

Polymorphic evolution sequence and evolutionary branching

Nicolas Champagnat*, Sylvie Méléard†

March 23, 2010

Abstract

We are interested in the study of models describing the evolution of a polymorphic population with mutation and selection in the specific scales of the biological framework of adaptive dynamics. The population size is assumed to be large and the mutation rate small. We prove that under a good combination of these two scales, the population process is approximated in the long time scale of mutations by a Markov pure jump process describing the successive trait equilibria of the population. This process, which generalizes the so-called trait substitution sequence, is called polymorphic evolution sequence. Then we introduce a scaling of the size of mutations and we study the polymorphic evolution sequence in the limit of small mutations. From this study in the neighborhood of evolutionary singularities, we obtain a full mathematical justification of a heuristic criterion for the phenomenon of evolutionary branching. This phenomenon corresponds to the situation where the population, initially essentially single modal, is driven by the selective forces to divide into two separate subpopulations. To this end we finely analyze the asymptotic behavior of 3-dimensional competitive Lotka-Volterra systems.

MSC 2000 subject classification: 92D25, 60J80, 37N25, 92D15, 60J75

Key-words: Mutation-selection individual-based model, fitness of invasion, adaptive dynamics, polymorphic evolution sequence, competitive Lotka-Volterra system, evolutionary branching.

1 Introduction

We consider an asexual population in which each individual's ability to survive and reproduce is characterized by a quantitative trait, such as the body size, the age at maturity, or the rate of food intake. Evolution, acting on the trait distribution of the population, is the consequence of three basic mechanisms: *heredity*, which transmits traits to new offsprings, *mutation*, driving a variation in the trait values in the population, and *selection* between

*EPI TOSCA, INRIA Sophia Antipolis – Méditerranée, 2004 route des Lucioles, BP. 93, 06902 Sophia Antipolis Cedex, France; e-mail: Nicolas.Champagnat@sophia.inria.fr.

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

these different trait values, which is due to the competition between individuals for limited resources or area. Adaptive dynamics models aim at studying the interplay between these different mechanisms [23, 27, 29]. One of the key features of these models, observed in many simulations, is the emergence of phenotypic separation without geographic separation in the population (Metz et al. [28]). In the sexual case, it is related to the phenomenon of sympatric speciation (Dieckmann and Doebeli [9]). When the initial population is monomorphic (i.e. all individuals have the same trait), simulations show that sometimes, one has the following phenomenon. The population stays essentially single-modal centered around a trait that evolves continuously until some random time. At this time it divides into two separate sub-populations. These sub-populations are still in interaction but are centered around distinct traits at a distance increasing with time. This phenomenon is called *Evolutionary Branching*. Our aim in this paper is to understand the dynamics of the process in long time scales and to highlight this evolutionary branching phenomenon. In particular, we want to give a rigorous proof of a result stated by Metz et al. [28, Section 3.2.5], where conditions on the parameters of the model allow one to predict whether evolutionary branching will occur or not.

Some mathematical approaches of this problem have been already developed. Some papers examined the stationary behavior of the population (Cressman and Hofbauer [7], Yu [32]). Partial results have also been obtained from deterministic models (Dieckmann et al. [11], Barles and Perthame [2], Desvillettes et al. [8]). Our approach is different. It is based on a stochastic individual-based model that details the ecological dynamics of each individual. In addition, it aims at describing how evolutionary branching occurs dynamically (i.e. in a non-stationary way): we first describe the approach of the population to a single modal equilibrium state, and next give a criterion to assess whether evolutionary branching occurs or not at this point.

We follow the basic description of adaptive dynamics based on the biologically motivated assumptions of rare mutations and a large population. Under these assumptions, we prove that the individual-based process can be approximated by a Markov pure jump process on the set of point measures on the trait space. The transitions of this process are given by the long time behavior of competitive Lotka-Volterra systems. They describe the succession of mutant invasions followed by a fast competition phase between the mutant population and the resident one. In the mutation time scale, and for large populations, the successful traits in the competition are given by the nontrivial equilibria of Lotka-Volterra systems which model the dynamics of the sizes of each sub-population corresponding to each resident or mutant trait. We thus generalize the situation introduced by Metz et al. [28] and mathematically developed by Champagnat [4], when the parameters of the model prevent the coexistence of two traits. In that case, the microscopic model converges to a monomorphic (one trait support) pure jump process, called Trait Substitution Sequence (TSS). This limit involves a timescale separation between the mutations and the population dynamics driving the competition between traits.

In this article, we relax the assumption of non-coexistence and obtain a *polymorphic evolution sequence* (PES), allowing coexistence of several traits in the population, from the same microscopic model described in Section 2.1. Simulations of an example are shown in Section 2.2 in order to illustrate the phenomenon of evolutionary branching. In Section 2.3, we introduce the deterministic competitive Lotka-Volterra systems describing the competition between traits. We prove in Section 2.4 that the PES takes the form of a

Markov jump process on the set of measures on the trait space \mathcal{X} that are finite sums of Dirac masses with positive weights, and we characterize the transitions of this process in terms of the long time behavior of competitive Lotka-Volterra systems. In Section 3, we explain why the assumptions ensuring the convergence to the PES are satisfied as long as no more than two traits coexist. In this case, the dynamics of the PES can be explicitly characterized. Next (Section 4), we study the transition from a monomorphic population to a stable dimorphic population, and give a full mathematical justification of the criterion for evolutionary branching proposed in [28], under the assumption of small mutation effects. To this end, we first show in Sections 4.1 that, away from evolutionary singularities, the support of the PES stays monomorphic and converges to an ODE known as the “canonical equation” (Dieckmann and Law [10]). Finally, in Section 4.2, we characterize in Theorem 4.9 the situations when evolutionary branching occurs by specializing to our situation the results of Zeeman [33] on the asymptotic behavior of 3-dimensional competitive Lotka-Volterra systems. This is the main result of the paper.

Let us stress the delicate combination of the limits. Here we are concerned by the combination of the limits of large populations and rare mutations, followed by a limit of small mutations. An alternative approach would be first to study the limit of large population alone, giving in the limit an integro-differential partial differential equation for the density of traits (Champagnat et al. [5]); and next to study a limit of small mutations on this equation with a proper time scaling that would lead to some dynamics on the set of finite sums of Dirac masses on the trait space. The second part of this program has already been partly studied by Dieckmann et al. [11] in a specific model, but is related to difficult problems on Hamilton-Jacobi equations with constraints (Barles and Perthame [2]). In this case, evolutionary branching is numerically observed, but not yet fully justified. Another approach would be to combine the three limits we consider directly at the level of the microscopic model, allowing one to study the evolutionary process on several time scales (Bovier and Champagnat [3]). This requires a finer analysis of the invasion and competition phases after the appearance of a new mutant. Note that all these approaches are based on the same idea of separation between the time scales of mutation and competition, whereas the model of Yu [32] does not satisfy this assumption. This may explain why his results are different from ours (in particular, he does not observe evolutionary branching in a model where the fitness satisfies our branching criterion). This shows the delicate influence of parameters scaling and of the specific ecological model on the phenomenon of evolutionary branching.

2 Models and Polymorphic Evolution Sequence (PES)

Let us introduce here the main models on which our approach is based and give our convergence result to the polymorphic evolution sequence.

2.1 The individual-based model

The microscopic model we use is an individual-based model with density-dependence, which has been already studied in ecological or evolutionary contexts by many authors [14, 5]. The trait space \mathcal{X} is assumed to be a **compact** subset of \mathbb{R}^l , $l \geq 1$. For any $x, y \in \mathcal{X}$, we introduce the following biological parameters

$\lambda(x) \in \mathbb{R}_+$ is the rate of birth from an individual holding trait x .

$\mu(x) \in \mathbb{R}_+$ is the rate of “natural” death for an individual holding trait x .

$r(x) := \lambda(x) - \mu(x)$ is the “natural” growth rate of trait x .

$K \in \mathbb{N}$ is a parameter scaling the population size and the resources.

$\frac{\alpha(x,y)}{K} \in \mathbb{R}_+$ is the competition kernel representing the pressure felt by an individual holding trait x from an individual holding trait y . It is not assumed to be a symmetric function.

$u_K p(x)$ with $u_K, p(x) \in (0, 1]$, is the probability that a mutation occurs in a birth from an individual with trait x . Small u_K means rare mutations.

$m(x, h)dh$ is the mutation law of a mutant trait $x + h \in \mathcal{X}$, born from an individual with trait x . Its support is a subset of the compact set $\{y - z : y, z \in \mathcal{X}\}$.

We consider, at any time $t \geq 0$, a finite number N_t of individuals, each of them holding a trait value in \mathcal{X} . Let us denote by x_1, \dots, x_{N_t} the trait values of these individuals. The state of the population at time $t \geq 0$, rescaled by K , is described by the finite point measure on \mathcal{X}

$$\nu_t^K = \frac{1}{K} \sum_{i=1}^{N_t} \delta_{x_i}, \quad (2.1)$$

where δ_x is the Dirac measure at x . Let $\langle \nu, f \rangle$ denote the integral of the measurable function f with respect to the measure ν and $\text{Supp}(\nu)$ denote its support.

Then $\langle \nu_t^K, \mathbf{1} \rangle = \frac{N_t}{K}$ and for any $x \in \mathcal{X}$, the positive number $\langle \nu_t^K, \mathbf{1}_{\{x\}} \rangle$ is called **the density** at time t of trait x .

Let \mathcal{M}_F denote the set of finite nonnegative measures on \mathcal{X} , equipped with the weak topology, and define

$$\mathcal{M}^K = \left\{ \frac{1}{K} \sum_{i=1}^n \delta_{x_i} : n \geq 0, x_1, \dots, x_n \in \mathcal{X} \right\}.$$

An individual holding trait x in the population ν_t^K gives birth to another individual with rate $\lambda(x)$ and dies with rate

$$\mu(x) + \int \alpha(x, y) \nu_t^K(dy) = \mu(x) + \frac{1}{K} \sum_{i=1}^{N_t} \alpha(x, x_i).$$

The parameter K scales the strength of competition, thus allowing the coexistence of more individuals in the population. A newborn holds the same trait value as its progenitor with probability $1 - u_K p(x)$, and with probability $u_K p(x)$, the newborn is a mutant whose trait value y is chosen according to $y = x + h$, where h is a random variable with law $m(x, h)dh$. In other words, the process $(\nu_t^K, t \geq 0)$ is a \mathcal{M}^K -valued Markov process with infinitesimal

generator defined for any bounded measurable function ϕ from \mathcal{M}^K to \mathbb{R} by

$$\begin{aligned}
L^K \phi(\nu) &= \int_{\mathcal{X}} \left(\phi \left(\nu + \frac{\delta_x}{K} \right) - \phi(\nu) \right) (1 - u_K p(x)) \lambda(x) K \nu(dx) \\
&+ \int_{\mathcal{X}} \int_{\mathbb{R}^l} \left(\phi \left(\nu + \frac{\delta_{x+h}}{K} \right) - \phi(\nu) \right) u_K p(x) \lambda(x) m(x, h) dh K \nu(dx) \\
&+ \int_{\mathcal{X}} \left(\phi \left(\nu - \frac{\delta_x}{K} \right) - \phi(\nu) \right) \left(\mu(x) + \int_{\mathcal{X}} \alpha(x, y) \nu(dy) \right) K \nu(dx). \tag{2.2}
\end{aligned}$$

For $\nu \in \mathcal{M}^K$, the integrals with respect to $K\nu(dx)$ in (2.2) correspond to sums over all individuals in the population. The first term (linear) describes the births without mutation, the second term (linear) describes the births with mutation, and the third term (non-linear) describes the deaths due to age or to competition. The density-dependent non-linearity of the third term models the competition in the population, and hence drives the selection process.

Let us denote by (A) the following three assumptions

(A1) λ , μ and α are measurable functions, and there exist $\bar{\lambda}, \bar{\mu}, \bar{\alpha} < +\infty$ such that

$$\lambda(\cdot) \leq \bar{\lambda}, \quad \mu(\cdot) \leq \bar{\mu} \quad \text{and} \quad \alpha(\cdot, \cdot) \leq \bar{\alpha}.$$

(A2) $r(x) = \lambda(x) - \mu(x) > 0$ for any $x \in \mathcal{X}$, and there exists $\underline{\alpha} > 0$ such that $\underline{\alpha} \leq \alpha(\cdot, \cdot)$.

(A3) There exists a function $\bar{m} : \mathbb{R}^l \rightarrow \mathbb{R}_+$ such that $m(x, h) \leq \bar{m}(h)$ for any $x \in \mathcal{X}$ and $h \in \mathbb{R}^l$, and $\int \bar{m}(h) dh < \infty$.

For fixed K , under (A1) and (A3) and assuming that $\mathbb{E}(\langle \nu_0^K, \mathbf{1} \rangle) < \infty$, the existence and uniqueness in law of a process on $\mathbb{D}(\mathbb{R}_+, \mathcal{M}^K)$ with infinitesimal generator L^K has been proved in [14]. Assumption (A2) prevents the population from exploding or becoming extinct too fast.

2.2 An example

The general birth-death-competition-mutation process described above has been heuristically studied in various ecological or evolutionary contexts. Let us illustrate the phenomenon of evolutionary branching with a simple example, where the parameters of the model are adapted from a classical model of competition for resources (Roughgarden [31], Dieckmann and Doebeli [9]). In this model, there is a single optimal trait value for the birth rate and a symmetric competition kernel. The parameters are the following:

$$\begin{aligned}
\mathcal{X} &= [-2, 2]; \quad \mu(x) \equiv 0; \quad p(x) \equiv p, \\
\lambda(x) &= \exp(-x^2/2\sigma_b^2), \\
\alpha(x, y) &= \tilde{\alpha}(x - y) = \exp(-(x - y)^2/2\sigma_a^2).
\end{aligned} \tag{2.3}$$

and $m(x, h)dh$ is the law of a $\mathcal{N}(0, \sigma^2)$ r.v. Y (centered Gaussian with variance σ^2) conditioned on $x + Y \in \mathcal{X}$.

The growth rate $\lambda(x)$ is maximal at $x = 0$ and there is local competition between traits, in the sense that $\alpha(x, y)$ is maximal for $x = y$ and is small when $|x - y|$ is large. If the competition kernel was flat ($\alpha \equiv 1$), evolution would favor mutant traits with maximal growth rate. However, if competition is local, numerical simulations of the microscopic model give different patterns, as shown in Fig. 2.1. In Fig. 2.1(b), the population, initially composed of traits concentrated around a single trait value, is driven by the evolutionary forces to states where the population is composed of two (or more) groups, concentrated around different trait values. This phenomenon of evolutionary branching has been observed in many biological models (see e.g. [28, 26, 19]), including populations with sexual reproduction (see e.g. [9]), for which this phenomenon is believed to be a possible mechanism leading to sympatric speciation (speciation without geographical separation).

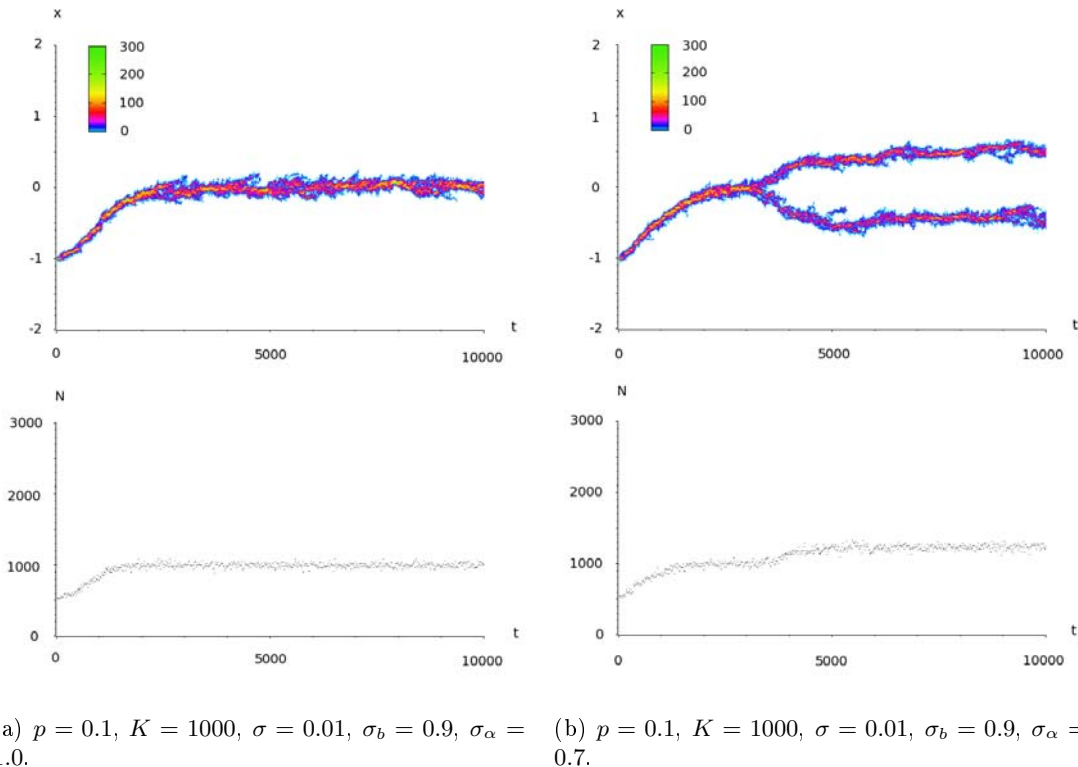


Figure 2.1: Numerical simulations of the trait distribution (upper panels) and population size (lower panels) of the microscopic model with parameters (2.3). The initial population is composed of K individuals all with trait -1.0 .

In this particular model, the possibility of evolutionary branching appears numerically to be governed by the values of σ_b and σ_α , which represent respectively the width of the trait region with high growth rate and the interaction range. In Fig. 2.1(a), $\sigma_\alpha > \sigma_b$ and there is no evolutionary branching, whereas in Fig. 2.1(b), $\sigma_\alpha < \sigma_b$ and evolutionary branching occurs. We observe in both simulations that, in a first phase, the population trait support is concentrated around a single trait value that converges to 0. In a second phase, new mutants are subject to two different selective pressures: high growth rate (traits close to 0) and competition (traits far from the rest of the population). If σ_α is small enough, the

decrease in competitive pressure compensates the loss of reproductive efficiency near 0 and allows the appearance of new branches.

In order to analyze the phenomenon of evolutionary branching, we are going to consider three biological asymptotics in the individual-based model: large population ($K \rightarrow +\infty$), rare mutations ($u_K \rightarrow 0$) and small mutation amplitude (h replaced by εh with $\varepsilon \rightarrow 0$). These scales and the biological heuristics of this approach were introduced in [28]. The combination of the first two scales (large population and rare mutation) will give convergence of the individual-based process to the so-called *polymorphic evolution sequence* (PES, Theorem 2.7 of Section 2.4.2). This convergence corresponds to approximating the simulated dynamics of Fig. 2.1(a) and (b) by the one of Fig. 2.2(a) and (b), respectively. Note that the probability of mutation p has been drastically reduced in Fig. 2.2. The reason why σ has been also increased is to make the jumps in the population state visible. Simulations with $\sigma = 0.01$ and $p = 0.0001$ would also show jumps, but of course much smaller ones. The biological heuristic of the convergence when $K \rightarrow \infty$ and $u_K \rightarrow 0$ is the following. Firstly, the assumption of rare mutations implies a separation between ecological (or population dynamics) and evolutionary (or mutation) time scales: the selection process has sufficient time between two mutations to eliminate disadvantaged traits. Secondly, the large population assumption allows one to assume a deterministic population dynamics between mutations, so that the outcome of the competition can be predicted. Then evolution proceeds by a succession of phases of mutant invasion and very short phases of competition between traits, and only few traits remain after competition between each mutation. The only randomness remaining in the system comes from the mutation times and the mutant traits appearing in the population.

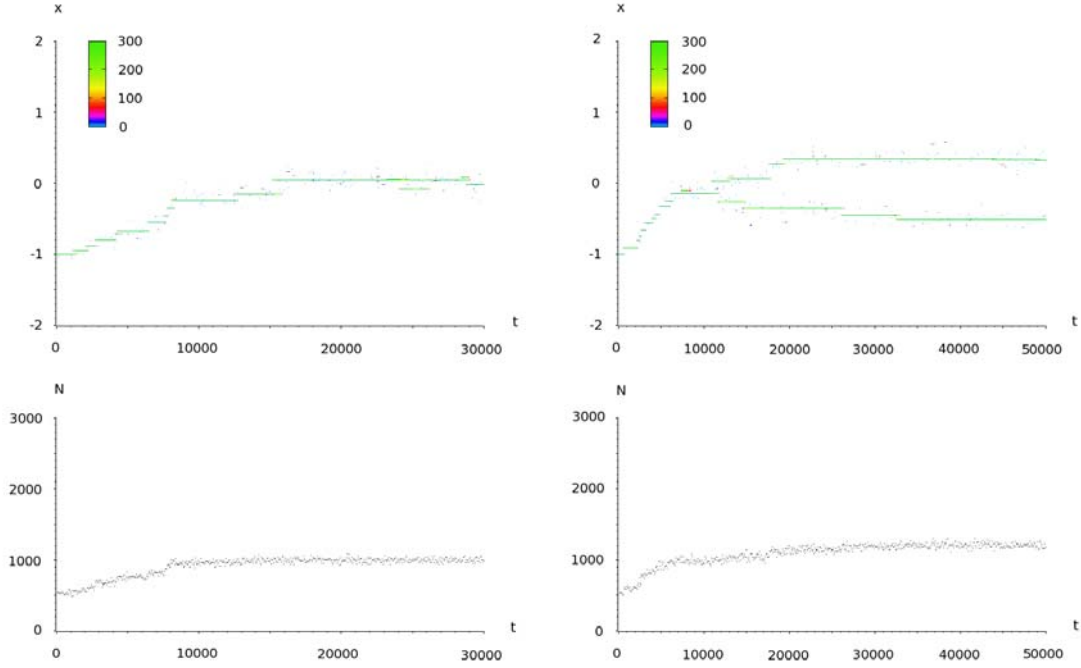
The PES is a convenient tool to study evolutionary branching: in Section 4.2, we give a precise definition for evolutionary branching in the PES (Definition 4.8), and we then prove a branching criterion (Theorem 4.9), which turns out to reduce to the variance criterion in this specific example.

