



**HAL**  
open science

# Stochastic models of evolving populations: diversity, spatial structures, cooperative behaviors and mating preferences

Hélène Leman

► **To cite this version:**

Hélène Leman. Stochastic models of evolving populations: diversity, spatial structures, cooperative behaviors and mating preferences. Probability [math.PR]. Ecole Nationale Supérieure de Lyon, 2024. tel-04616743

**HAL Id: tel-04616743**

**<https://hal.science/tel-04616743v1>**

Submitted on 19 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Manuscrit présenté pour l'obtention de L'HABILITATION A DIRIGER DES RECHERCHES

préparée à  
L'ENS DE LYON

**Hélène LEMAN**

Stochastic models of evolving populations: diversity,  
spatial structures, cooperative behaviors and mating  
preferences

**HDR présentée et soutenue à Lyon, le 7 juin 2024.**

*Après avis des rapporteurs :*

LUIS ALMEIDA, (Directeur de recherche, Sorbonne-Université)

VINCENT BANSAYE, (Professeur, Ecole Polytechnique)

ANITA WINTER, (Professor, Universität Duisburg-Essen)

*Jury de soutenance :*

LUIS ALMEIDA, Sorbonne-Université	Rapporteur
VINCENT BANSAYE, Ecole Polytechnique	Rapporteur
NICOLAS CHAMPAGNAT, Inria	Examineur
FLORENCE DÉBARRE, Sorbonne-Université	Examinatrice
EMMANUEL GRENIER, ENS de Lyon	Examineur
ANITA WINTER, Universität Duisburg-Essen	Rapporteuse



## REMERCIEMENTS

Je souhaite tout d'abord remercier mon jury qui a accepté d'être présent pour mon habilitation afin de m'aider à passer cette nouvelle étape dans ma carrière de chercheuse. Je suis tout particulièrement reconnaissante envers mes rapporteurs et rapportrice, Luis, Vincent et Anita, qui ont accepté la tâche fastidieuse de relire mon mémoire, d'autant que je sais qu'il s'agit d'une période de l'année assez chargée. Merci également à Emmanuel, qui nous quitte pour de nouvelles aventures, à Nicolas et Florence, qui ont accepté de voyager pour m'écouter parler de ma recherche !

Je profite également de cette page pour dire merci à ma directrice de thèse, Sylvie, qui m'a donné, il y a maintenant 12 ans, un sujet de recherche à l'interface entre probabilités, analyse et biologie que j'ai tout particulièrement apprécié et qui m'a permise de me lancer dans ces thématiques passionnantes. Merci aussi à Juan Carlos pour ces 2 belles années à Guanajuato et pour la découverte des CSBP. J'en profite pour remercier l'ensemble du CIMAT pour leur accueil dans ce merveilleux pays qu'est le Mexique.

Je souhaite également remercier l'ensemble de mes co-auteurs et co-auteurs, dont Camille, Charline, Manon et Violaine pour nos projets de spéciation et d'excursions, Sepideh pour la découverte des EDP, Nicolas et Laurent pour le C++, Elsa et Régis pour m'avoir initiée au paradoxe de la coopération, Juan Carlos pour les processus mexicains, David, Emmanuel et Gabriel pour l'apo(p)t(hé)ose, Sylvain, Chi et Thomas pour les limites un peu limite, Céline pour le SFS et sa bonne humeur. J'en profite pour remercier mes collaborateurs et collaboratrices actuelles qui ne font pas encore partie "officiellement" de la liste ci-dessus, notamment Aline pour la dormance, Amaury et Hélène pour les limites sans limite, Frédérique et Alice pour l'inactivation du X, Frédéric pour "sa" coopération et surtout Vincent pour son accueil à Lyon même s'il a fui l'ENS pour un grand périple jusqu'à l'ICJ avant mon arrivée !

Merci également aux étudiants que j'ai eu le plaisir d'accompagner et qui ont accepté de me faire confiance, notamment, Thuy, Céline, Josué et Vianney. Merci d'avoir mis de l'animation dans une période sans équipe Inria.

Merci aussi aux membres de l'UMPA, qui m'ont très bien accueillie à mon arrivée à Lyon et avec qui c'est toujours un plaisir de discuter et de perdre un peu de temps ! Merci tout particulièrement à Virginia, Magalie et Jessica qui ont tou-

jours une solution à mes problèmes administratifs et Micaël pour sa bienveillance informatique.

Merci à l'ancienne équipe Numed: Emmanuel, Paul, David et à la future équipe Casting: Pierre<sup>2</sup>, Céline, David, Karène, Sandra, qui va me mener vers de nouvelles aventures scientifiques. Au sein de l'Inria, je souhaite aussi particulièrement remercier Sylvie qui m'a permis de naviguer pendant les temps incertains où je n'avais plus de bateau!

D'une manière générale, merci à toute la communauté et à l'ensemble des chercheurs et chercheuses que j'ai pu croiser dans ces années de recherche et qui m'ont parfois m'influencée.

Pour finir, j'en profite pour remercier toute ma famille, qui est particulièrement bienveillante et chaleureuse, et notamment mes parents. Merci enfin à Pierre, Clément et Agathe, pour leurs sourires et leurs joies de vivre !

# CONTENTS

CONTENTS	5
LIST OF FIGURES	7
INTRODUCTION	9
<b>1 DIVERSITY, SPACE AND COOPERATION</b>	<b>13</b>
1.1 INTRODUCTION . . . . .	13
1.2 SPATIALLY STRUCTURED POPULATION . . . . .	13
1.2.1 Trait substitution sequence . . . . .	15
1.2.2 Diversity . . . . .	16
1.3 COOPERATIVE BEHAVIORS . . . . .	17
1.3.1 Soil bacteria and common good . . . . .	17
1.3.2 Mutualist species . . . . .	22
1.4 PERSPECTIVES . . . . .	27
1.4.1 Adaptive dynamics and limiting processes . . . . .	28
1.4.2 Cooperative behavior . . . . .	28
<b>2 MATING PREFERENCES AND DIVERSITY</b>	<b>29</b>
2.1 INTRODUCTION . . . . .	29
2.2 EMERGENCE OF ASSORTATIVE MATING . . . . .	32
2.3 SPECIATION AND ASSORTATIVE MATING . . . . .	38
2.3.1 Haploid population . . . . .	38
2.3.2 Diploid population . . . . .	46
2.4 DIVERSITY AND DISASSORTATIVE MATING . . . . .	47
2.5 PERSPECTIVES . . . . .	52
<b>3 CSBP WITH COMPETITION IN RANDOM ENVIRONMENT</b>	<b>53</b>
3.1 INTRODUCTION AND MODEL . . . . .	53
3.2 LÉVY ENVIRONMENT . . . . .	55
3.3 LOGISTIC COMPETITION IN A BROWNIAN ENVIRONMENT . . . . .	57

3.3.1	Case associated with a subordinator . . . . .	57
3.3.2	Second case . . . . .	60
3.3.3	General competition for branching diffusion in a Brownian environment . . . . .	62
3.4	RECENT RESULTS AND PERSPECTIVES . . . . .	62
4	GENETIC COMPOSITION OF TUMOR CELL POPULATIONS	63
4.1	INTRODUCTION . . . . .	63
4.2	RESCUED POPULATION . . . . .	64
4.3	PERSPECTIVES . . . . .	70
4.3.1	Convergence in law of the SFS . . . . .	70
4.3.2	Genetic composition of an exponentially growing population . . . . .	70
	BIBLIOGRAPHY	71

# LIST OF FIGURES

1.1	Succession of evolutionary branchings . . . . .	17
1.2	Schemes of the microbe-enzyme models of soil decomposition . . . . .	18
1.3	Patterns of selection on exoenzyme production for different soil diffusion rates. . . . .	21
1.4	Examples of interaction networks for plant-pollinator communities . . . . .	22
1.5	Graphical representation of $g_A$ and $g_P$ . . . . .	26
1.6	Phase plan and nullclines in case of one plant and one pollinator species . . . . .	27
2.1	Illustration of the frequency-dependent migration rates . . . . .	39
2.2	Projections of sets $\mathcal{D}_{5,5}^{\alpha,\alpha'}$ . . . . .	42
2.3	Influence of the migration parameters on the speed of reproductive isolation . . . . .	45
2.4	Mutation sizes and their effects on the disassortative advantage . . . . .	50
2.5	Evolution of the number of alleles maintained in the population . . . . .	51
4.1	Example of site frequency spectrum . . . . .	65
4.2	Empirical and theoretical expectation of the SFS for small value of $i$ . . . . .	66
4.3	Schemes of the SFS computation for large families . . . . .	68
4.4	Empirical expectation of the SFS for large families . . . . .	69





# INTRODUCTION

This manuscript presents a summary of my research work since my thesis. My research is at the interface between mathematics and biology. More precisely, a large part of my studies are based on probabilistic models, called individual-based models (or agent-based models), that describe varying size populations at an individual level. These models can be mathematically described by continuous-time Markov chains or using Poisson point measures. Generally, at a given time, we can compute the next event time (reproduction or death of an individual) by associating two exponential random variables with each individual, whose parameters are the reproduction and death rates of this individual. These rates may depend on any characteristics of the individual (genotype, phenotype, position, species...) and also on the composition of the overall population. The next event then corresponds to the one associated with the smallest outcome among all exponential random variables; and so on, resampling independent random variables for each event.

From a biological point of view these models are particularly interesting since they allow to include a large variety of interactions. For my part, I am particularly interested in understanding the dynamics of these processes in the case of explicit spatial structures, cooperative interactions, and mating preference mechanisms.

To get quantitative results I generally use scaling changes in space and/or time (large populations, rare mutations, long time scales, etc.), based on various biological assumptions. For example, under large population assumptions, the processes generally converge towards deterministic limits, represented by ordinary differential equations or partial differential equations. The study of these limits also brings interesting and new mathematical challenges.

I am thus using both stochastic and deterministic tools with the aim of understanding which evolutionary forces and natural mechanisms can generate and maintain diversity in populations, focusing mainly on the impact of spatial structures, cooperative behaviors, and mating mechanisms.

The first chapter of this manuscript deals with structured populations, including eventually cooperative mechanisms. The main aim of this part is to understand the impact of a spatial structure on the emergence of diversity in populations, or even the

impact on their survival in the case of cooperative individuals. Indeed, cooperators, that release common good in the environment that benefit the entire population, are vulnerable to cheating. Evolutionary forces could then doom the population to extinction. However, nature abounds with examples of cooperative species. Solving this paradox is an interesting challenge. In this chapter, I will present some answers related to the following works:

- H el ene Leman, Sylvie M el eard, and Sepideh Mirrahimi. Influence of a spatial structure on the long time behavior of a competitive Lotka-Volterra type system. *Discrete and Continuous Dynamical Systems-B*, 20(2):469-493, 2014.
- H el ene Leman. Convergence of an infinite dimensional stochastic process to a spatially structured trait substitution sequence. *Stochastics and Partial Differential Equations: Analysis and Computations*, 4:791-826, 2016.
- Elsa Abs, H el ene Leman, and R egis Ferri ere. A multi-scale eco-evolutionary model of cooperation reveals how microbial adaptation influences soil decomposition. *Communications biology*, 3(1):520, 2020.
- Sylvain Billiard, H el ene Leman, Thomas Rey, and Viet Chi Tran. Continuous limits of large plant-pollinator random networks and some applications. *arXiv preprint arXiv:2201.05219*, 2022.

Chapter 2 is devoted to my works related to mating preference mechanisms. The main goal here is to understand the influence of the mating preferences on the evolutionary forces and on the diversity of the populations. This chapter focus on the following works:

- Camille Coron, Manon Costa, H el ene Leman, and Charline Smadi. A stochastic model for speciation by mating preferences. *Journal of mathematical biology*, 76:1421-1463, 2018.
- H el ene Leman. A stochastic model for reproductive isolation under asymmetrical mating preferences. *Bulletin of mathematical biology*, 80(9):2502-2525, 2018.
- Charline Smadi, H el ene Leman, and Violaine Llaurens. Looking for the right mate in diploid species: How does genetic dominance affect the spatial differentiation of a sexual trait? *Journal of Theoretical Biology*, 447:154-170, 2018.
- Camille Coron, Manon Costa, Fabien Laroche, H el ene Leman, and Charline Smadi. Emergence of homogamy in a two-loci stochastic population model. *ALEA*, 18:469-508, 2021.
- Camille Coron, Manon Costa, H el ene Leman, Violaine Llaurens, and Charline Smadi. Origin and persistence of polymorphism in loci targeted by disassortative preference: a general model. *Journal of Mathematical Biology*, 86(1):4, 2023.

In Chapter 3, I present slightly different stochastic models based on continuous state branching processes. These are solution to stochastic differential equations and they represent the dynamics of the size of a population subject to the fluctuations of an external

random environment. In particular, I present conditions for extinction and formulas for the extinction time. The chapter is based on the following articles:

- Hélène Leman and Juan Carlos Pardo. Extinction and coming down from infinity of continuous-state branching processes with competition in a lévy environment. *Journal of Applied Probability*, 58(1):128-139, 2021.
- Hélène Leman and Juan Carlos Pardo. Extinction time of logistic branching processes in a Brownian environment. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 18:1859-1890, 2021

Finally, Chapter 4 is devoted to the study of tumor cell populations under treatment. In particular, I present characterizations of the distributions of neutral mutations found in those populations, that mutate frequently. This chapter corresponds to an opening chapter on works that I have started more recently. The following article will be presented in this chapter:

- Céline Bonnet and Hélène Leman. Site frequency spectrum of a rescued population under rare resistant mutations. *arXiv preprint arXiv:2303.04069*, 2023.

Finally, the following articles are not discussed to keep the manuscript coherent.

- Mireille Bossy, Nicolas Champagnat, Hélène Leman, Sylvain Maire, Laurent Violeau, and Mariette Yvinec. Monte carlo methods for linear and non-linear Poisson-Boltzmann equation. *ESAIM: Proceedings and Surveys*, 48:420-446, 2015.
- Diane Coursier, David Coulette, Hélène Leman, Emmanuel Grenier, and Gabriel Ichim. Live-cell imaging and mathematical analysis of the "community effect" in apoptosis. *Apoptosis*, 28(3-4):326-334, 2023.



# SPATIALLY STRUCTURED POPULATION AND COOPERATIVE BEHAVIOR: PROMOTING DIVERSITY



## 1.1 INTRODUCTION

Studying and understanding the selective mechanisms behind species diversity is a major challenge in evolutionary biology. The theory of evolution, proposed by Darwin [32], explains the emergence of new species from previous species mainly according to three principles: the *heritability* of individual traits, the *variability* of certain traits as a result of mutations or sexual reproduction, and finally the *selection* of individuals that are best able to reproduce in their environment. Diversity can therefore emerge when different traits, or even different species, can be simultaneously selected and thus coexist.

The spatial aspect of the environment plays a crucial role in the appearance and maintenance of species diversity. In particular, it has long been recognized that a heterogeneous environment facilitates the coexistence of species [89, 36]. Although complex by nature, it therefore seems essential to develop spatial models to understand in detail its effects on speciation phenomena, which correspond to the appearance of several species from a single one [47].

This first chapter focuses mainly on my work about spatially structured populations that may be subject to some form of cooperation. My aim is to understand how these spatial structures drive diversity within populations and species and, in particular, how they affect the evolution of the cooperative behaviors.

## 1.2 SPATIALLY STRUCTURED POPULATION

Let us start by introducing the first microscopic spatial model I worked on during my thesis. Since the results presented in this section are also included in my thesis, I will briefly outline them and refer readers to my thesis or to the two related articles [84, 82] for more details. This model describes the evolution of a spatially structured population and was first described by Champagnat and Méléard in [23].

The dynamics of the model follow a diffusive birth-and-death process, where the birth, death and movement of each individual  $i$  depend both on its position  $X_t^i$ , valued in a bounded, convex open set  $\mathcal{X} \subset \mathbb{R}^d$ , and on its phenotypic trait  $U_t^i$ , valued in a compact set  $\mathcal{U} \subset \mathbb{R}^q$ , at time  $t \geq 0$ . We represent the total population at any time  $t \geq 0$  by the following finite measure on the space  $\mathcal{X} \times \mathcal{U}$

$$\nu_t^K = \frac{1}{K} \sum_{i=1}^{N_t} \delta_{(X_t^i, U_t^i)} \in M_F(\mathcal{X} \times \mathcal{U}), \quad (1.2.1)$$

where  $\delta_y$  is the Dirac measure in  $y$ ,  $N_t$  is the number of individuals alive at time  $t$ , the space  $M_F(\mathcal{X} \times \mathcal{U})$  is the set of finite measures on  $\mathcal{X} \times \mathcal{U}$  and  $K$  is a scaling parameter. This parameter  $K$  quantifies the resources available in a habitat and measures the strength of competition between individuals; in other words, it represents the carrying capacity of the environment. We will be interested in taking  $K \rightarrow +\infty$  in the following under particular assumptions.

The phenotypic trait of an individual is fixed throughout its life, unlike its position. Each individual  $i$  of phenotypic trait  $U_t^i = u$  moves on  $\mathcal{X}$  according to the following stochastic differential equation

$$dX_t^i = \sqrt{2m^u} Id \cdot dB_t - n(X_t^i) dl_t, \quad (1.2.2)$$

where  $B$  is a  $\mathbb{R}^d$ -Brownian motion,  $l_t$  is the local time at the  $\mathcal{X}$  boundary and  $n$  is the outgoing normal of the  $\mathcal{X}$  bounded space. This corresponds to a pure diffusion reflected normally at the boundary of  $\mathcal{X}$ . The diffusion coefficient of each individual,  $m^u$ , may depend on its phenotypic trait.

Each individual with characteristics  $(x, u)$  reproduces at rate  $b(x, u)$ . The offspring is created at the position  $x$  of its parent at the time of birth. The trait of the new individual is identical to that of its parent with probability  $1 - q_K$ , and changes according to a kernel  $k(x, u, \cdot)$  with probability  $q_K$ . The parameter  $q_K$  thus normalizes the probability of mutation. We will be interested in rare mutations, i.e.  $q_K \rightarrow 0$ .

Finally, each individual dies at a rate

$$d(x, u) + \frac{1}{K} \sum_{i=1}^{N_t} c(u, X_t^i, U_t^i),$$

which takes into account a natural death rate  $d$  and a competition death rate through the kernel  $c$ .

In [23], the authors studied the behavior of the process in a limit of large population, i.e. when  $K$  becomes large. Since  $K$  scales for the carrying capacity of the environment or equivalently for the amount of resources, the larger  $K$  is, the smaller the strength of competition between two individuals,  $c/K$ , is. Under such scaling and when  $K$  tends to  $\infty$ , it can thus be shown that the dynamics of the stochastic process converges to the one of a deterministic process, at least for finite period of time. In the case of our interest,  $K$  tends to  $\infty$  and  $q_K$  tends to 0 simultaneously. The limiting deterministic process corresponds to the solution to

$$\begin{cases} \partial_t \xi_t(x, u) = m^u \Delta_x \xi_t(x, u) + \left( b(x, u) - d(x, u) - \int_{\mathcal{X}} c(u, y, v) \xi_t(dy, dv) \right) \xi_t(x, u), \\ \partial_n \xi_t(x, u) = 0, \text{ sur } [0, T] \times \partial \mathcal{X} \times \mathcal{U}. \end{cases} \quad (1.2.3)$$

In other words, the solution of this partial differential equation models the averaged behavior of the process under the assumptions of large population ( $K \rightarrow \infty$ ) and rare mutations ( $q_K \rightarrow 0$ ).

We studied this PDE in the case of a monomorphic population (i.e.  $\mathcal{U}$  restricted to  $\{u\}$ ) and the case of a dimorphic population (i.e.  $\mathcal{U}$  restricted to  $\{u, v\}$ ) in [84]. In particular, we were interested in the existence of stationary solutions, convergence to these solutions, and the criteria of coexistence and exclusion of a type in the dimorphic case.

### 1.2.1 Trait substitution sequence

In [84, 82], we were then interested in the process in the context of *adaptive dynamics*. The theory of adaptive dynamics was developed to take into account the fact that the adaptive landscapes of an ecosystem evolve as the individuals in this ecosystem evolve. When a mutant trait appears and becomes permanently established in an environment, the probability of survival and invasion of other traits is modified. If the ability of a mutant individual to establish itself in a given adaptive landscape can be quantified, this quantity is called the *fitness of invasion* of the trait into this landscape. A mutation is considered as advantageous in a particular landscape if its invasion fitness is positive, otherwise, it is considered as deleterious.

The theory of adaptive dynamics takes these phenomena into account under assumptions of large populations, and rare and small mutations, so that demographic and evolutionary time scales are separated. The population resulting from a mutant individual reaches demographic equilibrium before the occurrence of any new mutation. In this setting, Metz et al. [96] introduced the *Trait Substitution Sequence* process, which details successive fixations of advantageous mutations in a population using a process of jumps. This was then extensively studied from a mathematical point of view, see [19, 21, 30]. among others.

Here, we presented the case of spatially structured population and extended these results to this infinite-dimensional case. More precisely, under well chosen assumptions, which allow to maintain a monomorphic population at all times, we proved the following result of convergence on the measure  $(\nu_t^K)_{t \in [0, T]}$  that represents the spatially structured population.

**Theorem 1.2.1.** *Assume that*

$$\log(K) \ll \frac{1}{Kq_K} \ll e^{KV}, \text{ for all } V > 0, \text{ when } K \rightarrow \infty. \quad (1.2.4)$$

For all  $T > 0$ , the rescaled measure  $\left(\nu_t^K\right)_{t \in [0, T]}$  converges to a Markov process  $(\Lambda_t)_{t \geq [0, T]}$  when  $K \rightarrow +\infty$ . For all  $t$ ,  $\Lambda_t$  belongs to the space  $\{\bar{\xi}^u \delta_u, u \in \mathcal{U}\} \subset M_F(\mathcal{X} \times \mathcal{U})$ , where  $\bar{\xi}^u$  corresponds to the unique stationary distribution solution to (1.2.3) when considering that  $\mathcal{U}$  is restricted to  $\{u\}$ . The jump process  $(\Lambda_t)_{t \geq 0}$  jumps from equilibrium  $\bar{\xi}^u \delta_u$  to equilibrium  $\bar{\xi}^v \delta_v$  at the infinitesimal rate

$$\int_{\mathcal{X}} b(x, u) \phi^{vu}(x) k(x, u, v) \bar{\xi}^u(dx) dv,$$



where  $\phi^{vu}(x)$  corresponds to the probability that the offspring of an individual of trait  $v$ , which appeared at position  $x$  in a landscape made up entirely of  $u$  individuals, replaces this resident  $u$ -population (explicit formulation can be found in [82]).

The convergence holds in the sense of convergence of the infinite-dimensional distributions.

Assumption (1.2.4), introduced by Champagnat [19], guarantees that the demographic time scale and the evolutionary time scale are separate.

This theorem indicates that the limit process is a process of jumps through the set of equilibrium states associated with each trait. When a favorable mutant appears and invades, the demographic dynamics of the invasion phase are rapid. The limiting process thus jumps instantaneously to the equilibrium state characterized by the trait of this mutant.

**Ideas of proof.** The proof relies on the study of three phases in the population dynamics trajectories (mutant survival or extinction, mean-field phase, and resident extinction), as those detailed by Champagnat in [19].

The second phase (mean-field phase) consists in studying the process in large population in the case where resident and mutant are both present. In other words, it relies on the fine study of the deterministic differential equation (1.2.3) in the case of a monomorphic population (i.e.  $\mathcal{U} = \{u\}$ ) and the case of a dimorphic population (i.e.  $\mathcal{U} = \{u, v\}$ ). These results were presented in [84].

The first and third phases imply a fine study of the stochastic process. They first consist in verifying that the dynamics of the resident population remains close to a stationary state throughout the duration of the invasion (for the first phase) or the extinction (for the third phase) of the mutant. This is based in particular on results of large deviations. Secondly, it consists in studying the dynamics of the mutant itself, by approximating it with the dynamics of a branching Brownian motion living on a compact space. These results are exhibited in [82].

### 1.2.2 Diversity

Theorem 1.2.1 includes a strong assumption that drastically reduces the phenotypic diversity that could be obtained. Deleting this assumption, we conjecture that we could obtain a result similar to the one of Champagnat and Méléard in [24], i.e. the convergence towards some jump process representing a population of a finite number of traits. The main difficulty lies in the study of the resulting system of PDE that need to be investigated.

Simulations of the process described in the previous section under general assumptions (i.e. that do not satisfy those of Theorem 1.2.1) can, in some cases, produce a wide range of phenotypic traits. In the simulations shown in Figure 1.1, we have also added an assumption of small mutation amplitude. These simulations echo those of the seminal paper of Dieckmann-Doebeli [34], who obtained a very similar diversity pattern imitating speciation phenomena. It is essential to study this type of patterns to understand precisely the mechanism at work in natural diversity. I also refer the reader to chapter 3 of my thesis for further details.

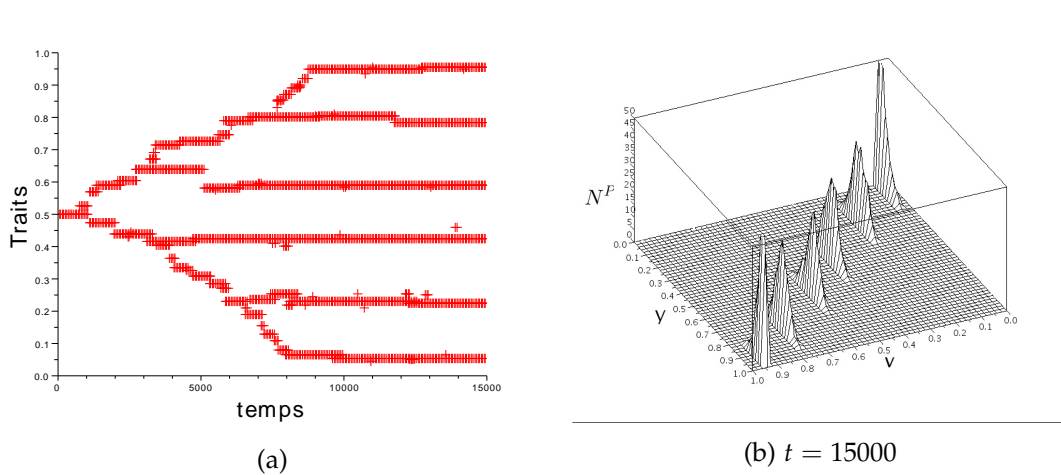


Figure 1.1 – (a) Traits of individuals along time. The initial population is monomorphic with trait  $v = 0.5$ . The population remains monomorphic for a time, then a succession of evolutionary branchings is observed, leading to polymorphism, i.e. the coexistence of several populations with different traits. (b) Density of the population on the space  $\mathcal{X} \times \mathcal{U}$  at the final time  $t = 15000$ .

In [20], Champagnat and Henry develop a method that can be applied to study the influence of the parameters on the final diversity obtained under these assumptions of the adaptive dynamics theory.

