Mathematical Biology

CrossMark

Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks

**Sylvain Billiard · Régis Ferrière ·
Sylvie Méléard · Viet Chi Tran**

Abstract How the neutral diversity is affected by selection and adaptation is investigated in an eco-evolutionary framework. In our model, we study a finite population in continuous time, where each individual is characterized by a trait under selection and a completely linked neutral marker. Population dynamics are driven by births and deaths, mutations at birth, and competition between individuals. Trait values influence ecological processes (demographic events, competition), and competition generates selection on trait variation, thus closing the eco-evolutionary feedback loop. The demographic effects of the trait are also expected to influence the generation and maintenance of neutral variation. We consider a large population limit with rare mutation, under the assumption that the neutral marker mutates faster than the trait under selec-

S. Billiard (✉)
Laboratoire EEP, UFR de Biologie, Université des Sciences
et Technologies Lille 1, Cité scientifique, 59655 Villeneuve d'Ascq Cedex, France
e-mail: sylvain.billiard@univ-lille1.fr

R. Ferrière
Eco-Evo-Math International Laboratory, Institut de Biologie de l'ENS,
CNRS INSERM ENS UMR 8197, 46 rue d'Ulm, 75230 Paris, France
e-mail: ferriere@biologie.ens.fr

R. Ferrière
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, USA

S. Méléard
CMAP, Ecole Polytechnique, CNRS, route de Saclay, 91128 Palaiseau Cedex, France
e-mail: sylvie.meleard@polytechnique.edu

V. C. Tran
Laboratoire P. Painlevé, UFR de Mathématiques, Université des Sciences
et Technologies Lille 1, Cité scientifique, 59655 Villeneuve d'Ascq Cedex, France
e-mail: chi.tran@math.univ-lille1.fr

tion. We prove the convergence of the stochastic individual-based process to a new measure-valued diffusive process with jumps that we call Substitution Fleming–Viot Process (SFVP). When restricted to the trait space this process is the Trait Substitution Sequence first introduced by Metz et al. (1996). During the invasion of a favorable mutation, a genetical bottleneck occurs and the marker associated with this favorable mutant is hitchhiked. By rigorously analysing the hitchhiking effect and how the neutral diversity is restored afterwards, we obtain the condition for a time-scale separation; under this condition, we show that the marker distribution is approximated by a Fleming–Viot distribution between two trait substitutions. We discuss the implications of the SFVP for our understanding of the dynamics of neutral variation under eco-evolutionary feedbacks and illustrate the main phenomena with simulations. Our results highlight the joint importance of mutations, ecological parameters, and trait values in the restoration of neutral diversity after a selective sweep.

1 Introduction

The science of biodiversity currently faces the challenge of understanding how ecological processes shape evolutionary change, and reciprocally how evolution affects the structure and function of ecological systems (Schoener 2011). Such eco-evolutionary feedbacks determine the dynamics of so-called adaptive traits -quantitative characters that are heritable yet mutable from parent to offspring (Dieckmann and Law 1996; Metz et al. 1996). Under the combined assumptions of large population and rare mutation scalings, the time evolution of an adaptive trait can be described as a sequence of mutant invasions, each being driven by positive selection in the ecological context set by the 'resident' value of the adaptive trait (Metz et al. 1992). The resulting evolutionary model is a jump process called the trait substitution sequence (TSS): every new mutant trait either goes extinct, or replaces the resident, causing the TSS to jump from the former resident population equilibrium to a new equilibrium (Metz et al. 1996; Champagnat 2006; Champagnat et al. 2008). In population genetics, these jumps are known as selective sweeps (Barton 1998; Stephan et al. 1992). Previous works support the view that the TSS as a model of long-term phenotypic evolution is relatively insensitive to the details of the genetic determination of the trait (Lloyd 1977; Christiansen and Loeschcke 1980; Hammerstein 1996; Weissing 1996; Matessi and Schneider 2009; Eshel et al. 1998).

Whereas eco-evolutionary feedbacks can result in variation of adaptive traits among populations (and even within populations when evolutionary branching occurs, Geritz et al. 1998), much of the molecular diversity measured by population geneticists involve DNA sequences of no known adaptive value, i.e. selectively neutral. A neutral sequence that is physically linked in the genome to the sequence that codes for the

adaptive trait is called a marker of that trait. A longstanding question in evolutionary theory is understanding how variation in such molecular markers evolves, and how patterns of neutral molecular evolution can be used to infer the history of trait mutation that have driven past adaptation.

When adaptive mutations are rare, adaptation proceeds as a series of selective sweeps: a trait mutation occurs while the population is monomorphic for the trait, and increases rapidly in frequency toward fixation. Following on from Kojima and Schaffer (1967), Smith and Haigh (1974) pointed out that selective sweeps purge genetic variation at linked sites: a particular marker allele goes to fixation as a consequence of linkage with the selected allele, a phenomenon they dubbed the 'hitchhiking effect'. Maynard Smith and Haigh's deterministic model was revisited in a stochastic approach by Ohta and Kimura (1975). These seminal studies of hitchhiking focused on the short-term dynamics of an interaction between two alleles at the locus under selection and two alleles at the neutral locus. Long-term dynamics were considered first by Kaplan et al. (1989) who developed a stochastic model for finite populations to describe the effect of recurrent hitchhiking. In order to describe stationary levels of nucleotide diversity at the marker locus, they used the infinite site model and a coalescent approach under the assumption of constant population size and constant selection coefficients. This has generated an abundant theoretical literature on modeling the impact of selection on neutral polymorphism (Barton 2000; Etheridge et al. 2006; Durrett and Schweinsberg 2004 and references therein). Recent deterministic models have relaxed the assumption of constant selection either because of the presence of genetic backgrounds (e.g. assuming a quantitative trait, Chevin and Hospital 2008) or in the case a parasite, because of the complexity of the demographic events involved in the life cycle (Schneider and Kim 2010). All previous models assume constant population size and constant selection, or that the population size is independent of the selective value of the individuals.

In this article, our goal is to relax these key assumptions. Under general ecological scenarios, eco-evolutionary feedbacks operate: as the adaptive trait evolves, population size and selection co-vary. The eco-evolutionary process of adaptive trait and neutral marker dynamics requires a rigorous mathematical framework, the foundation of which we establish here. We start with a 'microscopic', individual-based model where individuals have two heritable characteristics: (1) an adaptive trait that influences their intrinsic demographic rates and ecological interactions, and (2) a genetic marker that has no demographic or ecological effects, hence, is selectively neutral. This work focuses entirely on asexual populations and short genomic regions that remain perfectly linked to the loci under selection, neglecting recombination. The population is described by a measure according to which each individual is represented by a Dirac mass that weights its characters. This leads to study the population eco-evolutionary dynamics as a measure-valued stochastic process.

The dynamics are driven by competition between individuals, asexual reproduction without or with mutation, and death. Variation in population size and selection as the trait evolves are mediated by the demographic effects of change in the trait. These effects are expected to influence the generation and maintenance of neutral variation.

The effect of mutation on the marker can be continuous or discrete. Our framework thus encompasses a variety of conventional mutation models such as the two-alleles

model, the stepwise mutation model, and the continuous state mutation model. Our distinctive assumption here is that the marker mutation process is much faster than the trait mutation process but much slower than the ecological time-scale of birth and death events. This is supported by the fact that most mutations are neutral or nearly neutral (such as mutations involved in microsatellite variation). Therefore, there are three time scales in the model: the fast ecological time scale of birth and death events, the slow time scale of trait mutation, and an intermediate time scale of marker mutation. We study the joint process of trait and marker dynamics on the trait mutation time scale.

We are interested in limit theorems when the population carrying capacity goes to infinity. Then, the population size stabilizes in a neighborhood of the ecological equilibrium and jumps to another equilibrium when a successful trait mutant goes to fixation in the population. This is the TSS dynamics of the adaptive trait. It does not depend on the marker and has been mathematically proved by Champagnat (2006). The novelty in the model and in the proofs comes from the time-scales difference for the marker and trait mutations. The study of the marker distribution during the invasion period requires careful consideration of the individual process and of the different scales involved. In a first period, starting with the single invading mutant, we prove that the marker distribution remains close to a Dirac mass at the value of the initial mutant. Until the next jump of the TSS, the marker evolves as a stochastic distribution-valued process. In the case where the marker mutation effects are continuous and small, this is a Fleming–Viot process whose drift and covariance depend on the resident adaptive trait. In every cases, for any marker mutation model, the collated dynamics define a measure-valued diffusive process with jumps that we call Substitution Fleming–Viot Process (SFVP). The convergence of the microscopic process to the SFVP is shown both in the sense of finite dimensional distributions and in the sense of convergence of trait-marker-time measures, thus improving previous results of Champagnat (2006).

From a biological standpoint, we recover the conventional hitchhiking phenomenon: when a new mutant trait appears and sweeps through the population to fixation, the marker carried by the mutant individual is hitchhiked, and the marker distribution undergoes a genetical bottleneck. The mathematical construction of the SFVP process has new implications of biological relevance. Neutral diversity is restored after each adaptive jump, but as the adaptive trait evolves, population size, the mutation rate, genetic drift and demographic fluctuations change, which causes the rate of neutral polymorphism build-up and the moments of the marker distribution to change too. This suggests that the nature and structure of the whole eco-evolutionary feedback loop (i.e. how adaptive traits influence demographic rates and ecological interactions, and how ecological processes shape selection pressures on adaptive traits) may be important to explain the extreme disparities in genetic neutral diversity observed among species, even closely related ones and in the absence of differences in recombination profiles (Cutter and Payseur 2013). In fact, it is well-known that demographic differences due to external causes (demographic bottleneck or population expansion due to environmental changes) can affect neutral diversity of a population and that closely related species can show very different neutral diversity patterns. Here, we show that inter-

nal causes of demographic variation involved in adaptation can also affect species differently.

The article is organized as follows. In Sect. 2, we start with the model description. The stochastic individual-based process and its key assumptions are carefully described and examples are provided. A key parameter is K, an integer that gives the order of the population size and is used to rescale the mutation rates and kernels. By letting K go to infinity we study the large population limit of the stochastic process. The main theorem is enounced and discussed in Sect. 3, where biological implications are also highlighted. Time scale separations implied by the dependence in K of the trait and marker mutations lead to homogenization phenomena and then to the SFVP. Our mathematical analysis provides a precise description of the genetical bottleneck that occurs at each trait substitution. We show that the marker of the initial mutant individual dominates in the marker distribution of the mutant population until this population reaches a neighborhood of the new ecological equilibrium. Then, we present two numerical examples based on an ecological model adapted from Dieckmann and Doebeli (1999). In the first example, marker mutation is described by a continuous state model that leads to a piecewise Fleming–Viot process (Sect. 3.3) for the marker. In the second example, marker mutation follows a discrete two-allele model and the classical Wright–Fisher diffusion (3.5) is recovered. Further generalizations are discussed. The proof of the main theorem in the adaptive dynamics scaling is in Sect. 4. After having introduced a semi-martingale decomposition of our stochastic measure-valued process, we start with recalling and refining the result of Champagnat (2006) for the convergence of trait-marginals. For this purpose, we introduce the M1-topology on the Skorokhod space where the TSS lives, using some ideas of Collet et al. (2013). The second part of the proof focuses on the marker distribution in an invading mutant population. This gives the result on the genetical bottleneck. Then, between two trait substitutions, the dynamics of the marker converges to a diffusive measure-valued process. As a conclusion of the proof, we show the convergence to the SFVP in the space of trait-marker-time measures.

2 The stochastic model

We consider an asexual population driven by births and deaths where each individual is characterized by hereditary types: a phenotypic trait under selection and a neutral marker. The trait and marker spaces $\mathcal{X}$ and $\mathcal{U}$ are assumed to be compact subsets of $\mathbb{R}$. The type of individual i is thus a pair (x_i, u_i), $x_i \in \mathcal{X}$ being the trait value and $u_i \in \mathcal{U}$ its neutral marker. The individual-based microscopic model from which we start is a stochastic birth and death process with density-dependence whose demographic parameters are functions of the trait under selection and are independent of the marker. We assume that the population size scales with an integer parameter K tending to infinity while individuals are weighted with $\frac{1}{K}$. At any time $t \geq 0$, we have a finite number N_t^K of individuals, each of them holding trait and marker values in $\mathcal{X} \times \mathcal{U}$. Let us denote by $((x_1, u_1), \ldots, (x_{N_t^K}, u_{N_t^K}))$ the trait and marker values of these individuals. The state of the population at time $t \geq 0$, rescaled by K, is described by the point measure

$$v_t^K = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{(x_i,u_i)}, \tag{2.1}$$

where $\delta_{(x,u)}$ is the Dirac measure at (x,u). This measure belongs to the set of finite point measures on $\mathcal{X} \times \mathcal{U}$ with mass $1/K$. This set is a subset of the set $\mathcal{M}_F(\mathcal{X} \times \mathcal{U})$ of finite measures on $\mathcal{X} \times \mathcal{U}$, which is embedded with the weak convergence topology. We denote by $\langle v, f \rangle$ the integral of the measurable function f with respect to the measure v and by $\mathrm{Supp}(v)$ the support of v. Then $\langle v_t^K, 1 \rangle = \frac{N_t^K}{K}$.

