

A New Proof for the Convergence of an Individual Based Model to the Trait Substitution Sequence

Ankit Gupta · J.A.J. Metz · Viet Chi Tran

Received: 13 March 2012 / Accepted: 18 September 2013 / Published online: 28 September 2013
© Springer Science+Business Media Dordrecht 2013

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 the process is viewed in the timescale of mutations. Using averaging techniques due to Kurtz (in Lecture Notes in Control and Inform. Sci., vol. 177, pp. 186–209, 1992), we give a new proof of the convergence of the individual based model to the trait substitution sequence of Metz et al. (in Trends in Ecology and Evolution 7(6), 198–202, 1992), first worked out by Dieckman and Law (in Journal of Mathematical Biology 34(5–6), 579–612, 1996) and rigorously proved by Champagnat (in Theoretical Population Biology 69, 297–321, 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

A. Gupta (✉) · V.C. Tran
CMAP, Ecole Polytechnique, UMR CNRS 7641, Route de Saclay, 91128 Palaiseau Cédex, France
e-mail: gupta@cmap.polytechnique.fr

J.A.J. Metz
Mathematical Institute & Institute of Biology & NCB Naturalis, Leiden, Niels Bohrweg 1, 2333 CA,
Leiden, The Netherlands
e-mail: j.a.j.metz@biology.leidenuniv.nl

J.A.J. Metz
EEP, IIASA, Laxenburg, Austria

V.C. Tran
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
e-mail: chi.tran@math.univ-lille1.fr

Mathematics 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 was introduced by Dieckmann and Law [9] and Metz et al. [23] 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. [22] (with Metz et al. [23] as a follow up). Following Dieckmann and Law [9], the TSS 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 from one state to another. Champagnat’s proof is based on some fine estimates, including some fine large deviation results, to combine several approximations of the microscopic process. We separate the different time scales that are involved using averaging techniques due to Kurtz [17], 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 $x \in \mathbb{X}$, where \mathbb{X} is a compact subset of \mathbb{R}^d . We are interested in large populations. We assume that the population’s initial size is proportional to a parameter $K \in \mathbb{N}^* = \{1, 2, \dots\}$, to be interpreted as the area to which the population is confined. We will let K go to infinity while keeping the density constant by counting the 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 area. The population at time t can be described by the following point measure on \mathbb{X}

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

where $N^K(t)$ is the total number of individuals in the population at time t and where $x_i^t \in \mathbb{X}$ 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 positive continuous 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, and has the same 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 α 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:

$$(b(y) - d(y))\alpha(x, x) - (b(x) - d(x))\alpha(y, x) < 0 \tag{1.2}$$

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

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

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

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 consequence of this condition is that when the mutant population manages to reach a sufficiently large size, it wipes out the resident population. Of course the other possibility is that the mutant population becomes extinct quickly. Hence, the population should be monomorphic away from the mutation events.

In this paper, we use the methods of Kurtz [18], based on martingale problems, for separating time scales. We show that on 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. Our proof differs from the one in [2] in that we do not require comparisons with differential equations and large deviation results. Instead we use comparisons with branching processes to exhibit the stabilization of the population around the equilibria determined by the resident trait. Some of our arguments are similar in nature to those presented in [16] for a related model but with finite population sizes.

In Sect. 2, we describe the IBM introduced in [2]. The model accounts only for a trait-structure and otherwise has very simple dynamics. More general trait spaces (possibly functional, as first introduced in [10, 24]) are considered in [14] where the “invasion implies substitution” assumption is also relaxed (see also [3, 13, 23]). There exist several other possible generalisations of the IBMs that underlie the TSS, including for e.g. the physiological structure [11, 19], diploidy [7] or multi-resources [6]. We consider the process that counts the new traits appearing due to mutations, and the occupation measure Γ^K of the process X^K under a changed time scale. The tightness of this couple of processes is studied in Sect. 3. The limiting values are shown to satisfy an equation that is considered in Sect. 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 trait. From the consideration of monomorphic and dimorphic populations and constructing couplings with branching processes, we prove the convergence in distribution of $\{\Gamma^K\}$ to the occupation measure Γ of a pure jump Markov process that is called the 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 for e.g. [15]) and for any $M > 0$, the set $\{\mu \in \mathcal{M}_F(E) : \mu(E) \leq M\}$ is compact. For a measure μ , 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 $a \wedge b$. For any $a \in \mathbb{R}$, its positive part is denoted by $[a]^+$ and $\lfloor a \rfloor$ denotes the largest integer less than 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 $\mathcal{M}_{\mathbb{F}}(\mathbb{X})$ 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 derivatives. This class \mathbb{F}_b^2 is separable and it is known (see for example [8]) that it characterizes the convergence in distribution on $\mathcal{M}_F(\mathbb{X})$.

The class of càdlàg processes from \mathbb{R}_+ to E is denoted by $\mathbb{D}(\mathbb{R}_+, E)$.

The value at time t of a process X is denoted by $X(t)$ 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. For any $F_f \in \mathbb{F}_b^2$ let

$$\begin{aligned} L^K F_f(X) &= K \int_E b(x) \left[\int_{\mathbb{X}} \left(F_f \left(X + \frac{1}{K} \delta_y \right) - F_f(X) \right) M^K(x, dy) \right] X(dx) \\ &\quad + K \int_E (d(x) + \langle X, \alpha(x, \cdot) \rangle) \left(F_f \left(X - \frac{1}{K} \delta_x \right) - F_f(X) \right) X(dx), \end{aligned} \tag{2.1}$$

where M^K is the transition kernel given by

$$M^K(x, dy) = u_K p(x) m(x, dy) + (1 - u_K p(x)) \delta_x(dy). \tag{2.2}$$

Let $K \in \mathbb{N}^*$ be fixed. The martingale problem for L^K has a unique solution for any initial condition $X^K(0) \in \mathcal{M}_F(\mathbb{X})$ and it is possible to construct the solution of the martingale problem by considering a stochastic differential equation (SDE) driven by Poisson point processes which corresponds to the IBM used for simulations (see [4, 5]). The following estimate will be needed in the sequel. It is proved 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 make some assumptions on the initial condition.

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

- (A) There exists a $x_0 \in \mathbb{X}$ 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 (2.1), 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 so we require that $Ku_K \rightarrow 0$ as $K \rightarrow \infty$. Moreover as in [2], we will assume that the speed of arrival of new mutants is *slow* enough to give each mutant enough time to invade the population if it can (time of order $\log K$) and *fast* enough to ensure that a new mutant arrives while the resident population is near its equilibrium (time of order e^{cK} for some $c > 0$). We will therefore make the following assumption.

Assumption 2.3 For any $c > 0$

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

Consider the process

$$Z^K(t) = X^K\left(\frac{t}{Ku_K}\right), \quad t \geq 0. \tag{2.3}$$

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 $F_f \in \mathbb{F}_b^2$

$$\begin{aligned} \mathbb{L}^K F_f(Z) &= \int_{\mathbb{X}} p(x)b(x) \left[\int_{\mathbb{X}} \left(F_f\left(Z + \frac{1}{K}\delta_y\right) - F_f(Z) \right) m(x, dy) \right] Z(dx) \\ &\quad + \frac{1}{Ku_K} \left[\int_{\mathbb{X}} b(x)(1 - u_K p(x)) K \left(F_f\left(Z + \frac{1}{K}\delta_x\right) - F_f(Z) \right) Z(dx) \right. \\ &\quad \left. + \int_{\mathbb{X}} (d(x) + \langle Z, \alpha(x, \cdot) \rangle) K \left(F_f\left(Z - \frac{1}{K}\delta_x\right) - F_f(Z) \right) Z(dx) \right]. \end{aligned} \tag{2.4}$$

In the process Z^K we have compressed time so that the mutations occur at a 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 for e.g. [18]). 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 [18]. We introduce the occupation measure Γ^K defined for any $t \geq 0$ and for any set $A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X}))$ by

$$\Gamma^K([0, t] \times A) = \int_0^t \mathbb{1}_A(Z^K(s)) ds. \tag{2.5}$$

Kurtz’s techniques have been used in the context of measure-valued processes in [20, 21] for different population dynamic problems, but an additional difficulty arises here due to the presence of non-linearities 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 χ^K is a pure-jump Markov process that satisfies the following martingale problem. For any $F_f \in \mathbb{F}_b^2$

$$\begin{aligned}
 M_t^{\chi,K} &:= F_f(\chi^K(t)) - F_f(\chi^K(0)) \\
 &\quad - \int_0^t \int_{\mathbb{X}} p(x)b(x) \int_{\mathbb{X}} (F_f(\chi^K(s) + \delta_y) - F_f(\chi^K(s)))m(x, dy) Z^K(s, dx) ds \\
 &= F_f(\chi^K(t)) - F_f(\chi^K(0)) \\
 &\quad - \int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \left[\int_{\mathbb{X}} p(x)b(x) \int_{\mathbb{X}} (F_f(\chi^K(s) + \delta_y) \right. \\
 &\quad \left. - F_f(\chi^K(s)))m(x, dy)\mu(dx) \right] \Gamma^K(ds \times d\mu),
 \end{aligned}
 \tag{2.6}$$

