Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models

Nicolas Champagnat[a,b], Régis Ferrière[a,c,*], Sylvie Méléard[b]

[a]*Laboratoire d'Ecologie, Equipe Eco-Evolution Mathématique, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris Cedex 05, France*
[b]*Equipe MODAL'X Université Paris 10, 200 avenue de la République, 92001 Nanterre Cedex, France*
[c]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA*

Abstract

A distinctive signature of living systems is Darwinian evolution, that is, a propensity to generate as well as self-select individual diversity. To capture this essential feature of life while describing the dynamics of populations, mathematical models must be rooted in the microscopic, stochastic description of discrete individuals characterized by one or several adaptive traits and interacting with each other. The simplest models assume asexual reproduction and haploid genetics: an offspring usually inherits the trait values of her progenitor, except when a mutation causes the offspring to take a mutation step to new trait values; selection follows from ecological interactions among individuals. Here we present a rigorous construction of the microscopic population process that captures the probabilistic dynamics over continuous time of birth, mutation, and death, as influenced by the trait values of each individual, and interactions between individuals. A by-product of this formal construction is a general algorithm for efficient numerical simulation of the individual-level model. Once the microscopic process is in place, we derive different macroscopic models of adaptive evolution. These models differ in the renormalization they assume, i.e. in the limits taken, in specific orders, on population size, mutation rate, mutation step, while rescaling time accordingly. The macroscopic models also differ in their mathematical nature: deterministic, in the form of ordinary, integro-, or partial differential equations, or probabilistic, like stochastic partial differential equations or superprocesses. These models include extensions of Kimura's equation (and of its approximation for small mutation effects) to frequency- and density-dependent selection. A novel class of macroscopic models obtains when assuming that individual birth and death occur on a short timescale compared with the timescale of typical population growth. On a timescale of very rare mutations, we establish rigorously the models of "trait substitution sequences" and their approximation known as the "canonical equation of adaptive dynamics". We extend these models to account for mutation bias and random drift between multiple evolutionary attractors. The renormalization approach used in this study also opens promising avenues to study and predict patterns of life-history allometries, thereby bridging individual physiology, genetic variation, and ecological interactions in a common evolutionary framework.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Adaptive evolution; Individual-based model; Birth and death point process; Body size scaling; Timescale separation; Mutagenesis; Density-dependent selection; Frequency-dependent selection; Invasion fitness; Adaptive dynamics; Canonical equation; Nonlinear stochastic partial differential equations; Nonlinear PDEs; Large deviation principle

1. Introduction

Evolutionary biology has long received the enlightenment of mathematics. At the dawn of the twentieth century, Darwinian evolution was viewed essentially as a formal theory that could only be tested using mathematical and statistical techniques. The founding fathers of evolutionary genetics (Fisher, Haldane and Wright) used mathematical models to generate a synthesis between Mendelian genetics and Darwinian evolution that paved the way toward contemporary models of adaptive evolution. However, the development of a general and coherent framework for adaptive evolution modelling, built from the basic stochastic processes acting at the individual level, is far from complete (Page and Nowak, 2002). Mathematical

*Corresponding author. Fax: +33 1 4432 3885.
E-mail address: ferriere@biologie.ens.fr (R. Ferrière).

models of adaptive evolution are essentially phenomenological, rather than derived from the 'first principles' of individual birth, mutation, interaction and death. Here we report the rigorous mathematical derivation of macroscopic models of evolutionary dynamics scaling up from the microscopic description of demographic and ecological stochastic processes acting at the individual level. Our analysis emphasizes that different models obtain depending on how individual processes are renormalized, and provides a unified framework for understanding how these different models relate to each other.

Early models of adaptive evolution pictured the mutation-selection process as a steady ascent on a so-called 'adaptive landscape', thereby suggesting some solid ground over which the population would move, under the pressure of environmental factors (Wright, 1969). The next theoretical step was to recognize that the adaptive landscape metaphor misses one-half of the evolutionary process: although the environment selects the adaptations, these adaptations can shape the environment (Haldane, 1932; Pimentel, 1968; Stenseth, 1986; Metz et al., 1992). Therefore, there is no such thing as a pre-defined adaptive landscape; in fact, the fitness of a phenotype depends upon the phenotypic composition of the population, and selection generally is frequency-dependent (Metz et al., 1992; Heino et al., 1998). Throughout the last 50 years this viewpoint spread and affected not only the intuition of evolutionary biologists, but also their mathematical tools (Nowak and Sigmund, 2004). The notion of adaptive landscape across mathematical evolutionary theories is reviewed in Kirkpatrick and Rousset (2004).

Game theory was imported from economics into evolutionary theory, in which it became a popular framework for the construction of frequency-dependent models of natural selection (Hamilton, 1967; Maynard Smith and Price, 1973; Maynard Smith, 1982; Hofbauer and Sigmund, 1998, 2003; Nowak and Sigmund, 2004). With adaptive dynamics modelling, evolutionary game theory was extended to handle the complexity of ecological systems from which selective pressures emanate. However, the rare mutation and large population scenario assumed by adaptive dynamics modelling implies that the complexity of stochastic individual life-history events is subsumed into deterministic steps of mutant invasion-fixation, taking place in vanishingly small time by the whole population as a single, monomorphic entity (Metz et al., 1996; Dieckmann and Law, 1996). Thus, adaptive dynamics models make approximations that bypass rather than encompass the individual level (Nowak and Sigmund, 2004).

An alternate pathway has been followed by population and quantitative genetics, domains in which the emphasis was early on set on understanding the forces that maintain genetic variation (Bürger, 2000). The 'continuum-of-alleles' model introduced by Crow and Kimura (1964) does not impose a rare mutation scenario, but otherwise shares the same basic assumptions as in evolutionary game theory

and adaptive dynamics models: the genetic system involves one-locus haploid asexual individuals, and the effect of mutant alleles are randomly chosen from a continuous distribution. The mathematical study of the continuum-of-alleles model has begun only relatively recently (see Bürger, 1998, 2000 and Waxman, 2003 for reviews) in a frequency-independent selection framework. The mutation-selection dynamics of quantitative traits under frequency-dependent selection has been investigated thoroughly by Bürger (2005) and Bürger and Gimelfarb (2004) in the wake of Bulmer's (1974), Slatkin's (1979), Nagylaki's (1979), Christiansen and Loeschcke (1980) and Asmussen's (1983) seminal studies. After Matessi and Di Pasquale (1996), among others, had emphasized multilocus genetics as causing long-term evolution to depart from the Wrightian model, Bürger's (2005) approach has taken a major step forward in further tying up details of the genetic system with population demography.

Recent advances in probability theory (first applied in the population biological context by Fournier and Méléard, 2004 and Champagnat, 2004b) make the time ripe for attempting systematic derivation of macroscopic models of evolutionary dynamics from individual-based processes. By scaling up from the level of individuals and stochastic processes acting upon them to the macroscopic dynamics of population evolution, we aim at setting up the mathematical framework needed for bridging behavioral, ecological and evolutionary processes (Jansen and Mulder, 1999; Abrams, 2001; Ferrière et al., 2004; Dieckmann and Ferrière, 2004; Hairston et al., 2005). Our baseline model is a stochastic process describing a finite population of discrete interacting individuals characterized by one or several adaptive phenotypic traits. We focus on the simplest case of asexual reproduction and haploid genetics. The infinitesimal generator of this process captures the probabilistic dynamics, over continuous time, of birth, mutation and death, as influenced by the trait values of each individual and ecological interactions among individuals. The rigorous algorithmic construction of the population process is given in Section 2. This algorithm is implemented numerically and simulations are presented; they unveil qualitatively different evolutionary behaviors as a consequence of varying the order of magnitude of population size, mutation probability and mutation step size. These phenomena are investigated in the next sections, by systematically deriving macroscopic models from the individual-based process. Our first approach (Section 3) aims at deriving deterministic equations to describe the moments of trajectories of the point process, i.e. the statistics of a large number of independent realizations of the process. The model takes the form of a hierarchical system of moment equations embedded into each other; the competition kernels that capture individual interactions make it impossible, even in the simple mean-field case of random and uniform interactions among phenotypes, to find simple moment closures that would decorrelate the system.

The alternate approach involves renormalizing the individual-level process by means of a large population limit. Applied by itself, the limit yields a deterministic, nonlinear integro-differential equation (Section 4.1). For different scalings of birth, death and mutation rates, we obtain qualitatively different PDE limits, in which some form of demographic randomness may or may not be retained as a stochastic term (Section 4). More specifically, when combined with the acceleration of birth (hence the acceleration of mutation) and death and an asymptotic of small mutation steps, the large population limit yields either a deterministic nonlinear reaction–diffusion model, or a stochastic measure-valued process, depending on the acceleration rate of the birth-and-death process (Section 4.2.1). When this acceleration of birth and death is combined with a rare mutation limit, the large population approximation yields a nonlinear integro-differential equation, either deterministic or stochastic, depending again upon the acceleration rate of the birth-and-death process (Section 4.2.2). In Section 5, we assume that the ancestral population is monomorphic and that the timescale of ecological interactions and evolutionary change are separated: the birth-and-death process is fast while mutations are rare. In a large population limit, the process converges on the mutation timescale to a jump process over the trait space, which corresponds to the trait substitution sequence of adaptive dynamics modelling (Section 5.1). By rescaling the mutation step (making it infinitesimal) we finally recover a deterministic process driven by the so-called "canonical equation of adaptive dynamics" first introduced by Dieckmann and Law (1996) (Section 5.2).

Throughout the paper $E(\cdot)$ denotes mathematical expectation of random variables.

2. Population point process

Our model's construction starts with the microscopic description of a population in which the adaptive traits of individuals influence their birth rate, the mutation process, their death rate, and how they interact with each other and their external environment. Thus, mathematically, the population can be viewed as a stochastic interacting individual system (cf. Durrett and Levin, 1994). The phenotype of each individual is described by a vector of trait values. The trait space $\mathscr{X}$ is assumed to be a subset of l-dimensional real vectors and thus describes l real-valued traits. In the trait space $\mathscr{X}$, the population is entirely characterized by a *counting measure*, that is, a mathematical counting device which keeps track of the number of individuals expressing different phenotypes. The population evolves according to a Markov process on the set of such counting measures on $\mathscr{X}$; the Markov property assumes that the dynamics of the population after time t depends on the past information only through the current state of the population (i.e. at time t). The *infinitesimal generator* describes the mean behavior of this Markov process; it captures the birth and death events that each individual experiences while interacting with other individuals.

2.1. Process construction

We consider a population in which individuals can give birth and die at rates that are influenced by the individual traits and by interactions with individuals carrying the same or different traits. These events occur randomly, in continuous time. Reproduction is almost faithful: there is some probability that a mutation causes an offspring's phenotype to differ from her progenitor's. Interactions translate into a dependency of the birth and death rates of any focal individual upon the number of interacting individuals.

The population is characterized at any time t by the finite counting measure

$$v_t = \sum_{i=1}^{I(t)} \delta_{x_t^i}, \tag{2.1}$$

where δ_x is the Dirac measure at x. The measure v_t describes the distribution of individuals over the trait space at time t, where $I(t)$ is the total number of individuals alive at time t, and $x_t^1, \ldots, x_t^{I(t)}$ denote the individuals' traits. The time process v_t evolves in the set of all finite counting measures. Note that the total mass of the measure v_t is equal to $I(t)$. Likewise, $v_t(\Gamma)$ represents the number of individuals with traits contained in any subset Γ of the trait space, and

$$\int_{\mathscr{X}} \varphi(x) v_t(\mathrm{d}x) = \sum_{i=1}^{I(t)} \varphi(x_t^i),$$

which means that the total "mass" of individuals, each of them being "weighted" by the "scale" φ, is computed by integrating φ with respect to v_t over the trait space.

The population dynamics are driven by a birth-mutation-death process defined as follows. Individual mortality and reproduction are influenced by interactions between individuals. For a population whose state is described by the counting measure $v = \sum_{i=1}^{I} \delta_{x^i}$, let us define $d(x, U * v(x))$ as the death rate of individuals with trait x, $b(x, V * v(x))$ as the birth rate of individuals with trait x, where U and V are the interaction kernels affecting mortality and reproduction, respectively. Here $*$ denotes the convolution operator, which means that U and V give the "weight" of each individual when interacting with a focal individual, as a function of how phenotypically different they are. For example, $U * v(x) = \sum_{i=1}^{I} U(x - x^i)$. Mutation-related parameters are expressed as functions of the individual trait values only (although there would be no formal difficulty to include a dependency on the population state, in order to obtain adaptive mutagenesis models): $\mu(x)$ is the probability that an offspring produced by an individual with trait x carries a mutated trait, $M(x, z)$ is the mutation step kernel of the offspring trait $x + z$ produced by individuals with trait x. Since the mutant trait belongs to $\mathscr{X}$, we assume $M(x, z) = 0$ if $x + z$ does not belong to $\mathscr{X}$.

Thus, the individual processes driving the population adaptive evolution develop through time as follows:

- At $t = 0$ the population is characterized by a (possibly random) counting measure v_0. This measure gives the ancestral state of the population. Whether the ancestral state is monomorphic or polymorphic will prove mathematically important later on.
- Each individual has two independent random exponentially distributed "clocks": a birth clock with parameter $b(x, V * v_t(x))$, and a death clock with parameter $d(x, U * v_t(x))$. Assuming exponential distributions allows to reset both clocks to 0 every time one of them rings. At any time t:
- If the death clock of an individual rings, this individual dies and disappears.
- If the birth clock of an individual with trait x rings, this individual produces an offspring. With probability $1 - \mu(x)$ the offspring carries the same trait x; with probability $\mu(x)$ the trait is mutated.
- If a mutation occurs, the mutated offspring instantly acquires a new trait $x + z$, picked randomly according to the mutation step measure $M(x, z)\,dz$.

If $v = \sum_{i=1}^{I} \delta_{x^i}$ represents the population state at a given time t, the infinitesimal dynamics of the population after t is described by the following operator on the set of real bounded functions ϕ (so-called infinitesimal generator):

$$
L\phi(v) = \sum_{i=1}^{I} [1 - \mu(x^i)] b(x^i, V * v(x^i)) [\phi(v + \delta_{x^i}) - \phi(v)]
$$
$$
+ \sum_{i=1}^{I} \mu(x^i) b(x^i, V * v(x^i))
$$
$$
\times \int_{\mathbb{R}^l} [\phi(v + \delta_{x^i + z}) - \phi(v)] M(x^i, z)\,dz
$$
$$
+ \sum_{i=1}^{I} d(x^i, U * v(x^i)) [\phi(v - \delta_{x^i}) - \phi(v)]. \tag{2.2}
$$

The first term of (2.2) captures the effect on the population of birth without mutation; the second term, that of birth with mutation; and the last term, that of death. The density dependence of vital rates makes all terms nonlinear.

At this stage, a first mathematical step has to be taken: the formal construction of the process is required to justify the existence of a Markov process admitting L as infinitesimal generator. There is a threefold biological payoff to such a mathematical endeavor: (1) providing a rigorous and efficient algorithm for numerical simulations (given hereafter); (2) laying the mathematical basis to derive the moment equations of the process (Section 3); and (3) establishing a general method that will be used to derive macroscopic models (Sections 4 and 5).

We make the biologically natural assumption that all parameters, as functions of traits, remain bounded, except for the death rate. Specifically, we assume that for any population state $v = \sum_{i=1}^{I} \delta_{x^i}$, the birth rate $b(x, V * v(x))$ is

upper bounded by a constant $\bar{b}$, that the interaction kernels U and V are upper bounded by constants $\bar{U}$ and $\bar{V}$, and that there exists a constant $\bar{d}$ such that $d(x, U * v(x)) \leqslant (1 + I)\bar{d}$. The latter assumption means that the density dependence on mortality is "linear or less than linear". Lastly, we assume that there exist a constant C and a probability density $\bar{M}$ such that for any trait x, $M(x, z) \leqslant C\bar{M}(z)$. This is implied in particular if the mutation step distribution varies smoothly over a bounded trait space. These assumptions ensure that there exists a constant $\bar{C}$ such that for any population state described by the counting measure $v = \sum_{i=1}^{I} \delta_{x^i}$, the total event rate, i.e. the sum of all event rates, is bounded by $\bar{C}I(I + 1)$. Indeed, without density dependence, the per capita event rate should be upper bounded by $\bar{C}$, making the total event rate upper bounded by $\bar{C}I$ (since I is the size of the population); the influence of density dependence appears through the multiplicative term $I + 1$.

Let us now give an algorithmic construction of the population process $(v_t)_{t \geqslant 0}$. At any time t, we must describe the size of the population, and the trait vector of all individuals alive at that time. At time $t = 0$, the initial population state v_0 contains $I(0)$ individuals. The vector of random variables $\mathbf{X}_0 = (X_0^i)_{1 \leqslant i \leqslant I(0)}$ denotes the corresponding trait values. More generally the vector of traits of all individuals alive at time t is denoted by $\mathbf{X}_t$. We introduce the following sequences of independent random variables, which will drive the algorithm. First, the values of a sequence of random variables $(W_k)_{k \in \mathbb{N}^*}$ with uniform law on $[0, 1]$ will be used to select the type of birth or death events. Second, the times at which events may be realized will be described using a sequence of random variables $(\tau_k)_{k \in \mathbb{N}}$ with exponential law with parameter $\bar{C}$ (hence $E(\tau_k) = 1/\bar{C}$). Third, the mutation steps will be driven by a sequence of random variables $(Z_k)_{k \in \mathbb{N}}$ with law $\bar{M}(z)\,dz$.

