, it must be 0 at all times a.s. Hence for any  $F_f \in \mathbb{F}_b^2$ ,

$$\int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu) \Gamma(ds \times d\mu) = 0 \text{ a.s.}$$

Since  $\mathbb{F}_b^2$  is separable (2.14) also holds. □

## 4 Characterization of the limiting values

### 4.1 Dynamics without mutation

As in [2, 3], to understand of the information provided by (2.14), we are led to consider the dynamics of monomorphic and dimorphic populations. Our purpose in this section is to show

that (2.15) and Assumption 1.1 (B) characterize the state of the population between two rare mutant invasions. Because of Assumption 1.1 (B), we will see that two different traits cannot coexist in the long term and thus, we will see that it suffices to work with monomorphic or dimorphic initial populations (i.e. the support of  $Z_0^K$  is one or two singletons).

In Subsection 4.1.1, we consider monomorphic or dimorphic populations and show convergence of the occupation measures when the final composition in trait of the population is known. For instance, if the remaining trait is  $x_0$ , then the occupation measure of  $Z^K(dx, dt)$  converges to  $\widehat{n}_{x_0} \delta_{x_0}(dx) dt$ . In Subsection 4.1.2, we use couplings with birth and death processes to show that the distribution of the final trait composition of the population can be computed from the fitness of the mutant and the resident.

#### 4.1.1 Convergence of the occupation measure $\Gamma^K$ in absence of mutation

First, we show that the “invasion implies substitution” Assumption 1.1 (B) provides information on the behavior of a dimorphic process when we know which trait is fixed.

**Definition 4.1** *Let  $L_0^K$  be the operator  $L^K$  (given by (2.6)) with  $p(x) = 0$  for all  $x \in \mathbb{X}$ . We will denote by  $Y^K$  a process with generator  $L_0^K$  and with a initial condition that varies according to the case that is studied.  $Y^K$  has the same birth-death dynamics as a process with generator  $L^K$ , but there are no mutations.*

In this section we investigate how a process with generator  $L_0^K$  behaves at time scales of order  $1/Ku_K$ , when the population is monomorphic or dimorphic. We start by proving a simple proposition.

**Proposition 4.2** *For any  $x, y \in \mathbb{X}$  suppose that  $\pi \in \mathcal{P}(\mathcal{M}_F(\mathbb{X}))$  is such that*

$$\pi(\{\mu \in \mathcal{M}_F(\mathbb{X}) : \{x\} \subset \text{supp}(\mu) \subset \{x, y\}\}) = 1 \quad (4.24)$$

and

$$\int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu) \pi(d\mu) = 0 \quad (4.25)$$

for all  $F_f \in \mathbb{F}_b^2$ . Then for any  $A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X}))$  we have  $\pi(A) = \mathbb{1}_A(\widehat{n}_x \delta_x)$  where  $\widehat{n}_x$  has been defined in (2.18).

**Proof.** Since  $\pi$  satisfies (4.24), any  $\mu$  picked from the distribution  $\pi$  has the form  $\mu = n_x \delta_x + n_y \delta_y$  with  $n_x > 0$ . Let  $\Phi$  be the map from  $\mathcal{M}_F(\mathbb{X})$  to  $\mathbb{R}_+ \times \mathbb{R}_+$  defined by

$$\Phi(\mu) = (\langle \mu, \mathbb{1}_{\{x\}} \rangle, \langle \mu, \mathbb{1}_{\{y\}} \rangle).$$

Let  $\pi^* = \pi \circ \Phi^{-1} \in \mathcal{P}(\mathbb{R}_+ \times \mathbb{R}_+)$  be the image distribution of  $\pi$  by  $\Phi^{-1}$ . From (4.25), replacing  $\mathbb{B}F_f$  by its definition we obtain

$$\begin{aligned} 0 &= \int_{\mathcal{M}_F(\mathbb{X})} F'(\langle \mu, f \rangle) \left( \langle \mu, (b-d)f \rangle - \int_E f(x) \langle \mu, \alpha(x, \cdot) \rangle \mu(dx) \right) \pi(d\mu) \\ &= \int_{\mathbb{R}_+ \times \mathbb{R}_+} F'(f(x)n_x + f(y)n_y) [(b(x) - d(x) - \alpha(x, x)n_x - \alpha(x, y)n_y) n_x f(x) \\ &\quad + (b(y) - d(y) - \alpha(y, x)n_x - \alpha(y, y)n_y) n_y f(y)] \pi^*(dn_x, dn_y). \end{aligned}$$

This equation holds for all  $F_f \in \mathbb{F}_b^2$  only if the support of  $\pi^*$  consists of  $(n_x, n_y)$  with  $n_x > 0$  that satisfy  $(b(x) - d(x) - \alpha(x, x)n_x - \alpha(x, y)n_y)n_x = 0$  and  $(b(y) - d(y) - \alpha(y, x)n_x - \alpha(y, y)n_y)n_y = 0$ . The only possible solutions are  $(\hat{n}_x, 0)$  and

$$\begin{aligned} & (\tilde{n}_x, \tilde{n}_y) \tag{4.26} \\ &= \left( \frac{(b(x) - d(x))\alpha(y, y) - (b(y) - d(y))\alpha(x, y)}{\alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x)}, \frac{(b(y) - d(y))\alpha(x, x) - (b(x) - d(x))\alpha(y, x)}{\alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x)} \right). \end{aligned}$$

However due to Assumptions 1.1, either  $\tilde{n}^x$  or  $\tilde{n}^y$  is negative and hence  $(\tilde{n}^x, \tilde{n}^y)$  cannot be in the support of  $\pi^*$ . Therefore  $\pi^* (\{(\hat{n}_x, 0)\}) = 1$  and this proves the proposition.  $\square$

**Remark 4.3** Note that  $(0, 0)$ ,  $(\hat{n}_x, 0)$ ,  $(0, \hat{n}_y)$  and  $(\tilde{n}_x, \tilde{n}_y)$  are the stationary solutions of the following ordinary differential equation that approximates a large population with trait  $x$  and  $y$ :

$$\begin{aligned} \frac{dn_x}{dt} &= n_x(t) \left( b(x) - d(x) - \alpha(x, x)n_x(t) - \alpha(x, y)n_y(t) \right) \\ \frac{dn_y}{dt} &= n_y(t) \left( b(y) - d(y) - \alpha(y, x)n_x(t) - \alpha(y, y)n_y(t) \right). \end{aligned} \tag{4.27}$$

Heuristically, the ‘‘invasion implies substitution’’ assumption prevents two traits from coexisting in the long run. If we know which of the trait fixates, Proposition 4.2 provides the form of the solution  $\pi$  to (4.25). In this case, we can deduce the convergence of the occupation measure of  $Y^K(\cdot/Ku_K)$ .

**Corollary 4.4** Let  $x, y \in \mathbb{X}$ . For each  $K \in \mathbb{N}^*$ , let  $\{Y^K(t) : t \geq 0\}$  be a process with generator  $L_0^K$  and  $\text{supp}(Y^K(0)) = \{x, y\}$ . Let  $T > 0$ , and suppose that there exists a  $\delta > 0$  such that:

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( Y_t^K \{x\} < \delta \text{ for some } t \in \left[ 0, \frac{T}{Ku_K} \right] \right) = 0. \tag{4.28}$$

Then for any  $F_f \in \mathbb{F}_b^2$ ,

$$\int_0^T \int_{\mathcal{M}_F(\mathbb{X})} F_f(\mu) \Gamma_0^K(dt \times d\mu) := \int_0^T F_f \left( Y^K \left( \frac{t}{Ku_K} \right) \right) dt \Rightarrow T \times F_f(\hat{n}_x \delta_x)$$

as  $K \rightarrow \infty$ .

**Proof.** As in part (A) of Theorem 3.1 we can show that  $\{\Gamma_0^K : K \in \mathbb{N}^*\}$  is tight in the space  $\mathcal{P}(\mathcal{M}_F([0, T] \times \mathcal{M}_F(\mathbb{X})))$ . Let  $\Gamma_0$  be a limit point. Then from part (C) of Theorem 3.1 we get that

$$\int_0^T \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu) \Gamma_0(dt \times d\mu) = 0 \text{ for all } F_f \in \mathbb{F}_b^2 \text{ a.s.}, \tag{4.29}$$

where the operator  $\mathbb{B}$  is given by (2.15).

Since  $\text{supp}(Y^K(0)) \subset \{x, y\}$  we also have that  $\text{supp}(Y^K(t)) \subset \{x, y\}$  for all  $t \geq 0$ . Let

$$\mathcal{S}_\delta = \{\mu \in \mathcal{M}_F(\mathbb{X}) : \mu\{x\} \leq \delta\}.$$

Observe that  $\Gamma_0^K([0, T] \times \mathcal{S}_\delta) \leq T$  a.s. and

$$0 \leq \mathbb{E} (T - \Gamma_0^K([0, T] \times \mathcal{S}_\delta)) = Ku_K \mathbb{E} \left[ \int_0^{\frac{T}{Ku_K}} (1 - \mathbb{1}_{\mathcal{S}_\delta}(Y^K(t))) dt \right]$$

$$\leq T \mathbb{P} \left( Y_t^K \{x\} < \delta \text{ for some } t \in \left[ 0, \frac{T}{Ku_K} \right] \right).$$

Hence by (4.28) we get that  $\Gamma_0^K([0, T] \times \mathcal{S}_\delta)$  converges to  $T$  in  $L^1(d\mathbb{P})$ . Because  $\mathcal{S}_\delta$  is a closed set,  $\Gamma_0([0, T] \times \mathcal{S}_\delta) = T$  a.s.

