Stochastic individual-based models of adaptive dynamics and applications to cancer immunotherapy

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Martina Vera Baar

aus

Ibbenbüren

Bonn, Oktober 2016

Abstract

In this thesis stochastic individual-based models describing Darwinian evolution of asexual, competitive populations are studied. A specialization of these models is developed to describe tumor development under immunotherapy and an arising extended model is analyzed mathematically. In the first part (Chapter II) we consider a population with a large but non-constant population size characterized by a natural birth rate, a logistic death rate modeling competition, and a probability of mutation at each birth event. In this individual-based model the population state at a fixed time is given as a measure on the space of phenotypes and the evolution of the population is described by a continuous time, measure-valued Markov process. We investigate the long-term behavior of the system in the limits of large population size ($K \to \infty$), rare mutations ($u \to 0$), and small mutational effects ($\sigma \to 0$), proving convergence to the canonical equation of adaptive dynamics. This limit equation is an ODE that describes the evolution in time of the phenotypic value in a population consisting essentially of one single phenotype. The main difficulty is that we take the three limits simultaneously, i.e. $u = u_K$ and $\sigma = \sigma_K$, tend to zero with K, subject to conditions that ensure that the time scale of birth and death events remains separated from that of successful mutational events. This slows down the dynamics of the microscopic system and leads to serious technical difficulties that require the use of completely different methods than in comparable works where the limits are taken separately. More precisely, the time until a mutant phenotype fixates is diverging (in K) and thus, we cannot use the law of large numbers to approximate the stochastic system. In the second part (Chapter III) we propose an extension of the individual-based model, which broadens the range of biological applications. The primary motivation was to model cancer immunotherapy in order to simulate and describe qualitative the experiments reported in Landsberg et al. [92], where tumors resist immunotherapy through inflammation-induced reversible dedifferentiation. The main expansions are that we have three different actors in this context (T-cells, cytokines, and cancer cells), that we distinguish cancer cells by phenotype and genotype, that we include environment-dependent phenotypic plasticity, and that we take into account the therapy effects. With this new setup we are able to model various phenomena arising in immunotherapy. We argue why stochastic models may help to understand the resistance of tumors to therapeutic approaches and may have non-trivial consequences on tumor treatment protocols. Furthermore, we show that the interplay of genetic mutations and phenotypic switches on different time scales as well as the occurrence of metastability phenomena raise new mathematical challenges. The present thesis focuses more on these aspects. More precisely, we study the behavior of the individual-based model which includes phenotypic plasticity on a large (evolutionary) time scale and in the simultaneous limits of large populations ($K \to \infty$) and rare mutations ($u_K \to 0$), proving convergence to a Markov jump process, which is a generalization of the usual polymorphic evolution sequence. This can be seen as an extension of the results by Champagnat and Méléard (cf. [25, 30]).

Acknowledgement

For sure, this thesis would not have been possible without the help of many people. Therefore, I would like to use this acknowledgement to thank all of them.

First of all, I would like to express my sincere gratitude to my supervisor Prof. Dr. Anton Bovier for the encouragement to do my PhD and for the continuous support during my PhD study. Thank you very much for your open door conveying me that I am welcome to asks question, for your patience, guidance and support in presenting my research on conferences and in writing scientific publications, for sharing your enormous intuitions, your immense knowledge and your ideas among solving mathematical problems with me. Furthermore, I would like to say thank you for offering us "the best coffee in town".

My special thanks go to my co-author, Dr. Nicolas Champagnat, for helping me to handle all the necessary steps in the proof of the main result of this thesis. I am glad that I could profit from his expertise on my PhD topic especially during my stay in Nancy. I am very thankful that he answered all my emails in such detail and with such patience. I could not have imagined having a better co-author. Thank you for your enormous help, which ended in the successful publication of our work. Of course, I would like to express the warmest thanks also to my other co-authors, Prof. Dr. Thomas Tüting, Prof. Dr. Michael Hölzel, Meri Rogava, Dr. Loren Coquille and Dr. Hannah Mayer for many fruitful discussions about cancer immunotherapy and for providing experimental data. This work about cancer immunotherapy gave me a great opportunity to view beyond the horizon and get an expression of another pioneering field of research. Special thanks among this part of my thesis go to Hannah und Loren.

Besides my advisor and my co-authors, I would like to thank all members of the committee for being part of the jury, especially Prof. Dr. Anita Winter. Further, I thank Prof. Dr. Martin Rumpf for being my mentor.