Other simulations for small mutations are given in Fig. 2.3. In Fig. 2.3(a), we observe that for smaller σ_α , several evolutionary branching events can occur at different times. This shows that evolutionary branching may be a transient property, not necessarily well captured by invariant distribution properties. Note that, although our results on evolutionary branching are restricted to the first branching event, the PES obtained in Section 2.4.2 is well-defined in situations where more than two “branches” coexist in the population, as in Fig. 2.3(a). In Fig. 2.3(b), the size of mutation jumps σ is divided by 2 compared with Fig. 2.2(b). The simulation shows that the amount of time needed for the population to approach the trait value 0 is roughly multiplied by 4 with respect to Fig. 2.2(b). This will be mathematically justified in Theorem 4.4 of Section 4.1, where we prove that the support of the PES with time rescaled as t/σ^2 converges when $\sigma \rightarrow 0$ to the solution of an ordinary differential equation, known as the *canonical equation of adaptive dynamics*.

2.3 Lotka-Volterra systems

As the previous heuristic argument shows, we need to study the competition between a finite number of traits, say x_1, \dots, x_d , between two mutations. We are thus led to study the individual-based process without mutation ($u_K = 0$ for all $K \geq 1$).

So fix $x_1, \dots, x_d \in \mathcal{X}$ and assume that, for all $i \in \{1, \dots, d\}$, $\langle \nu_0^K, \mathbf{1}_{\{x_i\}} \rangle$ has bounded



(a) $p = 0.0001$, $K = 1000$, $\sigma = 0.08$, $\sigma_b = 0.9$, $\sigma_\alpha = 1.0$. (b) $p = 0.0001$, $K = 1000$, $\sigma = 0.08$, $\sigma_b = 0.9$, $\sigma_\alpha = 0.7$.

Figure 2.2: Numerical simulations of the trait distribution (upper panels) and population size (lower panels) of the microscopic model with parameters (2.3). The initial population is composed of K individuals all with trait -1.0 .

second-order moments and converges in distribution to $n_i(0) \in \mathbb{R}_+$. Then, as proved in [6, Thm.4.2], when $K \rightarrow +\infty$, the process $(\langle \nu_t^K, \mathbf{1}_{\{x_1\}} \rangle, \dots, \langle \nu_t^K, \mathbf{1}_{\{x_d\}} \rangle)$ converges in distribution for the Skorohod topology to the solution of the d -dimensional competitive Lotka-Volterra system $LV(d, \mathbf{x})$ with initial condition $(n_1(0), \dots, n_d(0))$, defined below.

Definition 2.1 For any $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{X}^d$, we denote by $LV(d, \mathbf{x})$ the competitive Lotka-Volterra system defined by

$$\dot{\mathbf{n}}(t) = F^{\mathbf{x}}(\mathbf{n}(t)), \quad t \geq 0, \quad (2.4)$$

where $\mathbf{n}(t) = (n_1(t), \dots, n_d(t))$,

$$F_i^{\mathbf{x}}(\mathbf{n}) := n_i G_i^{\mathbf{x}}(\mathbf{n}) \quad \text{where} \quad G_i^{\mathbf{x}}(\mathbf{n}) := r(x_i) - \sum_{j=1}^d \alpha(x_i, x_j) n_j. \quad (2.5)$$

The equilibria of $LV(d, \mathbf{x})$ are given by the intersection of hyperplanes $(P_i)_{1 \leq i \leq d}$, where P_i has equation either $n_i = 0$ or $G_i^{\mathbf{x}}(\mathbf{n}) = 0$. Let us introduce the following notion of coexisting traits.

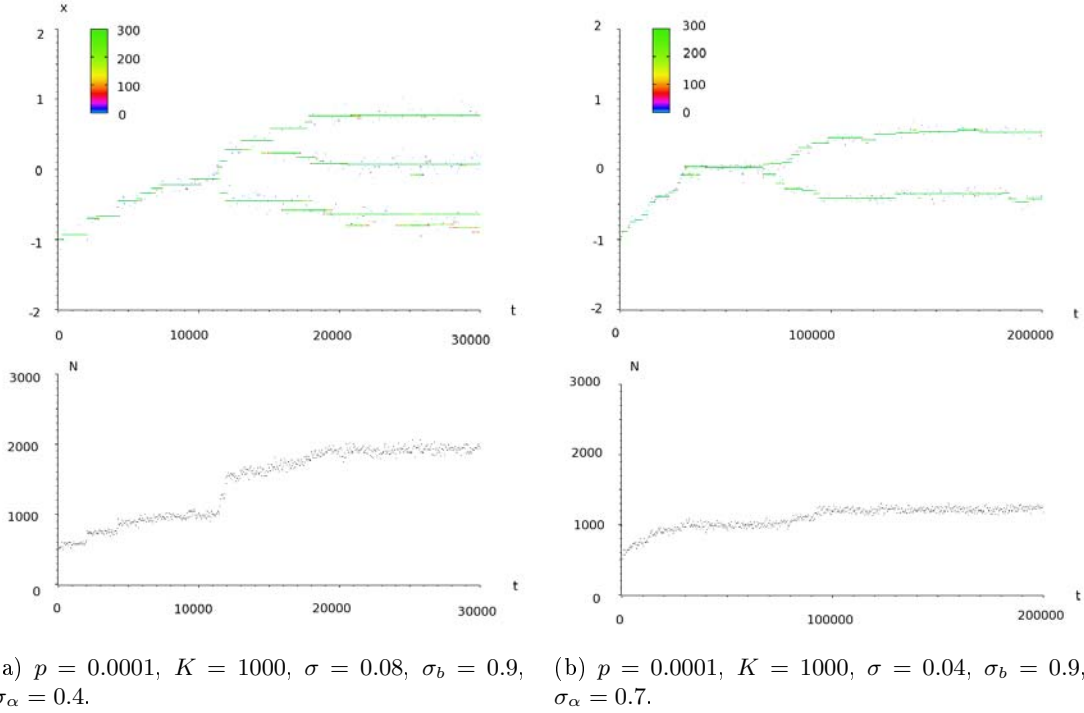


Figure 2.3: Numerical simulations of the trait distribution (upper panels) and population size (lower panels) of the microscopic model with parameters (2.3). The initial population is composed of K individuals all with trait -1.0 .

Definition 2.2 For any $d \geq 0$, we say that x_1, \dots, x_d *coexist* if $LV(d, \mathbf{x})$ admits a unique non-trivial equilibrium $\bar{\mathbf{n}}(\mathbf{x}) \in (0, \infty)^d$ which is locally strongly stable, in the sense that the eigenvalues of the Jacobian matrix of $F^{\mathbf{x}}$ at $\bar{\mathbf{n}}(\mathbf{x})$ have all (strictly) negative real part. In particular, for all $i \in \{1, \dots, d\}$,

$$G_i^{\mathbf{x}}(\bar{\mathbf{n}}(\mathbf{x})) = 0 \quad \text{and} \quad DF^{\mathbf{x}}(\bar{\mathbf{n}}(\mathbf{x})) = ((-\alpha(x_i, x_j)\bar{n}_i(\mathbf{x})))_{1 \leq i, j \leq d}. \quad (2.6)$$

Once $x_1, \dots, x_d \in \mathcal{X}$ coexist, the invasion of a mutant trait $y \in \mathcal{X}$ in this resident population is governed by its initial growth rate. This growth rate is given by the so-called fitness function defined by

$$f(y; \mathbf{x}) = f(y; x_1, \dots, x_d) = r(y) - \sum_{j=1}^d \alpha(y, x_j)\bar{n}_j(\mathbf{x}). \quad (2.7)$$

The stability of the equilibria of Lotka-Volterra systems is governed by the sign of this function.

Proposition 2.3 Assume that the traits $x_1, \dots, x_d \in \mathcal{X}$ coexist. Then

- (i) For any $i \in \{1, \dots, d\}$, $f(x_i; x_1, \dots, x_d) = 0$.

- (ii) If $f(y; x_1, \dots, x_d) < 0$, the equilibrium $(\bar{n}_1(\mathbf{x}), \dots, \bar{n}_d(\mathbf{x}), 0)$ of $LV(d+1, (x_1, \dots, x_d, y))$ is locally strongly stable, and if $f(y; x_1, \dots, x_d) > 0$, this equilibrium is unstable, in the sense that the Jacobian matrix of the system at this point has one positive eigenvalue.

Proof The first point is immediate. The second point comes from the following relation between Jacobian matrices of Lotka-Volterra systems

$$DF^{(x_1, \dots, x_d, y)}(\bar{n}_1(\mathbf{x}), \dots, \bar{n}_d(\mathbf{x}), 0) = \left(\begin{array}{c|c} DF^{\mathbf{x}}(\bar{\mathbf{n}}(\mathbf{x})) & \begin{array}{c} -\bar{n}_1(\mathbf{x})\alpha(x_1, y) \\ \vdots \\ -\bar{n}_d(\mathbf{x})\alpha(x_d, y) \end{array} \\ \hline \begin{array}{ccc} 0 & \dots & 0 \end{array} & f(y; \mathbf{x}) \end{array} \right).$$

Since x_1, \dots, x_d coexist, all the eigenvalues of $DF^{\mathbf{x}}(\bar{\mathbf{n}}(\mathbf{x}))$ have negative real parts. \square

Examples

1. In the monomorphic case ($d = 1$), the competitive Lotka-Volterra system $LV(1, x)$ takes the form of the so-called logistic equation

$$\dot{n}_x = n_x(r(x) - \alpha(x, x)n_x). \quad (2.8)$$

When $r(x) > 0$, the unique stable equilibrium of this equation is $\bar{n}(x) = r(x)/\alpha(x, x)$.

2. In the dimorphic case ($d = 2$), the system $LV(2, (x, y))$ takes the form

$$\begin{cases} \dot{n}_x = n_x(r(x) - \alpha(x, x)n_x - \alpha(x, y)n_y) \\ \dot{n}_y = n_y(r(y) - \alpha(y, x)n_x - \alpha(y, y)n_y). \end{cases} \quad (2.9)$$

Under Assumption (A2) and if $\alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x) \neq 0$, the equilibria of (2.9) are $(0, 0)$, $(\bar{n}(x), 0)$, $(0, \bar{n}(y))$ and a non-trivial equilibrium $\bar{\mathbf{n}}(x, y)$, which may possibly belong to $(0, \infty)^2$, with

$$\bar{n}_1(x, y) = \frac{r(x)\alpha(y, y) - r(y)\alpha(x, y)}{\alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x)}, \quad (2.10)$$

$$\bar{n}_2(x, y) = \frac{r(y)\alpha(x, x) - r(x)\alpha(y, x)}{\alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x)}. \quad (2.11)$$

The fitness function takes the form

$$f(y; x) = r(y) - \alpha(y, x)\bar{n}(x). \quad (2.12)$$

By Proposition 2.3, the stability of the equilibrium $(\bar{n}(x), 0)$ is governed by the sign of $f(y; x)$, and

$$f(x; x) = 0, \quad \forall x \in \mathcal{X}.$$

Moreover, the fitness function also gives a criterion for coexistence in the dimorphic case.

Proposition 2.4 *There is coexistence in the system $LV(2, (x, y))$ if and only if*

$$f(y; x) > 0 \quad \text{and} \quad f(x; y) > 0.$$

We refer to [24, Sec. 2.4.3] for a proof of this result.

3. In the trimorphic case ($d = 3$), the fitness of a mutant trait z in a population with two coexisting resident traits x and y is given by

$$f(z; x, y) = r(z) - \alpha(z, x)\bar{n}_1(x, y) - \alpha(z, y)\bar{n}_2(x, y), \quad (2.13)$$

where $\bar{n}_1(x, y)$ and $\bar{n}_2(x, y)$ are defined in (2.10) and (2.11).

2.4 Convergence to the Polymorphic Evolution Sequence (PES)

Our goal here is to examine the asymptotic behavior of the microscopic process when the population size grows to infinity as well as the mutation rate converges to 0, in a long time scale.

2.4.1 Assumptions

Let (B) denote the following Assumptions (B1) and (B2) on the Lotka-Volterra systems of Definition 2.1.

- (B1) Given any $\mathbf{x} = (x_1, \dots, x_d) \in \mathcal{X}^d$ such that x_1, \dots, x_d coexist, for Lebesgue almost any mutant trait $y \in \mathcal{X}$ such that $f(y; \mathbf{x}) > 0$, there exists a neighborhood $\mathcal{U} \subset \mathbb{R}^{d+1}$ of $(\bar{n}^1(\mathbf{x}), \dots, \bar{n}^d(\mathbf{x}), 0)$ such that all the solutions of $LV(d+1, (x_1, \dots, x_d, y))$ with initial condition in $\mathcal{U} \cap (0, \infty)^{d+1}$ converge as $t \rightarrow +\infty$ to a unique equilibrium in $(\mathbb{R}_+)^{d+1}$, denoted by

$$\mathbf{n}^*(x_1, \dots, x_d, y).$$

- (B2) Writing for simplicity $x_{d+1} = y$ and \mathbf{n}^* for $\mathbf{n}^*(x_1, \dots, x_{d+1})$, let

$$I(\mathbf{n}^*) := \{i \in \{1, \dots, d+1\} : n_i^* > 0\} \quad \text{and} \quad \mathbf{x}^* = (x_i; i \in I(\mathbf{n}^*)).$$

Then, for Lebesgue almost any mutant trait x_{d+1} as above, $\{x_i; i \in I(\mathbf{n}^*)\}$ coexist and

$$\text{for all } j \notin I(\mathbf{n}^*), \quad f(x_j; \mathbf{x}^*) < 0.$$

Assumption (B1) prevents cycles or chaotic dynamics in the Lotka-Volterra systems. Moreover, it also prevents situations as in Fig. 2.4, where the equilibrium \mathbf{n}^* is unstable. In this case, a solution of the Lotka-Volterra system $LV(d+1, (x_1, \dots, x_d, y))$ starting from a point in any neighborhood of $(\bar{n}^1(\mathbf{x}), \dots, \bar{n}^d(\mathbf{x}), 0)$, represented by the curved line in Fig. 2.4, need not converge to \mathbf{n}^* .

Definition 2.5 *An equilibrium \mathbf{n} of $LV(d, (x_1, \dots, x_d))$ is hyperbolic if the Jacobian matrix of $LV(d, (x_1, \dots, x_d))$ at \mathbf{n} has no eigenvalue with 0 real part.*

Assumption (B2) can also be replaced by one of the following two simpler assumptions.

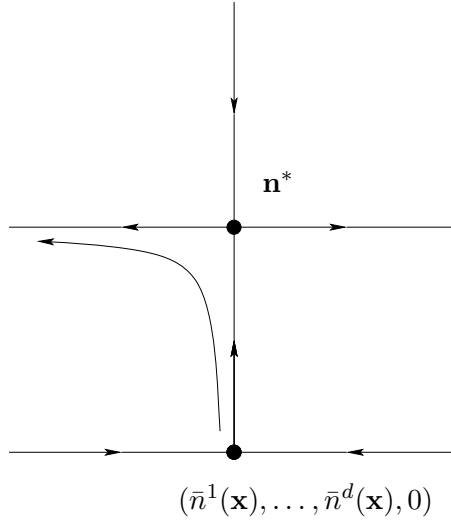


Figure 2.4: Assumption (B1) prevents such situation.

(B3) For Lebesgue almost any mutant trait x_{d+1} as in (B1), \mathbf{n}^* is hyperbolic.

(B4) For Lebesgue almost any mutant trait x_{d+1} as in (B1), \mathbf{n}^* is strongly locally stable.

Proposition 2.6 *Assumptions (B1) and (B2) are equivalent to Assumptions (B1) and (B3), and to Assumptions (B1) and (B4).*

Proof Let $k := \text{Card}(I(\mathbf{n}^*))$. Assume that x_1, \dots, x_{d+1} are reordered in a way such that $I(\mathbf{n}^*) = \{1, 2, \dots, k\}$. Then it is clear, by the definition of coexistence and the fact that

$$DF^{(x_1, \dots, x_{d+1})}(\mathbf{n}^*) = \left(\begin{array}{c|ccc} DF^{\mathbf{x}^*}(n_1^*, \dots, n_k^*) & (-\alpha(x_i, x_j)n_j^*)_{1 \leq i \leq k, k+1 \leq j \leq d+1} & & \\ \hline & f(x_{k+1}, \mathbf{x}^*) & & 0 \\ & & \ddots & \\ & 0 & & f(x_{d+1}, \mathbf{x}^*) \end{array} \right)$$

that (B2) implies (B4) which also trivially implies (B3). Assuming (B3), the stable manifold theorem (see e.g. [20] pp. 13–14) says that the set of points such that the solution of $LV(d+1, (x_1, \dots, x_{d+1}))$ started at this point converges to \mathbf{n}^* is a submanifold of $(0, \infty)^{d+1}$ of dimension l , where l is the number of eigenvalues of $DF^{(x_1, \dots, x_{d+1})}(\mathbf{n}^*)$ with negative real part. In particular, if $l < d + 1$, this manifold does not contain an open set of $(0, \infty)^{d+1}$, which is in contradiction with (B1). Therefore, $l = d + 1$, which implies (B2). \square

In Section 3, various situations ensuring Assumptions (B1) and (B2) (or (B3), or (B4)) will be discussed.

2.4.2 Definition of the PES and Convergence Theorem

We define $\mathcal{M}_0 \subset \mathcal{M}_F$ by

$$\mathcal{M}_0 := \left\{ \sum_{i=1}^d \bar{n}_i(\mathbf{x}) \delta_{x_i}; d \geq 1, x_1, \dots, x_n \in \mathcal{X} \text{ coexist} \right\}.$$

Theorem 2.7 *Assume (A) and (B). Take $x_1, \dots, x_d \in \mathcal{X}$ that coexist and assume that $\nu_0^K = \sum_{i=1}^d n_i^K \delta_{x_i}$ with $n_i^K \rightarrow \bar{n}_i(\mathbf{x})$ in probability for all $1 \leq i \leq d$. Assume finally that*

$$\forall V > 0, \quad \log K \ll \frac{1}{Ku_K} \ll \exp(VK), \quad \text{as } K \rightarrow \infty. \quad (2.14)$$

Then, $(\nu_{t/Ku_K}^K; t \geq 0)$ converges to the \mathcal{M}_0 -valued Markov pure jump process $(\Lambda_t; t \geq 0)$ defined as follows: $\Lambda_0 = \sum_{i=1}^d \bar{n}_i(\mathbf{x}) \delta_{x_i}$ and the process Λ jumps for all $j \in \{1, \dots, d\}$

$$\text{from } \sum_{i=1}^d \bar{n}_i(\mathbf{x}) \delta_{x_i} \text{ to } \sum_{i=1}^d n_i^*(x_1, \dots, x_d, x_j + h) \delta_{x_i} + n_{d+1}^*(x_1, \dots, x_d, x_j + h) \delta_{x_j+h}$$

with infinitesimal rate

$$p(x_j) \lambda(x_j) \bar{n}_j(\mathbf{x}) \frac{[f(x_j + h; \mathbf{x})]_+}{\lambda(x_j + h)} m(x_j, h) dh. \quad (2.15)$$

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

Remark that, in general, the convergence cannot hold for the Skorohod topology (even if the space \mathcal{M}_F is equipped with the weak topology). Indeed, it can be checked in explicit examples that the total mass of the limit process has jumps. This is in contradiction with the C -tightness of the sequence $(\nu_{t/Ku_K}^K, t \geq 0)$, which would hold in case of convergence for the Skorohod topology.

The infinitesimal generator of the process $(\Lambda_t, t \geq 0)$ is given by

$$\begin{aligned} \mathcal{L}\varphi \left(\sum_{i=1}^d \bar{n}_i(\mathbf{x}) \delta_{x_i} \right) &= \int_{\mathcal{X}} dh \sum_{j=1}^d p(x_j) \lambda(x_j) \bar{n}_j(\mathbf{x}) \frac{[f(x_j + h; \mathbf{x})]_+}{\lambda(x_j + h)} m(x_j, h) \times \\ &\left(\varphi \left(\sum_{i=1}^d n_i^*(x_1, \dots, x_d, x_j + h) \delta_{x_i} + n_{d+1}^*(x_1, \dots, x_d, x_j + h) \delta_{x_j+h} \right) - \varphi \left(\sum_{i=1}^d \bar{n}_i(\mathbf{x}) \delta_{x_i} \right) \right), \end{aligned} \quad (2.16)$$

for all measurable bounded function φ .