### 1.3 COOPERATIVE BEHAVIORS

In chapter 3 of my thesis, which corresponds to an opening chapter (not published), I looked at an extension of the previous model by considering a population representing a community of plants and pollinators with mutualistic-type interactions between species. This opened up my interest in the study of structured models incorporating positive interactions (mutualistic or cooperative). This is the subject of the next two sections, linked to two papers [1, 12] written in collaboration with Elsa Abs and Régis Ferrière for the first and Sylvain Billiard, Thomas Rey and Chi Tran for the second.

#### 1.3.1 Soil bacteria and common good

Soil micro-organisms play a crucial role in the terrestrial carbon cycle, decomposing soil organic matter and gradually transforming it into humus. This decomposition is actually a by-product of the metabolic activity of micro-organisms. They produce extracellular enzymes (exoenzymes) that convert complex compounds into smaller ones, that they need to assimilate for their growth and maintenance. By doing so, they face a "public good dilemma". Indeed, enzymes and their products are released outside the cell, making them accessible to all micro-organisms present locally. Micro-organisms that invest less energy in enzyme

production, known as "cheaters", can therefore take advantage of this external production. From an evolutionary point of view, we therefore expect cheaters to eventually invade the environment and enzyme production to decline accordingly, gradually leading to a vortex of extinction if final cheaters no longer produce enough enzymes to survive.

However, the microbial world provides numerous examples of communities based on the production of common goods that survive [51, 105, 81]. Therefore, conditions must exist that favor the evolution of exoenzyme production. Numerous studies have shown, for example, the importance of space and its heterogeneity [108, 3]. This understanding of the evolutionary stability of diffusive public goods in general, and in the case of degradative enzyme production in particular, remains incomplete. One limitation of previous models is their focus on two-way competition between two strains, typically a producing strain and a non-producing or "pure cheater" strain.

In this section, we set out to build a spatial model based on this enzyme production mechanism in order to gain a detailed understanding of how producers can resist invasion by strains that invest only slightly less into the common good. In other words, we adopt some assumptions of adaptive dynamics theory, assuming that mutations are rare and of small amplitude.

**Non-spatial models.** To this aim, we first developed a non-spatial individual-based model described by Figure 1.2(a), where processes act locally at the level of individual entities. Then, we derived a simplified hybrid stochastic-deterministic model by using appropri-

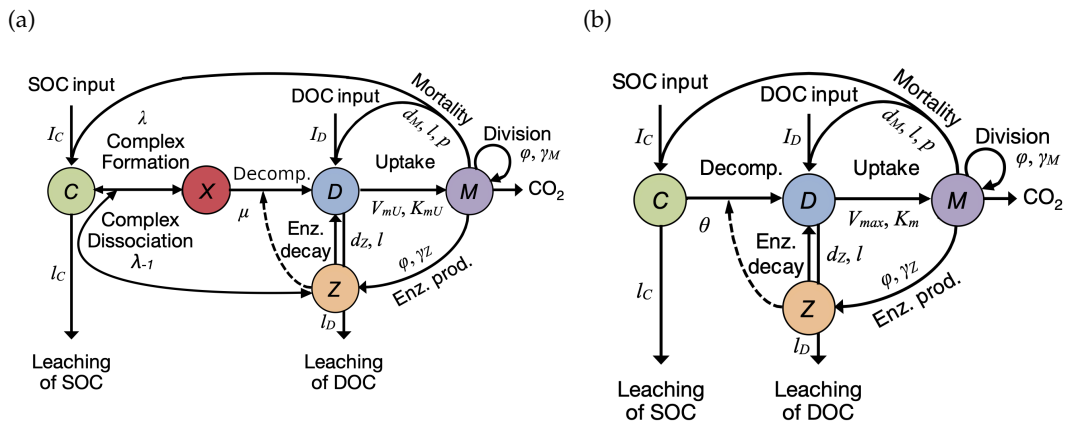


Figure 1.2 – Microbe-enzyme driven decomposition of soil organic matter: agents and processes. (a), Five-compartment model, with soil organic carbon (SOC, C), dissolved organic carbon (DOC, D), microbial cells (M), enzymes (Z) and SOC-enzyme complexes (X). (b), Four-compartment model, without the SOC-enzyme complexes (X). Plain arrows indicate carbon fluxes among compartments and in and out of the system. Dotted arrows indicate the exoenzyme concentration dependence of the decomposition rate.

ate scaling based on relevant biological assumptions, namely

1. dissociation and decomposition of the SOC-enzyme complexes ( $X$ ) are much faster than its formation;
2. a microbial cell ( $M$ ) is of the order of  $10^7$  to  $10^{10}$  times larger (in units of carbon mass) than one enzyme, one substrate (SOC) molecule, or one product (DOC) molecule; and in a given volume, the number of cells is between  $10^{-5}$  and  $10^{-10}$  times smaller than the number of molecules of SOC, DOC or enzyme;
3. Finally, the growth of individual cells ( $M$ ) is driven by events (resource uptake) that occur on the same timescale as the events affecting SOC, DOC, and enzymes.

All these assumptions and simplifications lead to a model described by Figure 1.2(b). In this model, the number of bacteria cells remains finite, and only the death events of bacteria cells remain stochastic at rate  $d_M$ . Between deaths, the amount  $S_i$  of biomass stored within cell  $i$  is governed by

$$\frac{dS_i(t)}{dt} = (1 - \varphi)\gamma_M V_{max} \frac{d(t)}{K_m + d(t)} \omega_M, \quad (1.3.1)$$

where all parameters are positive and  $d(t)$  denotes the biomass of DOC molecules at time  $t$ . Once  $S_i$  reaches the threshold  $\omega_M$ , the cell divides and both mother and daughter cells' reserve is set back to 0. Finally for a given number of cells,  $M$ , the change in enzyme ( $z(t)$ ), SOC ( $c(t)$ ) and DOC ( $d(t)$ ) are governed by

$$\begin{cases} \dot{c}(t) = I_C - I_C c(t) - \theta z(t) c(t) \\ \dot{d}(t) = I_D - I_D d(t) + \theta z(t) c(t) + (1 - l) d_{ZZ}(t) - V_{max} \frac{d(t)}{K_m + d(t)} \omega_M M \\ \dot{z}(t) = \varphi \gamma_Z V_{max} \frac{d(t)}{K_m + d(t)} \omega_M M - d_{ZZ}(t), \end{cases} \quad (1.3.2)$$

Notice in particular that the parameter  $\varphi$  models the trade-off between the investment in exoenzyme production and the one in cell growth.

A more recent publication [8] could lead us to further refine the microscopic model by detailing the consumption/seeking/enzyme-production stages of bacteria in order to rigorously derive the Mikaelis-Menten functions used in (1.3.1) and (1.3.2) to describe the consumption of DOC by the micro-organisms, which was not done in the publication [1] described in this section.

Studying numerically this non-spatial model, we observe that any initial level of microbial cooperation will be gradually eroded by the process of mutation-selection, driving the population towards a trait value at which extinction occurs. In finite populations, mutant success or failure becomes probabilistic. Due to random genetic drift, cheater phenotypes may fail to invade, and cooperator mutants may occasionally go to fixation. Long-term adaptive dynamics driven by rare mutation and selection in finite populations have been studied in a general framework by Champagnat and Lambert in [21]. They showed that the evolutionary trait dynamics can be described mathematically as a diffusion process whereby

a Brownian motion (white noise) is added to a trend driven by the deterministic selection gradient. Adapting to our case, it means that our non-spatial population is doomed to extinction even if fluctuations may change the speed of it.

**Spatial model.** We then extended the model to a spatial lattice. The model is then described from an algorithm point of view. The space is represented as a square regular lattice of microsites. In each microsite, the previous non-spatial model is running and the microsites are coupled through the diffusion of DOC and the dispersal of newborn cells. The diffusion of DOC corresponds to a discretization on the lattice of the following diffusion equation

$$\frac{d}{dt}d(\mathbf{x}, t) = \sigma_{\text{diff}}\Delta d(\mathbf{x}, t).$$

The dispersal of bacteria cells occurs at birth with probability  $p_{\text{disp}}$ . If one or more of the adjacent microsites is empty, the dispersing cell moves to one of them with equal probability. If all neighboring microsites are occupied, there is a probability  $p_{\text{open}}$  that a micro-disturbance of the soil strikes and opens one of the microsite by killing the local cell population therein, which then becomes occupied by the dispersing cell; otherwise, it stays in the microsite of its mother cell. These micro-disturbances model environmental stochasticity as done in [56, 87]. Moreover, with dispersal tied to micro-disturbances, resident and mutant strains do not mix within microsites. This is essential for the evolutionary stability of exoenzyme production. If resident and mutant strains were mixing within the microsites, any slightly cheating mutant would always invade and spread across the lattice, as for the non-spatial model.

We performed simulations of the process on pairwise resident-mutant competing strains. Precisely, our initial condition consists in a resident strain at equilibrium in each microsite except for 5% of the microsites at the center of the grid where there is a mutant strain characterized by a slightly different value of the parameter  $\varphi$ . We let the process evolve during a time  $T_{\text{max}} = 10^6$  and recorded the final mutant population size. We then computed the fitness of invasion corresponding to the average long-term growth (see Figure 1.3 for more details.). Repeating these simulations for different competing resident-mutant pairs and different values of the diffusion parameter  $\sigma_{\text{diff}}$ , we were able to conclude that

- ◇ spatial segregation of strains at microscale can promote the persistence of exoenzyme producers against invasion by cheaters (negative selection against cheating); and favor invasion by even stronger exoenzyme producers (positive selection for cooperation),
- ◇ and soil diffusivity is a key factor of the evolutionary stability of microbial decomposition. In particular, a low level of resource diffusion helps producers to sustain themselves.

**Ideas of proof.** Proofs only concern the simplifications of the microscopic model described previously in points 1 and 2. For point 1, the proof consisted in simplifying the complete 5-compartment individual-based model into a 4-compartment model where the SOC-enzyme complexes ( $X$ ) no longer appear. Indeed under the limits described in this

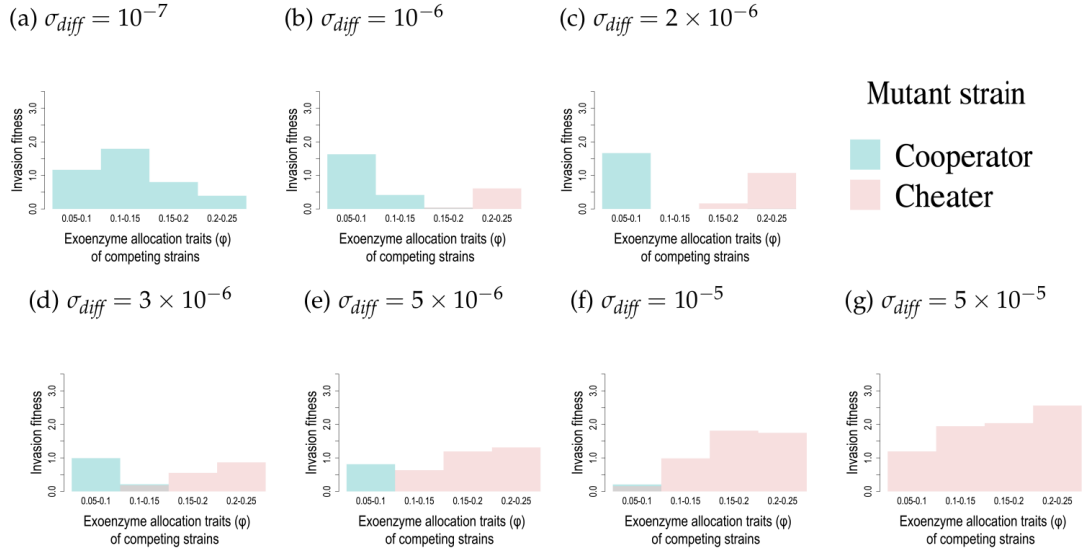


Figure 1.3 – Patterns of selection on exoenzyme production for different soil diffusion rates. Each graph shows the mutant invasion fitness across pairwise resident-mutant competing strains. Invasion fitness is measured as the product of the mutant survival probability and the average long-term growth rate of mutant populations. More precisely, the survival probability is estimated as the fraction of simulations with a non-extinct mutant population at  $T_{max}$ . The long-term growth is calculated as the average of  $(1/T_{max}) \log \frac{\text{final mutant population size}}{\text{initial mutant population size}}$  among the 20 runs for each pairwise competition test.  $T_{max} = 10^6$ . Red bars show invasion fitness of the cheater strain taken as mutant (with the lower  $\varphi$  value in the competing pair); blue bars show invasion fitness of the cooperator strain taken as mutant (with the higher  $\varphi$  value in the competing pair). Positive invasion fitness of cheater mutants (red bars) indicate selection against exoenzyme production. Positive invasion fitness of cooperator mutants (blue bars) indicate selection in favor of exoenzyme production. All constant parameters are set using values found in literature (see table 1 in [1] for more details). Mutant initial population size is set to 5% of the abundance of the resident population in the central microsites. We tested values of  $\sigma_{diff}$  between  $10^{-8}$  and  $10^{-4}$  and report results for  $\sigma_{diff}$  between  $10^{-7}$  and  $5 \times 10^{-5}$  as variation of  $\sigma_{diff}$  outside this range had no effect.

point, the species  $X$  is valued at 0 at almost all times. The main difficulty arises from the unbounded nature of some rates, and thus classical theorems do not apply directly. We had to precisely prove the tightness and identify the limit to conclude.

The limit described in point 2 fall within the assumptions of the theorems developed by Crudu et al. [31] and therefore a straightforward adaptation of these results was sufficient. Indeed, it corresponds to prove the convergence of a 4-compartments model to a piecewise deterministic Markov process (PDMP) where births and deaths of micro-organisms are still stochastic whereas the dynamics of SOC, DOC and enzymes are deterministic ones.

### 1.3.2 Mutualist species

This section is devoted to another type of mutualism corresponding to positive interactions between plant and pollinator communities.

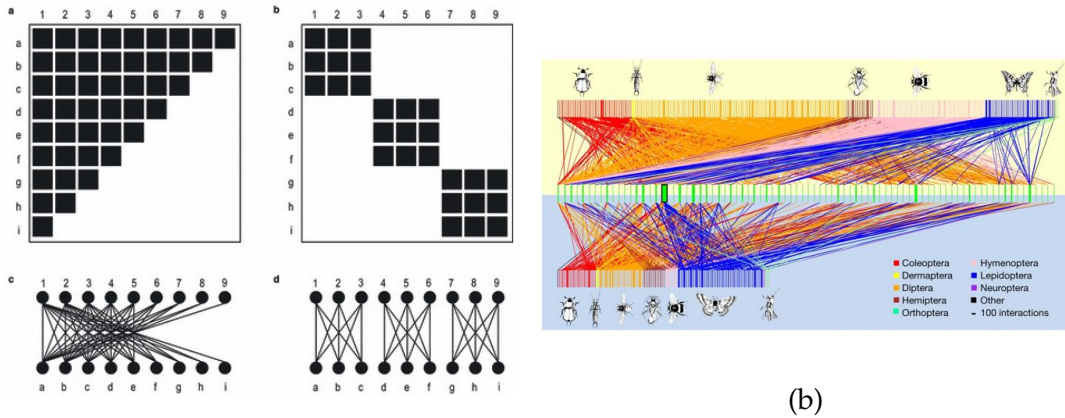
In this section, we present scaling limits of a model representing a community of plants and pollinators issued from [12]. Each plant (resp. pollinator) species is characterized by a trait  $x \in [0, 1]$  (resp.  $y \in [0, 1]$ ). This trait represents for example the degree of generalism of the species: a species is said to be generalist if it interacts with a large number of other species otherwise it said to be specialist.

Assuming a community of  $n$  plants and  $m$  pollinators species, we denote by  $P_t^{K,i}$  and  $A_t^{K,j}$  the size of the plant and pollinator species  $i$  and  $j$  at time  $t$ . The total plant and pollinator populations at time  $t$  can thus be represented by the following point measures:

$$\mathbf{P}_t^{K,n,m}(dx) = \frac{1}{nK} \sum_{i=1}^n P_t^{K,i} \delta_{x^i}(dx), \quad \mathbf{A}_t^{K,n,m}(dy) = \frac{1}{mK} \sum_{j=1}^m A_t^{K,j} \delta_{y^j}(dy). \quad (1.3.3)$$

As in Section 1.2,  $K > 0$  is a scaling parameter called the carrying capacity. It is a measure of the size of the system and it controls the abundance of the whole community that can be sustained by the environment.

The interactions between plants and pollinators are characterized by an exchange of resources [60]. For all  $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$ , we define  $G_{ij}^{n,m}$  as a Bernoulli random variable with parameter  $\phi(x^i, y^j)$ , where  $\phi$  is a continuous function on  $[0, 1]^2$  and such that  $(G_{ij}^{n,m})_{i,j}$  are independent. Individuals of the plant species  $i$  can interact with individuals of the pollinator species  $j$  if  $G_{ij}^{n,m} = 1$ . In other words, the adjacency matrix  $G^{n,m}$  models the ability for individuals to interact. Examples of (non-random) adjacency matrices and pollination network are drawn in Figure 1.4, and more detailed examples can be found in the article linked with this section [12].



(a)

(b)

Figure 1.4 – (a): Nested (left) or modular (right) bipartite networks, from Fontaine et al. [41]. (b) Pollination network for diurnal and nocturnal insect species, from Knop et al. [71].

If species  $i$  and  $j$  interact, we denote by  $c_{ij}^{n,m}$  the weight of the interaction.  $c_{ij}^{n,m}$  is a random variable with bounded variance and such that there exists a continuous function  $c^{n,m}$  satisfying

$$\mathbb{E} \left[ c_{ij}^{n,m} | (x^i, y^j)_{(i,j)} \right] = c^{n,m}(x^i, y^j).$$

It describes the intensity and frequency of the relation. From the point of view of the plant,  $c_{ij}^{n,m}$  can be interpreted as a measure of the pollination services received from the pollinators. From the point of view of the pollinators,  $c_{ij}^{n,m}$  measures the quantity and quality of nutrients collected from the plants.

The quantity of resources  $R$  gained by plants or pollinators is then modeled through a "mass-action model". At time  $t$ , a single individual of the plant species  $i$  interacting with pollinator species  $j$  is supposed to gain a quantity of resources proportional to the abundance of pollinators  $A_t^{K,j}/K$  weighted by the interaction efficiency  $c_{ij}^{n,m}$ , such that the total resource gained by a plant individual of species  $i$  through the pollination interactions is

$$R_t^{A,K,i} := \sum_{j=1}^m G_{ij}^{n,m} c_{ij}^{n,m} \frac{A_t^{K,j}}{K}.$$

Similarly, for a given pollinator of the species  $j$ , the resources gained from the interaction with the plants species is

$$R_t^{P,K,j} = \sum_{i=1}^n G_{ij}^{n,m} c_{ij}^{n,m} \frac{P_t^{K,i}}{K}.$$

Finally, the dynamics of the plant and pollinator populations is described as in Section 1.2 at the scale of the individual by an individual based model. We denote  $b^P(R)$  and  $b^A(R)$  the individual birth rate of plant and pollinator species, respectively. Each of them depends on the quantity of resources exchanged, i.e.  $R^P$  or  $R^A$ . Similarly, we denote  $d^P(R)$  and  $d^A(R)$  the individual death rates.

The plants and pollinators dynamics are also assumed to be affected by logistic competition among plants and among pollinators (within and between species competition). Hence, a plant with trait  $x \in [0, 1]$  suffers an additional death rate term due to competition which is

$$k \star \mathbf{P}_t^{K,n,m}(x) := \int_{[0,1]} k(x, x') d\mathbf{P}_t^{K,n,m}(dx') = \frac{1}{nK} \sum_{i=1}^n k(x, x^i) P_t^{K,i},$$

where  $k(x, x')$  is a continuous function and quantifies the competition pressure exerted by another plant of trait  $x'$ . Similarly, a pollinator with trait  $y' \in [0, 1]$  suffers an additional death rate due to competition with pollinators of trait  $y$  corresponding to  $h \star \mathbf{A}_t^{K,n,m}(y)$ , with  $h(y, y')$  a continuous function.

**Large number of individuals.** Following the work of Fournier and Méléard [44], it is possible to describe the previous model using Poisson point Measures. Under suitable assumptions (boundedness of the rates and converge of the initial condition) and when the



species abundances tend to infinity ( $K \rightarrow \infty$ ), the stochastic dynamics is well approximated by the following continuous dynamical system, during any finite period of time,

$$\begin{aligned} \forall 1 \leq i \leq n, \quad \frac{d\tilde{P}_t^i}{dt} &= \left( (b^P - d^P) \left( \sum_{j=1}^m G_{ij}^{n,m} c_{ij}^{n,m} \tilde{A}_t^j \right) - \frac{1}{n} \sum_{\ell=1}^n k(x^i, x^\ell) \tilde{P}_t^\ell \right) \tilde{P}_t^i \\ \forall 1 \leq j \leq m, \quad \frac{d\tilde{A}_t^j}{dt} &= \left( (b^A - d^A) \left( \sum_{i=1}^n G_{ij}^{n,m} c_{ij}^{n,m} \tilde{P}_t^i \right) - \frac{1}{m} \sum_{\ell=1}^m h(y^j, y^\ell) \tilde{A}_t^\ell \right) \tilde{A}_t^j. \end{aligned} \quad (1.3.4)$$

More precisely, the following proposition can be proved.

**Proposition 1.3.1.** *We consider a sequence  $(\mathbf{P}^{K,n,m}, \mathbf{A}^{K,n,m})_{K \in \mathbb{N}}$  of processes as in Definition 1.3.3, such that the initial conditions have a bounded third moment and converge almost surely. Then, for all  $T \geq 0$ ,*

$$\lim_{K \rightarrow \infty} \sup_{t \leq T} \sup_{i,j} \left\{ \left| \frac{P_t^{K,i}}{K} - \tilde{P}_t^i \right|, \left| \frac{A_t^{K,j}}{K} - \tilde{A}_t^j \right| \right\} = 0, \text{ a.s.}, \quad (1.3.5)$$

where  $(\tilde{P}_t^1, \dots, \tilde{P}_t^n, \tilde{A}_t^1, \dots, \tilde{A}_t^m)_{t \geq 0}$  is the unique solution of the system (1.3.4).

Equation (1.3.4) is similar to a classical Lotka-Volterra system applied to mutualistic interactions with competition. The functions  $b^P, b^A, d^P, d^A$  can take any form. As a consequence, this system of equations can capture a large variety of ecological situations. In particular, many ODE models published in the ecological literature are special cases of it (see [118, 11, 88]).

Proposition 1.3.1 corresponds to a law of large number and gives some average behavior of the process when there are a lot of individuals in the system. Similarly, a central limit theorem can be established in order to quantify the speed at which the convergence (1.3.5) holds, and to describe the fluctuations around the limit when  $K$  is large but not infinite (see Proposition 2.3 in [12]).

**Large number of species.** We then considered that the numbers of plant and pollinator species in the network tend to infinity. In other words, our aim was to obtain the equations describing the evolution of the population when  $n, m \rightarrow +\infty$ .

To this aim, we considered that the traits of plants and pollinators were chosen according to i.i.d. random variables, i.e.  $(x^i)_{i \in \{1, \dots, n\}}$  and  $(y^j)_{j \in \{1, \dots, m\}}$  are the ordered sequences of i.i.d. random variables. Moreover, we assumed that there exists a function  $c : [0, 1]^2 \mapsto \mathbb{R}$  to which the sequences of functions  $nc^{n,m}(\cdot, \cdot)$  and  $mc^{n,m}(\cdot, \cdot)$  converge uniformly, in  $L^\infty([0, 1]^2)$ . The idea in this assumption is that the function  $c(\cdot, \cdot)$  is a ‘‘harvesting function’’ underlying the matrix  $(c_{ij}^{n,m})_{i \in \llbracket 1, n \rrbracket, j \in \llbracket 1, m \rrbracket}$ .

Notice also that this assumption implies that  $n$  and  $m$  grow to infinity with a similar speed, i.e. there exists a sequence  $(\alpha_n)_{n \geq 1}$  such that

$$m = \alpha_n n \quad \text{and} \quad \lim_{n \rightarrow \infty} \alpha_n = 1. \quad (1.3.6)$$

Under these assumptions, we proved the tightness of the process.

**Proposition 1.3.2.** *If there exist deterministic continuous bounded densities  $\bar{p}_0$  and  $\bar{a}_0$  such that the following weak convergences hold:*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n \tilde{P}_0^i \delta_{x_i} \stackrel{w}{=} \bar{p}_0(x) dx, \quad \lim_{m \rightarrow +\infty} \frac{1}{m} \sum_{j=1}^m \tilde{A}_0^j \delta_{y_j} \stackrel{w}{=} \bar{a}_0(y) dy \quad a.s. \quad (1.3.7)$$

For any  $T \geq 0$ , and for  $n, m \rightarrow +\infty$ , the sequence of measure-valued processes

$$\left( \tilde{\mathbf{P}}_t^{n,m}(dx), \tilde{\mathbf{A}}_t^{n,m}(dy) \right)_{t \geq 0} := \left( \frac{1}{n} \sum_{i=1}^n \tilde{P}_t^i \delta_{x_i}, \frac{1}{m} \sum_{j=1}^m \tilde{A}_t^j \delta_{y_j} \right)_{t \geq 0}, \quad n, m \geq 1$$

are tight in  $\mathcal{C}([0, T], \mathcal{M}_F([0, 1]^2))$ , where  $\mathcal{M}_F^2([0, 1])$  is endowed with its weak topology.

Moreover, any limiting values admit densities with respect to the Lebesgue measure on  $[0, 1]$ .

We finally conjectured that there is only one limiting value which corresponds to the unique deterministic solution to, for any  $f \in \mathcal{C}([0, 1], \mathbb{R})$ ,

$$\begin{aligned} \int_0^1 f(x) d\tilde{\mathbf{P}}_t(x) &= \int_0^1 f(x) \bar{p}_0(x) dx \\ &\quad + \int_0^t \int_0^1 f(x) \left[ g^P \left( \int_0^1 c(x, y) \phi(x, y) \bar{a}_s(y) dy \right) - k \star \bar{p}_s(x) \right] \bar{p}_s(x) dx ds, \\ \int_0^1 f(y) d\tilde{\mathbf{A}}_t(y) &= \int_0^1 f(y) \bar{a}_0(y) dy \\ &\quad + \int_0^t \int_0^1 f(y) \left[ g^A \left( \int_0^1 c(x, y) \phi(x, y) \bar{p}_s(x) dx \right) - h \star \bar{a}_s(y) \right] \bar{a}_s(y) dy ds, \end{aligned} \quad (1.3.8)$$

where  $k \star v$  denotes in this case the convolution on two functions.

This remains as a conjecture for the moment since the identification of the limit is not proven.

Equation (1.3.8) is analogous to the ODE system given in Equation (1.3.4). However, here species are not considered as discrete but continuously distributed along a continuous trait. The connections are modeled by  $\phi$ . This function  $\phi : [0, 1]^2 \mapsto [0, 1]$  is a graphon (see [92]): it can be understood (in this case) as a graph on node sets  $[0, 1]$  for the plants and  $[0, 1]$  for the pollinators, where  $\phi(x, y)$  describes the density of connections between plants  $x$  and pollinator  $y$ . The term  $c\phi$  reflects both the topology ( $\phi$ ) and the intensity ( $c$ ) of plant-pollinator interactions throughout the community depending on the traits values  $x$  and  $y$  involved.