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 (χ, Γ) , where the slow component χ 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, 2.2 and 2.3 hold.*

(A) *There exists a process χ with paths in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathbb{X}))$ 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 (χ, Γ) satisfy the following. For all functions $F_f \in \mathbb{F}_b^2$,*

$$\begin{aligned}
 &F_f(\chi(t)) - F_f(\chi(0)) \\
 &\quad - \int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \left[\int_{\mathbb{X}} b(x)p(x) \int_{\mathbb{X}} (F_f(\chi(s) + \delta_y) \right. \\
 &\quad \left. - F_f(\chi(s)))m(x, dy)\mu(dx) \right] \Gamma(ds \times d\mu)
 \end{aligned}
 \tag{2.8}$$

is a square integrable martingale with respect to the filtration

$$\mathcal{F}_t = \sigma\{\chi(s), \Gamma([0, s] \times A) : s \in [0, t], A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X}))\}
 \tag{2.9}$$

and for any $t \geq 0$

$$\int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu)\Gamma(ds \times d\mu) = 0 \quad \text{for all } F_f \in \mathbb{F}_b^2 \text{ a.s.,}
 \tag{2.10}$$

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

$$\mathbb{B}F_f(\mu) = F'(\langle \mu, f \rangle) \int_{\mathbb{X}} (b(x) - (d(x) + \langle \mu, \alpha(x, \cdot) \rangle)) f(x)\mu(dx).
 \tag{2.11}$$

(B) *Let Γ be as in part (A). Then for any $t > 0$ and $A \in \mathcal{B}(\mathcal{M}_F(\mathbb{X}))$ we have*

$$\Gamma([0, t] \times A) = \int_0^t \mathbb{1}_A(\widehat{n}_{\chi'(s)}\delta_{\chi'(s)})ds,
 \tag{2.12}$$

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

$$\mathbb{C}f(x) = b(x)p(x)\widehat{n}_x \int_{\mathbb{X}} \frac{[\text{Fit}(y, x)]^+}{b(y)} (f(y) - f(x))m(x, dy), \tag{2.13}$$

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

$$\widehat{n}_x = (b(x) - d(x))/\alpha(x, x) \quad \text{and} \quad \text{Fit}(y, x) = b(y) - d(y) - \alpha(y, x)\widehat{n}_x. \tag{2.14}$$

Note that part (B) of the theorem characterizes the limiting occupation measure $\widehat{\Gamma}$. Once Γ is known, χ is characterized by the martingale problem (2.8). The next remark motivates the formula for \widehat{n}_x and the fitness function $\text{Fit}(y, x)$.

Remark 2.5 Consider a Lotka-Volterra system that results from the large number limit of the logistic competition process without mutation. If the system has two traits x and y , then the population sizes $n_x(t)$ and $n_y(t)$ corresponding to these traits evolve according to the system of ordinary differential equations given by

$$\begin{aligned} \frac{dn_x}{dt} &= n_x(t)(b(x) - d(x) - \alpha(x, x)n_x(t) - \alpha(x, y)n_y(t)) \\ \frac{dn_y}{dt} &= n_y(t)(b(y) - d(y) - \alpha(y, x)n_x(t) - \alpha(y, y)n_y(t)), \end{aligned} \tag{2.15}$$

where α is the competition kernel. Observe that $(\widehat{n}_x, 0)$ is a fixed point for this system. The fitness function $\text{Fit}(y, x)$ describes the growth rate of a negligible mutant population with trait y in an environment characterized by \widehat{n}_x . Furthermore, the fixed point $(\widehat{n}_x, 0)$ is asymptotically stable if and only if $\text{Fit}(y, x) < 0$. However, the analysis of such a dynamical system (done in [2]) is not necessary for our purpose.

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, $\sup_{K \geq 1} \mathbb{E}(\langle X^K(0), 1 \rangle^2) < \infty$ and $Ku_K \rightarrow 0$ as $K \rightarrow \infty$. 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 χ satisfies the martingale problem given by (2.8) and Γ satisfies (2.10).*

Proof To prove the tightness of $\{\chi^K : K \in \mathbb{N}^*\}$, we use a criterion from [12]. Let $n^K(t) = \langle \chi^K(t), 1 \rangle$ for $t \geq 0$. This process counts the number of mutations that occurred in the population. For any $T > 0$ we have by Lemma 2.1 that

$$\sup_{K \geq 1} \mathbb{E}(n^K(T)) \leq \|b\|_\infty T \sup_{K \geq 1, t \leq T} \mathbb{E}(\langle Z^K(t), 1 \rangle) \leq \|b\|_\infty T \sup_{K \geq 1, t \geq 0} \mathbb{E}(\langle X^K(t), 1 \rangle) < \infty. \tag{3.1}$$

From this estimate and the martingale problem (2.7), 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 $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ and 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

$$\sup_{K \geq 1, t \geq 0} \mathbb{P}(\langle Z^K(t), 1 \rangle > N_\epsilon) < \epsilon. \tag{3.2}$$

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 $T > 0$

$$\inf_{K \geq 1} \mathbb{E}(\Gamma^K([0, T] \times \mathcal{K}_\epsilon)) \geq (1 - \epsilon)T. \tag{3.3}$$

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 Fubini’s theorem and (3.2). From Lemma 1.3 of [18], $\{\Gamma^K : K \in \mathbb{N}^*\}$ 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 [18]. From part (A) we know that the distributions of $\{(\chi^K, \Gamma^K) : K \in \mathbb{N}^*\}$ are tight. Therefore there exists a subsequence $\{\eta_K\}$ along which (χ^K, Γ^K) converges in distribution to a limit (χ, Γ) . We can take the limit in (2.7) along this subsequence and show that $M_t^{\chi, K}$ converges in distribution to the martingale given by (2.8).

Let us now show that the limiting value Γ satisfies (2.10). From (2.4), for any $F_f \in \mathbb{F}_b^2$, we get that

$$\begin{aligned} m_t^{F, f, K} &= F_f(Z^K(t)) - F_f(Z^K(0)) - \int_0^t \mathbb{L}^K F_f(Z^K(s)) ds \\ &= F_f(Z^K(t)) - F_f(Z^K(0)) - \left(\frac{1}{Ku_K}\right) \int_0^t \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B} F_f(\mu) \Gamma^K(ds \times d\mu) \\ &\quad - \frac{\delta^{F, f, K}(t)}{Ku_K} \end{aligned} \tag{3.4}$$

is a martingale. Here the operator \mathbb{B} is defined by (2.11) and

$$\delta^{F, f, K}(t) = \int_0^t (Ku_K \mathbb{L}^K F_f(Z^K(s)) - \mathbb{B} F_f(Z^K(s))) ds. \tag{3.5}$$

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

$$\begin{aligned}
 & Ku_K \mathbb{L}^K F_f(\mu) - \mathbb{B}F_f(\mu) \\
 &= Ku_K \int_{\mathbb{X}} p(x)b(x) \left[\int_{\mathbb{X}} \left(F_f\left(\mu + \frac{1}{K}\delta_y\right) - F_f(\mu) \right) m(x, dy) \right] \mu(dx) \\
 &\quad + K \int_{\mathbb{X}} b(x) \left(F_f\left(\mu + \frac{1}{K}\delta_x\right) - F_f(\mu) - \frac{1}{K}F'(\langle\mu, f\rangle)f(x) \right) \mu(dx) \\
 &\quad + K \int_{\mathbb{X}} (d(x) + \langle\mu, \alpha(x, \cdot)\rangle) \left(F_f\left(\mu - \frac{1}{K}\delta_x\right) - F_f(\mu) + \frac{1}{K}F'(\langle\mu, f\rangle)f(x) \right) \mu(dx) \\
 &\quad - Ku_K \int_{\mathbb{X}} p(x)b(x) \left(F_f\left(\mu + \frac{1}{K}\delta_x\right) - F_f(\mu) \right) \mu(dx). \tag{3.6}
 \end{aligned}$$

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

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

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,

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

Pick any $T > 0$. This estimate along with Lemma 2.1 implies that as $K \rightarrow \infty$, $\delta^{F, f, K}(t)$ (given by (3.5)) converges to 0 in $L^1(d\mathbb{P})$, uniformly in $t \in [0, T]$. Multiplying (3.4) by Ku_K and letting $K \rightarrow \infty$, we get that along the subsequence η_K , the sequence of martingales $\{Ku_K m^{F, f, K} : K \in \mathbb{N}^*\}$ converges 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 $F_f \in \mathbb{F}_b^2$,

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

Separability of \mathbb{F}_b^2 ensures that (2.10) also holds. □

Theorem 3.1 shows tightness of the family of occupation measures $\{\Gamma^K : K \in \mathbb{N}^*\}$. In the next section we will prove that this family has a unique limit point Γ which is the occupation measure of the process $\{\hat{n}_{\chi'(t)}\delta_{\chi'(t)} : t \geq 0\}$ (see Theorem 2.4) that corresponds to the TSS. This demonstrates the convergence of Z^K to the TSS in the sense of occupation measures. Note that the process Z^K does not converge in the Skorokhod topology as $K \rightarrow \infty$ (see Proposition 1 in [2]), due to sharp transitions in the process at the time of mutations. Hence the convergence of Z^K to the TSS is shown in the sense of finite dimensional distributions in [2]. In our approach, instead of working with Z^K we work with its occupation measure Γ^K which remains relatively unaffected by the behaviour of Z^K over small time intervals. This mitigates the problem of having sharp mutation-induced transitions in Z^K and simplifies the analysis.