For any $t \geq 0$, we also introduce the trait marginal of the measure v_t^K on $\mathcal{X}$, denoted by X_t^K and defined by

$$X_t^K = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{x_i}.$$

Therefore, the population measure v_t^K writes

$$v_t^K(dx, du) = X_t^K(dx)\, \pi_t^K(x, du) \tag{2.2}$$

where $\pi_t^K(x, du)$ is the marker distribution for a given trait value x defined by

$$\pi_t^K(x, du) = \frac{\sum_{i=1}^{N_t^K} \mathbb{1}_{x_i=x} \delta_{u_i}}{\sum_{i=1}^{N_t^K} \mathbb{1}_{x_i=x}}. \tag{2.3}$$

Our purpose is to study the asymptotic behavior of the measure-valued process v^K at large times, when the trait and marker are inherited but mutations occur. The main interest of our model is that these mutations happen at different time scales for trait and marker, but both longer than the individuals lifetime scale. The trait mutates much slower than the marker and drives the evolution time scale. Thus, the limiting behavior results from the interplay of three time scales: births and deaths, trait mutations and marker mutations.

We describe the individuals' life history. The trait has an influence on the ability of individuals to survive (including competition with other ones) and to reproduce but the marker is neutral. The demographic parameters are thus functions of the trait only and are defined on $\mathcal{X}$.

Assumption 2.1   • An individual with trait x and marker u reproduces with birth rate given by $0 \leq b(x) \leq \bar{b}$, the function b being continuous.
  • Reproduction produces a single offspring which usually inherits the trait and marker of its ancestor except when a mutation occurs. Mutations on trait and marker occur independently with probabilities p_K and q_K respectively. Mutations are rare and the marker mutates much more often than the trait. We assume that

$$q_K = p_K\, r_K, \quad \text{with } p_K = \frac{1}{K^2}, \quad q_K (\log K)^2 \to_{K \to \infty} 0, \quad r_K \to_{K \to \infty} +\infty. \tag{2.4}$$

- When a trait mutation occurs, the new trait of the descendant is $x + k \in \mathcal{X}$ with k chosen according to the probability measure $m(x, k)dk$.
- When a marker mutation occurs, the new marker of the descendant is $u + h \in \mathcal{U}$ with h chosen according to the probability measure $G_K(u, dh)$.
  For any $u \in \mathcal{U}$, $G_K(u, .)$ is approximated as follows when K tends to infinity:

$$\lim_{K \to +\infty} \sup_{u \in \mathcal{U}} \left| \frac{r_K}{K} \int_{\mathcal{U}} (\phi(u + h) - \phi(u)) G_K(u, dh) - A\phi(u) \right| = 0, \quad (2.5)$$

where $(A, \mathcal{D}(A))$ is the generator of a Feller semigroup and $\phi \in \mathcal{D}(A) \subseteq \mathcal{C}_b(\mathcal{U}, \mathbb{R})$, the set of continuous bounded real functions on $\mathcal{U}$.
- An individual with trait x and marker u dies with intrinsic death rate $0 \leq d(x) \leq \bar{d}$, the function d being continuous. Moreover the individual experiences competition the effect of which is an additional death rate $\eta(x) C * \nu_t^K(x) = \frac{\eta(x)}{K} \sum_{i=1}^{N_t^K} C(x - x_i)$. The quantity $C(x - x_i)$ describes the competition pressure exerted by an individual with trait x_i on an individual with trait x. We assume that the functions C and η are continuous and that there exists $\underline{\eta} > 0$ such that

$$\forall x, y \in \mathcal{X}, \quad \eta(x) C(x - y) \geq \underline{\eta} > 0. \quad (2.6)$$

A classical choice of competition function is $C \equiv 1$ which is called "mean field case" or "logistic case". In that case the competition death rate is $\eta(x) N_t^K / K$.

Remark 2.2 Let us insist on the generality of Assumption (2.5) which allows a larger set of possible dynamics.

- Equation (2.5) is for example true for $\mathcal{U} = [u_1, u_2]$, G_K a centered Gaussian law (conditioned to $\mathcal{U}$) with variance $\sigma_K \to 0$ such that $\lim_K \sigma_K^2 \frac{r_K}{K} = \sigma^2$ and $A\phi = \frac{\sigma^2}{2} \phi''$ for $\phi \in \mathcal{C}^2$ with $\phi'(u_1) = \phi'(u_2) = 0$.
  Choosing for example $r_K = K^{3/2}$, $q_K = 1/\sqrt{K}$ and $\sigma_K^2 = 1/\sqrt{K}$ works. This choice can be seen as a continuous state generalization of the stepwise mutation model (Ohta and Kimura 1973).
- If in addition the distribution G_K has a non zero mean μ_K such that $\frac{r_K \mu_K}{K} \to \mu > 0$ corresponding to a mutational directional drift, then the operator A will be defined by $A\phi = \frac{\sigma^2}{2} \phi'' + \mu \phi'$.
- If we relax the compactness of $\mathcal{U}$ and assume that $\mathcal{U} = \mathbb{R}$, a third example consists in taking for G_K the law of a Pareto variable with index $\alpha \in (1, 2)$ divided by $K^{\eta/\alpha}$, for $\eta \in (0, 1]$. Then it has been proved by Jourdain et al. (2012) that

$$\lim_K \sup_u \left| K^\eta \int_{\mathbb{R}} (\phi(u + h) - \phi(u)) G_K(u, dh) - \frac{\alpha}{2} D^\alpha \phi(u) \right| = 0,$$

where

$$D^\alpha \phi(u) = \int_{\mathbb{R}} (\phi(u + h) - \phi(u) - h\phi'(u) \mathbb{1}_{|h| \leq 1}) \frac{dh}{|h|^{1+\alpha}}$$

is the fractional Laplacian with index α. Thus if we take r_K such that $\frac{r_K}{K^{1+\eta}}$ converges as K tends to infinity, and choose $A = D^\alpha$ in (2.5), Assumptions (2.4)–(2.5) will be satisfied as soon as $\eta < 1$.

- Another very interesting case is the discrete case when $\mathcal{U} = \{a, A\}$ is a set of two alleles. The mutation kernel is given by

$$G_K(u, dv) = \mathbb{1}_{u=a}\, q_a\, \delta_A(dv) + \mathbb{1}_{u=A}\, q_A\, \delta_a(dv). \tag{2.7}$$

In this case, (2.5) implies that r_K/K has a limit when $K \to +\infty$. Let $\bar{r}$ be this limit, then

$$A\phi(u) = \bar{r}\Big(\mathbb{1}_{u=a}\, q_a\big(\phi(A) - \phi(a)\big) + \mathbb{1}_{u=A}\, q_A\big(\phi(a) - \phi(A)\big)\Big). \tag{2.8}$$

We see that the ratio between the two mutation probabilities $r_K = q_K/p_K$ that allows convergence is highly dependent on the mutation distribution.

Note that since the demographic rates do not depend on the marker, the dynamics of the population distribution of the trait is independent of the marker distribution. But the dynamics of the marker distribution cannot be separated from the trait distribution as we shall see.

The process $(\nu_t^K, t \geq 0)$ is a càdlàg $\mathcal{M}_F(\mathcal{X} \times \mathcal{U})$-valued Markov process. Existence and uniqueness in law of the process can be adapted from Fournier and Méléard (2004) and Champagnat et al. (2008) under the assumption that $\mathbb{E}(\langle \nu_0^K, \mathbf{1}\rangle) < +\infty$. Moreover, Assumption (2.6) allows to prove as by Champagnat (Champagnat 2006, Lemma 1) that if for $p \geq 1$, $\sup_{K \in \mathbb{N}^*} \mathbb{E}(\langle \nu_0^K, \mathbf{1}\rangle^p) < +\infty$, then

$$\sup_{t \in \mathbb{R}_+, K \in \mathbb{N}^*} \mathbb{E}\big(\langle \nu_t^K, \mathbf{1}\rangle^p\big) < +\infty \tag{2.9}$$

which will be useful to study the tightness and convergence of the sequence.

3 Convergence to the Substitution Fleming–Viot Process

The adaptive trait mutation time scale is the slowest, equal to $\frac{1}{K p_K} = K$ by Assumptions (2.4). It scales the evolutionary time. So we shall consider the limiting behavior of $(\nu_{Kt}^K, t \geq 0)$. We will see in Sect. 4.2.1 that p_K of order $1/K^2$ is the only choice which leads to a non-trivial or non-degenerate marker dynamics.

Before stating our main result, we introduce several important ingredients which are used to describe the limit of $(\nu_{Kt}^K, t \geq 0)$ when $K \to +\infty$. We conclude the section with extensions and simulations.

3.1 Invasion fitness function

The large population behavior of the process $(\nu_t^K, t \geq 0)$ as K tends to infinity, can be studied by classical arguments and is given in the appendix. At the ecological time scale (of order 1), no mutation occurs in the asymptotic $K \to +\infty$. If the

initial population has a single adaptive trait x, then, in the limit $K \to +\infty$, the trait distribution remains δ_x since p_K and q_K vanish in the limit. The rescaled population size process $(N_t^K/K, t \geq 0)$ converges to the solution $(n_t, t \geq 0)$ of the ordinary differential equation

$$\frac{dn_t}{dt} = \big(b(x) - d(x) - \eta(x)C(0)n_t\big)n_t \tag{3.1}$$

which converges when t tends to infinity to the equilibrium

$$\widehat{n}_x = \frac{b(x) - d(x)}{\eta(x)C(0)}. \tag{3.2}$$

Conversely, at the adaptive trait-mutation time scale Kt, new mutant traits can invade. If they replace the previous traits, then the corresponding event is called "fixation".

The probability of fixation of a mutant trait y in a trait resident population x at equilibrium depends on the invasion fitness function $f(y; x)$:

$$f(y; x) = b(y) - d(y) - \eta(y)\, C(y - x)\,\widehat{n}_x. \tag{3.3}$$

This fitness function describes the initial growth of the mutant population. It does not depend on the neutral marker.

By simplicity we work under the assumption of 'invasion implies fixation', but this assumption will be relaxed in Sect. 3.5. When a mutant trait appears, either its line of descent replaces the resident population or it disappears. As a consequence, two traits cannot coexist in the long term.

Assumption 3.1 ("*Invasion implies fixation*") For all $x \in \mathcal{X}$ and for almost every $y \in \mathcal{X}$,

$$\text{either } \frac{b(y) - d(y)}{\eta(y)C(y - x)} < \frac{b(x) - d(x)}{\eta(x)C(0)},$$
$$\text{or } \frac{b(y) - d(y)}{\eta(y)C(y - x)} > \frac{b(x) - d(x)}{\eta(x)C(0)} \text{ and } \frac{b(x) - d(x)}{\eta(x)C(x - y)} < \frac{b(y) - d(y)}{\eta(y)C(0)}.$$

Remark 3.2 In the case of logistic populations with $C \equiv 1$, this assumption is satisfied as soon as $x \mapsto \widehat{n}_x$ is strictly monotonous.

3.2 Main theorem

Let us first give the definition of the Fleming–Viot process which will appear in our setting (see e.g. Dawson and Hochberg 1982; Dawson 1993; Donnelly and Kurtz 1996; Etheridge 2000). We recall that the operator A has been introduced in (2.5).

In the sequel, we denote by $\mathcal{P}(\mathcal{U})$ and $\mathcal{P}(\mathcal{X} \times \mathcal{U})$ the probability measure spaces respectively on $\mathcal{U}$ and on $\mathcal{X} \times \mathcal{U}$.

Definition 3.3 Let us fix $x \in \mathcal{X}$ and $u \in \mathcal{U}$. The Fleming–Viot process $(F_t^u(x, .), t \geq 0)$ indexed by x, started at time 0 with initial condition δ_u and associated with the mutation operator A is the $\mathcal{P}(\mathcal{U})$-valued process whose law is characterized as the unique solution of the following martingale problem. For any $\phi \in \mathcal{D}(A)$,

$$M_t^x(\phi) = \langle F_t^u(x, .), \phi \rangle - \phi(u) - b(x) \int_0^t \langle F_s^u(x, .), A\phi \rangle ds \qquad (3.4)$$

is a continuous square integrable martingale with quadratic variation process

$$\langle M^x(\phi) \rangle_t = \frac{b(x) + d(x) + \eta(x)C(0)\widehat{n}_x}{\widehat{n}_x} \int_0^t \left(\langle F_s^u(x, .), \phi^2 \rangle - \langle F_s^u(x, .), \phi \rangle^2 \right) ds$$

$$= \frac{2b(x)}{\widehat{n}_x} \int_0^t \left(\langle F_s^u(x, .), \phi^2 \rangle - \langle F_s^u(x, .), \phi \rangle^2 \right) ds. \qquad (3.5)$$

Let us now state our main theorem that describes the slow–fast dynamics of adaptive traits and neutral markers at the (trait) evolutionary time scale.