**Example of a nested graph of interactions.** We gave results on the limiting dynamics for the following particular forms of parameters, which are similar to those found in [59] or [88]:

$$\begin{aligned} g^P(R) &:= b^P(R) - d^P(R) = \frac{\alpha_P R}{\beta_P + \gamma_P R} - (d_P + \delta_P R), \\ g^A(R) &:= b^A(R) - d^A(R) = \frac{\alpha_A R}{\beta_A + \gamma_A R} - d_A, \end{aligned}$$

where all parameters are assumed to be positive and such that  $g^P$  and  $g^A$  are positive at least at one point (see graphical representations of  $g^P$  and  $g^A$  in Figure 1.5). The growth rate

$g^P$  reflects an interaction trade-off for the plant, indeed it is supposed that there is a cost for interacting with pollinators due to nectar production, leaves consumption, etc. On contrary, it is assumed that pollinators always increase their benefits when interacting with plants.

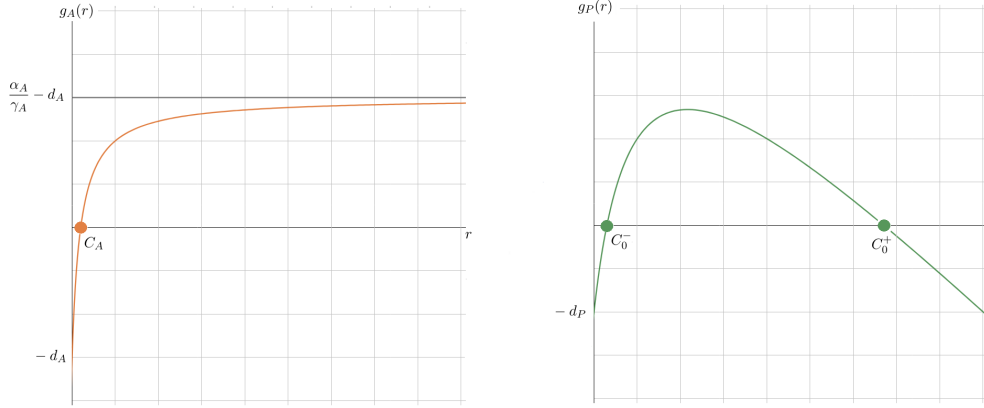


Figure 1.5 – Graphical representation of  $g_A$  and  $g_P$

We also assumed that the competition kernels  $k$  and  $h$  are constant functions, i.e. all species compete with each other with the same intensity, representing for example species that live on the same location, or that hatch and live at the same time of the year in the case of annual species.

We first discussed the case of the ODE (1.3.4) with  $n = m = 1$ . Mainly, depending on the parameter values, there are two possibilities which can be summarized in the two phase planes of Figure 1.6. It can be proven that either there is only the null equilibrium and all trajectories converge to it, or there is a attractive and a repulsive positive equilibrium in addition to the null equilibrium.

Our last results concerned the behavior of the kinetic equations (1.3.8).

We stated results for the particular case where for all  $x_0, y_0 \in [0, 1]$ ,  $y \mapsto \psi(x_0, y)$  and  $x \mapsto \psi(x, y_0)$  are increasing and continuous functions. From a biological point of view, this representing a community with nested interactions as those presented in Figure 1.4(a, left). Under these assumptions, we proved the following result.

**Proposition 1.3.3.** *System (1.3.8) does not admit non-null stationary state with densities w.r.t Lebesgue measure.*

Moreover, any non-null stationary state in  $L^1([0, 1]^2)$  is a couple of measures  $(\bar{P}^\infty, \bar{A}^\infty)$  such that

$$\begin{aligned} & \exists \bar{a}_0, \bar{p}_1 \in \mathbb{R}_+^*, \bar{p}_2 \in \mathbb{R}_+, \bar{x}_1, \bar{x}_2, \bar{y}_0 \in [0, 1], \begin{cases} \bar{P}^\infty = \bar{p}_1 \delta_{\bar{x}_1} + \bar{p}_2 \delta_{\bar{x}_2} \\ \bar{A}^\infty = \bar{a}_0 \delta_{\bar{y}_0} \end{cases} \\ & \text{with } \begin{cases} g^P(\bar{a}_0 \psi(\bar{x}_1, \bar{y}_0)) = g^P(\bar{a}_0 \psi(\bar{x}_2, \bar{y}_0)) = k(\bar{p}_1 + \bar{p}_2) \\ g^A(\bar{p}_1 \psi(\bar{x}_1, \bar{y}_0) + \bar{p}_2 \psi(\bar{x}_2, \bar{y}_0)) = h \bar{a}_0. \end{cases} \end{aligned} \quad (1.3.9)$$

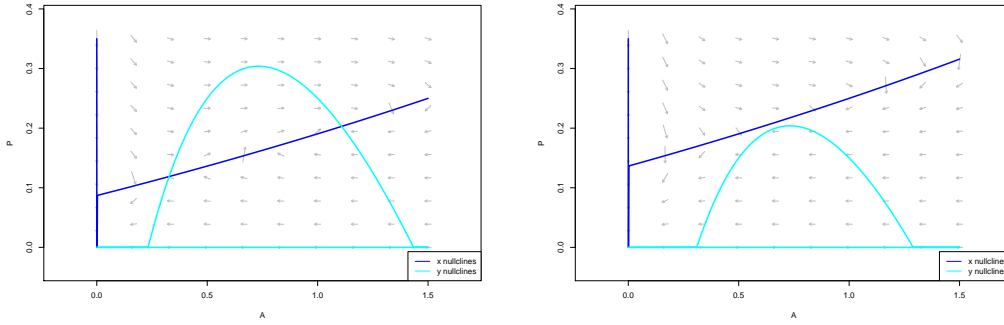


Figure 1.6 – Phase plan and nullclines of the system of ODE (1.3.4): nullclines for the pollinator dynamics in blue; nullclines for the plant dynamics in cyan. Right:  $d_A = 2$  and  $d_P = 1$ , the system has 3 stationary states: the null equilibrium, 1 stable positive equilibrium and 1 unstable positive equilibrium. Left:  $d_A = 3$  and  $d_P = 1.2$ , the unique equilibrium of the system is the null equilibrium. The other parameters are set to  $\alpha_A = 25$ ,  $\alpha_P = 9$ ,  $\beta_A = \beta_P = \gamma_A = \gamma_P = 1$  and  $\delta_P = 3$ .

All these stationary states are unstable, except the state

$$\begin{cases} \bar{p}^\infty = \frac{\max_{\mathbb{R}^+} g^P}{k} \delta_{x_0} \\ \bar{A}^\infty = \frac{\arg \max_{\mathbb{R}^+} g^P}{\psi(x_0, 1)} \delta_1 \end{cases} \quad (1.3.10)$$

if  $x_0$ , solution to  $g^A \left( \frac{\max_{\mathbb{R}^+} g^P}{k} \psi(x_0, 1) \right) \psi(x_0, 1) = h \cdot \arg \max_{\mathbb{R}^+} g^P$ , exists and is unique.

Finally, assuming that, for all initial conditions with positive densities w.r.t Lebesgue measure, the quantities  $\int_0^1 \psi(x, \cdot) \bar{p}_t(x) dx$ ,  $\int_0^1 \bar{p}_t(x) dx$ ,  $\int_0^1 \psi(\cdot, y) \bar{a}_t(y) dy$  and  $\int_0^1 \bar{a}_t(y) dy$  converge when  $t$  grows to infinity, then the trajectory converges to equilibrium (1.3.10).

This result shows that when the plant-pollinator network is nested and the competition among plants and among pollinators is constant, then the plant-pollinator community collapses to a single plant-pollinator species pair.

The proof of Proposition 1.3.3 is not restricted to the specified forms of  $g^P$  and  $g^A$ , but only to their shapes (successions of increases and decreases). In any case, System (1.3.8) has no stationary state with densities, all stationary states will be composed of Dirac measures. The maximal number of Dirac measures corresponds to the number of maximums of functions  $g^P$  and  $g^A$ . For example, the number of pollinator species and plant species is reduced to 1 when considering all type of growths given in Figure 2 of [59].

## 1.4 PERSPECTIVES

Understanding mechanisms underlying the evolution of a population is a major issue to explain better the natural diversity. Spatial aspects are assumed to play a key role in this dynamics and particularly in the evolution of the cooperative behaviors. Moreover, these

mechanisms give interesting challenges from a mathematical point of view. For these reasons, I am still particularly interested in continuing these studies.

#### 1.4.1 Adaptive dynamics and limiting processes

In ongoing projects with Amaury Lambert, H el ene Morlon, Josu e Tchouanti-Fotso and Thuy Vo, we are developing individual-based models with fixed population sizes describing the genetic evolution of populations subject to environmental changes. Using appropriate scaling, we are trying to prove that trait dynamics are likely to converge towards stochastic diffusion processes, under long time scales and small mutations assumptions. Indeed, at the macro-evolutionary scale, i.e. on geological time scales studies, the evolution of traits is traditionally modeled by a diffusion, as a Brownian motion or an Ornstein-Uhlenbeck process [78, 54]. Given the extensive use of these comparative phylogenetic approaches by evolutionary biologists, it seems important to interpret and justify them in terms of micro-evolutionary mechanisms (mutation, dispersal, adaptation, genetic divergence, introgression, genetic drift). Using adaptive dynamics approach, Champagnat and Lambert [21] gave a first answer using rare and small mutations assumptions in a model with a discrete population (so relaxing the large population assumption).

We are therefore seeking to develop this type of approach in two different contexts: (1) in the context of a changing environment and, at the same time, connecting the genetic structure of the population, and (2) for populations living on explicit spatial structures.

#### 1.4.2 Cooperative behavior

As presented in Section 1.3.1, the cooperative behaviors, especially these related to a "public good", are subject to a paradox. In a well mixed environment, the evolutionary forces doom populations to extinction, since they tend to select less and less cooperative individuals until the extinction of the species. Solving this paradox using a simple model, incorporating very few mechanisms, remains a fairly unexplored question, which is of particular interest to me.

For example, in collaboration with Vincent Calvez and Fr ed eric Chardard, we are working on a deterministic spatial model to study this puzzle. The dynamics of the population includes diffusion movements and frequent mutations of a trait that models the investment of individuals in cooperation. This model generates a propagating wave with a pulse shape across space. Indeed cooperators invade the empty space, while "cheaters" flood their domains and cause the extinction. We thus aim at studying and quantifying the speed of this propagating wave.

In a second project with C eline Bonnet, Aline Marguet and Charline Smadi, we are studying the possibility of a pleiotropy effect between cooperation and dormancy. To this aim, we are considering a model where cooperators have the ability to enter into dormancy to avoid extinction when facing the arrival of "cheaters". This type of mechanism ensures the survival of cooperators through an oscillating dynamics under specific parameters. Our main objective is thus to identify precisely conditions for survival or extinction.

## 2.1 INTRODUCTION

In addition to the mechanism of natural selection, described by Darwin in [32], and inherent to most evolutionary models, there exists also the mechanism of sexual selection, also described by Darwin in [32] but less studied then. This corresponds to the selection of the individuals best able to find a sexual partner. Indeed, some traits are selected by virtue of the advantage they confer in gaining access to sexual partners [18], since for instance that there is a phase of recognition of sexual partners before mating. Animals have developed complex recognition systems involving visual, olfactory, auditory and chemical signals, as well as highly elaborate behavioral sequences [113, 94, 109]. We speak of *assortative mating preference* or *homogamy* when individuals mate preferentially with those who have a certain degree of morphological and/or genetic similarity, and of *disassortative mating preference* or *heterogamy* in the opposite case. In all cases, these preferences and interactions evolve also under the influence of natural selection and can lead to reproductive isolation and eventual speciation. The mechanisms of natural selection and sexual selection sometimes come into conflict, with selection pressures favoring different aspects of individuals. Identifying the role of sexual selection itself as trigger of the evolution is a major issue [47].

In this chapter, we develop stochastic and deterministic models around three main questions related to sexual selection:

- ◇ Which conditions can lead to the emergence of assortative mating preference?
- ◇ What is the impact of assortative mating preference on the speciation of spatially structured species?
- ◇ Can we quantify the effect of disassortative mating preference on the appearance and persistence of polymorphism?

These questions echo also the problematic raised in the previous chapter in several aspects: diversity, speciation, impact of the spatial structure...

The models of this chapter are constructed as individual-based models [44], as those presented in the previous chapter. They are studied assuming large population and rare mutations assumptions, mimicking assumptions of adaptive dynamics theory presented at the beginning of Section 1.2.1.

Individuals are characterized at least by their phenotype and eventually by their genotype (see Section 2.2) or by their spatial position in a finite number of possible positions (see Section 2.3). The characteristics of the individuals will thus belong to some finite set  $\mathcal{G} := \{i_1, i_2, \dots, i_k\}$  of size  $k$ . The state of the population is characterized at each time  $t$  by a vector in  $\mathbb{N}^k$  giving the respective numbers of individuals carrying each of these characteristics. The dynamics of the total population is thus modeled by a multi-type birth-and-death process

$$(N^K(t), t \geq 0) := (N_{i_1}^K(t), N_{i_2}^K(t), \dots, N_{i_k}^K(t), t \geq 0)$$

with values in  $\mathbb{N}^k$ , integrating competition, Mendelian reproduction, assortative mating and (eventually) migration events. The parameter  $K > 0$  is a scaling parameter. It quantifies the environment's carrying capacity, which is a measure of the maximal population size that the environment can sustain for a long time. As indicated previously, we will be interested in an assumption of large population, which corresponds to studying the behavior of the system for large  $K$ , eventually infinite.

The loss by death is similar in all models considered in this chapter, as we consider only natural death and death by competition. Competition takes place only between individuals from the same patch. To state it, we thus introduce the relation  $i \sim j$  for  $i, j \in \mathcal{G}$ , meaning that  $i$  and  $j$  are two characteristics referring to the same patch. For models without spatial structure, all characteristics refer to the same patch and all individuals are in competition with each other. Finally, when the population is in state  $\mathbf{n} = (n_{i_1}, n_{i_2}, \dots, n_{i_k}) \in \mathbb{N}^k$ , the rate at which the population loses an individual of characteristic  $i \in \mathcal{G}$  by death, is equal to

$$d_i(\mathbf{n}) = n_i \left( d + \frac{c}{K} \sum_{j \sim i, j \in \mathcal{G}} n_j \right). \quad (2.1.1)$$

The parameters  $d \in \mathbb{R}^+$  and  $c > 0$  respectively model the natural and the competition death rates of individuals.

When there are several patches in the model, we define some migrations between patches. When an event of migration occurs, it results in the transfer of an individual from one characteristic  $i \in \mathcal{G}$  to another, more details will be given in Section 2.3. From a general point of view, when the population is in state  $\mathbf{n}$ , we can define the migration rate at which an individual of characteristic  $i \in \mathcal{G}$  migrates and thus acquires characteristic  $j \in \mathcal{G}$  as

$$m_{i,j}(\mathbf{n}). \quad (2.1.2)$$

This migration rate may depend on the characteristics. Indeed, it may include for instance the position of the individual, the mating preferences of the individual or the composition of the population. In particular, in [29, 83], we have been interested in situations where individuals carrying a phenotype at low frequency within patch have a greater migration rate. This models migration promoted by the local lack of suitable mates.

Finally, reproduction also occurs only between individuals living in the same patch. When the population is in state  $\mathbf{n}$ , the rate  $b_i(\mathbf{n})$  at which an individual of characteristic  $i \in \mathcal{G}$  appears in the population is given by

$$b_i(\mathbf{n}) := \sum_{j \sim i, j \in \mathcal{G}} \beta_j \left( \sum_{\ell \sim i} p_{j,\ell}^i \frac{n_\ell}{\sum_{\tilde{j} \sim i} n_{\tilde{j}}} \right) n_j. \quad (2.1.3)$$

The parameter  $\beta_j > 0$  represents the rate at which any individual of characteristic  $j \in \mathcal{G}$  (called first parent or choosing parent) mates, the second parent (also called the chosen parent) being chosen uniformly at random in the local population, i.e. among all individuals that belong to the same patch, hence the term  $\frac{n_\ell}{\sum_{\tilde{j} \sim i} n_{\tilde{j}}}$ . A mating where the first parent is of characteristic  $j$  and the second one is of characteristic  $\ell$  actually produces an offspring of characteristic  $i$  with probability  $p_{j,\ell}^i$ . This probability includes the sexual preferences of the first parent and the Mendelian rules of reproduction.

From a mathematical point of view, this model can be written using Poisson point measures as done in [44].

Similarly to results given in chapter 1, we can study the process in a limit of large population (i.e.  $K \rightarrow \infty$ ). Then the population dynamics is well approximated by a dynamical system. More precisely, the rescaled process

$$(\mathbf{Z}^K(t), t \geq 0) := \left( \frac{N_{i_1}^K(t)}{K}, \frac{N_{i_2}^K(t)}{K}, \dots, \frac{N_{i_k}^K(t)}{K}, t \geq 0 \right),$$

is close to the solution of the dynamical system

$$\dot{z}_i = b_i(\mathbf{z}) - (d + c \sum_{j \sim i} z_j) z_i - \sum_{j \in \mathcal{G}} m_{ij}(\mathbf{z}) + \sum_{j \in \mathcal{G}} m_{ji}(\mathbf{z}), \quad i \in \mathcal{G}, \quad (2.1.4)$$

where  $z = \sum_{j \in \mathcal{G}} z_j$  is the total size of the population. According to [26], this dynamical system has a unique solution starting from any point of  $\mathbb{R}_+^k$ . Hence, if we denote by

$$(\mathbf{z}^{(z^0)}(t), t \geq 0) = (z_{i_1}(t), \dots, z_{i_k}(t), t \geq 0)$$

this unique solution starting from  $\mathbf{z}(0) = \mathbf{z}^0 \in \mathbb{R}_+^k$ , we have the following result, which derives from Theorem 2.1 p 456 in [39].

**Lemma 2.1.1.** *Let  $T \in \mathbb{R}_+^*$ . Assume that the sequence  $(\mathbf{Z}^K(0), K \geq 1)$  converges in probability (resp. a.s.) to some deterministic vector  $\mathbf{z}^0 \in \mathbb{R}_+^k$  when  $K$  goes to infinity. Then*

$$\lim_{K \rightarrow \infty} \sup_{s \leq T} \|\mathbf{Z}^K(s) - \mathbf{z}^{(z^0)}(s)\|_\infty = 0 \quad \text{in probability (resp a.s.)}, \quad (2.1.5)$$

where  $\|\cdot\|_\infty$  denotes the  $L^\infty$ -Norm in  $\mathbb{R}^k$ .

The fine study of the limiting system (2.1.4) is essential to precisely understand the biological behavior of the population, and to analyze this behavior with respect to the different parameters. In this chapter, I present both results consisting in the precise study of deterministic limits, as well as results on more stochastic aspects, involving the invasion phases of certain mutant subpopulations arriving in a stationary resident population as it is the case in the next section.



## 2.2 EMERGENCE OF ASSORTATIVE MATING

This section is taken from the article [27] written in collaboration with Camille Coron, Manon Costa, Fabien Laroche et Charline Smadi.

We determine sufficient conditions to guarantee the emergence of assortative mating preference. Indeed, if a group of individuals starts to reproduce preferentially with a small part of the population, its reproductive success is likely to be reduced in comparison with individuals which continue to choose their partners indifferently. Our goal was thus to quantify precisely the costs and benefits required for individuals with preferential mating to obtain a selective advantage.

To this aim, we developed an individual-based model representing a haploid population that reproduce sexually, living all on the same patch. Individuals are characterized by their alleles at two loci located on different chromosomes. The first locus, where allele  $A$  or allele  $a$  can reside, codes for the phenotype of the individual. The second locus codes for the sexual preference: two alleles,  $p$  and  $P$ , can reside on it. If an individual carries the allele  $p$ , it is assumed to have no preference. It therefore mates indifferently with individuals of both phenotypes  $a$  and  $A$ , and when a reproductive event affects it, it selects its partner uniformly at random among all living individuals. In contrast, when an individual carries the  $P$  allele, it is assumed to present preferential mating with individuals carrying the same allele at the first locus as its own phenotype.

In this model, individuals are thus characterized by their genotypes included in the set  $\mathcal{G} := \{AP, Ap, aP, ap\}$ . To answer our specific question, all individuals are assumed to carry allele  $p$  initially and thus show no mating preference, except one individual that carries allele  $P$ . This is considered as the result of a mutation of allele  $p$  to allele  $P$ .

No spatial structure is considered in the model, so that the migration rates (2.1.2) are all equal to 0, and that individuals are all in competition with each other, i.e. the death rate (2.1.1) is

$$d_i(\mathbf{n}) = n_i \left( d + \frac{c}{K} n \right), \forall i \in \mathcal{G}, \quad (2.2.1)$$

with  $n = \sum_{j \in \mathcal{G}} n_j$ .

Two parameters  $\beta_1 \geq 0$  and  $\beta_2 \in [0, 1]$  respectively model the benefit and cost of the mating preference in the following way. Let set  $b$  a positive parameter, with  $b > d$ . All individuals are assumed to reproduce at rate  $b(1 + \beta_1)$ , i.e. to become the first parent for the mating, the second parent is then chosen at random, as explained in the introduction. The mating actually produces a descendant with probability:

- $1/(1 + \beta_1)$  when the first parent carries allele  $p$ ,
- 1 if the first parent carries allele  $P$  and both parents carry the same allele at locus 1,
- $(1 - \beta_2)/(1 + \beta_1)$  if the first parent carries allele  $P$  and the two parents carry different alleles at locus 1.

Notice that, with this choice of parameters, an individual that carries allele  $p$  actually reproduces at rate  $b$ , whatever the genotype of the second parent is, and thus do not show mating preference. The assumption  $b > d$  ensures that a population of such individuals can survive and is not subject to rapid extinction. Individuals with trait  $P$  reproduce more often (at rate  $b(1 + \beta_1)$ ) when surrounded by individuals with the same phenotype ( $A$  or  $a$ ), and less often (at rate  $b(1 - \beta_2)$ ) when surrounded by individuals with the opposite trait. The genome of the descendant is then chosen according to Mendelian rules, considering that the two loci are not on the same chromosome.

Finally, when the population is in state  $\mathbf{n}$ , the rate  $b_i(n)$  at which an individual with genotype  $i \in \mathcal{G}$  is born, is defined by

$$\begin{aligned}
b_{AP}(\mathbf{n}) &= b \left[ n_{AP} + \frac{1}{n} \left( \beta_1 n_{AP} \left( n_{AP} + \frac{n_{Ap}}{2} \right) \right. \right. \\
&\quad \left. \left. - \beta_2 \left( n_{AP} \left( n_{aP} + \frac{n_{ap}}{4} \right) + n_{Ap} \frac{n_{aP}}{4} \right) \right) + \frac{\Delta_{aP}}{2n} \right] \\
b_{Ap}(\mathbf{n}) &= b \left[ n_{Ap} + \frac{1}{n} \left( \beta_1 n_{Ap} \frac{n_{AP}}{2} - \beta_2 \left( n_{Ap} \frac{n_{aP}}{4} + n_{AP} \frac{n_{ap}}{4} \right) \right) - \frac{\Delta_{aP}}{2n} \right] \\
b_{aP}(\mathbf{n}) &= b \left[ n_{aP} + \frac{1}{n} \left( \beta_1 n_{aP} \left( n_{aP} + \frac{n_{ap}}{2} \right) \right. \right. \\
&\quad \left. \left. - \beta_2 \left( n_{aP} \left( n_{AP} + \frac{n_{Ap}}{4} \right) + n_{ap} \frac{n_{AP}}{4} \right) \right) - \frac{\Delta_{aP}}{2n} \right] \\
b_{ap}(\mathbf{n}) &= b \left[ n_{ap} + \frac{1}{n} \left( \beta_1 n_{ap} \frac{n_{aP}}{2} - \beta_2 \left( n_{ap} \frac{n_{AP}}{4} + n_{aP} \frac{n_{Ap}}{4} \right) \right) + \frac{\Delta_{aP}}{2n} \right],
\end{aligned} \tag{2.2.2}$$

where

$$\Delta_{aP} := n_{aP}n_{Ap} - n_{AP}n_{ap}.$$

As indicated previously, we assume that at time 0, the population is composed of individuals that carry allele  $p$ , i.e. show no mating preference, and of 1 individual of genotype  $\alpha P$ , with  $\alpha = A$  or  $a$ . The total size of the initial population is assumed to be close to  $K(b - d)/c$ , that means that it is close to its equilibrium state. Indeed any neighborhood of such an equilibrium is reached within a finite time as soon as the initial population size is of order  $K$ . We thus could relax this assumption and only assume that the  $p$ -population size is of order  $K$  and  $N_{Ap}^K(0) > N_{ap}^K(0)$ . This would however require more complex notations.

In the context described previously, we expressed conditions under which the mutant individuals with allele  $P$  survive and invade the resident population, until the extinction of the initial allele  $p$ . These conditions can be simply expressed using the parameters  $\beta_1, \beta_2$  and the initial proportions of individuals carrying alleles  $A$  and  $a$ , i.e.

$$\rho_A := \lim_{K \rightarrow \infty} \frac{N_{Ap}^K(0)}{N_p^K(0)} \quad \text{and} \quad \rho_a := 1 - \rho_A = \lim_{K \rightarrow \infty} \frac{N_{ap}^K(0)}{N_p^K(0)}. \tag{2.2.3}$$

The mutant individual has a positive probability to survive and invade the resident popula-

tion if and only if

$$\beta_1 > \beta_2 \quad \text{or} \quad \rho_A(1 - \rho_A) < \frac{\beta_1(\beta_2 + 2)}{2(\beta_1 + \beta_2)(\beta_1 + 2)}. \quad (2.2.4)$$

Notice that this condition gives two sufficient conditions. The first one imposes that the trade-off between the advantage for assortative mating ( $\beta_1$ ) and the loss for disassortative reproduction ( $\beta_2$ ) has to be favorable enough. The second condition requires a low level of initial allelic diversity at locus 1 (alleles  $A$  and  $a$ ). In particular, even if the advantage for assortative mating is very low, very asymmetrical initial conditions ( $\rho_A$  close to 0 or 1) will ensure the invasion of the mutation with positive probability.