Let  $\pi$  be the  $\mathcal{P}(\mathcal{M}_F(\mathbb{X}))$ -valued random variable defined by  $\pi(A) = \Gamma_0([0, T] \times A)/T$  for any  $A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X}))$ . Then  $\pi(\mathcal{S}_\delta) = 1$  and hence  $\pi$  satisfies (4.24) almost surely. Furthermore  $\int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu)\pi(d\mu) = 0$  for all  $F_f \in \mathbb{F}_b^2$ , almost surely. Therefore using Proposition 4.2 proves this corollary.  $\square$

#### 4.1.2 Fixation probabilities

We have seen in Corollary 4.4 that the behavior of a dimorphic population is known provided we know which trait fixates (see (4.28)). Following Champagnat et al. [2, 3], we can answer this question by using couplings with branching processes. We consider the process  $Y^K$  started with a monomorphic or dimorphic initial condition and examine the fixation probabilities in a time interval of order  $1/Ku_K$ .

We begin with some notation. For any  $x \in \mathbb{X}$  and  $\epsilon > 0$  let

$$\mathcal{N}_\epsilon(x) = \{\mu \in \mathcal{M}_F(\mathbb{X}) : \text{supp}(\mu) = \{x\} \text{ and } \langle \mu, 1 \rangle \in [\hat{n}_x - \epsilon, \hat{n}_x + \epsilon]\}. \quad (4.30)$$

Let  $K \in \mathbb{N}^*$ ,  $x, y \in \mathbb{X}$ . For the process  $Y^K$  of Definition 4.1 with initial condition  $Y^K(0) = (z_1^K/K)\delta_x + (z_2^K/K)\delta_y$ , we define the  $\mathbb{N}$ -valued processes  $N_1^K$  and  $N_2^K$  by  $N_1^K(t) = K \langle Y^K(t), \mathbb{1}_{\{x\}} \rangle$  and  $N_2^K(t) = K \langle Y^K(t), \mathbb{1}_{\{y\}} \rangle$ . Then  $\{N^K(t) = (N_1^K(t), N_2^K(t)) : t \geq 0\}$  is the  $\mathbb{N} \times \mathbb{N}$ -valued Markov jump process with transition rates given by

$$\begin{aligned} mb_1 &:= mb(x) && \text{from } (m, n) \text{ to } (m+1, n) \\ nb_2 &:= nb(y) && \text{from } (m, n) \text{ to } (m, n+1) \\ md_1(m, n) &:= m \left( d(x) + \alpha(x, x) \frac{m}{K} + \alpha(x, y) \frac{n}{K} \right) && \text{from } (m, n) \text{ to } (m-1, n) \\ nd_2(m, n) &:= n \left( d(y) + \alpha(y, x) \frac{m}{K} + \alpha(y, y) \frac{n}{K} \right) && \text{from } (m, n) \text{ to } (m, n-1). \end{aligned} \quad (4.31)$$

and with initial state  $(N_1^K(0), N_2^K(0)) = (z_1^K, z_2^K)$ . Notice that  $d_1, d_2$  are increasing functions in  $n$  and  $m$ .

Let also  $\mathcal{S}_K = [A_K, B_K] \times [C_K, D_K]$  be a subset of  $\mathbb{R}_+^2$  and define

$$T_{\mathcal{S}_K} = \inf \{t \geq 0 : N^K(t) \notin \mathcal{S}_K\}. \quad (4.32)$$

The results at the end of this section require comparisons with branching processes, which we recall in the next proposition whose proof is provided in the Appendix.

**Proposition 4.5** *For each  $K \in \mathbb{N}^*$  let  $A_K, B_K, C_K, D_K \in \mathbb{R}_+$  and let  $\mathcal{S}_K$  and  $(N_1^K, N_2^K)$  be defined as above, with  $N^K(0) \in \mathcal{S}_K$ . For parts (A) and (B), assume that  $\eta > 0$ ,  $\epsilon \in (0, 1)$  are such that  $[K\eta(1 - \epsilon), K\eta(1 + \epsilon)] \subset [A_K, B_K]$ . Let*

$$\begin{aligned} d_1^-(\eta, \epsilon) &= \inf_{K \in \mathbb{N}^*} \inf \{d_1(m, n) : (m, n) \in [K\eta(1 - \epsilon), K\eta(1 + \epsilon)] \times [C_K, D_K]\} \\ d_1^+(\eta, \epsilon) &= \sup_{K \in \mathbb{N}^*} \sup \{d_1(m, n) : (m, n) \in [K\eta(1 - \epsilon), K\eta(1 + \epsilon)] \times [C_K, D_K]\}. \end{aligned}$$

(A) If  $b_1 < d_1^-(\eta, \epsilon)$  and  $N_1^K(0) \leq K\eta$  then  $N_1^K$  exists  $[A_K, B_K)$  by  $A_K$  a.s.: for any  $T > 0$

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}_K} \leq \frac{T}{Ku_K}, N_1^K(T_{\mathcal{S}_K}) \geq B_K \right) = 0.$$

(B) If  $b_1 > d_1^+(\eta, \epsilon)$  and  $N_1^K(0) \geq K\eta$  then  $N_1^K$  exists  $[A_K, B_K)$  by  $B_K$  a.s.: for any  $T > 0$

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}_K} \leq \frac{T}{Ku_K}, N_1^K(T_{\mathcal{S}_K}) < A_K \right) = 0.$$

Now assume that there exist  $\epsilon_1, \epsilon_2 \in (0, 1)$  and  $\eta_1 < \eta_2$  such that  $[K\eta_i(1 - \epsilon_i), K\eta_i(1 + \epsilon_i)] \subset [A_K, B_K)$  for  $i = 1, 2$  and  $d_1^+(\eta_1, \epsilon_1) < b_1 < d_1^-(\eta_2, \epsilon_2)$ . Also suppose that  $N_1^K(0) \in [K\eta_1, K\eta_2]$ .

(C) Then for any  $T > 0$

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}_K} \leq \frac{T}{Ku_K}, N_1^K(T_{\mathcal{S}_K}) \notin [A_K, B_K) \right) = 0.$$

Let  $[C_K, D_K) = [1, K\epsilon)$  for some  $\epsilon > 0$  and define  $d_2^-(\epsilon), d_2^+(\epsilon)$

$$\begin{aligned} d_2^-(\epsilon) &:= \inf_{K \in \mathbb{N}^*} \inf \{ d_2(m, n) : (m, n) \in [A_K, B_K) \times [1, K\epsilon) \} \\ &\leq \sup_{K \in \mathbb{N}^*} \sup \{ d_2(m, n) : (m, n) \in [A_K, B_K) \times [1, K\epsilon) \} := d_2^+(\epsilon). \end{aligned}$$

Let  $t_K$  be any  $\mathbb{N}^*$ -valued sequence such that  $\log K \ll t_K \ll \frac{1}{Ku_K}$

(D) If  $b_2 < d_2^-(\epsilon)$  and  $N_2^K(0) \leq K\epsilon/2$  then

$$\lim_{K \rightarrow \infty} \mathbb{P} (T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) = 0) = 1.$$

(E) If  $b_2 > d_2^+(\epsilon)$  and  $N_2^K(0) = 1$  then

$$\begin{aligned} 1 - \frac{d_2^+(\epsilon)}{b_2} &\leq \lim_{K \rightarrow \infty} \mathbb{P} (T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) \geq \epsilon K) \leq 1 - \frac{d_2^-(\epsilon)}{b_2} \\ \text{and} \quad \frac{d_2^-(\epsilon)}{b_2} &\leq \lim_{K \rightarrow \infty} \mathbb{P} (T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) = 0) \leq \frac{d_2^+(\epsilon)}{b_2}. \end{aligned}$$

Using the results of Proposition 4.5, we can deduce the following result for the fixation probabilities for a branching process  $Y^K$  with a dimorphic initial condition and for the behavior at times of order  $1/Ku_K$ .

**Proposition 4.6** For each  $K \in \mathbb{N}^*$  let  $\{Y^K(t) : t \geq 0\}$  be a process with generator  $L_0^K$ . Pick two traits  $x, y \in \mathbb{X}$  initially present in the population and two  $\mathbb{N}^*$ -valued sequences  $\{z_1^K : K \in \mathbb{N}^*\}$  and  $\{z_2^K : K \in \mathbb{N}^*\}$  that give the sizes of the populations of trait  $x$  and  $y$ . Assume that

$$Y^K(0) = \frac{z_1^K}{K} \delta_x + \frac{z_2^K}{K} \delta_y.$$

Then we have the following for any  $T > 0$ .