We set $T_0 = 0$ and construct the process inductively over successive event steps $k \geqslant 1$ as follows. At step $k - 1$, the number of individuals is I_{k-1}, and the trait vector of these individuals is $\mathbf{X}_{T_{k-1}}$. We define $T_k = T_{k-1} + \tau_k / I_{k-1}(I_{k-1} + 1)$. The term $\tau_k / I_{k-1}(I_{k-1} + 1)$ represents the minimal amount of time between two events (birth or death) in a population of size I_{k-1} (this is because the total event rate is bounded by $\bar{C}I_{k-1}(I_{k-1} + 1)$). At time T_k, one chooses an individual $i_k = i$ uniformly at random among the I_{k-1} alive in the time interval $[T_{k-1}, T_k]$; this individual's trait is $X_{T_{k-1}}^i$. (If $I_{k-1} = 0$ then $v_t = 0$ for all $t \geqslant T_{k-1}$.) Because $\bar{C}(I_{k-1} + 1)$ gives an upper bound on the total event rate for each individual, one can decide of the fate of that individual by making use of the following rules:

- If
$$
0 \leqslant W_k \leqslant \frac{d(X_{T_{k-1}}^i, \sum_{j=1}^{I_{k-1}} U(X_{T_{k-1}}^i - X_{T_{k-1}}^j))}{\bar{C}(I_{k-1} + 1)}
$$
$$
= W_1^i(\mathbf{X}_{T_{k-1}}),
$$
then the chosen individual dies, and $I_k = I_{k-1} - 1$.

- If
$$W_1^i(\mathbf{X}_{T_{k-1}}) < W_k \leqslant W_2^i(\mathbf{X}_{T_{k-1}}),$$
where
$$W_2^i(\mathbf{X}_{T_{k-1}}) = W_1^i(\mathbf{X}_{T_{k-1}})$$
$$+ \frac{[1-\mu(X_{T_{k-1}}^i)]b(X_{T_{k-1}}^i, \sum_{j=1}^{I_{k-1}} V(X_{T_{k-1}}^i - X_{T_{k-1}}^j))}{\bar{C}(I_{k-1}+1)},$$
then the chosen individual gives birth to an offspring with the same trait, and $I_k = I_{k-1} + 1$.
- If
$$W_2^i(\mathbf{X}_{T_{k-1}}) < W_k \leqslant W_3^i(\mathbf{X}_{T_{k-1}}, Z_k),$$
where
$$W_3^i(\mathbf{X}_{T_{k-1}}, Z_k) = W_2^i(\mathbf{X}_{T_{k-1}})$$
$$+ \frac{\mu(X_{T_{k-1}}^i)b(X_{T_{k-1}}^i, \sum_{j=1}^{I_{k-1}} V(X_{T_{k-1}}^i - X_{T_{k-1}}^j))M(X_{T_{k-1}}^i, Z_k)}{\bar{C}\bar{M}(Z_k)(I_{k-1}+1)},$$
then the chosen individual gives birth to a mutant offspring with trait $X_{T_{k-1}}^i + Z_k$, and $I_k = I_{k-1} + 1$.
- If $W_k > W_3^i(\mathbf{X}_{T_{k-1}}, Z_k)$, nothing happens, and $I_k = I_{k-1}$.

Mathematically, it is necessary to justify that the individual-based process $(v_t)_{t \geqslant 0}$ is well defined on the whole time interval $[0, \infty)$; otherwise, the sequence T_k might converge to a finite accumulation point at which the population process would explode in finite time. To this end, one can use the fact that the birth rate remains bounded. This allows one to compare the sequence of jump times of the process v_t with that of a continuous-time Galton–Watson process. Since the latter converges to infinity, the same holds for the process v_t, which provides the necessary justification.

The process $(v_t)_{t \geqslant 0}$ is Markovian with generator L defined by (2.2). From this follows the classical probabilistic decomposition of v_t as a solution of an integro-differential equation governed by L and perturbed by a martingale process (Ethier and Kurtz, 1986). In particular, (2.2) entails that for any function φ, bounded and measurable on $\mathcal{X}$

$$\int_{\mathcal{X}} \varphi(x) v_t(\mathrm{d}x) = m_t(\varphi) + \int_{\mathcal{X}} \varphi(x) v_0(\mathrm{d}x)$$
$$+ \int_0^t \int_{\mathcal{X}} ((1-\mu(x))b(x, V * v_s(x))$$
$$- d(x, U * v_s(x)))\varphi(x) v_s(\mathrm{d}x) \, \mathrm{d}s$$
$$+ \int_0^t \int_{\mathcal{X}} \mu(x)b(x, V * v_s(x))$$
$$\times \left(\int_{\mathbb{R}^l} \varphi(x+z) M(x, z) \, \mathrm{d}z \right) v_s(\mathrm{d}x) \, \mathrm{d}s,$$
(2.3)

where the time process $m_t(\varphi)$ is a martingale (see appendix) which describes the random fluctuations of the Markov process v. For each t, the random variable $m_t(\varphi)$ has mean

zero and variance equal to:

$$\int_0^t \int_{\mathcal{X}} \left\{ \varphi^2(x)((1-\mu(x))b(x, V * v_s(x)) + d(x, U * v_s(x))) \right.$$
$$\left. + \mu(x)b(x, V * v_s(x)) \int_{\mathbb{R}^l} \varphi^2(x+z)M(x, z) \, \mathrm{d}z \right\} v_s(\mathrm{d}x) \, \mathrm{d}s.$$
(2.4)

This decomposition, developed in the appendix, will be the key to our approximation method. Eq. (2.3) can be understood as providing a general model for the 'phenotypic mass' of the population that can be associated with any given 'scale' φ, $\varphi(x)$ being the 'weight' of trait x in the phenotypic space $\mathcal{X}$.

2.2. Examples and simulations

A simple example assumes logistic density dependence mediated by the death rate only:
$$b(x, \zeta) = b(x), \quad d(x, \zeta) = d(x) + \alpha(x)\zeta,$$
(2.5)
where b, d and α are bounded functions. Then
$$d(x, U * v(x)) = d(x) + \alpha(x) \int_{\bar{\mathcal{X}}} U(x-y)v(\mathrm{d}y).$$
(2.6)

Note that, in the case where $\mu \equiv 1$, the individual-based model can also be interpreted as a model of "spatially structured population", where the trait is viewed as a spatial location and mutation is analogous to dispersal. This is the type of models studied by Bolker and Pacala (1997, 1999), Law et al. (2003) and Fournier and Méléard (2004). The well-known case $U \equiv 1$ corresponds to density dependence involving the total population size, and will be termed "mean field".

Kisdi (1999) has considered a version of (2.5)–(2.6) for which

$$\mathcal{X} = [0, 4], \quad d(x) = 0, \quad \alpha(x) = 1, \quad \mu(x) = \mu,$$

$$b(x) = 4 - x, \quad U(x-y) = \frac{2}{K}\left(1 - \frac{1}{1 + 1.2\exp(-4(x-y))}\right)$$
(2.7)

and $M(x, z)\,\mathrm{d}z$ is a centered Gaussian law with variance σ^2 conditioned to the fact that the mutant stays in $[0, 4]$. In this model, the trait x can be interpreted as body size; (2.7) means that body size has no effect on the mutation rate, influences the birth rate negatively, and creates asymmetrical competition reflected in the sigmoid shape of U (being larger is competitively advantageous). Thus, body size x is subject to (frequency-independent) stabilizing selection and mediates frequency- and density-dependent selection through intraspecific competition. As we shall see in Section 4, the constant K scaling the strength of competition also scales population size. Following Metz et al. (1996), we refer to K as the "system size".

We have performed simulations of this model by using the algorithm described in the previous section. The numerical results reported here (Figs. 1 and 2) are intended
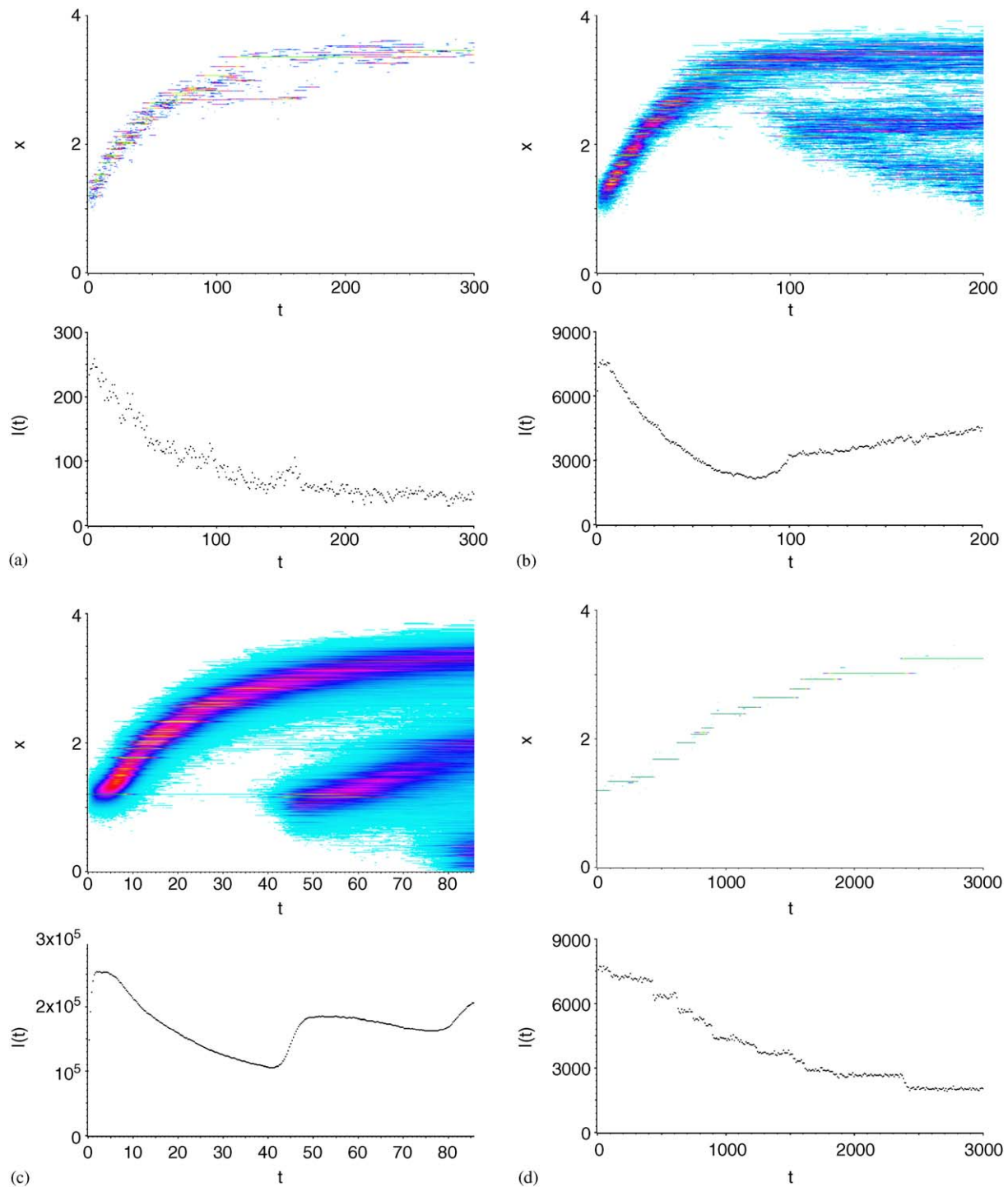
Fig. 1. Numerical simulations of trait distributions (upper panels, darker is higher frequency) and population size (lower panels). The initial population is monomorphic with trait value 1.2 and contains K individuals. (a–c) Qualitative effect of increasing system size (measured by parameter K). (d) Large system size and very small mutation probability (μ). Running time was chosen so that similar ranges of trait values were spanned by all simulated evolutionary trajectories: (a) $\mu = 0.03$, $K = 100$, $\sigma = 0.1$; (b) $\mu = 0.03$, $K = 3000$, $\sigma = 0.1$; (c) $\mu = 0.03$, $K = 100\,000$, $\sigma = 0.1$; (d) $\mu = 0.00001$, $K = 3000$, $\sigma = 0.1$.

to show that a wide variety of qualitative behaviors obtains for different combinations of the mutation parameters σ, μ and system size K. In each Figs. 1(a)–(d) and 2(a)–(d), the upper panel displays the distribution of trait values in the population at any time and the lower panel displays the dynamics of the total population size, $I(t)$.

These simulations hint at the different mathematical approximations that we establish in Sections 4 and 5. Fig. 1(a)–(c) represent the individual-based process $(v_t, I(t))$ with fixed μ and σ, and with an increasing system size K. As K increases, the fluctuations of the population size $I(t)$ (lower panels) are strongly reduced, which suggests the existence of a
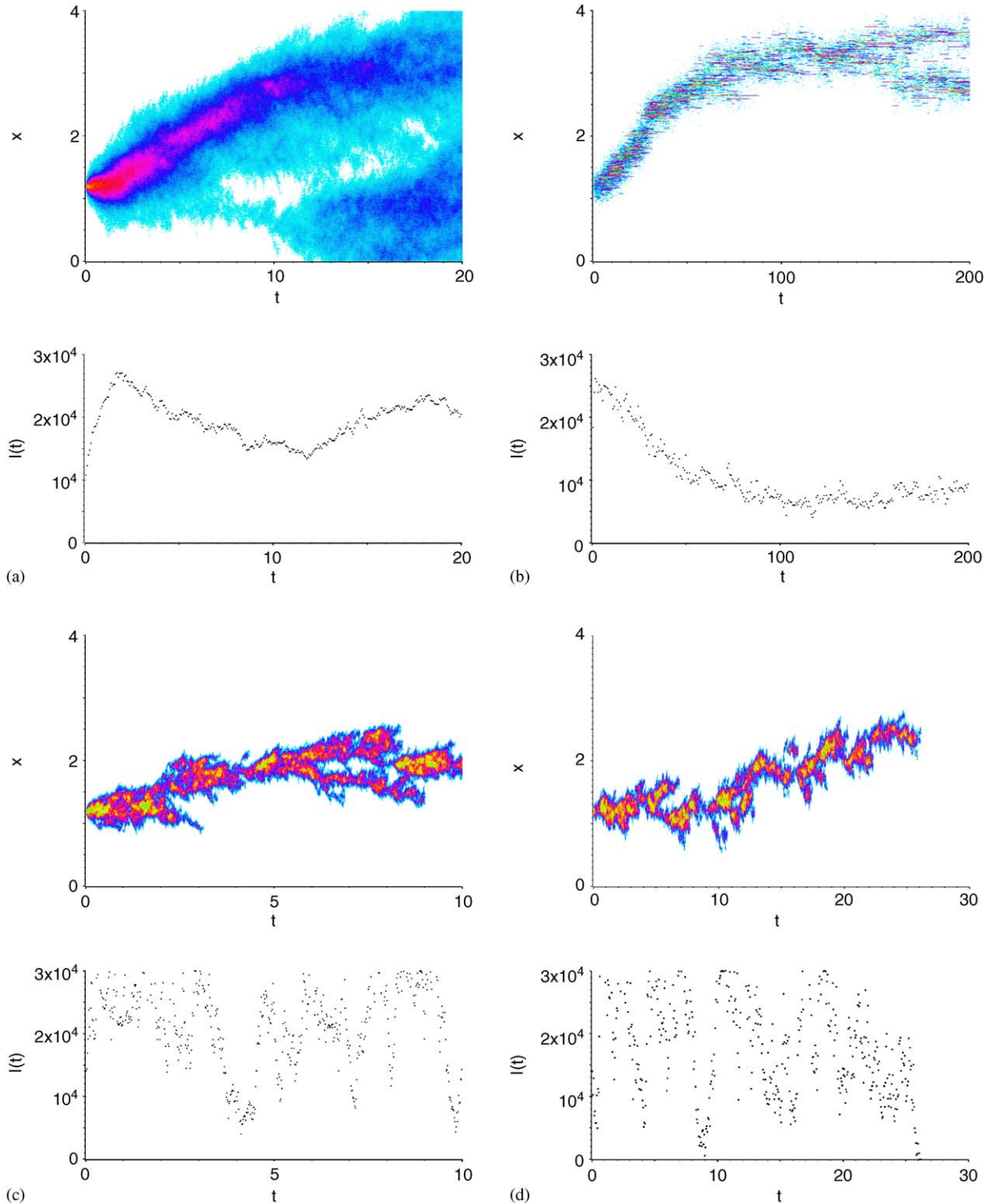
Fig. 2. Numerical simulations of trait distribution (upper panels, darker is higher frequency) and population size (lower panels) for accelerated birth and death and concurrently increased system size. Parameter η (between 0 and 1) relates the acceleration of demographic turnover and the increase of system size. (a) Rescaling mutation step. (b) Rescaling mutation probability. (c–d) Rescaling mutation step in the limit case $\eta = 1$; two samples for the same population. The initial population is monomorphic with trait value 1.2 and contains K individuals: (a) $\mu = 0.3$, $K = 10\,000$, $\sigma = 0.3/K^{\eta/2}$, $\eta = 0.5$; (b) $\mu = 0.1/K^\eta$, $K = 10\,000$, $\sigma = 0.1$, $\eta = 0.5$; (c) $\mu = 0.3$, $K = 10\,000$, $\sigma = 0.3/K^{\eta/2}$, $\eta = 1$; (d) $\mu = 0.3$, $K = 10\,000$, $\sigma = 0.3/K^{\eta/2}$, $\eta = 1$.

deterministic limit; and the support of the measure ν_t (upper panels) spreads over the trait space, which suggests the existence of a density for the limit measure (see Section 4.1). Fig. 1(d) illustrates the dynamics of the population on a long

timescale, when the mutation probability μ is very small. A qualitatively different phenomenon appears: the population remains almost monomorphic and the trait evolves according to a jump process, obtained in Section 5.