The condition (2.2.4) is obtained from the study of a branching process which is proved to be supercritical if and only if the matrix

$$J := \frac{b}{4} \begin{pmatrix} 2\beta_1\rho_A - (2 + \beta_2)\rho_a & (2 - \beta_2)\rho_a \\ (2 - \beta_2)\rho_A & 2\beta_1\rho_a - (2 + \beta_2)\rho_A \end{pmatrix} \quad (2.2.5)$$

has a positive eigenvalue. We denote by  $\lambda$  the maximum eigenvalue of (2.2.5), which is thus positive when (2.2.4) holds. Let also denote by  $(q_A, q_a)$  the smallest solution to the system of equations

$$\begin{aligned} u_A(s_A, s_a) &:= b(1 - s_A) + (J_{11} + b)(s_A^2 - s_A) + J_{12}(s_A s_a - s_A) = 0 \\ u_a(s_A, s_a) &:= b(1 - s_a) + (J_{22} + b)(s_a^2 - s_a) + J_{21}(s_A s_a - s_a) = 0. \end{aligned} \quad (2.2.6)$$

$q_\alpha$  will correspond to the extinction probability of the offspring of the mutant individual, it is equal to 1 as soon as (2.2.4) does not hold. Finally, to state our main result, let us introduce the following set of interest, for any  $\mu > 0$

$$S_\mu := \left[ \frac{b(1 + \beta_1) - d}{c} - \mu, \frac{b(1 + \beta_1) - d}{c} + \mu \right] \times \{0\} \times \{0\} \times \{0\}, \quad (2.2.7)$$

and let us define a stopping time that gives the time at which  $\mathbf{Z}^K$  reaches this set,

$$T_{S_\mu}^K := \inf\{t \geq 0, \mathbf{Z}^K(t) \in S_\mu\}. \quad (2.2.8)$$

as well as a stopping time which gives the time of extinction of the  $P$ -mutant population,

$$T_0^{P,K} := \inf\{t > 0, N_P^K(t) = 0\}, \quad (2.2.9)$$

We now state the main result

**Theorem 2.2.1.** *Assume that  $\lambda \neq 0$ , and*

$$\left( Z_{Ap}^K(0), Z_{ap}^K(0) \right) \xrightarrow{K \rightarrow \infty} \left( \rho_A \frac{b-d}{c}, (1 - \rho_A) \frac{b-d}{c} \right) \text{ in probability}$$

with  $\rho_A > 1/2$  and that for some  $\alpha \in \{A, a\}$

$$\left( N_{\bar{\alpha}P}^K(0), N_{\bar{\alpha}P}^K(0) \right) = (1, 0), \text{ with } \bar{\alpha} \in \{A, a\} \setminus \{\alpha\}.$$

Then there exists a Bernoulli random variable  $B$  with parameter  $1 - q_\alpha$  such that for any  $0 < \mu < (b(1 + \beta_1) - d)/c$ :

$$\lim_{K \rightarrow \infty} \left( \frac{T_{S_\mu}^K \wedge T_0^{P,K}}{\ln K}, \mathbf{1}_{\{T_{S_\mu}^K < T_0^{P,K}\}} \right) = B \times \left( \frac{1}{\lambda} + \frac{2}{b\beta_1}, 1 \right), \text{ in probability.} \quad (2.2.10)$$

Moreover,

$$\mathbf{1}_{\{T_0^{P,K} < T_{S_\mu}^K\}} \left\| \frac{\mathbf{N}^K(T_0^{P,K})}{K} - (0, \rho_A, 0, 1 - \rho_A) \frac{b-d}{c} \right\|_1 \xrightarrow{K \rightarrow \infty} 0 \text{ in probability,} \quad (2.2.11)$$

where  $\|\cdot\|_1$  stands for the  $L^1$ -norm.

Notice that if condition (2.2.4) does not hold,  $q_\alpha = 1$ , and the convergence in (2.2.10) is an almost sure convergence to  $(0, 0)$  meaning that the mutant population dies out in a time smaller than  $\ln K$ . In this case, the allelic proportions in the resident population do not vary.

Our result also ensures that when the mutant population invades (whatever allele  $a$  or  $A$  the first mutant carries), then the final population is monomorphic, and all individuals carry the allele  $a$  or  $A$  which was in the majority in the resident  $p$ -population. Only the mutant invasion probability depends on the allele carried by the first  $P$  individual.

We were not able to obtain an explicit formula in general for the extinction probability  $q_\alpha$  of the assortative mating mutation, solutions of (2.2.6), except in the particular case when there are only  $A$ - or  $a$ -individuals in the population before the arrival of the mutant ( $\rho_A \in \{0, 1\}$ ), see Proposition 2.3 in [27] for an exact formula. Then simulations show complex dependencies with respect to parameters.

**Ideas of proof** The proof of Theorem 2.2.1 relies on the study of three phases in the population dynamics trajectories (mutant survival or extinction, mean-field phase, and resident allele extinction). Although classical and used in several proofs preceding ours [19, 24], these three steps are more complex to study in our case because of the composition of the initial population which has the particularity of relying on a non-hyperbolic state. Indeed, the initial state corresponds to a  $p$ -population of size of order  $K$  and a  $P$ -population of size negligible with respect to  $K$ . Applying the deterministic approximation explained in the introduction of this chapter (see Lemma 2.1.5), we see that, for finite time, the dynamics of the initial population is well approximated by the dynamical system

$$\begin{cases} \dot{z}_{Ap} = z_{Ap}(b - d - c(z_{Ap} + z_{ap})) \\ \dot{z}_{ap} = z_{ap}(b - d - c(z_{Ap} + z_{ap})). \end{cases}$$

This system admits an infinity of equilibria:

- $(z_{Ap}, z_{ap}) = (0, 0)$ , which is unstable
- $(z_{Ap}, z_{ap}) = (\rho(b-d)/c, (1-\rho)(b-d)/c)$  for all  $\rho \in [0, 1]$ , which are non hyperbolic.

The initial state of the resident population is thus more variable than the ones of similar studies that rely on stable hyperbolic state. However, the equation giving the dynamics of the total population size  $z = z_{Ap} + z_{ap}$  is  $\dot{z} = z(b - d - cz)$ , and admits a globally stable equilibrium  $(b - d)/c$ .

Let us now briefly outline the three steps of the proof and the difficulties inherent in them.

**Step 1: mutant invasion.** The first step of the proof consists in studying the phase of invasion or extinction of the mutant population, which corresponds to a phase where the mutant population is negligible with respect to the carrying capacity  $K$ . To this aim, we set the following stopping time which gives the first time when the rescaled  $P$ -mutant population size reaches any threshold (from below or above):

$$T_\varepsilon^{P,K} := \inf \left\{ t > 0, N_P^K(t) = \lfloor \varepsilon K \rfloor \right\}, \quad (2.2.12)$$

where  $\lfloor x \rfloor$  is the integer part of  $x$ . We prove that, as long as the mutant population stays sufficiently small, i.e. before  $T_0^{P,K} \wedge T_\varepsilon^{P,K}$ , the resident population is not affected by it and stays close to its initial state characterized by its initial proportions (2.2.3) and its initial size, close to  $\frac{b-d}{c}$ .

Precisely, this relies on the study of the stopping time

$$U_\varepsilon^K := \inf \left\{ t \geq 0, \left| \frac{N_{Ap}^K(t)}{N_P^K(t)} - \frac{N_{Ap}^K(0)}{N_P^K(0)} \right| > \varepsilon \right\}, \quad (2.2.13)$$

which is the first time when the genetic proportions in the  $p$ -population deviate considerably from their starting values. Using martingales techniques, we prove that the proportions in the resident population do not vary substantially before the mutant population goes extinct or invades.

Then we prove that the total size of  $p$ -population is only slightly modified during this invasion (or extinction) phase, by focusing on the stopping time

$$R_\varepsilon^K := \inf \left\{ t \geq 0, \left| \frac{N_p^K(t)}{K} - \frac{b-d}{c} \right| > \varepsilon \right\}. \quad (2.2.14)$$

which gives the first time that the total size of the  $p$ -population deviates from its equilibrium state. This part of the proof relies on couplings between the process that gives the  $p$ -population total size and some birth-and-death processes, in the vein of [19].

The mutant population can then be compared with a branching process during the first times of the invasion, considering that the size and proportions of the resident population are fixed. In other words, by considering that  $(N_{Ap}^K, N_{ap}^K) = (K\rho_A \frac{b-d}{c}, K(1-\rho_A) \frac{b-d}{c})$ ,  $N^K = K \frac{b-d}{c}$  and by neglecting the second order terms in  $N_{Ap}^K$  and  $N_{ap}^K$ , the dynamics of the mutant population can be approximated by the one of  $\tilde{\mathbf{N}} = (\tilde{N}_A, \tilde{N}_a)$  a bi-type branching process

with the following transition rates:

$$\begin{aligned}
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A + 1, \bar{N}_a) & \text{at rate } & \bar{\beta}_{AA}\bar{N}_A + \bar{\beta}_{aA}\bar{N}_a \\
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A, \bar{N}_a + 1) & \text{at rate } & \bar{\beta}_{Aa}\bar{N}_A + \bar{\beta}_{aa}\bar{N}_a \\
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A - 1, \bar{N}_a) & \text{at rate } & b\bar{N}_A \\
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A, \bar{N}_a - 1) & \text{at rate } & b\bar{N}_a,
\end{aligned} \tag{2.2.15}$$

where for  $\alpha \in \{A, a\}$ ,  $\bar{\alpha} \in \{A, a\} \setminus \{\alpha\}$ ,

$$\bar{\beta}_{\alpha\alpha} := \frac{b}{2} \left( 1 + (\beta_1 + 1)\rho_\alpha - \frac{\beta_2}{2}\rho_{\bar{\alpha}} \right), \quad \bar{\beta}_{\alpha\bar{\alpha}} := \frac{b}{2} \left( 1 - \frac{\beta_2}{2} \right) \rho_{\bar{\alpha}}. \tag{2.2.16}$$

The extinction probabilities of the process  $\bar{\mathbf{N}}$  are given by

$$q_\alpha := \mathbb{P}(\exists t < \infty, \bar{\mathbf{N}}(t) = 0 | \bar{\mathbf{N}}(0) = \mathbf{e}_\alpha), \tag{2.2.17}$$

$\alpha \in \{A, a\}$ ,  $\mathbf{e}_A = (1, 0)$  and  $\mathbf{e}_a = (0, 1)$ , meaning that the process starts with only one individual of type  $A$  or  $a$ . According to classical results of branching process theory (see [4]), these extinction probabilities correspond to the smallest solution to the system (2.2.6). Moreover, the branching process  $\bar{\mathbf{N}}$  is supercritical (i.e.  $q_A$  and  $q_a$  are not equal to one) if and only if its mean matrix has a positive eigenvalue. A short computation shows that this matrix corresponds to matrix  $J$  defined in (2.2.5).

**Step 2: mean-field phase.** If the mutant population invades and its size reaches order  $K$ , the population dynamics is proved to enter a second phase during which it is well approximated by the dynamical system (2.1.4), as explained in the introduction of this chapter (see Lemma 2.1.5). An important question however is the initial condition of the limiting dynamical system.

We have seen that when (2.2.4) is satisfied, then the mutant population dynamics is close to that of the supercritical bi-type branching process  $\bar{\mathbf{N}}$  defined in (2.2.15). For such a process we are able to control the long time proportion of the different types of individuals. More precisely, Kesten-Stigum theorem (see [49] for instance) ensures the following property, if  $\lambda$  is positive:

$$\frac{(\bar{N}_A(t), \bar{N}_a(t))}{\bar{N}_A(t) + \bar{N}_a(t)} \xrightarrow{t \rightarrow \infty} (\pi_A, \pi_a) \quad \text{almost surely}$$

on the event of survival of  $\bar{\mathbf{N}}$ , where  $\pi$  is the positive left eigenvalue of  $J$  associated to  $\lambda$  such that  $\pi_A + \pi_a = 1$ .

In our proof, we show that, with a probability close to one for large  $K$ , if the mutant population reaches the size  $\varepsilon K$ , we may choose a time when the proportion of type  $A$  individuals in the  $P$ -population belongs to  $[\pi_A - \delta, \pi_A + \delta]$ , with  $\delta > 0$  small. In particular, the proportion  $\pi_A$  of individuals carrying allele  $A$  among the  $P$ -individuals is larger than  $1/2$  as soon as it is the case for the initial  $p$ -population.

We then study the limiting dynamical system (2.1.4) to prove that for an initial condition  $\mathbf{z}^0$  such that  $z_{AP}(0) > z_{ap}(0)$  and  $z_{AP}(0) > z_{aP}(0)$ , and if

$$\beta_1 > \beta_2 \quad \text{or} \quad \frac{(z_{AP}(0) + z_{aP}(0))(z_{AP}(0) + z_{ap}(0))}{z(0)^2} < \frac{\beta_1(\beta_2 + 2)}{2(\beta_1 + \beta_2)(\beta_1 + 2)}. \tag{2.2.18}$$

Then the solution  $\mathbf{z}^{(z^0)}$  of the system (2.1.4) converges as  $t \rightarrow \infty$  toward

$$\left( \frac{(1 + \beta_1)b - d}{c}, 0, 0, 0 \right).$$

Hence, the stochastic process follows a similar dynamics.

**Step 3: extinction of the resident population.** After the deterministic phase, the process is close to the state  $((b(\beta_1 + 1) - d)/c, 0, 0, 0)$ . When the sizes of populations  $aP$ ,  $Ap$  and  $ap$  become too small (of order smaller than  $K$  before rescaling), the mean field approximation stops being a good approximation, and we will again compare the dynamics of the small population sizes with those of branching processes (now subcritical). The birth and death rates of these branching processes provide the time to extinction of these small populations.

**Conclusion.** Combining all these steps, we are able to describe the invasion/extinction dynamics of the mutant population and prove Theorem 2.2.1.

## 2.3 SPECIATION AND ASSORTATIVE MATING

In this second section, some elements are given to answer the second question set out in the introduction, namely, what is the impact of assortative mating preference on the speciation of spatially structured species? Indeed, more and more evidence shows that mating preference is a mechanism that may lead to a reproductive isolation event [76, 15]. Initially, the role of ‘magic’ or ‘multiple effect’ traits, which associate both adaptation to an ecological niche and a mate preference, was studied deeply. It was shown that such traits may lead to speciation, using direct experimental evidence [95] or theoretical works [77, 121]. Then, studies focused on the particular role of mating preference during a speciation event [47], highlighting that (i) it may impede reproductive isolation [110, 111, 112], or, (ii) it may promote reproductive isolation. This promoting role may be secondary or primary. For example, the initial divergence in traits may be the result of natural selection in order to decrease hybridization and then be subjected to mating preference [102], producing speciation by reinforcement [50]. Our studies illustrate the direct and promoting role of assortative mating using theoretical studies and numerical simulations.

This section corresponds to the three articles [29, 83, 114] written in collaboration with Camille Coron, Manon Costa, Violaine Llaurens and Charline Smadi. The first two articles are explained briefly as they had already been detailed in a chapter of my thesis in a preliminary version. The first two articles are based on the modeling of a haploid population spatially structured on patches, and the third one is an extension of this model to a diploid population.

### 2.3.1 Haploid population

The population is divided into several patches. As in the two articles [29, 83], we focus the presentation on the case of two patches. The individuals are thus characterized by their

position (1 or 2 depending on the patch in which they are). Moreover, they are haploid and characterized by a diallelic locus ( $a$  or  $A$ ). The set of characteristics is thus

$$\mathcal{G} := \{(\alpha, i), \alpha \in \{a, A\}, i \in \{1, 2\}\}.$$

We assume frequency-dependent migration rate from one patch to another in such a way that individuals are more prone to move if they do not find a suitable mate. This hypothesis is relevant for all organisms with active mate searching [104, 120, 65]. The locus codes for the strength of the mating preference and simultaneously, the speed of migration, that depends on parameters  $m_\alpha$  such that the total migration rate of  $\alpha$ -individuals from patch 1 to patch 2 finally is

$$m_\alpha \left( \frac{n_{\bar{\alpha},1}}{n_{\alpha,1} + n_{\bar{\alpha},1}} \right) n_{\alpha,1}, \text{ with } \bar{\alpha} \in \{A, a\} \setminus \alpha. \quad (2.3.1)$$

Note that the migration rate does not depend on the other deme composition. Figure 2.1 illustrates the migration rates in the two patches case.

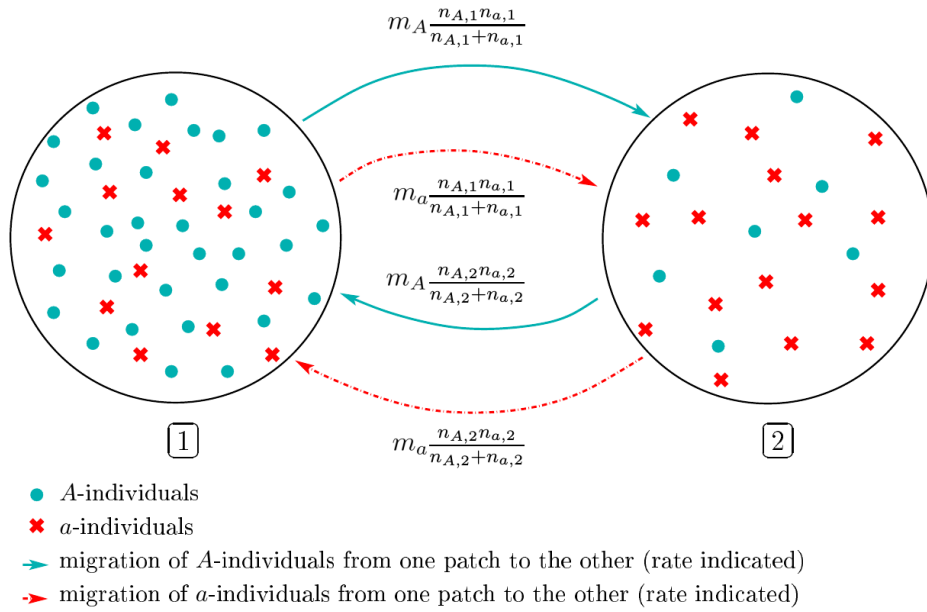


Figure 2.1 – Migration rates of  $A$  and  $a$  individuals in the two patches case.

Following the ideas given in the introduction, such that after encountering, two individuals that carry the same allele  $\alpha$  have a probability  $\beta_\alpha$ -times larger to mate and give birth to a viable offspring than two mating individuals with different traits, the total birth rate of  $\alpha$ -individuals in patch  $i$  is

$$bn_{\alpha,i} \frac{\beta_\alpha n_{\alpha,i} + n_{\bar{\alpha},i}}{n_{\alpha,i} + n_{\bar{\alpha},i}}. \quad (2.3.2)$$

Hence formula (2.3.2) models an assortative mating by phenotypic matching or recognition alleles [13, 64]. In this model, preference thus modifies the rate of mating and not only the distribution of genotypes, unlike what is usually assumed in classical generational models



where all individuals reproduce simultaneously at discrete times [97, 76, 70, 46, 17, 110]. The present model can be compared with these classical ones by computing the probabilities that an individual of trait  $\alpha$  in the patch  $i$  gives birth after encountering an individual of the same trait (resp. of the opposite trait) conditionally on the fact that this individual gives birth at time  $t$ , and we find

$$\frac{\beta_\alpha N_{\alpha,i}^K}{\beta_\alpha N_{\alpha,i}^K + N_{\bar{\alpha},i}^K} \quad \left( \text{resp.} \quad \frac{N_{\bar{\alpha},i}^K}{\beta_\alpha N_{\alpha,i}^K + N_{\bar{\alpha},i}^K} \right).$$

These probabilities are similar to the ones presented in [110], or in [48] for instance.

In this section, we assume that:

$$\beta_A > 1, \beta_a > 1, b > d > 0, c > 0, m_A \geq 0, m_a \geq 0.$$

Finally, in a large population approximation (see Lemma 2.1.5), the limiting dynamical system is

$$\begin{cases} \frac{d}{dt} z_{A,1}(t) = z_{A,1} \left[ b \frac{\beta_A z_{A,1} + z_{a,1}}{z_{A,1} + z_{a,1}} - d - c(z_{A,1} + z_{a,1}) - m_A \frac{z_{a,1}}{z_{A,1} + z_{a,1}} \right] + m_A \frac{z_{A,2} z_{a,2}}{z_{A,2} + z_{a,2}} \\ \frac{d}{dt} z_{a,1}(t) = z_{a,1} \left[ b \frac{\beta_a z_{a,1} + z_{A,1}}{z_{A,1} + z_{a,1}} - d - c(z_{A,1} + z_{a,1}) - m_a \frac{z_{A,1}}{z_{A,1} + z_{a,1}} \right] + m_a \frac{z_{A,2} z_{a,2}}{z_{A,2} + z_{a,2}} \\ \frac{d}{dt} z_{A,2}(t) = z_{A,2} \left[ b \frac{\beta_A z_{A,2} + z_{a,2}}{z_{A,2} + z_{a,2}} - d - c(z_{A,2} + z_{a,2}) - m_A \frac{z_{a,2}}{z_{A,2} + z_{a,2}} \right] + m_A \frac{z_{A,1} z_{a,1}}{z_{A,1} + z_{a,1}} \\ \frac{d}{dt} z_{a,2}(t) = z_{a,2} \left[ b \frac{\beta_a z_{a,2} + z_{A,2}}{z_{A,2} + z_{a,2}} - d - c(z_{A,2} + z_{a,2}) - m_a \frac{z_{A,2}}{z_{A,2} + z_{a,2}} \right] + m_a \frac{z_{A,1} z_{a,1}}{z_{A,1} + z_{a,1}}. \end{cases} \quad (2.3.3)$$

A direct computation implies that the following four points are stable equilibria of the system:

- equilibria with fixation of an allele (where only an allele is maintained in both patches)

$$(\zeta_A, 0, \zeta_A, 0), (0, \zeta_a, 0, \zeta_a), \quad (2.3.4)$$

- equilibria with maintenance of each allele in a different patch

$$(\zeta_A, 0, 0, \zeta_a), (0, \zeta_a, \zeta_A, 0), \quad (2.3.5)$$

with  $\zeta_\alpha := \frac{b\beta_\alpha - d}{c}$ ,  $\alpha \in \{A, a\}$ . These four equilibria describe states of reproductive isolation: once reaching one of these equilibria, migration rates equal zero and individuals do not migrate anymore. More specifically, observe that equilibria (2.3.5) are of particular interest to our problematic. Indeed, once reaching one of these equilibria, even if a small basal migration (i.e. constant migration) is added, the mating preferences and the frequency-dependent migration terms will prevent the populations of both demes to mix again, leading to *migration-selection balance* [67] but where selection is due to sexual selection and not to natural selection. Precisely, if an  $A$ -individual travels because of basal migration from patch 1 to patch 2, which is filled with  $a$ -individuals, its probability to reproduce will be significantly reduced in patch

2 and its migration rate to come back will be so high that it is quite unlikely that its offspring establish in patch 2. This reasoning, however, fails with equilibria (2.3.4).

We prove that all trajectories (except those starting from a set with empty interior) converge to one of these four equilibria as soon as the values of the migration rates are not too large. Numerics seem to show however that it is true for all parameters. We also detail the set of initial states that lead to each of these equilibria, i.e. we detail the basin of attraction of each equilibrium. The proofs in the symmetrical and asymmetrical cases are significantly different from each other and will be discussed in a more thorough manner in the next two subsections 2.3.1.1 and 2.3.1.2.

Let us denote by  $\mathcal{D}_{m_A, m_a}^{A,a}$  the set of initial conditions associated to a trajectory that converges to  $(\zeta_A, 0, 0, \zeta_a)$ . We define similarly all the sets  $\mathcal{D}_{m_A, m_a}^{\alpha, \alpha'}$  for  $\alpha, \alpha' \in \{A, a\}$ . When  $m_A = m_a = 0$ , these sets can be written as

$$\mathcal{D}_{0,0}^{A,a} = \left\{ \mathbf{z} \in \mathbb{R}^{\mathcal{E}}, (\beta_A - 1)z_{A,1} > (\beta_a - 1)z_{a,1} \text{ and } (\beta_A - 1)z_{A,2} < (\beta_a - 1)z_{a,2} \right\},$$

with similar expressions for the three other sets. Hence, without migration, the equilibrium reached depends on the initial number of individuals of each type and of the mating preference strengths. The basins of attraction  $\mathcal{D}_{m_A, m_a}^{\alpha, \alpha'}$  are continuously deformed with respect to  $m_A$  and  $m_a$ . By studying numerically these basins (see Figure 2.2 for some examples), we can conclude that, if  $\beta_A > \beta_a$ , large migration parameters tends to diminish the selective advantage of  $A$ -individuals by mixing the populations of both patches, except for the symmetrical case.

We are able to give an exact formulation of these basins, only in the symmetrical case. In this case, they never depend on  $m_A = m_a$  and can be written

$$\mathcal{D}_{m_A, m_A}^{A,a} = \mathcal{D}^{A,a} = \left\{ \mathbf{z} \in \mathbb{R}^{\mathcal{G}}, z_{A,1} > z_{a,1} \text{ and } z_{A,2} < z_{a,2} \right\},$$

with similar formulations for the three other sets. In other words, in each patch, the convergence will lead to state with a unique allele that corresponds to the one whom proportion at initial state was the highest.

Finally, the main result of this section gives the time before reproductive isolation occurs in the stochastic model. It describes the random time  $T_{\mathcal{B}_{A,a,\varepsilon}}^K$  that is the first time when the population process  $\mathbf{N}^K$  reaches the set

$$\mathcal{B}_{A,a,\varepsilon} := [(\zeta_A - \varepsilon)K, (\zeta_A + \varepsilon)K] \times \{0\} \times \{0\} \times [(\zeta_a - \varepsilon)K, (\zeta_a + \varepsilon)K],$$

with  $\varepsilon > 0$  and when  $K$  is large and starting from  $\mathcal{D}_{m_A, m_a}^{A,a}$ . It corresponds to the random time before (1) all  $a$ -individuals in patch 1 and all  $A$ -individuals in patch 2 get extinct, and (2) the population size in patch 1 is approximately  $K\zeta_A$  and the one in patch 2 is approximately  $K\zeta_a$ .