4 Characterization of the Limiting Values

4.1 Dynamics Without Mutation

As in [2, 4], to understand the information provided by (2.10), we need to consider the dynamics of monomorphic and dimorphic populations. Our purpose in this section is to show that under the time scale separation given by Assumption 2.3, the operator (2.11) and Assumption 1.1(B) characterize the state of the population between two mutant arrivals. Because of Assumption 1.1(B), we will see that two different traits cannot coexist in the long term and thus it suffices to work with monomorphic or dimorphic initial populations (i.e. the support of Z_0^K is one or two singletons).

In Sect. 4.1.1, we consider monomorphic or dimorphic populations and show convergence of the occupation measures when the final trait composition of the population is known. For instance, if the final trait is x_0 , then the occupation measure of $Z^K(dx, dt)$ converges to $\hat{n}_{x_0}\delta_{x_0}(dx)dt$. In Sect. 4.1.2, we use couplings with linear 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 Γ^K in the Absence of Mutation

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

Definition 4.1 Let L_0^K be the operator L^K (given by (2.1)) with $p(x) = 0$ for all $x \in \mathbb{X}$. We will denote by $\{Y^K(t) : t \geq 0\}$ a process with generator L_0^K and an initial condition that varies according to the case that is studied. This process has the same birth-death dynamics as a process with generator L^K , but there is no mutation.

In this section we investigate how a process with generator L_0^K behaves at time scales of order $(Ku_K)^{-1}$, 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

$$\pi(\{\mu \in \mathcal{M}_F(\mathbb{X}) : \{x\} \subset \text{supp}(\mu) \subset \{x, y\}\}) = 1 \tag{4.1}$$

and

$$\int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B} F_f(\mu) \pi(d\mu) = 0 \tag{4.2}$$

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 (2.14).

Proof Since π satisfies (4.1), any μ picked from the distribution π has the form $\mu = n_x \delta_x + n_y \delta_y$ with $n_x > 0$. Let Φ 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 of the distribution π under Φ^{-1} . Replacing the operator \mathbb{B} by its definition, we can rewrite (4.2) as

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

This equation can hold for all $F_f \in \mathbb{F}_b^2$ only if the support of π^* 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 $(\widehat{n}_x, 0)$ and

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

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

Remark 4.3 Note that $(0, 0)$, $(\widehat{n}_x, 0)$, $(0, \widehat{n}_y)$ and $(\widetilde{n}_x, \widetilde{n}_y)$ are the stationary solutions of the Lotka-Volterra system given by (2.15).

Heuristically, the ‘‘invasion implies substitution’’ assumption prevents two traits from coexisting in the long run. If we know which trait survives, then we know that it fixates and Proposition 4.2 provides the form of the solution π to (4.2). 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:*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(Y_t^K \{x\} < \delta \text{ for some } t \in \left[0, \frac{T}{Ku_K} \right] \right) = 0. \tag{4.3}$$

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

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

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 Γ_0 be a limit point. Then from part (C) of Theorem 3.1 we get that

$$\int_0^T \int_{\mathcal{M}_F(\mathbb{X})} \mathbb{B}F_f(\mu) \Gamma_0(dt \times d\mu) = 0 \quad \text{for all } F_f \in \mathbb{F}_b^2 \text{ a.s.,} \tag{4.4}$$

where the operator \mathbb{B} is given by (2.11).

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

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

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

$$\begin{aligned} 0 \leq \mathbb{E}(T - \Gamma_0^K([0, T] \times \mathcal{S}_\delta)) &= Ku_K \mathbb{E}\left[\int_0^{\frac{T}{Ku_K}} (1 - \mathbb{1}_{\mathcal{S}_\delta}(Y^K(t))) dt\right] \\ &\leq T \mathbb{P}\left(Y_t^K\{x\} < \delta \text{ for some } t \in \left[0, \frac{T}{Ku_K}\right]\right). \end{aligned}$$

Hence by (4.3) we get that $\Gamma_0^K([0, T] \times \mathcal{S}_\delta)$ converges to T in $L^1(d\mathbb{P})$. Because \mathcal{S}_δ is a closed set, $\Gamma_0([0, T] \times \mathcal{S}_\delta) = T$ a.s.

Let π 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 π satisfies (4.1) 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. □

4.1.2 Fixation Probabilities

We have seen in Corollary 4.4 that the behaviour of a dimorphic population is known provided we know which trait survives and then fixates. Following Champagnat et al. [2, 4], we can answer this question by using couplings with branching processes. This is done in Propositions 4.6 and 4.7, whose proofs are given in the Appendix. These propositions study populations evolving as Markov processes with generator L_0^K (see Definition 4.1). Proposition 4.6 shows that over a time period of order $(Ku_K)^{-1}$, a monomorphic population with a non-negligible initial size does not die and a monomorphic population that is initially near equilibrium remains near equilibrium. Proposition 4.7 considers a dimorphic population, where the resident population is near equilibrium while the mutant population has a small size. It shows that an unfavourable mutant will certainly die out quickly (in the evolutionary time scale), while a favourable mutant can invade the population with a positive probability whose value can be easily computed. Moreover after a successful invasion, the mutant population will not die in a time period of order $(Ku_K)^{-1}$. Formally, the process we consider in Propositions 4.6 and 4.7 can be defined as follows.

Definition 4.5 Pick two trait values $x, y \in \mathbb{X}$ and two \mathbb{N}^* -valued sequences $\{z_1^K : K \in \mathbb{N}^*\}$ and $\{z_2^K : K \in \mathbb{N}^*\}$. Let $\{Y^K(t) : t \geq 0\}$ be the process with generator L_0^K (see Definition 4.1) and initial condition

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

Here x and y should be seen as the resident trait and the mutant trait respectively.

For any $x \in \mathbb{X}$ and $\epsilon > 0$ let

$$\mathcal{N}_\epsilon(x) = \{ \mu \in \mathcal{M}_F(\mathbb{X}) : \text{supp}(\mu) = \{x\} \text{ and } \langle \mu, 1 \rangle \in [\widehat{n}_x - \epsilon, \widehat{n}_x + \epsilon] \}. \tag{4.5}$$

Proposition 4.6 (Behaviour of a monomorphic population) *Suppose that Assumptions 1.1 and 2.3 hold. Let $\{Y^K(t) : t \geq 0\}$ be the process given by Definition 4.5. Assume that $z_2^K = 0$ for all $K \in \mathbb{N}^*$. Then for any $T > 0$ we have the following.*

(A) *A monomorphic population with a non-negligible size does not die in a time of order $(Ku_K)^{-1}$: Suppose that for some $\epsilon > 0$, $z_1^K \geq K\epsilon$ for each $K \in \mathbb{N}^*$. Then for some $\delta > 0$*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\exists t \in \left[0, \frac{T}{Ku_K} \right], Y_t^K \{x\} < \delta \right) = 0. \tag{4.6}$$

(B) *A monomorphic population with a size around equilibrium remains near equilibrium for a time of order $(Ku_K)^{-1}$: Suppose that for some $\epsilon > 0$, $z_1^K \in [K(\widehat{n}_x - \epsilon), K(\widehat{n}_x + \epsilon)]$ for each $K \in \mathbb{N}^*$. Then*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\exists t \in \left[0, \frac{T}{Ku_K} \right], Y^K(t) \notin \mathcal{N}_{2\epsilon}(x) \right) = 0. \tag{4.7}$$

Proposition 4.7 (Behaviour of a dimorphic population) *Suppose that Assumptions 1.1 and 2.3 hold. Let $\{Y^K(t) : t \geq 0\}$ be the process given by Definition 4.5. We will assume that the resident population is near equilibrium, that is, for a small $\epsilon > 0$ we have $z_1^K \in [K(\widehat{n}_x - \epsilon), K(\widehat{n}_x + \epsilon)]$ for all $K \in \mathbb{N}^*$. Suppose that $\{t_K\}$ is any \mathbb{N}^* -valued sequence such that $\log K \ll t_K \ll 1/Ku_K$. Let $S_K = (\widehat{n}_x - 2\epsilon, \widehat{n}_x + 2\epsilon) \times (0, 2\epsilon)$ and let T_{S_K} be the stopping time*

$$T_{S_K} = \inf \{ t \geq 0 : Y^K(t) \notin S_K \}. \tag{4.8}$$

Then we have the following.