In Fig. 2, the underlying model involves accelerating the birth and death processes along with increasing system size, as if the population were made up of a larger number of smaller individuals, reproducing and dying at higher rates (see Section 4.2). Specifically, we take

$$b(x, \zeta) = K^\eta + b(x) \quad \text{and} \quad d(x, \zeta) = K^\eta + d(x) + \zeta,$$

where $b(x)$, $d(x)$, $\mu(x)$, $U(x)$ and $M(x, z)$ are defined as in (2.7). Note that the "demographic timescale" of population growth, that occurs at rate $b(x, \zeta) - d(x, \zeta)$, is unchanged.

There is a noticeable qualitative difference between Figs. 2(a)–(b), where $\eta = 1/2$, and Figs. 2(c)–(d), where $\eta = 1$. In the latter, we observe strong fluctuations in the population size, early extinction happened in many simulations (Fig. 2(d)) and the evolutionary pattern is finely branched, revealing that the stochasticity of birth and death persists and generates a new form of stochasticity in the large population limit (see Sections 4.2.1 and 4.2.2).

3. Moment equations

Moment equations have been introduced in theoretical population biology by Bolker and Pacala (1997, 1999) and Dieckmann and Law (2000) (referred hereafter as BPDL), following on from the seminal work of Matsuda et al. (1992), as handy analytical models for spatially extended populations. A similar approach has been proposed independently by McKane and Newman (2004) to model population dynamics when individual stochastic processes operate in spatially structured habitats. Hereafter, we use the analogy between population processes defined on trait space versus physical space to construct the moment equations of the population evolutionary dynamics. The "philosophy" of moment equations is germane to the principle of Monte-Carlo methods: computing the mean path of the point process from a large number of independent realizations. (The orthogonal stance, as we shall see in Section 4, is to model the behavior of a single trajectory while making the initial number of individuals become large).

Let us define the deterministic measure $E(v)$ associated with a random measure v by $\int_{\mathscr{X}} \varphi(x) E(v)(dx) = E(\int_{\mathscr{X}} \varphi(x) v(dx))$. Taking expectation in (2.3) and using $E(m_t(\varphi)) = 0$, one can obtain an equation for $\int_{\mathscr{X}} \varphi(x) E(v)(dx)$ involving the expectations of integrals with respect to $v(dx)$ or $v(dx)v(dy)$. This is a complicated equation involving an unresolved hierarchy of nonlinear terms. Writing an equation for $E(v(dx)v(dy))$ is feasible but yields integrals with respect to $v(dx)v(dy)v(dz)$, and so on. Whether this approach in general may eventually help describe the population dynamics in the trait space is still unclear.

Let us consider the case of logistic density dependence (see Section 2.2) where $d(x, \zeta) = d(x) + \alpha(x)\zeta$, $b(x, \zeta) = b(x)$ and $\mu(x) = 1$. Taking expectations in (2.3) with $\varphi \equiv 1$

yields:

$$N(t) = N(0) + \int_0^t \left\{ E\left(\int_{\mathscr{X}} [b(x) - d(x)] v_s(dx) \right) \right.$$
$$\left. - E\left(\int_{\mathscr{X} \times \mathscr{X}} \alpha(x) U(x - y) v_s(dx) v_s(dy) \right) \right\} ds, \quad (3.1)$$

where $N(t) = E(I(t))$ is the "mean" population size at time t. The specific case where b, d and α are independent of x, and U is symmetrical (cf. Law et al., 2003), corresponds to the BPDL model of spatial population dynamics. Eq. (3.1) then recasts into

$$\dot{N} = (b - d)N - \alpha \int_{\mathbb{R}^l} U(r) C_t(dr), \quad (3.2)$$

where C_t is defined at any time t as a "spatial covariance measure" (sensu BPDL) on $\mathbb{R}^l$, given by

$$\int_{\mathbb{R}^l} U(r) C_t(dr) = E\left(\int_{\mathscr{X} \times \mathscr{X}} U(x - y) v_t(dx) v_t(dy) \right). \quad (3.3)$$

A dynamic equation for this covariance measure then obtains by considering the quantities $\int_{\mathbb{R}^l} U(r) C_t(dr)$ as functions $\phi(v)$ in (2.2), but the equation involves moments of order 3, which prevents "closing" the model on lower-order variables. Even in the simplest mean-field case $U = 1$, we get

$$\dot{N}(t) = (b - d)N(t) - \alpha E\left(\int_{\mathscr{X} \times \mathscr{X}} v_t(dx) v_t(dy) \right). \quad (3.4)$$

Because of the expectation, the covariance term cannot be written as a function of the first-order moment $N(t)$, and, therefore, Eq. (3.4) does not simplify.

Even if there is no constructing a closed equation satisfied by $E(v)$, we are able to show, in the general case, the following important qualitative property: if the deterministic measure $E(v_0)$ of the initial population admits a density p_0 with respect to the Lebesgue measure, then for all $t \geqslant 0$, the deterministic measure $E(v_t)$ of the population has a probability density p_t. To see this, apply (2.3) to $\varphi = 1_A$ where A has zero Lebesgue measure. Taking expectations then yields $E(\int_{\mathscr{X}} \varphi(x) v_t(dx)) = \int_A E(v_t)(dx) = 0$, which gives the required result. As a consequence, the expectation of the total size of the population at time t is $N(t) = E(\int_{\mathscr{X}} v_t(dx)) = \int_{\mathscr{X}} p_t(x) dx$, and $p_t(x) dx / N(t)$ gives the probability of observing one individual at time t in a small ball centered in x with radius dx. In particular, this result implies that, when the initial trait distribution $E(v_0)$ has no singularity with respect to the Lebesgue measure, these singularities, such as Dirac masses, can only appear in the limit of infinite time.

This has biological implications on how one would analyze the process of population differentiation and phenotypic "packing" (Berstein et al., 1985). Such a population model should not be expected to converge in finite time towards neatly separated phenotypic peaks if the ancestral phenotypic distribution is even slightly spread out as opposed to being entirely concentrated on a set of distinct phenotypes. Thus, the biologically relevant

question that theory may address is not whether packing can arise from a continuous phenotypic distribution, but rather whether initial differentiation (ancestral phenotypic peaks) is amplified or buffered by the eco-evolutionary process.

4. Large-population renormalizations of the individual-based process

The moment equation approach outlined above is based on the idea of averaging a large number of independent realizations of the population process initiated with a finite number of individuals. If K denotes the initial number of individuals, taken as a measure of the "system size" *sensu* Metz et al. (1996), an alternative approach to deriving macroscopic models is to study the exact process by letting that system size become very large and making some appropriate renormalization. Several types of approximations can then be derived, depending on the renormalization of the process.

For any given system size K, we consider the set of parameters U_K, V_K, b_K, d_K, M_K, μ_K satisfying the previous hypotheses and being all continuous in their arguments. Let v_t^K be the counting measure of the population at time t. We define a renormalized population process $(X_t^K)_{t \geqslant 0}$ by

$$X_t^K = \frac{1}{K} v_t^K.$$

$(X_t^K)_{t \geqslant 0}$ is a measure-valued Markov process. As the system size K goes to infinity, the interaction kernels need be renormalized as $U_K(x) = U(x)/K$ and $V_K(x) = V(x)/K$. A biological interpretation of this renormalization is that larger systems are made up of smaller individuals, which may be a consequence of a fixed amount of available resources to be partitioned among individuals. Thus, the biomass of each individual scales as $1/K$, and the interaction kernels are renormalized in the same way, so that the interaction terms $U_K * v$ and $V_K * v$ that affect any focal individual stay of the same order of magnitude as the total biomass of the population.

Martingale theory allows one to describe the dynamics of X^K as the sum of a deterministic trajectory and a random fluctuation of zero expectation. The decomposition obtains by equations similar to (2.3) and (2.4), in which all coefficients depend on K, and the variance (2.4) of the martingale part is also divided by K. Deriving approximation limits for these two terms leads to the alternative choices of timescales that we present in this section. In particular, the nature (deterministic or stochastic) of the approximation can be determined by studying the variance of the random fluctuation term.

4.1. Large-population limit

Let us assume that, as K increases, the initial condition $X_0^K = (1/K)v_0^K$ converges to a finite deterministic measure

which has a density ξ_0 (when this does not hold, the following convergence results remain valid, but from a mathematical viewpoint the limit partial differential equations and stochastic partial differential equations have to be understood in a weak measure-valued sense). Moreover, we assume that $b_K = b$, $d_K = d$, $\mu_K = \mu$, $M_K = M$. Thus, the variance of the random fluctuation of X_t^K is of order $1/K$, and when the system size K becomes large, the random fluctuations vanish and the process $(X_t^K)_{t \geqslant 0}$ converges in law to a deterministic measure with density ξ_t satisfying the integro-differential equation with trait variable x and time variable t:

$$\partial_t \xi_t(x) = [(1 - \mu(x))b(x, V * \xi_t(x)) - d(x, U * \xi_t(x))]\xi_t(x)$$
$$+ \int_{\mathbb{R}^l} M(y, x - y)\mu(y)b(y, V * \xi_t(y))\xi_t(y)\, dy. \quad (4.1)$$

This result re-establishes Kimura's (1965) equation (see also Bürger, 2000 , p. 119, Eq. (1.3)) from microscopic individual processes, showing that the only biological assumption needed to scale up to macroscopic evolutionary dynamics is that of a large population. Importantly, Eq. (4.1) extends Kimura's original model to the case of frequency-dependent selection.

The convergence of X^K to the solution of (4.1) is illustrated by the simulations shown in Fig. 1(a)–(c). The proof of this result (adapted from Fournier and Méléard, 2004) strongly relies on arguments of tightness in finite measure spaces (Roelly, 1986). Desvillettes et al. (2004) suggest to refer to ξ_t as the population "number density"; then the quantity $n(t) = \int_{\mathscr{X}} \xi_t(x)\, dx$ can be interpreted as the "total population density" over the whole trait space. This means that if the population is initially seeded with K individuals, $Kn(t)$ approximates the number of individuals alive at time t, all the more closely as K is larger.

The case of logistic density-dependence with constant rates b, d, α leads to an interesting comparison with moment equations (cf. Section 3). Then (4.1) yields the following equation on $n(t)$:

$$\dot{n} = (b - d)n - \alpha \int_{\mathscr{X} \times \mathscr{X}} U(x - y)\xi_t(x)\, dx\xi_t(y)\, dy. \quad (4.2)$$

In the mean-field case $U \equiv 1$, the trait x becomes completely neutral, and the population dynamics are not influenced by the mutation distribution anymore—they are driven simply by the classical logistic equation of population growth:

$$\dot{n} = (b - d)n - \alpha n^2. \quad (4.3)$$

Comparing (4.3) with the first-moment equation (3.4) emphasizes the "decorrelative" effect of the large system size renormalization: in the moment equation model (3.4) that assumes finite population size, the correction term capturing the effect of correlations of population size across the trait space remains, even if one assumes $U \equiv 1$.

4.2. Large-population limit with accelerated births and deaths

We consider here an alternative limit obtained when large system size is combined with accelerated birth and death. This may be useful to investigate the qualitative differences of evolutionary dynamics across populations with allometric demographies: larger populations made up of smaller individuals. This leads in the limit to systems in which the organisms have short lives and reproduce fast while their colonies or populations grow or decline on a slow timescale. This applies typically to microorganisms, e.g. bacteria and plankton, including many pathogens within their hosts. Yoshida et al. (2003) have provided striking experimental evidence for rapid evolutionary changes taking place during long ecological cycles, and Thompson (1998) and Hairston et al. (2005) have emphasized the importance of convergent ecological and evolutionary time to understand temporal dynamics in ecological systems.

For mathematical simplicity, the trait space $\mathscr{X}$ is assumed here to be the whole $\mathbb{R}^l$. The boundedness assumptions on the rates d, b, and on the interaction kernel U (see Section 2) are maintained. We consider the acceleration of birth and death processes at a rate proportional to K^η while preserving the demographic balance; that is, the density-dependent birth and death rates scale with system size according to $b_K(x, \zeta) = K^\eta r(x) + b(x, \zeta)$ and $d_K(x, \zeta) = K^\eta r(x) + d(x, \zeta)$; hence $b_K(x, \zeta) - d_K(x, \zeta)$ is unchanged, equal to $b(x, \zeta) - d(x, \zeta)$. The allometric effect is parameterized by the positive and bounded function $r(x)$ and the constant η; $r(x)$ measures the contribution of the birth process to the phenotypic variability on the new timescale K^η. As before (cf. Section 4.1), the interaction kernels U and V are renormalized by K. Two interesting cases will be considered hereafter, in which the variance of the mutation effect $\mu_K M_K$ is of order $1/K^\eta$. That will ensure convergence of the deterministic part in (2.3). In the large-population renormalization (Section 4.1), the variance of fluctuations around the deterministic trajectory was of order $1/K$. Here, the variance of fluctuations is of order $K^\eta \times 1/K$, and hence stays finite provided that $\eta \in (0, 1]$, in which case tractable limits will ensue. If $\eta < 1$, the variance is zero and a deterministic model obtains. If $\eta = 1$, the variance does not vanish and the limit model is stochastic. These two cases are illustrated by the simulations shown in Fig. 2(a)–(d).

4.2.1. Accelerated mutations and small mutation steps
We consider here that the mutation probability is fixed ($\mu_K = \mu$), so that mutations are accelerated as a consequence of accelerating birth, while assuming infinitesimal steps: the mutation kernel $M_K(x, z)$ is the density of a random variable with mean zero and variance-covariance matrix $\Sigma(x)/K^\eta$ (where $\Sigma(x) = (\Sigma_{ij}(x))_{1 \leqslant i,j \leqslant l}$) (see the appendix for technical assumptions on Σ). For example, the mutation step density $M_K(x, z)$ is taken as the density

of a centered vector (dimension l) of independent Gaussian variables with mean 0 and variance $\sigma_K^2(x) = \sigma^2(x)/K^\eta$:

$$M_K(x, z) = \left(\frac{K^\eta}{2\pi\sigma^2(x)}\right)^{l/2} \exp[-K^\eta|z|^2/2\sigma^2(x)], \qquad (4.4)$$

where $\sigma^2(x)$ is positive and $\sigma\sqrt{r\mu}$ is assumed to be a Lipschitz function bounded over $\mathbb{R}^l$ and bounded away from 0. Thus, in larger systems (larger K), the phenotypic changes affecting mutants, as measured in the unchanged trait space, are smaller.

Let us assume that the initial condition $X_0^K = (1/K)v_0^K$ converges to a finite measure ξ_0. When $\eta < 1$, we can prove that the sequence of processes $(X^K)_{K \in \mathbb{N}^*}$ converges as K increases to a weak measure-valued solution of the deterministic partial differential equation

$$\partial_t \xi_t(x) = [b(x, V * \xi_t(x)) - d(x, U * \xi_t(x))]\xi_t(x)$$
$$+ \tfrac{1}{2}\Delta(\sigma^2 r\mu\xi_t)(x). \qquad (4.5)$$

This provides a new extension to frequency-dependent selection of the Fisher reaction–diffusion equation, which was known as an approximation of Kimura's equation for small mutation effects (Kimura, 1965). The evolutionary dynamics are monitored over the demographic timescale of the population, which can be thought of as the timescale over which 'typical' episodes of population growth or decline, as measured by $b(x, \zeta) - d(x, \zeta)$, take place. The 'typical' amount of population phenotypic change generated by mutation per unit time is $b_K\mu_K\sigma_K^2\xi_t = \mu(K^\eta r + b)(\sigma^2/K^\eta)\xi_t \approx \mu r\sigma^2\xi_t$. This rate indeed appears in the Laplacian diffusion term which corresponds to the Brownian approximation of the mutation process (Ewens, 2004).

When $\eta = 1$, the rescaling is similar to the one leading from a branching random walk to a superprocess (Dynkin, 1991) and an analogous argument gives rise to a (random) measure-valued process as macroscopic model. Indeed, the sequence of processes $(X^K)_{K \in \mathbb{N}^*}$ converges as K increases to a continuous process $(X_t)_{t \geqslant 0}$ where $X_0 = \xi_0$ and X_t is a finite measure which is formally a weak solution of the stochastic partial differential equation

$$\partial_t X_t(x) = [b(x, V * X_t(x)) - d(x, U * X_t(x))]X_t(x)$$
$$+ \tfrac{1}{2}\Delta(\sigma^2 r\mu X_t)(x) + \sqrt{2r(x)X_t(x)}\dot{W}. \qquad (4.6)$$

Here $\dot{W}$ is the so-called space-time white noise (Walsh, 1984). This term captures the effect of demographic stochasticity occurring in the 'super fast' birth-and-death process (i.e. with $\eta = 1$). The measure-valued process X is called *superprocess* and appears as a generalization of Etheridge's (2004) superprocess model for spatially structured populations. Here again, the Laplacian diffusion term corresponds to the Brownian approximation of the mutation process (Ewens, 2004). This specific approximation is recovered because of the appropriate time rescaling when making mutations smaller and more frequent.