**Theorem 2.3.1.** *There exists  $m_0 > 0$  such that the following holds for all  $m_A \leq m_0$  and  $m_a \leq m_0$ . Let  $\varepsilon_0 > 0$  and assume also that  $\mathbf{Z}^K(0) = \mathbf{N}^K(0)/K$  converges in probability to a deterministic vector*

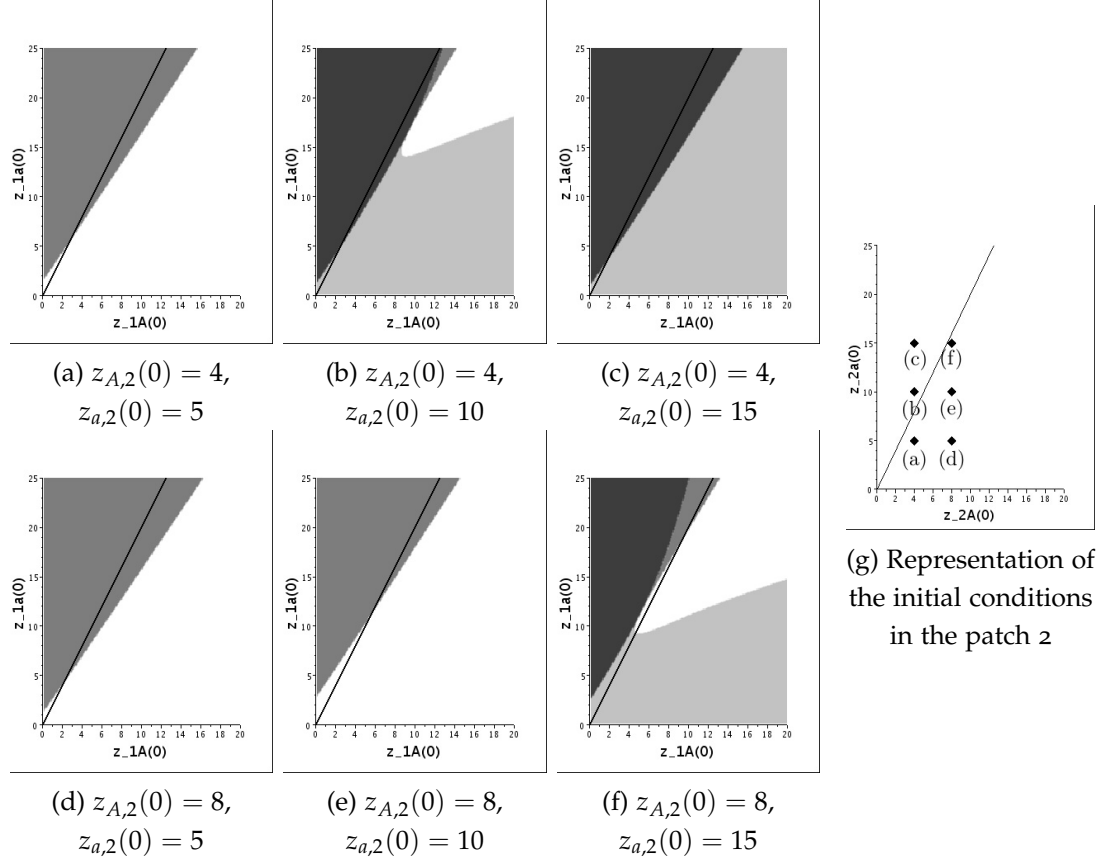


Figure 2.2 – (a-f): Projections of sets  $\mathcal{D}_{5,5}^{\alpha,\alpha'}$  on the planes characterized by the values of  $(z_{A,2}(0), z_{a,2}(0))$  given in captions. On each plane, the four sets from white to dark grey corresponds to initial conditions with convergence to  $(\zeta_A, 0, \zeta_A, 0)$ ,  $(\zeta_A, 0, 0, \zeta_a)$ ,  $(0, \zeta_a, \zeta_A, 0)$  and  $(0, \zeta_a, 0, \zeta_a)$  respectively. The black line is the solution of  $(\beta_A - 1)z_{A,1} - (\beta_a - 1)z_{a,1} = 0$ . (g): The black diamond points correspond to the initial conditions in patch 2 taken to obtain plots (a) to (f).

$\mathbf{z}^0 \in \mathcal{D}_{m_A, m_a}^{A,a}$  such that  $(z_{a,1}^0, z_{A,2}^0) \neq (0, 0)$ . Then there exist  $C_0 > 0$ ,  $M > 0$ , and  $V > 0$  depending only on  $(M, \varepsilon_0)$  such that, for any  $\varepsilon \leq \varepsilon_0$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \left| \frac{T_{\mathcal{B}_{A,a,\varepsilon}}^K}{\log K} - \frac{1}{\omega(A,a)} \right| \leq C_0 \varepsilon, N^K (T_{\mathcal{B}_{A,a,\varepsilon}}^K + t) \in \mathcal{B}_{A,a,M\varepsilon}; \forall t \leq e^{VK} \right) = 1, \quad (2.3.6)$$

where for all  $\alpha, \alpha' \in \{A, a\}$ ,

$$\omega(\alpha, \alpha') = \frac{1}{2} \left[ b(\beta_\alpha - 1 + \beta_{\alpha'} - 1) + m_\alpha + m_{\alpha'} - \sqrt{(b(\beta_\alpha - \beta_{\alpha'}) + (m_{\alpha'} - m_\alpha))^2 + 4m_\alpha m_{\alpha'}} \right]. \quad (2.3.7)$$

Similar results hold for the three other equilibria of (2.3.4) and (2.3.5).

Theorem 2.3.1 gives the first-order approximation of the time before reproductive isolation occurs. The latter is proportional to  $\log(K)$ , which is short compared to  $K$ , the order of

magnitude of the population size. Comparatively, the time scale needed for random genetic drift to cause the end of gene flow between two populations is of order  $K$  in many models (as for example in Wright-Fisher model, see chapter 7 in [55]). Theorem 2.3.1 ensures also that once the equilibrium is reached, the population sizes of both patches stay around  $K\zeta_\alpha$  during at least a long time of order  $e^{KV}$ . The assumption on initial condition  $((z_{a,1}^0, z_{A,2}^0) \neq (0, 0))$  is only needed to obtain the lower bound on the time  $T_{\beta_{A,a,\varepsilon}}^K$  given in (2.3.6). Otherwise, this time would be faster.

Finally,  $\omega(\alpha, \alpha) = b(\beta_\alpha - 1)$ , thus the time before reaching one of equilibria (2.3.4) does not depend on migration parameters unlike the time before reaching one of equilibria (2.3.5).

We discuss more specifically this theorem and its proofs in the two cases (symmetrical and asymmetrical) in the two following subsections. The proofs related to the behavior of the dynamical systems are substantially different.

### 2.3.1.1 Symmetrical case

The symmetrical case corresponds to the case where

$$\beta_A = \beta_a =: \beta \quad \text{and} \quad m_A = m_a =: m.$$

In this particular case, we are able to give more precise results concerning the behavior of the dynamical system (2.3.3).

When  $\beta > 1$ , there exists exactly 13 non-null equilibria of the system, but only the ones of (2.3.4) and (2.3.5) are hyperbolic attractive points. As described previously, almost all trajectories converge to one of these 4 points. Moreover for initial state not too large, we proved that this convergence occurs exponentially fast.

When  $\beta = 1$ , the dynamics is entirely different since there exists an infinite numbers of non-hyperbolic equilibria corresponding to a set of two segments and that contains the equilibria (2.3.4) and (2.3.5).

Finally, the proofs used here are sufficiently robust to take generalizations of the model to study:

- ecological differences between the two patches, i.e. the parameters  $b, d, c$  depend on the patch;
- more general migration rates  $m$  that depend on the size of the populations such that, for  $i \in \{1, 2\}$ , the migration from  $i$  to the other patch is symmetrical with respect to  $n_{A,i}$  and  $n_{a,i}$  (i.e.  $m(n_{A,i}, n_{a,i}) = m(n_{a,i}, n_{A,i})$ ) and  $m(n_{A,i}, n_{a,i}) \frac{n_{A,i} + n_{a,i}}{n_{A,i} n_{a,i}}$  is bounded;
- a general number  $N \in \mathbb{N}$  of patches.

In all of these generalizations, theorems similar to Theorem 2.3.1 can be written and the assortative mating influences the time needed to reach speciation in the same way.

**Ideas of proof.** As in the previous section, the proof relies on the study of different phases, implying different size scales of the population. Here, there are two phases of interest: a first phase of mean field dynamics and then a stochastic phase when some of the populations are microscopic.

**Phase 1. Study of (2.3.3).** Initially, the dynamics of the 4 populations is close to the one of the dynamical system (2.3.3), as recalled in Lemma 2.1.5. Thus a precise study of the trajectories of the system is needed, all results obtained are explained in the main text. The proof of these results relies on an argument of containment, i.e. we prove that all trajectories reach a set

$$\left\{ \mathbf{z} \in \mathbb{R}^G, (z_{A,1} + z_{a,1}, z_{A,2} + z_{a,2}) \in [z_{min}, z_{max}] \right\}, \text{ with } z_{min} > 0, z_{max} > 0,$$

in finite time, by using classical arguments on dynamical systems with a logistic term. Then the proof is completed by specifying Lyapounov functions of the form

$$V(\mathbf{z}) = \ln \left( \frac{z_{A,1} + z_{a,1}}{z_{A,1} - z_{a,1}} \right) + \ln \left( \frac{z_{a,2} + z_{A,2}}{z_{a,2} - z_{A,2}} \right)$$

and adapting arguments of [79].

**Phase 2. Microscopic populations.** Once the dynamical system is not a good approximation anymore, i.e. when some populations reach a threshold  $\varepsilon K$ . We adapt reasoning from [19] and use coupling arguments, i.e. the populations with a small size are compared with a subcritical branching process that is doomed to get extinct after a time proportional to  $\log(K)$ . The exact constant is determined using classical tools of branching processes [4].

### 2.3.1.2 Asymmetrical case

The study of the asymmetrical case ( $\beta_A \neq \beta_a$ ) is advantageous to precisely understand the impact of all parameters. In particular, we are interested in exploring the influence of migration rates and mating preference parameters on the process, using simulations and theoretical analysis.

**Time before differentiation.** According to Theorem 2.3.1 and starting from  $\mathbf{z}^0 \in \mathcal{D}_{m_A, m_a}^{A, a}$ , the stochastic process reaches a state of differentiation ( $A$ -individuals in patch 1 and  $a$ -individuals in patch 2) after a time of magnitude  $\log(K)\omega(A, a)^{-1}$ . Direct functional studies ensure that the constant of interest,  $\omega(A, a)^{-1}$ , is a decreasing function with respect to  $\beta_A$  and to  $\beta_a$  whatever the other parameters are (see Fig. 2.3, left). Hence, the stronger the sexual preference is, the faster the reproductive isolation is.

Then, if we consider that there exist  $\gamma_A, \gamma_a$  and  $m$  such that the migration rates are

$$m_A := \gamma_A m \quad \text{and} \quad m_a := \gamma_a m,$$

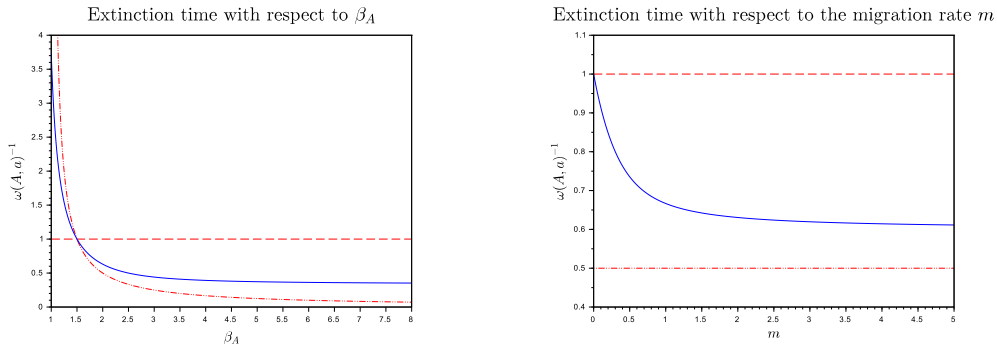


Figure 2.3 – Graphs of the constants in front of the times before reproductive isolation,  $\omega(A,a)^{-1}$  (blue line),  $\omega(A,A)^{-1}$  (red dashed line),  $\omega(a,a)^{-1}$  (red dashed-dotted line), with respect to  $\beta_A$  (left) and to  $m$  (right). The demographic parameters are  $\beta_a = 1.5$ ,  $b = 2$ ,  $d = 1$ ,  $c = 0.1$ ,  $\gamma_A = 1$ ,  $\gamma_a = \beta_a - 1 = 0.5$  and  $m = 2$  on the left and  $\beta_A = 2$  on the right.

we can show that  $\omega(A,a)^{-1}$  is a non-increasing function with respect to  $m$  (see Fig. 2.3, right). With these observations, our first conclusion is that a large migration rate seems to strengthen the assortative mating.

Now, considering  $m_A$  and  $m_a$  separately and assuming  $\beta_A > \beta_a$ , we show that  $\omega(A,a)^{-1}$  is a decreasing (resp. increasing) function with respect to  $m_A$  (resp.  $m_a$ ). The increase with respect to  $m_A$  highlights again the similarity of the migration and sexual preference effects. However, the decrease with respect to  $m_a$  is more surprising and highlights that the migration has not the same exact effect as the sexual preference but it implies more involved behavior. This is confirmed by the study of the basins of attraction (see Figure 2.1) and by the study of the minimal number of individuals needed for invasion. More precisely, we computed the minimal number of  $A$ -individuals such that they can survive (or even invade) in an environment of two patches filled with  $\zeta_a$   $a$ -individuals initially. This minimal number decreases when  $\beta_A$  increases. Moreover, if  $\beta_A$  and  $m$  are sufficiently large, the  $A$ -population replaces the resident  $a$ -population in both patches. This suggests again that individuals with a higher mating preference have a selective advantage.

Then comparing these minimal numbers for different values of  $m$ , we again conclude that the migration has a more intricate behavior, in the sense that it may favor the allele with the weakest mating preference in some cases.

**Ideas of proof.** The proof again relies on two phases of interest: a first phase of mean field dynamics and a stochastic phase when some of the populations are microscopic. However, the study of the dynamics of the system (2.3.3) is entirely different from the symmetrical case. Indeed, Lyapounov functions were not found in this case. We thus started with the study of the system without migration, then used a perturbation method to make  $m_A$  and  $m_a$  grow up and deduce results for some positive migration parameters. This method relies on the Implicit function Theorem, the fine study of the stable and unstable manifolds of all

the equilibria of the system without migration, and the continuity with respect to  $m_A$  and  $m_a$  of the flow of (2.3.3).

### 2.3.2 Diploid population

Finally, we present an extension of the model to a diploid population. Precisely, we focus on the symmetrical case and consider a population of diploid individuals characterized by a single phenotype controlled by their genotype at one bi-allelic locus ( $A$  and  $a$ ). The set of characteristics of this model is thus

$$\mathcal{G} = \{(AA, 1), (Aa, 1), (aa, 1), (AA, 2), (Aa, 2), (aa, 2)\}.$$

To study the effect of dominance on population differentiation, we compare two opposite scenarios: complete co-dominance and complete dominance. Dominance observed in natural populations can differ from these extreme cases, however by studying the limits of the dominance continuum we expected to cover the possible population dynamics.

In the complete dominance scenario, individuals with genotypes  $AA$  and  $Aa$  have the same phenotype,  $A$ , whereas individuals with genotype  $aa$  have the phenotype  $a$ . In the co-dominance scenario, heterozygotes express an intermediate phenotype between the two homozygotes and we thus consider two possibilities. (1) Preference expressed towards heterozygotes will be intermediate between assortative and disassortative mating (COD1) ( $\beta$  for pairs  $(AA, AA)$ ,  $(aa, aa)$  and  $(Aa, Aa)$ ,  $(\beta + 1)/2$  for pairs  $(AA, Aa)$  and  $(aa, Aa)$ , and 1 for pairs  $(AA, aa)$ ) and migration rate varies accordingly because the decision to leave the patch depends on the lack of preferred partners in the patch. (2) Heterozygotes express no preference towards any partners (COD2), the preference parameter is thus  $\beta$  for pairs  $(AA, AA)$  and  $(aa, aa)$ , and 1 for all other pairs. Because of this lack of preference, heterozygotes have no reason to discriminate against the individuals in their patch and thus do not migrate.

We introduce parameters  $p_\beta(\mathbf{g}, \mathbf{g}')$  that model the preference between two individuals with genotypes  $\mathbf{g}$  and  $\mathbf{g}' \in \{AA, Aa, aa\}$ . In other words, when an  $\mathbf{g}$ -individual reproduces as first parent (choosing one) with a second parent of genotype  $\mathbf{g}'$ , the probability that they actually mate and produce an offspring is proportional to  $p_\beta(\mathbf{g}, \mathbf{g}')$ . These preferences differ according to the model considered (see Table 2.1) and always belong to  $[1, \beta]$ . These parameters also influence the migration rates which are the product of these parameters with the proportion of unpreferred mates in the patch and with the total migration rate  $m$ .

	(COD1)	(COD2)	(DOM)
$p_\beta(AA, AA)$	$\beta$	$\beta$	$\beta$
$p_\beta(AA, Aa)$	$(\beta + 1)/2$	1	$\beta$
$p_\beta(Aa, Aa)$	$\beta$	1	$\beta$
$p_\beta(aa, Aa)$	$(\beta + 1)/2$	1	1
$p_\beta(aa, aa)$	$\beta$	$\beta$	$\beta$
$p_\beta(AA, aa)$	1	1	1

Table 2.1 – Preference functions assuming dominance hypothesis (COD1), (COD2) and (DOM).

Using numerical simulations and theoretical analysis, we identify two main conclusions.

The first one concerns the stability of the states where a type remains in the first patch and the other type remains in the second patch, corresponding to the modeling of differentiated populations. Contrary to the haploid case, these states are not stable as soon as the total migration rate  $m$  is too high. This result holds in all cases (co-dominant and dominant), but it is more substantial in co-dominant cases. This may be explained by the fact that the migration of heterozygotes has a major impact on the population behavior.

The second main point of interest concerns the existence of polymorphic equilibria. This persistence of polymorphisms within populations exists in co-dominant and dominant models but conclusions on this existence are slightly different. In the case of co-dominant models, polymorphic equilibria exist only for model (COD2). We conjecture that this polymorphism can emerge and be maintained through a trade-off between growth and migration, since migration persists in the limit. This limiting states are not possible in the other co-dominance case (model COD1), where heterozygotes can migrate and are half as preferred by homozygotes. Heterozygote behavior is thus a key parameter in the dynamics of population differentiation in mating traits.

Polymorphic equilibria can also emerge in the dominant case. However, the populations are almost phenotypically monomorphic in both patches. Indeed, heterozygotes express the preferred phenotype and therefore do not suffer from mate rejection in the patch where there are maintained, moreover they migrate at very low rate. This highlights how dominance may modulate spatial segregation and the emergence of well-separated species.

Finally, we conclude that the differences observed between haploid and diploid models highlights the need to consider the effect of ploidy in spatially-structured models of trait evolution. Indeed the presence of intermediate phenotypes can interfere in the differentiation process.

## 2.4 DIVERSITY AND DISASSORTATIVE MATING

In this section, we study mechanisms favoring the emergence and the persistence of polymorphism in loci targeted by disassortative mating preference. It corresponds to the article [28] written in collaboration with Camille Coron, Manon Costa, Violaine Llaurens and Charline Smadi. Our aim here is once again to give conditions for the emergence of diversity, however we focus on the possibility to observe such diversity on a same location and we are interested in disassortative mating for this part.

To this aim, we get back to haploid population, with all individuals living in a same patch and characterized by their allele at a single locus  $A$ . The set of characteristics is thus reduced to the  $k$  possible alleles

$$\mathcal{G} = \{1, 2, \dots, k\}.$$

Using the notation given in Equation (2.1.3), we have that  $p_{j,\ell}^i = 0$  as soon as  $i \notin \{j, \ell\}$ , and  $i \sim j$  for all  $i, j \in \mathcal{G}$  since there is a unique location for all individuals. In order to simplify



the study of the model, we introduce the parameter

$$b := \inf \left\{ \{ \beta_i p_{i,j}^i + \beta_j p_{j,i}^i \}_{i \neq j \in \{1, \dots, k\}}, \{ \beta_i p_{i,i}^i \} \right\},$$

called birth rate. We assume that  $b > 0$ , implying that it is impossible of having strict genetic incompatibilities between some pair of alleles. Adaptation can be however done to consider  $b = 0$ . For  $(i, j) \in \{1, \dots, k\}^2$ , we also introduce

$$s_{ij} := \frac{\beta_i p_{ij}^i + \beta_j p_{ji}^i}{b} - 1, \text{ if } i \neq j \quad \text{and} \quad s_{ii} := \frac{\beta_i p_{ii}^i}{b} - 1,$$

For each  $i, j \in \{1, \dots, k\}$  the parameter  $s_{ij}$  may be interpreted as the selective advantage of a pair of parents with genotypes  $i$  and  $j$  respectively. By construction  $s_{ij} = s_{ji}$  is positive or null. Note that the condition  $s_{ij} = 0$  does not correspond to a strict reproductive incompatibility of the pair  $(i, j)$  but to case where the pair  $(i, j)$  has the minimal birth rate in the population. Under large population assumption, the limiting dynamical system (2.1.4) can be rewritten as

$$\dot{z}_i(t) = z_i(t) \left( b \sum_{j=1}^k (1 + s_{ij}) \frac{z_j(t)}{z(t)} - d - cz(t) \right). \quad (2.4.1)$$

We assume that  $b \geq d > 0$  to ensure that any monomorphic population is able to maintain itself. Our analysis could be extended to avoid this assumption. However results and numerical simulations are more intricate.

We first give general results on this model concerning the possibility to obtain and maintain a large number of alleles. These conditions are based on the selective advantage of each pair of genotypes, i.e. on the following matrix  $M$  of selective advantages:

$$M := \begin{pmatrix} s_{11} & s_{12} & s_{13} & \dots & s_{1k} \\ s_{12} & s_{22} & s_{23} & \dots & s_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ s_{1k} & s_{2k} & \dots & s_{k-1,k} & s_{kk} \end{pmatrix}. \quad (2.4.2)$$

We say that a vector is positive ( $> 0$ ) if all its coordinates are positive and we state the following proposition.

**Proposition 2.4.1.** *Assume that  $\det(M) \neq 0$  and*

$$M^{-1} \mathbf{1} > 0, \quad \text{where} \quad \mathbf{1} = \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix}. \quad (2.4.3)$$

*The System (2.4.1) admits a unique positive equilibrium*

$$Z^* := \frac{1}{c} \left( b + \frac{b}{\mathbf{1}^T M^{-1} \mathbf{1}} - d \right) \frac{M^{-1} \mathbf{1}}{\mathbf{1}^T M^{-1} \mathbf{1}} \quad (2.4.4)$$

*where  $\mathbf{1}^T$  is the transpose of vector  $\mathbf{1}$ .*

*Furthermore, this equilibrium is globally asymptotically stable if and only if the matrix  $M$  has exactly 1 positive eigenvalue and  $k - 1$  negative eigenvalues.*

Note that the condition depends only on the disassortative advantage parameters  $s_{ij}$ , that represent the reproductive success of the different allelic pairs (this is true since  $b > d$ ). Using this proposition, we investigate particular cases to derive more specific conditions.

- For  $k = 2$ , two alleles  $A_1$  and  $A_2$  can be maintained in a population if and only if  $s_{11} - s_{12}$  and  $s_{22} - s_{12}$  are both negative, *i.e.* when disassortative matings produce more offspring than both assortative ones (result well documented, see [69]).
- For  $k = 3$  and with strict disassortative advantage ( $s_{ii} = 0$ ), the three alleles are maintained as soon as

$$s_{12} < s_{13} + s_{23}, \quad s_{13} < s_{12} + s_{23}, \quad s_{23} < s_{12} + s_{13}. \quad (2.4.5)$$

The three alleles are thus maintained when none of the parental pairs has a greater advantage than the sum of the advantages of the two other possible pairs of parental alleles. This condition was also identified as a necessary condition in the seminal paper of Lewontin [90] (here we prove also the sufficiency) and it is true for instance when  $s_{12} = s_{13} = s_{23} = s > 0$ .

- In the case of  $k$  symmetrical alleles (disassortative crosses are all characterized by the same advantage, and similarly for the assortative crosses), *i.e.*  $s_{ii} = \rho$  and  $s_{ij} = s$  if  $i \neq j$ , all alleles are maintained as soon as  $s > \rho$ .

Finally, this proposition highlights that the conditions for the persistence of a given level of allelic polymorphism at locus  $A$  depend on the relative reproductive advantages of disassortative *vs.* assortative crosses, but also on the relative reproductive success of the different disassortative pairs.

**Introduction of a new allele.** Proposition 2.4.1 also allows us to investigate the conditions for the emergence of a new arising allele, called mutant, in a resident population at equilibrium formed with pre-existing alleles. We thus consider a population with  $k$  alleles whose disassortative advantage matrix is denoted by  $M$  as previously. We assume that  $M$  satisfies the conditions of Proposition 2.4.1 (Equation (2.4.3)), *i.e.* the  $k$  alleles are maintained, as long as no mutation appears.

The mutant is characterized by new disassortative advantages  $S^T = (s_{k+1,1}, s_{k+1,2}, \dots, s_{k+1,k})$  and  $\sigma = s_{k+1,k+1}$ . We prove that the mutant invades if and only if

$$S^T M^{-1} \mathbf{1} > 1. \quad (2.4.6)$$

In this case, all alleles are maintained (mutant and resident alleles) as soon as the equilibrium with these  $k + 1$  alleles exists, *i.e.* if and only if

$$\bar{M}^{-1} \mathbf{1} > 0 \quad \text{with } \bar{M} = \begin{pmatrix} M & S \\ S^T & \sigma \end{pmatrix}.$$

In the general case, we do not know the long time behavior of the population when the equilibrium with all  $k + 1$  alleles does not exist. The dynamics can be quite involved; using

simulations, we can observe the loss of one or even several resident alleles after the introduction of the mutant one. The case  $k = 2$  and  $s_{11} = s_{22} = s_{33} = 0$  can however be entirely detailed:

- Either condition (2.4.5) holds, and the three alleles will always coexist, whatever the order of appearance of the different alleles.
- or for a given pair  $i, j \in \{1, 2, 3\}$ ,  $s_{ij} \geq s_{ik} + s_{kj}$ . Then an analysis of the dynamical system shows that this condition prevents the increase of polymorphism. More precisely alleles  $i$  and  $j$  will be maintained, while allele  $k$  becomes extinct, regardless of whether the indices  $i, j$  and  $k$  represent the mutant or the residents.

**Link with genetic distance.** In natural populations, the disassortative advantage is probably linked to genetic differences between pairs of alleles ([117]). We thus consider an extension of the general model by specifying the link between the selective advantage of the disassortative mating pairs with their genetic distance. To this aim, the genetic structure is modeled as the set of possible alleles  $\{0, 1\}^L$ , where  $L$  is the number of sites where mutations can occur in the locus  $A$  (Figure 2.4(a)). Then the selective advantage of the disassortative cross between two alleles  $x = (x_1, \dots, x_L) \in \{0, 1\}^L$  and  $y = (y_1, \dots, y_L) \in \{0, 1\}^L$  is defined by

$$s_{xy} = \left( \sum_{i=1}^L \mathbf{1}_{x_i \neq y_i} \right)^\alpha, \quad (2.4.7)$$

with  $\alpha > 0$ .  $\alpha$  thus modulates the relationship between this selective advantage and the genetic distance, corresponding to the number of different alleles (see Figure 2.4(b)).