(A) *A favorable mutant with a non-negligible size does not die in a time of order $(Ku_K)^{-1}$: Suppose that $\text{Fit}(y, x) > 0$ and $z_2^K > K\epsilon$ for all $K \in \mathbb{N}^*$. There exists an $\epsilon_0 > 0$ such that if $\epsilon < \epsilon_0$ then*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(\exists t \in \left[0, \frac{T}{Ku_K} \right], Y_t^K \{y\} < \frac{\epsilon}{2} \right) = 0. \tag{4.9}$$

(B) *An unfavorable mutant dies out in time t_K : Let $\text{Fit}(y, x) < 0$ and $z_2^K < K\epsilon$ for all $K \in \mathbb{N}^*$. There exists an $\epsilon_0 > 0$ such that if $\epsilon < \epsilon_0$ then*

$$\lim_{K \rightarrow \infty} \mathbb{P}(T_{S_K} \leq t_K, Y_{T_{S_K}}^K \{y\} = 0) = 1. \tag{4.10}$$

(C) A favorable mutant either dies out or invades the population in time t_K . The probability of invasion is the fitness of the mutant with respect to the resident trait divided by its birth rate: let $\text{Fit}(y, x) > 0$ and $z_2^K = 1$ for all $K \in \mathbb{N}^*$. Then there exist positive constants c, ϵ_0 such that for all $\epsilon < \epsilon_0$ we have

$$\limsup_{K \rightarrow \infty} \left| \mathbb{P}(T_{S_K} \leq t_K, Y_{T_{S_K}}^K \{y\} \geq 2\epsilon) - \frac{\text{Fit}(y, x)}{b(y)} \right| \leq c\epsilon, \tag{4.11}$$

$$\text{and } \limsup_{K \rightarrow \infty} \left| \mathbb{P}(T_{S_K} \leq t_K, Y_{T_{S_K}}^K \{y\} = 0) - \left(1 - \frac{\text{Fit}(y, x)}{b(y)}\right) \right| \leq c\epsilon. \tag{4.12}$$

Using Propositions 4.6 and 4.7, we can retrieve the state of the process in a large time window $[\epsilon/Ku_K, \epsilon^{-1}/Ku_K]$ from the initial condition. This allows us to understand what will happen to the population if we neglect the transitions due to rare mutation events.

Corollary 4.8 *Suppose that Assumptions 1.1 and 2.3 hold. For each $K \in \mathbb{N}^*$ let $\{Y^K(t) : t \geq 0\}$ be a process with generator L_0^K .*

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

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[\frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K} \right] \right) = 1. \tag{4.13}$$

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

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

(C) *Suppose that for some $x, y \in \mathbb{X}$ and $\epsilon > 0$ we have $\text{Fit}(y, x) > 0$, $\text{supp}(Y^K(0)) = \{x, y\}$, $Y_0^K \{x\} \in [\widehat{n}_x - \epsilon, \widehat{n}_x + \epsilon]$ 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 We first prove part (A). Let Y^K be the process with generator L_0^K such that $\text{supp}(Y^K(0)) = \{x\}$ and $Y_0^K \{x\} > \epsilon$. Part (A) of Proposition 4.6 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 $F_f \in \mathbb{F}_b^2$,

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

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) consider the set $\mathcal{S}_K = (\hat{n}_x - 2\epsilon, \hat{n}_x + 2\epsilon) \times (0, 2\epsilon)$ and let $T_{\mathcal{S}_K}$ be given by (4.8). Let the sequence $\{t_K\}$ be as in Proposition 4.7 and consider the event $E^K(\epsilon) = \{T_{\mathcal{S}_K} \leq t_K, Y_{T_{\mathcal{S}_K}}^K\{y\} = 0\}$. Since $\text{Fit}(y, x) < 0$, part (B) of Proposition 4.7 says that as $K \rightarrow \infty$, the probability of the event $E^K(\epsilon)$ approaches 1 for a sufficiently small $\epsilon > 0$. Note that $Ku_K t_K \rightarrow 0$ as $K \rightarrow \infty$. The proof of part (B) follows from part (A) of this corollary along with the strong Markov property at time $T_{\mathcal{S}_K}$.

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

$$\begin{aligned} & \mathbb{P}\left(Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[\frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K}\right]\right) \\ &= \sum_{i=1}^3 \mathbb{P}\left(Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[\frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K}\right]; E_i^K(\epsilon)\right), \end{aligned} \tag{4.14}$$

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 (4.14) corresponding to $i = 1$. On the event $E_1^K(\epsilon)$, we have $Y_{T_{\mathcal{S}_K}}^K\{x\} \in (\hat{n}_x - 2\epsilon, \hat{n}_x + 2\epsilon)$ and $Y_{T_{\mathcal{S}_K}}^K\{y\} = 0$. The strong Markov property at time $T_{\mathcal{S}_K}$, along with part (A) of this corollary and part (C) of Proposition 4.7 imply that this term converges to $1 - \text{Fit}(y, x)/b(y)$ as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$.

The term corresponding to $i = 2$ in (4.14) can be written as

$$\begin{aligned} & \mathbb{P}\left(Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[\frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K}\right]; E_2^K(\epsilon)\right) \\ &= \mathbb{E}\left(\mathbb{1}_{\{T_{\mathcal{S}_K} \leq t_K; Y_{T_{\mathcal{S}_K}}^K\{y\} \geq 2\epsilon\}} \mathbb{P}\left(Y^K(t) \in \mathcal{N}_\epsilon(x) \text{ for all } t \in \left[\frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K}\right] \middle| \mathcal{F}_{T_{\mathcal{S}_K}}\right)\right) \end{aligned} \tag{4.15}$$

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 (A) of Proposition 4.7, the probability of the process $\{Y_t^K\{y\} : t \geq 0\}$ going below ϵ between times $T_{\mathcal{S}_K}$ and $T_{\mathcal{S}_K} + \epsilon^{-1}/Ku_K$ tends to 0 as $K \rightarrow \infty$. Hence the probability of the event $\{\exists t \in [T_{\mathcal{S}_K}, T_{\mathcal{S}_K} + \epsilon^{-1}/Ku_K] : y \notin \text{supp}(Y_t^K)\}$ also tends to 0 as $K \rightarrow \infty$. Note that if $y \in \text{supp}(Y_t^K)$ then $Y^K(t) \notin \mathcal{N}_\epsilon(x)$. Conditioning by $\mathcal{F}_{T_{\mathcal{S}_K}}$ and using the strong Markov property shows that the term corresponding to $i = 2$ in (4.14) converges to 0 as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$.

Part (C) of Proposition 4.7 implies that

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_1^K(\epsilon) \cup E_2^K(\epsilon)) = 1.$$

Hence

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_3^K(\epsilon)) = 0$$

which shows that the term corresponding to $i = 3$ in (4.14) converges to 0.

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

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

is similar. This completes the proof of part (C) of the corollary. □

4.2 Proof of Theorem 2.4: Convergence to the TSS

We now have 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 (χ, Γ) be a limiting value satisfying (2.8) and (2.10). If we can prove that (2.12) holds for the process $\{\chi'(t) : t \geq 0\}$ (introduced in the statement of Theorem 2.4), then the distribution of Γ is uniquely determined, which in turn uniquely determines the distribution of χ due to the martingale problem given by (2.8). Hence we only have to prove part (B) of Theorem 2.4.

Since (χ, Γ) is a limiting value, due to Prohorov’s theorem, there exists a subsequence $\{(\tilde{\chi}^K, \tilde{\Gamma}^K)\}$ that converges in distribution to (χ, Γ) as $K \rightarrow \infty$. The Skorokhod representation theorem (see for e.g. [1]) says that on the same probability space as (χ, Γ) , we can construct a sequence which we again denote by $\{(\chi^K, \Gamma^K)\}$ with an abuse of notation, such that $(\tilde{\chi}^K, \tilde{\Gamma}^K)$ converges to (χ, Γ) a.s. and (χ^K, Γ^K) has the same marginal distributions as $\{(\tilde{\chi}^K, \tilde{\Gamma}^K)\}$.