The proof of the first convergences makes use of techniques very similar to those used in Section 4.1. The

proof of the second statement requires additional results that are specific to superprocesses (Evans and Perkins, 1994) in order to establish uniqueness of the limit process. Both proofs are expounded in the appendix, for the general case of the mutation kernel with covariance matrix $\Sigma(x)/K^\eta$, and the corresponding general results can be stated as follows. When $\eta < 1$, the process X^K converges to the solution of the following deterministic reaction–diffusion equation:

$$\partial_t \xi_t(x) = [b(x, V * \xi_t(x)) - d(x, U * \xi_t(x))]\xi_t(x)$$
$$+ \frac{1}{2} \sum_{1 \leqslant i,j \leqslant l} \partial_{ij}^2 (r\mu\Sigma_{ij}\xi_t)(x), \qquad (4.7)$$

where $\partial_{ij}^2 f$ denotes the second-order partial derivative of f with respect to x_i and x_j ($x = (x_1, \ldots, x_d)$). When $\eta = 1$, the limit is the following stochastic partial differential equation:

$$\partial_t X_t(x) = [b(x, V * X_t(x)) - d(x, U * X_t(x))]X_t(x)$$
$$+ \frac{1}{2} \sum_{1 \leqslant i,j \leqslant l} \partial_{ij}^2 (r\mu\Sigma_{ij}X_t)(x) + \sqrt{2r(x)X_t(x)}\dot{W}.$$
$$(4.8)$$

The simulations displayed in Fig. 2(c)–(d), compared with Fig. 2(a), give a flavor of the complexity of the dynamics that the individual process can generate under the conditions leading to the superprocess models (4.6) or (4.8). Two distinctive features are the fine branching structure of the evolutionary pattern, and the wide fluctuations in population size that occur in parallel. In fact, replicated simulations show that the system often undergo rapid extinction (Fig. 2(d)). The results of our simulations suggest that the super fast timescale involved here is a general cause for complex population dynamics on the demographic timescale and for extinction driven by the joint ecological and evolutionary processes, which is germane to the phenomenon of *evolutionary suicide* described in adaptive dynamics theory (Dercole et al., 2002; Ferrière et al., 2002; Dieckmann and Ferrière, 2004); this may be largely independent of the ecological details of the system. This conjecture is supported by the mathematical proof of almost sure extinction obtained by Etheridge (2004) in her study of a related superprocess describing spatial population dynamics.

4.2.2. Rare mutations

Here, the mutation step density M is kept constant, while the mutation rate is decelerated proportionally to $1/K^\eta$: $\mu_K = \mu/K^\eta$. Thus only births without mutation are accelerated. As in Section 4.2.1, the macroscopic model keeps track of the phenotypic distribution over the population demographic timescale, which coincides here with the mutation timescale. Again, the limit model can be deterministic or stochastic, depending on whether the allometric parameter η is less than 1 or equal to 1, respectively.

Let us assume that the initial condition $X_0^K = (1/K)v_0^K$ converges to the finite measure with density ξ_0. When $\eta < 1$, the sequence of processes $(X^K)_{K \in \mathbb{N}^*}$ converges, as K increases, to a measure-valued process with density ξ_t solution of the following deterministic nonlinear integro-differential equation:

$$\partial_t \xi_t(x) = [b(x, V * \xi_t(x)) - d(x, U * \xi_t(x))]\xi_t(x)$$
$$+ \int_{\mathbb{R}^l} M(y, x - y)\mu(y)r(y)\xi_t(y)\,\mathrm{d}y. \qquad (4.9)$$

This equation is similar to (4.1), where the allometric effect rate r appears in lieu of the birth rate b in the mutation term; this is because the per capita mutation rate is equal to $\mu_K b_K = (\mu/K^\eta)(K^\eta r + b) \approx \mu r$, while the mutation step density is kept constant. Simulations of the individual process under the conditions leading to this model are shown in Fig. 2(b). When $\eta = 1$, we show, by arguments similar to those involved in Section 4.2.1, that the limit model is a measure-valued (random) process, which obtains as weak solution of the stochastic integro-differential equation

$$\partial_t X_t(x) = [b(x, V * X_t(x)) - d(x, U * X_t(x))]X_t(x)$$
$$+ \int_{\mathbb{R}^l} M(y, x - y)\mu(y)r(y)X_t(\mathrm{d}y)$$
$$+ \sqrt{2r(x)X_t(x)}\dot{W}, \qquad (4.10)$$

where $\dot{W}$ is a space-time white noise.

Eqs. (4.9) and (4.10) are obtained in a limit of rare mutations, with accelerated birth and death, on a timescale such that the order of magnitude of the *individual* mutation rate remains constant. In the next section, we study the behavior of the population process in a limit of rare mutations and accelerated birth and death, on an even longer timescale, such that the order of magnitude of the *total* mutation rate in the population remains constant. This assumption of extremely rare mutations leads to a different class of stochastic models which will provide a description of the population dynamics on a slow evolutionary timescale.

5. Renormalization of the monomorphic process and adaptive dynamics

Metz et al. (1996) have introduced an asymptotic of rare mutations to approximate the process of adaptive evolution with a monomorphic jump process. The jump process describes evolutionary trajectories as *trait substitution sequences* developing over the timescale of mutations. Dieckmann and Law (1996) have further introduced ingenious heuristics to achieve a deterministic approximation for the jump process, solution to the so-called *canonical equation of adaptive dynamics*. Metz et al.'s notion of trait substitution sequences and Dieckmann and Law's canonical equation form the core of the current theory of *adaptive dynamics*. In this section, we present a rigorous derivation of the stochastic trait substitution

sequence jump process from the individual-based model initiated with a monomorphic ancestral condition. Our derivation emphasizes how the mutation scaling should compare to the system size (K) in order to obtain the correct timescale separation between mutant invasion events (taking place on a short timescale) and mutation occurrences (defining the evolutionary timescale). Next we recover a generalized canonical equation as an approximation of the jump process in an asymptotic of small mutation steps. We also propose a diffusion approximation of the jump process which allows one to study the timescale on which a change of basin of attraction for an evolutionary trajectory can occur, providing insights into patterns of macroevolutionary change (for related theoretical considerations, see Rand and Wilson, 1993).

5.1. Jump process construction from IBM

The mathematically rigorous construction of the jump process from the individual-based model requires that we first study the behavior of a monomorphic population in the absence of mutation, and next the behavior of a dimorphic population, involving competition, after a mutation has occurred. In the limit of large system size $(K \to \infty)$ without mutation $(\mu \equiv 0)$, with only trait x present at time $t = 0$, we have $X_0^K = n_0^K(x)\delta_x$ and $X_t^K = n_t^K(x)\delta_x$ for any time t. Using the same scaling parameters as in Section 4.1 ($U_K = U/K$, $V_K = V/K$, b_K and d_K independent of K), the convergence result stated therein tells us that $n_t^K(x)$ approaches $n_t(x)$ when K becomes large, and Eq. (4.1) (in its weak form) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t} n_t(x) = \rho_1(x, n_t(x))n_t(x) \qquad (5.1)$$

where $\rho_1(x, n_t(x)) = b(x, V(0)n_t(x)) - d(x, U(0)n_t(x))$. We will assume that $\rho_1(x, 0) > 0$, that $\rho_1(x, n) \to -\infty$ when $n \to +\infty$, and that, for any trait x, this differential equation possesses a *unique* positive equilibrium $\bar{n}(x)$, necessarily satisfying $b(x, V(0)\bar{n}(x)) = d(x, U(0)\bar{n}(x))$. Then, it takes only elementary calculus to prove that any solution to (5.1) with positive initial condition converges to $\bar{n}(x)$. In the case of linear logistic density-dependence introduced in Section 2.2 ($b(x, \zeta) = b(x)$ and $d(x, \zeta) = d(x) + \alpha(x)\zeta$), the equilibrium monomorphic density $\bar{n}(x)$ is $(b(x) - d(x))/\alpha(x)U(0)$.

When the population is dimorphic with traits x and y, i.e. when $X_0^K = n_0^K(x)\delta_x + n_0^K(y)\delta_y$, we can define $n_t(x)$ and $n_t(y)$ for any t as before. Then $\xi_t = n_t(x)\delta_x + n_t(y)\delta_y$ satisfies Eq. (4.1), which recast into the following system of coupled ordinary differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_t(x) = \rho_2(x, y, n_t(x), n_t(y))n_t(x),$$
$$\frac{\mathrm{d}}{\mathrm{d}t} n_t(y) = \rho_2(y, x, n_t(x), n_t(y))n_t(y), \qquad (5.2)$$

where $\rho_2(x, y, n, n') = b(x, V(0)n + V(x-y)n') - d(x, U(0)n + U(x-y)n')$. Note that $\rho_2(x, y, n, 0) = \rho_1(x, n)$. The system

(5.2) possesses two (non-trivial) equilibria on the boundary of $\mathbb{R}_+ \times \mathbb{R}_+$, $(\bar{n}(x), 0)$ and $(0, \bar{n}(y))$, which must be stable in the horizontal and vertical direction, respectively. We then state as a rule that "y invades x" if the equilibrium $(\bar{n}(x), 0)$ of (5.2) is unstable in the vertical direction; this can be shown to occur if $\rho_2(y, x, 0, \bar{n}(x)) > 0$ (Ferrière and Gatto, 1995; Geritz et al., 2002; Rinaldi and Scheffer, 2000). We further say that "invasion of x by y implies fixation" if $\rho_2(y, x, 0, \bar{n}(x)) > 0$ entails that all orbits of the dynamical system (5.2) issued from sufficiently small perturbations of the equilibrium $(\bar{n}(x), 0)$ in the positive orthant converge to $(0, \bar{n}(y))$. Our construction requires that this property holds for almost any mutant trait y borne out from x. Geritz et al. (2002) and Geritz (2004) have actually proved that this is true for general models when the mutant trait is close to the resident and the resident is sufficiently far from special trait values corresponding to *branching points* or *extinction points* of the trait space $\mathcal{X}$. From a biological viewpoint, the quantity $\rho_2(y, x, 0, \bar{n}(x))$ is the *fitness* of mutant y in a resident population of trait x at equilibrium (Metz et al., 1992), which we will hereafter denote by $f(y, x)$ and refer to as the *fitness function*. Note that the fitness function f satisfies the usual property that $f(x, x) = 0$ for any trait value x.

The heuristics of trait substitution sequence models (Metz et al., 1996) assume that a monomorphic population reaches its ecological (deterministic) equilibrium before the first mutation occurs. As a mutant arises, it competes with the resident trait, and sufficient time is given to the ecological interaction for sorting out the winner before a new mutant appears. In the simplest case, only one trait survives: either the mutant dies out (due to stochasticity or selective inferiority), or it replaces the resident trait (due to stochasticity or selective superiority). Therefore, on a long timescale, the evolutionary dynamics can be described as a succession of "mutation-invasion" events corresponding to jumps in the trait space.

These heuristics raise conflicting demands on the mutation rate that only a full mathematical treatment can resolve. First, mutation events should be rare enough so that the next mutant is unlikely to appear until the previous mutant's fixation or extinction is settled. Second, mutation events should be frequent enough, so that the next mutation is not delayed beyond the time when the resident population size is likely to have stochastically drifted away from its equilibrium. Large deviation theory (Dembo and Zeitouni, 1993) and results on Galton-Watson processes can be used to determine the correct mutation timescale for which both conditions are satisfied. The mathematical work is reported in Champagnat (2004a), and the end result of biological interest is that, if the mutation probability is taken as $\mu_K(x) = u_K\mu(x)$, where u_K converges to zero when K goes to infinity, then the mutation probability and the system size should scale according to

$$\mathrm{e}^{-CK} \ll u_K \ll \frac{1}{K \log K} \quad \text{for any } C > 0. \qquad (5.3)$$

Eq. (5.3) implies in particular that Ku_K tends to 0 as K tends to infinity; therefore, for each time t, the time change t/Ku_K represents a long time scaling. This slow timescale is that of the mutation process: the population size is of the order of K, and the per capita mutation rate is proportional to u_K, hence the population mutation rate is of the order of Ku_K. Conditions (5.3) may be rewritten as

$$\log K \ll \frac{t}{Ku_K} \ll e^{CK} \quad \text{for any } t, C > 0,$$

and obtains because $\log K$ is the typical time of growth and stabilization of a successful mutant, and $\exp(CK)$ is the typical time over which the resident population is likely to drift stochastically away from deterministic equilibrium (problem of exit from a domain, Freidlin and Wentzel, 1984).

Under assumption (5.3), the method developed in Champagnat (2004a) can be adapted to prove that, as the system size K becomes large, the process $X^K_{t/Ku_K} = (1/K)v^K_{t/Ku_K}$ converges, when the initial distribution is monomorphic with trait x, to the process $\bar{n}(Y_t)\delta_{Y_t}$ in which the population is at any time monomorphic. The time process involved, $(Y_t)_{t\geqslant0}$, is Markovian and satisfies $Y_0 = x$. This is a jump process with infinitesimal generator L given, for all bounded measurable function $\varphi: \mathscr{X} \to \mathbb{R}$, by

$$L\varphi(x) = \int_{\mathbb{R}^l} (\varphi(x + z) - \varphi(x))[g(x + z, x)]_+ M(x, z)\,dz \tag{5.4}$$

where

$$g(y, x) = \mu(x)b(x, V(0)\bar{n}(x))\bar{n}(x)\frac{f(y, x)}{b(y, V(y - x)\bar{n}(x))} \tag{5.5}$$

($[z]_+$ denotes the positive part: $[z]_+ = 0$ if $z \leqslant 0$; $[z]_+ = z$ if $z > 0$). The generator's form (5.4) means that the process Y_t jumps from state x with rate $R(x) = \int_{\mathbb{R}^l}[g(x + z, x)]_+ M(x, z)\,dz$ to the new state $x + z$, where z follows the law $[g(x + z, x)]_+ M(x, z)\,dz/R(x)$. The simulation shown in Fig. 1(d) illustrates this convergence result and the behavior of the process $(Y_t)_{t\geqslant0}$.

The expression for g given by Eq. (5.5) can be understood as follows. When the population is monomorphic with trait x, its density reaches a given neighborhood of its equilibrium $\bar{n}(x)$ in finite time, i.e. within an infinitesimal time with respect to the timescale of mutation; this can be shown by using results on stochastic comparison between (v^K_t) and logistic birth-and-death processes. Then, the population size being close to $\bar{n}(x)$, the population mutation rate is close to $u_K\mu(x)b(x, V(0)\bar{n}(x))K\bar{n}(x)$. Therefore, on the mutation timescale, the mutation rate is given by $\mu(x)b(x, V(0)\bar{n}(x))\bar{n}(x)$, which yields one part of (5.5). The other part deals with the invasion of a mutant trait y, which can be divided into three phases (Fig. 3), as is done classically by population geneticists dealing with selective sweeps (Kaplan et al., 1989; Durrett and Schweinsberg, 2004). Initially, there is only one mutant individual; the
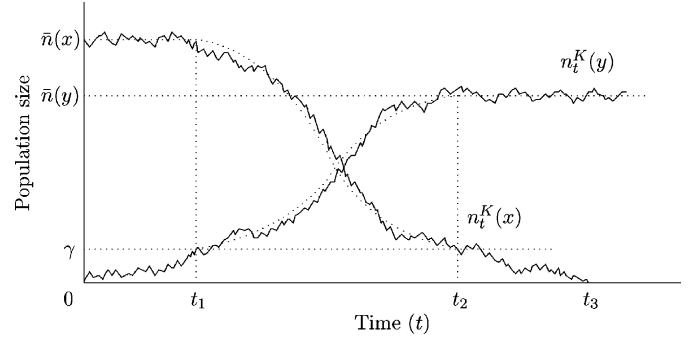

Fig. 3. The three phases of the invasion and fixation of a mutant trait y in a monomorphic population with trait x. Plain curves represent the resident and mutant densities $n^K_t(x)$ and $n^K_t(y)$, respectively. Dotted curves represent the solution of Eq. (5.2) with initial state $n_0(x) = \bar{n}(x)$ and $n_0(y) = \varepsilon$.

population it spawns may go extinct quickly even if its fitness $f(y, x)$ is positive, due to demographic stochasticity. To estimate the probability of such early extinction, we compare the number of mutant individuals, as long as it is small with respect to the resident population size, with Galton–Watson processes, with constant birth and death rates. This first phase ends when the mutant density reaches a fixed small level γ and corresponds in Fig. 3 to the time interval $[0, t_1]$. Here, the mutant birth rate is close to $b(y, V(y - x)\bar{n}(x))$, and the mutant death rate is close to $d(y, U(y - x)\bar{n}(x))$. The probability of survival of a Galton–Watson process with these parameters is given classically as $[f(y, x)]_+/b(y, V(y - x)\bar{n}(x))$, which yields the second part of (5.5). If invasion occurs, which is possible only if $f(y, x) > 0$, the resident and mutant densities get close to the solution of Eq. (5.2), represented by the dotted curves between t_1 and t_2 in Fig. 3; this is phase 2. Since we assume that "invasion implies fixation", the resident density converges to 0 and reaches level γ in bounded time. The third phase (between times t_2 and t_3 in Fig. 3) is analyzed by means of a comparison argument between the number of resident individuals and a Galton–Watson process similar to the previous one, which allows us to prove that the resident population goes extinct in infinitesimal time with respect to the mutation timescale. These arguments can be expounded formally by adapting the method of Champagnat (2004a). The times t_1 and $t_3 - t_2$ are of the order of $\log K$, while $t_2 - t_1$ only depends on γ.

The mathematical derivation of the jump process model (5.4) and (5.5) emphasizes that the rare-mutation and large-population limits must be taken simultaneously if one is to model evolutionary dynamics as a stochastic trait substitution sequence. On the other hand, the large population limit by itself can only yield the deterministic model (4.1) (generalized Kimura's equation). On the other hand, the dynamics of a finite population on the mutation timescale are trivial under the rare mutation scenario: the population goes immediately extinct almost surely on that timescale. This follows from the fact that the individual-based process v (cf. Section 2.1) that drives the dynamics of

the total (finite) population size $I(t)$ is stochastically bounded by a logistic birth-death process with birth and death rates of order $I(t)$ and $I(t)^2$, respectively; this process goes almost surely extinct in finite time. Therefore, it is always possible to pick u small enough so that extinction occurs instantaneously on the mutation timescale set by t/u.

In order to extend this result to the case where the coexistence of several traits is possible, i.e. when the "invasion-implies-fixation" assumption is relaxed, the probabilistic component of our approach can easily be generalized. The major difficulty is an analytical one: we would need to assume that, for any k and for any set of traits $\{x_1, \ldots, x_k\}$, the k-morphic system of coupled differential equations, that generalizes (5.2), admits a unique stable equilibrium towards which any solution with positive initial condition converges. Such an assumption is very restrictive and excludes the possibility of nonequilibrium attractors. As far as we know, no such analytic condition has been established even for restricted classes of ecological models.