Theorem 3.4 *We work under Assumptions* 2.1 *and* 3.1. *The initial conditions are* $v_0^K(dy, dv) = n_0^K \delta_{(x_0, u_0)}(dy, dv)$ *with* $\lim_{K \to \infty} n_0^K = \widehat{n}_{x_0}$ *and for any* $\epsilon > 0$, $\sup_{K \in \mathbb{N}^*} \mathbb{E}((n_0^K)^{2+\epsilon}) < +\infty$.
Then, the population process $(v_{Kt}^K, t \geq 0)$ *converges in law to the* $\mathcal{M}_F(\mathcal{X} \times \mathcal{U})$-*valued process* $(V_t(dy, dv), t \geq 0)$ *defined by*

$$V_t(dy, dv) = \widehat{n}_{Y_t} \, \delta_{Y_t}(dy) \, F_t^{U_t}(Y_t, dv), \qquad (3.6)$$

where the process $((Y_t, U_t), t \geq 0)$ *on* $\mathcal{X} \times \mathcal{U}$, *started at* (x_0, u_0), *jumps at time t from* (x, u) *to* $(x + k, v)$ *with the jump measure*

$$b(x)\widehat{n}_x \frac{[f(x+k; x)]_+}{b(x+k)} F_t^u(x, dv) \, m(x, k)dk. \qquad (3.7)$$

The convergence holds in the sense of finite dimensional distributions on $\mathcal{M}_F(\mathcal{X} \times \mathcal{U})$.
In addition, the convergence also holds in the space of trait-marker-time measures, i.e. the measure $v_{Kt}^K(dy, dv)dt$ *on* $\mathcal{X} \times \mathcal{U} \times [0, T]$ *converges weakly to the measure* $\widehat{n}_{Y_t} \, \delta_{Y_t}(dy) \, F_t^{U_t}(Y_t, dv)dt$ *for any* $T > 0$. $\qquad \square$

Definition 3.5 The limiting measure-valued process $(V_t(dy, dv), t \geq 0)$ is called Substitution Fleming–Viot Process. It generalizes the Trait Substitution Sequence (TSS) introduced by Metz et al. (1996).

We observe that the Substitution Fleming–Viot Process includes the three qualitative behaviors due to the three different time scales: deterministic equilibrium for the size of the population (driven by the ecological birth and death events), transitory diffusive behavior for the marker distribution (driven by marker mutation), jump process for the trait distribution (driven by adaptive trait mutation).

Remark 3.6 Equations (3.4)–(3.5) have important biological implications regarding neutral genetic diversity. Once the fixation of a favorable mutation has occurred and the population is monomorphic for the selected trait, the evolution of the neutral marker distribution is described by a Fleming–Viot process whose law is given by the martingale in (3.4). The bracket of the martingale in (3.5) shows that the stochastic fluctuations with time of the marker distribution are due to randomness in births and deaths and mutations. The multiplicative factor $2b(x)/\widehat{n}_x$ in (3.5) depends on the trait value x, and on the assumed ecological model which determines the relationships between x, the death and birth rates and the competition kernel. Notice that $2b(x)/\widehat{n}_x$ corresponds to the quotient of variance (here $2b(x)$) divided by effective size N_e (here $\widehat{n}_x$) that appears in the usual Wright Fisher equation. The quantity $\widehat{n}_x$ corresponds to the mass of the population when there is an infinite number of small individuals; if the size of the population is of order K, it means that there is approximately $\widehat{n}_x K$ individuals of weights $1/K$. The right term in (3.4), (i.e. the drift term in a mathematical sense) involves the generator A and is associated with the mutation model as seen in Assumption (2.5). The generator A describes the speed at which the neutral diversity is restored. For instance in a continuous state model, if $A\phi = \frac{\sigma^2}{2}\phi''$, we recover the heat equation whose solutions have a variance in t. In a discrete state model similar to (2.8), this equation gives the growth of the support.

In short, (3.4)–(3.5) shows that the distribution of the neutral marker depends on ecological processes and their parameters: every changes in x will result in changes in the distribution of the neutral marker, through changes in birth, death and mutation rates, in competition and equilibrium population size. This result is biologically relevant and important since it differs from the assumptions of classical genetic hitchhiking models, in which selection and population size remain constant, leading to the fact that the neutral diversity restoration will not depend on the trait substitution and its history. In examples below, we will give more detailed results regarding the distribution of the neutral marker changes.

The proof of Theorem 3.4 is the subject of Sect. 4.

The trait dynamics in the limit of Theorem 3.4 is the Trait Substitution Sequence obtained by Theorem 1 in Champagnat (2006) whose assumptions are satisfied. Our main contribution in Theorem 3.4 is to prove that at the adaptive trait mutation time scale, a homogeneization phenomenon takes place. There is a deterministic limit for the fastest process (the births and deaths leading to $\widehat{n}_x$), and stochastic limits for the two slower processes. The limiting process $(V_t, t \geq 0)$ is a measure-valued process with jumps (corresponding to trait mutations) and diffusion (corresponding to marker dynamics). If the population is trait-monomorphic with trait x, the jump measure is

$$b(x)\widehat{n}_x \int_{\mathcal{X}-\{x\}} \frac{[f(x+k; x)]_+}{b(x+k)} m(x, k)dk,$$

where $\mathcal{X} - \{x\} = \{y - x, \ y \in \mathcal{X}\}$. When a jump occurs at t, the process jumps from (x, u) to $(x + k, v)$ where k is chosen in $m(x, k)dk$ and v is chosen at time t in the marker distribution $F_t^u(x, dv)$.

The marker distribution is the second fastest-evolving component, but marker muta-
tions are assumed small (2.5), allowing to recover a non-degenerate Fleming–Viot
superprocess parameterized by the trait of the population but with jumps. Between the
jumps, this superprocess is the pathwise limit of the marker dynamics where traits are
fixed. The jumps are hitchhiking events due to the trait mutations (see in another con-
text Etheridge et al. 2006). There is a bottleneck at each successful invasion-fixation
of mutant traits. Indeed, the individuals present at the fixation time are all descendants
of the successful initial mutant. The trait and marker of the latter alone determine
the state of the new mutant population, hence creating the bottleneck for the whole
population genealogy. This result is biologically intuitive since we assume that the
neutral marker and the trait are completely linked, but the mathematical proof of these
phenomena is the hardest part of the proof of Theorem 3.4, and we will show that our
results still have biological interest. Extending this model to the case of recombination
is a challenging problem for future work (see Smadi 2014 in this direction).

It is also worth to notice that contrarily to other extensions of the TSS (e.g. the TSS
with age-structure of Méléard and Tran 2009 or the Polymorphic Evolution Sequence
for a multi-resource chemostat in Champagnat et al. 2014) that usually jump from
an equilibrium to another, the marker distribution is here described by a stochastic
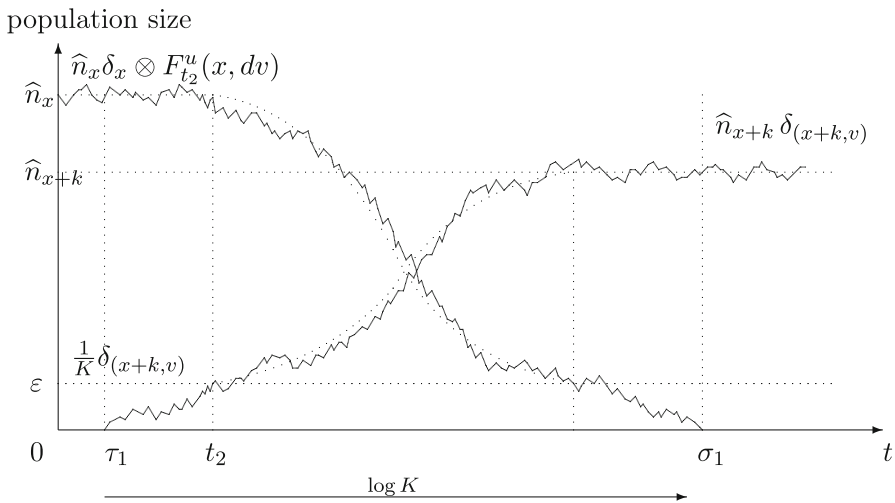
Fig. 1 Invasion and fixation of a successful trait mutant. In the population of resident trait x and marker
distribution $F_t^u(x, dv)$, a mutant trait $x + k$ appears at time τ_1. Let v be the marker of the mutant individ-
ual. As in Champagnat et al. (2008), the fluctuations of the resident population can be neglected in first
approximation and the mutant population evolves as a birth and death process with rates $b(x + k)$ and
$d(x + k) + \eta(x + k)C(k)\widehat{n}_x$, independent of the marker distribution. When the mutant population reaches
a sufficient size ε at time t_2, with probability $[f(x + k, x)]_+/b(x + k)$, the 'invasion implies fixation'
assumption leads to the replacement of the former population in a time t_K such that $t_K/\log(K) \to \infty$.
This time interval is too short to allow other marker mutant to appear in non-negligible proportion, with
large probability. Thus, when the mutant population has fixed, at time σ_1, it is close to $\widehat{n}_{x+k}\delta_{(x+k,v)}$. Before
the next adaptive trait mutation occurs, the marker mutates a lot, since marker mutations happen on a faster
scale. The dynamics of the marker distribution is then the one of a Fleming–Viot superprocess started at δ_v
and with statistics depending on $x + k$

process and not an equilibrium measure. This is due to the fact that the time scales of the trait and marker mutations are assumed different: in the time scale of marker mutations, the trait mutations are too rare and not seen.

An illustration of the invasion and fixation phenomena is summed up in Fig. 1.

3.3 An example from Dieckmann and Doebeli (1999)

Let us first illustrate our model by simulations based on an example inspired from Roughgarden (1979) and Dieckmann and Doebeli (1999). Here $\mathcal{X} = [-1, 1], \mathcal{U} = [-2, 2]$ and $K = 1,000$ in all the simulations. The individual dynamics is characterized by

- the birth rate $b(x) = \exp(-x^2/2\sigma_b^2)$ with $\sigma_b = 0.9$. The probability of mutation of the trait and marker are respectively $p_K = 1/K^2$ and $q_K = 1/\sqrt{K}$. The adaptive trait mutation kernel $m(x, k)dk$ is a Gaussian law with mean 0 and variance 0.1, conditioned to $[-1, 1]$. The marker mutation kernel $G_K(u, dh)$ is a Gaussian law with mean 0 and variance $\sigma_K^2 = 1/\sqrt{K}$, conditioned to $[-2, 2]$.
- symmetric competition for resources, with $\eta(x) = 1$ and $C(x - y) = \exp(-(x - y)^2/2\sigma_C^2)$, $\sigma_C = 0.8$.

Here, the 'optimal trait' is $x = 0$ where the birth rate has its maximum and the population is governed by local competition. We start with the initial condition: $x_0 = -1, u_0 = 0$.

The simulations (see Fig. 2) illustrate Theorem 3.4. They show the replacement of a resident population by a mutant population. In Fig. 2a, the dynamics of the support of the marker distribution is represented. The mutant and resident populations are pictured together and separately to better observe the extinction of the resident population (black) and the expansion of the mutant population from one individual (light). The invasion started around time 3,175 and after time 3,250, the mutant population has totally replaced the resident one.

In Fig. 2b, the histograms of traits and markers at three times during the invasion are represented simultaneously, to underline the hitchhiking effect of the marker during the 'invasion implies fixation' phase. We can see that the distribution of the marker during the fixation remains close to a Dirac mass at the marker value of the initial mutant (red line). This illustrates the bottleneck phenomenon, the existence of which we prove rigorously [Eq. (4.8) of Proposition 4.5].

Let now focus on the Dieckmann–Doebeli's example and highlight the biological implications regarding the eco-evolutionary feedback on the distribution of the neutral marker. Here, $\widehat{n}_x = b(x) - d(x)$ and therefore the Fleming–Viot process $F_t^u(x, .)$ is the solution of the martingale problem (3.4) with $A\phi = \frac{\sigma^2}{2}\phi''$ and with bracket (3.5) given for all $\phi \in \mathcal{C}(\mathcal{U}, \mathbb{R})$ by $2\frac{b(x)}{b(x)-d(x)} \int_0^t \left(\langle F_s^u(x, .), \phi^2 \rangle - \langle F_s^u(x, .), \phi \rangle^2 \right) ds$.