(A) A monomorphic population with a sufficiently large size does not die:  
 Suppose that for some  $\epsilon > 0$ ,  $z_1^K \geq K\epsilon$  and  $z_2^K = 0$  for each  $K \in \mathbb{N}^*$ . Then for some  $\delta > 0$

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \exists t \in \left[ 0, \frac{T}{Ku_K} \right], Y_t^K \{x\} < \delta \right) = 0. \quad (4.33)$$

(B) A monomorphic population with a size around  $\hat{n}_x$  remains there:  
 Suppose that for some  $\epsilon > 0$ ,  $z_1^K \in [K(\hat{n}_x - \epsilon), K(\hat{n}_x + \epsilon)]$  and  $z_2^K = 0$  for each  $K \in \mathbb{N}^*$ .  
 Then

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \exists t \in \left[ 0, \frac{T}{Ku_K} \right], Y^K(t) \notin \mathcal{N}_{2\epsilon}(x) \right) = 0. \quad (4.34)$$

(C) A favorable mutant with a non-negligible size in a resident population near equilibrium fixates a.s.:

Suppose that  $\text{Fit}(y, x) > 0$  and for some  $\epsilon > 0$ ,  $z_1^K < K(\hat{n}_x + \epsilon)$  and  $z_2^K > K\epsilon$  for all  $K \in \mathbb{N}^*$ . Then there exists an  $\epsilon_0 > 0$  such that if  $\epsilon < \epsilon_0$  then

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \exists t \in \left[ 0, \frac{T}{Ku_K} \right], Y_t^K \{y\} < \frac{\epsilon}{2} \right) = 0. \quad (4.35)$$

For parts (D) and (E), we consider a small mutant population in a resident population near its equilibrium. We assume that for some  $\epsilon > 0$  we have  $z_1^K \in [K(\hat{n}_x - \epsilon), K(\hat{n}_x + \epsilon)]$  and  $z_2^K \in [1, K\epsilon]$  for all  $K \in \mathbb{N}^*$ . Let  $t_K$  be any  $\mathbb{N}^*$ -valued sequence such that  $\log K \ll t_K \ll 1/Ku_K$ . Let  $\mathcal{S}_K = [K(\hat{n}_x - 2\epsilon), K(\hat{n}_x + 2\epsilon)] \times [1, 2K\epsilon]$  and let  $T_{\mathcal{S}_K}$  be the associated stopping time (4.32).

(D) An unfavorable mutant dies out in a time  $t_K$ .

Let  $\text{Fit}(y, x) < 0$ . There exists an  $\epsilon_0 > 0$  such that if  $\epsilon < \epsilon_0$  then

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}_K} \leq t_K, Y_{T_{\mathcal{S}_K}}^K \{y\} = 0 \right) = 1. \quad (4.36)$$

(E) A favorable mutant invades with probability  $\text{Fit}(y, x)/b(y)$ .

Let  $\text{Fit}(y, x) > 0$  and  $z_2^K = 1$  for all  $K \in \mathbb{N}^*$ . Then there exist positive constants  $c, \epsilon_0$  such that for all  $\epsilon < \epsilon_0$  we have

$$\lim_{K \rightarrow \infty} \left| \mathbb{P} \left( T_{\mathcal{S}_K} \leq t_K, Y_{T_{\mathcal{S}_K}(\epsilon)}^K \{y\} \geq 2\epsilon \right) - \frac{\text{Fit}(y, x)}{b(y)} \right| \leq c\epsilon, \quad (4.37)$$

$$\text{and} \quad \lim_{K \rightarrow \infty} \left| \mathbb{P} \left( T_{\mathcal{S}_K} \leq t_K, Y_{T_{\mathcal{S}_K}(\epsilon)}^K \{y\} = 0 \right) - \left( 1 - \frac{\text{Fit}(y, x)}{b(y)} \right) \right| \leq c\epsilon. \quad (4.38)$$

**Proof.** Let us prove part (A). Since  $z_2^K = 0$ ,  $N_2^K(t) = 0$  for all  $t \geq 0$  and  $K \in \mathbb{N}^*$ . Note that  $b(x) - d(x) > 0$  and so we can pick a  $\delta > 0$  such that  $b(x) - d(x) - 2\alpha(x, x)\delta > 0$  and  $2\delta < \epsilon$ . Let  $\mathcal{S}'_K = [K\delta, \infty) \times \{0\}$  and let  $T_{\mathcal{S}'_K}$  be given by (4.32). Then

$$\begin{aligned} \lim_{K \rightarrow \infty} \mathbb{P} \left( Y_t^K \{x\} < \delta \text{ for some } t \in \left[ 0, \frac{T}{Ku_K} \right] \right) &\leq \lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}'_K} \leq \frac{T}{Ku_K} \right) \\ &= \lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}'_K} \leq \frac{T}{Ku_K}, N_1^K \left( T_{\mathcal{S}'_K} \right) < K\delta \right). \end{aligned}$$

The last equality above follows from the fact that the only way to exit the set  $\mathcal{S}'_K$  is by having  $N_1^K$  go below  $K\delta$ . Observe that on the set  $[K\delta, 2K\delta] \times \{0\}$ , the supremum of  $d_1$  is bounded above by  $d(x) + 2\alpha(x, x)\delta$  which is less than  $b(x)$ . Therefore by part (B) of Proposition 4.5:

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}'_K} \leq \frac{T}{Ku_K}, N_1^K (T_{\mathcal{S}'_K}) < K\delta \right) = 0,$$

which proves part (A).

For part (B) we can choose  $\epsilon$  sufficiently small such that  $\hat{n}_x > 2\epsilon$ . Let  $\eta_1 = \hat{n}_x - \epsilon$  and  $\eta_2 = \hat{n}_x + \epsilon$ . We can find  $\epsilon_1, \epsilon_2 > 0$  such that  $[\eta_1(1 - \epsilon_1), \eta_1(1 + \epsilon_1)] \subset [\hat{n}_x - 2\epsilon, \hat{n}_x]$ ,  $[\eta_2(1 - \epsilon_2), \eta_2(1 + \epsilon_2)] \subset [\hat{n}_x, \hat{n}_x + 2\epsilon]$  and on the set  $[K\eta_1(1 - \epsilon_1), K\eta_1(1 + \epsilon_1)] \times \{0\}$  the supremum of  $d_1$  is strictly below  $b(x)$  while on the  $[K\eta_2(1 - \epsilon_2), K\eta_2(1 + \epsilon_2)] \times \{0\}$  the infimum of  $d_1$  is strictly above  $b(x)$ . Let  $\mathcal{S}'_K = [K(\hat{n}_x - 2\epsilon), K(\hat{n}_x + 2\epsilon)] \times \{0\}$  and let  $T_{\mathcal{S}'_K}$  be given by (4.32). From part (B) of Proposition 4.5 we get

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( T_{\mathcal{S}'_K} \leq \frac{T}{Ku_K}, N_1^K (T_{\mathcal{S}'_K}) \notin [K(\hat{n}_x - 2\epsilon), K(\hat{n}_x + 2\epsilon)] \right) = 0.$$

Observe that  $\text{supp}(Y^K(t)) = \{x\}$  for all  $t \geq 0$ . Hence this limit proves part (B).

For part (C) note that  $\text{Fit}(y, x) > 0$ . We can choose  $\epsilon_0 > 0$  such that  $d(y) + \alpha(y, x)(\hat{n}_x + 2\epsilon_0) + 2\alpha(y, y)\epsilon_0 = b(y)$ . Now let  $\epsilon < \epsilon_0$  and assume that  $z_1^K < K(\hat{n}_x + \epsilon)$  and  $z_2^K > K\epsilon$  for all  $K \in \mathbb{N}^*$ , as stated in the proposition. Define the set  $\mathcal{S}'_K = [0, K(\hat{n}_x + 2\epsilon)] \times [K\epsilon/2, \infty)$  and let  $T_{\mathcal{S}'_K}$  be given by (4.32). It is easy to see that

$$\begin{aligned} & \lim_{K \rightarrow \infty} \mathbb{P} \left( Y_t^K \{y\} < \frac{\epsilon}{2} \text{ for some } t \in \left[ 0, \frac{T}{Ku_K} \right] \right) \\ & \leq \lim_{K \rightarrow \infty} \left[ \mathbb{P} \left( T_{\mathcal{S}'_K} \leq \frac{T}{Ku_K}, N_1^K (T_{\mathcal{S}'_K}) \geq K(\hat{n}_x + 2\epsilon) \right) + \mathbb{P} \left( T_{\mathcal{S}'_K} \leq \frac{T}{Ku_K}, N_2^K (T_{\mathcal{S}'_K}) < \frac{K\epsilon}{2} \right) \right]. \end{aligned}$$

On the set  $[K(\hat{n}_x + \epsilon/2), K(\hat{n}_x + 3\epsilon/2)] \times [K\epsilon/2, \infty)$ , the infimum of  $d_1$  is greater than  $b(x)$ . Part (A) of Proposition 4.5 shows that the first limit on the right is 0. On the set  $[0, K(\hat{n}_x + 2\epsilon)] \times [K\epsilon/2, 3K\epsilon/4)$ , the supremum of  $d_2$  is less than  $b(y)$ . We can use part(B) of Proposition 4.5, to see that the second limit on the right is also 0. This proves part (C).