5.2. Canonical equation and extensions

In order to perform the small mutation renormalization of the jump process constructed in the previous section, we introduce a (small) parameter $\varepsilon > 0$ by which the mutation step is multiplied, and we define a family of Markov jump processes $\{(x_t^\varepsilon)_{t \geqslant 0}\}_{\varepsilon > 0}$ with infinitesimal generator

$$L^\varepsilon \varphi(x) = \frac{1}{\varepsilon^2} \int_{\mathbb{R}^l} (\varphi(x + \varepsilon z) - \varphi(x))[g(x + \varepsilon z, x)]_+ M(x, z) \, \mathrm{d}z.$$
(5.6)

This model assumes the unusual time scaling by ε^{-2} which is required to avoid the process becoming constant in the limit of infinitesimal mutation steps ($\varepsilon \to 0$) as a consequence of $g(x, x) = 0$. The approach we used to study the renormalization of the individual point process in Sections 4.1 and 4.2 applies here again to prove (Champagnat et al., 2001; Champagnat, 2004b) convergence and recover the canonical equation of adaptive dynamics: under general regularity assumptions on g and M, when $\varepsilon \to 0$, the family of processes $\{(x_t^\varepsilon)_{t \geqslant 0}\}_{\varepsilon \geqslant 0}$ converges to the unique solution $(\tilde{x}_t)_{t \geqslant 0}$ to the (deterministic) ordinary differential equation

$$\frac{\mathrm{d}\tilde{x}}{\mathrm{d}t} = \int_{\mathbb{R}^l} z[z \cdot \nabla_1 g(\tilde{x}, \tilde{x})]_+ M(\tilde{x}, \mathrm{d}z),$$
(5.7)

where $\nabla_1 g$ denotes the gradient of $g(x, y)$ with respect to the first variable x.

In the case where $M(x, \cdot)$ is a symmetrical measure on $\mathbb{R}^l$ for any trait x in $\mathscr{X}$, (5.7) recast into the classical form of the "canonical equation" (Dieckmann and Law, 1996):

$$\frac{\mathrm{d}\tilde{x}}{\mathrm{d}t} = \frac{1}{2} \Sigma(\tilde{x}) \nabla_1 g(\tilde{x}, \tilde{x}),$$
(5.8)

where $\Sigma(x)$ denotes the variance-covariance matrix of the mutation kernel $M(x, \cdot)$. Champagnat (2004b) proved a

similar result for polymorphic populations in which a generalized "invasion-implies-fixation" principle holds, which is true away from branching points (or extinction points). Thus, the scope of the canonical equation appears to be as broad as the "invasion-implies-fixation" principle can be.

The equilibria of Eqs. (5.7) and (5.8) satisfy $\nabla_1 g(x, x) = 0$ and are classically called "evolutionary singularities".

5.3. Higher-order approximation

The large population assumption that goes along with the invasion-implies-fixation principle entails that adaptive change may only be directional—in the direction determined by $\nabla_1 g(x, x)$. However, in large yet finite populations, stochasticity may cause a mutant to invade even if its fitness is negative, so that adaptive evolution may proceed in any direction of the trait space. To account for this feature, we introduce a new model of adaptive dynamics in the form of a stochastic differential equation driving a diffusion process. The infinitesimal generator of this diffusion is a first-order approximation in ε of the generator (5.6) of the directional jump process, in the limit of small mutation jumps. Interestingly, the second-order differential operator obtained in this way possesses degenerate and discontinuous coefficients, rendering the classical theory of diffusion processes (see e.g. Karatzas and Shreve, 1988) non applicable. These bad regularity properties come from the non smooth function $[\cdot]_+$ appearing in the rescaled generator (5.6). The weak existence of solutions to this stochastic differential equation has been proved in Champagnat (2004b) under general regularity assumptions on g and M.

In the special case where the trait space $\mathscr{X}$ equals $\mathbb{R}$ and the mutation law $M(x, \cdot)$ is symmetrical, let $\sigma^2(x)$ be the variance of $M(x, \cdot)$, and $M_3(x) = \int_0^\infty z^3 M(x, z) \, \mathrm{d}z$; then the stochastic differential equation is

$$\mathrm{d}X_t^\varepsilon = [B_1(X_t^\varepsilon) + \varepsilon B_2(X_t^\varepsilon)] \, \mathrm{d}t + \sqrt{\varepsilon A(X_t^\varepsilon)} \, \mathrm{d}W_t,$$
(5.9)
where

$$B_1(x) = \frac{1}{2} \sigma^2(x) \partial_1 g(x, x),$$

$$B_2(x) = \frac{1}{2} M_3(x) \mathrm{sign}[\partial_1 g(x, x)] \partial_1^2 g(x, x),$$

$$A(x) = M_3(x) |\partial_1 g(x, x)|,$$

and W is a standard Brownian motion. Brownian motion theory then suggests to probe the behavior of this one-dimensional process when the canonical equation possesses multiple locally stable equilibria. The results of Champagnat (2003) can be applied here to show that when an ancestral population is surrounded by an attracting ESS (evolutionarily stable strategy) and an attracting branching point (cf. e.g. Geritz et al., 1998), this one-dimensional evolutionary process will almost surely home in at the ESS, rather than going through the branching point. This

mathematical result substantiates the numerical observation that branching is usually a very slow phenomenon; thus, when mutations steps are small, branching points are so difficult to reach as to leave time for the system to stabilize at an ESS if there is one within mutational reach. Therefore, to ensure that a population with unknown monomorphic ancestral state undergoes evolutionary branching, all of the attracting evolutionary singularities should be branching points.

In general, the issue of evolutionary dynamics drifting away from trajectories predicted by the canonical equation can be investigated by considering the asymptotic of the probability of 'rare events' for the sample paths of the diffusion. By 'rare events' we mean diffusion paths drifting far away from the canonical equation. The probability of such rare events is governed by a *large deviation principle* (Wentzel, 1976a,b; Freidlin and Wentzel, 1984): when ε goes to 0, the probability that the sample path of the diffusion process is close to a given rare path φ decreases exponentially to 0 with rate $I(\varphi)$, where the 'rate function' I can be expressed in terms of the parameters of the diffusion. The difficulty lies in the fact that the diffusion coefficient A is null at the evolutionary singularities and that the drift term B_2 is discontinuous at the same points, and the same problem arises for any value of the dimension l of the trait space. The large deviation principle has been obtained by Champagnat (2003) for any value of the trait space dimension, and implies in particular that the paths of X_t^ε converge in probability to the solution of the canonical equation (5.8) when ε goes to 0.

This result can be used to study the long-time behavior of the diffusion process when there are multiple attractive evolutionary singularities and the dimension of the trait space $\mathscr{X}$ is 2 or greater. Let us introduce the 'quasi-potential' $H(x, y)$ as the minimum of the rate function I over all the trajectories linking x to y. When ε is small, the most likely path followed by the diffusion when exiting the basin of attraction G of some evolutionary singularity x^*, is the one minimizing the rate function I over all the trajectories linking x^* to the boundary of G. Therefore, the time needed to exit G can be shown (Champagnat, 2003) to be of the order of or greater than $\exp[H/\varepsilon]$ for small ε, where H is the minimum of $H(x^*, y)$ over all the y in the boundary of G. Moreover, the exit event occurs with probability converging to 1 in any neighborhood of special points of the basin's boundary where the quasi-potential $H(x^*, \cdot)$ is minimum, so that one can predict the next basin of attraction visited by the diffusion. From a biological standpoint, this result provides a quantitative tool for analyzing the macroevolutionary notion of punctuated equilibria (Rand and Wilson, 1993; Stanley, 1979). The model generally predicts that the order of magnitude of the time spent in the neighborhood of evolutionary equilibria, between rapid evolutionary moves, is the exponential of the inverse of the mutation step standard deviation. This theory also predicts the sequence order of evolutionary singularities (equilibria or general attractors) that the evolutionary process is most likely to visit (Freidlin and Wentzel, 1984).

6. Discussion and conclusion

Martingale and large deviation theories provided us with the new probabilistic tools which were necessary for deriving and unifying models of evolutionary dynamics from stochastic nonlinear processes operating at the individual level. Different macroscopic models obtain depending on the renormalizations applied to the stochastic individual-based model. Hereafter we review the different models thus obtained and highlight how some of them relate to models previously known in quantitative genetics and evolutionary ecology. Then we review the biological insights that one can gain from the very construction of these models. Finally, we outline some promising directions for the analysis and further extensions of these models.

6.1. Unifying macroscopic models of evolutionary dynamics

A Monte-Carlo approach yields a hierarchy of equations describing the dynamics of the moments of population number in trait space. A similar approach has been taken heuristically by Bolker and Pacala (1997, 1999), Dieckmann and Law (2000) and McKane and Newman (2004) to construct macroscopic models of population dynamics in physical (geographic) space while accounting for individual dispersal. Our mathematical derivation sheds light on the structural features of the model which makes the problem of moment closure so challenging, especially the fact that, in general, the covariance measure may not have a density.

Alternatively to the Monte-Carlo approach, various macroscopic models obtain for different timescale separations, under the common assumption of the system size being large. The large-population limit by itself yields a generalization to frequency- and density-dependent selection scenarios of Kimura's (1965) continuum-of-alleles model (a nonlinear integro-differential equation). The assumption of small mutational effects, under which Kimura derived a diffusion approximation of his model, can be made while simultaneously accelerating the individual process of birth and death. This leads to separating the timescale of individual birth and death (assumed to be fast) from the timescale of population demography (over which significant population growth or decline occurs). This timescale separation may be most appropriate to study the interplay of ecological and evolutionary processes in microorganisms (Turchin, 2003; MacLean, 2005), including pathogens in which the concern of rapid evolution urges the need for appropriate modelling tools. The resulting model is a reaction–diffusion equation similar to Kimura's approximation and generalized to frequency- and density-dependent selection. Interestingly, the scaling exponent (η, between 0 and 1) which defines the proper acceleration of birth and death as the population

size is made larger, has no effect on the macroscopic dynamics, except when $\eta = 1$ which corresponds to maximum birth-death acceleration. In this case, the macroscopic model is structurally different, as it takes the form of a stochastic partial differential equation. Simulations of the individual process in this case (Fig. 2(c)–(d)) show that the evolutionary dynamics has a finely branched, fractal structure (we suspect that its Hausdorff dimension is between 1 and 2); the population displays wild fluctuations in total size, and faces a high risk of rapid extinction—a phenomenon akin to evolutionary suicide (Dieckmann and Ferrière, 2004).

The separation of the (fast) individual birth and death timescale and slow population demography can also be assumed under a rare- (rather than small-) mutation scenario. A model similar to Kimura's integro-differential equation obtains, but in which the 'loss' of individual births with any given trait due to mutation is not apparent. This reflects the fact that when the birth process is fast while the mutation probability becomes infinitesimal, the change in the frequency of any given trait due to mutation is caused predominantly by 'incoming' mutants born from progenitors carrying other trait values. As before, the macroscopic model does not depend upon the birth-death acceleration exponent η, except when birth and death are made maximally fast ($\eta = 1$), in which case the model is a stochastic nonlinear integro-differential equation.

The previous limits make the timescales of 'typical' mutation steps and 'typical' variation in population size coincide. An alternative approach is to assume that variation in population size occurs on a fast timescale compared to the timescale of mutation steps. This is the basis for modelling evolutionary dynamics as stochastic trait substitution sequences (Metz et al., 1996), which underlies the adaptive dynamics approach. We show that such trait substitution sequences are trajectories of a jump process which obtains under the assumption of an ancestral population being monomorphic. Our approach clearly isolates and solves the two key issues raised by the heuristics of the original derivation of adaptive dynamics models (Dieckmann and Law, 1996). One issue is underscored by the construction of the infinitesimal generator of the jump process. In Dieckmann–Law's heuristics, the population growth of a mutant is described by a Galton–Watson branching process, which appropriately assumes that the mutant population is finite; at the same time, however, the mutant branching process is parameterized by the density of the resident population which is assumed to be infinitely large. Resolving the tension between these conflicting assumptions requires that the whole system be regarded as finite, which then raises the issue that the resident population, being large yet finite, may stochastically drift away from the deterministic equilibrium predicted by the infinitely large population limit. The issue is taken care of by using large deviation theory to specify the appropriate mutation timescale over which this is unlikely to happen.

The second issue with the canonical equation heuristics (Dieckmann and Law, 1996) was the notion that the solution to the equation should describe the mean trait value in the population for small mutation steps. Our derivation shows that, in fact, the canonical equation drives the *exact* path of the jump process in the limit of *infinitesimal* mutation steps, which provides a mathematical justification for Dieckmann and Law's "mean path" interpretation. Our derivation further implies that the canonical equation orbits describe population change on a 'super long' timescale: firstly, time t is scaled as t/Ku_K where $u_K = o(1/K)$ is the order of magnitude of mutation probability (rare mutation assumption) and secondly, time is scaled by $1/\varepsilon^2$ where ε is the order of magnitude of mutation steps. This rescaling may be taken as a formal definition for the notion of a 'macroevolutionary' timescale.

6.2. Biological insights from the process of model construction

A general conclusion that emerges from this work is that how timescales of individual processes compare to each other can have a major impact on the structure of macroscopic models (integro-differential equations versus reaction–diffusion equations, deterministic versus stochastic), hence on the evolutionary dynamics predicted by these models. This was lucidly anticipated by Barton and Polechova (2005) in a commentary of the limitations of adaptive dynamics models, and is herein illustrated by simulations (Figs. 1 and 2), that hint at a whole array of strikingly different qualitative behaviors:

- In large populations, the process of diversification is faster (compare Figs. 1(a) and (b)) and can turn from gradual (Fig. 1(b)) to discontinuous and step-wise (Fig. 1(c)).
- For a given system size, extreme mutation rarity changes the prediction of diversifying dynamics in which both stabilizing selection and disruptive selection play strong roles (Fig. 1(b)), to a pattern dominated by stabilizing selection in which the population remains essentially monomorphic (Fig. 1(d)).
- When rare mutations occur in a large population of individuals reproducing and dying at high rates, phenotypic diversification occurs, although more slowly and to a lesser extent (Fig. 2(b)).
- In a large population of individuals reproducing and dying fast, mutations that are small rather than rare will cause a discontinuous pattern of diversification similar to Fig. 1(c), see Fig. 2(a). However, as mutation parameters μ and σ become smaller, the tendency for diversification is strongly limited, which demonstrates that the timescale of diversification is highly sensitive to the mutation pattern.
- However, maximal acceleration of birth and death associated with minimal mutation steps in large systems

generates yet another type of evolutionary dynamics, involving a high rate of diversification of the population into fine phenotypic clusters combined with a high rate of extinction of these clusters; the resulting evolutionary pattern has a remarkable fractal structure, as seen in Fig. 2(c). In particular, the pattern shown in that figure suggests that there might exist a general scaling relationship between the width of phenotypic clusters and their rate of extinction.

A simplistic aspect of our models resides in the absence of physiological, social or environmental structure: the individual life history is reduced to the simplest possible birth and death process. Although this simplifying assumption narrows down the scope of our theory, it has the merit of making the models thus obtained directly comparable to the classical models of population genetics, quantitative genetics and adaptive dynamics, which were derived under a similar assumption. Much generality, however, is kept in our treatment of density dependence and in the way adaptive traits influence individual processes.

When considering the acceleration of birth and death, the most general case we could handle involved rescaling the birth rate as

$$b_K = b\left(x, \frac{U}{K} * v\right) + r(x)K^\eta \qquad (6.1)$$

(and similar rescaling for the death rate). This relationship between birth and system size lends itself to interpretation within the context of allometries and life-history scaling (Calder, 1984; Charnov, 1993; Brown et al., 2004). Eq. (6.1) means that for large system size, density dependence is effectively 'felt' by any individual as a consequence of the large number of competitors (rather than the individual effect of each competitor), which suggest that large system size implicitly comes along with individuals being small. Let us introduce the notion of a taxonomic size' (or mass) m to express such a relationship more precisely. One can think of m as a macroscopic parameter (characteristic of a species or taxonomic group) which is fixed on the microevolutionary scale over which x can change adaptively. It is known (Damuth, 1987; Belgrano et al., 2002) that the system size, K, scales with the $-3/4$ power of (taxonomic) body mass, m. Thus, across the range of large system sizes, i.e., small organisms, the birth rate given by Eq. (6.1) is approximately equal to $r(x)m^{-3\eta/4}$. Thus, this model makes it possible to explore the adaptive evolution of the intercept, $r(x)$, (at least over a range of sufficiently small mass) of the power law that scales birth rate with body mass—an issue currently attracting much attention from life-history biologists (Economo et al., 2005; McCarthy and Enquist, 2005).

The idea that most mutations are caused by molecular processes that ultimately are consequences of metabolism have led Gillooly et al. (2005) to predict that point mutations occur at a rate proportional to metabolic rate, i.e. the rate at which energy and materials are taken up from the environment and used for maintenance, growth and reproduction. The metabolic rate, and therefore the point mutation rate should scale as the $-\frac{1}{4}$ power of body mass, m, hence as the $+\frac{1}{3}$ power of system size, K (Gillooly et al., 2001). Provided that the mutation step variance scales as $K^{-\eta}$, this pattern is consistent with model (6.1) with $\eta = \frac{1}{3}$. However, when considering a large population of individuals subject to fast birth and death processes and rare mutation, the mutation rate, $b_K\mu_K$, is approximately equal to $r\mu$ and hence is independent of system size. Such tension with Gillooly et al.'s (2005) predictions may stem from their disputably combining the metabolic approach to scaling mutation rates with Kimura's neutral theory of molecular evolution (Kimura, 1968); in fact, body mass and metabolic rates do change in a process of non-neutral evolution, which warrants seeking an extension of Gillooly et al.'s (2005) theory by making use of the framework presented here.