If the death rate is a constant $d(x) = d$, then the multiplicative factor in the bracket (3.5), $b(x)/(b(x) - d)$, decreases when $b(x)$ increases. Heuristically we expect that the stochastic fluctuations in time of the distribution of the neutral marker decrease when the trait x approaches the evolutionary stable strategy (ESS, see Maynard Smith 1982) and $b(x)$ increases, since the equilibrium size is greater and the diffusion coeffi-
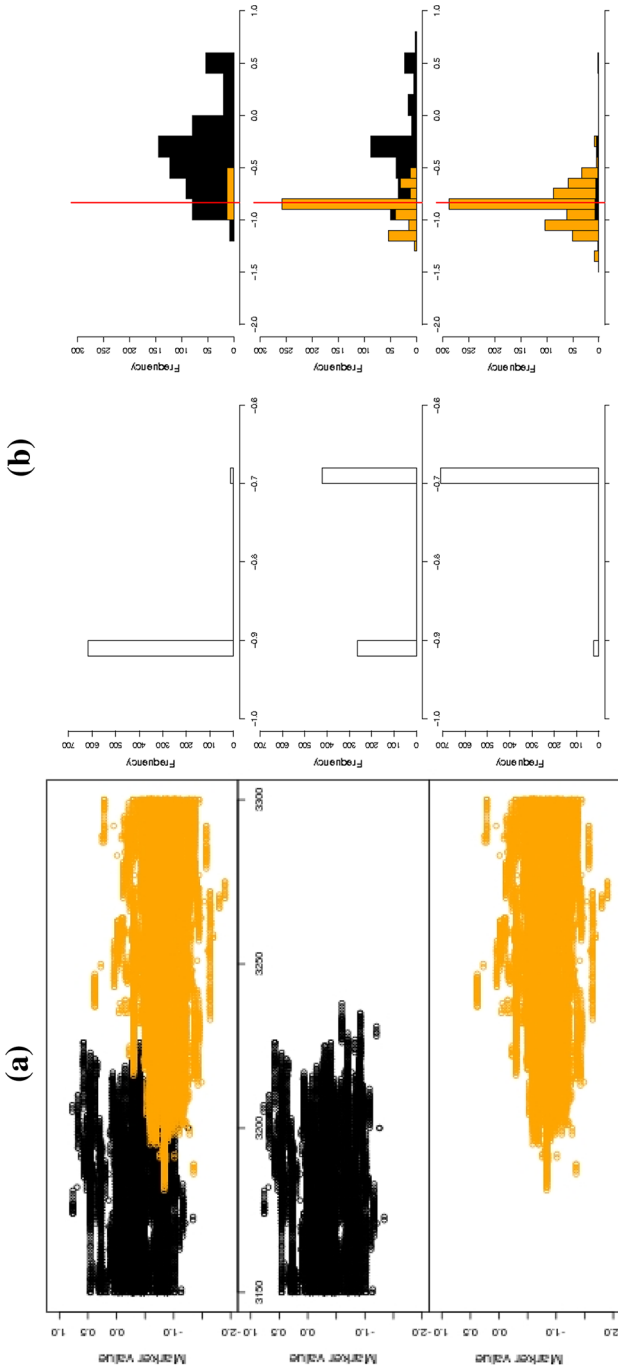
Fig. 2 We consider a resident trait-monomorphic population (*black*) in which a mutant trait (*light*) appears and goes to fixation. Here the intrinsic death rate $d(x) = 0$. **a** Evolution of the support of the marker distribution with respect to time; the support of the mutant population is *lighter*. The mutant and resident populations are shown separately and together. **b** Distributions of the traits (*left*) and markers (*right*) in the population at three times during the invasion. The marker and trait values are in abscissa and frequencies are in ordinate. The marker value of the initial mutant is indicated by the *red line*. When the mutant trait appears, the resident population is quickly invaded by the mutant population during a transition period. In **b**, we can see that if the support of the marker distribution for the resident population remains wide (see also **a**), the size of the resident population decreases quickly. In the *second column* of (**b**), we see that the marker distribution in the mutant population remains spiked at the marker value of the first mutant individual during the whole transition period. After invasion (see **a**), the spread of the marker distribution follows the Fleming–Viot process (3.4)–(3.5). On **a**, we see that for the Fleming–Viot process, the support of the marker distribution spreads slowly

cient is lower. The drift term is $b(x)\frac{\sigma^2}{2}\int_0^t \langle F_s^u(x, .), \phi'' \rangle ds$ and thus the multiplicative factor $b(x)$ increases when approaching the ESS, contrarily to the multiplicative factor of the bracket (3.5). In the case $d \equiv 0$, the Fleming–Viot process has a constant diffusion coefficient and the bracket (3.5) does not depend on x. The Fleming–Viot process depends only on the trait x through the drift term. Notice that this is true for any mutation model satisfying (2.5). This simple result illustrates how the ecological processes can shape the neutral diversity.

3.4 Corollary: the Wright–Fisher evolutionary process

There exists a version of the SFVP in the case when the marker space $\mathcal{U}$ is discrete. Assume for instance that there exist only two alleles of the marker trait, denoted by a and A, so that $\mathcal{U} = \{a, A\}$. In this case, we apply Theorem 3.4 with the mutation kernel G_K defined in (2.7) and $r_K / K \rightarrow \bar{r} > 0$ when $K \rightarrow +\infty$.

Proposition 3.7 *We work under Assumptions 2.1 and 3.1 with probabilities q_A and q_a to mutate from marker A to marker a and from marker a to marker A. Moreover, we consider similar initial conditions v_0^K as in Theorem 3.4. Then, the population process $(v_{Kt}^K, t \geq 0)$ converges in law to the $\mathcal{M}_F(\mathcal{X} \times \{a, A\})$-valued process*

$$(\widehat{n}_{Y_t} \left(W_t^a \, \delta_{(Y_t, a)}(dy, du) + (1 - W_t^a) \, \delta_{(Y_t, A)}(dy, du)\right), t \geq 0),$$

where $(Y_t, t \geq 0)$ is the TSS process that jumps from x to $x + k$ in $\mathcal{X}$ with the jump measure $b(x)\widehat{n}_x \frac{[f(x+k;x)]_+}{b(x+k)} m(x, k)dk$ and where $(W_t^a, t \geq 0)$ is the following Wright–Fisher jump process that represents the proportion of alleles a in the population of trait Y_t at time t. Between jumps, it satisfies the usual Wright–Fisher equation with mutations

$$dW_t^a = \bar{r}\, b(Y_t)\big(q_A(1 - W_t^a) - q_a W_t^a\big)dt + \sqrt{\frac{2b(Y_t)}{\widehat{n}_{Y_t}}}\, W_t^a\, (1 - W_t^a)dB_t \quad (3.8)$$

$(B_t, t \geq 0)$ being a standard Brownian motion. It jumps with the TSS and at jump time t, the process $(W_t^a, 1 - W_t^a)$ goes to $(1, 0)$ with probability W_t^a and to $(0, 1)$ with probability $1 - W_t^a$. □

An illustration of this proposition is given in Fig. 3.

This result can be generalized to discrete marker spaces $\mathcal{U} = \{a_1, \ldots a_m\}$, by introducing the transition probabilities q_{ij} to mutate from a_i to a_j, $i, j \in \{1, \ldots, m\}$. An application is when the marker corresponds to the genetical sequence of n nucleotides (A, T, G or C for each position). In this case, $m = \text{Card}\,\mathcal{U} = 4^n$.

Traditionally in a population genetics framework, the evolution in finite populations of the diversity at a neutral marker is described as a diffusion process with two fixed parameters: the population size and the mutation "rate" (e.g. Crow and Kimura 1970). The population size is related to what is called the "genetic drift" and generally refers to the random sampling of gametes performed for reproduction at the beginning
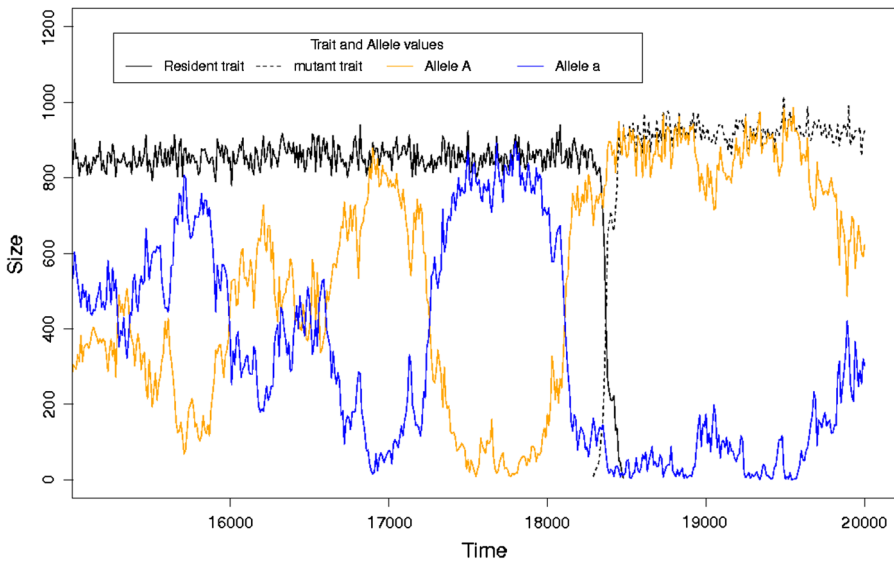
Fig. 3 Evolution of sizes of the subpopulations with markers a and A. The simulation uses individual-based algorithms. The proportions of marker alleles a and A follow Wright–Fisher diffusions while the size of the population stabilizes around the equilibrium given by the trait value. A trait mutant appears around time 18,290, invades and fixes into the population. Before the appearance of this mutant trait, fluctuations in the marker distribution are due to (fast) marker mutation, birth and death stochastic events. At the time when the mutant trait appears, the A-allele frequency is 85 %, giving a high probability for an A-allele hitchhike. This is the case in the simulation. After fixation time (around time 18,490), the a-allele population is extinct. It is regenerated by mutations of the marker but get extinct three times before taking up around time 19,600

of each generation, and the higher the population size, the lower the genetic drift. Under this framework, genetic drift induces stochastic fluctuations in the frequencies of the alleles A and a and can cause the decrease of neutral genetic diversity when an allele is randomly lost. On the other hand, mutation introduces continuously alleles A and a in the population and thus allows the restoration and the maintenance of the neutral genetic diversity. It is important to note that under the population genetics framework, mutation rates and population size are fixed and do not depend on the ecological processes and their parameters, neither on the trait value when the population is monomorphic for the trait under selection. As a consequence, those parameters do not change as successive selective sweeps occur especially during the adaptation process. Here we can use Eq. (3.8) and try to compare the classical population genetics results about the distribution of the neutral genetic diversity and the one in our model.

In an eco-evolutionary framework, (3.8) first shows that mutation rates and population size, i.e. the genetic drift, are not fixed and depend on the ecological processes and on the trait value x. The mutation rates are $\bar{r}\, b(Y_t)q_A$ and $\bar{r}\, b(Y_t)q_a$ in our framework while it is only q_A and q_a under a population genetics framework (e.g. Crow and Kimura 1970). The genetic drift, i.e. the equilibrium population size, is given by $1/\widehat{n}_{Y_t}$ while it is a constant $1/n$ in population genetics framework. Second, (3.8) shows that extra ecological processes affect the distribution of the neutral marker since in the left-hand side there is the term $2b(Y_t)$. This term can be interpreted as the effect of demographic stochasticity, which is not taken into account in population genetics.

3.5 Extensions to co-existing traits

The work of Champagnat and Méléard (2011) generalizes the TSS to the case of coexisting trait values, when Assumption 3.1 is relaxed. They define a polymorphic TSS called polymorphic evolutionary sequence (PES) and denoted by $(X_t)_{t \geq 0} \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathcal{X}))$. When a mutant trait y appears in a resident population of trait x_0 at time t_1, either its descendent line is killed with probability $1 - [f(y; x_0)/b(y)]_+$, or it survives. In that case, we can have coexistence of y and x_0 when there is a positive globally stable non-trivial equilibrium $(n^*_{x_0, y}, n^*_{y, x_0})$ to the Lotka–Volterra system defined in (5.4). Therefore the population jumps from $X_{t_1-} = \widehat{n}_{x_0} \delta_{x_0}$ to

$$X_{t_1} = n^*_{x_0, y} \delta_{x_0}(dx) + n^*_{y, x_0} \delta_y(dx).$$

For a probability π, a trait measure X and $x \in \mathcal{X}$, let us denote by $F_t(\pi, x, X, du)$ the Fleming–Viot process started at π, evolving in the trait distribution X and parameterized by x.