We call this process *Polymorphic Evolution Sequence* (PES), by analogy with the so-called ‘‘Trait Substitution Sequence’’ (TSS) of Section 6.4 of [28] (see also Section 3.1 below). Recall that the equilibrium \mathbf{n}^* needs not to have all nonzero coordinates, which means that the number of traits in the support of the PES may not be monotonous.

Note that it follows from Assumption (A) and from (2.6) that the jump rates of the process Λ are bounded. Moreover, by Assumption (B2), for almost all mutant traits y such that

$f(y; \mathbf{x}) > 0$, $\sum_{i=1}^d n_i^*(x_1, \dots, x_d, y) \delta_{x_i} + n_{d+1}^*(x_1, \dots, x_d, y) \delta_y \in \mathcal{M}_0$. Thus, the PES is well-defined on \mathbb{R}_+ and belongs almost surely to \mathcal{M}_0 for all time.

We now give the general idea of the proof, extending the biological heuristics of [28]. We refer to Appendix A for the detailed proof.

Let us roughly describe the successive steps of mutation, invasion and competition. The two steps of the invasion of a mutant in a given population are firstly the stabilization of the resident population before the mutation and secondly the invasion of the mutant population after the mutation.

Fix $\eta > 0$. In the first step, assuming that d traits x_1, \dots, x_d that coexist are present, we prove that the population densities $(\langle \nu_t^K, \mathbf{1}_{\{x_1\}} \rangle, \dots, \langle \nu_t^K, \mathbf{1}_{\{x_d\}} \rangle)$ belong to the η -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$ with high probability for large K until the next mutant y appears. To this aim, we use large deviation results on the problem of exit from a domain [15] to prove that the time needed for the population densities to leave the η -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$ is bigger than $\exp(VK)$ for some $V > 0$ with high probability. Therefore, until this exit time, the rate of mutation from trait x_i in the population is close to $u_K p(x_i) \lambda(x_i) K \bar{n}_i(\mathbf{x})$ and thus, the first mutation appears before this exit time if one assumes that

$$\frac{1}{Ku_K} \ll e^{VK}.$$

In particular, the mutation rate from trait x_i on the time scale t/Ku_K is close to

$$p(x_i) \lambda(x_i) \bar{n}_i(\mathbf{x}).$$

In the second step, we divide the invasion of a given mutant trait y into 3 phases shown in Fig. 2.5, in a similar way as done classically by population geneticists dealing with selective sweeps [25].

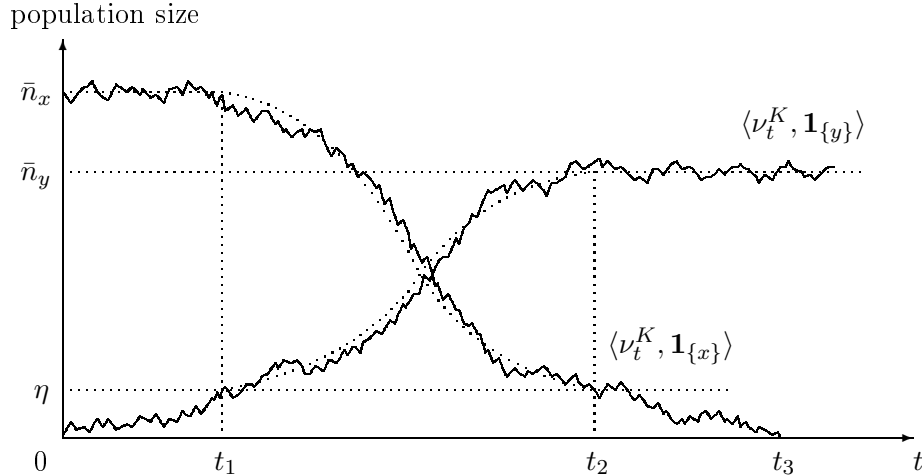


Figure 2.5: The three steps of the invasion of a mutant trait y in a monomorphic population with trait x .

In the first phase (between time 0 and t_1 in Fig. 2.5), the number of mutant individuals is small, and the resident population stays close to its equilibrium density $\bar{\mathbf{n}}(\mathbf{x})$. Therefore,

the dynamics of the mutant individuals is close to a branching process with birth rate $\lambda(y)$ and death rate $\mu(y) + \sum_{i=1}^d \alpha(y, x_i) \bar{n}_i(\mathbf{x})$. Hence, the growth rate of this branching process is equal to the fitness $f(y; \mathbf{x})$ of (2.7), describing the ability of the initially rare mutant trait y to invade the equilibrium resident population with traits x_1, \dots, x_d . If this fitness is positive (i.e. if the branching process is super-critical), the probability that the mutant population reaches density $\eta > 0$ at some time t_1 is close to the probability that the branching process reaches ηK , which is itself close to its survival probability $[f(y; \mathbf{x})]_+ / \lambda(y)$ when K is large.

In the second phase (between time t_1 and t_2 in Fig. 2.5), we use the fact that, when $K \rightarrow +\infty$, the population densities $(\langle \nu_t^K, \mathbf{1}_{\{x_1\}} \rangle, \dots, \langle \nu_t^K, \mathbf{1}_{\{x_d\}} \rangle, \langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle)$ are close to the solution of the Lotka-Volterra system $LV(d+1, (x_1, \dots, x_d, y))$ with the same initial condition, on any time interval $[0, T]$. Assumption (B1) ensures that, if η is sufficiently small, then any solution to the Lotka-Volterra system starting in some neighborhood of $(\bar{n}_1(\mathbf{x}), \dots, \bar{n}_d(\mathbf{x}), 0)$ converges to a new equilibrium $\mathbf{n}^* \in \mathbb{R}^{d+1}$ as time goes to infinity. Therefore, the population densities reach with high probability the η -neighborhood of \mathbf{n}^* at some time t_2 .

Finally, in the last phase, we use the same idea as in the first phase: since \mathbf{n}^* is a strongly locally stable equilibrium (Assumption (B2)), we can approximate the densities of the traits x_j such that $n_j^* = 0$ by branching processes which are sub-critical. Therefore, they reach 0 in finite time and the process comes back to the first step until the next mutation. We will prove that the duration of these three phases is of order $\log K$. Therefore, under the assumption

$$\log K \ll \frac{1}{Ku_K},$$

the next mutation occurs after these three phases with high probability.

3 Particular cases and extensions of the PES

In this section, we discuss various situations when Assumptions (B1) and (B2) are satisfied allowing one to explicitly obtain the PES.

3.1 The "no coexistence" case: an extension of the trait substitution sequence (TSS)

In this section we characterize the case when the PES is well defined until the first co-existence time of two different traits. Let us introduce the assumption

(C1) For all $x \in \mathcal{X}$, the set of y such that $f(y; x) = 0$ has Lebesgue measure 0,

where $f(y; x)$ is defined in (2.12). It will be shown below that Assumption (B) with $d = 1$ (only one resident trait) follows from (C1).

Let us now introduce the following killed PES $(\Lambda_t^{(1)}, t \geq 0)$ as a Markov jump process on $\mathcal{M}_0 \cup \{\partial\}$, where ∂ is a cemetery state, with infinitesimal generator $\mathcal{L}^{(1)}$ defined as follows.

For a bounded measurable function φ on $\mathcal{M}_0 \cup \{\partial\}$, we have $\mathcal{L}^{(1)}\varphi(\partial) = 0$ and

$$\begin{aligned} & \mathcal{L}^{(1)}\varphi(\bar{n}(x)\delta_x) \\ &= \int_{\mathcal{X}} \left(\varphi(\bar{n}(x+h)\delta_{x+h}) - \varphi(\bar{n}(x)\delta_x) \right) p(x)\lambda(x)\bar{n}(x) \frac{[f(x+h;x)]_+}{\lambda(x+h)} \mathbf{1}_{\{f(x;x+h)<0\}} m(x,h) dh \\ & \quad + \int_{\mathcal{X}} \left(\varphi(\partial) - \varphi(\bar{n}(x)\delta_x) \right) p(x)\lambda(x)\bar{n}(x) \mathbf{1}_{\{f(x;x+h)>0, f(x+h;x)>0\}} m(x,h) dh. \end{aligned} \quad (3.1)$$

By construction, the killed PES $(\Lambda_t^{(1)}, t \geq 0)$ is always monomorphic before killing. Once the killed PES reaches the cemetery state ∂ , it no longer jumps.

We have the following corollary of Theorem 2.7.

Corollary 3.1 *With the same assumption and notation as in Theorem 2.7, except that Assumption (B) is replaced by Assumption (C1) and that $d = 1$, let*

$$\tau_K := \inf\{t \geq 0 : \text{Supp}(\nu_t^K) = \{x, y\} \text{ such that } (x, y) \text{ coexist}\}.$$

Then the process

$$\left(\nu_{\frac{t}{Ku_K}}^K \mathbf{1}_{\{\frac{t}{Ku_K} \leq \tau_K\}} + \partial \mathbf{1}_{\{\frac{t}{Ku_K} > \tau_K\}}, t \geq 0 \right) \quad (3.2)$$

converges as $K \rightarrow +\infty$ to the killed PES $(\Lambda_t^{(1)}, t \geq 0)$ with initial condition $\Lambda_0^{(1)} = \bar{n}(x)\delta_x$. The convergence is understood in the same sense as in Theorem 2.7.

Proof Let us first prove that Assumption (B) for $d = 1$ is implied by (C1). The asymptotic behavior of 2-dimensional competitive Lotka-Volterra systems is well-known (see e.g. [24]):

- if $f(x;y) > 0$ and $f(y;x) < 0$, any solution of $LV(2, (x, y))$ starting from $\mathbb{R}_+ \times (0, \infty)$ converges to $\mathbf{n}^*(y, x) = (\bar{n}(x), 0)$,
- if $f(x;y) < 0$ and $f(y;x) > 0$, any solution of $LV(2, (x, y))$ starting from $(0, \infty) \times \mathbb{R}_+$ converges to $\mathbf{n}^*(x, y) = (0, \bar{n}(y))$,
- if $f(x;y) > 0$ and $f(y;x) > 0$, any solution of $LV(2, (x, y))$ starting from $(0, \infty)^2$ converges to $\mathbf{n}^*(x, y) = \mathbf{n}^*(y, x) = \bar{\mathbf{n}}(x, y)$,
- if $f(x;y) < 0$ and $f(y;x) < 0$, $(\bar{n}(x), 0)$ and $(0, \bar{n}(y))$ are both locally strongly stable.

Moreover, all the equilibria are hyperbolic if and only if $f(y;x) \neq 0$ and $f(x;y) \neq 0$. Therefore, Assumption (C1) implies Assumption (B) for $d = 1$ since $m(x, h)dh$ is absolutely continuous w.r.t. Lebesgue's measure.

Therefore, the generator (3.1) corresponds to defining the killed PES $\Lambda^{(1)}$ as the PES, and send it to the cemetery state ∂ as soon as a mutant trait $x+h$ appears, that coexists with the resident trait $x \in \mathcal{X}$. Note that ∂ is reached as soon as a mutant *appears*, that could coexist with the resident trait, even if this mutant actually does not *invade* the population. That explains why the invasion probability $[f(x+h;x)]_+/\lambda(x+h)$ does not appear in the last line of (3.1).

Note that under Assumptions (A) and (C1), the killed PES $(\Lambda_t^{(1)}, t \geq 0)$ is almost surely well-defined. The convergence of the processes (3.2) to $\Lambda^{(1)}$ is thus obtained by a proof easily adapted from the one of Theorem 2.7. \square

Remark 3.2 *The killed PES generalizes the so-called “Trait Substitution Sequence” (TSS), introduced in [28], and rigorously studied in [4]. The TSS is the support of the limiting process of the sequence (3.2) in the case when the parameters of the microscopic model prevent the coexistence of any two traits. Such an assumption, known as “Invasion-Implies-Fixation” (IIF) principle [17] is given by:*

(IIF) *for all $x \in \mathcal{X}$, almost all $y \in \mathcal{X}$ such that $f(y; x) > 0$ satisfy $f(x; y) < 0$.*

Hence, the PES Λ has the form

$$\Lambda_t = \bar{n}(X_t)\delta_{X_t}, \quad t \geq 0,$$

where X is called the TSS and is a Markov pure jump process on \mathcal{X} with infinitesimal generator

$$L\varphi(x) = \int_{\mathbb{R}^d} (\varphi(x+h) - \varphi(x))p(x)\lambda(x)\bar{n}(x)\frac{[f(x+h;x)]_+}{\lambda(x+h)}m(x,h)dh. \quad (3.3)$$

The killed PES $(\Lambda_t^{(1)}, t \geq 0)$ prevents the coexistence of two or more traits. Therefore, this process is not suited to our study of evolutionary branching in Section 4. To this end, we need to examine a more general situation.

3.2 The “no triple coexistence” case

In this section we characterize the case when the PES is well defined until the first coexistence time of three different traits. To this aim, we first extend the fitness function $f(z; x, y)$ of (2.13) to any $x, y \in \mathcal{X}$ such that $f(x; y)f(y; x) > 0$ (and not only for the ones that coexist). It can be easily checked that $\alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x)$ cannot be 0 under this condition. This extension is needed in the results of Zeeman [33] we use below.

We can now introduce the following assumption:

(C2) For all $x, y \in \mathcal{X}$ that coexist, the set of z such that $f(x; z) = 0$, $f(z; x) = 0$, $f(y; z) = 0$, $f(z; y) = 0$, $f(x; y, z) = 0$ or $f(y; x, z) = 0$ (when these last quantities are defined) has Lebesgue measure 0.

We denote by (C) the assumption (C1)+(C2).

In the case when the mutant z invades a resident population with coexisting traits x and y (i.e. if x, y, z satisfy $f(x; y) > 0$, $f(y; x) > 0$ and $f(z; x, y) > 0$ by Propositions 2.3 and 2.4), one needs to study the asymptotic behavior of the 3-dimensional competitive Lotka-Volterra system $LV(3, (x, y, z))$. The complete classification of such 3-dimensional systems has been done by Zeeman [33] in terms of the signs of the fitnesses involved in (C2). The case of coexisting x and y and invading z leads to 10 classes of asymptotic behaviors represented in Figure 3.1, labelled with the same numbers as in [33]. The precise meaning

of these pictures is given in Appendix B. Classes 7 and 8 correspond to the extinction of both resident traits and survival of the mutant; classes 9 to 12 correspond to the extinction of one resident trait and the coexistence of the two others; classes 26, 29, 31, 33 correspond to the coexistence of the three traits. Thus the triple coexistence case is described by the set of traits

$$C_{\text{coex}} := \{(x, y, z) \in \mathcal{X}^3 : LV(3, (x, y, z)) \text{ belongs to classes 26, 29, 31 or 33}\}. \quad (3.4)$$

As can be checked from Fig. 3.1, if x and y coexist and $f(z; x, y) > 0$, then $(x, y, z) \in C_{\text{coex}}$ if and only if both of the following properties are satisfied

- (P1) If $f(y; x, z)$ is well-defined, then $f(x; z)$, $f(z; x)$ and $f(y; x, z)$ have all the same sign.
(P2) If $f(x; y, z)$ is well-defined, then $f(y; z)$, $f(z; y)$ and $f(x; y, z)$ have all the same sign.

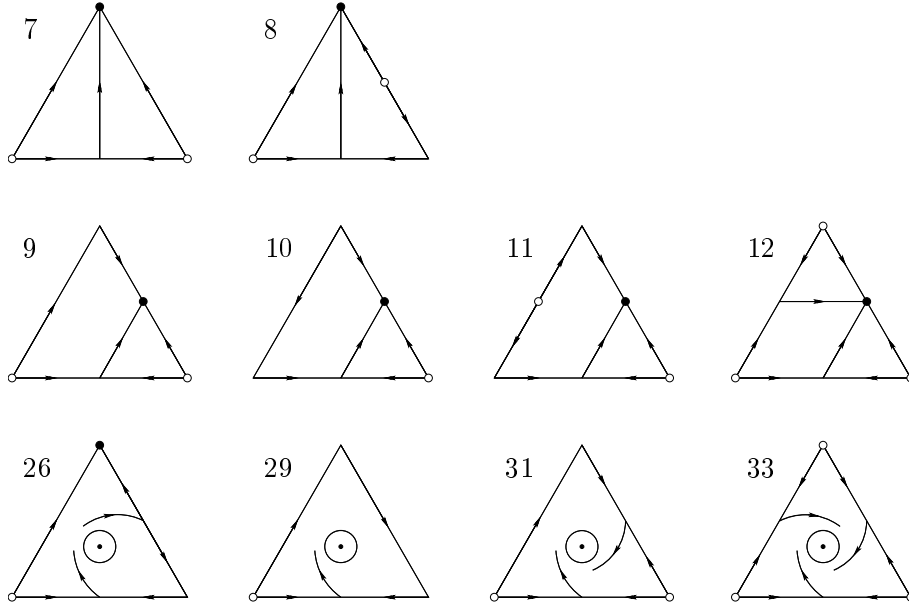


Figure 3.1: The different equivalence classes of phase portraits on Σ of 3-dimensional competitive Lotka-Volterra systems where two traits coexist and the last one has a positive fitness w.r.t. the two others.

Similarly as in Section 3.1, we define the killed PES $(\Lambda_t^{(2)}, t \geq 0)$ as a Markov pure jump process on $\mathcal{M}_0 \cup \{\partial\}$, with infinitesimal generator $\mathcal{L}^{(2)}$. The latter is given by (2.16) for $d = 1$, and for $d = 2$ and coexisting x_1, x_2 , it is modified as follows. Let $\nu := \sum_{i=1}^2 \bar{n}_i(x_1, x_2) \delta_{x_i}$, then

$$\begin{aligned} \mathcal{L}^{(2)}\varphi(\nu) &= \int_{\mathbb{R}^l} \sum_{j=1}^2 \left(\varphi \left(\sum_{i=1}^2 n_i^*(x_1, x_2, x_j + h) \delta_{x_i} + n_3^*(x_1, x_2, x_j + h) \delta_{x_j+h} \right) - \varphi(\nu) \right) \times \\ & p(x_j) \lambda(x_j) \bar{n}_j(x_1, x_2) \frac{[f(x_j + h; x_1, x_2)]_+}{\lambda(x_j + h)} \mathbf{1}_{\{(x_1, x_2, x_j+h) \notin C_{\text{coex}}\}} m(x_j, h) dh \\ & + \int_{\mathbb{R}^l} \sum_{j=1}^2 (\varphi(\partial) - \varphi(\nu)) p(x_j) \lambda(x_j) \bar{n}_j(x_1, x_2) \mathbf{1}_{\{(x_1, x_2, x_j+h) \in C_{\text{coex}}\}} m(x_j, h) dh. \end{aligned} \quad (3.5)$$

In (3.5), for $(x, y, z) \notin C_{\text{coex}}$, the notation $\mathbf{n}^*(x, y, z)$ denotes the unique stable equilibrium of $LV(\mathfrak{3}, (x, y, z))$ given by classes 7 to 12 in Figure 3.1.

This generator defines the killed PES as the PES, and sends it to the cemetery state as soon as a mutant trait x_3 appears in a dimorphic population of traits $x_1, x_2 \in \mathcal{X}$ such that the Lotka Volterra dynamics associated with traits x_1, x_2, x_3 belongs to classes 26, 29, 31 or 33. Notice that the killed PES's support has at most two traits at each time.

This killed PES is the limit of the individual-based process killed at the first triple-coexistence time.

Corollary 3.3 *With the same assumption and notation as in Theorem 2.7, except that Assumption (B) is replaced by Assumption (C) and that $d \in \{1, 2\}$, let*

$$\tilde{\tau}_K := \inf\{t \geq 0 : \text{Supp}(\nu_t^K) = \{x, y, z\} \text{ such that } (x, y, z) \in C_{\text{coex}}\}.$$

Then the process

$$\left(\nu_{\frac{t}{Ku_K}}^K \mathbf{1}_{\{\frac{t}{Ku_K} \leq \tilde{\tau}_K\}} + \partial \mathbf{1}_{\{\frac{t}{Ku_K} > \tilde{\tau}_K\}}, t \geq 0 \right) \quad (3.6)$$

converges as $K \rightarrow +\infty$ to the killed PES $(\Lambda_t^{(2)}, t \geq 0)$ with initial condition $\Lambda_0^{(2)} = \sum_{i=1}^d \bar{n}_i(\mathbf{x}) \delta_{x_i}$.

Proof We first need to check that Assumption (C) implies Assumption (B) for $d = 2$ and for all $(x, y, z) \notin C_{\text{coex}}$. Since no pattern as in Fig. 2.4 occurs in diagrams 7 to 12 in Fig. 3.1, we see that Assumption (B1) is always satisfied for $(x, y, z) \notin C_{\text{coex}}$. Moreover, as before, all the steady states are hyperbolic if all the 2- and 3-dimensional fitnesses are nonzero. Thus Assumption (B) is satisfied for all $(x, y, z) \in \mathcal{X}^3 \setminus C_{\text{coex}}$ as soon as Assumption (C2) is satisfied.

Remark that Assumption (B1) could be violated in C_{coex} , either if the interior equilibrium is surrounded by a stable cycle, or in diagram 26 in the case when the pattern of Figure 2.4 may appear.