All mutation parameters are assumed to be potentially influenced by trait values, which makes it possible to study the evolution of the mutagenesis process itself. An important feature of our theory is that it does not assume that mutational effects are symmetrical around the progenitor's trait values. Relaxing the symmetry assumption on the mutation distribution leads to extending the canonical equations to the case of biased mutation. This may be important for several reasons (Pomiankowski et al., 1991), two of them being structural: the measurement of any mutation bias (or the lack thereof) is scale dependent. If mutation is unbiased given one definition of a character and one scale of measurement then it must be biased for many other definitions and scales. Also, mutation bias may be trait-dependent: there may be little bias away from evolutionary singularities, but strong bias close to singularities that are extreme points of the trait set. There are genetic reasons for which the opposite may happen too. Mukai (1964) predicted that during episodes of directional selection, characters should show a negative correlation between mutation bias and the direction of selection; there may be weak mutation bias close to the adaptive optimum, but strong bias away from the optimum. Mutagenesis studies provide further evidence for the occurrence of mutation bias in quantitative traits (see e.g. Pomiankowski et al., 1991 , and Lai and Mackay, 1990). Whole-genome sequence analysis in a single-stranded DNA bacteriophage has recently provided direct evidence for the effect of mutation bias on the course of adaptive evolution (Rokyta et al., 2005).

One aspect that deterministic limits of selection dynamics fail to account for is random drift—the stochastic fixation of an allele with negative fitness, or as a corollary, the stochastic extinction of an allele with positive fitness. Our extension of the canonical equation as a diffusion

process represents a first step towards general incorporation of this effect in macroscopic models derived from individual stochastic processes. The mathematical nature of the diffusion model suggests to focus analysis on the consequences of multistability in the canonical equation, that is, the coexistence of several attractive evolutionary singularities. This focus hinges on the celebrated issue of disentangling the ecological and evolutionary causes and consequences of empirical records of punctualism, i.e., 'saltatory' dynamics characterized by long phases of phenotypic stasis interspersed with rapid transitions (e.g. Stanley, 1979). Specifically, even though our diffusion model of adaptive dynamics assumes that the population remains monomorphic, the model's properties underscore the possibility of contrasting macroevolutionary patterns dominated either by punctuated equilibria, or by radiation events. Radiation occurs when the population diversifies into two or more phenotypic branches (which may correspond to speciation in sexually reproducing species) (e.g. Schluter, 2000). It might be possible to develop a diffusion model allowing for evolutionary branching, for which our analysis could be extended to compare the time needed for phenotypes to diverge around branching points with the time taken by the population to jump between the basins of these branching points. Although branching may be a ubiquitous property of attractive evolutionary singularities (Doebeli and Dieckmann, 2000), there might be ecological and genetic conditions under which evolution would proceed through punctuated equilibria and long-term evolutionary cycles, rather than branching and radiation, in species that yet possess multiple branching points.

6.3. Conclusions

Recently Page and Nowak (2002) showed that apparently disparate deterministic models of evolutionary dynamics are, in fact, part of a single unified framework, while emphasizing that "stochastic, spatial or individual-based approaches (…) are notoriously more difficult to develop and analyze, and therefore much less amenable to any attempt of unification". By making use of advanced mathematical tools of probability theory (martingales, large deviations), this work achieves unification of quantitative-genetics and adaptive-dynamics models, together with other, previously unknown, macroscopic models of adaptive evolution. Our presentation underlines biological insights gained from the process of model construction itself. The obvious step to be taken next is model analysis, e.g. establishing existence and uniqueness of stationary solutions and characterizing their structure (in the wake of Bürger and Bomze's (1996) study of Kimura's equation), but this is not trivial. Even the numerical simulation of trajectories of the stochastic macroscopic models prove to raise formidable computational difficulties. Further directions for future research are prompted by the challenge of

relaxing two pivotal assumptions of our approach: the large system size approximation, and the simplistic structure of the trait space. Thus, the next generation of models should (i) assume large yet finite populations, and therefore study birth-mutation-death processes conditional on population non-extinction (e.g. Gosselin, 2001); and (ii) acknowledge the complex structure of real phenotypic space and genotype-phenotype mapping (e.g. Stadler et al., 2001) to account for developmental noise and plasticity in variable environments (e.g. Pigliucci, 2003), and therefore use probability measures on infinite-dimensional function spaces to model individual traits. This agenda delineates some of the new frontiers facing the mathematical exploration of adaptive evolution.

Appendix A

We develop here the proof of the convergence results to (4.5) and (4.6) stated in Section 4.2.1. The other results of Section 4 obtain by similar arguments.

We denote by $M_F = M_F(\mathbb{R}^l)$ the set of finite measures on $\mathbb{R}^l = \mathcal{X}$ endowed with the weak topology, and by $\mathbb{D}([0, T], M_F)$ the set of right-continuous and left-limited functions from $[0, T]$ to M_F, endowed with the Skorohod topology. We will also use, when it is convenient, the notation $\langle v, f \rangle$ for $\int f(x)v(\mathrm{d}x)$.

We rely on all the assumptions of Section 2.1, of the beginning of Section 4 and of Section 4.2.1. In particular, the variance-covariance matrix of $M_K(x, z)$ is given by $\Sigma(x)/K^\eta$, and we will assume that the third-order moment of $M_K(x, z)$ is bounded by $C/K^{\eta+\varepsilon}$ uniformly in x for some constants C and $\varepsilon > 0$. Let us also assume that, if $\sqrt{\Sigma}$ denotes the symmetrical square root matrix of Σ, the function $\sqrt{\Sigma r \mu}$ is bounded and Lipschitz. We will also assume that $b(x, \zeta)$ and $d(x, \zeta)$ are globally Lipschitz with respect to the second variable, that X_0^K converges in law to ξ_0 for the weak topology in M_F and that

$$\sup_K E[\langle X_0^K, 1 \rangle^3] < +\infty. \tag{A.2}$$

Since $X^K = (1/K)v^K$, and using (2.2), the process X^K is a Markov process with generator

$$
\begin{aligned}
L^K \phi(v) =\ & K \int_{\mathbb{R}^l} (K^\eta r(x) + b(x, V * v(x)))(1 - \mu(x)) \\
& \times \left(\phi\left(v + \frac{1}{K}\delta_x\right) - \phi(v) \right) v(\mathrm{d}x) \\
& + K \int_{\mathbb{R}^l} (K^\eta r(x) + b(x, V * v(x)))\mu(x) \\
& \times \int_{\mathbb{R}^l} \left(\phi\left(v + \frac{1}{K}\delta_{x+z}\right) - \phi(v) \right) M_K(x, z)\,\mathrm{d}z\,v(\mathrm{d}x) \\
& + K \int_{\mathbb{R}^l} (K^\eta r(x) + d(x, U * v(x))) \\
& \times \left(\phi\left(v - \frac{1}{K}\delta_x\right) - \phi(v) \right) v(\mathrm{d}x).
\end{aligned}
\tag{A.3}
$$

A simple computation using the boundedness of b and the Gronwall Lemma allows us to prove the following lemma (see the proof of Theorem 3.1 in Fournier and Méléard, 2004):

Lemma A.1. *For any K, if $E[\langle X_0^K, 1\rangle^3] < +\infty$, then, for any $T > 0$,*

$$
E\left(\sup_{t\in[0,T]} \langle X_t^K, 1\rangle^3 \right) < +\infty.
$$

By standard probabilistic arguments, one obtains a martingale decomposition for functions of X^K: for any bounded and measurable functions ϕ on M_F such that $|\phi(v)| + |L^K\phi(v)| \leqslant C(1 + <v, 1>^3)$, the process

$$
\phi(X_t^K) - \phi(X_0^K) - \int_0^t L^K\phi(X_s^K)\,\mathrm{d}s
\tag{A.4}
$$

is a martingale. In particular, applying this property to $\phi(v) = \langle v, f\rangle$ and $\phi(v) = \langle v, f\rangle^2$ for each measurable bounded function f, we obtain

$$
\begin{aligned}
& \int_{\mathbb{R}^l} f(x) X_t^K(\mathrm{d}x) \\
& = m_t^{K,f} + \int_{\mathbb{R}^l} f(x) X_0^K(\mathrm{d}x) + \int_0^t \int_{\mathbb{R}^l} (b(x, V * X_s^K(x)) \\
& \quad - d(x, U * X_s^K(x)))f(x) X_s^K(\mathrm{d}x)\,\mathrm{d}s \\
& \quad + \int_0^t \int_{\mathbb{R}^l} \mu(x)(K^\eta r(x) + b(x, V * X_s^K(x))) \\
& \quad \times \left(\int_{\mathbb{R}^l} f(x + z) M_K(x, z)\,\mathrm{d}z - f(x) \right) X_s^K(\mathrm{d}x)\,\mathrm{d}s,
\end{aligned}
\tag{A.5}
$$

where $m_t^{K,f}$ is a martingale with variance (quadratic variation)

$$
\langle m^{K,f}\rangle_t = \frac{1}{K}\left\{ \int_0^t \int_{\mathbb{R}^l} \mu(x)(K^\eta r(x) + b(x, V * X_s^K(x))) \right.
$$

$$
\begin{aligned}
& \times \left(\int_{\mathbb{R}^l} f^2(x + z) M_K(x, z)\,\mathrm{d}z - f^2(x) \right) X_s^K(\mathrm{d}x)\,\mathrm{d}s \\
& + \int_0^t \int_{\mathbb{R}^l} (2K^\eta r(x) + b(x, V * X_s^K(x)) \\
& \left. + d(x, U * X_s^K(x)))f^2(x) X_s^K(\mathrm{d}x)\,\mathrm{d}s \right\}
\end{aligned}
\tag{A.6}
$$

Then the convergence results of Section 4.2.1 can be stated as follows.

Theorem A.2. (1) *Assume all the assumptions above and $0 < \eta < 1$. Assume also that the measure ξ_0 is deterministic.*

Then, for each $T > 0$, the sequence of processes (X^K) belonging to $\mathbb{D}([0, T], M_F)$ converges (in law) to the deterministic continuous function $(\xi_t)_{t\geqslant 0} \in C([0, T], M_F)$, unique solution satisfying $\sup_{t\in[0,T]} \langle \xi_t, 1\rangle < \infty$, of the integro-differential equation written in its weak form: for each function $f \in C_b^2(\mathbb{R}^l)$,

$$
\begin{aligned}
\langle \xi_t, f\rangle =\ & \langle \xi_0, f\rangle \\
& + \int_0^t \int_{\mathbb{R}^l} (b(x, V * \xi_s(x)) - d(x, U * \xi_s(x)))f(x)\xi_s(\mathrm{d}x)\,\mathrm{d}s \\
& + \int_0^t \int_{\mathbb{R}^l} \frac{1}{2}\mu(x)r(x) \sum_{1\leqslant i,j\leqslant l} \Sigma_{ij}(x)\partial_{ij}^2 f(x)\xi_s(\mathrm{d}x)\,\mathrm{d}s
\end{aligned}
\tag{A.7}
$$

(2) *Assume moreover that there exists $c > 0$ such that $r(x)\mu(x)s^*\Sigma(x)s \geqslant c\|s\|^2$ for each x and $s \in \mathbb{R}^l$. Then for each $t > 0$, the measure ξ_t has a density with respect to the Lebesgue measure.*

Remark A.3. In case (2), Eq. (A.7) may then be written as

$$
\begin{aligned}
\partial_t \xi_t(x) =\ & (b(x, V * \xi_t(x)) - d(x, U * \xi_t(x)))\xi_t(x) \\
& + \frac{1}{2} \sum_{1\leqslant i,j\leqslant l} \partial_{ij}^2 (r\mu \Sigma_{ij} \xi_t)(x),
\end{aligned}
\tag{A.8}
$$

which yields the Fisher reaction–diffusion equation (4.5) in the case where $\Sigma(x) = \sigma^2(x)\mathrm{Id}$.

Theorem A.4. *Assume all the assumptions above and $\eta = 1$. Here, ξ_0 may be stochastic. Then, for each $T > 0$, the sequence of processes (X^K) converges in law in $\mathbb{D}([0, T], M_F)$ to the unique (in law) continuous superprocess $X \in C([0, T], M_F)$, defined by the following conditions:*

$$
\sup_{t\in[0,T]} E[\langle X_t, 1\rangle^3] < \infty,
\tag{A.9}
$$

and for any $f \in C_b^2(\mathbb{R}^l)$,

$$
\begin{aligned}
\bar{m}_t^f =\ & \langle X_t, f\rangle - \langle X_0, f\rangle - \frac{1}{2}\int_0^t \int_{\mathbb{R}^l} \mu(x)r(x) \\
& \times \sum_{1\leqslant i,j\leqslant l} \Sigma_{ij}(x)\partial_{ij}^2 f(x) X_s(\mathrm{d}x)\,\mathrm{d}s \\
& - \int_0^t \int_{\mathbb{R}^l} f(x)[b(x, V * X_s(x)) \\
& - d(x, U * X_s(x))] X_s(\mathrm{d}x)\,\mathrm{d}s
\end{aligned}
\tag{A.10}
$$

is a continuous martingale with quadratic variation

$$\langle \bar{m}^f \rangle_t = 2 \int_0^t \int_{\mathbb{R}^l} r(x) f^2(x) X_s(\mathrm{d}x)\, \mathrm{d}s. \tag{A.11}$$

Remark A.5. The conditions characterizing the process X above can be formally rewritten as Eq. (4.8) of Section 4.2.1, but not rigorously since, apart from the case $l = 1$, we suspect that X_t has a.s. no density with respect to the Lebesgue measure.

Proof of Theorem A.2. (1) We divide the proof into six steps. Let us fix $T > 0$.

Step 1: Let us first show uniqueness for (A.7).

First, we need to define the evolution equation associated with (A.7). It is easy to prove that if ξ is a solution of (A.7) satisfying $\sup_{t \in [0,T]} \langle \xi_t, 1 \rangle < \infty$, then for each test function $\psi_t(x) = \psi(t, x) \in C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^l)$, one has

$$\langle \xi_t, \psi_t \rangle = \langle \xi_0, \psi_0 \rangle + \int_0^t \int_{\mathbb{R}^l} (b(x, V * \xi_s(x))$$
$$- d(x, U * \xi_s(x)))\psi(s, x)\xi_s(\mathrm{d}x)\, \mathrm{d}s + \int_0^t \int_{\mathbb{R}^l} (\partial_s \psi(s, x)$$
$$+ \frac{1}{2} r(x)\mu(x) \sum_{i,j} \Sigma_{ij}(x)\partial_{ij}^2 \psi_s(x))\xi_s(\mathrm{d}x)\, \mathrm{d}s. \tag{A.12}$$

Now, since the function $\sqrt{\Sigma r \mu}$ is Lipschitz continuous and bounded, we may define the transition semigroup (P_t) with infinitesimal generator $f \mapsto \frac{1}{2} r\mu \sum_{i,j} \Sigma_{ij} \partial_{ij}^2 f$. Then, for each function $f \in C_b^2(\mathbb{R}^l)$ and fixed $t > 0$, choosing $\psi(s, x) = P_{t-s}f(x)$ yields

$$\langle \xi_t, f \rangle = \langle \xi_0, P_t f \rangle + \int_0^t \int_{\mathbb{R}^l} (b(x, V * \xi_s(x))$$
$$- d(x, U * \xi_s(x)))P_{t-s}f(x)\xi_s(\mathrm{d}x)\, \mathrm{d}s, \tag{A.13}$$

since $\partial_s \psi(s, x) + \frac{1}{2}\sigma^2(x)r(x)\mu(x)\Delta_x \psi(s, x) = 0$ for this choice.

We now prove the uniqueness of a solution of (A.13).

Let us consider two solutions $(\xi_t)_{t \geq 0}$ and $(\bar{\xi}_t)_{t \geq 0}$ of (A.13) satisfying $\sup_{t \in [0,T]} \langle \xi_t + \bar{\xi}_t, 1 \rangle = A_T < +\infty$. We consider the variation norm defined for μ_1 and μ_2 in M_F by

$$\|\mu_1 - \mu_2\| = \sup_{f \in L^\infty(\mathbb{R}^l),\ \|f\|_\infty \leq 1} |\langle \mu_1 - \mu_2, f \rangle|. \tag{A.14}$$

Then, we consider some bounded and measurable function f defined on $\mathbb{R}^l$ such that $\|f\|_\infty \leq 1$ and obtain

$$|\langle \xi_t - \bar{\xi}_t, f \rangle| \leq \int_0^t \left| \int_{\mathbb{R}^l} (b(x, V * \xi_s(x)) - d(x, U * \xi_s(x))) \right.$$
$$\left. \times P_{t-s}f(x)[\xi_s(\mathrm{d}x) - \bar{\xi}_s(\mathrm{d}x)] \right| \mathrm{d}s$$
$$+ \int_0^t \left| \int_{\mathbb{R}^l} (b(x, V * \xi_s(x)) - b(x, V * \bar{\xi}_s(x))) \right.$$
$$\left. \times P_{t-s}f(x)\bar{\xi}_s(\mathrm{d}x) \right| \mathrm{d}s$$

$$+ \int_0^t \left| \int_{\mathbb{R}^l} (d(x, U * \xi_s(x)) - d(x, U * \bar{\xi}_s(x))) \right.$$
$$\left. \times P_{t-s}f(x)\bar{\xi}_s(\mathrm{d}x) \right| \mathrm{d}s. \tag{A.15}$$

Since $\|f\|_\infty \leq 1$, then $\|P_{t-s}f\|_\infty \leq 1$ and for all $x \in \mathbb{R}^l$,

$$|(b(x, V * \xi_s(x)) - d(x, U * \xi_s(x)))P_{t-s}f(x)| \leq \bar{b} + \bar{d}(1 + \bar{U}A_T).$$

Moreover, b and d are Lipschitz continuous in their second variable with respective constants K_b and K_d. We then obtain from (A.15) that

$$|\langle \xi_t - \bar{\xi}_t, f \rangle| \leq [\bar{b} + \bar{d}(1 + \bar{U}A_T) + K_b A_T \bar{V} + K_d A_T \bar{U}]$$
$$\times \int_0^t \|\xi_s - \bar{\xi}_s\|\, \mathrm{d}s. \tag{A.16}$$

Taking the supremum over all functions f such that $\|f\|_\infty \leq 1$, and using Gronwall's Lemma, we finally deduce that for all $t \leq T$, $\|\xi_t - \bar{\xi}_t\| = 0$. Uniqueness holds.