Assuming $(\chi^K, \Gamma^K) \rightarrow (\chi, \Gamma)$ a.s. as $K \rightarrow \infty$, we now try to identify (χ, Γ) . The main idea is that between subsequent appearances of new mutants, our process $\{X^K(t) : t \geq 0\}$ behaves like the process considered in Corollary 4.8. When a fit mutant appears, it either gets extinct 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 population can be inferred from the occupation measure, because the population is monomorphic and the size is shown to reach an equilibrium.

Throughout this section whenever we say “as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$ ” we mean that the limit $K \rightarrow \infty$ is taken first and the limit $\epsilon \rightarrow 0$ is taken next. For $K \in \mathbb{N}^*$, $i \in \mathbb{N}^*$ let τ_i^K and τ_i be the i -th jump times of the process χ^K and χ respectively. For convenience we define $\tau_0^K = \tau_0 = 0$. Since $(\chi^K, \Gamma^K) \rightarrow (\chi, \Gamma)$ a.s., for any $m \in \mathbb{N}^*$, $(\tau_1^K, \dots, \tau_m^K) \rightarrow (\tau_1, \dots, \tau_m)$ a.s. Using (2.8) and Lemma 2.1, we know that $\tau_i - \tau_{i-1} > 0$ almost surely for each $i \in \mathbb{N}^*$. Thus

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(\tau_i^K - \tau_{i-1}^K > \epsilon) = 1. \tag{4.16}$$

Pick an arbitrary $x_{\text{arb}} \in \mathbb{X}$. For any $i \in \mathbb{N}^*$ and $\epsilon > 0$ define

$$R_i^{K, \epsilon} = \mathbb{1}_{\{\tau_i^K - \tau_{i-1}^K > \epsilon\}} \frac{\int_{(\tau_{i-1}^K + \epsilon)}^{\tau_i^K} \int_{\mathcal{M}_F(\mathbb{X})} \langle \mu, x \rangle \Gamma^K(ds \times d\mu)}{\int_{(\tau_{i-1}^K + \epsilon)}^{\tau_i^K} \int_{\mathcal{M}_F(\mathbb{X})} \langle \mu, 1 \rangle \Gamma^K(ds \times d\mu)} + \mathbb{1}_{\{\tau_i^K - \tau_{i-1}^K \leq \epsilon\}} x_{\text{arb}}. \tag{4.17}$$

Note that if $\text{supp}(Z^K(s)) = \{x\}$ for all $s \in [\tau_{i-1}^K + \epsilon, \tau_i^K)$, then we have $R_i^{K,\epsilon} = x$. Heuristically, $R_i^{K,\epsilon}$ is an estimator for the trait that fixates between the $(i - 1)$ th and the i th mutation. Define an event

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

where for any $x \in \mathbb{X}$, $\mathcal{N}_\epsilon(x)$ is defined by (4.5). The next proposition shows how $R_i^{K,\epsilon}$ and $E_i^K(\epsilon)$ behave as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$.

Proposition 4.9 *Suppose that Assumptions 1.1, 2.2 and 2.3 hold. Then for any $i \in \mathbb{N}^*$*

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_i^K(\epsilon)) = 1. \tag{4.18}$$

Furthermore for any measurable set $A \subset \mathbb{X}$, $x \in \mathbb{X}$ and $i \in \mathbb{N}^*$ we have

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(R_{i+1}^{K,\epsilon} \in A | R_i^{K,\epsilon} = x) \\ &= \mathbb{1}_A(x) \int_{\mathbb{X}} \left(1 - \frac{[\text{Fit}(y, x)]^+}{b(y)}\right) m(x, dy) + \int_A \frac{[\text{Fit}(y, x)]^+}{b(y)} m(x, dy). \end{aligned} \tag{4.19}$$

Proof 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

$$Z^K(\tau_{i-1}^K + t) = Y_i^K\left(\frac{t}{Ku_K}\right) \text{ for all } t \in [0, \tau_i^K - \tau_{i-1}^K). \tag{4.20}$$

For $i = 1$, we obtain from part (B) of Assumption 2.2 and part (A) of Corollary 4.8 that

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(Y_1^K(t) \in \mathcal{N}_\epsilon(x_0) \text{ for all } t \in \left[\frac{\epsilon}{Ku_K}, \frac{\epsilon^{-1}}{Ku_K}\right]\right) = 1. \tag{4.21}$$

This limit along with (4.16) and (4.20) imply that $R_1^{K,\epsilon} \rightarrow x_0$ with probability converging to 1 as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$. Moreover (4.18) holds for $i = 1$ thanks to (4.20) and (4.21).

For any $i \in \mathbb{N}^*$, let U_i^K denote the type of the new mutant that appears at time τ_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 $Z_{\tau_i^K}^K\{x\} \in [\hat{n}_x - \epsilon, \hat{n}_x + \epsilon]$. Using parts (B) and (C) of Corollary 4.8 and (4.20), we obtain that

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

and $\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_{i+1}^K(\epsilon), R_{i+1}^{K,\epsilon} = y | E_i^K(\epsilon), R_i^{K,\epsilon} = x, U_i^K = y) = \frac{[\text{Fit}(y, x)]^+}{b(y)}$.

Since the distribution of U_i^K conditionally to $\{E_i^K(\epsilon), R_i^{K,\epsilon} = x\}$ is $m(x, dy)$, for any measurable set $A \subset \mathbb{X}$ we get

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_{i+1}^K(\epsilon), R_{i+1}^{K,\epsilon} \in A | E_i^K(\epsilon), R_i^{K,\epsilon} = x) \\ &= \mathbb{1}_A(x) \int_{\mathbb{X}} \left(1 - \frac{[\text{Fit}(y, x)]^+}{b(y)}\right) m(x, dy) + \int_A \frac{[\text{Fit}(y, x)]^+}{b(y)} m(x, dy). \end{aligned} \tag{4.22}$$

Taking $A = \mathbb{X}$ in (4.22) we obtain

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_{i+1}^K(\epsilon) | E_i^K(\epsilon)) = 1. \tag{4.23}$$

This relation shows that if (4.18) holds for some $i \in \mathbb{N}^*$, then it also holds for $(i + 1)$. Since we have already shown that (4.18) holds for $i = 1$, by induction we can conclude that (4.18) holds for all $i \in \mathbb{N}^*$. From (4.22) and (4.23) we can deduce that for any measurable set $A \subset \mathbb{X}$, $x \in \mathbb{X}$ and $i \in \mathbb{N}^*$

$$\lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(R_{i+1}^{K,\epsilon} \in A | R_i^{K,\epsilon} = x) = \lim_{\epsilon \rightarrow 0} \lim_{K \rightarrow \infty} \mathbb{P}(E_{i+1}^K(\epsilon), R_{i+1}^{K,\epsilon} \in A | E_i^K(\epsilon), R_i^{K,\epsilon} = x).$$

Hence (4.19) holds due to (4.22). This completes the proof of the proposition. □

We now complete the proof of part (B) of Theorem 2.4 by using Proposition 4.9.

Proof of Theorem 2.4(B) For each $i \in \mathbb{N}^*$ define

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

Since $(\chi^K, \Gamma^K) \rightarrow (\chi, \Gamma)$ a.s. for any positive integer m we must have $(R_1^{K,\epsilon}, \dots, R_m^{K,\epsilon}) \rightarrow (R_1, \dots, R_m)$ a.s. as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$. Proposition 4.9 implies that $\{R_i : i \in \mathbb{N}^*\}$ is a Markov chain over \mathbb{X} with $R_1 = x_0$ and transition probabilities given by

$$\mathbb{P}(R_{i+1} \in A | R_i = x) = \mathbb{1}_A(x) \int_{\mathbb{X}} \left(1 - \frac{[\text{Fit}(y, x)]^+}{b(y)} \right) m(x, dy) + \int_A \frac{[\text{Fit}(y, x)]^+}{b(y)} m(x, dy) \tag{4.24}$$

for any measurable set $A \subset \mathbb{X}$, $x \in \mathbb{X}$ and $i \in \mathbb{N}^*$.