I really enjoyed my PhD time in Bonn and it was great that I could be part of the probability theory group at the University of Bonn. In particular, I would like to thank Hannah for the personal as well as mathematical support, I received from her since my first semester. Thank you for almost ten years friendship. Moreover, I am very grateful to Anna for proofreading the introduction. Many thanks also go to the other PhD students, Adela, Anna, Carina, Daniel, Kaveh, Lisa, Patrick, Peter, Raphael and Rebecca, and Postdocs, Constanza, Evangelina, Loren, Nicola and Sebastian. My special thanks go of course to our secretary Mei-Ling. Finally, I would like to express my gratitude to Sebastian and my family for the love, the encouragement and support I constantly received from them.

Contents

Chapter I

Introduction

In this thesis we study stochastic-individual based models for Darwinian evolution of asexual reproducing, competitive populations. Furthermore, we specialize these models to be able to describe tumor development under treatment with immunotherapy and investigate the extended individual-based models which arise from this applications mathematically. Biological evolution, explaining the origin and the variation of species, is very complex and a result of various underlying processes such as reproduction, variation by mutation and recombination of genetic material, competition between individuals and species, and selection of the most adaptive traits. Cancer immunotherapy harnesses and enhances a patient's own immune system to treat cancer and is one of the most promising new cancer treatment approaches [34]. Also the mechanisms behind cancer immunotherapy are driven by various underlying processes, such as interaction between immune cells, cancer cells, and cytokines or the phenotypic and genotypic heterogeneity of cancer cells. Much of the mathematical work in evolution theory as well as in cancer immunology has taken place on a deterministic level, using dynamical systems and differential equations [7]. Our aim is to study stochastic models which describe the system on the level of individuals and use them to make predictions about the macroscopic long-term behavior of the system, which incorporate the random effects of the microscopic level. An evolutionary example for this is that a mutant, which appears in a large population and is fitter than the other individuals, can die out accidentally with a certain probability and therefore does not invade the population. With respect to cancer immunotherapy a further example is that there can be randomness concerning whether a therapy destroys all cancer cells or some remain, which might then lead to a relapse. Studying these models is challenging from the mathematical point of view and requires to establish new methods, which may also be useful to solve problems in other models as we will see e.g. in Chapter II.

The introduction is organized as follows. In the first section we give a short introduction and historical overview of the biological theory of evolution and describe some of the main mathematical formalizations and modeling approaches of evolution theory which were established in the last century: population genetics (including quantitative genetics), evolutionary game theory, and adaptive dynamics. We explain why stochastic individual-based models of adaptive dynamics are a convenient tool to study mathematically certain aspects of the biological theory of evolution and give some background information about these density-dependent Markov processes. Then, we give a rough overview of the different therapeutic approaches of cancer immunotherapy and argue why an expansion of the stochastic individual-based models of adaptive dynamics is a good choice for modeling cancer immunotherapy.

In the second section we begin with studying a simple individual-based model describing evolution of a population only as a result of interaction between individuals but ignoring

variation, i.e. describing development due to ecological effects only, not long-term evolution. We give three examples with concrete parameters and explain why the stochastic process can be approximated for large populations by a deterministic function. After this we define the stochastic individual-based model, which is the foundation of the models we study in this thesis and describes the Darwinian evolution of a population including the effects of interaction and variation. Variation is modeled by the possibility that a mutation can occur at each birth event with a certain probability. In this type of models the evolution of the population is described by a continuous time, measure-valued Markov process. Examples and simulations are provided at the end of the section.

In the third section we give an overview of the thesis and present the main results. Theorem II.4.1, a convergence result for the stochastic individual-based model, will appear soon in *Annals of Applied Probability* as a joint work with A. Bovier and N. Champagnat. In this publication, which is the content of Chapter II, we investigate the long-term evolution of the system in the limit of large population size combined with rare mutations and small mutational effects, proving the convergence to the canonical equation of adaptive dynamics – in one step. In Chapter III we propose a model for cancer immunotherapy and present an example which qualitatively models the experiment of Landsberg et al. [92], where tumors escape cancer immunotherapy by phenotypic plasticity in presence of certain cytokines. Furthermore, we study the influence of phenotypic plasticity on the long-term evolution of asexually reproducing populations in the limits of a large population size combined with rare mutations, proving the convergence to the extended version of the polymorphic evolution sequence. Parts of this chapter are already published in *Scientific Reports* as a joint work with L. Coquille, H. Mayer, M. Hölzel, M. Rogava, T. Tüting, and A. Bovier.