Under Assumptions (A) and (C), the killed PES $(\Lambda_t^{(2)}, t \geq 0)$ is almost surely well-defined. The convergence of the processes (3.6) to $\Lambda^{(2)}$ is thus obtained by a proof easily adapted from the one of Theorem 2.7. \square

We will show in Section 4 that the killed PES obtained in this section is sufficient to study the phenomenon of evolutionary branching when $\mathcal{X} \subset \mathbb{R}$.

3.3 Back to our example

Let us prove that Assumption (C) is satisfied by the parameters of the example of Section 2.2.

Recall the definition of the function $\tilde{\alpha}$ in (2.3). If Assumption (C1) were not true, then, as the functions λ and $\tilde{\alpha}$ can be extended as analytic functions on \mathbb{C} , one should have

$$f(y; x) = 0, \quad \forall y \in \mathcal{X} = [-2, 2],$$

for at least one $x \in [-2, 2]$. In view of Formula (2.12) for $f(y; x)$, this implies that the functions r (which is λ in this case) and $y \mapsto \tilde{\alpha}(y - x)$ are linearly dependent, which cannot hold unless $\sigma_b = \sigma_\alpha$.

Similarly, if Assumption (C2) were not true, one would have either

$$f(x; z) = 0, \quad \forall z \in \mathcal{X} = [-2, 2],$$

for some $x \in [-2, 2]$, or

$$f(x; y, z) = 0, \quad \forall z \in \mathcal{X} = [-2, 2],$$

for some $x \neq y \in [-2, 2]$. In the first case, this would mean that the function $z \mapsto \lambda(z)\tilde{\alpha}(x - z)$ is constant, as $\alpha(z, z) = \tilde{\alpha}(0) = 1$. But this does not hold. In the second case, in view of (2.13), (2.10) and (2.11), one would have that the family of functions

$$\{1; z \mapsto \tilde{\alpha}(z - y)^2; z \mapsto \lambda(z)\tilde{\alpha}(y - z); z \mapsto \lambda(z)\tilde{\alpha}(x - z); z \mapsto \tilde{\alpha}(x - z)\tilde{\alpha}(y - z)\}$$

is linearly dependent. After some algebra, we want to prove that, if for some $a_0, \dots, a_4 \in \mathbb{R}$,

$$a_0 e^{z^2/\sigma_\alpha^2} + a_1 e^{2yz/\sigma_\alpha^2} + a_2 e^{(x+y)z/\sigma_\alpha^2} + e^{-(\sigma_b^{-2} - \sigma_\alpha^{-2})z^2/2} \left(a_3 e^{yz/\sigma_\alpha^2} + a_4 e^{xz/\sigma_\alpha^2} \right) = 0$$

for all $z \in \mathbb{R}$, then $a_0 = \dots = a_4 = 0$. If $\sigma_b < \sigma_\alpha$, this can be easily deduced from asymptotic considerations when $z \rightarrow +\infty$ and $-\infty$. The case $\sigma_b > \sigma_\alpha$ can be handled similarly.

Therefore, we have proved that, provided $\sigma_b \neq \sigma_\alpha$, the example of Section 2.2 satisfies Assumption (C).

Note that in higher dimensions l , the previous argument, based on analytic functions properties, cannot apply. However, the same method could apply for other large classes of functions, for example if the functions λ and $\tilde{\alpha}$ were polynomials in $x \in \mathcal{X}$.

4 Evolutionary branching and small jumps

Our main goal in this section is to study the PES of Theorem 2.7 in order to establish an evolutionary branching criterion. We will assume, in all that follows, that the initial population is monomorphic (at time 0, all individuals have the same trait).

We have seen in Section 3.1 that, as long as there is no coexistence of two traits in the population, the support of the PES Λ is reduced to a single trait and $\Lambda = \Lambda^{(1)}$, where $\Lambda^{(1)}$ is the killed PES with generator (3.1). In this section, our aim is to characterize the traits around which coexistence is possible and how evolutionary branching can then occur, as observed in Fig. 2.2(b). To do so, following a general idea of the biological literature [28, 10, 18, 17, 11, 16], a key assumption is that the mutation amplitude is small. Under this assumption, we study the behavior of the PES on large time scales, which allows us to observe global evolutionary dynamics.

In Subsection 4.1, we prove that when ε tends to zero, the TSS with small mutation steps scaled by ε , converges on the time scale $\frac{t}{\varepsilon^2}$, to the solution of a (deterministic) ODE, called *the canonical equation of adaptive dynamics*, or, more simply *the canonical equation*. We then obtain a similar result for the PES in dimension $l = 1$, and we show that evolutionary branching can only occur on a longer time scale and in the neighborhood of specific points of the trait space called *evolutionary singularities*. In Subsection 4.2, we give a precise definition of evolutionary branching and we prove the main result of this section, giving a criterion for evolutionary branching in the limit of small mutational jumps.

First, we introduce the following additional Assumptions (A'):

(A'1) There exists $\bar{\varepsilon} > 0$ such that for all $x \in \mathcal{X}$ and $h \in \text{Supp}(m(x, \cdot))$, $x + \varepsilon h \in \mathcal{X}$ for $\varepsilon \leq \bar{\varepsilon}$. This holds for example if \mathcal{X} is convex for $\bar{\varepsilon} = 1$.

(A'2) The map $(x, h) \mapsto m(x, h)$ is Lipschitz continuous on $\mathcal{X} \times \mathbb{R}^d$.

(A'3) The functions $\lambda(\cdot)$ and $\mu(\cdot)$ are \mathcal{C}^3 on \mathcal{X} and the function $\alpha(\cdot, \cdot)$ is \mathcal{C}^4 on \mathcal{X}^2 .

Later in this section, we will also need Assumption (A''):

(A'') We assume that the trait space is one-dimensional ($l = 1$) and that, for any x in the interior of \mathcal{X} , $\int_{\mathbb{R}_-} m(x, h)dh > 0$ and $\int_{\mathbb{R}_+} m(x, h)dh > 0$.

The assumption that $l = 1$ is required as our proof of the evolutionary branching criterion is based on monotonicity properties, which can be only used in dimension 1. The second part of this assumption means that mutants can appear from a trait x on both sides of x , which is biologically relevant.

Finally, let us introduce the parameter $\varepsilon \in (0, \bar{\varepsilon}]$ scaling the size of mutation. Thanks to Assumption (A'1), it is possible to define a PES in which mutational jumps are scaled by the parameter ε , by replacing in its generator (2.16) $m(x_j, h)dh$ by $m(x_j, h)dh \circ H_\varepsilon^{-1}$ for all $j \in \{1, \dots, d\}$, where $H_\varepsilon(h) = \varepsilon h$. Under Assumptions (A) and (B), by Theorem 2.7, this “rescaled PES” ($\Lambda_t^\varepsilon, t \geq 0$) is well-defined for all time. If only Assumptions (A) and (C) are satisfied, Corollary 3.3 only guarantees that the rescaled PES exists until the first time of coexistence. In this case, we denote by ($\Lambda_t^{(2),\varepsilon}, t \geq 0$) the corresponding “rescaled killed PES”. Finally, we do a time scaling of order $1/\varepsilon^2$ to obtain the process

$$\tilde{\Lambda}_t^\varepsilon = \begin{cases} \Lambda_{t/\varepsilon^2}^\varepsilon & \text{if Assumptions (A) and (B) are satisfied} \\ \Lambda_{t/\varepsilon^2}^{(2),\varepsilon} & \text{if only Assumptions (A) and (C) are satisfied.} \end{cases}$$

Since both Λ_t^ε and $\Lambda_t^{(2),\varepsilon}$ agree as long as there is no triple coexistence, and since we will only be interested in the sequel to the cases when the PES is monomorphic or dimorphic, we will not need to distinguish between these two cases.

4.1 The PES and the Canonical Equation of Adaptive Dynamics

Doing a similar time scaling as for $\tilde{\Lambda}^\varepsilon$, we can define for all $\varepsilon \in (0, 1]$, the ε -rescaled TSS ($X_t^\varepsilon, t \geq 0$) by modifying the generator (3.3) as follows: for any bounded measurable function φ ,

$$L^\varepsilon \varphi(x) := \frac{1}{\varepsilon^2} \int_{\mathbb{R}^l} (\varphi(x + \varepsilon h) - \varphi(x)) [g(x + \varepsilon h; x)]_+ m(x, h) dh, \quad (4.1)$$

where

$$g(y; x) = p(x) \lambda(x) \bar{n}(x) \frac{f(y; x)}{\lambda(y)}, \quad \forall x, y \in \mathcal{X}.$$

Formally speaking, the multiplicative term ε^{-2} takes into account that the integral term is of order ε^2 , since $g(x; x) = f(x; x) = 0$.

A key fact is that, as a consequence of Corollary 3.3, the process X^ε can be coupled with the process $\tilde{\Lambda}^\varepsilon$ in such a way that

$$\tilde{\Lambda}_t^\varepsilon = \bar{n}(X_t^\varepsilon)\delta_{X_t^\varepsilon}, \quad \forall t < \tau^\varepsilon, \quad (4.2)$$

where τ^ε is the first coexistence time:

$$\begin{aligned} \tau^\varepsilon &:= \inf\{t \geq 0 : \text{Supp}(\tilde{\Lambda}_t^\varepsilon) = 2\} \\ &= \inf\{t \geq 0 : f(X_t^\varepsilon; X_{t-}^\varepsilon) > 0 \text{ and } f(X_{t-}^\varepsilon; X_t^\varepsilon) > 0\}, \end{aligned} \quad (4.3)$$

by Proposition 2.4. We say that *coexistence occurs* in the PES $\tilde{\Lambda}$ if $\tau^\varepsilon < +\infty$.

Let us now state the convergence theorem of the rescaled TSS to the canonical equation of adaptive dynamics. Its proof, based on a standard uniqueness-compactness method, is given in Appendix C.

Theorem 4.1 *Assume (A) and (A'). Suppose also that the family of initial states $\{X_0^\varepsilon\}_{0 < \varepsilon \leq \bar{\varepsilon}}$ is bounded in \mathbb{L}^2 and converges in law to a random variable X_0 as $\varepsilon \rightarrow 0$.*

Then for each $T > 0$, the sequence $(X^\varepsilon)_\varepsilon$ converges when $\varepsilon \rightarrow 0$, for the Skorohod topology of $\mathbb{D}([0, T], \mathcal{X})$, to the process $(x(t), t \leq T)$ with initial state X_0 and with deterministic sample paths, unique solution of the ordinary differential equation, known as canonical equation of adaptive dynamics:

$$\frac{dx(t)}{dt} = \int_{\mathbb{R}^l} h[h \cdot \nabla_1 g(x(t); x(t))]_+ m(x(t), h) dh, \quad (4.4)$$

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

Remark 4.2 *In the case when $m(y, \cdot)$ is a symmetrical measure on \mathbb{R}^l for all $y \in \mathcal{X}$, Equation (4.4) gets the classical form, heuristically introduced in [10],*

$$\frac{dx(t)}{dt} = \frac{1}{2} K(x(t)) \nabla_1 g(x(t); x(t)), \quad (4.5)$$

where $K(x)$ is the covariance matrix of $m(x, h)dh$.

Let us now introduce some definitions and notation. In dimension $l = 1$, the canonical equation (4.4) reads

$$\frac{dx(t)}{dt} = \int_{\mathbb{R}} h[h\partial_1 g(x(t); x(t))]_+ m(x(t), h) dh. \quad (4.6)$$

The equilibria of this equation are given by the points x^* such that either $\partial_1 g(x^*; x^*) = 0$, or $\int_{\mathbb{R}_+} m(x^*, h) dh = 0$ and $\partial_1 g(x^*; x^*) > 0$, or $\int_{\mathbb{R}_-} m(x^*, h) dh = 0$ and $\partial_1 g(x^*; x^*) < 0$ (because of the positive part involved in the integral). We will concentrate on the points such that $\partial_1 g(x^*; x^*) = 0$, or equivalently, $\partial_1 f(x^*; x^*) = 0$, since

$$\partial_1 g(x; x) = p(x)\bar{n}(x)\partial_1 f(x; x).$$

Remark that, since $f(x; x) = 0$ for all $x \in \mathcal{X}$,

$$\partial_1 f(x; x) + \partial_2 f(x; x) = 0, \quad \forall x \in \mathcal{X} \quad (4.7)$$

$$\partial_{11} f(x; x) + 2\partial_{12} f(x; x) + \partial_{22} f(x; x) = 0, \quad \forall x \in \mathcal{X}. \quad (4.8)$$

Therefore, $\partial_1 f(x^*; x^*) = \partial_2 f(x^*; x^*) = 0$.

Definition 4.3 *The points $x^* \in \mathcal{X}$ such that $\partial_1 g(x^*; x^*) = 0$, or equivalently, $\partial_1 f(x^*; x^*) = \partial_2 f(x^*; x^*) = 0$ are called evolutionary singularities (ES).*

Note that under Assumption (A''), all equilibria of (4.6) are evolutionary singularities, except possibly the points of the boundary $\partial\mathcal{X}$ of \mathcal{X} .

Recall the definition (4.3) of the first coexistence time τ^ε . For any $\eta > 0$, we define the entrance time of the process in a η -neighborhood of an ES x^* ,

$$\theta_\eta^\varepsilon(x^*) = \inf\{t \geq 0, \text{Supp}(\tilde{\Lambda}_t^\varepsilon) \cap (x^* - \eta, x^* + \eta) \neq \emptyset\}. \quad (4.9)$$

Theorem 4.4 *Assume that (A), (A'), (A'') and either (B) or (C) hold. Assume also that $\tilde{\Lambda}_0^\varepsilon = \bar{n}(x_0)\delta_{x_0}$ where x_0 is not an ES. Let $x(t)$ be the solution to (4.4) such that $x(0) = x_0$ and let $x^* := \lim_{t \rightarrow +\infty} x(t)$. This limit exists and is finite as $l = 1$ and \mathcal{X} is compact, and x^* is an equilibrium of (4.4). Assume that x^* is an ES. Then,*

(i) *For any $T > 0$,*

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P}(\tau^\varepsilon > T) = 1.$$

(ii) *For any $T > 0$, the process $\tilde{\Lambda}_t^\varepsilon$ converges as $\varepsilon \rightarrow 0$ to the (deterministic) process $\bar{n}(x(t))\delta_{x(t)}$ for the Skorohod topology on $\mathbb{D}([0, T], \mathcal{M}_F)$, where \mathcal{M}_F is equipped with the weak topology.*

(iii) *For any $\eta > 0$, there exists $\varepsilon_0 > 0$ such that, for all $\varepsilon < \varepsilon_0$,*

$$\mathbb{P}(\theta_\eta^\varepsilon(x^*) \leq \tau^\varepsilon) = 1,$$

and the process $X_{t \wedge \theta_\eta^\varepsilon(x^)}^\varepsilon$ is a.s. monotone, where X_t^ε is defined by (4.2) for all $t < \tau^\varepsilon$.*

The proof of this result is based on Theorem 4.1, the coupling (4.2) and the following lemma.

Lemma 4.5 *Assume (A), (A') and (A'').*

- (1) *The solution $x(t)$ of (4.4) starting from a point that is not an equilibrium cannot attain an ES in finite time.*
- (2) *Fix $\eta > 0$ and any connected subset S of $\{y \in \mathcal{X} : |\partial_1 f(y; y)| > \eta\}$. Then, for all $y, z \in S$ close enough, $f(y; z)f(z; y) < 0$ and $(z - y)f(z; y)$ has constant sign.*
- (3) *Assume that $x_0 = x(0)$ is not an ES and let $I_T = \{x(t), t \in [0, T]\}$. Then, for any sufficiently small $\eta' > 0$, for any y at a distance to I_T smaller than η' and for any z sufficiently close to y , $f(y; z)f(z; y) < 0$ and $(z - y)f(z; y)$ has constant sign.*

Proof (1) Let C be a constant such that $x \mapsto \int_{\mathbb{R}} h[h \cdot \partial_1 g(x; x)]_+ m(x, h) dh$ is C -Lipschitz. Then, for any ES x^* ,

$$\left| \frac{d}{dt} (x(t) - x^*)^2 \right| \leq 2 |\dot{x}(t)| |x(t) - x^*| \leq 2C (x(t) - x^*)^2.$$

Thus, $|x(t) - x^*| \geq |x(0) - x^*| \exp(-Ct) > 0$.

(2) Since $\partial_1 f(y; y)$ is continuous, $\partial_1 f(y; y)$ has constant sign for all $y \in S$. So fix $y \leq z$ in S and assume that $\partial_1 f(y; y) > \eta$ and $\partial_1 f(y; y) > \eta$. Since $f(y; y) = 0$, a second-order Taylor-Lagrange expansion of $f(z; y)$ at (y, y) implies that $f(z; y) \geq \eta(z - y)/2$ provided that $0 \leq z - y \leq \eta/C'$, where $C' > 0$ is a constant uniformly bounding the second-order derivatives of $f(\cdot; \cdot)$ on the compact set \mathcal{X}^2 . Similarly, $f(y; z) \leq \eta(y - z)/2$, and one obtains reversed inequalities when $\partial_1 f(y; y) < -\eta$ and $\partial_1 f(y; y) < -\eta$. In any case, $f(y; z)f(z; y) < 0$ if $|z - y| \leq \eta/C'$ and $(z - y)f(z; y)$ has constant sign.

(3) Remark first that, from Point (1), $\eta := \inf_{y \in I_T} |\partial_1 f(y; y)| > 0$. Therefore, for $\eta' > 0$ sufficiently small, $\{y \in \mathcal{X} : \text{dist}(y, I_T) \leq \eta'\} \subset \{y \in \mathcal{X} : |\partial_1 f(y; y)| > \eta/2\}$. The result then follows from Point (2). \square

Proof of Theorem 4.4 (i) Recall from (4.2) that, before the stopping time τ^ε , the support of $\tilde{\Lambda}_t^\varepsilon$ is a singleton whose dynamics is that of the rescaled TSS $(X_t^\varepsilon, t \geq 0)$. Because of Theorem 4.1, when $\varepsilon \rightarrow 0$, the TSS is close to the canonical equation. In particular, for all $\eta' > 0$, its values on the time interval $[0, T]$ belong to the set $\{x \in \mathcal{X} : \text{dist}(x, I_T) \leq \eta'\}$ with probability converging to 1. Moreover, since \mathcal{X} is compact, $\text{Supp}(m(x, \cdot)) \subset \{x - y; x, y \in \mathcal{X}\}$ has diameter bounded by $2\text{Diam}(\mathcal{X})$. Therefore, the distance between a mutant trait and the trait of its progenitor in the rescaled PES $\tilde{\Lambda}^\varepsilon$ is a.s. less than $2\varepsilon\text{Diam}(\mathcal{X})$. Hence, because of Lemma 4.5 (3), for η' and ε small enough, on the event $\{\sup_{0 \leq t \leq T} \|X_t^\varepsilon - x(t)\| < \eta'\}$, no mutant in the PES can coexist with the resident trait before time T and thus $\tau^\varepsilon > T$.

(ii) This is an immediate consequence of (4.2), Theorem 4.1 and Point (i).

(iii) Fix $\eta > 0$. Let S be the connected component of $\mathcal{X} \setminus \bigcup_{y^* \text{ is an ES}} (y^* - \eta, y^* + \eta)$ containing x_0 and let θ_S^ε denote the first exit time of the support of the PES $\tilde{\Lambda}^\varepsilon$ from S . Similarly as in Point (i), we deduce from Lemma 4.5 (2) that $\theta_S^\varepsilon \leq \tau^\varepsilon$ a.s. if η and ε are small enough. Moreover, the generator (4.1) of the TSS X^ε involves the *positive part* of $f(x + \varepsilon h; x)$, where h is drawn according to $m(x, h)dh$. Therefore, taking ε small enough and using Lemma 4.5 (2) again, a jump in the TSS before time θ_S^ε is only possible towards larger trait values if $(z - y)f(z; y) > 0$ in Lemma 4.5 (2), or towards smaller trait values conversely. Thus, the TSS process is a.s. monotone before time θ_S^ε . By definition of the point x^* , we then see that the exit of the TSS from the interval S can only occur in the direction of x^* , i.e. $X_{\theta_S^\varepsilon}^\varepsilon \in (x^* - \eta, x^* + \eta)$. Hence $\theta_S^\varepsilon = \theta_\eta^\varepsilon(x^*)$ a.s., and the proof of Theorem 4.4 is completed. \square

Remark 4.6 *Theorem 4.4 allows one to describe the phase of convergence before evolutionary branching in Fig. 2.1(b) and 2.2(b). This result also explains why the time of convergence is approximately multiplied by 4 when the mutation size σ is divided by 2, in Fig. 2.2(b) and 2.3(b).*

Theorem 4.4 implies that, when the initial population is monomorphic and away from evolutionary singularities, evolutionary branching can only occur in the neighborhood of an evolutionary singularity, and after time T/ε^2 with probability converging to 1 when $\varepsilon \rightarrow 0$, for all $T > 0$.

The next result shows that we can restrict to the ES that are not repulsive for the canonical equation.