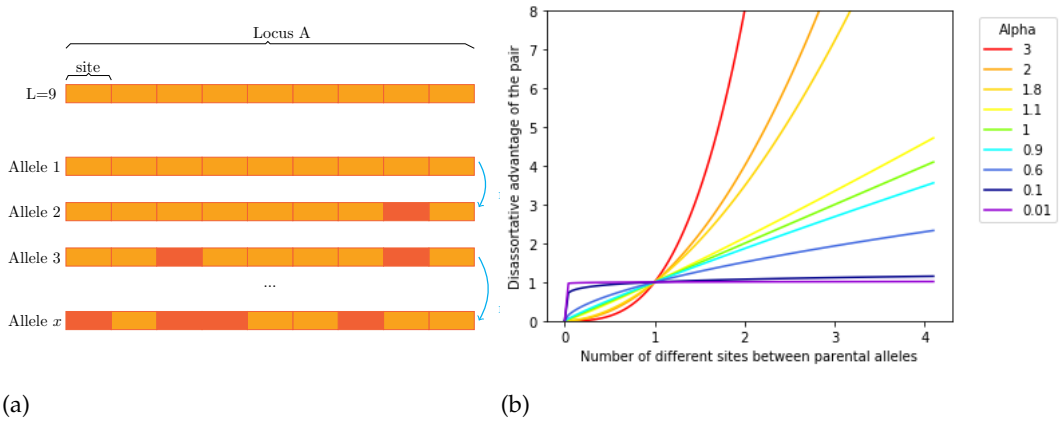


Figure 2.4 – Mutation sizes and their effects on the disassortative advantage. (a): The locus  $A$  contains  $L$  sites where mutations can occur. We model either point mutations, whereby one mutation leads to a change at a single site or other mutation kernels, where a mutation event can simultaneously affect several sites within the locus  $A$ . (b): relationship between the number of differing sites between both parents and the disassortative advantage. The parameter  $\alpha$  determines the shape of the relation.

We investigate this model and find very different behavior depending on the value of  $\alpha$ , i.e. for  $\alpha < 1$  and for  $\alpha > 1$ .

For  $\alpha < 1$ , we prove that any resident population with two alleles  $A_1$  and  $A_2$  can be invaded by any new mutant  $A_3$  at distance 1 of either  $A_1$  or  $A_2$  (i.e. in the case of point mutations) and will lead to a population with 3 alleles. Moreover, the population with all possible alleles maintained exists and is (globally) stable. We have no theoretical proof that the successive introduction of point mutations may lead to a population with all possible alleles. However, simulations seem to support this conjecture (see Figure 2.5). We thus

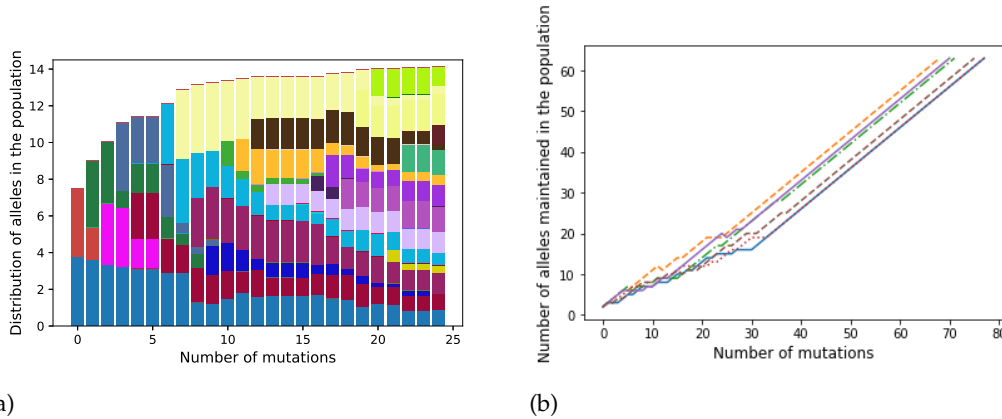


Figure 2.5 – Evolution of the number of alleles maintained in the population, assuming point mutations and convex shape of the function determining the fitness of allelic pairs ( $\alpha \leq 1$ ). From an initial population with two alleles, we numerically introduce successive mutations and track their invasion success through time. (a) shows the distribution of alleles in the population through time. Each color corresponds to a given allele and the height of the bar is the number of individuals carrying each allele within the population at a each time. (b) gives the number of alleles maintained at equilibrium after each mutation until the total number of alleles is reached. Each line corresponds to a different numerical simulation ( $n = 6$ ). Here  $L = 6$  and  $\alpha = 0.6$  such that there are  $2^6 = 64$  possible alleles.

conclude that when  $\alpha < 1$ , the specific shape of (2.4.7) may thus stabilize the polymorphism, by preventing large variations in disassortative advantages among co-existing alleles.

For  $\alpha \geq 1$ , the dynamics of invasion is very different. Firstly, the population with all possible alleles maintained exists but is unstable.

Then in the case of 3 alleles, when introducing a new mutant  $A_3$  at distance 1 of one of the two resident alleles  $A_1$  or  $A_2$  only two alleles will remain. These correspond to the most different alleles, regardless of whether there are the mutant or the resident alleles. Hence, introducing successive point mutations, the population will remain composed of two alleles with a larger and larger genetic distance. The evolution ends when the most differentiate alleles are present, i.e. alleles  $(0, \dots, 0)$  and  $(1, \dots, 1)$ . Moreover, such a population cannot be invaded by any new allele whatever the genetic distance of the mutant allele.

However, a modification of the the mutation kernel changes drastically the conclusion. We develop a specific example showing that, with a mutation kernel allowing large mutations (i.e. more than one site at a time), coexistence of more than two alleles can be observed in the case where  $\alpha \geq 1$  and  $L = 3$ . Simulations were also performed to explore other pa-

rameters and seem to confirm this conclusion. This study highlights the importance of the mutation kernel when studying evolutionary processes.

**Ideas of proofs.** The proof of the main result, Proposition 2.4.1 relies on two steps. We first specify the positive equilibrium when it exists and then prove an equivalence between its local stability and the condition on the eigenvalues of  $M$ , by studying the Jacobian of the system and by using tools of linear and bilinear algebra. We finally prove the global stability by specifying a Lyapounov function of the form

$$V(\mathbf{z}) := \sum_{\ell=1}^k \left( \frac{z_{\ell}}{z} - \frac{z_{\ell}^*}{z^*} \ln \left( \frac{z_{\ell}}{z} \right) \right) = 1 - \sum_{\ell=1}^k \frac{z_{\ell}^*}{z^*} \ln \left( \frac{z_{\ell}}{z} \right), \quad (2.4.8)$$

where  $\mathbf{z}^*$  is the positive equilibrium given in (2.4.3), and using again bilinear algebra to prove the negativeness of its derivative. When  $k = 3$ , the computations are easier and everything can be specified using this same Lyapounov function.

To find the condition of invasion of a new mutant allele (2.4.6), we simply compute the Jacobian matrix of the dynamical system of dimension  $k + 1$  around the point  $(\mathbf{z}^*, 0)$ . To prove that this condition is equivalent to the stability of the equilibrium with  $k + 1$  alleles as soon as it exists, we use again tools of linear and bilinear algebra, as Cramer's rule, the Eigenvalue Interlacing Theorem and the Schur complement.

## 2.5 PERSPECTIVES

Understanding the relations between sexual preferences, and more generally sexual reproduction, and the emergence and preservation of diversity remain a central question in evolutionary biology. The framework developed in this chapter is particularly adapted to explore how the different evolutionary forces (natural selection, sexual selection..) interact and to quantify their effects.

I am continuing to investigate these questions and related ones in collaboration with Camille Coron, Manon Costa, Violaine Llaurens and Charline Smadi. For example, we are actually interested in understanding and quantifying how family structures, emerging in some species with sexual reproduction, affect species diversity in general.

# EXTINCTION TIMES OF CONTINUOUS STATE BRANCHING PROCESSES WITH COMPETITION IN RANDOM ENVIRONMENT

# 3

## 3.1 INTRODUCTION AND MODEL

In this chapter, I present different type of population dynamics models based on *continuous state branching processes* (CSBP). These models are continuous time Markov processes with càdlàg (right continuous with left limits) paths and values in  $[0, \infty)$ . They were introduced by Jirina [63] and studied by many authors since then. They represent the dynamics of a large population in the sense that they are proved to be scaling limits in time and space of Galton-Watson processes (see Lamperti [75]). In this sense, the value represents some density of the population. However, despite being in a large population limit, the dynamics of such a population remain stochastic.

Numerous generalizations have been developed around this model. In this chapter, we are particularly interested in the addition of a term representing competition between individuals of the population, as well as a term representing an exogenous random environment acting on the population.

CSBP with competition have been considered by several authors. In [74], Lambert introduced and studied the so-called logistic branching process, where the competition has a quadratic form. More general competition mechanisms were considered since then (see for example [6, 103]).

CSBP in random environment have also been intensively studied since they were introduced by Smith and Wilkinson [116] (see for instance [2, 9]).

More recently, both types of generalizations have been combined and CSBP with competition in a random environment have become the focus of interest. In particular, Pardo and Palau [101] introduced them as strong solutions of stochastic differential equations (SDE). And Bansaye et al. [7] proved that they also arise as the scaling limit of discrete population models in random environment.

More precisely, continuous state branching processes (CSBP) with competition in random environment can be defined as the unique strong solution (up to explosion) of the following

SDE

$$\begin{aligned} Z_t = Z_0 + \int_0^t bZ_s ds - \int_0^t g(Z_s) ds + \int_0^t \sqrt{2\gamma^2 Z_s} dB_s^{(b)} + \int_0^t Z_{s-} dS_s^{(e)} \\ + \int_0^t \int_{[1,\infty)} \int_0^{Z_{s-}} zN^{(b)}(ds, dz, du) + \int_0^t \int_{(0,1)} \int_0^{Z_{s-}} z\tilde{N}^{(b)}(ds, dz, du). \end{aligned} \quad (3.1.1)$$

$g$  is a function on  $[0, \infty)$  with  $g(0) = 0$ .  $B^{(b)}$  is a standard Brownian motion.  $N^{(b)}$  is a Poisson random measure defined on  $\mathbb{R}_+^3$  independent from  $B^{(b)}$  and with intensity measure  $ds\mu(dz)du$ .  $\mu$  is a measure concentrated on  $(0, \infty)$  satisfying

$$\int_{(0,\infty)} (1 \wedge z^2) \mu(dz) < \infty. \quad (3.1.2)$$

$\tilde{N}^{(b)}$  is the compensated measure associated to  $N^{(b)}$ .  $S^{(e)}$  is a Lévy process with no negative jumps smaller than  $-1$  which is independent of  $B^{(b)}$  and  $N^{(b)}$ . Precisely,  $S^{(e)}$  can be written as follows

$$S_t^{(e)} = \sigma B_t^{(e)} + \int_0^t \int_{(-1,1)^c} (e^z - 1) N^{(e)}(ds, dz) + \int_0^t \int_{(-1,1)} (e^z - 1) \tilde{N}^{(e)}(ds, dz), \quad (3.1.3)$$

where  $\sigma \geq 0$ ,  $B^{(e)} = (B_t^{(e)}, t \geq 0)$  is a standard Brownian motion and  $N^{(e)}$  denotes a Poisson random measure, independent of  $B^{(e)}$ , taking values on  $\mathbb{R}_+ \times \mathbb{R}$  with intensity  $ds\pi(dz)$  satisfying

$$\int_{\mathbb{R} \setminus \{0\}} (1 \wedge z^2) \pi(dz) < \infty, \quad (3.1.4)$$

and  $\tilde{N}^{(e)}$  denotes the compensated version of  $N^{(e)}$ .

The existence and uniqueness of the SDE (3.1.1) were established for example by Palau and Pardo in [101]. We studied the solution to (3.1.1) up to explosion, which means that  $\infty$  and  $0$  are thus considered here as absorbing boundaries.

When  $g = 0$  and  $S^{(e)} \equiv 0$ , the solution corresponds to the CSBP (see for example [45] or [73] for a review on the subject). It represents the dynamics of a large number of individuals without interactions and it is generally referred to as the branching mechanism, when dealing with CSBP with interactions. The law of this CSBP is fully characterized by the triplet  $(b, \gamma, \mu)$  or by the function

$$\psi(\lambda) = -b\lambda + \gamma^2 \lambda^2 + \int_{(0,\infty)} \left( e^{-\lambda u} - 1 + \lambda u \mathbf{1}_{\{u < 1\}} \right) \mu(du). \quad (3.1.5)$$

This function is connected to the Laplace transform of the CSBP, which we denote by  $(Y_t, t \geq 0)$ , as

$$\mathbb{E}[e^{\lambda Y_t} | Y_0 = x] = e^{-xu_t(\lambda)} \quad \text{for all } (x, t, \lambda) \in \mathbb{R}_+^3,$$

with  $u$  the unique solution to the differential equation

$$\partial_t u_t(\lambda) + \psi(u_t(\lambda)) = 0, \quad u_0(\lambda) = \lambda.$$

The term including function  $g$  represents competition between individuals. We assume that  $g$  is positive and increasing. In section 3.3, I will focus on the case where  $g(x) = cx^2$  representing a quadratic competition.

Finally, the Lévy process  $S^{(e)}$  models the external environment. Note that the environment can have both positive and negative effects on the dynamics of the population. Moreover, each event affects a randomly selected proportion of individuals in the same way. In Section 3.3, I will focus on the case where  $S^{(e)}$  is a Brownian motion.

In the two following sections, I present results on the extinction time and long-time behavior of CSBP with competition in random environment. They correspond to the work in [85, 86] written in collaboration with Juan Carlos Pardo.

### 3.2 LÉVY ENVIRONMENT

In this section, I give results on time extinction and coming down from infinity for general terms  $g$  and  $S^{(e)}$  under the following assumptions.

**Assumption 3.2.1.** *The measure  $\mu$  satisfies the integral condition*

$$\int_{(0,\infty)} (z \wedge z^2) \mu(dz) < \infty.$$

*The function  $g \in C^1$  is non-negative, non-decreasing and it is not the null function. Moreover, we assume that*

$$\int_0^\infty \frac{1}{g(z)} dz < \infty. \quad (3.2.1)$$

These are technical assumptions. However, the assumption on  $\mu$  allows us to compensate large jumps and to compute the term  $-\psi'(0+) = b + \int_1^\infty z \mu(dz)$  that represents the Malthusian parameter of the exponential growth. Assumptions on  $g$  confirm that it truly represents a competition, which increases with the size of the population, and in particular (3.2.1) ensures that it cannot be a linear term.

In order to state the main result of this section, let us recall the definition of a strong Feller process. We denote the sets of bounded measurable and bounded continuous functions on  $[0, \infty)$  by  $\mathcal{B}_b([0, \infty))$  and  $C_b([0, \infty))$ , respectively. A process is a  $C_b$ -Feller, if its semigroup  $(P_t, t \geq 0)$  satisfies the following points

- i) for any  $f \in C_b([0, \infty))$ ,  $P_t f \in C_b([0, \infty))$ , for  $t \geq 0$ ; and
- ii) for any  $f \in C_b([0, \infty))$  and  $x \geq 0$ ,  $P_t f(x) \rightarrow f(x)$ , as  $t$  goes to 0.

If in addition, for any  $f \in \mathcal{B}_b([0, \infty))$ ,  $P_t f \in C_b([0, \infty))$ , for  $t \geq 0$ , the process is said to be strong Feller (see Kallenberg [66] for example).



Following ideas of He et al. [57] and Le [80], we give sufficient condition such that the process, solution to (3.1.1), becomes extinct and *comes down from infinity*. More precisely, the result is the following one.

**Theorem 3.2.2.** *Assume that Assumptions 3.2.1 hold and that Grey's condition holds, i.e.*

$$\int^{\infty} \frac{d\lambda}{\psi(\lambda)} < \infty, \quad (3.2.2)$$

then

$$\sup_{x>0} \mathbb{E}_x[T_0] < \infty, \quad (3.2.3)$$

Moreover the process  $Z$  comes down from infinity, in the sense that  $\infty$  is a continuous entrance point, i.e.

$$\lim_{M \rightarrow \infty} \lim_{x \rightarrow \infty} \mathbb{P}_x(T_M < t) = 1 \quad \text{for all } t > 0,$$

where  $T_M = \inf\{t \geq 0 : Z_t \leq M\}$  (assuming that  $\inf\{\emptyset\} = \infty$ ). Finally, the process is strong Feller and it can be extended into a  $C_b$ -Feller process on  $[0, \infty]$ , i.e. its semigroup satisfies also conditions (i) and (ii) above for any function  $f \in C_b([0, \infty))$  with a limit at  $\infty$ .

In other words, the condition on the competition parameter (3.2.1) is so strong that the process becomes extinct regardless the long time behavior of the environment. (3.2.3) was also obtained by Le [80] for fixed environments using similar conditions for the competition mechanism  $g$ .

Grey's condition (3.2.2) is equivalent to the extinction of the CSBP associated to the branching mechanism (without competition nor environment) with positive probability. It is thus natural to find it in this setting.

Notice that a linear term could be added in the definition of  $S^{(e)}$  in (3.1.3). All results would be identical since from a mathematical point of view, this is equivalent to a change in the value of  $b$ .

Similarly, we may relax the assumptions on  $g$  by only assuming that there exist  $\theta \geq 0$  such that the map  $x \mapsto (\theta - b)x + g(x)$  is non-negative and non-decreasing. The same calculations could be then performed by modifying the values of  $b$  and  $g$  and recovering the current setting. Indeed, the condition (3.2.1) for the new function  $g$  then follows from Lemma 2.3 in Le and Pardoux [122].

**Ideas of proof** We first proved coupling results, namely the solution to (3.1.1) can be stochastically dominated by a process satisfying a similar SDE but without competition and starting eventually from a higher level. Using these couplings and the results in He et al. [57], we deduced that the process, solution to (3.1.1), gets extinct with positive probability. Adding few computations on  $|\mathbb{P}_t f(y) - \mathbb{P}_t f(x)|$ , we proved the Feller properties.

Our second step was to show that there exists  $M > 0$  such that

$$\sup_{x \geq 0} \mathbb{E}_x[T_M] = \sup_{x \geq M} \mathbb{E}_x[T_M] < \infty.$$

To this aim, we applied Itô formula to the process  $Z$  and the function

$$G(z) = \int^z \frac{dw}{g(w)}.$$

In addition with the coupling arguments and the fact that such process starting from  $M$  has a positive probability to get extinct, Equation (3.2.3) was proved. The remaining properties, namely  $\infty$  is an entrance point and the semigroup can be extended into a  $C_b$ -Feller process on  $[0, \infty]$ , were directly deduced from it.

### 3.3 LOGISTIC COMPETITION IN A BROWNIAN ENVIRONMENT

This section is devoted to the case where

$$g(z) = cz^2 \quad \text{and} \quad S^{(e)} = \sigma B^{(e)}. \quad (3.3.1)$$

These assumptions enable us to obtain much more detailed results than in the previous section concerning the long-time behavior and extinction time of the solution to (3.1.1). These results are mainly based on a Lamperti transformation, similar to that obtained by Lambert [74] in the case of the logistic Feller process and by Ma et al. [93] in the case of a CSBP with competition. In our case, this transformation links, by using a time change, the solution (3.1.1) to a Feller diffusion which is perturbed by a Lévy process. This Lévy process corresponds to the branching mechanism. If it is a subordinator, then the process obtained through the Lamperti transformation turns out to be a CB-process with immigration for which many results are known.

On the basis of these remarks, notice that the results obtained will differ depending on whether the branching mechanism is associated with a subordinator or not. The results are therefore summarized and presented according to these two cases.

#### 3.3.1 Case associated with a subordinator

In this section, we are interested in characterizing the long time behavior of the process under the particular case of a branching mechanism associated with the Laplace transform of a subordinator, i.e.

$$\psi(z) = -\delta z - \int_{(0, \infty)} (1 - e^{-zu}) \mu(du), \quad (3.3.2)$$

where

$$\int_{(0, \infty)} (1 \wedge u) \mu(du) < \infty \quad \text{and} \quad \delta := b - \int_{(0, 1)} u \mu(du) \geq 0.$$

We also introduce, the function

$$\omega(x) = cx + \frac{\sigma^2 x^2}{2},$$

where we recall that  $\sigma > 0$  is the variance of the Brownian environment and  $c \geq 0$  is the competitive parameter. Notice that the case without competition (i.e.  $c = 0$ ) can also be considered and results on this case are also provided.

Our first result gives a necessary and sufficient condition under which the process  $Z$  is conservative, i.e. that  $Z$  does not explode at finite time a.s.

**Theorem 3.3.1.** *The process  $Z$ , with branching mechanism given by (3.3.2), is conservative if and only if*

$$\mathcal{I} := \int_0^1 \frac{1}{\omega(z)} \exp \left\{ \int_z^1 \frac{\psi(u)}{\omega(u)} du \right\} dz = \infty.$$

In this setting, we also have the following identity for the total population size of the process  $Z$  up to time  $T_a = \inf\{t \geq 0 : Z_t \leq a\}$ , the first hitting time of  $Z$  at  $a$ .

**Proposition 3.3.2.** *For every  $\lambda > 0$  and  $x \geq a \geq 0$ , we have*

$$\mathbb{E}_x \left[ \exp \left\{ -\lambda \int_0^{T_a} Z_s ds \right\} \right] = \frac{f_\lambda(x)}{f_\lambda(a)},$$

with

$$f_\lambda(x) := \int_0^\infty \frac{dz}{\omega(z)} \exp \left\{ -xz + \int_\ell^z \frac{\lambda - \psi(u)}{\omega(u)} du \right\}, \quad x \geq 0,$$

where  $\ell$  is an arbitrary constant larger than 0.

Two very different long-time behaviors emerge for the process depending on the value of  $\sigma$  with respect to  $\delta$ , i.e. comparing the fluctuations due to the environment ( $\sigma$ ) with the growth of the branching mechanism ( $\delta$ ). The results are summarized below. The following theorem establishes conditions such that 0 is polar for  $Z$ , i.e. such that  $\mathbb{P}_x(T_0 < \infty) = 0$  for all  $x > 0$ . Moreover, it gives conditions for the recurrence or transience of the process that we defined using the following definition (see for instance Chapter X of [106] or Definition 1 in [35]).

**Definition 3.3.3.** *Assume that 0 is polar, the process  $Z$  is said to be recurrent if there exists  $x > 0$  such that*

$$\mathbb{P}_x \left( \liminf_{t \rightarrow \infty} |Z_t - x| = 0 \right) = 1.$$

On the other hand, the process is said to be transient if

$$\mathbb{P}_x \left( \lim_{t \rightarrow \infty} Z_t = \infty \right) = 1, \quad \text{for every } x > 0.$$

Observe that if the property of recurrence is satisfied for a particular  $x > 0$ , it is also true for all  $x > 0$ .

**Theorem 3.3.4.** *Assume that  $\sigma > 0$ .*

(1) *If  $\sigma^2 > 2\delta$ , then  $Z$  converges to 0 with positive probability, i.e*

$$\mathbb{P}_x \left( \lim_{t \rightarrow \infty} Z_t = 0 \right) > 0, \quad \text{for } x > 0.$$

*If we also assume that  $\mathcal{I} = \infty$ , then the process converges to 0 a.s.*

(2) *If  $2\delta \geq \sigma^2$  and  $c > 0$ . Then the point 0 is polar, i.e.  $\mathbb{P}_x(T_0 < \infty) = 0$  for all  $x > 0$ .*

(2a) Moreover if

$$\int_0^1 \frac{dz}{z} \exp \left\{ - \int_z^1 \int_0^\infty \frac{(1 - e^{-us})}{\omega(u)} \mu(ds) du \right\} = \infty \quad (3.3.3)$$

$Z$  is recurrent. Additionally,

(2a.i) if  $2\delta > \sigma^2$  then the process  $Z$  is positive recurrent and its invariant distribution  $\rho$  has a finite expected value if and only if

$$\int_1^\infty \log(u) \mu(du) < \infty. \quad (3.3.4)$$

(2a.ii) if  $2\delta = \sigma^2$  and (3.3.4) holds, we are able to exhibit two exclusive (but not exhaustive) conditions  $(\partial)$  and  $(\tilde{\partial})$  under which  $Z$  is either positive recurrent, or null-recurrent and convergent to 0 in probability, respectively.

(2b) Finally, if (3.3.3) is not satisfied, then  $Z$  explodes at finite time a.s. Hence, it is transient.

It is important to note that (3.3.3) is satisfied as soon as (3.3.4) holds.

The log-moment condition (3.3.4) is necessary and sufficient for the generalized Ornstein-Uhlenbeck process  $R$  to possess an invariant distribution [107]. This condition thus emerges through the Lamperti transformation.

In the cases where an invariant measure exists, it corresponds to the size-biased distribution of  $\nu$ , i.e.  $\rho(dz) = \left( \int_{(0,\infty)} s^{-1} \nu(ds) \right)^{-1} z^{-1} \nu(dz)$ , where  $\nu$  is defined by the following lemma.

**Lemma 3.3.5.** Assume that  $\sigma^2, c > 0$  and that the branching mechanism  $\psi$ , given by (3.3.2), satisfies the log-moment condition (3.3.4). Then

$$\int_{(0,\infty)} e^{-\lambda z} \nu(dz) = \exp \left( \int_0^\lambda \frac{\psi(u)}{\omega(u)} du \right), \quad \lambda \geq 0,$$

defines a unique probability measure  $\nu$  on  $(0, \infty)$  which is infinitely divisible. In addition, it is self-decomposable whenever  $\int_0^\infty \mu(dz) \leq \delta$ .

Finally, I choose not to exhibit precisely the critical case where  $\sigma^2 = 2\delta$ . Indeed, the conditions  $(\partial)$  and  $(\tilde{\partial})$  are cumbersome and unfortunately not exhaustive. However, readers can refer to the article [86] for further details.

Before turning to the opposite case, where the branching mechanism is not associated with a subordinator, I detail a brief example where the branching mechanism is such that  $\psi(z) = -c_\alpha z^\alpha$ , for  $z \geq 0$ , with  $\alpha \in (0, 1)$  and  $c_\alpha > 0$ . It corresponds to the negative of a stable subordinator. Straightforward computations lead to  $\mathcal{I}$  is finite or infinite depending on whether  $c = 0$  or  $c > 0$ . In other words, if there is presence of competition the associated process  $Z$  is conservative and moreover the process becomes extinct a.s., since  $\sigma^2$  is always positive ( $\delta = 0$  in this case). If there is no competition, the process  $Z$  explodes with positive

probability. This result is consistent with the ones of [100] where the rate of explosion was determined explicitly.

Finally, we also point out that the previous results are consistent with the behaviors found in Proposition 2.1 in [40] where  $\psi(z) = -bz$ .

### 3.3.2 Second case

Finally, we consider the case where the process  $X$  is not a subordinator and that it satisfies (3.3.4). Here, we present very different results than the ones of the previous section as we give a formula of the Laplace transform of the extinction time, in the case where there is extinction a.s.. However, conditions for extinction are not discussed here and have not been determined at the time of writing the paper associated with this section.

Precisely, we assume the following.

**Assumption 3.3.6.** *The branching mechanism  $\psi$  satisfies that there exists  $\vartheta \geq 0$  such that  $\psi(z) > 0$  for any  $z \geq \vartheta$  and it satisfies the condition (3.3.4).*

Our main result in this section provides a complete characterization of the Laplace transform of the stopping times

$$T_a = \inf\{t \geq 0 : Z_t \leq a\}, \quad \text{for } a \geq 0,$$

as long as  $T_0$  is finite a.s. and that  $c > 0$ .