Note that this introductory chapter gives the historical context and a more extensive overview about different model approaches. The most relevant parts, directly related to the mathematical work presented in this thesis, are also given in the introductions of Chapters II and III in a condensed way. The two main chapters are related to each other but can be read independently.

I.1   Modeling Darwinian evolution and cancer immunotherapy

The modern theory of biological evolution has its source in Charles Darwin's book *On the Origin of Species* [36], published in 1859, where he outlined the famous basic principles of evolution which were later summarized by the phrase *survival of the fittest*. The scientific theory of evolution is based on the principle of natural selection, which was independently also conceived and described by Wallace [37]. It can be described by the following three basic mechanisms:

Heredity: individuals can reproduce and pass their traits from generation to generation

Variation: traits vary among individuals with respect to morphology, physiology, etc.

Natural Selection: different traits have different rates of survival and reproduction (fitness)

Individuals which are better adapted to the environment survive and reproduce more likely and thus transmit their traits to more descendants than less adapted individuals. This produces the process of natural selection and has the consequence that disadvantaged traits disappear over time. These central ideas of Darwin and Wallace have remained largely unchanged. What was missing in Darwin's theory of natural selection was a proper scientific theory of inheritance which explains the variation among individuals on which natural selection can act. To obtain that selection modifies populations gradually over a long period of

time, as suggested by Darwin, a continuous supply of variation is necessary. During the 19th century the idea of blending inheritance, which is based on the hypothesis that offsprings have characteristics that are intermediate between their parents, was quite common, but Darwin had reservations regarding this idea. Moreover, as was pointed out by Jenkin [82], variation decays rapidly over time under blending inheritance. A proper mechanism of inheritance, which is essentially still accepted today even if our modern understanding of heredity is much more complex, was provided in 1866 by Gregor Mendel [100], who studied the reproduction of peas. His predictions about how traits are inherited from one generation to the next led to the formalism of Mendel's law of inheritance. In other words, Mendel devised the mechanism of heredity for sexual reproduction that was missing in the theory by Darwin. The Mendelian inheritance is based on phenotypic traits, which are determined by genes. Each gene of a diploid organism consists of two alleles, one from each parent. Furthermore, it is based on three laws: *segregation*, alleles segregate from each other and each gamete (sexual reproductive cell, e.g. egg or sperm cell) carries only one allele for each gene; *independent assortment*, genes segregate independently during the formation of gametes; and *dominance*, one allele dominates the other in inheritance unless both are recessive. (Note that until the 1920's it was not clear how Darwin's and Mendel's theory could be combined, see Subsection I.1.1 for more details.)

Darwin's and Mendel's works are the foundation of the evolutionary theory we still use today. Each organism on earth is characterized by a genotype, which contains the full heredity information and is encoded in the DNA, and a phenotype, which describes an organism's actual properties, such as morphological or physiological properties. The distinction between genotype and phenotype is fundamental in the study of inheritance of traits and their evolution. In a population consisting of individuals with different phenotypes, the individuals interact with each other, they compete e.g. for resources (area, nutrients, water, food, etc.), or with other species (host, parasite, predator, prey, etc.). This has of course an effect on the reproduction and survival ability for each individual. In other words, the selection process acts on the phenotypes and is a consequence of the competition between the different actors. Thus, adaptation of phenotypes depends on the outer environment and on the composition of the rest of the population. The mechanism of heredity is given in two forms, either by asexual reproduction, i.e. an organism just copies its genome, which results in two genotypic identical organisms as long as there was no error in the process of replication, or by sexual reproduction, i.e. the genomes of two gametes recombine and form an organism which includes genetic material from both gametes. Asexual reproduction is the primary form of reproduction for single-celled organisms (e.g. bacteria), but also many plants and fungi reproduce asexually. More complex organisms usually reproduce sexually. Variation in sexually reproducing species is generated by the recombination of two different genotypes and by mutations in the genome of the offsprings. In species reproducing asexually, variation is generated by mutations only. A *mutation* is a permanent alteration of the genome of an organism and results for example from errors in the process of replication. Note that there is general agreement that mutations are the ultimate source of variation [40]. Without mutations the genotype frequencies in sexually reproducing population would remain constant after a relatively short time such that the population would not evolve anymore (cf. Hardy-Weinberg principle).