Proposition 4.7 *Under the assumptions of Theorem 4.4, coexistence of two traits in the PES $\tilde{\Lambda}^\varepsilon$ can only occur in the neighborhood of evolutionary singularities $x^* \in \mathcal{X}$ which satisfy*

$$\partial_{22}f(x^*; x^*) \geq \partial_{11}f(x^*; x^*). \quad (4.10)$$

More precisely, for any neighborhood \mathcal{U} of the set of evolutionary singularities satisfying (4.10), for all ε small enough,

$$\mathbb{P}(\tau^\varepsilon < +\infty \quad \text{and} \quad \text{Supp}(\Lambda_{\tau^\varepsilon}^\varepsilon) \notin \mathcal{U}) = 0.$$

Proof Let us remark that an ES such that

$$\partial_{11}f(x^*; x^*) + \partial_{12}f(x^*; x^*) > 0. \quad (4.11)$$

is always a repulsive point for the canonical equation, in the sense that, for any solution $x(t)$ of the canonical equation starting sufficiently close from x^* , the distance between $x(t)$ and x^* is non-decreasing in the neighborhood of time 0. In other words, there exists a neighborhood \mathcal{U} of x^* such that no solution of the canonical equation (4.6) starting out of \mathcal{U} can enter \mathcal{U} . To see this, it suffices to observe that (4.11) implies that there exists η_{x^*} with

- $\partial_1 g(x; x) > 0$ if $x \in (x^*, x^* + \eta_{x^*}]$,
- $\partial_1 g(x; x) < 0$ if $x \in [x^* - \eta_{x^*}, x^*)$.

Observe that, by (4.8), (4.11) is equivalent to $\partial_{11}f(x^*; x^*) - \partial_{22}f(x^*; x^*) > 0$.

Let S be the set of repulsive ES and define $\mathcal{V} = \cup_{x^* \in S} (x^* - \eta_{x^*}, x^* - \eta_{x^*})$. Fix \mathcal{U} as in the statement of Proposition 4.7 and assume (without loss of generality) that $\mathcal{U} \cap \mathcal{V} = \emptyset$ and $x \notin \mathcal{U} \cup \mathcal{V}$. Let $[a, b]$ be any connected component of $\mathcal{X} \setminus (\mathcal{U} \cup \mathcal{V})$. Since $\partial_1 f(y, y) \neq 0$ for all $y \in [a, b]$, Lemma 4.5 (2) shows that coexistence never happens in a monomorphic population with trait in $\mathcal{X} \setminus (\mathcal{U} \cup \mathcal{V})$ if ε is sufficiently small. Similarly, for ε sufficiently small, no mutant in \mathcal{V} born from a monomorphic population with trait not belonging to \mathcal{V} has a positive fitness. Therefore, the TSS cannot drive the population inside \mathcal{V} starting from outside. Thus, if coexistence occurs in finite time, it must occur in \mathcal{U} . This completes the proof of Proposition 4.7. \square

4.2 Evolutionary branching criterion

In this section we prove a criterion of evolutionary branching.

4.2.1 Definition and main result

We first need to precisely define what we mean by evolutionary branching.

Definition 4.8 *Fix $\varepsilon > 0$ and x^* an ES. For all $\eta > 0$, we say that there is η -branching at x^* for the PES $\tilde{\Lambda}^\varepsilon$ if*

- *there exists $t_1 > 0$ such that the support of the PES at time t_1 is composed of a single point belonging to $[x^* - \eta, x^* + \eta]$ (i.e. $\theta_\eta^\varepsilon(x^*) < \infty$)*

- there exists $t_2 > t_1$ such that the support of the PES at time t_2 is composed of exactly 2 points separated by a distance of more than $\eta/2$
- between t_1 and t_2 , the support of the PES is always a subset of $[x^* - \eta, x^* + \eta]$, and is always composed of at most 2 traits, and has nondecreasing (in time) diameter.

We only consider *binary* evolutionary branching. We will actually prove in Corollary 4.14 that the subdivision of a single branch into three branches (or more) in the neighborhood of an evolutionary singularity is a.s. impossible under the assumptions of the following theorem. Note that the notion of evolutionary branching requires the coexistence of two traits (i.e. $\tau^\varepsilon < +\infty$), but also that these two traits diverge from one another.

Our main result is the following.

Theorem 4.9 *Assume (A), (A'), (A'') and either (B) or (C). Assume also that $\tilde{\Lambda}_0^\varepsilon = \bar{n}(x)\delta_x$ and that the canonical equation with initial condition x converges to an ES x^* in the interior of \mathcal{X} such that*

$$\partial_{22}f(x^*; x^*) > \partial_{11}f(x^*; x^*) \quad (4.12)$$

$$\text{and } \partial_{22}f(x^*; x^*) + \partial_{11}f(x^*; x^*) \neq 0. \quad (4.13)$$

Then, for all sufficiently small η , there exists $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$,

- (a) if $\partial_{11}f(x^*; x^*) > 0$, $\mathbb{P}(\eta\text{-branching at } x^* \text{ for } \tilde{\Lambda}^\varepsilon) = 1$.
- (b) if $\partial_{11}f(x^*; x^*) < 0$, $\mathbb{P}(\eta\text{-branching at } x^* \text{ for } \tilde{\Lambda}^\varepsilon) = 0$. Moreover,

$$\mathbb{P}(\forall t \geq \theta_\eta^\varepsilon(x^*), \text{Card}(\text{Supp}(\tilde{\Lambda}_t^\varepsilon)) \leq 2 \text{ and } \text{Supp}(\tilde{\Lambda}_t^\varepsilon) \subset (x^* - \eta, x^* + \eta)) = 1,$$

where $\theta_\eta^\varepsilon(x^*)$ has been defined in (4.9).

This criterion appeared for the first time in [28, Section 3.2.5] with a heuristic justification. We see that, locally around x^* , one of the two following events can occur almost surely: either there is binary evolutionary branching and the two branches diverge monotonously, or there is no evolutionary branching, and the population stays forever inside any neighborhood of x^* . Coexistence can occur in this case, but cannot drive the support of the population away from a small neighborhood of x^* . We will actually prove that, in this case, as soon as there is coexistence of two traits in the population, the diameter of the support of the PES cannot increase before it reaches 0 (i.e. until the next time when the population becomes monomorphic).

The proof of this result is given in the following subsections. In Section 4.2.3, we prove regularity results on the 2- and 3-dimensional fitness functions and give their second order expansions in the neighborhood of evolutionary singularities. A first corollary of this result is given in Section 4.2.4 where, using the results of M.-L. Zeeman [33] and Fig. 3.1, we show that no triple coexistence can occur in the neighborhood of evolutionary singularities. Finally, a case by case study of the zone of coexistence and of the signs of fitness functions in the neighborhood of an evolutionary singularity allows us to conclude the proof in Section 4.2.5.

Before coming to the proof and in order to illustrate the difference between coexistence and evolutionary branching, we state a result that will be needed in the course of the proof of Theorem 4.9.

Proposition 4.10 *Assume (A2) and that λ , μ and α are \mathcal{C}^2 . Let $x^* \in \mathcal{X}$ be any ES.*

- (a) *If $\partial_{11}f(x^*; x^*) + \partial_{22}f(x^*; x^*) > 0$, then for all neighborhood \mathcal{U} of x^* , there exist $x, y \in \mathcal{U}$ that coexist.*
- (b) *If $\partial_{11}f(x^*; x^*) + \partial_{22}f(x^*; x^*) < 0$, then there exists a neighborhood \mathcal{U} of x^* such that any $x, y \in \mathcal{U}$ do not coexist.*

This shows that the criterion of evolutionary branching ($\partial_{11}f(x^*; x^*) > 0$) is different from the criterion of coexistence ($\partial_{11}f(x^*; x^*) + \partial_{22}f(x^*; x^*) > 0$). If one assumes as in Theorem 4.9 that $\partial_{22}f(x^*; x^*) > \partial_{11}f(x^*; x^*)$, the evolutionary branching condition $\partial_{11}f(x^*; x^*) > 0$ implies the coexistence criterion $\partial_{11}f(x^*; x^*) + \partial_{22}f(x^*; x^*) > 0$, as expected. The proof of this proposition is deferred until Subsection 4.2.5.

4.2.2 Example

Let us come back to the example introduced in Subsection 2.2. The parameters of the model satisfy Assumptions (A), (A'), (A'') and (C) (see Subsection 3.3).

The fitness function is

$$\begin{aligned} f(y; x) &= \lambda(y) - \alpha(y, x)\bar{n}(x) \\ &= \exp\left(-\frac{y^2}{2\sigma_b^2}\right) - \exp\left(-\frac{(x-y)^2}{2\sigma_\alpha^2}\right) \exp\left(-\frac{x^2}{2\sigma_b^2}\right). \end{aligned}$$

Computation gives

$$\partial_1 f(x^*; x^*) = -\frac{x^*}{\sigma_b^2} \exp\left(-\frac{(x^*)^2}{2\sigma_b^2}\right) = 0 \iff x^* = 0.$$

Moreover, $\partial_{11}f(0; 0) = \frac{1}{\sigma_\alpha^2} - \frac{1}{\sigma_b^2}$ and $\partial_{22}f(0; 0) = \frac{1}{\sigma_\alpha^2} + \frac{1}{\sigma_b^2}$. Thus, the coexistence criterion of Proposition 4.10 (a) is always satisfied. We furthermore observe that (4.12) and (4.13) hold, and that

$$\partial_{11}f(0; 0) > 0 \iff \sigma_\alpha < \sigma_b.$$

Then if $\sigma_\alpha < \sigma_b$, we have almost surely branching and if $\sigma_\alpha > \sigma_b$, we have only coexistence. This is consistent with Fig. 2.2 (a) and (b).

4.2.3 Trait smoothness of fitnesses around evolutionary singularities

The problem of local expansion of fitness functions has been already studied in [12] for general models. In this section, we establish regularity and expansion results on our 2- and 3-dimensional fitness functions in the neighborhood of evolutionary singularities. To this aim, we need the following lemma.

Lemma 4.11 *Let $h(x, y, z)$ be a \mathcal{C}^k function for $k \geq 1$ defined on \mathcal{X}^3 such that $h(x, x, z) = 0$ for all $x, z \in \mathcal{X}$. Then, the function*

$$(x, y, z) \mapsto \frac{h(x, y, z)}{x - y}$$

can be extended on $\{x = y\}$ as a \mathcal{C}^{k-1} function $\hat{h}(x, y, z)$ on \mathcal{X}^3 by setting $\hat{h}(x, x, z) = \partial_1 h(x, x, z)$ for all $x, z \in \mathcal{X}$.

Proof Taylor's formula with integral remainder yields

$$\frac{h(x, y, z)}{x - y} = \int_0^1 \partial_1 h(y + (x - y)u, y, z) du$$

for all $x \neq y$. The right-hand side also has a sense for $x = y$ and defines a \mathcal{C}^{k-1} function on \mathcal{X}^3 . \square

Let $x^* \in \mathcal{X}$ be an ES as in the statement of Theorem 4.9. By Assumptions (A) and (A''), the 2-dimensional fitness function $f(y; x)$ defined in (2.12) is well-defined for all $x, y \in \mathcal{X}$ and is a \mathcal{C}^3 function. We can extend the definition of the 3-dimensional fitness function

$$f(z; x, y) = r(z) - \alpha(z, x)\bar{n}_1(x, y) - \alpha(z, y)\bar{n}_2(x, y),$$

where $\bar{n}_i(x, y)$, $i = 1, 2$, are defined in (2.10) and (2.11) to all $x, y \in \mathcal{X}$ such that

$$\alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x) \neq 0,$$

and the following result holds for this extended 3-dimensional fitness function. We will also use the notation

$$a = \partial_{11}f(x^*; x^*) \quad \text{and} \quad c = \partial_{22}f(x^*; x^*). \quad (4.14)$$

Note that, by (4.8),

$$\partial_{12}f(x^*; x^*) = -\frac{a + c}{2}. \quad (4.15)$$

Proposition 4.12 *Under the assumptions (A2), (A'3) and (4.13), the following properties hold.*

(i) *For all $x, y \in \mathcal{X}$ in a neighborhood of x^* ,*

$$x \neq y \implies \alpha(x, x)\alpha(y, y) \neq \alpha(x, y)\alpha(y, x).$$

This implies in particular that $\bar{\mathbf{n}}(x, y)$ (defined in Definition 2.2) and $f(\cdot; x, y)$ are well-defined for such x, y .

(ii) *When $x, y \rightarrow x^*$ in such a way that $x \neq y$,*

$$\bar{n}_1(x, y) + \bar{n}_2(x, y) \longrightarrow \bar{n}(x^*) = \frac{r(x^*)}{\alpha(x^*, x^*)}; \quad (4.16)$$

$$f(z; x, y) \longrightarrow f(z; x^*), \quad \forall z \in \mathcal{X}. \quad (4.17)$$

(iii) *With the notation (4.14), as $x, y \rightarrow x^*$,*

$$f(y; x) = \frac{1}{2}(x - y)(c(x - x^*) - a(y - x^*)) + o(|x - y|(|x - x^*| + |y - x^*|)). \quad (4.18)$$

(iv) *The function $f(z; x, y)$ can be extended as a \mathcal{C}^2 function on $\{(x, y, z) : z \in \mathcal{X}, x, y \in \mathcal{U}\}$ for some neighborhood \mathcal{U} of x^* in \mathcal{X} . Still denoting by $f(z; x, y)$ the extended function, as $x, y \rightarrow x^*$,*

$$f(z; x, y) = \frac{a}{2}(z - x)(z - y) + o(|z - x||z - y|). \quad (4.19)$$

Proof Let $D(x, y) := \alpha(x, x)\alpha(y, y) - \alpha(x, y)\alpha(y, x)$. It follows from Lemma 4.11 that $D(x, y)/(x - y)$ can be extended on \mathcal{X}^2 as a \mathcal{C}^3 function, which has value

$$(\partial_1\alpha(x, x))\alpha(x, x) + (\partial_2\alpha(x, x))\alpha(x, x) - (\partial_1\alpha(x, x))\alpha(x, x) - \alpha(x, x)(\partial_2\alpha(x, x)) = 0$$

at the point (x, x) . Therefore, Lemma 4.11 can be applied once more to prove that $D(x, y)/(x - y)^2$ can be extended as a \mathcal{C}^2 function $\hat{D}(x, y)$ on \mathcal{X}^2 . Hence, an elementary computation involving the second-order Taylor expansion of $D(x, y)$ yields that

$$D(x, y) = (x - y)^2(\alpha(x^*, x^*)\partial_{12}\alpha(x^*, x^*) - \partial_1\alpha(x^*, x^*)\partial_2\alpha(x^*, x^*)) + o(|x - y|^2).$$

Thus, Point (i) follows from the fact that $\alpha(x^*, x^*)\partial_{12}\alpha(x^*, x^*) \neq \partial_1\alpha(x^*, x^*)\partial_2\alpha(x^*, x^*)$, which is a consequence of (4.13). Indeed, one can check that

$$a = r''(x^*) - r(x^*)\frac{\partial_{11}\alpha(x^*, x^*)}{\alpha(x^*, x^*)}$$

$$\text{and } c = -r''(x^*) + 2r'(x^*)\frac{\partial_1\alpha(x^*, x^*)}{\alpha(x^*, x^*)} + r(x^*)\frac{\alpha(x^*, x^*)(\partial_{11}\alpha(x^*, x^*) + 2\partial_{12}\alpha(x^*, x^*)) - 2\partial_1\alpha(x^*, x^*)(\partial_1\alpha(x^*, x^*) + \partial_2\alpha(x^*, x^*))}{\alpha(x^*, x^*)^2}.$$

Using the fact that

$$r'(x^*) = r(x^*)\frac{\partial_1\alpha(x^*, x^*)}{\alpha(x^*, x^*)} \quad (4.20)$$

since x^* is an ES, we have that

$$\alpha^2(x^*, x^*)(a + c) = 2r(x^*)(\alpha(x^*, x^*)\partial_{12}\alpha(x^*, x^*) - \partial_1\alpha(x^*, x^*)\partial_2\alpha(x^*, x^*)).$$

Hence,

$$\alpha(x^*, x^*)\partial_{12}\alpha(x^*, x^*) - \partial_1\alpha(x^*, x^*)\partial_2\alpha(x^*, x^*) \neq 0 \iff a + c \neq 0.$$

In particular, this implies that the function $\hat{D}(x, y)$ is non-zero in a neighborhood of x^* . For Point (ii), observe that

$$\bar{n}_1(x, y) + \bar{n}_2(x, y) = \frac{r(x)\frac{\alpha(y, y) - \alpha(y, x)}{x - y} + r(y)\frac{\alpha(x, x) - \alpha(x, y)}{x - y}}{(x - y)\hat{D}(x, y)}.$$

By the proof of Lemma 4.11, the numerator can be extended as a \mathcal{C}^3 function $h(x, y)$ by setting

$$h(x, y) = -r(x)\int_0^1 \partial_2\alpha(y, y + (x - y)u)du + r(y)\int_0^1 \partial_2\alpha(x, y + (x - y)u)du$$

for all $x, y \in \mathcal{X}$. In particular, $h(x, x) = 0$ for all $x \in \mathcal{X}$. Therefore, Lemma 4.11 can be applied once more to prove that $\bar{n}_1(x, y) + \bar{n}_2(x, y)$ can be extended as a \mathcal{C}^2 function in the neighborhood of x^* and that

$$\lim_{x, y \rightarrow x^*, x \neq y} \bar{n}_1(x, y) + \bar{n}_2(x, y) = \frac{\frac{\partial h}{\partial x}(x^*, x^*)}{\hat{D}(x^*, x^*)} = \frac{r(x^*)\partial_{12}\alpha(x^*, x^*) - r'(x^*)\partial_2\alpha(x^*, x^*)}{\alpha(x^*, x^*)\partial_{12}\alpha(x^*, x^*) - \partial_1\alpha(x^*, x^*)\partial_2\alpha(x^*, x^*)}.$$

Hence, (4.16) and then (4.17) follow from (4.20).

Point (iii) is obtained from the fact that $f(x; x) = 0$, from Lemma 4.11 and from the second-order Taylor expansion of $f(y; x)$. In this computation, one must use the fact that x^* is an ES and (4.15).

The fact that $f(z; x, y)$ is \mathcal{C}^2 in $\mathcal{U} \times \mathcal{U} \times \mathcal{X}$ can be proven exactly as the regularity of $\bar{n}_1(x, y) + \bar{n}_2(x, y)$ above, observing that

$$f(z; x, y) = r(z) - \frac{r(x) \frac{\alpha(z, x)\alpha(y, y) - \alpha(z, y)\alpha(y, x)}{x-y} + r(y) \frac{\alpha(z, y)\alpha(x, x) - \alpha(z, x)\alpha(x, y)}{x-y}}{(x-y)\hat{D}(x, y)}.$$

Therefore, using the fact that $f(x; x, y) = f(y; x, y) = 0$, Lemma 4.11 can be applied twice to prove that

$$f(z; x, y) = \frac{\gamma}{2}(z-x)(z-y) + o(|z-x||z-y|)$$

for some constant $\gamma \in \mathbb{R}$. The second-order Taylor expansion of $f(z; x, y)$ shows that $\gamma = \partial_{11}f(x^*; x^*, x^*)$. Now, because of (4.17), $\partial_{11}f(z; x^*, x^*) = \partial_{11}f(z; x^*)$ for all $z \in \mathcal{X}$. Hence $\gamma = a$, which ends the proof of Point (iv). \square

Remark 4.13 *Let us remark that, if x^* is not an evolutionary singularity, Point (ii) of Proposition 4.12 need not to be true anymore, which may be surprising for the intuition and which has been a source of errors in some biological works.*

Moreover, if x^ is an ES but Assumption (4.13) ($a + c \neq 0$) is not true, Point (ii) of Proposition 4.12 may also fail. Indeed, in the case when $\alpha(x, x)\partial_{12}\alpha(x, x) \neq \partial_1\alpha(x, x)\partial_2\alpha(x, x)$ for $x \neq x^*$,*

$$\begin{aligned} \bar{n}_1(x, x) + \bar{n}_2(x, x) &= \frac{r(x)\partial_{12}\alpha(x, x) - r'(x)\partial_2\alpha(x, x)}{\alpha(x, x)\partial_{12}\alpha(x, x) - \partial_1\alpha(x, x)\partial_2\alpha(x, x)} \\ &= \frac{r(x^*)(\partial_{112}\alpha(x^*, x^*) + \partial_{122}\alpha(x^*, x^*)) - r'(x^*)\partial_{22}\alpha(x^*, x^*) - r''(x^*)\partial_2\alpha(x^*, x^*) + o(1)}{\alpha(x^*, x^*)(\partial_{112}\alpha(x^*, x^*) + \partial_{122}\alpha(x^*, x^*)) - \partial_2\alpha(x^*, x^*)\partial_{11}\alpha(x^*, x^*) - \partial_1\alpha(x^*, x^*)\partial_{22}\alpha(x^*, x^*) + o(1)} \end{aligned}$$

as $x \rightarrow x^*$. This expression involves $r''(x^*)$, whose value is not imposed by the assumptions. Therefore, changing the function r in such a way that $r(x^*)$ and $r'(x^*)$ are fixed but $r''(x^*)$ changes also changes the value of $\lim_{x, y \rightarrow x^*} \bar{n}_1(x, y) + \bar{n}_2(x, y)$.

4.2.4 On triple coexistence in the neighborhood of x^*

Points (iii) and (iv) of Proposition 4.12 allow one to determine the signs of the 2- and 3-dimensional fitnesses in a trimorphic population with traits x, y, z close to x^* . Combining this with the classification of Zeeman [33] (see Section 3.2 and Figure 3.1) gives the following corollary.