For part (D), observe that  $\text{Fit}(y, x) < 0$  and let  $\epsilon_0 > 0$  satisfy  $d(y) + \alpha(y, x)(\hat{n}_x - 2\epsilon_0) = b(y)$ . Pick an  $\epsilon \in (0, \epsilon_0)$ . On the set  $[K(\hat{n}_x - 2\epsilon), K(\hat{n}_x + 2\epsilon)] \times [1, 2K\epsilon)$  the infimum of  $d_2$  is greater than  $b(y)$ . Part (D) of Proposition 4.5 proves part (D).

For part (E), note that  $\text{Fit}(y, x) > 0$  and let  $\epsilon_0 > 0$  satisfy  $d(y) + \alpha(y, x)(\hat{n}_x + 2\epsilon_0) + 2\alpha(y, y)\epsilon_0 = b(y)$ . Pick  $\epsilon \in (0, \epsilon_0)$ . On the set  $[K(\hat{n}_x - 2\epsilon), K(\hat{n}_x + 2\epsilon)] \times [1, 2K\epsilon)$  the supremum of  $d_2$  is less than  $d_2^+(\epsilon) := d(y) + \alpha(y, x)(\hat{n}_x + 2\epsilon) + 2\alpha(y, y)\epsilon$  and the infimum of  $d_2$  is greater than  $d_2^-(\epsilon) := d(y) + \alpha(y, x)(\hat{n}_x - 2\epsilon)$ . Both  $d_2^+(\epsilon)$  and  $d_2^-(\epsilon)$  are less than  $b(y)$ . Using part (E) of Proposition 4.5 proves part (E).  $\square$

From the preceding proposition, we can retrieve the state of the process on a large window  $[\epsilon/Ku_K, \epsilon^{-1}/Ku_K]$  given the initial condition. This allows us to understand what will happen if we neglect the transitions and large time rare events.

**Corollary 4.7** *Let us consider the process  $Y^K$  of Definition 4.1.*

(A) Suppose that for some  $x \in \mathbb{X}$ ,  $\text{supp}(Y^K(0)) = \{x\}$  and  $Y_0^K\{x\} > \epsilon$  for all  $K \in \mathbb{N}^*$ . Then

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) = 1. \quad (4.39)$$

(B) Suppose that for some  $x, y \in \mathbb{X}$  such that  $\text{Fit}(y, x) < 0$ , we have  $\text{supp}(Y^K(0)) = \{x, y\}$  with  $Y_0^K\{x\} \in [\hat{n}_x - \epsilon, \hat{n}_x + \epsilon]$  and  $Y_0^K\{y\} < \epsilon$  for all  $K \in \mathbb{N}^*$ . Then

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) = 1.$$

(C) Suppose that for some  $x, y \in \mathbb{X}$  such that  $\text{Fit}(y, x) > 0$ , we have  $\text{supp}(Y^K(0)) = \{x, y\}$  with  $Y_0^K\{x\} \in [\hat{n}_x - \epsilon, \hat{n}_x + \epsilon]$  and  $Y_0^K\{y\} = 1/K$  for all  $K \in \mathbb{N}^*$ . Then

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) \\ &= 1 - \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(y) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) \\ &= 1 - \frac{\text{Fit}(y, x)}{b(y)}. \end{aligned}$$

**Proof.** Let us first consider part (A). Let  $\epsilon > 0$  and  $Y^K$  be the process of Definition 4.1, with the initial condition stated in the statement of part (A). Proposition 4.6 (A) implies that for some  $\delta > 0$

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( Y_t^K\{x\} < \delta \text{ for some } t \in \left[ 0, \frac{\epsilon^{-1}}{Ku_K} \right] \right) = 0.$$

From Corollary 4.4 we know that for any  $t \geq 0$  and  $F_f \in \mathbb{F}_b^2$ ,

$$\int_0^t \mathbb{F}_f \left( Y^K \left( \frac{s}{Ku_K} \right) \right) ds \Rightarrow tF_f(\hat{n}_x \delta_x)$$

as  $K \rightarrow +\infty$ . Hence if we define  $\sigma_\epsilon^K = \inf\{t \geq 0 : Y^K(t) \in \mathcal{N}_{\epsilon/2}(x)\}$ , then  $Ku_K \sigma_\epsilon^K \rightarrow 0$  in probability as  $K \rightarrow \infty$ . Now let the process  $\{\tilde{Y}^K(t) : t \geq 0\}$  be given by  $\tilde{Y}_\epsilon^K(t) = Y_\epsilon^K(t + \sigma_\epsilon^K)$ . By the strong Markov property, this process also has generator  $L_0^K$ . Moreover its initial state is inside  $\mathcal{N}_{\epsilon/2}(x)$ . Using part (B) of Proposition 4.6 proves part (A).

For part (B), fix an  $\epsilon > 0$  and consider the process  $Y^K$  with the initial condition specified in the statement. Let us consider the stopping time  $T_{\mathcal{S}_K}$  associated with  $\mathcal{S}_K = [K(\hat{n}_x - 2\epsilon), K(\hat{n}_x + 2\epsilon)] \times [K\epsilon, 2K\epsilon]$  by (4.32) and a sequence  $\{t_K\}$  as in Proposition 4.6. Since  $\text{Fit}(y, x) < 0$ , thanks to Proposition 4.6 (D):

$$\begin{aligned} & \lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) \\ &= \lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] ; T_{\mathcal{S}_K} \leq t_K ; Y_{T_{\mathcal{S}_K}}^K\{y\} = 0 \right). \quad (4.40) \end{aligned}$$



Let  $\{\tilde{Y}^K(t) : t \geq 0\}$  be the process given by  $\tilde{Y}^K(t) = Y^K(T_{\mathcal{S}_K} + t)$ . By the strong Markov property, this process also has generator  $L_0^K$ . On the event  $\{T_{\mathcal{S}_K} \leq t_K ; Y_{T_{\mathcal{S}_K}}^K\{y\} = 0\}$ ,  $\tilde{Y}^K(0)$  is such that  $\text{supp}(\tilde{Y}^K(0)) = \{x\}$  and  $\tilde{Y}_0^K\{x\} \in [\hat{n}_x - 2\epsilon, \hat{n}_x + 2\epsilon]$ . Applying Part (A) of the corollary for  $\tilde{Y}^K$  provides:

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \tilde{Y}^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] ; T_{\mathcal{S}_K} \leq t_K ; Y_{T_{\mathcal{S}_K}}^K\{y\} = 0 \right) = 1.$$

Moreover, since  $Ku_K t_K \rightarrow 0$ , we obtain (4.39) when  $T_{\mathcal{S}_K} < t_K$ , which proves part (B).

For part (C), fix an  $\epsilon > 0$  and define  $Y^K$  with the initial condition specified in the statement. Let us consider  $\mathcal{S}_K$  and  $T_{\mathcal{S}_K}$  as in part (B). We can write

$$\begin{aligned} & \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) \\ &= \sum_{i=1}^3 \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] ; E_i^K(\epsilon) \right), \end{aligned} \quad (4.41)$$

where  $E_1^K(\epsilon) = \{T_{\mathcal{S}_K} \leq t_K, Y_{T_{\mathcal{S}_K}}^K\{y\} = 0\}$ ,  $E_2^K(\epsilon) = \{T_{\mathcal{S}_K} \leq t_K, Y_{T_{\mathcal{S}_K}}^K\{y\} \geq 2\epsilon\}$  and  $E_3^K(\epsilon) = (E_1^K(\epsilon) \cup E_2^K(\epsilon))^c$ .

Let us consider the term in (4.41) corresponding to  $i = 1$ . Since  $t_K < \epsilon/Ku_K$  for sufficiently large  $K$  we have

$$\begin{aligned} & \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] ; E_1^K(\epsilon) \right) \\ &= \mathbb{E} \left( \mathbb{1}_{\{T_{\mathcal{S}_K} \leq t_K; Y_{T_{\mathcal{S}_K}}^K\{y\} = 0\}} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \middle| \mathcal{F}_{T_{\mathcal{S}_K}} \right) \right) \end{aligned} \quad (4.42)$$

On the event  $E_1^K(\epsilon)$ , we also have  $Y_{T_{\mathcal{S}_K}}^K\{x\} \in [\hat{n}_x - 2\epsilon, \hat{n}_x + 2\epsilon]$ . Thus, applying the strong Markov property, part (A) and the fact that  $Ku_K t_K \rightarrow 0$ , the probability inside the expectation in the r.h.s. of (4.42) converges to 1. From part (E) of Proposition 4.6, the term in (4.41) corresponding to  $i = 1$  converges to  $1 - \text{Fit}(y, x)/b(y)$  as  $K \rightarrow \infty$  and  $\epsilon \rightarrow 0$ .