Define two \mathbb{X} -valued processes as

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

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

We now show that the process $\{\chi'(t) : t \geq 0\}$ uniquely characterizes the distribution of the limiting occupation measure Γ . For any $F_f \in \mathbb{F}_b^2$, $i \in \mathbb{N}^*$ and $t \geq 0$, define the real-valued variables

$$\begin{aligned} \rho_i^{K,\epsilon} &= \int_{(\tau_{i-1}^K + \epsilon) \wedge t}^{\tau_i^K \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} F_f(\mu) \Gamma^K(ds \times d\mu) \quad \text{and} \\ \rho_i &= \int_{\tau_{i-1} \wedge t}^{\tau_i \wedge t} \int_{\mathcal{M}_F(\mathbb{X})} F_f(\mu) \Gamma(ds \times d\mu). \end{aligned} \tag{4.25}$$

Then certainly $\rho_i^{K,\epsilon} \rightarrow \rho_i$ a.s. as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$. Suppose F_f has the form $F_f(\mu) = F(\langle \mu, f \rangle)$ for some $f \in C_b(\mathbb{X}, \mathbb{R})$ and $F \in C_b^2(\mathbb{R}, \mathbb{R})$. Then for any $\mu \in \mathcal{N}_\epsilon(x)$

$$|F_f(\mu) - F_f(\widehat{n}_x \delta_x)| = |F(\langle \mu, f \rangle) - F(\widehat{n}_x f(x))| \leq \|F'\|_\infty \|f\|_\infty \epsilon.$$

Therefore, on the event $E_i^K(\epsilon)$,

$$\begin{aligned} \left| \rho_i^{K,\epsilon} - \int_{(\tau_{i-1}^K + \epsilon) \wedge t}^{\tau_i^K \wedge t} F(\widehat{n}_{R_i^{K,\epsilon}} f(R_i^{K,\epsilon})) ds \right| &= \left| \int_{(\tau_{i-1}^K + \epsilon) \wedge t}^{\tau_i^K \wedge t} (F_f(Z_s^K) - F(\widehat{n}_{R_i^{K,\epsilon}} f(R_i^{K,\epsilon}))) ds \right| \\ &\leq \int_{(\tau_{i-1}^K + \epsilon) \wedge t}^{\tau_i^K \wedge t} |F_f(Z_s^K) - F(\widehat{n}_{R_i^{K,\epsilon}} f(R_i^{K,\epsilon}))| ds \\ &\leq \|F'\|_\infty \|f\|_\infty \epsilon (\tau_i^K \wedge t - (\tau_{i-1}^K + \epsilon) \wedge t). \end{aligned}$$

Since

$$\int_{(\tau_{i-1}^K + \epsilon) \wedge t}^{\tau_i^K \wedge t} F(\widehat{n}_{R_i^{K,\epsilon}} f(R_i^{K,\epsilon})) ds = (\tau_i^K \wedge t - (\tau_{i-1}^K + \epsilon) \wedge t) F(\widehat{n}_{R_i^{K,\epsilon}} f(R_i^{K,\epsilon}))$$

converges to $(\tau_i \wedge t - \tau_{i-1} \wedge t) F_f(\widehat{n}_{R_i} \delta_{R_i})$ as $K \rightarrow \infty$ and $\epsilon \rightarrow 0$, we have that

$$\rho_i = (\tau_i \wedge t - \tau_{i-1} \wedge t) F_f(\widehat{n}_{R_i} \delta_{R_i}) = \int_{\tau_{i-1} \wedge t}^{\tau_i \wedge t} F_f(\widehat{n}_{\chi'(s)} \delta_{\chi'(s)}) ds.$$

This is true for each $i \in \mathbb{N}^*$ and this ensures that (2.12) holds with χ' defined as above.

Recall that the process $\{\chi(t) : t \geq 0\}$ satisfies the martingale problem given by (2.8). Using (2.12) we obtain that for any $F_f \in \mathbb{F}_b^2$

$$\begin{aligned} F_f(\chi(t)) - F_f(\chi(0)) - \int_0^t b(\chi'(s)) p(\chi'(s)) \widehat{n}_{\chi'(s)} \\ \times \left(\int_{\mathbb{X}} (F_f(\chi(s) + \delta_y) - F_f(\chi(s))) m(\chi'(s), dy) \right) ds, \end{aligned}$$

is a martingale. This shows that if $\chi'(t) = x$, then the next jump time of the process χ is exponentially distributed with parameter $b(x)p(x)\widehat{n}_x$. Therefore for each $i \in \mathbb{N}^*$, $(\tau_i - \tau_{i-1})$ is exponentially distributed with parameter $b(R_i)p(R_i)\widehat{n}_{R_i}$. Since $\{R_i : i \in \mathbb{N}^*\}$ is a Markov chain with $R_1 = x_0$ and transition probabilities given by (4.19), we can conclude that $\{\chi'(t) : t \geq 0\}$ is a Markov process with generator \mathbb{C} (given by (2.13)) and initial state x_0 . This completes the proof of part (B) of Theorem 2.4. □

Acknowledgements The authors thank Sylvie Méléard for invaluable discussions. We also wish to give special thanks to a reviewer whose suggestions were very helpful in improving the presentation of our paper. 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. V.C.T. was also supported in part by the Labex CEMPI (ANR-11-LABX-0007-01).

Appendix

Our main aim in this section is to prove Propositions 4.6 and 4.7. The proof will involve several stages. We start by recalling some useful results for one-dimensional branching processes (Lemma A.2). We then consider a two-dimensional birth and death process, where the two components interact by influencing each other’s death rate. In Proposition A.3

we present some coupling relationships between the components of this process and one-dimensional branching processes. These relationships enable us to prove Propositions A.4 and A.5, which give some important estimates for the two-dimensional process. Finally we show how these estimates can be used in proving Propositions 4.6 and 4.7.

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 $\mathbf{P}(b, d, K)$ where $b \neq d$. Let $\{t_K\}$ be any \mathbb{N}^* -valued sequence satisfying $\log K \ll t_K$. 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 $\sup_{K \geq 1} \mathbb{E}(\sigma_K^2) < \infty$.
- (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 ∞ 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

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

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

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

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, Theorem 4]. □

Proposition A.3 *Let $\{N(t) = (N_1(t), N_2(t)) : t \geq 0\}$ be a $\mathbb{N} \times \mathbb{N}$ -valued pure jump Markov process with the following transition rates.*

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

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

$$d_i^- \leq \inf\{d_i(m, n) : (m, n) \in \mathcal{S} \cap (\mathbb{N} \times \mathbb{N})\} \leq \sup\{d_i(m, n) : (m, n) \in \mathcal{S} \cap (\mathbb{N} \times \mathbb{N})\} \leq d_i^+,$$

for $i = 1, 2$. 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 $\{N(t) : t \geq 0\}$, 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 the direct coupling of these processes. □

We now introduce some notation that will be used in Propositions A.4 and A.5. For each $K \in \mathbb{N}^*$, let $\{N^K(t) = (N_1^K(t), N_2^K(t)) : t \geq 0\}$ be the $\mathbb{N} \times \mathbb{N}$ -valued Markov process described in Proposition A.3. We allow the functions d_1 and d_2 to depend on K and so we denote them as d_1^K and d_2^K respectively. Corresponding to any set $\mathcal{S}_K \subset \mathbb{R}^2$ define the stopping time $T_{\mathcal{S}_K}$ by

$$T_{\mathcal{S}_K} = \inf\{t \geq 0 : N^K(t) \notin \mathcal{S}_K\}. \tag{A.3}$$

Proposition A.4 (Exit of N_1^K from sets away from 0) *Suppose that Assumption 2.3 is satisfied. Let $\mathcal{S}_K = [A_K, B_K] \times [C_K, D_K]$ for some positive real numbers A_K, B_K, C_K and D_K . Assume that $(N_1^K(0), N_2^K(0)) \in \mathcal{S}_K$ and there exist constants $\eta > 0, \epsilon \in (0, 1)$ such that*

$[K\eta(1 - \epsilon), K\eta(1 + \epsilon)] \subset [A_K, B_K]$. Let $d_1^-(\eta, \epsilon)$ and $d_1^+(\eta, \epsilon)$ be positive constants that satisfy

$$d_1^-(\eta, \epsilon) \leq \inf_{K \in \mathbb{N}^*} \inf \{d_1^K(m, n) : (m, n) \in [K\eta(1 - \epsilon), K\eta(1 + \epsilon)] \times [C_K, D_K]\} \tag{A.4}$$

and $d_1^+(\eta, \epsilon) \geq \sup_{K \in \mathbb{N}^*} \sup \{d_1^K(m, n) : (m, n) \in [K\eta(1 - \epsilon), K\eta(1 + \epsilon)] \times [C_K, D_K]\}$. (A.5)

Let T_{S_K} be given by (A.3). Then we have the following.

(A) If $b_1 < d_1^-(\eta, \epsilon)$ and $N_1^K(0) \leq K\eta$ then in a time interval of order $(Ku_K)^{-1}$, N_1^K does not exit the interval $[A_K, B_K]$ by B_K almost surely: For any $T > 0$

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

(B) If $b_1 > d_1^+(\eta, \epsilon)$ and $N_1^K(0) \geq K\eta$ then in a time interval of order $(Ku_K)^{-1}$, N_1^K does not exit the interval $[A_K, B_K]$ by A_K almost surely: For any $T > 0$

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

(C) 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 $i = 1, 2$. Let $d_1^+(\eta_1, \epsilon_1)$ be a positive constant that satisfies (A.5) for $\eta = \eta_1$ and $\epsilon = \epsilon_1$. Similarly let $d_1^-(\eta_2, \epsilon_2)$ be a positive constant that satisfies (A.4) for $\eta = \eta_2$ and $\epsilon = \epsilon_2$.