Corollary 4.14 *Assume (A''). For any ES x^* satisfying (4.12), (4.13) and such that $\partial_{11}f(x^*; x^*) \neq 0$, there exists a neighborhood \mathcal{U} of x^* such that, for all distinct $x, y, z \in \mathcal{U}$, $(x, y, z) \notin C_{\text{coex}}$, where C_{coex} is defined in (3.4).*

Proof Let us assume for simplicity that $x^* = 0$. We shall distinguish between the cases $a > 0$ and $a < 0$, and prove in each case that the fitnesses cannot have any of the sign configuration corresponding to the classes 26, 29, 31 and 33 in the neighborhood of x^* . Since all these classes contain the pattern of Fig. B.1, we have (possibly after relabelling the traits x, y, z) that $f(x; y) \geq 0$, $f(y; x) \geq 0$, $f(z; x, y) \geq 0$ and $x < y$. Consider first the case $a < 0$. It follows from Proposition 4.12 (iv) that the function $f(\cdot; \cdot, \cdot)$ has the shape of Fig. 4.1 (a) in the neighborhood of x^* . In particular, this implies that $x < z < y$, $f(z; x, y) > 0$, $f(x; y, z) < 0$ and $f(y; x, z) < 0$ as soon as x, y, z are sufficiently close to x^* . In view of Fig. 3.1, these conditions are incompatible with classes 31 and 33. Moreover, $\partial_{11}f(x; y) < 0$ for all x, y sufficiently close to x^* . Therefore, by Lemma 4.11,

$$\frac{\partial}{\partial x} \left(\frac{f(x; y)}{y - x} \right) = - \int_0^1 u \partial_{11} f(y + u(x - y); y) du \quad (4.21)$$

is positive for all x, y sufficiently close to x^* . Hence, since $x < z < y$, we have $f(z; y)/(y - z) > f(x; y)/(y - x) \geq 0$ and thus $f(z; y) > 0$. Similarly, $f(z; x) > 0$. Together with $f(z; x, y) > 0$, these conditions are incompatible with classes 26 and 29. This ends the proof in the case where $a < 0$.

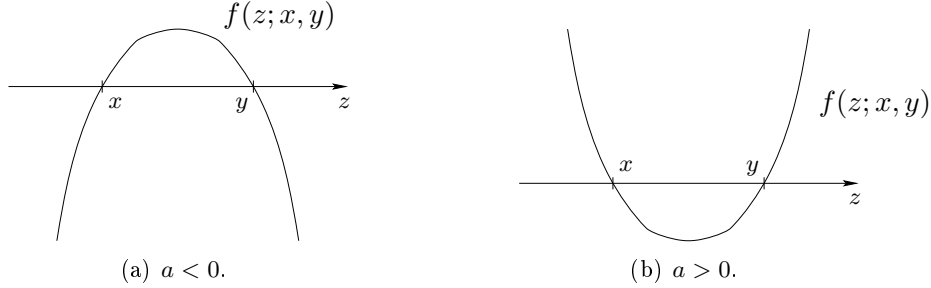


Figure 4.1: The shape of the 3-dimensional fitness as a function of the sign of a .

In the case when $a > 0$, by Proposition 4.12 (iv), $f(\cdot; \cdot, \cdot)$ has the shape of Fig. 4.1 (b) in the neighborhood of x^* . Therefore, $z \notin [x, y]$. Assume for example that $z < x < y$. By Proposition 4.12 (iv) again, $f(x; y, z) < 0$ and $f(y; x, z) > 0$. These conditions are incompatible with class 33. Moreover, using the fact that $\partial_{11}f(x; y) > 0$ for all x, y sufficiently close to x^* , it follows from the fact that (4.21) is negative that $f(z; y)/(y - z) > f(x; y)/(y - x) \geq 0$ and thus that $f(z; y) > 0$. Similarly, because of Assumption (4.12), $\partial_{22}f(x; y) > 0$ for all x, y sufficiently close to x^* . Therefore, by Lemma 4.11,

$$\frac{\partial}{\partial x} \left(\frac{f(y; x)}{y - x} \right) = - \int_0^1 u \partial_{22} f(y; y + u(x - y)) du < 0$$

for all x, y sufficiently close to x^* . Thus, $f(y; x) \geq 0$ implies that $f(y; z) > 0$. Together with the fact that $f(x; y, z) < 0$, these conditions are incompatible with classes 26, 29 and 31.

In the case when $x < y < z$, the method above proves that $f(x; z) > 0$, $f(z; x) > 0$ and $f(y; x, z) < 0$, which is again incompatible with classes 26, 29, 31 and 33. This ends the proof of Corollary 4.14. \square

4.2.5 The main proofs

We prove here Proposition 4.10, which gives a criterion for the coexistence of two traits in the neighborhood of x^* , and we end the proof of Theorem 4.9. The proof of Proposition 4.10 is based on the study of the region of double coexistence, defined as $\{(x, y) \in \mathcal{X} : f(x; y) > 0 \text{ and } f(y; x) > 0\}$ in the neighborhood of x^* . The proof of Theorem 4.9 is based on a case-by-case study that extends the proof of Corollary 4.14.

Proof of Proposition 4.10 It follows from Proposition 4.12 (iii) that the set of $(x, y) \in \mathcal{X}$ such that $f(y; x) = 0$ is composed of the line $\{y = x\}$ and of a C^2 -curve γ in the neighborhood of x^* , containing (x^*, x^*) and admitting as tangent at this point the line $\{a(y - x^*) = c(x - x^*)\}$. The curve γ is given by the Implicit Function Theorem applied to $f(y; x)/(y - x)$, which is a C^2 -function by Lemma 4.11. Since $a < c$, the curves γ and $\{y = x\}$ divide \mathcal{X}^2 in the neighborhood of (x^*, x^*) into 4 regions. Moreover, because of (4.18), $f(y; x)$ changes sign when the point (x, y) changes region by crossing either the line $\{y = x\}$ or the curve γ .

It is elementary from a case-by-case study to check that coexistence can occur in the neighborhood of x^* if $c > a > 0$, $a > c > 0$, $-a < c < 0 < a$ and $a < 0 < -a < c$, and that coexistence cannot occur in the neighborhood of x^* if $c < -a < 0 < a$, $c < a < 0$, $a < c < 0$ and $a < 0 < c < -a$. The cases when coexistence is possible are represented in Fig. 4.2. In these figures, the curve γ is represented by its tangent line $\{a(y - x^*) = c(x - x^*)\}$ and the sign of $f(y; x)$ is represented by $+$ and $-$ signs depending on the position of (x, y) with respect to γ and $\{y = x\}$. The sign of $f(x; y)$ is obtained by an axial symmetry of the figure with axis $\{y = x\}$. In Fig. 4.2, we denote by γ^s the mirror image of the curve γ with respect to this axis. The region of coexistence is the one where $f(y; x) > 0$ and $f(x; y) > 0$.

Note that the expansion of $f(y; x)$ done when proving Proposition 4.12 (iii) does not make use of any assumption on a and c . Therefore, a similar study can be done to treat the degenerate cases. One easily obtains that coexistence is possible in the neighborhood of (x^*, x^*) if $c = a > 0$, $c = 0$ and $a > 0$ or $a = 0$ and $c > 0$. Similarly, coexistence cannot occur in the neighborhood of (x^*, x^*) if $c = a < 0$, $c = 0$ and $a < 0$ or $a = 0$ and $c < 0$. The case $c = -a$ is undetermined and depends on higher-order expansions of the fitness function. \square

Proof of Theorem 4.9: $\theta_\eta^\varepsilon(x^*) < \infty$ a.s.

Let us first prove that, for all $\eta > 0$, for ε sufficiently small, $\theta_\eta^\varepsilon(x^*) < \infty$ a.s. This is a consequence of the fact that the TSS, which is equal to the support of the PES before time $\theta_\eta^\varepsilon(x^*)$, is monotone (Theorem 4.4 (iii)), and that, because of Assumption (A''), the TSS has always a positive probability to have ‘‘big’’ jumps.

More precisely, consider first the case where $x_0 < x^*$ and assume that $\mathbb{P}(\theta_\eta^\varepsilon(x^*) = \infty) > 0$. Then, a.s. on this event, $\tau^\varepsilon = \infty$ and the TSS process stays for all positive time in the interval $I := [x_0, x^* - \eta]$. As in the proof of Lemma 4.5 (2), since $\partial_1 g(x; x) > 0$ for all $x \in I$, by Assumption (A''), for all $x \in I$, $h \geq 0$ in the support of $m(x, h)dh$, and $\varepsilon \leq \varepsilon_0$ where $\varepsilon_0 > 0$ is small enough,

$$\int_0^{+\infty} [g(x + \varepsilon h; x)]_+ m(x, h) dh > 0, \quad \forall x \in I.$$

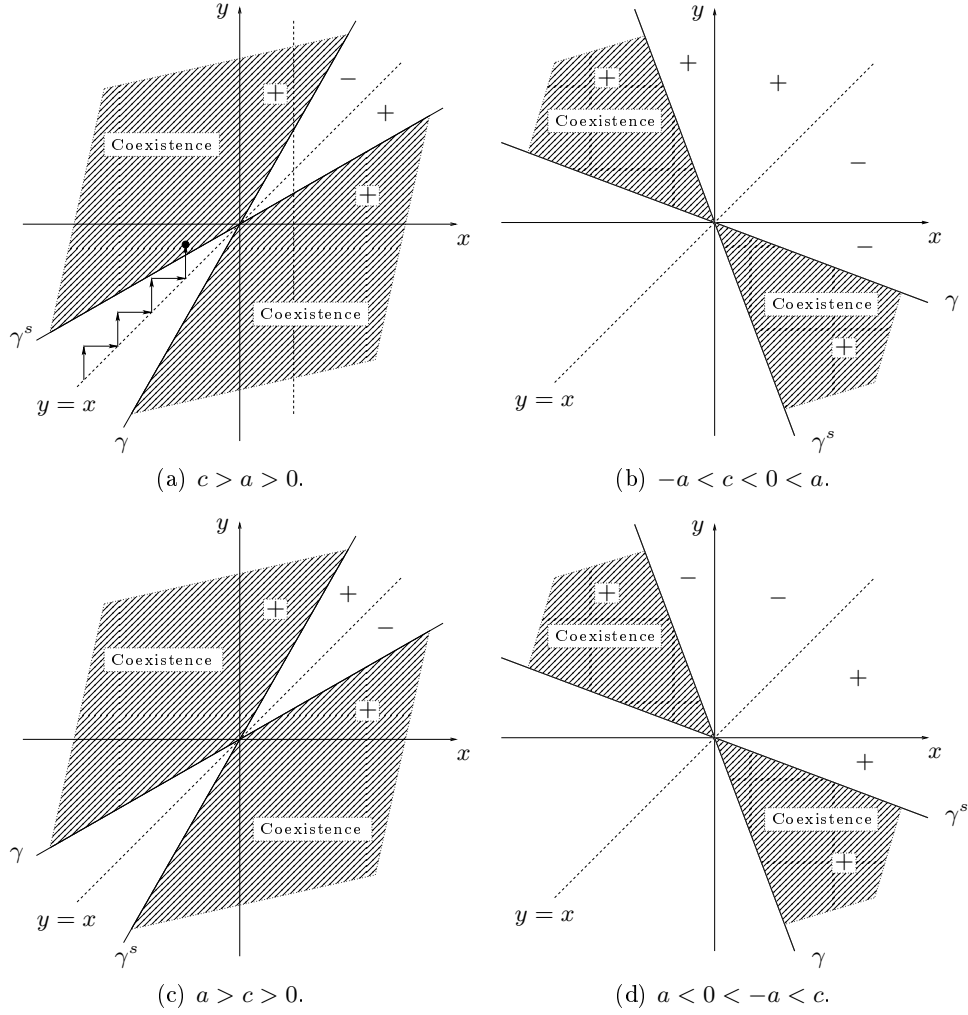


Figure 4.2: In the four cases when coexistence is possible, these figures show the sign configuration of $f(y; x)$ depending on the position of (x, y) with respect to the curve γ and the line $\{y = x\}$ and the region of coexistence. For convenience, we assumed $x^* = 0$.

Fix $\varepsilon \leq \varepsilon_0$. Because of Assumption (A'), the previous quantity is continuous w.r.t. $x \in I$. Therefore, there exists $\beta > 0$ such that

$$\int_0^{+\infty} [g(x + \varepsilon h; x)]_+ m(x, h) dh \geq \beta, \quad \forall x \in I.$$

Since $|g(x + \varepsilon h; x)| \leq C\varepsilon|h| \leq C|h|$ for some constant C and since $m(\cdot, \cdot)$ is bounded, there exists $\beta' > 0$ such that

$$\int_{\beta'}^{+\infty} [g(x + \varepsilon h; x)]_+ m(x, h) dh \geq \beta/2, \quad \forall x \in I.$$

In view of the generator (4.1) of the TSS, this means that, in the TSS, the rate of jump at a distance bigger than β' is uniformly bounded from below in I . Therefore, a.s. on the

event $\{\theta_\eta^\varepsilon(x^*) = \infty\}$, there must be infinitely many jumps in the TSS at a distance bigger than β' . Since the TSS is a.s. monotone, this is a contradiction. A similar argument gives the same conclusion when $x_0 > x^*$. \square

Proof of Theorem 4.9 (b): case $a = \partial_{11}f(x^*; x^*) < 0$

It follows from Theorems 4.4 that for any fixed $\eta > 0$, for ε small enough, the PES stays monomorphic until it reaches $(x^* - \eta, x^* + \eta)$. Moreover, as shown in the proof of Proposition 4.7, no mutant out of $(x^* - \eta, x^* + \eta)$ can invade the population as long as it is monomorphic with support inside this interval.

Now, by Proposition 4.10, when $a < 0$, coexistence may happen in the rescaled PES if $c = \partial_{22}f(x^*; x^*) > -a$. In this case, at the first coexistence time τ^ε , the two traits x and y that coexist belong to $(x^* - \eta, x^* + \eta)$ and are distant of less than $2\varepsilon \text{Diam}(\mathcal{X})$ since $m(x, \cdot)$ has support in $\{x - y; x, y \in \mathcal{X}\}$.

Let us examine what happens when a mutant trait z invades this population. Remind that we showed in the proof of Corollary 4.14 that, if $a < 0$, $x < y$, $f(x; y) > 0$, $f(y; x) > 0$ and $f(z; x, y) > 0$, then $f(x; y, z) < 0$, $f(y; x, z) < 0$, $f(z; y) > 0$ and $f(z; x) > 0$. Examining Fig. 3.1, we see that these conditions are incompatible with all classes except classes 7 and 9. Therefore, once the mutant z invades, the new state of the rescaled PES can be either $\bar{n}(z)\delta_z$ in the case of class 7, or either $\bar{n}_1(x, z)\delta_x + \bar{n}_2(x, z)\delta_z$ or $\bar{n}_1(y, z)\delta_y + \bar{n}_2(y, z)\delta_z$ in the case of class 9. In particular, we see that either the population becomes monomorphic again, or it stays dimorphic, but the distance between the two traits of the support of the PES has decreased. In addition, in both cases, the support of the new state of the PES is a subset of $(x^* - \eta, x^* + \eta)$. Hence, η -branching, as defined in Definition 4.8, cannot occur as soon as $\varepsilon < \eta/(4\text{Diam}(\mathcal{X}))$. Moreover, this shows that, after time $\theta_\eta^\varepsilon(x^*)$, the support of the PES is always included in $(x^* - \eta, x^* + \eta)$. This ends the proof of Theorem 4.9 (b). \square

Proof of Theorem 4.9 (a): case $a > 0$

By Proposition 4.10, when $a > 0$, under the assumptions of Theorem 4.9, we are in the situation of Fig. 4.2 (a), and hence coexistence is always possible in the neighborhood of x^* . Fix $\eta > 0$. We are going to prove that, if η is small enough, then for ε small enough,

- (i) the first time of coexistence τ^ε is a.s. finite and $\text{Supp}(\tilde{\Lambda}_{\tau^\varepsilon}^\varepsilon) \subset (x^* - \eta, x^* + \eta)$ a.s.;
- (ii) after time τ^ε , the distance between the two points of the support of the rescaled PES is non-decreasing and becomes a.s. bigger than $\eta/2$ in finite time, before exiting the interval $(x^* - \eta, x^* + \eta)$.

These two points will clearly imply Theorem 4.9 (a).

For Point (i), observe first that, by Proposition 4.7, if $\tau^\varepsilon < +\infty$, then $\text{Supp}(\tilde{\Lambda}_{\tau^\varepsilon}^\varepsilon) \subset (x^* - \eta, x^* + \eta)$. Thus we only have to prove that $\mathbb{P}(\tau^\varepsilon < \infty) = 1$.

Recall that Fig. 4.2 (a) represents the sign configuration of the fitness function $f(y; x)$. The curves γ and $\{x = y\}$ are the set of zeros of $f(y; x)$. Recall that γ is C^2 and that the slope of γ at (x^*, x^*) is c/a . Recall also the definition of γ^s as the symmetric of γ w.r.t. the line $\{x = y\}$. Since $c > a > 0$ by assumption, the three curves only have a single intersection point (x^*, x^*) , and any vertical line sufficiently close to x^* (as the dashed vertical line in Fig. 4.2 (a)) only has a single intersection point with each of the three curves γ , $\{x = y\}$

and γ^s , the closest from x^* being the intersection point with γ^s , and the farthest, with γ . In view of the vertical line in Fig. 4.2 (a), because of the sign configuration of $f(y; x)$, when x and y are close enough to x^* , the only mutant traits y that can invade the resident trait x in the PES are either closer to x^* than x (i.e. are below the line $\{x = y\}$ when $x > x^*$, or above this line when $x < x^*$), or coexist with x . In particular, if x and y are on opposite sides of x^* , they necessarily coexist.

This shows that, after time $\theta_\eta^\varepsilon(x^*)$ and before time τ^ε , the support of the PES (i.e. the TSS) is monotonous. Therefore, the PES moves according to the arrows in Fig. 4.2 (a), by a succession of mutant invasions (vertical arrows) and fixations (horizontal arrows), until a vertical arrow enters the zone of coexistence. Since the TSS monotonously approaches x^* before τ^ε , this picture explains why coexistence is going to occur almost surely. The proof we give below is based on this argument, taking into account the additional difficulty that the jump rate is almost zero in that case.

Fix $\varepsilon < \varepsilon_0$. Taking ε_0 small enough, it follows from (4.18) and from the inequality $c > a > 0$ that, for all x sufficiently close from x^* and h in the support of $m(x, h)dh$ such that $(x - x^*)h \leq 0$,

$$\begin{aligned} [g(x + \varepsilon h; x)]_+ &\geq C\varepsilon h \left((a - c)(x - x^*) + a\varepsilon h + O(|x - x^*| + \varepsilon|h|) \right) \\ &\geq \frac{C(c - a)}{2}\varepsilon|h|(|x - x^*| + \varepsilon|h|), \end{aligned}$$

for some $C > 0$. Similarly as for the proof of the fact that $\theta_\eta^\varepsilon(x^*) < \infty$ a.s., we then deduce from Assumptions (A'2) and (A'') that there exist constants $\beta, \beta' > 0$ such that, for any $x \in [x^* - \eta, x^* + \eta]$

$$\begin{aligned} \int_{\beta'}^{+\infty} [g(x + \varepsilon h, x)]_+ m(x, h)dh &\geq \beta\varepsilon(|x - x^*| + \varepsilon) \quad \text{if } x \leq x^*, \\ \int_{-\infty}^{-\beta'} [g(x + \varepsilon h, x)]_+ m(x, h)dh &\geq \beta\varepsilon(|x - x^*| + \varepsilon) \quad \text{if } x \geq x^*. \end{aligned}$$

Now, it follows from (4.18) that $\int_{\mathbb{R}} [g(x + \varepsilon h, x)]_+ m(x, h)dh \leq C\varepsilon(|x - x^*| + \varepsilon)$ for some constant $C > 0$. This shows that each jump in the TSS before time τ^ε has a probability bigger than β/C to be larger than β' . Moreover, the total jump rate in the TSS from any point of $[x^* - \eta, x^* + \eta]$ is lower bounded by $\beta\varepsilon^2 > 0$. Therefore, if $\mathbb{P}(\tau^\varepsilon = \infty) > 0$, almost surely on this event, the TSS monotonously would approach x^* from one side of x^* and would have infinitely many jumps, among which infinitely many would be bigger than $\beta' > 0$. This is impossible and thus $\mathbb{P}(\tau^\varepsilon = \infty) = 0$.

For Point (ii), assume that the rescaled PES is dimorphic at some time t , with support $\{x, y\}$, $x < y$. Let us examine what happens when a mutant trait z invades this population. Remind that we assume $a > 0$. In this case, as shown in the proof of Corollary 4.14, reducing η if necessary, for all $x, y, z \in (x^* - \eta, x^* + \eta)$ such that $x < y$, $f(x; y) > 0$, $f(y; x) > 0$ and $f(z; x, y) > 0$, then

- either $z < x < y$ and $f(x; y, z) < 0$, $f(y; x, z) > 0$, $f(z; y) > 0$ and $f(y; z) > 0$,
- or $x < y < z$ and $f(x; y, z) > 0$, $f(y; x, z) < 0$, $f(z; x) > 0$ and $f(x; z) > 0$.