For the term corresponding to  $i = 2$ , we condition in a way similar to (4.42). On the event  $E_2^K(\epsilon)$ ,  $Y_{T_{\mathcal{S}_K}}^K\{y\} \geq 2\epsilon$  and  $Y_{T_{\mathcal{S}_K}}^K\{x\} \in [\hat{n}_x - 2\epsilon, \hat{n}_x + 2\epsilon]$ . From part (C) of Proposition 4.6, the probability of the process  $Y^K\{y\}$  going below  $\epsilon$  between  $T_{\mathcal{S}_K}$  and  $T_{\mathcal{S}_K} + \epsilon^{-1}/Ku_K$  tends to 0 when  $K \rightarrow +\infty$ . Hence, the condition (4.28) of Corollary 4.4 is satisfied and as a consequence, the stopping time:

$$\sigma_{K,\epsilon} = \inf \{t \geq T_{\mathcal{S}_K} : \tilde{Y}_t^K\{x\} < \epsilon \text{ and } \tilde{Y}_t^K\{y\} \in [\hat{n}_y - \epsilon, \hat{n}_y + \epsilon]\}$$

satisfies  $Ku_K \sigma_{K,\epsilon} \rightarrow 0$  in probability. Conditioning by  $\mathcal{F}_{\sigma_{K,\epsilon}}$ , and using part (B), we show that the term corresponding to  $i = 2$  in (4.41) converges to 0.

The term for  $i = 3$  converges to 0 since  $\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_3^K(\epsilon)) = 0$ .

Gathering the results for  $i \in \{1, 2, 3\}$ , we easily get

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) = 1 - \frac{\text{Fit}(y, x)}{b(y)}.$$

The proof that  $\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P} \left( Y^K(t) \in \mathcal{N}_\epsilon(y) \text{ for all } t \in \left[ \frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) = \text{Fit}(y, x)/b(y)$  is similar. This completes the proof of part (C) of the corollary.  $\square$

## 4.2 Proof of Theorem 2.4: convergence to the TSS

We have now the tools to prove Theorem 2.4. By Theorem 3.1, the distributions of  $\{(\chi^K, \Gamma^K) : K \in \mathbb{N}^*\}$  are tight. Let  $(\chi, \Gamma)$  be a limiting value satisfying (2.12) and (2.14). Using Prohorov's theorem, there exists a subsequence  $\{(\tilde{\chi}^K, \tilde{\Gamma}^K)\}$  that converges in distribution to  $(\chi, \Gamma)$ . By Skorokhod representation theorem (see e.g. [1]), there exists on the same probability space as  $(\chi, \Gamma)$  a sequence again denoted by  $\{(\chi^K, \Gamma^K)\}$  with an abuse of notation, that converges a.s. to  $(\chi, \Gamma)$  and which has the same marginal distributions as  $\{(\tilde{\chi}^K, \tilde{\Gamma}^K)\}$ . Observe that the process  $\{\chi'(t) : t \geq 0\}$  (introduced in the statement of Theorem 2.4) uniquely determines  $\Gamma$  (through (2.16)) and  $\Gamma$  uniquely determines the process  $\{\chi(t) : t \geq 0\}$  (through the martingale problem given by (2.12)). It thus remains to prove part (B) of Theorem 2.4.

The main idea to identify the limiting values is that between subsequent appearances of new mutants, our process  $X^K$  behaves like the process considered in Corollary 4.7. When a fit mutant appears, it either gets eradicated quickly or the process stabilizes around the new monomorphic equilibrium characterized by the mutant trait. Between two rare mutations, the trait and size of the monomorphic equilibrium can be inferred from the occupation measure, because the population is monomorphic and because its size is shown to reach an equilibrium.

For  $K \in \mathbb{N}^*$ ,  $i \in \mathbb{N}^*$  let  $\tau_i^K$  and  $\tau_i$  be the  $i$ -th jump times of the process  $\chi^K$  and  $\chi$  respectively. For convenience we define  $\tau_0^K = \tau_0 = 0$ . Since  $(\chi^K, \Gamma^K) \rightarrow (\chi, \Gamma)$  a.s., then for any  $i \in \mathbb{N}^*$ ,  $(\tau_{i-1}^K, \tau_i^K) \rightarrow (\tau_{i-1}, \tau_i)$  a.s. Using (2.12) and Lemma 2.1, we have that almost surely  $\tau_i - \tau_{i-1} \in (0, +\infty)$  for each  $i \in \mathbb{N}^*$ . Thus

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(\tau_i^K - \tau_{i-1}^K \in [\epsilon, \epsilon^{-1}]) = 1. \quad (4.43)$$

In the rest of the proof we will assume that we are always on the set  $\{\forall i \in \mathbb{N}^*, \tau_i^K - \tau_{i-1}^K \geq \epsilon\}$ .

For  $\epsilon > 0$  and  $t > 0$  define:

$$R_i^{K,\epsilon}(t) = \frac{\mathbb{1}_{\{\tau_i^K - \tau_{i-1}^K \geq \epsilon\}} \int_{(\tau_{i-1}^K + \epsilon) \wedge t}^{\tau_i^K \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} \langle \mu, x \rangle \Gamma^K(ds \times d\mu)}{\int_{(\tau_{i-1}^K + \epsilon) \wedge t}^{\tau_i^K \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} \langle \mu, 1 \rangle \Gamma^K(ds \times d\mu)}$$

$$R_i(t) = \frac{\int_{\tau_{i-1} \wedge t}^{\tau_i \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} \langle \mu, x \rangle \Gamma(dt \times d\mu)}{\int_{\tau_{i-1} \wedge t}^{\tau_i \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} \langle \mu, 1 \rangle \Gamma(dt \times d\mu)}.$$

Note that if  $\text{supp}(Z^K(s)) = \{x\}$  for all  $s \in [\tau_{i-1}^K \wedge t + \epsilon, \tau_i^K \wedge t)$ , then  $R_i^{K,\epsilon}(t) = x$ , and the same holds for  $Z$  and  $R_i$ . Heuristically,  $R_i^{K,\epsilon}$  and  $R_i$  are the estimators of the trait of the monomorphic population that fixes between the  $(i-1)$ th and the  $i$ th mutations.

Let us define:

$$\chi'^{K,\epsilon}(t) = x_0 \mathbb{1}_{t < \epsilon} + \sum_{i=1}^{+\infty} R_i^{K,\epsilon}(t) \mathbb{1}_{[\tau_{i-1}^K + \epsilon, \tau_i^K + \epsilon)}(t) \quad \text{and} \quad \chi'(t) = \sum_{i=1}^{+\infty} R_i(t) \mathbb{1}_{[\tau_{i-1}, \tau_i)}(t).$$

The a.s. convergence  $(\chi^K, \Gamma^K) \rightarrow (\chi, \Gamma)$  also implies that  $\chi'^{K,\epsilon}$  converges to  $\chi'$  as  $K \rightarrow +\infty$  and  $\epsilon \rightarrow 0$  in the Skorokhod space  $\mathbb{D}([0, T], \mathbb{X})$  for any  $T > 0$ .

For any  $i \in \mathbb{N}^*$  and  $\epsilon > 0$  define an event

$$E_i^K(\epsilon) = \left\{ Z^K(t) \in \mathcal{N}_\epsilon(R_i^{K,\epsilon}(t)) \text{ for all } t \in [\tau_{i-1}^K + \epsilon, \tau_i^K) \right\},$$

where for any  $x \in \mathbb{X}$ ,  $\mathcal{N}_\epsilon(x)$  is defined by (4.30). On  $E_i^K(\epsilon)$ ,  $R_i^{K,\epsilon}(t) = R_i^{K,\epsilon}(\tau_{i-1}^K + \epsilon)$  for all  $t \in [\tau_{i-1}^K + \epsilon, \tau_i^K)$  and with an abuse of notation, we will write  $R_i^{K,\epsilon}$  for this value.

Now, thanks to the convergence of the  $\tau_i^K$ 's and Lemma 2.1, the proof of part (B) of Theorem 2.4 is done if we prove the following proposition:

**Proposition 4.8** *Under Assumptions 1.1 and 2.2, we have for all  $i \in \mathbb{N}^*$  that:*

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_i^K(\epsilon)) = 1. \quad (4.44)$$

Indeed, if Proposition 4.8 is proved, then  $\Gamma$  is identified. To see this, let  $F_f \in \mathbb{F}_b^2$ ,  $i \in \mathbb{N}^*$  and  $t \geq 0$ , and define two real-valued variables

$$\rho^K = \int_{\tau_{i-1}^K \wedge t}^{\tau_i^K \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} F_f(\mu) \Gamma^K(ds \times d\mu) \text{ and } \rho = \int_{\tau_{i-1} \wedge t}^{\tau_i \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} F_f(\mu) \Gamma(ds \times d\mu).$$

Then certainly  $\rho^K \Rightarrow \rho$  as  $K \rightarrow \infty$ . Take any open set  $B \in \mathbb{R}$ . Then

$$\begin{aligned} \lim_{K \rightarrow \infty} \mathbb{P}(\rho^K \in B) &= \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(\rho^K \in B, E_i^K(\epsilon)) \\ &= \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}\left(\int_{\tau_{i-1}^K \wedge t}^{\tau_i^K \wedge t} F_f\left(\widehat{n}_{R_i^{K,\epsilon}} \delta_{R_i^{K,\epsilon}}\right) ds \in B, E_i^K(\epsilon)\right) \\ &= \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}\left((\tau_i^K \wedge t - \tau_{i-1}^K \wedge t) F_f\left(\widehat{n}_{R_i^{K,\epsilon}} \delta_{R_i^{K,\epsilon}}\right) \in B\right) \\ &= \mathbb{P}((\tau_i \wedge t - \tau_{i-1} \wedge t) F_f(\widehat{n}_{R_i} \delta_{R_i}) \in B). \end{aligned}$$

This proves (2.16).