If $d_1^+(\eta_1, \epsilon_1) < b_1 < d_1^-(\eta_2, \epsilon_2)$ and $N_1^K(0) \in [K\eta_1, K\eta_2]$ then in a time interval of order $(Ku_K)^{-1}$, N_1^K does not exit the interval $[A_K, B_K]$ almost surely: For any $T > 0$

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

Proof We first 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 $b_1 < d_1^-(\eta, \epsilon)$, by Proposition A.3 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 γ_K is the first time 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 $c > 0$

$$\mathbb{P}(N_1^K(\sigma_K) > K\eta(1 + \epsilon)) \leq \mathbb{P}(B_+^K(\gamma_K) > K\eta(1 + \epsilon)) \leq e^{-cK}. \tag{A.6}$$

Before N_1^K exits $[A_K, B_K]$ from above, this process crosses the interval $[K\eta(1 - \epsilon), K\eta]$ several times. Let ρ_K be the number (possibly 0) of these passages in the interval $[0, T_{S_K}]$. Because of (A.6), for any $n \in \mathbb{N}$,

$$\mathbb{P}(\rho_K < n) \leq ne^{-cK}. \tag{A.7}$$

Let $n_K = \lfloor 1/Ku_K \rfloor^2$. Then

$$\begin{aligned} & \mathbb{P}\left(T_{S_K} \leq \frac{T}{Ku_K}, N_1^K(T_{S_K}) \geq B_K\right) \\ & \leq \mathbb{P}\left(T_{S_K} \leq \frac{T}{Ku_K}, N_1^K(T_{S_K}) \geq K\eta(1 + \epsilon), \rho_K \geq n_K\right) + \mathbb{P}(\rho_K < n_K) \\ & \leq \mathbb{P}\left(\sum_{i=1}^{n_K} \tau_{K,i} < \frac{T}{Ku_K}\right) + \mathbb{P}(\rho_K < n_K), \end{aligned} \tag{A.8}$$

where the $\tau_{K,1}, \dots, \tau_{K,n_K}$ denotes the durations of the first n_K passages of N_1^K from $K\eta(1 - \epsilon)$ to $K\eta$. The last term in the r.h.s. of (A.8) tends to 0 with K due to our choice of n_K , (A.7) and Assumption 2.3. Using the couplings of Proposition A.3, 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 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 part (B) of Lemma A.2, $\inf_{K \geq 1} \mathbb{E}(\tau_{K,1}) > 0$ and $\sup_{K \geq 1} \mathbb{E}(\tau_{K,1}^2) < \infty$. Applying Chebychev's inequality we get

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

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). □

Proposition A.5 (Exit of N_2^K from sets close to 0) *Suppose that Assumption 2.3 is satisfied. Let the process $\{N^K(t) = (N_1^K(t), N_2^K(t)) : t \geq 0\}$ and the set $S_K = [A_K, B_K] \times [C_K, D_K]$ be as in Proposition A.4. Let $[C_K, D_K] = [1, K\epsilon]$ for some $\epsilon > 0$ and let $\{t_K\}$ be any \mathbb{N}^* -valued sequence such that $\log K \ll t_K \ll \frac{1}{Ku_K}$. Suppose $d_2^-(\epsilon)$ and $d_2^+(\epsilon)$ are positive constants that satisfy*

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

Let T_{S_K} be given by (A.3). Then we have the following.

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

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

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

$$\liminf_{K \rightarrow \infty} \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) \geq \epsilon K) \geq 1 - \frac{d_2^+(\epsilon)}{b_2},$$

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

Proof We first prove part (A). Since $b_2 < d_2^-(\epsilon)$ and $N_2^K(0) \leq K\epsilon/2$, Proposition A.3 shows that N_2^K can be dominated on $[0, T_{S_K}]$ by a subcritical branching process B_+^K with distribution $\mathbf{P}(b_2, d_2^-(\epsilon), \lfloor K\epsilon/2 \rfloor)$. Let

$$\gamma_K^+ = \inf\{t \geq 0 : B_+^K(t) = 0 \text{ or } B_+^K(t) \geq K\epsilon\}. \tag{A.9}$$

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

$$\begin{aligned} \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) = 0) &\leq \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) = 0, T_{S_K} \geq \gamma_K^+) + \mathbb{P}(\gamma_K^+ \leq t_K, T_{S_K} < \gamma_K^+) \\ &\leq \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) = 0, T_{S_K} = \gamma_K^+) + \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) = 0, T_{S_K} < \gamma_K^+) \\ &\quad + \mathbb{P}(T_{S_K} \leq t_K, N_1^K(T_{S_K}) \notin [A_K, B_K), T_{S_K} < \gamma_K^+) \\ &\leq \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) = 0) + \mathbb{P}(T_{S_K} \leq t_K, N_1^K(T_{S_K}) \notin [A_K, B_K)). \end{aligned} \tag{A.10}$$

Due to part (C) of Proposition A.4, the limit of the second term in the r.h.s. of (A.10) is 0. Since part (C) of Lemma A.2 tells us that the limit of the l.h.s. is 1, this completes the proof of part (A).

We now prove part (B). Since $b_2 > d_2^+(\epsilon) > d_2^-(\epsilon)$ and $N_2^K(0) = 1$, using Proposition A.3, 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 $t \leq T_{S_K}$ almost surely. Let γ_K^+ be given by (A.9) and γ_K^- be the corresponding stopping time for B_-^K . The fact that B_-^K is below N_2^K until time T_{S_K} , (A.10) and part (C) of Proposition A.4 allow us to prove that

$$\lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) = 0) \leq \liminf_{K \rightarrow \infty} \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) = 0) \tag{A.11}$$

$$\text{and } \limsup_{K \rightarrow \infty} \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) = 0) \leq \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^- \leq t_K, B_-^K(\gamma_K^-) = 0). \tag{A.12}$$

Since B_+^K is above N_2^K until time T_{S_K} , we obtain

$$\limsup_{K \rightarrow \infty} \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) \geq K\epsilon) \leq \lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^+ \leq t_K, B_+^K(\gamma_K^+) \geq K\epsilon). \tag{A.13}$$

Finally, by a proof similar to the one of (A.10), we have

$$\lim_{K \rightarrow \infty} \mathbb{P}(\gamma_K^- \leq t_K, B_-^K(\gamma_K^-) \geq K\epsilon) \leq \liminf_{K \rightarrow \infty} \mathbb{P}(T_{S_K} \leq t_K, N_2^K(T_{S_K}) \geq K\epsilon). \tag{A.14}$$

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

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

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

These relations along with (A.11), (A.12), (A.13) and (A.14) prove part (B). □

Remark A.6 Let $\{Y(t) : t \geq 0\}$ be the process described in Definition 4.5. Define the $\mathbb{N} \times \mathbb{N}$ -valued process $\{N^K(t) = (N_1^K(t), N_2^K(t)) : t \geq 0\}$ 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 $\{N^K(t) : t \geq 0\}$ is an instance of the $\mathbb{N} \times \mathbb{N}$ -valued Markov process considered in Propositions A.4 and A.5 with $b_1 = b(x)$, $b_2 = b(y)$,

$$d_1^K(m, n) = \left(d(x) + \alpha(x, x) \frac{m}{K} + \alpha(x, y) \frac{n}{K} \right) \quad \text{and}$$

$$d_2^K(m, n) = \left(d(y) + \alpha(y, x) \frac{m}{K} + \alpha(y, y) \frac{n}{K} \right).$$

Using Propositions A.4 and A.5 we now give the proofs of Propositions 4.6 and 4.7.

Proof of Proposition 4.6 Corresponding to the process $\{Y^K(t) : t \geq 0\}$ let $\{N^K(t) : t \geq 0\}$ be the $\mathbb{N} \times \mathbb{N}$ -valued Markov process described in Remark A.6.

We first 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 $S'_K = [K\delta, \infty) \times \{0\}$ and let $T_{S'_K}$ be given by (A.3). The only way N^K can exit the set S'_K is by having N_1^K go below $K\delta$. Therefore

$$\lim_{K \rightarrow \infty} \mathbb{P} \left(Y_t^K \{x\} < \delta \text{ for some } t \in \left[0, \frac{T}{Ku_K} \right] \right) = \lim_{K \rightarrow \infty} \mathbb{P} \left(T_{S'_K} \leq \frac{T}{Ku_K}, N_1^K(T_{S'_K}) < K\delta \right). \tag{A.15}$$

Observe that on the set $[K\delta, 2K\delta] \times \{0\}$, the supremum of d_1^K is bounded above by $d(x) + 2\alpha(x, x)\delta$ which is less than $b(x)$. Therefore by part (B) of Proposition A.4, the limit on the r.h.s. of (A.15) is 0. This proves part (A).