Step 2: Next, we establish some moment estimates. First, we check that for all $T < \infty$,

$$\sup_K \sup_{t \in [0,T]} E[\langle X_t^K, 1 \rangle^3] < \infty. \tag{A.17}$$

To this end, we use (A.4) with $\phi(v) = \langle v, 1 \rangle^3$. (To be completely rigorous, one should first use $\phi(v) = \langle v, 1 \rangle^3 \wedge A$, make A tend to infinity and then use Lemma A.1 to ensure the convergence of all terms except the last, and apply the monotone convergence theorem to the last term). Taking expectation, we obtain that for all $t \geq 0$, all K,

$$E[\langle X_t^K, 1 \rangle^3] = E[\langle X_0^K, 1 \rangle^3]$$
$$+ \int_0^t E\left[\int_{\mathbb{R}^l} [K^{\eta+1}r(x) + Kb(x, V * X_s^K(x))] \right.$$
$$\times \left\{ \left[\langle X_s^K, 1 \rangle + \frac{1}{K} \right]^3 - \langle X_s^K, 1 \rangle^3 \right\} X_s^K(\mathrm{d}x) \right] \mathrm{d}s$$
$$+ \int_0^t E\left[\int_{\mathbb{R}^l} \{K^{\eta+1}r(x) + Kd(x, U * X_s^K(x))\} \right.$$
$$\times \left\{ \left[\langle X_s^K, 1 \rangle - \frac{1}{K} \right]^3 - \langle X_s^K, 1 \rangle^3 \right\} X_s^K(\mathrm{d}x) \right] \mathrm{d}s.$$

Neglecting the non-positive death term involving d, we get

$$E[\langle X_t^K, 1 \rangle^3] \leq E[\langle X_0^K, 1 \rangle^3]$$
$$+ \int_0^t E\left[\int_{\mathbb{R}^l} K^{\eta+1}r(x)\left\{ \left[\langle X_s^K, 1 \rangle + \frac{1}{K} \right]^3 \right. \right.$$
$$\left. \left. + \left[\langle X_s^K, 1 \rangle - \frac{1}{K} \right]^3 - 2\langle X_s^K, 1 \rangle^3 \right\} X_s^K(\mathrm{d}x) \right] \mathrm{d}s$$
$$+ \int_0^t E\left[\int_{\mathbb{R}^l} Kb(x, V * X_s^K(x)) \right.$$
$$\times \left\{ \left[\langle X_s^K, 1 \rangle + \frac{1}{K} \right]^3 - \langle X_s^K, 1 \rangle^3 \right\} X_s^K(\mathrm{d}x) \right] \mathrm{d}s.$$

But for all $x \geq 0$, all $\varepsilon \in (0,1]$, $(x+\varepsilon)^3 - x^3 \leq 6\varepsilon(1+x^2)$ and $|(x+\varepsilon)^3 + (x-\varepsilon)^3 - 2x^3| = 6\varepsilon^2 x$. We finally obtain

$$
\begin{aligned}
E\left[\langle X_t^K, 1\rangle^3\right] &\leq E\left[\langle X_0^K, 1\rangle^3\right] \\
&+ C \int_0^t E\left[\langle X_s^K, 1\rangle + \langle X_s^K, 1\rangle^2 + \langle X_s^K, 1\rangle^3\right] ds.
\end{aligned}
$$

Assumption (A.2) and Gronwall's Lemma allow us to conclude that (A.17) holds.

Next, we need to check that

$$
\sup_K E\left(\sup_{t \in [0,T]} \langle X_t^K, 1\rangle^2 \right) < \infty. \tag{A.18}
$$

Applying (A.5) with $f \equiv 1$, we obtain

$$
\begin{aligned}
\langle X_t^K, 1\rangle = \langle X_0^K, 1\rangle &+ \int_0^t \int_{\mathbb{R}^l} [b(x, V * X_s^K(x)) \\
&- d(x, U * X_s^K(x))] X_s^K(dx)\, ds + m_t^{K,1}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\sup_{s \in [0,t]} \langle X_s^K, 1\rangle^2 \\
&\leq C\left(\langle X_0^K, 1\rangle^2 + \bar{b} \int_0^t \langle X_s^K, 1\rangle^2\, ds + \sup_{s \in [0,t]} |m_s^{K,1}|^2 \right).
\end{aligned}
$$

Thanks to (A.2), Doob's inequality and Gronwall's Lemma, there exists a constant C_t not depending on K such that

$$
E\left(\sup_{s \in [0,t]} \langle X_s^K, 1\rangle^2 \right) \leq C_t (1 + E[\langle m^{K,1}\rangle_t]).
$$

Using now (A.6), we obtain, for some other constant C_t not depending on K,

$$
E[\langle m^{K,1}\rangle_t] \leq C \int_0^t (E[\langle X_s^K, 1\rangle] + E[\langle X_s^K, 1\rangle^2])\, ds \leq C_t
$$

thanks to (A.17). This concludes the proof of (A.18).

Step 3: We first endow M_F with the vague topology, the extension to the weak topology being handled in Step 6 below. To show the tightness of the sequence of laws $Q^K = \mathcal{L}(X^K)$ in $\mathcal{P}(\mathbb{D}([0,T], M_F))$, it suffices, following Roelly (1986), to show that for any continuous bounded function f on $\mathbb{R}^l$, the sequence of laws of the processes $\langle X^K, f\rangle$ is tight in $\mathbb{D}([0,T], \mathbb{R})$. To this end, we use Aldous' criterion (Aldous, 1978) and Rebolledo's criterion (see Joffe and Métivier, 1986). We have to show that

$$
\sup_K E\left(\sup_{t \in [0,T]} |\langle X_t^K, f\rangle| \right) < \infty, \tag{A.19}
$$

and the tightness of the laws of the predictable quadratic variation of the martingale part and of the finite variation part of the semimartingales $\langle X^K, f\rangle$, respectively.

Since f is bounded, (A.19) is a consequence of (A.18). Let us thus consider a couple (S, S') of stopping times

satisfying a.s. $0 \leq S \leq S' \leq S + \delta \leq T$. Using (A.5), we get

$$
\begin{aligned}
&E[\langle m^{K,f}\rangle_{S'} - \langle m^{K,f}\rangle_S] \\
&\quad \leq E\left[C \int_S^{S+\delta} (\langle X_s^K, 1\rangle + \langle X_s^K, 1\rangle^2)\, ds \right] \leq C\delta,
\end{aligned}
$$

the last inequality coming from (A.18), and a similar inequality is obtained for the finite variation part. Therefore, the sequence $Q^K = \mathcal{L}(X^K)$ is uniformly tight.

Step 4: Let us now denote by Q the limiting law of a subsequence of Q^K. We still denote this subsequence by Q^K. Let $X = (X_t)_{t \geq 0}$ a process with law Q. We remark that by construction, almost surely,

$$
\sup_{t \in [0,T]} \sup_{f \in L^\infty(\mathbb{R}^l), \|f\|_\infty \leq 1} |\langle X_t^K, f\rangle - \langle X_{t^-}^K, f\rangle| \leq \frac{1}{K}.
$$

This implies that the process X is a.s. strongly continuous.

Step 5: The time $T > 0$ is fixed. Let us now check that, almost surely, the process X is the unique solution of (A.7). Thanks to (A.18), it satisfies $\sup_{t \in [0,T]} \langle X_t, 1\rangle < +\infty$ a.s., for each T. We now fix a function $f \in C_b^3(\mathbb{R}^l)$ (the extension of (A.7) to any function f in C_b^2 is not hard) and some $t \leq T$.

For $v \in C([0,T], M_F)$, we set

$$
\begin{aligned}
\Psi_t^1(v) = \langle v_t, f\rangle - \langle v_0, f\rangle &- \int_0^t \int_{\mathbb{R}^l} (b(x, V * v_s(x)) \\
&- d(x, U * v_s(x))) f(x) v_s(dx)\, ds,
\end{aligned}
$$

$$
\Psi_t^2(v) = - \int_0^t \int_{\mathbb{R}^l} \frac{1}{2} \mu(x) r(x) \sum_{i,j} \Sigma_{ij}(x) \partial_{ij}^2 f(x) v_s(dx)\, ds. \tag{A.20}
$$

We have to show that

$$
E_Q[|\Psi_t^1(X) + \Psi_t^2(X)|] = 0. \tag{A.21}
$$

By (A.5), we know that for each K,

$$
m_t^{K,f} = \Psi_t^1(X^K) + \Psi_t^{2,K}(X^K),
$$

where

$$
\begin{aligned}
\Psi_t^{2,K}(X^K) = - \int_0^t \int_{\mathbb{R}^l} &\mu(x)(K^\eta r(x) + b(x, V * X_s^K(x))) \\
&\times \left(\int_{\mathbb{R}^l} f(x+z) M_K(x,z)\, dz - f(x) \right) X_s^K(dx)\, ds.
\end{aligned} \tag{A.22}
$$

Moreover, (A.18) implies that for each K,

$$
\begin{aligned}
E\left[|m_t^{K,f}|^2 \right] = E\left[\langle m^{K,f}\rangle_t \right] &\leq \frac{C_f K^\eta}{K} E\left[\int_0^t \{\langle X_s^K, 1\rangle + \langle X_s^K, 1\rangle^2\}\, ds \right] \\
&\leq \frac{C_{f,T} K^\eta}{K}, \tag{A.23}
\end{aligned}
$$

which goes to 0 as K tends to infinity, since $0 < \eta < 1$. Therefore,

$$
\lim_K E[|\Psi_t^1(X^K) + \Psi_t^{2,K}(X^K)|] = 0.
$$

Since X is a.s. strongly continuous, since $f \in C_b^3(\mathbb{R}^l)$ and thanks to the continuity of the parameters, the functions

Ψ_t^1 and Ψ_t^2 are a.s. continuous at X. Furthermore, for any $v \in \mathbb{D}([0, T], M_F)$,

$$|\Psi_t^1(v) + \Psi_t^2(v)| \leqslant C_{f,T} \sup_{s \in [0,T]} (1 + \langle v_s, 1 \rangle^2). \tag{A.24}$$

Hence using (A.17), we see that the sequence $(\Psi_t^1(X^K) + \Psi_t^2(X^K))_K$ is uniformly integrable, and thus

$$\lim_K E(|\Psi_t^1(X^K) + \Psi_t^2(X^K)|) = E(|\Psi_t^1(X) + \Psi_t^2(X)|). \tag{A.25}$$

Now we have to deal with $\Psi_t^{2,K}(X^K) - \Psi_t^2(X^K)$. The convergence of this term is due to the fact that the measure $M_K(x, z) \, \mathrm{d}z$ has mean 0, variance $\Sigma(x)/K^\eta$, and third moment bounded by $C/K^{\eta+\varepsilon}$ ($\varepsilon > 0$) uniformly in x. Indeed, if $Hf(x)$ denotes the Hessian matrix of f at x,

$$\int_{\mathbb{R}^l} f(x+z) M_K(x, z) \, \mathrm{d}z$$
$$= \int_{\mathbb{R}^l} \left(f(x) + z \cdot \nabla f(x) + \frac{1}{2} z^* Hf(x) z + \mathrm{O}(z^3) \right) M_K(x, z) \, \mathrm{d}z$$
$$= f(x) + \frac{1}{2} \sum_{i,j} \frac{\Sigma_{ij}(x)}{K^\eta} \partial_{ij}^2 f(x) + \mathrm{o}\left(\frac{1}{K^\eta} \right). \tag{A.26}$$

where $K^\eta \mathrm{o}(1/K^\eta)$ tends to 0 uniformly in x (since f is in C_b^3), as K tends to infinity. Then,

$$\Psi_t^{2,K}(X^K) = -\int_0^t \int_{\mathbb{R}^l} \mu(x)(K^\eta r(x) + b(x, V * X_s^K(x)))$$
$$\times \left(\frac{1}{2} \sum_{i,j} \frac{\Sigma_{ij}(x)}{K^\eta} \partial_{ij}^2 f(x) + \mathrm{o}\left(\frac{1}{K^\eta} \right) \right) X_s^K(\mathrm{d}x) \, \mathrm{d}s,$$

and

$$|\Psi_t^{2,K}(X^K) - \Psi_t^2(X^K)| \leqslant C_f \langle X_s^K, 1 \rangle \left(\frac{1}{K^\eta} + K^\eta \mathrm{o}\left(\frac{1}{K^\eta} \right) \right).$$

Using (A.18), we conclude the proof of (A.21).

Step 6: The previous steps imply that the sequence $(X^K)_K$ converges to ξ in $\mathbb{D}([0, T], M_F)$, where M_F is endowed with the vague topology. To extend the result to the case where M_F is endowed with the weak topology, we use a criterion proved in Méléard and Roelly (1993): since the limiting process is continuous, it suffices to prove that the sequence $(\langle X^K, 1 \rangle)_K$ converges to $\langle \xi, 1 \rangle$ in law, in $\mathbb{D}([0, T], \mathbb{R}^l)$. One may of course apply Step 5 with $f \equiv 1$, which concludes the proof.

(2) Let us now assume the non-degeneracy property $r(x)\mu(x)s^*\Sigma(x)s \geqslant c\|s\|^2$ for each $x, s \in \mathbb{R}^l$. That implies that for each time $t > 0$, the transition semigroup $P_t(x, \mathrm{d}y)$ introduced in Step 1 of this proof has for each $x \in \mathbb{R}^l$ a density function $p_t(x, y)$ with respect to the Lebesgue measure. Hence (A.13) writes

$$\int_{\mathbb{R}^l} f(x) \xi_t(\mathrm{d}x) = \int_{\mathbb{R}^l} \left(\int_{\mathbb{R}^l} f(y) p_t(x, y) \, \mathrm{d}y \right) \xi_0(\mathrm{d}x)$$
$$+ \int_0^t \int_{\mathbb{R}^l} (b(x, V * \xi_s(x)) - d(x, U * \xi_s(x)))$$
$$\times \left(\int_{\mathbb{R}^l} f(y) p_{t-s}(x, y) \, \mathrm{d}y \right) \xi_s(\mathrm{d}x) \, \mathrm{d}s.$$

Using the boundedness of f and of the parameters, and $\sup_{t \leqslant T} \langle \xi_t, 1 \rangle < +\infty$, we apply Fubini's theorem and deduce that

$$\int_{\mathbb{R}^l} f(x) \xi_t(\mathrm{d}x) = \int_{\mathbb{R}^l} H_t(y) f(y) \, \mathrm{d}y$$

where $H \in L^\infty([0, T], L^1(\mathbb{R}^l))$, which implies that ξ_t has a density with respect to the Lebesgue measure for each time $t \leqslant T$.

Proof of Theorem A.4. We use a similar method as previously. Steps 2, 3, 4 and 6 of this proof are essentially the same. Therefore, we have to prove the uniqueness (in law) of the solution to the martingale problem (A.9)–(A.11) (Step 1), and that any accumulation point of the sequence of laws of X^K is solution of (A.9)–(A.11) (Step 5).

Step 1: This uniqueness result is well-known for the super-Brownian process (defined by a similar martingale problem, but with $b = d = 0$, $r = \mu = 1$ and $\Sigma = \mathrm{Id}$, see Roelly, 1986). As in Etheridge (2004), we use the version of Dawson's Girsanov transform obtained in Evans and Perkins (1994, Theorem 2.3), which implies the uniqueness of the law of X, provided that the condition

$$E\left(\int_0^t \int_{\mathbb{R}^l} [b(x, V * X_s(x)) - d(x, U * X_s(x))]^2 X_s(\mathrm{d}x) \, \mathrm{d}s \right) < +\infty$$

is satisfied. This is easily obtained from $\sup_{t \in [0,T]} E[\langle X_t, 1 \rangle^3] < \infty$.

Step 5: Let us call $Q^K = \mathcal{L}(X^K)$ and denote by Q a limiting value of the tight sequence Q^K, and by $X = (X_t)_{t \geqslant 0}$ a process with law Q. Because of Step 4, X belongs a.s. to $C([0, T], M_F)$. Let us show that X satisfies conditions (A.9)–(A.11). Equation (A.9) is straightforward from (A.18). Let us consider a function f in $C_b^3(\mathbb{R}^l)$, and prove that the process $\bar{m}_t^f$ defined by (A.10) is a martingale (the extension to every function in C_b^2 is not hard). For $0 \leqslant s_1 \leqslant \cdots \leqslant s_n < s < t$, some continuous bounded maps $\phi_1, \ldots, \phi_n$ on M_F, we introduce the function Ψ from $\mathbb{D}([0, T], M_F)$ into $\mathbb{R}$ defined by

$$\Psi(v) = \phi_1(v_{s_1}) \ldots \phi_n(v_{s_n}) \left\{ \langle v_t, f \rangle - \langle v_s, f \rangle \right.$$
$$- \frac{1}{2} \int_s^t \int_{\mathbb{R}^l} \mu(x) r(x) \sum_{i,j} \Sigma_{ij}(x) \partial_{ij}^2 f(x) v_u(\mathrm{d}x) \, \mathrm{d}u$$
$$- \int_s^t \int_{\mathbb{R}^l} f(x)[b(x, V * v_u(x))$$
$$\left. - d(x, U * v_u(x))] v_u(\mathrm{d}x) \, \mathrm{d}u \right\}. \tag{A.27}$$

Our aim is to show that

$$E(\Psi(X)) = 0. \tag{A.28}$$

It follows from (A.5) that

$$0 = E\left[\phi_1(X_{s_1}^K) \ldots \phi_n(X_{s_n}^K) \{ m_t^{K,f} - m_s^{K,f} \} \right] = E[\Psi(X^K)] - A_K, \tag{A.29}$$

where A_K is defined by

$$A_K = E\left[\phi_1(X_{s_1}^K)\ldots\phi_n(X_{s_n}^K)\int_s^t\int_{\mathbb{R}^l}\mu(x)\left\{r(x)K\right.\right.$$

$$\times\left[\int_{\mathbb{R}^l}M_K(x,z)f(x+z)\,\mathrm{d}z - f(x)\right.$$

$$\left.-\sum_{i,j}\frac{\Sigma_{ij}(x)}{2K}\partial_{ij}^2 f(x)\right] + b(x, V*X_u^K(x))$$

$$\left.\left.\times\left[\int_{\mathbb{R}^l}M_K(x,z)f(x+z)\,\mathrm{d}z - f(x)\right]\right\}X_u^K(\mathrm{d}x)\,\mathrm{d}u\right].$$

It turns out from (A.26) that A_K tends to 0, as K grows to infinity, and from (A.18), that the sequence $(|\Psi(X^K)|)_K$ is uniformly integrable. Hence,

$$\lim_K E(|\Psi(X^K)|) = E_Q(|\Psi(X)|). \tag{A.30}$$

Thus, (A.28) holds, and therefore, $\bar{m}^f$ is a martingale.