To this aim, we introduce the functions

$$m(\lambda) := \int_0^\lambda \frac{\psi(u)}{\omega(u)} du \quad \text{and} \quad \mathbb{I}(\lambda) := \int_0^\lambda e^{m(u)} du, \quad \text{for } \lambda \geq 0, \quad (3.3.5)$$

From Assumptions 3.3.6,  $m$  is increasing on  $(\vartheta, \infty)$  implying that  $\mathbb{I}(\cdot)$  is a bijection from  $\mathbb{R}_+$  into itself. We denote its inverse by  $\varphi$  and a simple computation provides

$$\varphi'(z) = \exp(-m \circ \varphi(z)). \quad (3.3.6)$$

The formulation of the Laplace transform of  $T_a$  will be written using the non-negative solution  $y_\lambda$  that vanishes at  $\infty$  to the following Riccati equation

$$y'(z) = y(z)^2 - \lambda \left( \frac{\varphi'(z)}{\sqrt{\omega(\varphi(z))}} \right), \quad \text{with } \lambda > 0. \quad (3.3.7)$$

Using ideas similar to Lemma 2.1 in [74], we deduce the following Lemma on this equation.

**Lemma 3.3.7.** *For any  $\lambda > 0$ , there exists a unique non-negative solution  $y_\lambda$  to the equation (3.3.7) such that it vanishes at  $\infty$ . Moreover,  $y_\lambda$  is positive on  $(0, \infty)$ , and for any  $z$  sufficiently small or large,  $y_\lambda(z) \leq \frac{\sqrt{\lambda}\varphi'(z)}{\sqrt{\omega(\varphi(z))}}$ . As a consequence,  $y_\lambda$  is integrable at 0, and it decreases initially and ultimately.*

The main result thus gives the Laplace transform of  $T_a$ .

**Theorem 3.3.8.** *Assume that  $c > 0$  and that Assumption 3.3.6 holds. Then the function*

$$h_\lambda(x) := 1 + \lambda \int_0^\infty \frac{e^{-xz-m(z)}}{\omega(z)} e^{-\int_0^z y_\lambda(v)dv} \int_0^z e^{m(u)+2\int_0^u y_\lambda(v)dv} du dz \quad (3.3.8)$$

*is well defined and positive for any  $x > 0$  and  $\lambda > 0$  and it is a non-increasing  $C^2$ -function on  $(0, \infty)$ .*

*Furthermore, if  $\mathbb{P}_x(T_0 < \infty) = 1$ , for any  $x > 0$  then  $h_\lambda$  is also well-defined at 0 with*

$$h_\lambda(0) = \exp \left\{ \int_0^\infty y_\lambda(v)dv \right\} < \infty,$$

*and, for any  $x \geq a \geq 0$ ,*

$$\mathbb{E}_x \left[ e^{-\lambda T_a} \right] = \frac{h_\lambda(x)}{h_\lambda(a)}, \quad (3.3.9)$$

*and, for  $x > 0$ ,*

$$\mathbb{E}_x[T_0] = \int_0^\infty du e^{m(u)} \int_u^\infty \frac{e^{-m(z)}}{\omega(z)} (1 - e^{-zx}) dz. \quad (3.3.10)$$

Notice that we can combine the previous results with Theorem 3.2.2 in case of a logistic branching process satisfying assumptions 3.2.1 and 3.3.6 and Grey's condition (3.2.2). In this case, the process can be extended into a Feller process on  $[0, \infty]$  and we are able to compute the expectation of the extinction time.

$$\mathbb{E}_\infty \left[ e^{-\lambda T_a} \right] = \frac{1}{h_\lambda(a)} \quad \text{and} \quad \mathbb{E}_\infty[T_0] = \int_0^\infty du e^{m(u)} \int_u^\infty \frac{e^{-m(z)}}{\omega(z)} dz.$$

**Ideas of proof.** As explained in the main text, the proofs rely on a Lamperti transformation that links the process  $Z$  with the following process  $R$ , via a change of time,

$$dR_t = \mathbf{1}_{\{R_{r^-} > 0; r \leq t\}} dX_t - \mathbf{1}_{\{R_{r^-} > 0; r \leq t\}} cR_t dt + \mathbf{1}_{\{R_{r^-} > 0; r \leq t\}} \sigma \sqrt{R_t} dW_t, \quad (3.3.11)$$

with  $R_0 = x$ ,  $X$  a spectrally positive Lévy process with characteristics  $(-b, \gamma, \mu)$  and  $W$  a standard Brownian motion independent of  $X$ .

When  $X$  is a subordinator,  $R$  is then a Feller diffusion with immigration. We were thus able to use the results of Duhalde et al. [35] and link the long time behavior of  $R$  with the one of  $Z$ . We then focused on the case of recurrence with (3.3.4) satisfied. In order to find the possible invariant measure, we searched for measures  $\rho$  satisfying

$$\int_0^\infty \mathcal{U}f(z)\rho(dz) = 0,$$

for any  $f$  in the domain of  $\mathcal{U}$  the infinitesimal generator of  $Z$ . Focusing on functions  $f(z) = e^{-\lambda z}$  and following ideas of [74], we deduce that the Laplace transform of  $z\rho(dz)$  should be proportional to  $\exp\left(\int_0^\lambda \frac{\psi(u)}{\omega(u)} du\right)$ . We then exhibited conditions under which such measure is finite, which gave us when the process is positive recurrent or not.

The second case was also proved by adapting proofs of Lambert [74].

### 3.3.3 General competition for branching diffusion in a Brownian environment

Finally, notice that in the case of a branching diffusion in a Brownian environment, i.e. when the branching mechanism is associated with a Brownian motion with drift, we may study the process by using classical results of diffusion process. With this in mind, we gave necessary and sufficient conditions for the a.s. extinction of the process and we provided a new formulation for the Laplace transform of  $T_a$ . I refer readers to the manuscript [86] for more detailed information.

## 3.4 RECENT RESULTS AND PERSPECTIVES

Notice that the above picture is not complete, and that some cases have been left unsolved due to a lack of success in controlling them. More recently, Foucart and Vidmar [43] provided some duality results and their techniques are sufficient to complete the picture and characterize the long time behavior of the CSBP with logistic competition in a random environment in almost all cases.

Now that the picture is complete for this case, an interesting extension would be to add an explicit spatial structure to the model, echoing the work of the previous chapters. For example, starting from a study similar to [40, 58], we could extend our model to a patch environment where migration is very rapid between patches. We could then compare extinction conditions in this patchy environment with the non-spatial model. As explained in the previous chapter, studying the impact of explicit spatial structures on the evolutionary process of a population is of particular interest, as environment and space are key factors explaining species diversity.

# GENETIC COMPOSITION OF TUMOR CELL POPULATIONS

# 4

## 4.1 INTRODUCTION

More recently, I came across the study of tumor cell diversity, around a major issue: How can tumor cell diversity be used to facilitate and improve oncology treatments? The first goal is thus to develop a detailed understanding of the mechanisms at work in a tumor, of the interactions between the different cell types and their environment that includes treatment.

In many medical studies, tumors are now considered as a complex ecosystem, with different types of tumor cells interacting with each other and with their environment. A detailed understanding of this ecosystem and its development seems essential for optimizing therapies, with the following objectives:

- ◇ better identify the early phases of tumors,
- ◇ limiting doses and frequencies of drugs in case of established tumors,
- ◇ avoid relapses, by limiting as far as possible the appearance of cells resistant to known drugs.

With these objectives in mind, I developed stochastic processes that model dynamics of tumor cell populations. These cells are generally subject to relatively frequent mutations, since they are deficient in the capacity to repair DNA damages. One of the main goals is thus to identify and study the genetic composition of these populations.

This chapter corresponds to an opening chapter. In the next section, I give a first example of study where cells are submitted to a treatment and eventually escape the extinction due to a rescue event induced by the emergence of a resistant mutation. Finally, in section 4.3, I detail various perspectives corresponding to current or future projects in this fast-growing and important field.



## 4.2 RESCUED POPULATION

In this section, we aim at studying the impact of resistance acquisition on the distribution of neutral mutations in a cell population under therapeutic pressure. This section is based on the preprint [14] written in collaboration with Céline Bonnet.

More precisely, we study a population with a large initial size and composed of cancer cells sensitive to a specific treatment, i.e. its dynamics follows the one of sub-critical birth and death process whose size decreases exponentially fast. While the treatment is administered, these cells can become resistant to it through the acquisition of a mutation, producing a second population whose dynamics follows a supercritical process even under the treatment. The overall population thus may avoid extinction. Such dynamics are commonly called *rescue* dynamics and are the subject of many works using multitype branching processes as [5, 38, 61, 68, 72, 42, 99], among others. Studying these dynamics is particularly important in the context of oncology. Indeed treatments, as chemotherapy, exert significant selection pressures on cells and can thus favor the emergence of resistant cells, as justified by some works of Ollier et al. [98].

More precisely, the model of this section corresponds to a bi-type branching process indexed by a scaling parameter  $N$  and representing a population of sensitive (type 0) and resistant (type 1) cells to a given treatment. At any time  $t \geq 0$ , the number of sensitive cells is denoted  $Z_0^N(t)$  and the one of resistant cells is denoted  $Z_1^N(t)$ . Initially, the cells all carry type 0, associated with a negative growth rate. The initial state of the process is thus

$$(Z_0^N(0), Z_1^N(0)) = (N, 0).$$

Mutations towards type 1 are assumed to be rare, and lead to the survival of cells under the treatment, i.e. type 1 is associated with a positive growth rate, and thus models the acquisition of a resistance.

Cells are also subject to frequent *neutral mutations* that are accumulated by inheritance at each division. A *neutral mutation* corresponds to a mutation that has no impact on the birth and death rates of the cell.

Precisely, the process follows the following dynamics:

- Each sensitive cell divides at rate  $b_0$  and dies at rate  $d_0$ . We denote by  $\lambda_0 := d_0 - b_0 > 0$  the absolute value of the growth rate of sensitive cells.
- Each resistant cell divides at rate  $b_1$  and dies at rate  $d_1$ . Its growth rate is positive and is denoted by  $\lambda_1 := b_1 - d_1 > 0$ .
- At each division, the cell is replaced by two daughter cells:
  - ◊ Each daughter cell inherits the neutral mutations of their mother in addition to an independent random number of new neutral mutations  $N_\omega$  such that

$$\mathbb{E}[N_\omega] = \omega/2 \geq 0.$$

- ◇ Each of the two daughter cells may become resistant with probability  $\gamma/N^\alpha$ , with  $0 < \alpha \leq 1$ , independently from one another. The parameter  $\alpha$  models the rarity of the occurrence of resistances. Depending on its value the expected number of rescue events, which is of order  $N^{1-\alpha}$ , is finite or infinite.

Notice that, considering all the dynamics, the exact growth rate of sensitive cells is  $-(\lambda_0 + 2b_0\gamma/N^\alpha)$ , which is negative.

We aim at describing the distribution of neutral mutations in the resistant cell population in a limit of large initial sensitive population (i.e.  $N \rightarrow \infty$ ), rare advantageous mutations (i.e.  $\gamma/N^\alpha \rightarrow 0$ ), and at a large time, precisely after the characteristic time of extinction of sensitive cells, i.e. at

$$t_N := t \log(N), \text{ with } t > 0.$$

To characterize this distribution, we study the *site frequency spectrum* (SFS) which counts, for all  $i \in \mathbb{N}^*$  and  $t \geq 0$ , the number of neutral mutations carried by exactly  $i$  resistant cells at time  $t$ . This sequence is denoted by  $(S_i(t))_{i \in \mathbb{N}, t \geq 0}$  (see Figure 4.1 for an example). Precisely,

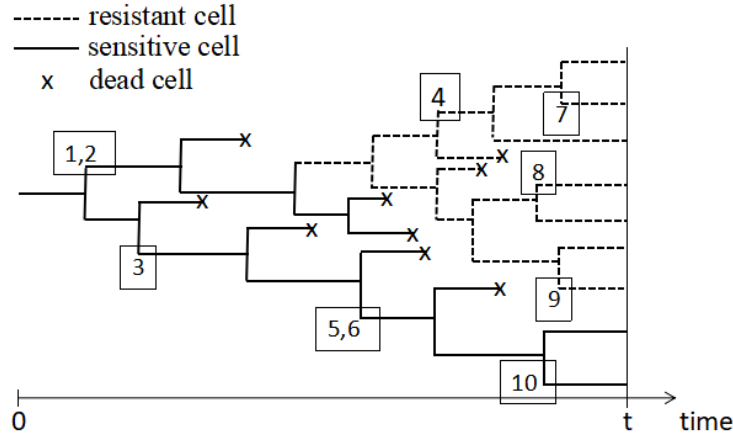


Figure 4.1 – Example of progeny starting from one sensitive cell and including neutral mutations. Solid lines correspond to sensitive genealogy, dashed lines to resistant one, cross to dead cells and framed numbers to neutral mutations appeared at each division. In this example, the SFS associated with the resistant cells is thus  $S_1(t) = 3$ ,  $S_3(t) = 1$ ,  $S_7(t) = 2$  and for all  $i \notin \{1, 3, 7\}$ ,  $S_i(t) = 0$ .

we give asymptotically-equivalent expressions of the expected SFS for finite  $i \in \mathbb{N}$  and also for  $i$  proportional to  $e^{\lambda_1 t_N}$ , i.e. proportional to the size of the resistant population at time  $t_N$ . In both cases,  $S_i^N(t)$  is computed by separating the quantity into two different contributions, similarly to what has been done in [33, 119]. These contributions are linked to the resistant cells issued from a resistant mutation event, i.e. these are resistant cells whose mother cell is a sensitive one. These specific resistant cells are called *ancestral resistant* cells in the following. Finally,  $S_i^N(t)$  is the sum of

- (1)  $\bar{S}_i^N(t)$  that counts neutral mutations that appeared in a resistant cell and
- (2)  $\underline{S}_i^N(t)$  that counts neutral mutations that appeared in a sensitive cell, and which are transmitted to resistant cells as a hitch-hiking effect ([115]).

**Number of mutations shared by a small number of cells.** Firstly, the expected SFS is computed for fixed  $i$ , i.e. it describes the number of mutations shared by a small number of resistant cells at time  $t_N$ .

**Theorem 4.2.1.** For all  $i \in \mathbb{N}$ ,  $t \geq 0$ ,

$$\mathbb{E} \left[ S_i^N(t_N) \right] \underset{N \rightarrow \infty}{\sim} I(i) \frac{2b_0\gamma\omega}{\lambda_1 + \lambda_0} N^{\lambda_1 t + 1 - \alpha}, \text{ with } I(i) := \int_0^1 \frac{1-y}{1-d_1 y/b_1} y^{i-1} dy. \quad (4.2.1)$$

As a corollary of the proof of this theorem, we deduce also

$$\mathbb{E} \left[ S_i^N(t_N) \right] \underset{N \rightarrow +\infty}{\sim} \mathbb{E} \left[ \bar{S}_i^N(t_N) \right]. \quad (4.2.2)$$

This is consistent with the intuition since mutations appeared in sensitive cells are expected to be shared by a large amount of cells and therefore not to appear in the computation of the term (4.2.1). The equivalence relation (4.2.2) is illustrated in Figure 4.2.

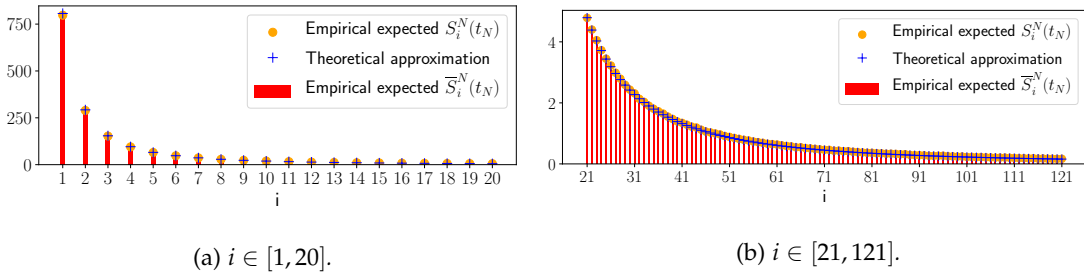


Figure 4.2 – Empirical and theoretical expectation of  $S_i^N(t_N)$  and  $\bar{S}_i^N(t_N)$  for small value of  $i$ , i.e.  $i \in [1, 20]$  in (a) and  $i \in [21, 121]$  in (b). The orange bullets and the red bars correspond to the empirical expectation over 50 000 realizations, respectively, of  $S_i^N(t_N)$  and  $\bar{S}_i^N(t_N)$ . The blue crosses correspond to the theoretical approximation given by Theorem 4.2.1.  $N = 500$ .

Approximation (4.2.1) can be seen as the contribution of each *ancestral resistant* cell multiplied by the total number of such cells appeared before  $t_N$ , which is a quantity that we were able to compute in [14]. Precisely,

$$\mathbb{E} \left[ \bar{S}_i^N(t_N) \right] \underset{N \rightarrow \infty}{\sim} \underbrace{\frac{2b_0\gamma}{\lambda_0} N^{1-\alpha}}_{\text{number of ancestral resistant cells}} \underbrace{\frac{\omega\lambda_0}{\lambda_1 + \lambda_0} I(i) N^{\lambda_1 t}}_{\text{contribution of a cell}}. \quad (4.2.3)$$

This decomposition highlights the impact of the dynamics of the sensitive cells on the expected SFS. In [52] and [22], an equivalent of the expected SFS was obtained for a birth-and-death process starting with one resistant cell. Notice that, since the resistance events all

appear in a negligible time relatively to the time of interest  $t_N$ , one may think that the contribution of an ancestral resistant cell would correspond to the one of a resistant cell living alone at time 0, i.e.  $\omega I(i)N^{\lambda_1 t}$ . However, notice that the growth rate of the resistant population modifies the constant parameter. Indeed the population growth is exponentially rapid, then starting the process a small amount of time after has already an impact. We thus conclude that the *rescue* dynamics has a significant effect on the SFS, although the order size of the approximation ( $N^{\lambda_1 t + 1 - \alpha}$ ) and its shape with respect to  $i$  are not directly impacted. The fraction  $\frac{\lambda_0}{\lambda_1 + \lambda_0}$  can be interpreted as a loss coefficient due to the rescued dynamics. When  $\lambda_0$  is large with respect to  $\lambda_1$ , the process  $Z_0^N$  is extinct quickly and mutations appear almost instantaneously, such that there is almost no loss ( $\lambda_0/(\lambda_0 + \lambda_1) \simeq 1$ ). When  $\lambda_1$  is large with respect to  $\lambda_0$ , any delay in the emergence of the *ancestral resistant cell* will have a huge impact and the loss will be large ( $\lambda_0/(\lambda_0 + \lambda_1) \simeq 0$ ).

**Mutations shared by a large number of cells.** As  $t_N$  increases with  $N$ , the order size of the total population at  $t_N$  is  $e^{\lambda_1 t_N}$  conditioned to the *rescue*. Hence, to study mutations that affect a large number of resistant cells, we compute the SFS for some  $i$  depending on  $N$  as  $i \sim N^{\lambda_1 t}$ . More precisely, the next result gives  $S_x(t)$  the number of mutations carried by a number of resistant cells larger than  $x e^{\lambda_1 t}$  at time  $t$ .

**Theorem 4.2.2.** *Let  $t \geq 0$ ,  $x \in (0, \infty]$ , and set  $S_x(t) := \sum_{i \geq x e^{\lambda_1 t_N}} S_i(t)$ . Then*

$$\mathbb{E} \left[ S_x^N(t_N) \right] \underset{N \rightarrow \infty}{\sim} b_0 \gamma \omega \lambda_1 \left( K(x) + L(x) \right) N^{1-\alpha}.$$

where

$$K(x) := \frac{2}{\lambda_0 + \lambda_1} \int_0^\infty (e^{-(\lambda_1 + \lambda_0)s} - 1) e^{\lambda_1 s} e^{-x \frac{\lambda_1}{b_1} e^{\lambda_1 s}} ds,$$

$$L(x) := \frac{1}{b_1} \int_0^{+\infty} (1 + 2b_0 s) e^{-\lambda_0 s} e^{-x \frac{\lambda_1}{b_1} e^{\lambda_1 s}} ds.$$

Contrary to Theorem 4.2.1,  $\mathbb{E}[\bar{S}_x^N(t_N)]$  and  $\mathbb{E}[\underline{S}_x^N(t_N)]$  both contribute to the asymptotically-equivalent expression of  $\mathbb{E}[S_x^N(t_N)]$ . Their contributions are represented respectively by functions  $K$  and  $L$ .

For  $x > 0$ , the theoretical approximation of  $\mathbb{E}[\underline{S}_x^N(t_N)]$  when  $N$  is large is

$$N^{1-\alpha} b_0 \gamma \omega \lambda_1 L(x) = \underbrace{\frac{2b_0 \gamma}{\lambda_0} N^{1-\alpha}}_{\substack{\text{number of} \\ \text{ancestral} \\ \text{resistant cells}}} \int_0^\infty \underbrace{\frac{\omega}{2} (1 + 2b_0 s)}_{\Delta_1} \underbrace{\frac{\lambda_1}{b_1} e^{-x \frac{\lambda_1}{b_1} e^{\lambda_1 s}}}_{\Delta_2} \underbrace{\lambda_0 e^{-\lambda_0 s}}_{\Delta_3} ds.$$

This approximation can thus be seen as the expected number of ancestral resistant cells multiplied by the contribution of one *ancestral resistant cell*. The proof of Theorem 4.2.2 also allows us to give a natural interpretation of each of the underlined terms (see Figure 4.3 for a scheme).  $\Delta_3$  corresponds to the limiting density of time of occurrence of an *ancestral resistant cell*.  $\Delta_2$  can be seen as an approximation of the probability that a cell, appeared at time  $s$ , has

more than  $xe^{\lambda_1 t_N}$  offspring at time  $t_N$ . Indeed,

$$\mathbb{P}\left(\tilde{Z}(t_N - s) > xe^{\lambda_1 t_N} \mid \tilde{Z}(0) = 1\right) \xrightarrow[N \rightarrow \infty]{} \Delta_2,$$

with  $\tilde{Z}$  a birth-and-death process with parameters  $b_1$  and  $d_1$ . Finally, the factor  $\Delta_1 = \frac{\omega}{2}(1 + 2b_0s)$  represents the mean number of mutations carried by a cell appeared at time  $s$ , which is proportional to the number of times a sensitive cell divides before becoming resistant at time  $s$ . Indeed,  $\omega/2$  is the mean number of mutations that a cell gets after one division.  $(1 + 2b_0s)$  corresponds to the mean number of divisions found in the ancestral lineage of an *ancestral resistant* cell that appeared at time  $s$ . The factor 2 in front of  $b_0$  has already been met in previous works on branching processes (see [10]).

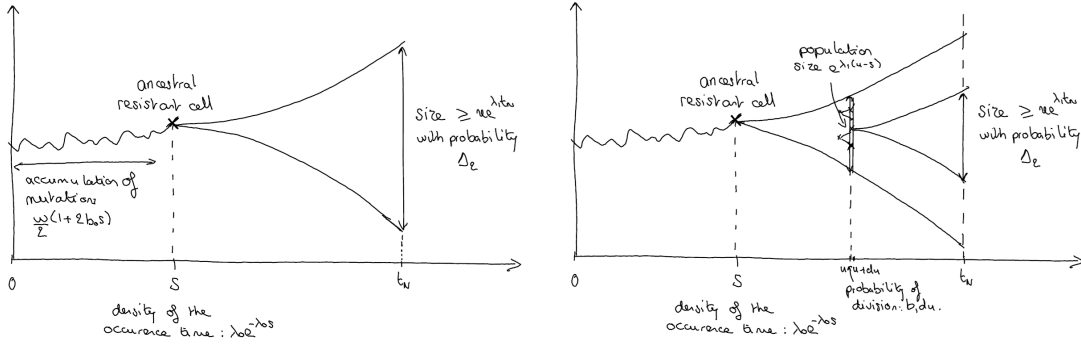


Figure 4.3 – Schemes of the computations of the number of mutations shared by a large number of cells. (Left) Scheme for the computation of  $\mathbb{E}[\underline{S}_x^N(t_N)]$ . (Right) Scheme for the computation of  $\mathbb{E}[\overline{S}_x^N(t_N)]$ .

The approximation of  $\mathbb{E}[\overline{S}_x^N(t_N)]$  can also be decomposed into several contributions to simply understand the result obtained. As previously, this approximation can be seen as the expected number of ancestral resistant cells multiplied by the contribution of one *ancestral resistant* cell, as

$$b_0 \gamma \omega \lambda_1 K(x) N^{1-\alpha} = \underbrace{\frac{2b_0 \gamma}{\lambda_0} N^{1-\alpha}}_{\text{number of ancestral resistant cells}} \int_0^\infty \underbrace{\left( \int_0^u \omega b_1 e^{\lambda_1(u-s)} \lambda_0 e^{-\lambda_0 s} ds \right)}_{\Omega_1} \underbrace{\frac{\lambda_1}{b_1} e^{-x \frac{\lambda_1}{b_1} e^{\lambda_1 u}}}_{\Delta_2} du.$$

The integral term  $\Omega_1$  represents the number of mutations, from resistant divisions, that appeared at time  $u$  (see Figure 4.3 for a scheme). Indeed, the time of emergence of an *ancestral resistant* cell has, in the limit, an exponential distribution of parameter  $\lambda_0$ . Then  $\omega$  represents the mean number of mutations due to one division,  $b_1 du$  represents the probability that a resistant cell divides in a interval  $[u, u + du]$  and  $e^{\lambda_1(u-s)}$  represents the number of progeny at time  $u$  of an *ancestral resistant* cell appeared at time  $s$ . The factor  $\Delta_2$  is the same as the one discussed above.

Finally, we gave partial results on the number of mutations carried by exactly  $i_N$  resistant cells, with  $i_N \sim e^{\lambda_1 t_N}$  when  $N \rightarrow \infty$ . We were able to derive the approximation of  $\mathbb{E}[\bar{S}_{i_N}^N(t_N)]$ , which corresponds to  $b_0 \gamma \omega \lambda_1 K'(x) N^{1-\lambda_1 t - \alpha}$ , but not the one of  $\mathbb{E}[\underline{S}_{i_N}^N(t_N)]$ . Indeed, we are not able to control the influence of the kinship events between *ancestral resistant* cells to obtain this quantity. Simulations seem to show that both contributions are of the same size order (see Figure 4.4) and that we need both computations to conclude on the value of  $\mathbb{E}[S_{i_N}(t_N)]$ .