Let π_0 be the initial marker distribution of the monomorphic population of trait x_0. Before the time t_1 of appearance of the first mutant, the marker distribution evolves as $(F_t(\pi_0, x_0, \widehat{n}_{x_0} \delta_{x_0}, du))_{t \geq 0}$. Let $\pi_{t_1} = F_{t_1}(\pi_0, x_0, \widehat{n}_{x_0} \delta_{x_0}, du)$ be the marker distribution at t_1 and let V_1 be a random variable drawn in the distribution π_{t_1}. After t_1 and before the occurence of the second trait-mutation at t_2, the population evolves as

$$n^*_{x_0, y} \delta_{x_0}(dx) F_{t-t_1}(\pi_{t_1}, x_0, X_{t_1}, du) + n^*_{y, x_0} \delta_y(dx) F_{t-t_1}(\delta_{V_1}, y, X_{t_1}, du).$$

The processes $F_t(\pi_{t_1}, x_0, X_{t_1}, du)$ and $F_t(\delta_{V_1}, y, X_{t_1}, du)$ are independent generalizations of the Fleming–Viot process defined in Definition 3.3 conditionally on π_{t_1}, X_{t_1} and V_1. Indeed their semimartingale decompositions are respectively:

$$\langle F_t(\pi_{t_1}, x_0, X_{t_1}, .), \phi \rangle = \langle \pi_{t_1}, \phi \rangle + b(x_0) \int_0^t \langle F_s(\pi_{t_1}, x_0, X_{t_1}, .), A\phi \rangle \, ds + M_t^1(\phi);$$

$$\langle F_t(\delta_{V_1}, y, X_{t_1}, .), \phi \rangle = \phi(V_1) + b(y) \int_0^t \langle F_s(\delta_{V_1}, y, X_{t_1}, .), A\phi \rangle \, ds + M_t^2(\phi),$$

$$(3.9)$$

where $M^1(\phi)$ and $M^2(\phi)$ are independent square integrable martingales such that

$$\langle M^1(\phi) \rangle_t = \frac{b(x_0) + d(x_0) + \eta(x_0)C(0)n^*_{x_0, y} + \eta(x_0)C(x_0 - y)n^*_{y, x_0}}{n^*_{x_0, y} + n^*_{y, x_0}}$$

$$\times \int_0^t \left(\langle F_s(\pi_{t_1}, x_0, X_{t_1}, .), \phi^2 \rangle - \langle F_s(\pi_{t_1}, x_0, X_{t_1}, .), \phi \rangle^2 \right) ds$$

$$= \frac{2b(x_0)}{n^*_{x_0, y} + n^*_{y, x_0}} \int_0^t \left(\langle F_s(\pi_{t_1}, x_0, X_{t_1}, .), \phi^2 \rangle - \langle F_s(\pi_{t_1}, x_0, X_{t_1}, .), \phi \rangle^2 \right) ds,$$

$$\langle M^2(\phi)\rangle_t = \frac{b(y) + d(y) + \eta(y)C(y - x_0)n^*_{x_0,y} + \eta(y)C(0)n^*_{y,x_0}}{n^*_{x_0,y} + n^*_{y,x_0}}$$

$$\times \int_0^t \left(\langle F_s(\delta_{V_1}, y, X_{t_1}, .), \phi^2 \rangle - \langle F_s(\delta_{V_1}, y, X_{t_1}, .), \phi \rangle^2 \right) ds$$

$$= \frac{2b(y)}{n^*_{x_0,y} + n^*_{y,x_0}} \int_0^t \left(\langle F_s(\delta_{V_1}, y, X_{t_1}, .), \phi^2 \rangle - \langle F_s(\delta_{V_1}, y, X_{t_1}, .), \phi \rangle^2 \right) ds.$$

$$(3.10)$$

At time t_2, when a third trait appears in the population, the system can evolve to three two or just one coexisting traits, depending on the new trait equilibrium of the Lotka equations that is reached. For each of the traits, the marker distribution evolves as a generalization of the Fleming–Viot processes above.

Remark 3.8 The above equations show that, when there is coexistence of two traits in the population, the markers in the subpopulations defined by the two traits evolve independently but with parameters depending on the two co-existing traits. Thus, when there is a diversification event in the population, the distribution of the neutral diversity in one of the two subpopulations does not evolve as completely forgetting the other one, as it is usually assumed. The parameters of the underlying Fleming–Viot process depend on the complete trait distribution.

We present in Fig. 4 simulations in the case of coexistence, with the same model and parameters as in Sect. 3.3, except $\sigma_C = 0.7$ and the initial condition: $x_0 = -0.1$. The simulations (see Fig. 4) show the appearance of a new mutant trait (yellow) in a population of two coexisting traits (black and blue).

4 Proof of Theorem 3.4

Let us sketch the proof. In this section, we will suppose that Assumptions 2.1, 3.1 are satisfied and the initial conditions are $v_0^K(dy, dv) = n_0^K \delta_{(x_0, u_0)}(dy, dv)$ with $\lim_{K \to \infty} n_0^K = \widehat{n}_{x_0}$ and $\sup_{K \in \mathbb{N}^*} \mathbb{E}((n_0^K)^3) < +\infty$.

First, we recall results due to Champagnat et al. (2008) that provide the finite marginal convergence of the trait process $(X_{Kt}^K; t \geq 0)$. We extend these results to obtain the weak convergence of the measures $(X_{Kt}^K(dx)dt; K \geq 0)$ in $\mathcal{M}_F(\mathcal{X} \times [0, T])$ embedded with the weak convergence topology. This corresponds to the convergence of $(X_{Kt}^K; t \geq 0)$ as a trait-marker-time measure, as developed by Kurtz (1992). Secondly, we include the fast component (the marker) and prove the tightness of the sequence $(v_{Kt}^K(dx, du)dt; K \geq 0)$ in $\mathcal{M}_F(\mathcal{X} \times \mathcal{U} \times [0, T])$. We then consider a subsequence, again denoted by $(v_{Kt}^K(dx, du)dt, K \geq 0)$ with an abuse of notation, that converges to a limit $\Gamma(dt, dx, du) \in \mathcal{M}_F([0, T] \times \mathcal{X} \times \mathcal{U})$ that we have to identify. This derivation is done in several steps. When a successful mutant appears in the monomorphic population with trait x, the transition period to fixation is to be considered carefully. It has been proved by Champagnat (2006) that these transitions are of order $\log(K)$. We prove that during this time interval, the marker distribution in the mutant subpopulation remains a Dirac mass at the value of the initial mutant.
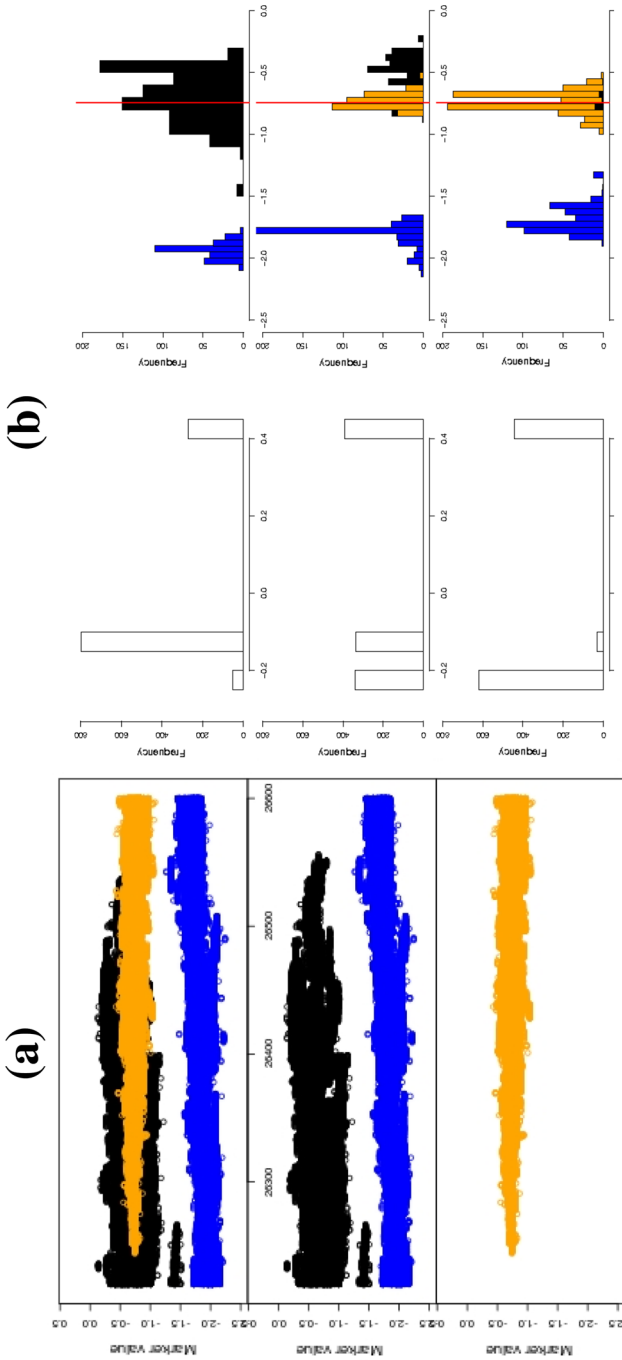
Fig. 4 Neutral marker dynamics in a trait-dimorphic population. Evolution of markers with respect to time. **a** The mutant population (*yellow*) and resident populations (*black* and *blue*) are shown separately and together. **b** Distributions of the traits and markers in the population at three times during the invasion. The marker value of the initial mutant is indicated by the *red line*

This results from the combined effects of small or rare marker mutations, large population and slow take-off of the new mutant population. Then, we show that in a trait monomorphic population with value x, the marker distribution converges to a Fleming–Viot superprocess parameterized by x.

4.1 Semimartingale decomposition of ν^K

Let us introduce some notation to keep forthcoming formula simple. For $\nu \in \mathcal{M}_F(\mathcal{X} \times \mathcal{U})$ and $\phi(x, u) \in \mathcal{C}(\mathcal{X} \times \mathcal{U}, \mathbb{R})$, we define the (nonlinear) generators B^K and $D^K(\nu)$ such that

$$
\begin{aligned}
B^K \phi(x, u) = {}& (1 - p_K)(1 - q_K) b(x) \phi(x, u) \\
& + p_K(1 - q_K) b(x) \int_{\mathcal{X}} \phi(x + k, u) m(x, k) dk \\
& + q_K(1 - p_K) b(x) \int_{\mathcal{U}} \phi(x, u + h) G_K(u, dh) \\
& + p_K q_K b(x) \int_{\mathcal{X} \times \mathcal{U}} \phi(x + k, u + h) m(x, k) dk \, G_K(u, dh) \quad (4.1)
\end{aligned}
$$

$$
D^K(\nu) \phi(x, u) = \big(d(x) + \eta(x) C * \nu(x) \big) \phi(x, u). \tag{4.2}
$$

The process $\langle \nu_{\cdot}^K, \phi \rangle$ is a square integrable semi-martingale and we give its characteristics.

Proposition 4.1 *For a continuous bounded function $\phi(x, u)$ on $\mathcal{X} \times \mathcal{U}$, the process*

$$
M_t^{K, \phi} = \langle \nu_t^K, \phi \rangle - \langle \nu_0^K, \phi \rangle - \int_0^t ds \int_{\mathcal{X} \times \mathcal{U}} \nu_s^K(dx, du) \big(B^K - D^K(X_s^K) \big) \phi(x, u) \tag{4.3}
$$

is a square integrable martingale with previsible quadratic variation

$$
\langle M^{K, \phi} \rangle_t = \frac{1}{K} \int_0^t ds \int_{\mathcal{X} \times \mathcal{U}} \nu_s^K(dx, du) \big(B^K + D^K(X_s^K) \big) \phi^2(x, u). \tag{4.4}
$$

Proof The dynamics being given in Sect. 2, the proof can be adapted from Fournier and Méléard (2004, Lemma 5.2). One main step consists in showing that there exists a Poisson point measure driving the measure-valued processes ν^K for all $K \in \mathbb{N}^*$. □

4.2 Convergence of the trait-marginal in the trait mutation time scale

As previously emphasized, the trait dynamics is described by the measure-valued process X^K which does not depend on the markers. This process has been fully studied in Champagnat (2006) and Champagnat et al. (2008). In this section, we recall the finite marginal convergence result obtained in these papers. We give some additional

properties concerning the topology involved. This result shows a time scale separation with successive fixations of successful mutants, under Assumptions 2.1 and 3.1. Notice that the time scale assumption is

$$\forall V > 0, \quad \log K \ll \frac{1}{K p_K} \ll \exp(V K), \quad \text{as } K \to \infty, \tag{4.5}$$

which is realized in our case for $p_K = 1/K^2$.

Theorem 4.2 *Under Assumptions 2.1 and 3.1, let us also assume that the initial population is trait-monomorphic: $X_0^K = n_0^K \delta_x$ for $x \in \mathcal{X}$ and $n_0^K \to \widehat{n}_x$ in probability and $\sup_{K \in \mathbb{N}^*} \mathbb{E}((n_0^K)^3) < +\infty$.*

Then, the sequence $(X_{Kt}^K; t \geq 0)$ converges to the pure jumps singleton measure-valued Markov process $(\widehat{n}_{Y_t} \delta_{Y_t}; t \geq 0)$ defined as follows: $Y_0 = x$, and the process Y jumps from x to $x + k$ with jump measure $b(x) \widehat{n}_x \frac{[f(x+k;x)]_+}{b(x+k)} m(x, k) dk$.

The convergence holds in the sense of finite dimensional distributions on $\mathcal{M}_F(\mathcal{X})$ equipped with the topology of total variation.

This theorem has been proved by Champagnat (2006) for the logistic case and generalized by Champagnat et al. (2008).

The trait-marginal process $(X_{Kt}^K; t \in [0, T])$ does not converge in $\mathbb{D}([0, T], \mathcal{M}_F(\mathcal{X}))$ embedded with the Skorokhod topology. Indeed, the size of jumps is upper-bounded by $\frac{1}{K}$ and nevertheless the limiting total mass process has jumps, preventing trajectorial tightness (at least in the $J1$-topology). Following the idea of Kurtz (1992) and as developed in Méléard and Tran (2012) and Gupta et al. (2014), a weaker topology consists in forgetting the process point of view and considering the measure $X_{Kt}^K(dx)dt$ in $\mathcal{M}_F([0, T] \times \mathcal{X})$ embedded with the topology of weak convergence. This convergence strengthens the result of Theorem 4.2 but in a topology weaker than the Skorohod topology.

To achieve this, as in Collet et al. (2013), we first introduce the M_1-topology on $\mathbb{D}([0, T], \mathbb{R}_+)$. It is weaker than the usual J_1-topology and allows monotonous processes with jumps tending to 0 to converge to processes with jumps (see Skorohod 1956). For a càdlàg function h on $[0, T]$, the continuity modulus for the M_1-topology is given by

$$w_\delta(h) = \sup_{\substack{0 \leq t_1 \leq t \leq t_2 \leq T; \\ 0 \leq t_2 - t_1 \leq \delta}} d(h(t), [h(t_1), h(t_2)]). \tag{4.6}$$

Note that if the function h is monotone, then $w_\delta(h) = 0$.