Examining Fig. 3.1, we see that both situations are only compatible with classes 9, 10, 11 and 12. In the case where $z < x < y$, the inequalities $f(y; z) > 0$, $f(z; y)$ and $f(x; y, z) < 0$ imply that the equilibrium $\bar{\mathbf{n}}(y, z)$ is stable, and in the case $x < y < z$, the inequalities $f(x; z) > 0$, $f(z; x) > 0$ and $f(y; x, z) < 0$ imply that the equilibrium $\bar{\mathbf{n}}(x, z)$ is stable. Now, in all the classes 9 to 12, there is only one stable equilibrium point (represented by a filled dot \bullet). Therefore, with the notation of Assumption (B),

$$\mathbf{n}^*(x, y, z) = \begin{cases} (0, \bar{n}_1(y, z), \bar{n}_2(y, z)) & \text{if } z < x < y, \\ (\bar{n}_1(x, z), 0, \bar{n}_2(x, z)) & \text{if } x < y < z. \end{cases} \quad (4.22)$$

In other words, once the mutant z invades, the new state of the rescaled PES is $\bar{n}_1(x, z)\delta_x + \bar{n}_2(x, z)\delta_z$ if $x < y < z$ or $\bar{n}_1(y, z)\delta_y + \bar{n}_2(y, z)\delta_z$ if $z < x < y$. In both cases, the distance between the two existing traits have increased.

In both cases, defining θ' as the first time where one of the points of the support of the PES leaves $(x^* - \eta, x^* + \eta)$, we see that for all $t \in [\tau^\varepsilon, \theta')$, the support of the PES $\tilde{\Lambda}_t^\varepsilon$ is always composed of exactly two traits $\{Y_t^\varepsilon, Z_t^\varepsilon\}$, with $Y_t^\varepsilon < Z_t^\varepsilon$. Moreover, the process $(Y_t^\varepsilon, t \in [\tau^\varepsilon, \theta'))$ is non-increasing, and the process $(Z_t^\varepsilon, t \in [\tau^\varepsilon, \theta'))$ is non-decreasing. In particular, the diameter $Z_t^\varepsilon - Y_t^\varepsilon$ of the support of the PES is non-decreasing. Hence, in order to complete the proof, it suffices to prove that

$$\theta' < \infty \quad \text{a.s.} \quad \text{and} \quad \text{Diam}(\text{Supp}(\tilde{\Lambda}_{\theta'}^\varepsilon)) > \eta/2.$$

The fact that $\theta' < \infty$ a.s. can be proved using Assumptions (A'') and (4.18) in a similar way as for Point (i) above. The lower bound of the diameter of the PES is a consequence of the inequality

$$\tau^\varepsilon > \theta_{\kappa\varepsilon}^\varepsilon(x^*) \quad \text{a.s.}, \quad \text{where} \quad \kappa = \frac{4c\text{Diam}(\mathcal{X})}{c-a}, \quad (4.23)$$

proved below. This shows that $Y_{\tau^\varepsilon}^\varepsilon$ and $Z_{\tau^\varepsilon}^\varepsilon$ belong to $[x^* - \kappa'\varepsilon, x^* + \kappa'\varepsilon]$, where $\kappa' := \kappa + 2\text{Diam}(\mathcal{X})$. Since the processes Y^ε and Z^ε are monotone, we deduce that

$$Y_{\theta'}^\varepsilon \leq x^* + \kappa'\varepsilon \quad \text{and} \quad Z_{\theta'}^\varepsilon \geq x^* - \kappa'\varepsilon.$$

Since one of the two numbers $Y_{\theta'}^\varepsilon$ and $Z_{\theta'}^\varepsilon$ must be outside of the interval $[x^* - \eta, x^* + \eta]$, we finally obtain that $\text{Diam}(\tilde{\Lambda}_{\theta'}^\varepsilon) \geq \eta - \kappa'\varepsilon > \eta/2$ if ε is small enough.

Hence, it only remains to prove (4.23). For any $x, y \in \mathbb{R}$ such that

$$|x - x^*| \geq \frac{2c|x - y|}{c-a}, \quad (4.24)$$

we have

$$|y - x^*| \geq |x - x^*| - |y - x| \geq \frac{c+a}{2c}|x - x^*|$$

and

$$|y - x^*| \leq |x - x^*| + |y - x^*| \leq \left(1 + \frac{c-a}{2c}\right)|x - x^*|.$$

Since $0 < a < c$, we have

$$\frac{c+a}{2c} > \frac{a}{c} \quad \text{and} \quad 1 + \frac{c-a}{2c} < 1 + \frac{c-a}{a} = \frac{c}{a}.$$

Now, $\{(y-x^*) = (c/a)(x-x^*)\}$ is tangent to γ at (x^*, x^*) and $\{(y-x^*) = (a/c)(x-x^*)\}$ is tangent to γ^s at (x^*, x^*) . Therefore, for any x, y sufficiently close to x^* and satisfying (4.24), the point (x, y) is between the two curves γ and γ^s (i.e. below γ and above γ^s if $x > x^*$, and below γ^s and above γ if $x < x^*$, see Fig. 4.2 (a)). In particular, x and y do not coexist. Therefore, at time τ^ε , the two traits $Y_{\tau^\varepsilon}^\varepsilon$ and $Z_{\tau^\varepsilon}^\varepsilon$ of the support of the PES must satisfy

$$|Y_{\tau^\varepsilon}^\varepsilon - x^*| \vee |Z_{\tau^\varepsilon}^\varepsilon - x^*| \leq \frac{2c}{c-a} |Y_{\tau^\varepsilon}^\varepsilon - Z_{\tau^\varepsilon}^\varepsilon| \leq \frac{4c\varepsilon \text{Diam}(\mathcal{X})}{c-a} = \kappa\varepsilon,$$

since one of the traits $Y_{\tau^\varepsilon}^\varepsilon$ and $Z_{\tau^\varepsilon}^\varepsilon$ is a mutant born from the other trait. \square

A Proof of Theorem 2.7

The proof of this result is very similar to the proof of [4, Thm.1]. We do not repeat all the details and restrict ourselves to the steps that must be modified. The general idea of the proof follows closely the heuristic argument of Section 2.4.2. Its skeleton is similar to the one in [4] for monomorphic populations.

For all $\varepsilon > 0$, $t > 0$, and $\Gamma \subset \mathcal{X}$ measurable, let

$$A_{\varepsilon,d}(t, \Gamma) := \left\{ \text{Supp}(\nu_{t/Ku_K}) \subset \Gamma \text{ has } d \text{ elements that coexist, say } x_1, \dots, x_d, \right. \\ \left. \text{and } \forall 1 \leq i \leq d, |\langle \nu_{t/Ku_K}, \mathbf{1}_{\{x_i\}} \rangle - \bar{n}_i(\mathbf{x})| < \varepsilon \right\}.$$

To prove Theorem 2.7, we establish that for all $\varepsilon > 0$, $t > 0$ and $\Gamma \subset \mathcal{X}$ measurable,

$$\lim_{K \rightarrow +\infty} \mathbb{P}(A_{\varepsilon,d}(t, \Gamma)) = \mathbb{P}(\text{Supp}(\Lambda_t) \subset \Gamma \text{ and has } d \text{ elements}). \quad (\text{A.1})$$

where $(\Lambda_t, t \geq 0)$ is defined in Theorem 2.7. The first ingredient of the proof is the following proposition, which generalizes Theorem 3 (a) and (b) of [4].

Proposition A.1 *Assume that, for any $K \geq 1$, $\text{Supp}(\nu_0^K) = \{x_1, \dots, x_d\}$ and $\langle \nu_0^K, \mathbf{1}_{\{x_i\}} \rangle \in C$ a.s., where C is a compact subset of \mathbb{R}_+ . Let $\phi(t, (n_1, \dots, n_d))$ denote the value at time t of the solution of $LV(d, \mathbf{x})$ with initial condition (n_1, \dots, n_d) . Then, for all $T > 0$,*

$$\lim_{K \rightarrow +\infty} \sup_{1 \leq i \leq d, t \in [0, T]} \left| \langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle - \phi_i(t, (\langle \nu_0^K, \mathbf{1}_{\{x_1\}} \rangle, \dots, \langle \nu_0^K, \mathbf{1}_{\{x_d\}} \rangle)) \right| = 0 \quad \text{a.s.} \quad (\text{A.2})$$

This result is a direct corollary of Theorem 11.2.1 of [13], except for two small difficulties. The first one is that Theorem 11.2.1 of [13] assumes that the function $\mathbf{n} \mapsto F^{\mathbf{x}}(\mathbf{n})$ involved in the definition (2.4) of the Lotka Volterra system is uniformly Lipschitz on \mathbb{R}_+^d , which is not the case. However, observe first that, if $n_i \leq M$ for some $M > 0$ for all $i \in \{1, \dots, d\}$, then $\phi_i(t, (n_1, \dots, n_d)) \leq M \vee (2\bar{\lambda}/\underline{\alpha})$ for all $t \geq 0$. Indeed, if there is equality for some $t \geq 0$ and $i \in \{1, \dots, d\}$, then $\dot{\phi}_i(t, (n_1, \dots, n_d)) < 0$. Therefore, the coefficients of the system $LV(d, \mathbf{x})$ are uniformly Lipschitz on the set of states that can be attained by the solution of the system starting from any initial conditions in a compact set. The second difficulty is that Theorem 11.2.1 of [13] only implies that (A.2) holds on the event where there is no mutation between 0 and T . In Lemma 2 (a) of [4], it is proved that for general initial condition ν_0^K , the probability of mutation on the time interval $[0, T]$ converges to 0, thus the conclusion follows.

The second ingredient is the following exponential deviation estimate on the so-called “problem of exit from an attracting domain” [15]. It generalizes Theorem 3 (c) of [4].

Proposition A.2 *Let $x_1, \dots, x_d \in \mathcal{X}$ coexist. Then there exist constants $c, V > 0$ such that, for any sufficiently small $\varepsilon > 0$, if $(\langle \nu_0^K, \mathbf{1}_{\{x_i\}} \rangle)_{1 \leq i \leq d}$ belongs to the $(\varepsilon/2)$ -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$, the time of exit of $(\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle)_{1 \leq i \leq d}$ from the ε -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$ is bigger than $e^{VK} \wedge \tau$ with probability converging to 1, where τ denotes the first mutation time. Moreover, the previous result also holds if, for all $i \in \{1, \dots, d\}$, the death rate of an individual with trait x_i*

$$\mu(x_i) + \sum_{j=1}^d \alpha(x_i, x_j) \langle \nu_t^K, \mathbf{1}_{\{x_j\}} \rangle \quad (\text{A.3})$$

is perturbed by an additional random process that is uniformly bounded by $c\varepsilon$.

Such results are fairly standard and can be proved in a variety of ways. We let the proof to the reader. The first part of this proposition is used to prove that, when the first mutation occurs, the population densities have never left the ε -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$ and the second is used to prove that, after the first mutation, as long as the mutant population is small, the resident population densities do not leave the ε -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$. In this case, the additional term in (A.3) is $\alpha(x_i, y) \langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$, where y is the mutant trait, which is smaller than $\bar{\alpha}\varepsilon$ if $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle \leq \varepsilon$.

From these two results can be deduced the following lemma, which is the extension of Lemma 2 (b) and (c) of [4]. The proof is a simple copy of the argument in [4].

Lemma A.3 *Let $\text{Supp}(\nu_0^K) = \{x_1, \dots, x_d\}$ that coexist and let τ denote the first mutation time. There exists ε_0 such that, if $(\langle \nu_0^K, \mathbf{1}_{\{x_i\}} \rangle)_{1 \leq i \leq d}$ belongs to the ε_0 -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$, then, for any $\varepsilon < \varepsilon_0$,*

$$\lim_{K \rightarrow +\infty} \mathbb{P} \left(\tau > \log K, \sup_{1 \leq i \leq d, t \in [\log K, \tau]} |\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle - \bar{n}_i(\mathbf{x})| < \varepsilon \right) = 1,$$

$$Ku_K \tau \xrightarrow[K \rightarrow \infty]{\mathcal{L}} \text{Exp} \left(\sum_{j=1}^d p(x_j) \lambda(x_j) \bar{n}_j(\mathbf{x}) \right)$$

$$\text{and } \lim_{K \rightarrow +\infty} \mathbb{P}(\text{at time } \tau, \text{ the mutant is born from trait } x_i) = \frac{p(x_i) \lambda(x_i) \bar{n}_i(\mathbf{x})}{\sum_{j=1}^d p(x_j) \lambda(x_j) \bar{n}_j(\mathbf{x})}$$

for all $i \in \{1, \dots, d\}$, where $\xrightarrow{\mathcal{L}}$ denotes the convergence in law of real r.v. and $\text{Exp}(u)$ denotes the exponential law with parameter u .

The fourth ingredient is the following lemma, which is an extension of Lemma 3 of [4].

Lemma A.4 *Let $\text{Supp}(\nu_0^K) = \{x_1, \dots, x_d, y\}$ where x_1, \dots, x_d coexist and y is a mutant trait that satisfy Assumption (B). Let τ denote the first next mutation time, and define*

$$\begin{aligned} \tau_1 &= \inf \{ t \geq 0 : \forall i \in I(\mathbf{n}^*), |\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle - n_i^*| < \varepsilon \text{ and } \forall i \notin I(\mathbf{n}^*), \langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle = 0 \} \\ \tau_2 &= \inf \{ t \geq 0 : \langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle = 0 \text{ and } \forall i \in \{1, \dots, d\}, |\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle - \bar{n}_i(\mathbf{x})| < \varepsilon \}. \end{aligned}$$

Assume that $\langle \nu_0^K, \mathbf{1}_{\{y\}} \rangle = 1/K$ (a single initial mutant). Then, there exists ε_0 such that for all $\varepsilon < \varepsilon_0$, if $(\langle \nu_0^K, \mathbf{1}_{\{x_i\}} \rangle)_{1 \leq i \leq d}$ belongs to the ε -neighborhood of $\bar{\mathbf{n}}(\mathbf{x})$,

$$\lim_{K \rightarrow +\infty} \mathbb{P}(\tau_1 < \tau_2) = \frac{[f(y; \mathbf{x})]_+}{\lambda(y)}, \quad \lim_{K \rightarrow +\infty} \mathbb{P}(\tau_2 < \tau_1) = 1 - \frac{[f(y; \mathbf{x})]_+}{\lambda(y)}$$

$$\text{and } \forall \eta > 0, \quad \lim_{K \rightarrow +\infty} \mathbb{P} \left(\tau_1 \wedge \tau_2 < \frac{\eta}{Ku_K} \wedge \tau \right) = 1.$$

The proof of this lemma is similar to the proof of Lemma 3 in [4]. The main steps are the following. Assume first that $\varepsilon < 1/2$. We introduce the following stopping times:

$$\begin{aligned} R_\varepsilon^K &= \inf\{t \geq 0 : \exists i \in \{1, \dots, d\}, |\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle - \bar{n}_i(\mathbf{x})| \geq \varepsilon\} \\ S_\varepsilon^K &= \inf\{t \geq 0 : \langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle \geq \varepsilon\} \\ S_0^K &= \inf\{t \geq 0 : \langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle = 0\}. \end{aligned}$$

R_ε^K is the time of drift of the resident population away from its equilibrium, S_ε^K is the time of *invasion* of the mutant trait (time t_1 in Fig. 2.5) and S_0^K is the time of extinction of the mutant trait. By the second part of Proposition A.2, it can be proven exactly as in [4] that there exists $\rho, V > 0$ and $c < 1$ such that, for K large enough,

$$\mathbb{P} \left(\frac{\rho}{Ku_K} < \tau \right) \geq 1 - \varepsilon \quad \text{and} \quad \mathbb{P}(S_\varepsilon^K \wedge \tau \wedge e^{KV} < R_{\varepsilon/c}^K) \geq 1 - \varepsilon.$$

Then, on $[0, \tau \wedge S_\varepsilon^K \wedge R_{\varepsilon/c}^K]$, by computing lower and upper bounds on the death rate of a mutant individual, it can be easily checked that, for K large enough, almost surely,

$$\frac{\Lambda_t^{1, \varepsilon}}{K} \leq \langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle \leq \frac{\Lambda_t^{-1, \varepsilon}}{K}$$

where, for $i = 1$ or -1 , $\Lambda^{i, \varepsilon}$ is a continuous-time branching process such that $\Lambda_0^{i, \varepsilon} = 1$ and with birth rate $(1 - i\varepsilon)\lambda(y)$ and death rate

$$\mu(y) + \sum_{j=1}^d \alpha(y, x_j) \bar{n}_j(\mathbf{x}) + i(d+1) \bar{\alpha} \frac{\varepsilon}{c}.$$

Next, we use the results of Theorem 4 of [4] on branching processes in order to control the probability that $\Lambda^{i, \varepsilon}/K$ exceeds ε before it reaches 0, and to upper bound the time at which one of these events happens. As in [4], we obtain that there exists $C > 0$ such that, for all $\eta > 0$, $\varepsilon > 0$ sufficiently small and K large enough,

$$\begin{aligned} \mathbb{P} \left(\tau_2 < \tau \wedge \frac{\eta}{Ku_K} \wedge S_\varepsilon^K \wedge R_{\varepsilon/c}^K \right) &\geq 1 - \frac{[f(y; \mathbf{x})]_+}{\lambda(y)} - C\varepsilon \\ \mathbb{P} \left(S_\varepsilon^K < \tau \wedge \frac{\eta}{Ku_K} \wedge S_0^K \wedge R_{\varepsilon/c}^K \right) &\geq \frac{[f(y; \mathbf{x})]_+}{\lambda(y)} - C\varepsilon. \end{aligned} \tag{A.4}$$

On the event $\{S_\varepsilon^K < \tau \wedge S_0^K \wedge R_{\varepsilon/c}^K\}$, we introduce for $\varepsilon' > 0$ the stopping times

$$\begin{aligned} T_\varepsilon^K &= \inf\{t \geq S_\varepsilon^K : \forall i \in \{1, \dots, d\}, |\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle - n_i^*| < \varepsilon^2 \text{ and } |\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle - n_{d+1}^*| < \varepsilon^2\}, \\ U_{\varepsilon, \varepsilon'}^K &= \inf\{t \geq T_\varepsilon^K : \exists i \in I(\mathbf{n}^*), |\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle - n_i^*| \geq \varepsilon'\}, \\ V_\varepsilon^K &= \inf\{t \geq T_\varepsilon^K : \exists i \notin I(\mathbf{n}^*), \langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle \geq \varepsilon\}. \end{aligned}$$

We next use the Markov property at time S_ε^K and apply Proposition A.1 as in [4] to obtain a constant $C' > C$ such that, for K large enough,

$$\mathbb{P}\left(S_\varepsilon^K < T_\varepsilon^K < \tau \wedge \frac{\eta}{Ku_K}\right) \geq \frac{[f(y; \mathbf{x})]_+}{\lambda(y)} - C'\varepsilon.$$

Next, we can use again Proposition A.2 to prove that there exists $V' > 0$, $C'' > C'$ and $c' < 1$ such that

$$\mathbb{P}(S_\varepsilon^K < T_\varepsilon^K < V_\varepsilon^K \wedge \tau \wedge e^{KV'} < U_{\varepsilon, \varepsilon/c'}^K) \geq \frac{[f(y; \mathbf{x})]_+}{\lambda(y)} - C''\varepsilon.$$

In a last step, we can as before prove that, for all $t \in [T_\varepsilon^K, U_{\varepsilon, \varepsilon/c'}^K \wedge V_\varepsilon^K]$ and for all $i \notin I(\mathbf{n}^*)$,

$$\langle \nu_t^K, \mathbf{1}_{\{x_i\}} \rangle \leq \frac{\tilde{\Lambda}_t^{i, \varepsilon}}{K},$$

where $\tilde{\Lambda}^{i, \varepsilon}$ is a continuous-time branching process such that $\tilde{\Lambda}_{T_\varepsilon^K}^{i, \varepsilon} = \lceil \varepsilon^2 K \rceil$ and with birth rate $\lambda(x_i)$ and death rate

$$\mu(x_i) + \sum_{j \in I(\mathbf{n}^*)} \alpha(x_i, x_j) n_j^* - \text{Card}(I(\mathbf{n}^*)) \bar{\alpha} \frac{\varepsilon}{c'}.$$

Since, by Assumption (B2), $f(x_i; \mathbf{x}^*) < 0$, this branching process is sub-critical if ε is small enough. Hence, with arguments similar to the ones in [4] (especially the results of Theorem 4), we can prove that there exist $C''' > 0$ such that, for all $\eta > 0$, $\varepsilon > 0$ sufficiently small and K large enough,

$$\mathbb{P}\left(S_\varepsilon^K < \tau_1 < \tau \wedge \frac{\eta}{Ku_K} \wedge U_{\varepsilon, \varepsilon/c'}^K\right) \geq \frac{[f(y; \mathbf{x})]_+}{\lambda(y)} - C'''\varepsilon.$$

Combining this with (A.4), we obtain Lemma A.4 by letting ε go to 0.