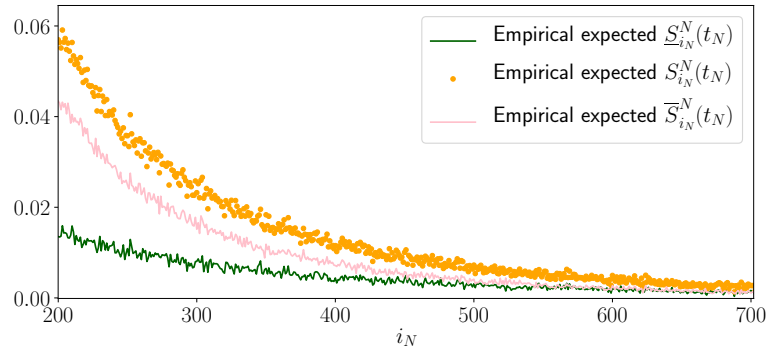


Figure 4.4 – Empirical expectation of  $\underline{S}_{i_N}^N(t_N)$ ,  $S_{i_N}^N(t_N)$  and  $\bar{S}_{i_N}^N(t_N)$  for large values of  $i_N$ , i.e  $i_N \in [200, 700]$ . The orange bullets, the pink and green lines correspond to the empirical expectation over 50 000 realizations, respectively, of  $S_{i_N}^N(t_N)$ ,  $\bar{S}_{i_N}^N(t_N)$ , and  $\underline{S}_{i_N}^N(t_N)$ .  $N = 500$

**Ideas of proofs.** As explained previously, the quantities were computed by dividing into two contributions:  $\bar{S}_i^N(t)$ , the number of neutral mutations that appeared in resistant cells, and  $\underline{S}_i^N(t)$ , the ones that appeared in sensitive cells. Moreover, each of both contributions were considered as the sum of

- (1) the part provided by the progenies (of the initial sensitive cells) that include exactly one *ancestral resistant* cell,
- (2) and the part provided by the progenies that carry two or more *ancestral resistant* cells.

With this in mind, we studied first the law of the emergence time of an *ancestral resistant* cell conditioned on belonging to a progeny that includes exactly one *ancestral resistant* cell. Precisely, such time was proved to be an exponential time whose parameter converges to  $\lambda_0$  when  $N \rightarrow \infty$ . This allowed us to compute the first parts (1) described above.

Secondly, we computed the expected number of multiple ancestral resistant cells in a progeny, which helped us to conclude that the second parts (2) described above are negligible when  $N \rightarrow \infty$  with respect to the first ones (1) in every situation of our interest.

## 4.3 PERSPECTIVES

### 4.3.1 Convergence in law of the SFS

The results presented above have the disadvantage of being based on the expectation of the Site Frequency Spectrum. However, in biological experiments, we rarely have access to information on the expectation that will require a large number of replicates. Instead, the experiments correspond to a single realization of the process. It would therefore be interesting to obtain more precise information. For example, in a recent preprint [53], the authors give almost sure convergences of the number of mutations shared by a small number of cells, within the framework of an exponentially growing population starting with one cell. Similarly, with Céline Bonnet, we are currently working on extending our results to obtain stronger convergences of the number of mutations shared by a large number of cells in the context of rescue dynamics. A significant difference in behavior is already expected for mutation rates of order  $1/N^\alpha$  with  $\alpha < 1$  and those of order  $1/N$ . In the first case, we will observe averaging due to the large number of mutations, which will bring us back to the work presented previously, whereas the second case is much more challenging, since the number of mutations will be finite a.s.

Other extensions are also under consideration in order to take into account more realistic models, by studying more complex population dynamics with interactions between cells for example, or by considering that more than one mutation is required before the acquisition of resistance (see the preprint [91] for example).

### 4.3.2 Genetic composition of an exponentially growing population

As part of his thesis, Vianney Brouard focused on precisely describing the genetic composition of an exponentially growing population subject to rare mutations. Precisely, the mutation probabilities are modeled as negative powers of a parameter  $n$  whereas the typical population sizes of interest are positive powers of  $n$ . Such models or similar ones are commonly used to model the expansion of tumors and the appearance of resistance [37, 25]. Vianney Brouard was able to describe the first-order asymptotics of the size of each subpopulation [16] at any time, in the case of neutral or deleterious mutations.

In an ongoing project with Vianney Brouard, we are extending these results for selective mutations. This is of particular interest in this context, as one can imagine that resistance or more aggressive cells can appear spontaneously. However, it remains challenging as the mathematical techniques and tools needed are different from the previous works of this subject.

In the longer term, we could also study this model under a changing environment, modeling for example the administration of various treatments during certain periods. This is particularly appropriate in the context of oncology, where current research is focusing on the effects of alternating treatments [62].

# BIBLIOGRAPHY

- [1] Elsa Abs, H el ene Leman, and R egis Ferri ere. A multi-scale eco-evolutionary model of cooperation reveals how microbial adaptation influences soil decomposition. *Communications biology*, 3(1):520, 2020. 17, 19, 21
- [2] Valeriy I Afanasyev, Christian B oinghoff, G otz Kersting, and Vladimir A Vatutin. Limit theorems for weakly subcritical branching processes in random environment. *Journal of Theoretical Probability*, 25(3):703–732, 2012. 53
- [3] Steven D Allison. Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. *Ecology Letters*, 8(6):626–635, 2005. 18
- [4] K.B. Athreya and P.E. Ney. *Branching processes*. Springer-Verlag Berlin, Mineola, NY, 1972. Reprint of the 1972 original [Springer, New York; MR0373040]. 37, 44
- [5] Ricardo BR Azevedo and Peter Olofsson. A branching process model of evolutionary rescue. *Mathematical Biosciences*, 341:108708, 2021. 64
- [6] Mamadou Ba and Etienne Pardoux. Branching processes with interaction and a generalized ray–knight theorem. *Annales de l’IHP Probabilit es et statistiques*, 51(4):1290–1313, 2015. 53
- [7] Vincent Bansaye, Maria-Emilia Caballero, and Sylvie M el eard. Scaling limits of population and evolution processes in random environment. *Electron. J. Probab*, 24(19):1–38, 2019. 53
- [8] Vincent Bansaye and Bertrand Cloez. From the distributions of times of interactions to preys and predators dynamical systems. *Journal of Mathematical Biology*, 87(1):1–41, 2023. 19
- [9] Vincent Bansaye, Juan Carlos Pardo, and Charline Smadi. Extinction rate of continuous state branching processes in critical L evy environments. *ESAIM Probab. Stat.*, 25:346–375, 2021. MR4291369. 53
- [10] Vincent Bansaye and Viet Chi Tran. Branching Feller diffusion for cell division with parasite infection. *ALEA : Latin American Journal of Probability and Mathematical Statistics*, 8:95–127, 2011. 68



- [11] J. Bascompte, P. Jordano, C. Melian, and J.M. Olesen. The nested assembly of plant-animal mutualistic networks. Proceedings of the National Academy of Sciences, 100(16):9383–9387, 2003. 24
- [12] Sylvain Billiard, H el ene Leman, Thomas Rey, and Viet Chi Tran. Continuous limits of large plant-pollinator random networks and some applications. arXiv preprint arXiv:2201.05219, 2022. 17, 22, 24
- [13] Andrew R Blaustein. Kin recognition mechanisms: phenotypic matching or recognition alleles? The American Naturalist, 121(5):749–754, 1983. 39
- [14] C eline Bonnet and H el ene Leman. Site frequency spectrum of a rescued population under rare resistant mutations. arXiv preprint arXiv:2303.04069, 2023. 64, 66
- [15] J W Boughman. Divergent sexual selection enhances reproductive isolation in sticklebacks. Nature, 411(6840):944–948, 2001. 38
- [16] Vianney Brouard. Genetic composition of supercritical branching populations under power law mutation rates. arXiv preprint arXiv:2309.12055, 2023. 70
- [17] Reinhard B urger and Kristan A Schneider. Intraspecific competitive divergence and convergence under assortative mating. The American Naturalist, 167(2):190–205, 2006. 40
- [18] F. C ezilly and D. Allain e. Thomas F., Lef evre T., Raymond M., Biologie  evolutive, chapter La s election sexuelle., pages 387–422. De Boeck, 2010. 29
- [19] N Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. Stochastic Processes and their Applications, 116(8):1127–1160, 2006. 15, 16, 35, 36, 44
- [20] Nicolas Champagnat and Beno t Henry. A probabilistic approach to dirac concentration in nonlocal models of adaptation with several resources. The Annals of Applied Probability, 29(4):2175–2216, 2019. 17
- [21] Nicolas Champagnat and Amaury Lambert. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. The Annals of Applied Probability, 17(1):102–155, 2007. 15, 19, 28
- [22] Nicolas Champagnat, Amaury Lambert, and Mathieu Richard. Birth and death processes with neutral mutations. International Journal of Stochastic Analysis, 2012, 2012. 66
- [23] Nicolas Champagnat and Sylvie M el eard. Invasion and adaptive evolution for individual-based spatially structured populations. Journal of Mathematical Biology, 55(2):147–188, 2007. 13, 14
- [24] Nicolas Champagnat and Sylvie M el eard. Polymorphic evolution sequence and evolutionary branching. Probability Theory and Related Fields, 151:45–94, 2011. 16, 35

- [25] David Cheek and Tibor Antal. Mutation frequencies in a birth–death branching process. The Annals of Applied Probability, 28(6):3922–3947, 2018. 70
- [26] C. Chicone. Ordinary Differential Equations with Applications. Number 34 in Texts in Applied Mathematics. Springer-Verlag New York, 2006. 31
- [27] Camille Coron, Manon Costa, Fabien Laroche, H el ene Leman, and Charline Smadi. Emergence of homogamy in a two-loci stochastic population model. ALEA, 18:469–508, 2021. 32, 35
- [28] Camille Coron, Manon Costa, H el ene Leman, Violaine Llaurens, and Charline Smadi. Origin and persistence of polymorphism in loci targeted by disassortative preference: a general model. Journal of Mathematical Biology, 86(1):4, 2023. 47
- [29] Camille Coron, Manon Costa, H el ene Leman, and Charline Smadi. A stochastic model for speciation by mating preferences. Journal of mathematical biology, 76:1421–1463, 2018. 30, 38
- [30] M Costa, C Hauzy, N Loeuille, and S M el eard. Stochastic eco-evolutionary model of a prey-predator community. Journal of mathematical biology, 72(3):573–622, 2015. 15
- [31] Alina Crudu, Arnaud Debussche, Aur elie Muller, Ovidiu Radulescu, et al. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. The Annals of Applied Probability, 22(5):1822–1859, 2012. 21
- [32] Charles Darwin. On the origin of species: facsimile of the first edition. LONDON: JOHN MURRAY, ALBEMARLE STREET., 1859. 13, 29
- [33] Khanh N Dinh, Roman Jaksik, Marek Kimmel, Amaury Lambert, and Simon Tavar e. Statistical inference for the evolutionary history of cancer genomes. Statistical Science, 35(1), 2020. 65
- [34] M. Doebeli and U. Dieckmann. Speciation along environmental gradients. Nature, 421:259–263, 2003. 16
- [35] Xan Duhalde, Cl ement Foucart, and Chunhua Ma. On the hitting times of continuous-state branching processes with immigration. Stochastic Process. Appl., 124(12):4182–4201, 2014. MR3264444. 58, 61
- [36] R. Durrett and S. Levin. The importance of being discrete (and spatial). Theoretical population biology, 46(3):363–394, 1994. 13
- [37] Richard Durrett and Richard Durrett. Branching process models of cancer. Springer, 2015. 70
- [38] Richard Durrett and Stephen Moseley. Evolution of resistance and progression to disease during clonal expansion of cancer. Theoretical population biology, 77(1):42–48, 2010. 64

- [39] S. Ethier and T. Kurtz. Markov processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence. 31
- [40] Steven N. Evans, Alexandru Hening, and Sebastian J. Schreiber. Protected polymorphisms and evolutionary stability of patch-selection strategies in stochastic environments. J. Math. Biol., 71(2):325–359, 2015. MR3367678. 60, 62
- [41] C. Fontaine, P.R. Guimaraes Jr., S. Kéfi, and E. Thébault. The ecological and evolutionary implications of merging different types of networks. Ecology Letters, 14(11):1170–1181, 2011. 22
- [42] Jasmine Foo and Kevin Leder. Dynamics of cancer recurrence. The Annals of Applied Probability, 23(4):1437 – 1468, 2013. 64
- [43] Clément Foucart and Matija Vidmar. Continuous-state branching processes with collisions: first passage times and duality. Stochastic Processes and their Applications, 167:104230, 2024. 62
- [44] N Fournier and S Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. The Annals of Applied Probability, 14(4):1880–1919, 2004. 23, 29, 31
- [45] Zongfei Fu and Zenghu Li. Stochastic equations of non-negative processes with jumps. Stochastic Processes and their Applications, 120(3):306–330, 2010. 54
- [46] S Gavrillets. Fitness Landscapes and the Origin of Species. Princeton University Press, 2004. 40
- [47] S Gavrillets. Models of speciation: Where are we now? Journal of heredity, 105(S1):743–755, 2014. 13, 29, 38
- [48] S Gavrillets and C R B Boake. On the evolution of premating isolation after a founder event. The American Naturalist, 152(5):706–716, 1998. 40
- [49] Hans-Otto Georgii and Ellen Baake. Supercritical multitype branching processes: the ancestral types of typical individuals. Advances in Applied Probability, 35(4):1090–1110, 2003. 37
- [50] Hans-Rolf Gregorius. Characterization and Analysis of Mating Systems. Citeseer, 1989. 38
- [51] Ashleigh S Griffin, Stuart A West, and Angus Buckling. Cooperation and competition in pathogenic bacteria. Nature, 430(7003):1024–1027, 2004. 18
- [52] Einar Bjarki Gunnarsson, Kevin Leder, and Jasmine Foo. Exact site frequency spectra of neutrally evolving tumors: A transition between power laws reveals a signature of cell viability. Theoretical Population Biology, 142:67–90, 2021. 66

- [53] Einar Bjarki Gunnarsson, Kevin Leder, and Xuanming Zhang. Limit theorems for the site frequency spectrum of neutral mutations in an exponentially growing population. arXiv preprint arXiv:2307.03346, 2023. 70
- [54] Thomas F Hansen. Stabilizing selection and the comparative analysis of adaptation. Evolution, 51(5):1341–1351, 1997. 28
- [55] Daniel L Hartl, Andrew G Clark, and Andrew G Clark. Principles of population genetics, volume 116. Sinauer associates Sunderland, 1997. 43
- [56] Alan Hastings. Disturbance, coexistence, history, and competition for space. Theoretical population biology, 18(3):363–373, 1980. 20
- [57] Hui He, Zenghu Li, and Wei Xu. Continuous-state branching processes in lévy random environments. Journal of Theoretical Probability, 31:1952–1974, 2018. 56
- [58] Alexandru Hening and Dang H Nguyen. The competitive exclusion principle in stochastic environments. Journal of mathematical biology, 80:1323–1351, 2020. 62
- [59] J.N. Holland and J.L. Bronstein. Mutualism. In Sven Erik Jørgensen and Brian D. Fath, editors, Encyclopedia of Ecology, pages 2485–2491. Academic Press, Oxford, 2008. 25, 27
- [60] J.N. Holland and D.L DeAngelis. A consumer?resource approach to the density-dependent population dynamics of mutualism. Ecology, 91:1286–1295, 2010. 22
- [61] Yoh Iwasa, Martin A Nowak, and Franziska Michor. Evolution of resistance during clonal expansion. Genetics, 172(4):2557–2566, 2006. 64
- [62] Haojie Jin, Liqin Wang, and René Bernards. Rational combinations of targeted cancer therapies: background, advances and challenges. Nature Reviews Drug Discovery, 22(3):213–234, 2023. 70
- [63] Miloslav Jiřina. Stochastic branching processes with continuous state space. Czechoslovak Mathematical Journal, 8(2):292–313, 1958. 53
- [64] A G Jones and N L Ratterman. Mate choice and sexual selection: What have we learned since darwin? PNAS, 106(1):10001–10008, 2009. 39
- [65] Jure Jugovic, Mitja Crne, and Martina Luznik. Movement, demography and behaviour of a highly mobile species: A case study of the black-veined white, *aporia crataegi* (lepidoptera: Pieridae). European Journal of Entomology, 114:113, 2017. 39
- [66] Olav Kallenberg and Olav Kallenberg. Foundations of modern probability, volume 2. Springer, 1997. 55
- [67] Samuel Karlin and James McGregor. Polymorphisms for genetic and ecological systems with weak coupling. Theoretical population biology, 3(2):210–238, 1972. 40
- [68] David G Kendall. Birth-and-death processes, and the theory of carcinogenesis. Biometrika, 47(1/2):13–21, 1960. 64

- [69] Motoo Kimura and Tomoko Ohta. CHAPTER EIGHT. Breeding Structure of Populations, pages 117–140. Princeton University Press, 2020. 49
- [70] M Kirkpatrick. Sexual selection and the evolution of female choice. Evolution, 41:1–12, 1982. 40
- [71] E. Knop, L. Zoller, R. Ryser, C. Gerpe, M. Hörler, and C. Fontaine. Artificial light at night as a new threat to pollination. Nature, 548(7666):206–209, 2017. 22
- [72] Natalia Komarova. Stochastic modeling of drug resistance in cancer. Journal of theoretical biology, 239(3):351–366, 2006. 64
- [73] Andreas E Kyprianou. Introductory lectures on fluctuations of Lévy processes with applications. Springer Science & Business Media, 2006. 54
- [74] Amaury Lambert. The branching process with logistic growth. Ann. Appl. Probab., 15(1A):1506–1535, 2005. 53, 57, 60, 61
- [75] John Lamperti. The limit of a sequence of branching processes. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 7(4):271–288, 1967. 53
- [76] R Lande. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences, 78(6):3721–3725, 1981. 38, 40
- [77] R Lande and M Kirkpatrick. Ecological speciation by sexual selection. Journal of Theoretical Biology, 133(1):85–98, 1988. 38
- [78] Russell Lande. Natural selection and random genetic drift in phenotypic evolution. Evolution, pages 314–334, 1976. 28
- [79] J. LaSalle. Some extensions of liapunov’s second method. Circuit Theory, IRE Transactions on, 7(4):520–527, 1960. 44
- [80] Vi Le. Processus de branchement avec interaction. PhD thesis, Aix-Marseille, 2014. 56
- [81] Mickael Le Gac and Michael Doebeli. Environmental viscosity does not affect the evolution of cooperation during experimental evolution of colicigenic bacteria. Evolution, 64(2):522–533, 2010. 18
- [82] H el ene Leman. Convergence of an infinite dimensional stochastic process to a spatially structured trait substitution sequence. Stochastics and Partial Differential Equations: Analysis and Computations, 4:791–826, 2016. 13, 15, 16
- [83] H el ene Leman. A stochastic model for reproductive isolation under asymmetrical mating preferences. Bulletin of mathematical biology, 80(9):2502–2525, 2018. 30, 38
- [84] H el ene Leman, Sylvie M el eard, and Sepideh Mirrahimi. Influence of a spatial structure on the long time behavior of a competitive lotka-volterra type system. Discrete and Continuous Dynamical Systems-B, 20(2):469–493, 2014. 13, 15, 16
- [85] H el ene Leman and Juan Carlos Pardo. Extinction and coming down from infinity of continuous-state branching processes with competition in a l evy environment. Journal of Applied Probability, 58(1):128–139, 2021. 55

- [86] H el ene Leman and Juan Carlos Pardo. Extinction time of logistic branching processes in a brownian environment. ALEA: Latin American Journal of Probability and Mathematical Statistics, 18:1859–1890, 2021. 55, 59, 62
- [87] Brian A Lerch, Akshata Rudrapatna, Nasser Rabi, Jonas Wickman, Thomas Koffel, and Christopher A Klausmeier. Connecting local and regional scales with stochastic metacommunity models: Competition, ecological drift, and dispersal. Ecological Monographs, 93(4):e1591, 2023. 20
- [88] J. Jelle Lever, Egbert H. van Nes, Marten Scheffer, and Jordi Bascompte. The sudden collapse of pollinator communities. Ecology letters, 17(3):350–359, 2014. 24, 25
- [89] S. Levin. Dispersion and population interactions. American Naturalist, pages 207–228, 1974. 13
- [90] RC Lewontin, LR Ginzburg, and SD Tuljapurkar. Heterosis as an explanation for large amounts of genic polymorphism. Genetics, 88(1):149–169, 1978. 49
- [91] Aaron Li, Danika Kibby, and Jasmine Foo. A comparison of mutation and amplification-driven resistance and their impacts on tumor recurrence. arXiv preprint arXiv:2305.19423, 2023. 70
- [92] L. Lov asz. Large Networks and Graph Limits. Number 60 in American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 2012. 25
- [93] Rugang Ma. Lamperti transformation for continuous-state branching processes with competition and applications. Statist. Probab. Lett., 107:11–17, 2015. MR3412750. 57
- [94] T. Mendelson and K. Shaw. Sexual behaviour: rapid speciation in an arthropod. Nature, 433(7024):375–376, 2005. 29
- [95] R M Merrill, R W R Wallbank, V Bull, P C A Salazar, J Mallet, M Stevens, and C D Jiggins. Disruptive ecological selection on a mating cue. Proceedings of the Royal Society of London B: Biological Sciences, 279(1749):4907–4913, 2012. 38
- [96] J. Metz, S. Geritz, G. Mesz ena, F. Jacobs, and JS. Van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. Stochastic and spatial structures of dynamical systems, 45:183–231, 1996. 15
- [97] P O’Donald. Assortive mating in a population in which two alleles are segregating. Heredity, 15:389–396, 1960. 40
- [98] Edouard Ollier, Pauline Mazzocco, Damien Ricard, Gantian Kaloshi, Ahmed Idbaih, Agusti Alentorn, Dimitri Psimaras, J er ome Honnorat, Jean-Yves Delattre, Emmanuel Grenier, et al. Analysis of temozolomide resistance in low-grade gliomas using a mechanistic mathematical model. Fundamental & clinical pharmacology, 31(3):347–358, 2017. 64
- [99] H Allen Orr and Robert L Unckless. The population genetics of evolutionary rescue. PLoS genetics, 10(8):e1004551, 2014. 64

- [100] S. Palau and J. C. Pardo. Continuous state branching processes in random environment: the Brownian case. *Stochastic Process. Appl.*, 127(3):957–994, 2017. MR3605717. 60
- [101] Sandra Palau and Juan Carlos Pardo. Branching processes in a lévy random environment. *Acta Applicandae Mathematicae*, 153:55–79, 2018. 53, 54
- [102] Tami M Panhuis, Roger Butlin, Marlene Zuk, and Tom Tregenza. Sexual selection and speciation. *Trends in Ecology & Evolution*, 16(7):364–371, 2001. 38
- [103] Étienne Pardoux. *Probabilistic Models of Population Evolution: Scaling Limits, Genealogies and Interactions*, volume 1. Springer, 2016. 53
- [104] R J H Payne and D C Krakauer. Sexual selection, space, and speciation. *Evolution*, 51(1):1–9, 1997. 39
- [105] Paul B Rainey and Katrina Rainey. Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, 425(6953):72–74, 2003. 18
- [106] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999. MR1725357. 58
- [107] Ken-iti Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. MR1739520. 59
- [108] Joshua P Schimel and Michael N Weintraub. The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. *Soil Biology and Biochemistry*, 35(4):549–563, 2003. 18
- [109] O. Seehausen, Y. Terai, I. Magalhaes, K. Carleton, H. Mrosso, R. Miyagi, I. van der Sluijs, M. Schneider, M. Maan, H. Tachida, H. Imai, and N. Okada. Speciation through sensory drive in cichlid fish. *Nature*, 455(7213):620–626, 2008. 29
- [110] M R Servedio. Limits to the evolution of assortative mating by female choice under restricted gene flow. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1703):179–187, 2010. 38, 40
- [111] M R Servedio and R Bürger. The counterintuitive role of sexual selection in species maintenance and speciation. *Proceedings of the National Academy of Sciences*, 111(22):8113–8118, 2014. 38
- [112] M R Servedio and R Bürger. The effects of sexual selection on trait divergence in a peripheral population with gene flow. *Evolution*, 69(10):2648–2661, 2015. 38
- [113] K. Shaw and Y. Parsons. Divergence of mate recognition behavior and its consequences for genetic architectures of speciation. *The American Naturalist*, 159(S3):S61–S75, 2002. 29

- [114] Charline Smadi, Hélène Leman, and Violaine Llaurens. Looking for the right mate in diploid species: How does genetic dominance affect the spatial differentiation of a sexual trait? Journal of Theoretical Biology, 447:154–170, 2018. 38
- [115] John Maynard Smith and John Haigh. The hitch-hiking effect of a favourable gene. Genetics Research, 23(1):23–35, 1974. 66
- [116] Walter L Smith and William E Wilkinson. On branching processes in random environments. The Annals of Mathematical Statistics, pages 814–827, 1969. 53
- [117] Thorsten Stefan, Louise Matthews, Joaquin M Prada, Colette Mair, Richard Reeve, and Michael J Stear. Divergent allele advantage provides a quantitative model for maintaining alleles with a wide range of intrinsic merits. Genetics, 212(2):553–564, 2019. 50
- [118] E. Thébault and C. Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science, 329(5993):853–856, 2010. 24
- [119] Hwai-Ray Tung and Rick Durrett. Signatures of neutral evolution in exponentially growing tumors: A theoretical perspective. PLOS Computational Biology, 17(2):e1008701, 2021. 65
- [120] J Albert C Uy, Gail L Patricelli, and Gerald Borgia. Complex mate searching in the satin bowerbird *ptilonorhynchus violaceus*. The American Naturalist, 158(5):530–542, 2001. 39
- [121] G. Van Doorn, A. Noest, and P. Hogeweg. Sympatric speciation and extinction driven by environment dependent sexual selection. Proceedings of the Royal Society of London B: Biological Sciences, 265(1408):1915–1919, 1998. 38
- [122] Le Vi and Etienne Pardoux. Height and the total mass of the forest of genealogical trees of a large population with general competition. ESAIM: Probability and Statistics, 19:172–193, 2015. 56



## **Stochastic models of evolving populations: diversity, spatial structures, cooperative behaviors and mating preferences**

Ce manuscrit présente un résumé de mes travaux de recherche qui se situent à l'interface entre les mathématiques et la biologie. Plus précisément, la majeure partie de mes travaux repose sur des modèles probabilistes qui décrivent, à un niveau individuel, la dynamique de populations soumises à différentes forces évolutives. Il est alors possible d'étudier différentes renormalisations de ces processus - mimant des hypothèses biologiques de grandes populations, de temps long, de mutations rares,... - afin d'obtenir des convergences vers des limites macroscopiques, qu'elles soient déterministes ou stochastiques. Ces limites sont également analysées pour obtenir des résultats quantifiés sur les comportements des populations modélisées.

En utilisant à la fois des outils stochastiques et déterministes, mon objectif est ainsi de comprendre les forces évolutives et les mécanismes naturels contribuant à l'émergence et au maintien de la diversité au sein des populations. Le premier chapitre se concentre sur l'impact des structures spatiales et des comportements coopératifs. Le deuxième chapitre explore le rôle de la reproduction sexuée, notamment autour de la question des liens entre préférences sexuelles et spéciation d'espèces. Le troisième chapitre présente des résultats sur les temps d'extinction d'un modèle macroscopique stochastique de dynamique de populations soumises à un environnement externe. Enfin, le quatrième chapitre se penche sur l'étude de populations de cellules tumorales soumises à des mutations neutres fréquentes, en particulier, en présentant des caractérisations des distributions de ces mutations pour des populations soumises à un traitement.