Proposition 4.3 *Let us consider a continuous, monotonous and non-negative function g. Then, under Assumptions 2.1 and 3.1, the process $(R_t^K, t \in [0, T])$ defined by*

$$R_t^K = \int g(x) X_{Kt}^K(dx)$$

converges in law in the sense of the Skorohod M_1-topology to the process $(R_t, t \in [0, T])$ where $R_t = \widehat{n}_{Y_t} g(Y_t)$.

Proof Assume that g is non-decreasing. From Theorem 4.2, finite dimensional distributions of $(R_t^K, t \in [0, T])$ converge to those of $(\widehat{n}_{Y_t} g(Y_t), t \in [0, T])$. By Skorohod (1956, Theorem 3.2.1), it remains to prove that for all $\eta > 0$,

$$\lim_{\delta \to 0} \limsup_{K \to \infty} \mathbb{P}(w_\delta(R_\cdot^K) > \eta) = 0,$$

where w_δ has been defined in (4.6).

The mutation rate in $(R_t^K, t \in [0, T])$ being bounded, the probability that two mutations occur within a time less than δ is $o(\delta)$. It is therefore enough to study the case where there is at most one mutation in the time interval $[0, \delta]$. Following Champagnat (2006), the path of R^K can be decomposed into several subpaths, each of them being closed to a large population deterministic measure-valued function ξ (see Proposition 5.1 in the Appendix) with a probability tending to 1. Away from invading mutations and for a trait-monomorphic population with trait x, $\langle \xi_{Kt}, g \rangle = g(x) n_{Kt}(x)$ where $n_\cdot(x)$ is the solution of the logistic Eq. (3.1). We can easily check that $t \to n_t(x)$ converges monotonously to its stable equilibrium $\widehat{n}_x$ and then $\langle \xi_{Kt}, g \rangle$ is monotonous and the modulus of continuity tends to 0. Around an invading mutant y, $\langle \xi_{Kt}, g \rangle$ is close to $n_{Kt}(x)g(x) + n_{Kt}(y)g(y)$ where $(n_t(x), n_t(y))$ is solution of the Lotka–Volterra system (5.4) with an initial condition close to $(\widehat{n}_x, 0)$. The mutant y invades if the fitness function $f(y; x)$ is positive (and $f(x; y)$ is negative). From Assumption 3.1, an easy study of the Lotka–Volterra system (see for example the Appendix in Champagnat 2004, Fig. b p.187), shows that either $n_t(x)$ and $n_t(y)$ are increasing or $\dot{n}_t(x) < 0$; $\dot{n}_t(y) > 0$. In that case, $n_t(x)g(x) + n_t(y)g(y)$ is the sum of two monotonous functions and the modulus of continuity tends also to 0   □

Corollary 4.4 *The sequence of random measures $X_{Kt}^K(dx)dt$ converges in law to the random measure $\widehat{n}_{Y_t} \delta_{Y_t}(dx)dt$ in $\mathcal{M}_F([0, T] \times \mathcal{X})$ embedded with the weak convergence topology.*

Proof It is enough to prove the convergence in law of $\int h(t)e^{-qx} X_{Kt}^K(dx)dt$ to $\int h(t)e^{-qx}\widehat{n}_{Y_t} \delta_{Y_t}(dx)dt$ for a measurable bounded function h and $q \in \mathbb{Q}$. In Skorohod (1956), it is proved that if x_K converges to x in $\mathbb{D}([0, T], \mathbb{R})$ embedded with the M_1-topology, then for t outside a denumerable set, $x_K(t)$ converges to $x(t)$. Then it follows by Lebesgue's Theorem that $\int_0^T H(t, x_K(t))dt$ converges to $\int_0^T H(t, x(t))dt$, as soon as H is bounded and continuous. We apply this result to the process $(\int_{\mathcal{X}} e^{-qx} X_{Kt}^K(dx), t \geq 0)$ and the function

$$H_M(t, y) = h(t)(y \wedge M),$$

for any $M > 0$. Estimate (2.9) (with $p = 1$) allows to conclude.   □

4.2.1 Marker distribution in a new adaptive trait mutant population

In this section, we study the transition of the marker distribution when a new mutant adaptive trait appears in a monomorphic population with trait x_0. We consider this

phenomenon at the ecological time scale and we prove that the fixation of the mutant trait creates a genetical bottleneck.

Let K be fixed. Initially we have a trait monomorphic population with trait x_0 and a marker distribution $\pi^K(x_0, du)$. Then an individual (x_0, v) from this population gives birth to an individual with mutant trait y and marker v (v has been chosen according to $\pi^K(x_0, du)$). We consider the process $(v_t^K; t \geq 0)$ started at

$$
\begin{aligned}
v_0^K(dx, du) &= X_0^K(dx)\pi_0^K(x, du) \\
&= \frac{1}{K}\delta_{(y,v)}(dx, du) + \frac{N_0^K - 1}{K}\delta_{x_0}(dx)\pi_0^K(x_0, du).
\end{aligned}
$$

Proposition 4.5 *Under Assumptions* 2.1 *and* 3.1, *let us consider a mutant* (y, v) *appearing in a monomorphic population with trait* x_0 *and marker distribution* $\pi_0^K(x_0, du)$. *Let us assume that* $f(y; x_0) > 0$, *where the fitness function has been defined in* (3.3). *There exists* $\varepsilon > 0$ *such that for any sequence* $(t_K; K \in \mathbb{N}^*)$ *with* $\lim_{K \to +\infty} t_K / \log K = +\infty$ *and* $\lim_{K \to +\infty} t_K / K = 0$ *(for example* $t_K = (\log K)^2$), *we have*

$$
\lim_{K \to +\infty} \mathbb{P}(\langle v_{t_K}^K, \mathbb{1}_y \rangle > \varepsilon) = \frac{f(y; x_0)}{b(y)} \quad \text{and} \quad \lim_{K \to +\infty} \mathbb{P}(\langle v_{t_K}^K, \mathbb{1}_y \rangle = 0) = 1 - \frac{f(y; x_0)}{b(y)}. \tag{4.7}
$$

Further, for the marker distribution, we can prove that

$$
\lim_{K \to +\infty} \mathbb{P}(\pi_{t_K}^K(y, du) = \delta_v(du)\langle v_{t_K}^K, \mathbb{1}_y \rangle > \varepsilon) = \frac{f(y; x_0)}{b(y)}. \tag{4.8}
$$

The Eq. (4.8) tells us that when the mutant trait survives in the resident population of trait x_0, then by the time t_K it needs to reach a non-negligible size, its marker distribution is still a Dirac mass at y. The assumption $q_K(\log K)^2$ in Assumption 2.1 ensures this. This assumption is not very restrictive as $(\log K)^2$ is a very slow growth rate. Additional comments are given after the proof.

Proof Properties (4.7) have been proved in Champagnat (2006) and Champagnat et al. (2008) and depend only on the trait distribution. We consider test functions $\phi(x, u)$ of the form $\mathbb{1}_y(x)g(u)$ with $g \in \mathcal{C}^2(\mathcal{U}, \mathbb{R})$ such that $\|g\|_\infty + \|g''\|_\infty \leq 1$. Starting from Proposition 4.1 and using Itô's formula with jumps, we obtain as soon as the population with trait y survives,

$$
\begin{aligned}
\int_{\mathcal{U}} g(u)\pi_{t_K}^K(y, du) &= \frac{\langle v_{t_K}^K, \mathbb{1}_y g \rangle}{\langle v_{t_K}^K, \mathbb{1}_y \rangle} \\
&= g(v) + M_{t_K}^{K,g} + q_K(1 - p_K)b(y)\int_0^{t_K}\left(1 - \frac{1}{K\langle v_s^K, \mathbb{1}_y \rangle + 1}\right) \\
&\quad \times \int_{\mathcal{U}} \pi_s^K(y, du)\int_{\mathcal{U}}(g(u + h) - g(u))G_K(u, dh)\,ds \tag{4.9}
\end{aligned}
$$

where $M^{K,g}$ is a square integrable martingale with previsible quadratic variation:

$$\langle M^{K,g}\rangle_{t_K} = \frac{1}{K}\int_0^{t_K} ds\left\{ b(y)(1-q_K)(1-p_K)\right.$$

$$\times \frac{\langle v_s^K, \mathbb{1}_y\rangle}{\left(\langle v_s^K, \mathbb{1}_y\rangle + \frac{1}{K}\right)^2}\int_{\mathcal{U}}\left(g(u)-\langle \pi_s^K, g\rangle\right)^2\pi_s^K(y,du)$$

$$+\left(d(y)+\eta(y)C*v_s^K(y)\right)\frac{\langle v_s^K, \mathbb{1}_y\rangle}{\left(\langle v_s^K, \mathbb{1}_y\rangle - \frac{1}{K}\right)^2}\int_{\mathcal{U}}\left(g(u)-\langle \pi_s^K, g\rangle\right)^2\pi_s^K(y,du)$$

$$+b(y)q_K(1-p_K)\frac{\langle v_s^K, \mathbb{1}_y\rangle}{\left(\langle v_s^K, \mathbb{1}_y\rangle + \frac{1}{K}\right)^2}\int_{\mathcal{U}}\pi_s^K(y,du)$$

$$\left.\times \int_{\mathcal{U}} G_K(u,dh)\left(g(u+h)-\langle \pi_s^K, g\rangle\right)^2\right\}. \tag{4.10}$$

The third term in the right hand side of (4.9) is of order t_K/K. Indeed thanks to (2.4) and (2.5), it is upper bounded by

$$\frac{t_K}{K}\,\bar{b}\,\|Ag\|_\infty.$$

Equation (4.10) needs more attention. As soon as the mass $\langle v_s^K, \mathbb{1}_y\rangle$ of the mutant population is of order 1, the variance of $M_{t_K}^{K,g}$ is in t_K/K which tends to zero when $K\to+\infty$. However, since we start from 1 individual, we have to separate the time interval $[0,t_K]$ into 2 parts. Let us introduce a sequence (s_K) such that $s_K\leq t_K$ for any K and

$$\log K \ll s_K \ll \frac{1}{\sqrt{q_K}}.$$

This is possible thanks to the assumption that $q_K(\log K)^2\to 0$. Notice that s_K can be equal to t_K. Using Assumption 3.1, we can prove as in Champagnat (2006, Lemma 3) that there exists $\varepsilon_0 > 0$ such that

$$\lim_{K\to\infty}\mathbb{P}\left(\forall s\in[s_K,t_K],\ \langle v_s^K, \mathbb{1}_y\rangle \geq \varepsilon_0\right) = \frac{f(y;x_0)}{b(y)}.$$

It turns immediately out that

$$\mathbb{E}\left(\frac{\mathbb{1}_{\{\forall s\in[s_K,t_K],\ \langle v_s^K,\mathbb{1}_y\rangle>0\}}}{K}\int_{s_K}^{t_K}\frac{b(y)+d(y)+\eta(y)C*v_s^K(y)}{\langle v_s^K, \mathbb{1}_y\rangle}\right.$$

$$\left.\int_{\mathcal{U}}\left(g(u)-\langle \pi_s^K, g\rangle\right)^2\pi_s^K(y,du)ds\right) \leq C\,\frac{t_K}{K}. \tag{4.11}$$

Before time s_K, the population size with trait y is not large enough and $\frac{1}{K\langle v_s^K, \mathbb{1}_y\rangle}$ can only be upper bounded by 1. Therefore we have to control the expectation of the variance of g under π_s^K. The expected number of marker mutations at time s along a lineage is sq_K and the variance of such mutation is bounded by $\|g\|_\infty^2 = \sup\{g(h)^2, h \in \mathcal{U}\}$. Then

$$\mathbb{E}\left(\int_{\mathcal{U}} \left(g(u) - \langle \pi_s^K, g\rangle\right)^2 \pi_s^K(y, du)\right) \le s \, q_K \, \|g\|_\infty^2, \qquad (4.12)$$

and

$$\mathbb{E}\left(\frac{1}{K}\int_0^{s_K} \frac{b(y) + d(y) + \eta(y)C * v_s^K(y)}{\langle v_s^K, \mathbb{1}_y\rangle} \int_{\mathcal{U}} \left(g(u) - \langle \pi_s^K, g\rangle\right)^2 \pi_s^K(y, du) ds\right)$$
$$\le C \frac{(s_K)^2 r_K}{K^2}.$$

The upper bound converges to 0 by the assumption on q_K. The third term of (4.10) is treated similarly. This concludes the proof. $\qquad\square$

Remark 4.6 For $q_K = 1/\sqrt{K}$, let us notice that the rate of appearance of mutant markers in a population of size K is of order $Kq_K = \sqrt{K}$ which does not tend to zero. This means that many mutant markers appear in the population of trait y during the t_K time interval following the first mutant (y, v). However, heuristically, since in a tree the mass is concentrated around the leaves, the mutants do not appear with the same probability along the time interval and mutations are mostly observed after the time s_K when the mutant population (y, v) is already large. Moreover, using that the marker mutation step and/or marker mutation frequency is small we obtain that the mutant markers remain in negligible proportion between s_K and t_K.