**Proof.**[Proof of Proposition 4.8] For each  $i \in \mathbb{N}^*$ , we can construct a process  $\{Y_i^K(t) : t \geq 0\}$  with generator  $L_0^K$  such that

$$Z^K(\tau_i^K + t) = Y_i^K\left(\frac{t}{Ku_K}\right) \text{ for all } t \in [0, \tau_{i+1}^K - \tau_i^K). \quad (4.45)$$

For  $i = 1$ , we obtain from Assumption 2.2 and part (A) of Corollary 4.7 that

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(Y_0^K(t) \in \mathcal{N}_\epsilon(x_0) \text{ for all } t \in \left[\frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K}\right]\right) = 1, \quad (4.46)$$

from which we have (4.44) for  $i = 1$  thanks to (4.45) and (4.43).

We now proceed by induction. Let us assume that for  $i \geq 1$ , we have (4.44). Let  $U_i^K$  denote the type of the new mutant that appears at time  $\tau_i^K$ . Pick  $x, y \in \mathbb{X}$ . On the event  $\{E_i^K(\epsilon), R_i^{K,\epsilon} = x, U_i^K = y\}$ , we have  $\text{supp}(Z^K(\tau_i^K)) = \{x, y\}$ ,  $Z_{\tau_i^K}^K\{y\} = 1/K$  and  $Z_{\tau_i^K}^K\{x\} \in [\widehat{n}_x - \epsilon, \widehat{n}_x + \epsilon]$ . Using parts (B) and (C) of Corollary 4.7 and (4.45), we obtain that

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}\left(E_{i+1}^K(\epsilon), R_{i+1}^{K,\epsilon} = x \mid E_i^K(\epsilon), R_i^{K,\epsilon} = x, U_i^K = y\right) &= \left(1 - \frac{[\text{Fit}(y, x)]^+}{b(y)}\right) \\ \text{and } \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}\left(E_{i+1}^K(\epsilon), R_{i+1}^{K,\epsilon} = y \mid E_i^K(\epsilon), R_i^{K,\epsilon} = x, U_i^K = y\right) &= \frac{[\text{Fit}(y, x)]^+}{b(y)}. \end{aligned}$$

Since the distribution of  $U_i^K$  conditionally to  $\{E_i^K(\epsilon), R_i^{K,\epsilon}(\tau_i^K) = x\}$  is  $m(x, dy)$ ,

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P} \left( E_{i+1}^K(\epsilon), R_{i+1}^{K,\epsilon} \in A \mid E_i^K(\epsilon), R_i^{K,\epsilon} = x \right) \\ &= \mathbb{1}_A(x) \int_{\mathbb{X}} \left( 1 - \frac{[\text{Fit}(y, x)]^+}{b(y)} \right) m(x, dy) + \int_A \frac{[\text{Fit}(y, x)]^+}{b(y)} m(x, dy). \end{aligned} \quad (4.47)$$

This along with (4.46) for  $i = 1$  implies that for any  $i \in \mathbb{N}^*$  we have  $\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P} (E_i^K(\epsilon)) = 1$ . This concludes the proof of Proposition 4.8 and the proof of Theorem 2.4.  $\square$

## A Appendix.

We start with recalling some useful estimates for birth and death process (Lemma A.2). Coupling facts between the components of a 2D-birth and death processes on the quadrant (see (4.31)) and 1D-birth and death processes are then given in Proposition A.4. Both results are then used to obtain estimates for the 2D-birth and death processes on the quadrant.

**Definition A.1** For any  $b, d \in \mathbb{R}_+$  and  $n \in \mathbb{N}$ , let  $\mathbf{P}(b, d, n)$  denote the law of the  $\mathbb{N}$ -valued continuous time branching process starting at  $n$  with birth rate  $b$  and death rate  $d$ . For convenience, we will consider  $\mathbf{P}(b, \infty, n)$  to be the law of the process that is 0 at all times.

**Lemma A.2** For each positive integer  $K$  let  $\{B^K(t) : t \geq 0\}$  be a continuous time branching process with law  $\mathbf{P}(b, d, K)$  with  $b \neq d$ . Pick an  $\epsilon \in [0, 1]$  and define a stopping time

$$\sigma_K = \inf \{t \geq 0 : B^K(t) \leq K(1 - \epsilon) \text{ or } B^K(t) > K(1 + \epsilon)\}.$$

Then we have the following.

(A) If  $b < d$  then  $\mathbb{P}(B^K(\sigma_K) > K(1 + \epsilon)) \leq \exp(-K\epsilon \log(d/b))$   
and if  $b > d$  then  $\mathbb{P}(B^K(\sigma_K) \leq K(1 - \epsilon)) \leq \exp(-K\epsilon \log(b/d))$ .

(B) If  $bd = 0$  and  $\epsilon \notin \{0, 1\}$  then  $\inf_{K \geq 1} \mathbb{E}(\sigma_K) > 0$  and  $\sup_{K \geq 1} \mathbb{E}(\sigma_K^2) < \infty$ .  
In parts (C) and (D) let  $t_K$  be any  $\mathbb{N}^*$ -valued sequence satisfying  $\log K \ll t_K$

(C) If  $\epsilon = 1$  and if  $b < d$  then  $\lim_{K \rightarrow \infty} \mathbb{P}(\sigma_K \leq t_K, B^K(\sigma_K) = 0) = 1$ .

(D) Let  $\{Y(t) : t \geq 0\}$  be a branching process with law  $\mathbf{P}(b, d, 1)$ , where  $b > d$ . For any  $\epsilon > 0$  define

$$\gamma_K = \inf \{t \geq 0 : Y(t) = 0 \text{ or } Y(t) \geq K\epsilon\}.$$

Then

$$1 - \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K \leq t_K, Y(\gamma_K) \geq K\epsilon) = \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K \leq t_K, Y(\gamma_K) = 0) = \frac{d}{b}.$$

**Proof.** Note that  $\sigma_K < \infty$  a.s. since a branching process either goes to 0 or to  $\infty$  almost surely. We can easily check that  $M_K(t) = (d/b)^{B^K(t \wedge \sigma_K)}$  is a bounded martingale. Then by the optional sampling theorem we get

$$\mathbb{E}(M_K(\sigma_K)) = \mathbb{E}(M_K(0)) = \left(\frac{d}{b}\right)^K. \quad (\text{A.1})$$

If  $b < d$  then  $\mathbb{E}(M_K(\sigma_K)) \geq \mathbb{P}(B^K(\sigma_K) > K(1 + \epsilon)) (d/n)^{K(1+\epsilon)}$  and if  $b > d$  then  $\mathbb{E}(M_K(\sigma_K)) \geq \mathbb{P}(B^K(\sigma_K) \leq K(1 - \epsilon)) (d/b)^{K(1-\epsilon)}$ . Substituting these estimates in (A.1) proves part (A).

For part (B) we assume that  $d = 0$  and  $b > 0$ . The case where  $b = 0$  and  $d > 0$  is similar. We can also assume that  $K(1 + \epsilon)$  is a positive integer. Since  $d = 0$ , the process  $B^K$  is monotonically increasing and hence  $B^K(\sigma_K) = K(1 + \epsilon)$ . We can show with Dynkin's formula that  $B^K(t) - b \int_0^t B^K(s) ds$  is a martingale, and obtain by the optional sampling theorem that

$$K = \mathbb{E}(B^K(\sigma_K)) - b\mathbb{E}\left(\int_0^{\sigma_K} B^K(s) ds\right) = K(1 + \epsilon) - b\mathbb{E}\left(\int_0^{\sigma_K} B^K(s) ds\right).$$

Since  $B^K(t) \in [K, K(1 + \epsilon)]$  for  $t \leq \sigma_K$  we have  $K\mathbb{E}(\sigma_K) \leq \mathbb{E}(\int_0^{\sigma_K} B^K(s) ds) \leq K(1 + \epsilon)\mathbb{E}(\sigma_K)$ , and hence

$$\frac{\epsilon}{b(1 + \epsilon)} \leq \inf_{K \geq 1} \mathbb{E}(\sigma_K) \leq \sup_{K \geq 1} \mathbb{E}(\sigma_K) \leq \frac{\epsilon}{b}. \quad (\text{A.2})$$