It remains to show (A.11). To this end, we first check that

$$\bar{N}_t^f = \langle X_t, f\rangle^2 - \langle X_0, f\rangle^2 - \int_0^t\int_{\mathbb{R}^l}2r(x)f^2(x)X_s(\mathrm{d}x)\,\mathrm{d}s$$

$$-\int_0^t 2\langle X_s, f\rangle\int_{\mathbb{R}^l}f(x)[b(x, V*X_s(x))$$

$$-d(x, U*X_s(x))]X_s(\mathrm{d}x)\,\mathrm{d}s$$

$$-\int_0^t 2\langle X_s, f\rangle\int_{\mathbb{R}^l}\frac{1}{2}\mu(x)r(x)\sum_{i,j}\Sigma_{ij}(x)\partial_{ij}^2 f(x)X_s(\mathrm{d}x)\,\mathrm{d}s$$
$$\tag{A.31}$$

is a martingale. This can be done exactly as for $\bar{m}_t^f$, using the semimartingale decomposition of $\langle X_t^K, f\rangle^2$, given by (A.4) applied to $\phi(v) = \langle v, f\rangle^2$. Next, Itô's formula applied to (A.10) implies that

$$\langle X_t, f\rangle^2 - \langle X_0, f\rangle^2 - \langle\bar{M}^f\rangle_t$$

$$-\int_0^t 2\langle X_s, f\rangle\int_{\mathbb{R}^l}f(x)[b(x, V*X_s(x))$$

$$-d(x, U*X_s(x))]X_s(\mathrm{d}x)\,\mathrm{d}s$$

$$-\int_0^t 2\langle X_s, f\rangle\int_{\mathbb{R}^l}\frac{1}{2}r(x)\mu(x)\sum_{i,j}\Sigma_{ij}(x)\partial_{ij}^2 f(x)X_s(\mathrm{d}x)\,\mathrm{d}s$$

is a martingale. Comparing this formula with (A.31), we obtain (A.11). $\quad\square$

References

Abrams, P.A., 2001. Modelling the adaptive dynamics of traits involved in inter- and intraspecific interactions: an assessment of three methods. Ecol. Lett. 4, 166–175.

Aldous, D., 1978. Stopping times and tightness. Ann. Probab. 6, 335–340.

Asmussen, M.A., 1983. Density-dependent selection incorporating intraspecific competition. II. A diploid model. Genetics 103, 335–350.

Barton, N.H., Polechova, J., 2005. The limitations of adaptive dynamics as a model of evolution. J. Evol. Biol. 18, 1186–1190.

Belgrano, A., Allen, A.P., Enquist, B.J., Gillooly, J.F., 2002. Allometric scaling of maximum population density: a common rule for marine phytoplankton and terrestrial plants. Ecol. Lett. 5, 611–613.

Berstein, H., Byerly, H.C., Hopf, F.A., Michod, R.E., 1985. Sex and the emergence of species. J. Theor. Biol. 117, 665–690.

Bolker, B.M., Pacala, S.W., 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. Theor. Popul. Biol. 52, 179–197.

Bolker, B.M., Pacala, S.W., 1999. Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. Am. Nat. 153, 575–602.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., West, G.B., 2004. Toward a metabolic theory of ecology. Ecology 85, 1771–1789.

Bulmer, M.G., 1974. Density-dependent selection and character displacement. Am. Nat. 108, 45–58.

Bürger, R., 1998. Mathematical properties of mutation-selection models. Genetica 102/103, 279–298.

Bürger, R., 2000. The Mathematical Theory of Selection, Recombination, and Mutation. Wiley, Chichester.

Bürger, R., 2005. A multilocus analysis of intraspecific competition and stabilizing selection on a quantitative trait. J. Math. Biol. 50, 355–396.

Bürger, R., Bomze, I.M., 1996. Stationary distributions under mutation-selection balance: structure and properties. Adv. Appl. Probab. 28, 227–251.

Bürger, R., Gimelfarb, A., 2004. The effects of intraspecific competition and stabilizing selection on a polygenic trait. Genetics 167, 1425–1443.

Calder III, W.A., 1984. Size, Function, and Life History. Harvard University Press, Cambridge.

Champagnat, N., 2003. Large deviations for diffusion models of adaptive dynamics. Preprint MODALX 03/23, University of Paris X.

Champagnat, N., 2004a. A microscopic interpretation for adaptive dynamics trait substitution sequence models. Preprint MODALX 04/20, University of Paris X.

Champagnat, N., 2004b. Étude mathématique de modèles stochastiques d'évolution issus de la théorie écologique des dynamiques adaptatives. Ph.D. Thesis, University of Paris X.

Champagnat, N., Ferrière, R., Ben Arous, G., 2001. The canonical equation of adaptive dynamics: a mathematical view. Selection 2, 71–81.

Charnov, E.L., 1993. Life History Invariants. Oxford University Press, Oxford.

Christiansen, F.B., Loeschcke, V., 1980. Evolution and intraspecific exploitative competition. II. One-locus theory for small additive gene effects. Theor. Popul. Biol. 18, 297–313.

Crow, J.F., Kimura, M., 1964. The theory of genetic loads. In: Geerts, S. (Ed.), Proceedings of the XI International Congress of Genetics. Pergamon Press, Oxford, pp. 495–505.

Damuth, J., 1987. Interspecific allometry of population density in mammals and other animals: the independence of body mass and population energy use. Biol. J. Linn. Soc. 31, 193–246.

Dembo, A., Zeitouni, O., 1993. Large deviations techniques and applications. 2nd ed. John and Barlett Publishers, Boston.

Dercole, F., Ferrière, R., Rinaldi, S., 2002. Ecological bistability and evolutionary reversals under asymmetrical competition. Evolution 56, 1081–1090.

Desvillettes, L., Prévost, C., Ferrière, R., 2004. Infinite dimensional reaction-diffusion for evolutionary population dynamics. Preprint CMLA, Ecole Normale Supérieure de Cachan.

Dieckmann, U., Ferrière, R., 2004. Adaptive dynamics and evolving biodiversity. In: Ferrière, R., Dieckmann, U., Couvet, D. (Eds.), Evolutionary Conservation Biology. Cambridge University Press, Cambridge, pp. 188–224.

Dieckmann, U., Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. J. Math. Biol. 34, 579–612.

Dieckmann, U., Law, R., 2000. Relaxation projections and the method of moments. In: Dieckmann, U., Law, R., Metz, J.A.J. (Eds.), The

Geometry of Ecological Interactions: Symplifying Spatial Complexity. Cambridge University Press, Cambridge, pp. 412–455.

Doebeli, M., Dieckmann, U., 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am. Nat. 156, S77–S101.

Durrett, R., Levin, S., 1994. The importance of being discrete (and spatial). Theor. Pop. Biol. 46, 363–394.

Durrett, R., Schweinsberg, J., 2004. Approximating selective sweeps. Theor. Pop. Biol. 66, 129–138.

Dynkin, E.B., 1991. Branching particle systems and superprocesses. Ann. Probab. 19, 1157–1194.

Economo, E.P., Kerkhoff, A.J., Enquist, B.J., 2005. Allometric growth, life-history invariants and population energetics. Ecol. Lett. 8, 353–360.

Etheridge, A., 2004. Survival and extinction in a locally regulated population. Ann. Appl. Probab. 14, 188–214.

Ethier, S.N., Kurtz, T.G., 1986. Markov Processes: Characterization and Convergence. Wiley, New York.

Evans, S.N., Perkins, E.A., 1994. Measure-valued branching diffusions with singular interactions. Can. J. Math. 46, 120–168.

Ewens, J.W., 2004. Mathematical population genetics. I. Theoretical introduction. Interdisciplinary Applied Mathematics. vol. 27, second ed. Springer, Berlin.

Ferrière, R., Gatto, M., 1995. Lyapunov exponents and the mathematics of invasion in oscillatory or chaotic populations. Theor. Popul. Biol. 48, 126–171.

Ferrière, R., Bronstein, J.L., Rinaldi, S., Law, R., Gauduchon, M., 2002. Cheating and the evolutionary stability of mutalism. Proc. R. Soc. Lond. B 269, 773–780.

Ferrière, R., Dieckmann, U., Couvet, D., 2004. An introduction to evolutionary conservation biology. In: Ferrière, R., Dieckmann, U., Couvet, D. (Eds.), Evolutionary Conservation Biology. Cambridge University Press, Cambridge, pp. 1–14.

Fournier, N., Méléard, S., 2004. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. Ann. Appl. Probab. 14, 1880–1919.

Freidlin, M.I., Wentzel, A.D., 1984. Random Perturbations of Dynamical Systems. Springer, Berlin.

Geritz, S.A.H., 2004. Resident-invader dynamics and the coexistence of similar strategies. J. Math. Biol. 50, 67–82.

Geritz, S.A.H., Kisdi, E., Meszena, G., Metz, J.A.J., 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol. Ecol. 12, 35–57.

Geritz, S.A.H., Gyllenberg, M., Jacobs, F.J.A., Parvinen, K., 2002. Invasion dynamics and attractor inheritance. J. Math. Biol. 44, 548–560.

Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M., Charnov, E.L., 2001. Effects of size and temperature on metabolic rate. Science 293, 2248–2251.

Gillooly, J.F., Allen, A.P., West, G.B., Brown, J.H., 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. Proc. Natl Acad. Sci. USA 102, 140–145.

Gosselin, F., 2001. Asymptotic behavior of absorbing Markov chains conditional on nonabsorption for applications in conservation biology. Ann. Appl. Probab. 11, 261–284.

Hairston Jr., N.G., Ellner, S.P., Geber, M.A., Yoshida, T., Fox, J.A., 2005. Rapid evolution and the convergence of ecological and evolutionary time. Ecol. Lett. 8, 1114–1127.

Haldane, J.B.S., 1932. The Causes of Evolution. Harper, London.

Hamilton, W.D., 1967. Extraordinary sex ratios. Science 156, 477–488.

Heino, M., Metz, J.A.J., Kaitala, V., 1998. The enigma of frequency-dependence. Trends Ecol. Evol. 13, 367–370.

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Oxford University Press, Oxford.

Hofbauer, J., Sigmund, K., 2003. Evolutionary game dynamics. Bull. Am. Math. Soc. 40, 479–519.

Jansen, V.A.A., Mulder, G.S.E.E., 1999. Evolving biodiversity. Ecol. Lett. 2, 379–386.

Joffe, A., Métivier, M., 1986. Weak convergence of sequences of semimartingales with applications to multitype branching processes. Adv. Appl. Probab. 18, 20–65.

Kaplan, N.L., Hudson, R.R., Langley, C.H., 1989. The hitch-hiking effect revisited. Genetics 123, 887–899.

Karatzas, I., Shreve, S., 1988. Brownian Motion and Stochastic Calculus. Springer, New York.

Kimura, M., 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl Acad. Sci. USA 54, 731–736.

Kimura, M., 1968. Evolutionary rate at the molecular level. Nature 217, 624–626.

Kirkpatrick, M., Rousset, F., 2004. Wright meets AD: not all landscapes are adaptive. J. Evol. Biol. 18, 1166–1169.

Kisdi, E., 1999. Evolutionary branching under asymmetric competition. J. Theor. Biol. 197, 149–162.

Lai, C., Mackay, T., 1990. Hybrid dysgenesis-induced quantitative variation the X chromosome of Drosophila melanogaster. Genetics 124, 627–636.

Law, R., Murrell, D.J., Dieckmann, U., 2003. Population growth in space and time: Spatial logistic equations. Ecology 84, 252–262.

MacLean, R.C., 2005. Adaptive radiation in microbial microcosms, J. Evol. Biol. 18, 1–11.

Matessi, C., Di Pasquale, C., 1996. Long-term evolution of multilocus traits. J. Math. Biol. 34, 613–653.

Matsuda, H., Ogita, N., Sasaki, A., Sato, K., 1992. Statistical mechanics of populations: The lattice Lotka-Volterra model. Progr. Theor. Phys. 88, 1035–1049.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Maynard Smith, J., Price, G., 1973. Logic of animal conflict. Nature 246, 15–18.

McCarthy, M.C., Enquist, B.J., 2005. Organismal size, metabolism and the evolution of complexity in metazoans. Evol. Ecol. Res. 7, 1–16.

McKane, A.J., Newman, T.J., 2004. Stochastic models in population biology and their deterministic analogs. Phys. Rev. E 70, 041902.

Méléard, S., Roelly, S., 1993. Sur les convergences étroite ou vague de processus à valeurs mesures. C. R. Acad. Sci. Paris Sér. I Math. 317, 785–788.

Metz, J.A.J., Nisbet, R.M., Geritz, S.A.H., 1992. How should we define 'fitness' for general ecological scenarios? Trends Ecol. Evol. 7, 198–202.

Metz, J.A.J., Geritz, S.A.H., Meszéna, G., Jacobs, F.A.J., van Heerwaarden, J.S., 1996. 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. North-Holland, Amsterdam, pp. 183–231.

Mukai, T., 1964. Polygenic mutation affecting quantitative character of Drosophila melanogaster. In: Mutations in Quantitative Traits. Proceedings of Gamma Field Symposium, vol. 3, Ministry of Agriculture, Japan, pp. 13–29.

Nagylaki, T., 1979. Dynamics of density- and frequency-dependent selection. Proc. Natl Acad. Sci. USA 76, 438–441.

Nowak, M.A., Sigmund, K., 2004. Evolutionary dynamics of biological games. Science 303, 793–799.

Page, K., Nowak, M.A., 2002. Unifying evolutionary dynamics. J. Theor. Biol. 219, 93–98.

Pigliucci, M., 2003. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Ecol. Lett. 6, 265–272.

Pimentel, D., 1968. Population regulation and genetic feedback. Science 159, 1432–1437.

Pomiankowski, A., Iwasa, Y., Nee, S., 1991. The evolution of costly mate preference I. Fisher and biased mutation. Evolution 45, 1422–1430.

Rand, D.A., Wilson, H.B., 1993. Evolutionary catastrophes, punctuated equilibria and gradualism in ecosystem evolution. Proc. R. Soc. London B 253, 137–141.

Rinaldi, S., Scheffer, M., 2000. Geometric analysis of ecological models with slow and fast processes. Ecosystems 3, 507–521.

Roelly, S., 1986. A criterion of convergence of measure-valued processes: Application to measure branching processes. Stoch. Stoch. Rep. 17, 43–65.

Rokyta, D.R., Joyce, P., Caudle, S.B., Wichman, H.A., 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nature Genet. 37, 441–444.

Schluter, D., 2000. The Ecology of Adaptive Radiation. Oxford University Press, Oxford.

Slatkin, M., 1979. Frequency- and density-dependent selection on a quantitative character. Genetics 93, 755–771.

Stadler, B.M.R., Stadler, P.F., Wagner, G.P., Fontana, W., 2001. The topology of the possible: Formal spaces underlying patterns of evolutionary change. J. Theor. Biol. 213, 241–274.

Stanley, S.M., 1979. Macroevolution: Pattern and Process. Freeman, San Francisco.

Stenseth, N.C., 1986. Darwinian evolution in ecosystems: A survey of some ideas and difficulties together with some possible solutions. In: Castri, J.L., Karlquist, A. (Eds.), Complexity, Language and Life: Mathematical Approaches. Springer, Berlin, pp. 105–145.

Thompson, J.N., 1998. Rapid evolution as an ecological process. Trends Ecol. Evol. 13, 329–332.

Turchin, P., 2003. Evolution in population dynamics. Nature 424, 257–258.

Walsh, J.B., 1984. An introduction to stochastic partial differential equations. École d'été de Saint-Flour. Lecture Notes in Mathematics, vol. 1180. Springer, Berlin, pp. 265–437.

Waxman, D., 2003. Numerical and exact solutions for continuum of alleles models. J. Math. Biol. 46, 225–240.

Wentzel, A.D., 1976a. Rough limit theorems on large deviations for Markov random processes, I. Theor. Probab. Appl. 21, 227–242.

Wentzel, A.D., 1976b. Rough limit theorems on large deviations for Markov random processes, II. Theor. Probab. Appl. 21, 499–512.

Wright, S., 1969. Evolution and the Genetics of Populations II. Chicago University Press, Chicago.

Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F., Hairston Jr., N.G., 2003. Rapid evolution drives ecological dynamics in a predator–prey system. Nature 424, 303–306.