4.3 Convergence of the marker distribution process in a trait-monomorphic population

For $K \in \mathbb{N}^*$, we introduce, as in Champagnat (2006), the following sequence of stopping times τ_k^K and θ_k^K:

$$\tau_0^K = 0, \qquad \theta_0^K = 0$$
$$\tau_{k+1}^K = \inf\{t > \tau_k^K, \ \mathrm{Card}\big(\mathrm{supp}(\bar{X}_t^K)\big) = \mathrm{Card}\big(\mathrm{supp}(\bar{X}_{t-}^K)\big) + 1\}$$
$$\theta_k^K = \inf\{t > \tau_k^K, \ \mathrm{Card}\big(\mathrm{supp}(\bar{X}_t^K)\big) = 1\}.$$

The times τ_k^K's are the times of appearance of the successive mutant traits in the population and the θ_k^K's are the times at which the population returns to monomorphic state. These times are possibly infinite, if the corresponding sets are empty. It has been proved in Champagnat (2006) that for t_K be such that $\lim_{K\to+\infty} t_K/\log(K) = +\infty$ and $\lim_{K\to+\infty} t_K/K = 0$,

$$\lim_{K \to +\infty} \mathbb{P}\left(\forall k \ge 0, \tau_k^K \wedge KT \le \theta_k^K \wedge KT \le (\tau_k^K + t_K) \wedge KT \le \tau_{k+1}^K \wedge KT\right) = 1.$$
(4.13)

Proposition 4.7 *Take the process* $(v_{Kt}^K; t \in [0, T])$ *started with the monomorphic initial condition* $v_0^K(dx, du) = n_0^K \delta_{(x_0, u_0)}(dx, du)$, *where* $\lim_{K \to +\infty} n_0^K = \widehat{n}_{x_0} > 0$ *and* $\sup_{K \in \mathbb{N}^*} \mathbb{E}((n_0^K)^3) < +\infty$.
(i) In the trait-mutation time scale, the time of first mutation converges in distribution as follows:
$$\lim_{K \to +\infty} \tau_1^K / K = \tau_1,$$
(4.14)

where τ_1 *is an exponential time with parameter* $b(x_0)\widehat{n}_{x_0}$.
(ii) Let us consider the processes $(\pi_{K(t \wedge \tau_1^K)}^K; t \in [0, T])$ *stopped at the time of first mutation. When* $K \to +\infty$, *this sequence converges in distribution in* $\mathbb{D}([0, T], \mathcal{P}(\mathcal{U}))$ *to the Fleming–Viot process* $F^{u_0}(x_0, du)$ *(see Definition* 3.3*) and stopped at the independent exponential time* τ_1.

Proof First of all, the trait and marker mutations are independent. Thus, the stopping time τ_1^K is independent of the marker distribution $\pi^K(x_0, du)$. The results of Champagnat and coauthors Champagnat (2006) and Champagnat et al. (2008) are unchanged and give (4.14). Moreover, by Champagnat et al. (2008, Lemma 5.4)

$$\lim_{K \to +\infty} \mathbb{P}\left(\sup_{s \in [\log K, \tau_1^K]} \langle X_s^K, \mathbf{1} \rangle \ge \frac{\widehat{n}_{x_0}}{2}\right) = 1.$$
(4.15)

Let $\phi \in \mathcal{C}(\mathcal{U}, \mathbb{R})$. Since the population is trait-monomorphic with trait x_0, then $\langle \pi_{Kt}^K(x_0, du), \phi(u) \rangle = \frac{\langle v_{Kt}^K, \phi \rangle}{\langle v_{Kt}^K, \mathbf{1} \rangle}$. Thus, from Proposition 4.1 and Itô's formula, we get that in the time scale Kt

$$\langle \pi_{K(t \wedge \tau_1^K)}^K(x_0, .), \phi \rangle = \langle \pi_0^K(x_0, .), \phi \rangle + H_{K(t \wedge \tau_1^K)}^{K, \phi} + b(x_0) q_K (1 - p_K)$$
$$\times \int_0^{t \wedge \tau_1^K} \left(1 - \frac{1}{K \langle v_{Ks}^K, \mathbf{1} \rangle + 1}\right) \frac{r_K}{K}$$
$$\times \int_{\mathcal{U} \times \mathcal{U}} \left(\phi(u + h) - \phi(u)\right) G_K(u, dh) K \pi_{Ks}^K(x_0, du) ds$$
(4.16)

where $H^{K, \phi}$ is a square integrable martingale with quadratic variation is

$$\langle H^{K, \phi} \rangle_{K(t \wedge \tau_1^K)} = \int_0^{t \wedge \tau_1^K} b(x_0)(1 - q_K)(1 - p_K) \frac{\langle v_{Ks}^K, \mathbf{1}_{x_0} \rangle}{\left(\langle v_{Ks}^K, \mathbf{1} \rangle + \frac{1}{K}\right)^2}$$
$$\times \int_{\mathcal{U}} \pi_{Ks}^K(x_0, du) \left(\phi(u) - \langle \pi_{Ks}^K(x_0, .), \phi \rangle\right)^2$$
$$+ \left(d(x_0) + \eta(x_0) C * X_{Ks}^K(x_0)\right) \frac{\langle v_{Ks}^K, \mathbf{1}_{x_0} \rangle}{\left(\langle v_{Ks}^K, \mathbf{1} \rangle - \frac{1}{K}\right)^2}$$

$$\times \int_{\mathcal{U}} \left(\phi(u) - \langle \pi^K_{Ks}(x_0, .), \phi \rangle \right)^2$$

$$+ b(x_0) q_K (1 - p_K) \frac{\langle \nu^K_{Ks}, \mathbb{1}_{x_0} \rangle}{\left(\langle \nu^K_{Ks}, \mathbf{1} \rangle + \frac{1}{K} \right)^2}$$

$$\times \int_{\mathcal{U}} \pi^K_{Ks}(x_0, du) \left(\phi(u + h) - \langle \pi^K_{Ks}(x_0, .), \phi \rangle \right)^2 \tag{4.17}$$

The computation shows that the order of the quadratic variation of π^K_t is $\frac{1}{K}$. Thus at time scale Kt, this order will be 1. That justifies Assumption (2.4) for p_K which is the only choice to get a non degenerate diffusive limit.

Let us introduce a process $(\widetilde{\pi}^{K,x_0}_t(du), t \geq 0)$ coupled with $(\pi^K_t(x_0, du), t \geq 0)$, on the same probability space and driven by the same Poisson point measures, that satisfies the following properties. The dynamics of $\widetilde{\pi}^{K,x_0}_.(du)$ is given by (4.16)–(4.17) but without the stopping times τ^K_1 and we have that $\forall t \geq 0,\ \pi^K_{K(t \wedge \tau^K_1)}(x_0, du) = \widetilde{\pi}^{K,x_0}_{K(t \wedge \tau^K_1)}(du)$. In a nutshell, $(\widetilde{\pi}^{K,x_0}_t, t \geq 0)$ corresponds to the process $(\pi^K_t(x_0, .), t \geq 0)$ that is obtained by setting the trait mutation kernel to the Dirac mass at 0.

Thanks to (2.5), (4.15) and using that $\widetilde{\pi}^{K,x_0}$ is a probability-valued process, (4.16) and (4.17) imply that for any $\phi \in \mathcal{C}(\mathcal{U}, \mathbb{R})$, the distribution sequence of $(\langle \widetilde{\pi}^{K,x_0}_{K.}, \phi \rangle; K \in \mathbb{N}^*)$ is uniformly tight in $\mathbb{D}([0, T], \mathbb{R})$. By Roelly's criterion (1986, Theorem 2.1), this implies the tightness of the sequence of the laws of $(\widetilde{\pi}^{K,x_0}_{K.}; K \in \mathbb{N}^*)$ in $\mathbb{D}([0, T], \mathcal{P}(\mathcal{U}))$.

Let us consider a limiting value $(\bar{\pi}_t(du); t \in [0, T])$ of the tight sequence and a subsequence, again denoted by $\widetilde{\pi}^{K,x_0}_{K.}(du)$, that converges to $\bar{\pi}_.(du)$. By Assumption (4.15) and since individuals have weight $1/K$, the limiting laws only charge $\mathcal{C}([0, T], \mathcal{P}(\mathcal{U}))$.

It remains to identify $\bar{\pi}_.(du)$. Let $0 < s < t < T$, let $k \in \mathbb{N}$ and $0 < s_1 \leq \cdots s_k < s < t$, let $\phi_1, \cdots \phi_k$ be bounded continuous function on $\mathcal{P}(\mathcal{X} \times \mathcal{U})$ and $\phi \in \mathcal{C}(\mathcal{U}, \mathbb{R})$. We define the following bounded functional on $\mathbb{D}([0, T], \mathcal{P}(\mathcal{U}))$

$$\Psi_{s,t}(Y) = \phi_1(Y_{s_1}) \cdots \phi_k(Y_{s_k}) \left\{ \langle Y_t, \phi \rangle - \langle Y_s, \phi \rangle - \int_s^t du \int_{\mathcal{U}} Y_u(du) b(x_0) A\phi(u) \right\}$$

On the one hand, using (4.16), we obtain that

$$\Psi_{s,t}(\widetilde{\pi}^{K,x_0}_{K.}) = \phi_1(\widetilde{\pi}^{K,x_0}_{Ks_1}) \cdots \phi^K_k(\widetilde{\pi}^{K,x_0}_{Ks_k}) \left\{ H^{K,\phi}_{Kt} - H^{K,\phi}_{Ks} + \varepsilon^K_{Kt} \right\}$$

where

$$\varepsilon^K_{Kt} = \int_0^t ds \int_{\mathcal{U}} \widetilde{\pi}^{K,x_0}_{Ks}(du) \left[b(x_0) \frac{r_K}{K} \int_{\mathcal{U}} \left(\phi(u + h) - \phi(u) \right) G_K(u, dh) \right]$$

$$- \int_0^t ds \int_{\mathcal{U}} \widetilde{\pi}^{K,x_0}_{Ks}(du) b(x_0) A\phi(u) \tag{4.18}$$

tends to 0 in $\mathbb{L}^1$ when $K \to +\infty$. Thus,

$$\lim_{K \to +\infty} \mathbb{E}\left(\Psi_{s,t}(\widetilde{\pi}_{K}^{K,x_0})\right) = 0. \tag{4.19}$$

On the other hand, using (2.9) and the convergence of $(\widetilde{\pi}_{K(\cdot \wedge \tau_1^K)}^{K,x_0}(du); K \in \mathbb{N}^*)$ to $\bar{\pi} \in \mathcal{C}([0, T], \mathcal{M}_F(\mathcal{U}))$, we get

$$\mathbb{E}\left(\Psi_{s,t}(\bar{\pi})\right) = \lim_{K \to +\infty} \mathbb{E}\left(\Psi_{s,t}(\widetilde{\pi}_{K}^{K,x_0}(du))\right). \tag{4.20}$$

This shows that $\mathbb{E}\left(\Psi_{s,t}(\bar{\pi})\right) = 0$ and hence the process $M^{x_0}(\phi)$ defined in (3.4) is a martingale obtained as the uniform limit in time of $H_{Kt}^{K,\phi}$, when $K \to +\infty$. Moreover, the bracket (4.17) converges to

$$\int_0^t \frac{b(x_0) + d(x_0) + \eta(x_0)\widehat{n}_{x_0}}{\widehat{n}_{x_0}} \int_{\mathcal{U}} \bar{\pi}(du)\left[\left(\phi(u) - \langle \bar{\pi}_s, \phi \rangle\right)^2\right] ds$$
$$= \int_0^t \frac{2b(x_0)}{\widehat{n}_{x_0}}\left[\langle \bar{\pi}_s, \phi^2 \rangle - \langle \bar{\pi}, \phi \rangle^2\right] ds. \tag{4.21}$$

Indeed, the integral in (4.17) can be separated into two integrals, one between 0 and $\frac{\log K}{K} \wedge t \wedge \tau_1^K$ and the other between $\frac{\log K}{K} \wedge t \wedge \tau_1^K$ and $t \wedge \tau_1^K$. The second integral converges to (4.21), but some caution is needed for the first integral since the ratios $\langle \nu_{Ks}^K, \mathbb{1}_{x_0} \rangle / \left(\langle \nu_{Ks}^K, \mathbb{1} \rangle \pm \frac{1}{K}\right)^2$ are of order K. Using the same arguments as for (4.12), we can upper bound the integral between 0 and $\frac{\log K}{K} \wedge t \wedge \tau_1^K$ by

$$CK \int_0^{\log K/K} sq_K ds = C \frac{(\log K)^2}{K} q_K \to_{K \to +\infty} 0.$$

Using Theorem 3.12 p. 432 of Jacod and Shiryaev (1987), that provides the convergence of $H_{K}^{K,\phi}$ to the solution of the martingale problem (3.4)–(3.5) with $x = x_0$.

By the independence of $\widetilde{\pi}_{K}^{K,x_0}(du)$ and τ_1^K, τ_1 is independent of $\bar{\pi}_\cdot(du)$ and $\bar{\pi}_{\cdot \wedge \tau_1} = \pi_{\cdot \wedge \tau_1}^{u_0}(x_0, du)$. This concludes the proof. $\qquad\square$

4.4 Conclusion

Using Theorem 4.2, Proposition 4.5 and 4.7, we prove the first part of Theorem 3.4, for the convergence in finite distribution. Let us now consider the convergence in the space of trait-marker-time measures.