From the fact that  $tB^K(t) - \int_0^t B^K(s)(bs + 1)ds$  is also a martingale, we obtain that

$$K\mathbb{E}\left(\frac{b}{2}\sigma_K^2 + \sigma_K\right) \leq \mathbb{E}\left(\int_0^{\sigma_K} B^K(s)(bs + 1)ds\right) = \mathbb{E}(\sigma_K B^K(\sigma_K)) = K(1 + \epsilon)\mathbb{E}(\sigma_K).$$

This along with (A.2), proves part (B). Parts (C) and (D) follow directly from [2, Th. 4].  $\square$

**Definition A.3** Suppose that  $d$  is a function from  $\mathbb{N} \times \mathbb{N}$  to  $\mathbb{R}_+$ . Then for any set  $\mathcal{S} \subset \mathbb{R}_+ \times \mathbb{R}_+$  we define

$$\inf d(\mathcal{S}) = \inf \{d(m, n) : (m, n) \in \mathcal{S} \cap (\mathbb{N} \times \mathbb{N})\}$$

$$\text{and } \sup d(\mathcal{S}) = \sup \{d(m, n) : (m, n) \in \mathcal{S} \cap (\mathbb{N} \times \mathbb{N})\}.$$

**Proposition A.4** Let  $\{N(t) = (N_1(t), N_2(t))\}$  be a  $\mathbb{N} \times \mathbb{N}$ -valued pure jump Markov process with the following transition rates.

$$\begin{array}{ll} mb_1 & \text{from } (m, n) \text{ to } (m + 1, n) \\ nb_2 & \text{from } (m, n) \text{ to } (m, n + 1) \\ md_1(m, n) & \text{from } (m, n) \text{ to } (m - 1, n) \\ nd_2(m, n) & \text{from } (m, n) \text{ to } (m, n - 1). \end{array}$$

Here  $b_1, b_2$  are positive constants and  $d_1, d_2$  are functions from  $\mathbb{N} \times \mathbb{N}$  to  $\mathbb{R}_+$ . Suppose that there is a set  $\mathcal{S} \subset \mathbb{R}_+ \times \mathbb{R}_+$  and constants  $d_1^+, d_1^-, d_2^+, d_2^- \in [0, \infty]$  such that

$$d_1^- \leq \inf d_1(\mathcal{S}) \leq \sup d_1(\mathcal{S}) \leq d_1^+ \text{ and}$$

$$d_2^- \leq \inf d_2(\mathcal{S}) \leq \sup d_2(\mathcal{S}) \leq d_2^+$$

Assume that  $(N_1(0), N_2(0)) \in \mathcal{S}$  and let  $T_{\mathcal{S}}$  be the random time defined by

$$T_{\mathcal{S}} = \inf\{t \geq 0 : N(t) \notin \mathcal{S}\}.$$

Let  $z_1^+, z_1^-, z_2^+, z_2^-$  be positive integers satisfying  $z_1^- \leq N_1(0) \leq z_1^+$  and  $z_2^- \leq N_2(0) \leq z_2^+$ . Then on the same probability space as  $N$ , we can construct four  $\mathbb{N}$ -valued processes  $B_1^+, B_1^-, B_2^+$  and  $B_2^-$  with laws  $\mathbf{P}(b_1, d_1^+, z_1^+)$ ,  $\mathbf{P}(b_1, d_1^-, z_1^-)$ ,  $\mathbf{P}(b_2, d_2^+, z_2^+)$ ,  $\mathbf{P}(b_2, d_2^-, z_2^-)$  such for all  $t \leq T_{\mathcal{S}}$  the following relations are satisfied almost surely,

$$B_1^-(t) \leq N_1(t) \leq B_1^+(t) \quad \text{and} \quad B_2^-(t) \leq N_2(t) \leq B_2^+(t).$$

**Proof.** The proof follows from direct coupling of these processes.  $\square$

**Proof.** [Proof of Proposition 4.5] We prove part (A). Without loss of generality, we can assume that  $K\eta, K\eta(1 - \epsilon), K\eta(1 + \epsilon)$  are positive integers and  $N_1^K(0) = K\eta$ . Define

$$\sigma_K = \inf \{t \geq 0 : N^K(t) \notin [K\eta(1 - \epsilon), K\eta(1 + \epsilon)] \times [C_K, D_K]\}.$$

In order for the process  $N_1^K$ , started at  $K\eta$ , to go beyond level  $B_K$  it must exit the interval  $[K\eta(1 - \epsilon), K\eta(1 + \epsilon)]$  from above. The probability to go from  $K\eta$  to  $K\eta(1 + \epsilon)$  is exponentially small. Indeed, since  $b < d_1^-(\eta, \epsilon)$ , by Proposition A.4 we can construct a coupled subcritical branching process  $B_+^K$  with law  $\mathbf{P}(b_1, d_1^-(\eta, \epsilon), K)$  such that  $B_+^K(t) \geq N_1^K(t)$  for all  $t \leq \sigma_K$  almost surely. Hence if  $\gamma_K$  is the first time the process  $B_+^K$  leaves the set  $[K\eta(1 - \epsilon), K\eta(1 + \epsilon)]$  then from part (A) of Lemma A.2 we obtain that for some  $c > 0$

$$\mathbb{P}(N_1^K(\sigma_K) > K\eta(1 + \epsilon)) \leq \mathbb{P}(B_+^K(\gamma_K) > K\eta(1 + \epsilon)) \leq e^{-cK}. \quad (\text{A.3})$$

Before  $N_1^K$  exists  $[A_K, B_K)$  from above, this process crosses the interval  $[K\eta(1 - \epsilon), K\eta]$  several times. Let  $\rho_K$  be the number (possibly 0) of these passages in the interval  $[0, \mathcal{T}_{S_K}]$ . Because of (A.3), for any  $n \in \mathbb{N}$ ,

$$\mathbb{P}(\rho_K < n) \leq ne^{-cK}. \quad (\text{A.4})$$

Let  $n_K = \lceil 1/Ku_K \rceil^2$ . Then,

$$\begin{aligned} & \mathbb{P}\left(\mathcal{T}_{S_K} \leq \frac{T}{Ku_K}, N_1^K(\mathcal{T}_{S_K}) \geq B_K\right) \\ & \leq \mathbb{P}\left(\mathcal{T}_{S_K} \leq \frac{T}{Ku_K}, N_1^K(\mathcal{T}_{S_K}) \geq K\eta(1 + \epsilon), \rho_K \geq n_K\right) + \mathbb{P}(\rho_K < n_K) \\ & \leq \mathbb{P}\left(\sum_{i=1}^{n_K} \tau_{K,i} < \frac{T}{Ku_K}\right) + \mathbb{P}(\rho_K < n_K), \end{aligned} \quad (\text{A.5})$$

where the  $\tau_{K,i}$  denotes the durations of the  $n_K$ 's first passages of  $N_1^K$  from  $K\eta$  to  $K\eta(1 + \epsilon)$ . The last term in the r.h.s. of (A.5) tends to 0 with  $K$  due to our choice of  $n_K$ , (A.4) and (2.3). Using the couplings of Proposition A.4, it is possible to dominate  $N_1^K$  by a pure birth process with distribution  $\mathbf{P}(b_1, 0, K\eta(1 - \epsilon))$  on each of its excursions of  $N_1^K$  from  $K\eta(1 - \epsilon)$  to  $K\eta$ . Thus, each  $\tau_{K,i}$  can be bounded above by the time  $\bar{\sigma}_{K,i}$  needed by a pure birth process with distribution  $\mathbf{P}(b_1, 0, K\eta(1 - \epsilon))$  to reach  $K\eta$ . The  $\bar{\sigma}_{K,i}$ 's can be chosen to be i.i.d. and by Lemma A.2, Part (B),  $\inf_{K \geq 1} \mathbb{E}(\tau_{K,1}) > 0$  and  $\sup_{K \geq 1} \mathbb{E}(\tau_{K,1}^2) < \infty$ . Applying Chebychev's inequality we get

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n_K} \tau_{K,i} \leq \frac{T}{Ku_K}\right) & \leq \mathbb{P}\left(\frac{1}{n_K} \sum_{i=1}^{n_K} \bar{\sigma}_{K,i} - \mathbb{E}(\bar{\sigma}_{K,i}) \leq \frac{T}{Ku_K n_K} - \mathbb{E}(\bar{\sigma}_{K,1})\right) \\ & \leq \frac{\mathbb{E}(\bar{\sigma}_{K,1}^2)}{n_K \left(\mathbb{E}(\bar{\sigma}_{K,1}) - \frac{T}{Ku_K n_K}\right)^+}. \end{aligned}$$

As  $K \rightarrow \infty$ ,  $T/(Ku_K n_K) \rightarrow 0$  and  $n_K \rightarrow \infty$ . This proves part (A). Proof of part (B) is similar and part (C) is a direct consequence of parts (A) and (B).