For part (B) we can choose an $\epsilon > 0$ sufficiently small such that $\widehat{n}_x > 2\epsilon$. Let $\eta_1 = \widehat{n}_x - \epsilon$ and $\eta_2 = \widehat{n}_x + \epsilon$. We can find $\epsilon_1, \epsilon_2 > 0$ such that $[\eta_1(1 - \epsilon_1), \eta_1(1 + \epsilon_1)] \subset [\widehat{n}_x - 2\epsilon, \widehat{n}_x]$, $[\eta_2(1 - \epsilon_2), \eta_2(1 + \epsilon_2)] \subset [\widehat{n}_x, \widehat{n}_x + 2\epsilon]$ and on the set $[K\eta_1(1 - \epsilon_1), K\eta_1(1 + \epsilon_1)] \times \{0\}$ the supremum of d_1^K is strictly below $b(x)$ while on the set $[K\eta_2(1 - \epsilon_2), K\eta_2(1 + \epsilon_2)] \times \{0\}$ the infimum of d_1^K is strictly above $b(x)$. Let $S'_K = [K(\widehat{n}_x - 2\epsilon), K(\widehat{n}_x + 2\epsilon)] \times \{0\}$ and let $T_{S'_K}$ be given by (A.3). From part (C) of Proposition A.4 we get

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

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

Proof of Proposition 4.7 Corresponding to the process $\{Y^K(t) : t \geq 0\}$ let $\{N^K(t) : t \geq 0\}$ be the $\mathbb{N} \times \mathbb{N}$ -valued Markov process described in Remark A.6.

We first prove part (A). Note that $\text{Fit}(y, x) > 0$. We can choose an $\epsilon_0 > 0$ such that $d(y) + \alpha(y, x)(\widehat{n}_x + 2\epsilon_0) + 2\alpha(y, y)\epsilon_0 = b(y)$. Now let $\epsilon < \epsilon_0$ and assume that $z_1^K < K(\widehat{n}_x + \epsilon)$ and $z_2^K > K\epsilon$ for all $K \in \mathbb{N}^*$. Define the set $S'_K = [0, K(\widehat{n}_x + 2\epsilon)) \times [K\epsilon/2, \infty)$ and let $T_{S'_K}$ be given by (A.3). It is easy to see that

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

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

For part (B), observe that $\text{Fit}(y, x) < 0$ and let $\epsilon_0 > 0$ satisfy $d(y) + \alpha(y, x)(\widehat{n}_x - 2\epsilon_0) = b(y)$. Pick an $\epsilon \in (0, \epsilon_0)$. On the set $[K(\widehat{n}_x - 2\epsilon), K(\widehat{n}_x + 2\epsilon)] \times [1, 2K\epsilon)$ the infimum of d_2^K is greater than $b(y)$. Part (A) of Proposition A.5 proves part (B).

For part (C), note that $\text{Fit}(y, x) > 0$ and let $\epsilon_0 > 0$ satisfy $d(y) + \alpha(y, x)(\widehat{n}_x + 2\epsilon_0) + 2\alpha(y, y)\epsilon_0 = b(y)$. Pick an $\epsilon \in (0, \epsilon_0)$. On the set $[K(\widehat{n}_x - 2\epsilon), K(\widehat{n}_x + 2\epsilon)] \times [1, 2K\epsilon)$ the supremum of d_2^K is less than $d_2^+(\epsilon) := d(y) + \alpha(y, x)(\widehat{n}_x + 2\epsilon) + 2\alpha(y, y)\epsilon$ and the infimum of d_2^K is greater than $d_2^-(\epsilon) := d(y) + \alpha(y, x)(\widehat{n}_x - 2\epsilon)$. Both $d_2^+(\epsilon)$ and $d_2^-(\epsilon)$ are less than $b(y)$. Part (B) of Proposition A.5 proves part (C). □

References

1. Billingsley, P.: *Convergence of Probability Measures*. Wiley, New York (1968)
2. Champagnat, N.: A microscopic interpretation for adaptative dynamics trait substitution sequence models. *Stoch. Process. Appl.* **116**(8), 1127–1160 (2006)
3. Champagnat, N., Méléard, S.: Polymorphic evolution sequence and evolutionary branching. *Probab. Theory Relat. Fields* **151**, 45–94 (2011)
4. Champagnat, N., Ferrière, R., Méléard, S.: Individual-based probabilistic models of adaptative evolution and various scaling approximations. In: Dalang, R.C., Dozzi, M., Russo, F. (eds.) *Proceedings of the 5th Seminar on Stochastic Analysis, Random Fields and Applications*, Ascona, May 2005. *Progress in Probability*, vol. 59, pp. 75–114. Birkhäuser, Basel (2006)
5. Champagnat, N., Ferrière, R., Méléard, S.: Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models via timescale separation. *Theor. Popul. Biol.* **69**, 297–321 (2006)
6. Champagnat, N., Jabin, P.-E., Méléard, S.: Adaptation in a stochastic multi-resources chemostat model (2013). <http://arxiv.org/abs/1302.0552>
7. Collet, P., Méléard, S., Metz, J.A.J.: A rigorous model study of the adaptative dynamics of Mendelian diploids (2011). <http://arxiv.org/abs/1111.6234>
8. Dawson, D.A.: Measure-valued Markov processes. In: Hennequin, P.-L. (ed.) *Ecole d'Été de probabilités de Saint-Flour XXI. Lecture Notes in Math.*, vol. 1541, pp. 1–260. Springer, New York (1993)
9. Dieckmann, U., Law, R.: The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* **34**(5–6), 579–612 (1966)
10. Dieckmann, U., Heino, M., Parvinen, K.: The adaptive dynamics of function-valued traits. *J. Theor. Biol.* **241**(2), 370–389 (2006)

11. Durinx, M., Metz, J.A.J., Meszéna, G.: Adaptive dynamics for physiologically structured models. *J. Math. Biol.* **56**(5), 673–742 (2008)
12. Ethier, S.N., Kurtz, T.G.: *Markov Processes, Characterization and Convergence*. Wiley, New York (1986)
13. Geritz, S.A.H., Kisdi, É., Meszéna, G., Metz, J.A.J.: Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57
14. Gupta, A., Metz, J.A.J., Tran, V.C.: Work in progress
15. Kallenberg, O.: *Random Measures*. Academic Press, New York (1983)
16. Klebaner, F.C., Sagitov, S., Vatutin, V.A., Haccou, P., Jagers, P.: Stochasticity in the adaptive dynamics of evolution: the bare bones. *J. Biol. Dyn.* **5**(2), 147–162 (2011)
17. Kurtz, T.G.: *Approximation of Population Processes*. SIAM, Philadelphia (1981)
18. Kurtz, T.G.: Averaging for martingale problems and stochastic approximation. In: Karatzas, I., Ocone, D. (eds.) *Applied Stochastic Analysis*, New Brunswick, NJ, 1991. *Lecture Notes in Control and Inform. Sci.*, vol. 177, pp. 186–209. Springer, Berlin (1992)
19. Méléard, S., Tran, V.C.: Trait substitution sequence process and canonical equation for age-structured populations. *J. Math. Biol.* **58**(6), 881–921 (2009)
20. Méléard, S., Tran, V.C.: Slow and fast scales for superprocess limits of age-structured populations. *Stoch. Process. Appl.* **122**(1), 250–276 (2012)
21. Méléard, S., Metz, J.A.J., Tran, V.C.: Limiting Feller diffusions for logistic populations with age-structure. In: *Proceedings of the 58th World Statistics Congress—ISI2011*. ISI (2011). Preprint HAL 00595928
22. Metz, J.A.J., Nisbet, R.M., Geritz, S.A.H.: How should we define fitness for general ecological scenarios. *Trends Ecol. Evol.* **7**(6), 198–202 (1992)
23. Metz, J.A.J., Geritz, S.A.H., Meszéna, G., Jacobs, F.A.J., Van Heerwaarden, J.S.: Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In: Van Strien, S.J., Verduyn Lunel, S.M. (eds.) *Stochastic and Spatial Structures of Dynamical Systems*. KNAW Verhandelingen Afd. Natuurkunde, Eerste reeks, vol. 45, pp. 183–231. North Holland, Amsterdam (1996)
24. Parvinen, K., Dieckmann, U., Heino, M.: Function-valued adaptive dynamics and the calculus of variations. *J. Math. Biol.* **52**(1), 1–26 (2006)