Corollary 4.8 *The family* $(\nu_{Kt}^K(dx, du) dt, K \in \mathbb{N}^*)$ *is uniformly tight in* $\mathcal{M}_F(\mathcal{X} \times \mathcal{U} \times [0, T])$ *embedded with the weak convergence topology and converges in distribution to the measure* $V_t(dx, du)dt$, *where* V *is defined in Theorem* 3.4.

Proof Since the space $\mathcal{X} \times \mathcal{U} \times [0, T]$ is compact, it is sufficient to prove that

$$\sup_{K \in \mathbb{N}^*} \mathbb{E} \left(\int_0^T \langle \nu_{Kt}^K, 1 \rangle dt \right) < +\infty$$

which is a consequence of Fubini's theorem since

$$\mathbb{E} \left(\int_0^T \langle \nu_{Kt}^K, 1 \rangle dt \right) \leq T \sup_{K \in \mathbb{N}^*, t \in \mathbb{R}_+} \mathbb{E} \left(\langle \nu_t^K, 1 \rangle \right).$$

Estimate (2.9) concludes the proof of tightness.

Let us now consider continuous functions $\phi \in \mathcal{C}(\mathcal{X} \times [0, T], \mathbb{R})$ and $g \in \mathcal{C}(\mathcal{U}, \mathbb{R})$, where the stopping times τ_k^K and θ_k^K have been introduced in Sect. 4.3. Then

$$\int_0^T \int_{\mathcal{X} \times \mathcal{U}} \phi(x, t) g(u) \nu_{Kt}^K(dx, du) \, dt = \int_0^T \langle \pi_{Kt}^K(x, .), g \rangle \phi(x, t) X_{Kt}^K(dx) dt$$

$$= \sum_{k \geq 0} \int_{(\theta_k^K \wedge T)/K}^{(\tau_{k+1}^K \wedge T)/K} \langle \pi_{Kt}^K(x, .), g \rangle \phi(x, t) X_{Kt}^K(dx) dt$$

$$+ \int_{(\tau_{k+1}^K \wedge T)/K}^{(\theta_{k+1}^K \wedge T)/K} \langle \pi_{Kt}^K(x, .), g \rangle \phi(x, t) X_{Kt}^K(dx) dt. \tag{4.22}$$

The limit (4.13) implies that the second term of the right hand side of (4.22) converges to 0. Given X^K, the processes $(\pi_{Kt}^K(x, .); t \in [\theta_k^K/K, \tau_{k+1}^K/K))$, for $k \geq 0$, in the first term of the r.h.s. of (4.22) are independent and, by Proposition 4.7, they converge in distribution in $\mathbb{D}([0, T], \mathbb{R})$ to the Fleming–Viot processes (3.4)–(3.5) with the initial conditions described by the jumps of the extended TSS (Y, U). Corollary 4.4 and dominated convergence theorem allows us to conclude the proof. $\square$

5 Appendix: Limit theorems for the trait and marker distributions in the ecological time scale

If we let $K \to +\infty$ without changing the time scale, we obtain:

Proposition 5.1 *Assume that the sequence $(\nu_0^K(dx, du); K \in \mathbb{N}^*)$ converges in probability to the measure $\xi_0(dx, du)$ when $K \to +\infty$. Then the sequence of processes $(\nu_t^K(dx, du); t \geq 0)_{K \in \mathbb{N}^*}$ converges in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathcal{X} \times \mathcal{U}))$ to the deterministic process $\xi \in \mathcal{C}(\mathbb{R}_+, \mathcal{M}_F(\mathcal{X} \times \mathcal{U}))$ defined for every $\phi(x, u) \in \mathcal{C}(\mathcal{X} \times \mathcal{U}, \mathbb{R})$ by:*

$$\langle \xi_t, \phi \rangle = \langle \xi_0, \phi \rangle + \int_0^t \int_{\mathcal{X} \times \mathcal{U}} \big(b(x) - d(x) - \eta(x) C * \xi_s(x) \big) \phi(x, u) \xi_s(dx, du) \, ds.$$

(5.1)

Proof The proofs proceed in a classical tightness-uniqueness way (Fournier and Méléard 2004).   □

Notice that no mutation can be seen at this scale. To see the trait mutations, we have to consider the process at the mutation scale Kt (Champagnat et al. 2008).

Trait-monomorphic case From Proposition 5.1:

Corollary 5.2 *Assume that the initial population is trait-monomorphic* $v_0^K(dx, du)$ $= n_0^K \delta_{x_0}(dx) \pi_0^K(x_0, du)$ *where* $\lim_{K \to +\infty} n_0^K = n_0$ *and* $\lim_{K \to +\infty} \pi^K(x_0, du) =$ $\pi_0(x_0, du)$ *in probability. Then the sequence* $(v^K; K \in \mathbb{N}^*)$ *converges, in probability and uniformly on every compact time interval* $[0, T]$ *with* $T > 0$*, to* $(v_t(dx, du) =$ $n_t(x) \delta_{x_0}(dx) \pi_0(x_0, du); t \geq 0)$ *where* $n_t(x)$ *is the deterministic solution of the logistic equation*

$$\frac{dn_t}{dt} = \big(b(x) - d(x) - \eta(x) C(0) n_t(x) \big) n_t(x)$$

(5.2)

which converges when t *tends to infinity to*

$$\widehat{n}_x = \frac{b(x) - d(x)}{\eta(x) C(0)}.$$

(5.3)

Proof For the part of the proof dealing with ξ, we refer to Champagnat et al. (2008).   □

Trait-dimorphic case From Proposition 5.1:

Corollary 5.3 *Assume that the initial population is trait-dimorphic*

$$v_0^K(dx, du) = n_0^K(x_1) \delta_{x_1}(dx) \pi_0^K(x_1, du) + n_0^K(x_2) \delta_{x_2}(dx) \pi_0^K(x_2, du)$$

where $n_0^K(x)$ *is the number of individuals with trait* x *renormalized by* K*. We assume that for* $x \in \{x_1, x_2\}$*,* $\lim_{K \to +\infty} n_0^K(x) = n_0(x) > 0$ *and* $\lim_{K \to +\infty} \pi^K(x, du) =$ $\pi_0(x, du)$ *in probability. We also assume that* (x_1, x_2) *satisfies the Assumption* 3.1.

Then the sequence $(v^K; K \in \mathbb{N}^*)$ *converges, in probability and uniformly on every compact time interval* $[0, T]$ *with* $T > 0$*, to*

$$(v_t(dx, du) = n_t(x_1) \delta_{x_1}(dx) \pi_0(x_1, du) + n_t(x_2) \delta_{x_2}(dx) \pi_0(x_2, du); t \geq 0)$$

where $(n_t(x_1), n_t(x_2))$ *solves the system*

$$\frac{dn_t(x_1)}{dt} = \big(b(x_1) - d(x_1) - \eta(x_1)(C(0) n_t(x_1) + C(x_1 - x_2) n_t(x_2)) \big) n_t(x_1)$$

$$\frac{dn_t(x_2)}{dt} = \big(b(x_2) - d(x_2) - \eta(x_2)(C(x_2 - x_1) n_t(x_1) + C(0) n_t(x_2)) \big) n_t(x_2).$$

(5.4)

whose only stable equilibrium is $(0, \widehat{n}_{x_2})$*, with* $\widehat{n}_{x_2}$ *defined in* (3.2).

It can be seen that the conditional distributions of the marker, given the trait x_1 or x_2 remain constant.

Proof The convergence in large population of $(\nu^K; K \in \mathbb{N}^*)$ is a consequence of Proposition 5.1. □

References

Barton N (1998) The effect of hitchhiking on neutral genealogies. Genet Res 72:123–133

Barton N (2000) Genetic hitchhiking. Phil Trans R Soc Lond B 355:1553–1562

Champagnat N (2004) Etude mathématique de modèles stochastiques d'évolution issus de la théorie écologique des dynamiques adaptatives, phd thesis. Ph.D. thesis, Université Paris X-Nanterre

Champagnat N (2006) A microscopic interpretation for adaptative dynamics trait substitution sequence model. Stoch Process Appl 116:1127–1160

Champagnat N, Méléard S (2011) Polymorphic evolution sequence and evolutionary branching. Probab Theory Relat Fields 151:45–94

Champagnat N, Ferrière R, Méléard S (2008) From individual stochastic processes to macroscopic models in apative evolution. Stoch Models 24(Suppl 1):2–44

Champagnat N, Jabin PE, Méléard S (2014) Adaptation in a stochastic multi-resources chemostat model. J Math Pures Appl 6:755–788

Chevin LM, Hospital F (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180:1645–1660

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

Collet P, Méléard S, JAJ M (2013) A rigorous model study of the adaptative dynamics of Mendelian diploids. J Math Biol 67:569–607

Crow J, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New york

Cutter A, Payseur B (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. Nature 14:262–274

Dawson D (1993) Mesure-valued markov processes. Springer, editor, Ecole d'Eté de probabilités de Saint-Flour XXI, vol 1541 of Lectures Notes in Math., New York

Dawson D, Hochberg K (1982) Wandering random measures in the Fleming–Viot model. Ann Probab 10:554–580

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. Nature 400:354–357

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

Donnelly P, Kurtz T (1996) A countable representation of the Fleming–Viot measure-valued diffusion. Ann Probab 24:698–742

Durrett R, Schweinsberg J (2004) Approximating selective sweeps. Theor Popul Biol 66:129–138

Eshel I, Feldman M, Bergman A (1998) Long-term evolution, short-term evolution, and population genetic theory. J Theor Biol 191:391–396

Etheridge A (2000) An introduction to superprocesses, volume 20 of University Lecture Series. American Mathematical Society, Providence

Etheridge A, Pfaffelhuber P, Wakolbinger A (2006) An approximate sampling formula under genetic hitchhiking. Ann Appl Probab 16:685–729

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

Geritz S, Kisdi E, Meszéna G, Metz J (1998) Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree. Evol Ecol 12:35–57

Gupta A, Metz J, Tran V (2014) A new proof for the convergence of an individual based model to the trait substitution sequence. Acta Appl Math 131:1–27

Hammerstein P (1996) Darwinian adaptation, population genetics and the streetcar theory of evolution. J Math Biol 34:511–532

Jacod J, Shiryaev A (1987) Limit theorems for stochastic processes. Springer, Berlin

Jourdain B, Méléard S, Woyczynsk W (2012) Lévy flights in evolutionary ecology. J Math Biol 65:677–707

Kaplan N, Hudson R, Langley C (1989) The hitchhiking effect revisited. Genetics 123:887–899

Kojima K, Schaffer H (1967) Survival process of linked mutant genes. Evolution 21:518–531

Kurtz T (1992) Averaging for martingale problems and stochastic approximation. In: Applied stochastic analysis (New Brunswick, NJ, 1991), Lectures Notes in Control and Inform Sci, vol 177, Springer, Berlin

Lloyd D (1977) Genetic and phenotypic models of natural selection. J Theor Biol 69:543–560

Matessi C, Schneider K (2009) Optimization under frequency-dependent selection. Theor Popul Biol 72:1–12

Maynard Smith J (1982) Evolution and the theory of games. Cambridge University Press, New York

Méléard S, Tran V (2009) Trait substitution sequence process and canonical equation for age-structured populations. J Math Biol 58:881–921

Méléard S, Tran V (2012) Slow and fast scales for superprocess limits of age-structured populations. Stoch Process Appl 122:250–276

Metz J, Nisbet R, Geritz S (1992) How should we define 'fitness' for general ecological scenarios? Trends Ecol Evol 7:198–202

Metz J, Geritz S, Meszéna G, Jacobs F, Van Heerwaarden J (1996) Adaptative dynamics, a geometrical study of the consequences of nearly faithful reproduction. In: Van Strien SJ, Verduyn Lunel SM (eds) Stochastic and spatial structures of dynamical systems. North Holland, Amsterdam, The Netherlands

Ohta T, Kimura M (1973) Model of mutation appropriate to estimate number of electrophoretically detectable alleles in a finite population. Genet Res 22:201–204

Ohta T, Kimura M (1975) Effects of selected linked locus on heterozygosity of neutral alleles (hitchhiking effect). Genet Res 25:313–326

Roelly S (1986) A criterion of convergence of measure-valued processes: application to measure branching processes. Stochastics 17:43–65

Roughgarden J (1979) Theory of population genetics and evolutionary ecology: an introduction. Macmillan, New York

Schneider K, Kim Y (2010) An analytical model for genetic hitchhiking in the evolution of antimalarial drug resistance. Theor Popul Biol 78:93–103

Schoener T (2011) The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. Science 331:426–429

Skorohod A (1956) Limit theorems for stochastic processes. Theory Probab Appl 1:261–290

Smadi C (2014) An eco-evolutionary approach of adaptation and recombination in a large population of varying size. arXiv:14024104

Smith JM, Haigh J (1974) The hitchhiking effect of a favourable gene. Genet Res 23:23–35

Stephan W, Wiehe T, Lenz W (1992) The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor Popul Biol 41:237–254

Weissing F (1996) Genetic versus phenotypic models of selection: can genetics be neglected in a long-term perspective? J Math Biol 34:533–555