We now prove part (D). Using Proposition A.4 and since  $b_2 < d_2^-(\epsilon)$  and  $N_2^K(0) \leq K\epsilon/2$ ,  $N_2^K$  can be dominated on  $[0, T_{\mathcal{S}_K}]$  by a subcritical branching process  $B_+^K$  with distribution  $\mathbf{P}(b_2, d_2^-(\epsilon), [K\epsilon/2])$ . Let

$$\gamma_K^+ = \inf\{t \geq 0 : B_+^K(t) = 0 \text{ or } B_+^K(t) \geq K\epsilon\}. \quad (\text{A.6})$$

The coupling ensures that if  $T_{\mathcal{S}_K} \geq \gamma_K^+$  and  $B_+^K(\gamma_K^+) = 0$  then  $N_2^K(\gamma_K^+) = 0$  and  $T_{\mathcal{S}_K} = \gamma_K^+$ . Since  $B_+^K(t) \in (0, K\epsilon)$  for  $t \in [0, \gamma_K^+)$  we can also see that if  $T_{\mathcal{S}_K} < \gamma_K^+$  then either  $N_2^K(T_{\mathcal{S}_K}) = 0$  or  $N_1^K(T_{\mathcal{S}_K}) \notin [A_K, B_K)$ . Hence

$$\begin{aligned} & \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) = 0) \\ & \leq \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) = 0, T_{\mathcal{S}_K} \geq \gamma_K^+) + \mathbb{P}(\gamma_K^+ \leq t_K, T_{\mathcal{S}_K} < \gamma_K^+) \\ & \leq \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) = 0, T_{\mathcal{S}_K} = \gamma_K^+) + \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) = 0, T_{\mathcal{S}_K} < \gamma_K^+) \\ & \quad + \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_1^K(T_{\mathcal{S}_K}) \notin [A_K, B_K), T_{\mathcal{S}_K} < \gamma_K^+) \\ & \leq \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) = 0) + \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_1^K(T_{\mathcal{S}_K}) \notin [A_K, B_K)). \end{aligned} \quad (\text{A.7})$$

Due to part (C), the limit of the second term in the r.h.s. of (A.7) is 0. Since part (C) of Lemma A.2 tells us that the limit of the l.h.s. is 1, this proves part (D).

We now prove part (E). Since  $b_2 > d_2^+(\epsilon) > d_2^-(\epsilon)$  and  $N_2^K(0) = 1$ , using Proposition A.4, we can construct two coupled supercritical branching processes  $B_-^K, B_+^K$  with distributions  $\mathbf{P}(b_2, d_2^+(\epsilon), 1), \mathbf{P}(b_2, d_2^-(\epsilon), 1)$  such that  $B_-^K(t) \leq N_2^K(t) \leq B_+^K(t)$  for all  $t \leq T_{\mathcal{S}_K}$  almost surely. Let  $\gamma_K^+$  be given by (A.6) and  $\gamma_K^-$  be the corresponding stopping time for  $B_-^K$ . The fact that  $B_-^K$  is below  $N_2^K$  until time  $T_{\mathcal{S}_K}$ , (A.7) and part (C) allow us to prove that:

$$\begin{aligned} \lim_{K \rightarrow +\infty} \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) = 0) & \leq \lim_{K \rightarrow +\infty} \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) = 0) \\ & \leq \lim_{K \rightarrow +\infty} \mathbb{P}(\gamma_K^- \leq t_K, B_-^K(\gamma_K^-) = 0). \end{aligned} \quad (\text{A.8})$$

Since  $B_+^K$  is above  $N_2^K$  until time  $T_{\mathcal{S}_K}$ , we obtain:

$$\lim_{K \rightarrow +\infty} \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) \geq K\epsilon) \leq \lim_{K \rightarrow +\infty} \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) \geq K\epsilon). \quad (\text{A.9})$$

Finally, a proof similar to the one of (A.7), we can prove that:

$$\lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^- \leq t_K, B_-^K(\gamma_K^-) \geq K\epsilon) \leq \lim_{K \rightarrow \infty} \mathbb{P}(T_{\mathcal{S}_K} \leq t_K, N_2^K(T_{\mathcal{S}_K}) \geq K\epsilon). \quad (\text{A.10})$$

From part (D) of Lemma A.2 and Assumption 2.3 we can deduce that

$$\begin{aligned} 1 - \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) \geq K\epsilon) & = \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) = 0) = \frac{d_2^-(\epsilon)}{b_2} \\ \text{and } 1 - \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^- \leq t_K, B_-^K(\gamma_K^-) \geq K\epsilon) & = \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^- \leq t_K, B_-^K(\gamma_K^-) = 0) = \frac{d_2^+(\epsilon)}{b_2}. \end{aligned}$$

These relations along with (A.8), (A.9) and (A.10) prove part (E).  $\square$

**Acknowledgments:** The authors thank Sylvie Méléard for invaluable discussion. This work benefited from the support of the ANR MANEGE (ANR-09-BLAN-0215), from the Chair ‘‘Modélisation Mathématique et Biodiversité’’ of Veolia Environnement-Ecole Polytechnique-Museum National d’Histoire Naturelle-Fondation X.

## References

- [1] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1968. 4.2
- [2] N. Champagnat. A microscopic interpretation for adaptative dynamics trait substitution sequence models. *Stochastic Processes and their Applications*, 116(8): 1127–1160, 2006. 1, 1, 2, 4.1, 4.1.2, A
- [3] N. Champagnat, R. Ferrière, and S. Méléard. Individual-based probabilistic models of adaptative evolution and various scaling approximations. In R.C. Dalang, M. Dozzi and F. Russo, editors, *Proceedings of the 5th seminar on Stochastic Analysis, Random Fields and Applications, Ascona, May 2005*, Volume 59 of Progress in Probability, pages 75–114, Birkhauser, Basel, 2006. 2, 4.1, 4.1.2
- [4] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models via timescale separation. *Theoretical Population Biology*, 69: 297–321, 2006. 2
- [5] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields* 151: 45–94, 2011 1
- [6] P. Collet, S. Méléard, and J.A.J. Metz. A rigorous model study of the adaptative dynamics of Mendelian diploids. 2011. <http://arxiv.org/abs/1111.6234>. 1
- [7] D. A. Dawson. Measure-valued markov processes. In P.-L. Hennequin, editor, *Ecole d’Eté de probabilités de Saint-Flour XXI*, volume 1541 of *Lectures Notes in Math.*, pages 1–260, Springer, New York, 1993. 1
- [8] U. Dieckmann, M. Heino and K. Parvinen The adaptive dynamics of function-valued traits. *Journal of theoretical Biology* 241(2): 370–389, 2006. 1
- [9] U. Dieckmann and R. Law. The dynamical theory of coevolution: A derivation from stochastic ecological processes. *Journal of Mathematical Biology*, 34(5–6): 579–612, 1966. 1
- [10] M. Durinx, J.A.J. Metz, and G. Meszéna. Adaptive dynamics for physiologically structured models. *Journal of Mathematical Biology*, 56(5): 673–742, 2008. 1
- [11] S.N. Ethier and T.G. Kurtz. *Markov Processus, Characterization and Convergence*. John Wiley & Sons, New York, 1986. 3
- [12] S.A.H. Geritz, É. Kisdi, G. Meszéna and J.A.J. Metz Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, 12: 35–57. 1
- [13] A. Gupta, J.A.J. Metz, and V.C. Tran. work in progress. 1
- [14] O. Kallenberg. *Random Measures*. Academic Press, 1983. 1
- [15] T.G. Kurtz. *Approximation of population processes*. SIAM, Philadelphia, PA, 1981. 1
- [16] T.G. Kurtz. Averaging for martingale problems and stochastic approximation. In I. Karatzas and D. Ocone, editors, *Applied stochastic analysis (New Brunswick, NJ, 1991)*, volume 177 of *Lectures Notes in Control and Inform. Sci.*, pages 186–209, Springer, Berlin, 1992. 1, 2, 3
- [17] S. Méléard, J.A.J. Metz, and V.C. Tran. Limiting Feller diffusions for logistic populations with age-structure. In *Proceedings of the 58th World Statistics Congress - ISI2011*. ISI, 2011. Preprint HAL 00595928. 2
- [18] S. Méléard and V.C. Tran. Trait substitution sequence process and canonical equation for age-structured populations. *Journal of Mathematical Biology*, 58(6): 881–921, 2009. 1
- [19] S. Méléard and V.C. Tran. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and their Applications*, 122(1): 250–276, January 2012. 2
- [20] J.A.J. Metz, R.M.Nisbet and S.A.H. Geritz How should we define fitness for general ecological scenarios. *Trends in Ecology and Evolution*, 7(6): 198–202, 1992. 1
- [21] J.A.J. Metz, S.A.H. Geritz, G. Meszéna, F.A.J. Jacobs, and J.S. Van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In S.J. Van Strien and S.M. Verduyn Lunel, editors, *Stochastic and Spatial Structures of Dynamical Systems*, Volume 45 of *KNAW Verhandeligen Afd. Natuurkunde, Eerste reeks*, pages 183–231, North Holland, Amsterdam, 1996. 1, 1
- [22] K. Parvinen, U. Dieckmann and M. Heino Function-valued adaptive dynamics and the calculus of variations. *Journal of Mathematical Biology*, 52(1): 1–26, 2006. 1