Finally, (A.1) is deduced from these lemmas exactly as in [4] and similarly, the proof of Theorem 2.7 from (A.1). \square

B Explanation of Figure 3.1

Let us comment the classification of the asymptotic behavior of 3-dimensional competitive Lotka-Volterra systems done by Zeeman [33]. Any 3-dimensional competitive Lotka-Volterra system admits an invariant hypersurface Σ called carrying simplex, such that any non-zero solution of the system is asymptotic as $t \rightarrow +\infty$ to one in Σ (cf. [21]). Σ is a Lipschitz submanifold of \mathbb{R}_+^3 homeomorphic to the unit simplex in \mathbb{R}_+^3 by radial projection. Moreover, Σ is a global attractor for the dynamics in $\mathbb{R}_+^3 \setminus \{0\}$ ([22, Thm.3]). In particular, one can deduce from the asymptotic behavior of trajectories on Σ the asymptotic behavior of trajectories starting in a neighborhood of Σ .

Zeeman obtained a full classification of the topological equivalence classes of the 3-dimensional competitive Lotka-Volterra systems by determining the 33 topological equivalence classes

of those systems restricted on their carrying simplex. (In an equivalence class, the trajectories of two systems are related by a homeomorphism of \mathbb{R}_+^3). For a given system $LV(3, (x, y, z))$, the equivalence class to which it belongs is determined by the sign of the 2-dimensional fitnesses $f(x; y)$, $f(y; x)$, $f(x; z)$, $f(z; x)$, $f(y; z)$, $f(z; y)$ and of the 3-dimensional fitnesses $f(x; y, z)$, $f(y; x, z)$, $f(z; x, y)$ when they are defined. The equivalence classes of [33] are characterized by drawing on the unit simplex of \mathbb{R}_+^3 the fixed points and the limit cycles of the system, joined by their stable and unstable manifolds¹.

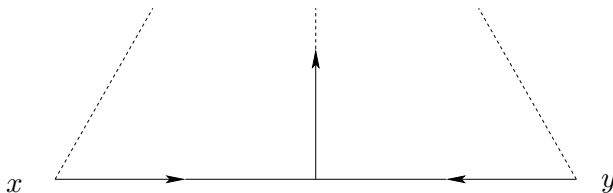


Figure B.1: The pattern on the carrying simplex that corresponds to the case where the two (resident) traits x and y coexist and the third (mutant) trait z has a positive fitness $f(z; x, y)$.

The signs of the fitnesses correspond to the arrows in each diagram. For example, $f(y; x) > 0$ means that, on the edge of the simplex that reach x and y , there is an arrow starting from x in the direction of y . In other words, the unstable manifold of $(\bar{n}(x), 0, 0)$ contains (a part of) the edge of the simplex that reach x to y . Similarly, $f(z; x, y) > 0$ means that $f(x; y)f(y; x) > 0$, i.e. that $LV(3, (x, y, z))$ has as fixed point $(\bar{n}_1(x, y), \bar{n}_2(x, y), 0)$ with $\bar{n}_1(x, y) > 0$ and $\bar{n}_2(x, y) > 0$, represented as the midpoint of the edge of the simplex linking x and y , and that this fixed point has an unstable manifold pointing in the direction of the interior of the simplex. The situation represented in Fig. B.1 corresponds to this case, when x and y coexist.

In order to check if Assumption (B) holds, we only need to restrict to the equivalence classes in which two traits coexist (the resident traits, say x and y), and the third (mutant) trait (say z) satisfy $f(z; x, y) > 0$. This situation corresponds to the cases when the carrying simplex has one side containing the pattern of Fig. B.1. Among the 33 equivalence classes of [33], there are only 10 of them that satisfy this requirement, shown in Fig 3.1. We label them with the same numbers as in [33]. In Fig. 3.1, the figures obtained by exchanging x and y belong to the same equivalence class. An attracting fixed point of $LV(3, (x, y, z))$ is represented by a filled dot \bullet , a repulsive fixed point by an empty dot \circ , a saddle point by the intersection of its stable and unstable manifolds. When the interior fixed point (the non-trivial equilibrium) is not a saddle point, it can be either stable or unstable. Depending on cases, this equilibrium can also be surrounded by one or several stable or unstable cycles. In particular, the sign of the fitnesses is not sufficient to determine the precise asymptotic behavior of the system near the interior equilibrium. The undetermined type of these equilibria is represented in Fig. 3.1 by the symbol \odot .

¹The stable manifold of an equilibrium is the set of starting points of the Lotka-Volterra system such that the solution converges to this equilibrium. The unstable manifold is defined in the same way, but for the time-reversed system.

C Proof of Theorem 4.1

We are actually going to prove that Theorem 4.1 holds under weaker assumptions, that we omitted in Section 4 to simplify the presentation.

Theorem C.1 *The statement of Theorem 4.1 holds if Assumptions (A'2) and (A'3) are replaced by*

(A'4) *The map $x \mapsto \int m(x, h)dh$ is Lipschitz continuous from \mathcal{X} to the set of probability measures $\mathcal{P}(\mathbb{R}^l)$, for the Wasserstein metric*

$$\rho(P_1, P_2) = \inf \left\{ \int_{\mathbb{R}^l \times \mathbb{R}^l} |x-y| R(dx, dy); R \in \mathcal{P}(\mathbb{R}^l \times \mathbb{R}^l) \text{ with marginals } P_1 \text{ and } P_2 \right\}.$$

(Recall that \mathcal{X} is a compact set of \mathbb{R}^l .)

(A'5) *The function*

$$g(y; x) = p(x)\lambda(x)\bar{n}(x)\frac{f(y; x)}{\lambda(y)}$$

is continuous on \mathcal{X}^2 (and thus bounded), and of class \mathcal{C}^1 with respect to its first coordinate, where $f(\cdot; \cdot)$ is defined in (2.12). Since \mathcal{X} is a compact set of \mathbb{R}^l , there exists a constant $G > 0$ such that $\forall x, y \in \mathcal{X}$, $[g(y; x)]_+ \leq G$.

Proof The proof follows a classical uniqueness-compactness argument.

(i) **Uniqueness of the solution of Equation (4.4) with given initial condition.**

Let us show that $a(x) = \int_{\mathbb{R}^l} h[h \cdot \nabla_1 g(x; \mathbf{x})]_+ m(x, h)dh$ is Lipschitz continuous on \mathcal{X} . We have

$$\begin{aligned} \|a(x) - a(x')\| &\leq \int_{\mathbb{R}^l} \|h\| \times |[h \cdot \nabla_1 g(x; x)]_+ - [h \cdot \nabla_1 g(x'; x')]_+| m(x, h)dh \\ &\quad + \left\| \int_{\mathbb{R}^l} h[h \cdot \nabla_1 g(x'; x')]_+ (m(x, h) - m(x', h))dh \right\|. \end{aligned} \quad (\text{C.1})$$

Because of $|[a]_+ - [b]_+| \leq |a - b|$ and Assumption (A'), the first term of the right hand side of (C.1) is bounded by some constant times $\|x - x'\|$.

If we denote by ξ the vector $\nabla_1 g(x'; x')$ and $\psi(h) = h[h \cdot \xi]_+$, then

$$\|\psi(h) - \psi(h')\| \leq \|(h - h')[h \cdot \xi]_+\| + \|h'([h \cdot \xi]_+ - [h' \cdot \xi]_+)\| \leq 2\|\xi\| \|h - h'\| (\|h\| + \|h'\|).$$

Thus, using the dual form of the Kantorovich-Rubinstein metric (see Rachev [30]) and (A'), one obtains that the second term of the right-hand side of (C.1) is also bounded by some constant times $\|x - x'\|$. Hence Cauchy-Lipschitz Theorem can be applied and $(x(t), t \geq 0)$ is uniquely defined.

(ii) **The processes X^ε , $\varepsilon > 0$, with generator L^ε can be constructed on the same probability space.**

Recall the definition of \bar{m} in Assumption (A3).

Lemma C.2 Assume (A) and (A'). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $N(dh, d\theta, ds)$ be a point Poisson measure on $\mathbb{R}^l \times [0, 1] \times \mathbb{R}_+$ with intensity $G\bar{m}(h)dhd\theta ds$. Let $\varepsilon > 0$ and denote by N^ε the image measure of N by the mapping $s \mapsto \varepsilon^2 s$. Let X_0^ε be a \mathcal{X} -valued random variable, independent of N . Then the process X^ε defined by

$$X_t^\varepsilon = X_0^\varepsilon + \int_{\mathbb{R}^l \times [0, 1] \times [0, t]} (\varepsilon h) \mathbf{1}_{\left\{ \theta \leq \frac{[g(X_{s-}^\varepsilon + \varepsilon h; X_{s-}^\varepsilon)]_+ m(X_{s-}^\varepsilon, h)}{G} \right\}} N^\varepsilon(dh, d\theta, ds), \quad (\text{C.2})$$

is a jump Markov process with generator L^ε . Its law will be denoted by $\mathbf{P}_{X_0^\varepsilon}^\varepsilon$.

Indeed, for a bounded function φ on \mathcal{X} ,

$$\begin{aligned} \varphi(X_t^\varepsilon) &= \varphi(X_0^\varepsilon) \\ &\quad + \int_{\mathbb{R}^l \times [0, 1] \times [0, t]} (\varphi(X_{s-}^\varepsilon + \varepsilon h) - \varphi(X_{s-}^\varepsilon)) \mathbf{1}_{\left\{ \theta \leq \frac{1}{\varepsilon^2} \frac{[g(X_{s-}^\varepsilon + \varepsilon h; X_{s-}^\varepsilon)]_+ m(X_{s-}^\varepsilon, h)}{G} \right\}} N^\varepsilon(dh, d\theta, ds) \\ &= \varphi(X_0^\varepsilon) + M_t^{\varepsilon, \varphi} + \int_0^t \int_{\mathbb{R}^l \times [0, 1]} \frac{1}{\varepsilon^2} (\varphi(X_s^\varepsilon + \varepsilon h) - \varphi(X_s^\varepsilon)) g(X_s^\varepsilon + \varepsilon h; X_s^\varepsilon) m(X_s^\varepsilon, h) dh d\theta ds, \end{aligned}$$

where $M^{\varepsilon, \varphi}$ is a martingale.

(iii) **Tightness of the sequence of laws $\{\mathbf{P}_{X_0^\varepsilon}^\varepsilon\}_{\varepsilon > 0}$ of the processes $(X_t^\varepsilon, t \geq 0)$**

We will use Aldous' criterion [1]. Let τ be a stopping time less than T and (δ_ε) positive numbers converging to 0 when $\varepsilon \rightarrow 0$. We remark that $|g(x + \varepsilon h; x)| \leq \varepsilon C \|h\|$, by an expansion of g with respect to its first variable and the fact that $g(x; x) = 0$ and since $\nabla_1 g$ is bounded by a constant C . We have

$$\mathbb{E}(\|X_{\tau+\delta_\varepsilon}^\varepsilon - X_\tau^\varepsilon\|) = \mathbb{E} \left(\int_\tau^{\tau+\delta_\varepsilon} \int_{\mathbb{R}^l} \|\varepsilon h\| [g(X_{s-}^\varepsilon + \varepsilon h; X_{s-}^\varepsilon)]_+ m(X_{s-}^\varepsilon, h) dh \frac{ds}{\varepsilon^2} \right) \leq CM_2 \delta_\varepsilon,$$

where $M_2 = \int \|h\|^2 \bar{m}(h) dh$. Then, for any $\alpha > 0$,

$$\mathbb{P}(\|X_{\tau+\delta_\varepsilon}^\varepsilon - X_\tau^\varepsilon\| > \alpha) \leq \frac{CM_2}{\alpha} \delta_\varepsilon \rightarrow 0 \quad \text{when } \varepsilon \rightarrow 0.$$

This gives the first part of Aldous' criterion. For the second part, we have to prove the uniform tightness of the laws of $(\sup_{t \leq T} \|X_t^\varepsilon\|)_{\varepsilon > 0}$. We use Itô's formula to write $(X_t^\varepsilon)^2$ from (C.2). Next, using Cauchy-Schwarz' and Doob's inequalities, we obtain by a localization argument and Gronwall's lemma that $\mathbb{E}(\sup_{t \leq T} \|X_t^\varepsilon\|^2) \leq C_T (\mathbb{E}(\|X_0^\varepsilon\|^2) + 1)$, where C_T is a constant depending on T , M_2 and G . Since $(X_0^\varepsilon)_{0 < \varepsilon \leq 1}$ is bounded in \mathbb{L}^2 , the tightness of the laws of $(\sup_{t \leq T} \|X_t^\varepsilon\|)_{\varepsilon > 0}$ follows.

(iv) **Convergence of the generators.**

Let us now prove that

$$\forall \varphi \in \mathcal{C}_b^2(\mathcal{X}), \quad L^\varepsilon \varphi \rightarrow L^0 \varphi \quad \text{uniformly on } \mathcal{X}, \quad (\text{C.3})$$

where L^ε is defined in (3.3) and L^0 is defined by

$$L^0 \varphi(\mathbf{x}) = \int_{\mathbb{R}^l} (h \cdot \nabla \varphi(x)) [h \cdot \nabla_1 g(x; x)]_+ m(x, h) dh,$$

where $\nabla\varphi(x)$ is the gradient vector of $\varphi(x)$. We have,

$$\begin{aligned} |L^\varepsilon\varphi(x) - L^0\varphi(x)| &\leq \int_{\mathbb{R}^l} [h \cdot \nabla_1 g(x; x)]_+ \times \left| \frac{\varphi(x + \varepsilon h) - \varphi(x)}{\varepsilon} - h \cdot \nabla\varphi(x) \right| m(x, h) dh \\ &+ \int_{\mathbb{R}^l} \left| \frac{\varphi(x + \varepsilon h) - \varphi(x)}{\varepsilon} \right| \times \left| \left[\frac{g(x + \varepsilon h; x)}{\varepsilon} \right]_+ - [h \cdot \nabla_1 g(x; x)]_+ \right| m(x, h) dh. \end{aligned} \quad (\text{C.4})$$

Let us call I_1 and I_2 the quantities inside the integral in the first and the second term, respectively. Now, φ is \mathcal{C}^1 , $g(x; x) = 0$ and by Assumption (A'), $g(x; y)$ is \mathcal{C}^1 with respect to the first variable x . So, we can find θ_1, θ_2 and θ_3 in $[0, 1]$ depending on x and h such that

$$\begin{aligned} I_1 &= [h \cdot \nabla_1 g(x; x)]_+ \times |h \cdot \nabla\varphi(x + \theta_3 \varepsilon h) - h \cdot \nabla\varphi(x)|; \\ I_2 &= |h \cdot \nabla\varphi(x + \theta_1 \varepsilon h)| \times |[h \cdot \nabla_1 g(x + \theta_2 \varepsilon h; x)]_+ - [h \cdot \nabla_1 g(x; x)]_+|. \end{aligned}$$

Since φ is in \mathcal{C}_b^2 , and because of Assumption (A'), we can choose a number C such that $\nabla\varphi$ and $\nabla_1 g$ are both C -Lipschitz and bounded by C on \mathcal{X} and \mathcal{X}^2 respectively. Then

$$\begin{aligned} I_1 &\leq C\|h\| \times \|h\|C\|\theta_3 \varepsilon h\| \leq \varepsilon C^2 \|h\|^3; \\ I_2 &\leq C\|h\| \times |h \cdot \nabla_1 g(x + \theta_2 \varepsilon h, x) - h \cdot \nabla_1 g(x, x)| \leq \varepsilon C^2 \|h\|^3. \end{aligned}$$

It remains to put these bounds in Equation (C.4) to obtain:

$$|L^\varepsilon\varphi(x) - L^0\varphi(x)| \leq 2\varepsilon C^2 \int_{\mathbb{R}^l} \|h\|^3 m(x, h) dh.$$

We conclude using Assumption (A').

(v) **Martingale problem for limiting distributions.**

Once we have the uniform tightness and the convergence of the generators, it is standard (cf. Ethier-Kurtz [13] Lemma 4.5.1) to deduce that any accumulation point \mathbf{P} of the family of laws $\{\mathbf{P}_{X_0^\varepsilon}^\varepsilon\}$ on $\mathbb{D}([0, T], \mathcal{X})$ is the law of the process X solution to (4.4) with initial state X_0 . The theorem then follows from the uniqueness proved in (i). \square

Acknowledgements We thank Michel Benaïm and Salome Martinez who brought to our attention the reference [33] and for helpful discussions concerning Lotka-Volterra systems. We warmly thank the anonymous referees for very valuable comments, thanks to which the paper has been considerably improved.

References

- [1] D. Aldous. Stopping times and tightness. *Ann. Probab.*, 6(2):335–340, 1978.
- [2] G. Barles and B. Perthame. Concentrations and constrained Hamilton-Jacobi equations arising in adaptive dynamics. In *Recent developments in nonlinear partial differential equations*, volume 439 of *Contemp. Math.*, pages 57–68. Amer. Math. Soc., Providence, RI, 2007.

- [3] A. Bovier and N. Champagnat. Time scales in adaptive dynamics: directional selection, fast and slow branching. In prep.
- [4] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Process. Appl.*, 116(8):1127–1160, 2006.
- [5] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic evolution. *Theor. Pop. Biol.*, 69:297–321, 2006.
- [6] N. Champagnat, R. Ferrière, and S. Méléard. From individual stochastic processes to macroscopic models in adaptive evolution. *Stoch. Models*, 24(suppl. 1):2–44, 2008.
- [7] R. Cressman and J. Hofbauer. Measure dynamics on a one-dimensional continuous trait space: theoretical foundations for adaptive dynamics. *Theor. Pop. Biol.*, 67:47–59, 2003.
- [8] L. Desvillettes, P.-E. Jabin, S. Mischler, and G. Raoul. On selection dynamics for continuous structured populations. *Commun. Math. Sci.*, 6(3):729–747, 2008.
- [9] U. Dieckmann and M. Doebeli. On the origin of species by sympatric speciation. *Nature*, 400:354–357, 1999.
- [10] U. Dieckmann and R. Law. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.*, 34(5-6):579–612, 1996.
- [11] O. Diekmann, P.-E. Jabin, S. Mischler, and B. Perthame. The dynamics of adaptation: An illuminating example and a Hamilton-Jacobi approach. *Theor. Pop. Biol.*, 67:257–271, 2005.
- [12] M. Durinx, J. A. J. Metz, and G. Meszéna. Adaptive dynamics for physiologically structured population models. *J. Math. Biol.*, 56(5):673–742, 2008.
- [13] S. N. Ethier and T. G. Kurtz. *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.
- [14] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004.
- [15] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984. Translated from the Russian by Joseph Szücs.
- [16] S. A. H. Geritz. Resident-invader dynamics and the coexistence of similar strategies. *J. Math. Biol.*, 50(1):67–82, 2005.
- [17] S. A. H. Geritz, M. Gyllenberg, F. J. A. Jacobs, and K. Parvinen. Invasion dynamics and attractor inheritance. *J. Math. Biol.*, 44(6):548–560, 2002.

- [18] S. A. H. Geritz, J. A. J. Metz, E. Kisdi, and G. Meszén. The dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.*, 78:2024–2027, 1997.
- [19] S. A. H. Geritz, E. van der Meijden, and J. A. J. Metz. Evolutionary dynamics of seed size and seedling competitive ability. *Theor. Pop. Biol.*, 55:324–343, 1999.
- [20] J. Guckenheimer and P. Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.
- [21] M. W. Hirsch. Systems of differential equations which are competitive or cooperative. III. Competing species. *Nonlinearity*, 1(1):51–71, 1988.
- [22] M. W. Hirsch. On existence and uniqueness of the carrying simplex for competitive dynamical systems. *Journal of Biological Dynamics*, 2(2):169–179, 2008.
- [23] J. Hofbauer and K. Sigmund. Adaptive dynamics and evolutionary stability. *Appl. Math. Lett.*, 3(4):75–79, 1990.
- [24] J. Istas. *Mathematical modeling for the life sciences*. Universitext. Springer-Verlag, Berlin, 2005.
- [25] N. L. Kaplan, R. R. Hudson, and C. H. Langley. The hitch-hiking effect revisited. *Genetics*, 123:887–899, 1989.
- [26] E. Kisdi. Evolutionary branching under asymmetric competition. *J. Theor. Biol.*, 197:149–162, 1999.
- [27] P. Marrow, R. Law, and C. Cannings. The coevolution of predator-prey interactions: ESSs and red queen dynamics. *Proc. R. Soc. Lond. B*, 250:133–141, 1992.
- [28] J. A. J. Metz, S. A. H. Geritz, G. Meszén, F. J. A. Jacobs, and J. S. van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In *Stochastic and spatial structures of dynamical systems (Amsterdam, 1995)*, Konink. Nederl. Akad. Wetensch. Verh. Afd. Natuurk. Eerste Reeks, 45, pages 183–231. North-Holland, Amsterdam, 1996.
- [29] J. A. J. Metz, R. M. Nisbet, and S. A. H. Geritz. How should we define 'fitness' for general ecological scenarios? *Trends Ecol. Evol.*, 7:198–202, 1992.
- [30] S. T. Rachev. *Probability metrics and the stability of stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1991.
- [31] J. A. Roughgarden. *Theory of population genetics and evolutionary ecology: an introduction*. Macmillan, New York, 1979.
- [32] F. Yu. Stationary distributions of a model of sympatric speciation. *Ann. Appl. Probab.*, 17:840–874, 2007.
- [33] M. L. Zeeman. Hopf bifurcations in competitive three-dimensional Lotka-Volterra systems. *Dynam. Stability Systems*, 8(3):189–217, 1993.