Until today there are still lots of open problems in the theory of evolution. Some examples are how the adaptation of an individual to the environment should be quantified, how the environment influences the phenotype giving rise to phenotypic plasticity, and how the complex map between genotype and phenotype, linking heredity and ecological influences, works. More general problems concern the mechanisms behind natural selection, which results in the

survival of the fittest on long time scales, and separation, which gives rise to new species. Hence, simplifications and approximations are necessary to understand the complex mechanism of biological evolution. In the following, we describe some of the main mathematical formalizations and modeling approaches of evolution theory which were established in the last century and have different simplifications in order to focus on different aspects of evolution.

I.1.1   Population genetics

The theory of *population genetics* focuses on the genetic differences within a population and studies changes in the allele frequencies, but usually ignores that individuals interact with each other or, more general, are influenced by their environment. This simplification makes it possible to model a realistic inheritance law in order to understand the complex patterns of genetic variation. Population genetics has its origin in the fundamental work of Fisher, Haldane, and Wright, developed already about one hundred years ago. However, the current research in this field is still based and strongly influenced by their work. This pioneering work also laid the foundation of the *modern evolutionary synthesis*, in which the disagreements between Darwin's and Mendel's theories were overcome.

Among the scientists at the end of the 19th century there was already a disagreement whether the process of evolutionarily changes is gradual, as Darwin argued, or occurs in jumps, as e.g. Huxley believed. However, with the rediscovery of Mendel's work in 1900 many scientists rather believed in a non-Darwinian evolution process through jumps and some even thought that Mendel's work refuted Darwin's idea of natural selection. This lead to a discord between Darwinism and Mendelism. The problem was that Darwin focused on the evolution of complex organisms, where the selection process acts on a large number of slight variants, and Mendel focused in his studies on the inheritance of discrete phenotypic traits determined by a single gene. In the early years of the twentieth century, eventually, Fisher, Haldane, and Wright solved this problem by establishing a theoretical framework which integrates the inheritance principles of Mendel in the Darwinian theory of the natural selection. More precisely, Fisher showed in [58] that the correlation between relatives, measured by biometric properties, can be explained by multiple Mendelian factors and random non-genetic influences and that thus Mendel's inheritance theory agrees with the theory of natural selection: The discontinuous jump character of Mendelian genetics disappears if traits depend on many genes, where each has only small contributions, and results in almost continuous variation and thus gradual evolution. Haldane developed a mathematical theory of natural and artificial selection in Mendelian populations, providing expressions for the evolutionary changes caused by slow and rapid selection, in which traits depend on a single or on several genes and generations do or do not overlap [70, 71]. In these models selection is acting on the differences in survival ability, reproduction or mortality due to Mendelian genes. Furthermore, the interaction of natural selection with mutation as well as with migration is mathematically analyzed and metastable phenomena caused by genes which are disadvantageous alone but advantageous together are mentioned. In [117] Wright described mathematically how the random process of reproduction changes the gene frequency in finite populations, analyzed the interplay between this random *genetic drift* and mutation, migration, and selection of various sorts, and synthesized these processes into a single formula for the stationary distribution (cf. [11]). (Note that this stationary distribution can be seen as an expansion of the *Hardy-Weinberg principle*, developed in 1908 by Hardy and Weinberg, independently from each other, and stating that the genetic variation within populations remains constant from one generation to the next in the absence of other disruption processes such as mutation.) Furthermore, in 1932 Wright introduced the concept of an evolutionary or *adaptive landscape*, where selection drives populations upwards

on this fitness landscape towards a local peak while the genetic drift can push the population away from such a local maximum and could potentially cause a peak shift [118].

The historical part about population genetics above is mainly taken from the introductory books [52] and [56]. Note that both books focus on the purely mathematical theory and less on population genetics itself. (See [47, 35] for alternative introductions.)

In general, population genetics incorporates experimental, observational and theoretical aspects and is largely quantitative. Unfortunately, the complexity of nature often has the consequence that the mathematical models of populations genetics, which are necessarily based on simplifications, are eventuality obsolete because of new findings from experiments or observations [56]. However, the purely mathematical theory of population genetics is a very large area of applied mathematical research, provides detailed models of the genome structure and the mechanisms of inheritance, and may help to make quantitative statements of new findings qualitative. Apart from studying the forward going evolutionary process, as the classical theory surrounding the Wright-Fisher model does, since the 1980's it is also common to focus on the retrospective analysis, i.e. to look backwards in time e.g. to the most recent common ancestor [52, 56]. In particular, the retrospective models surrounding the Kingman coalescent process have on the one hand rich mathematical structure and provide on the other hand the necessary tools for the interpretation of genetic data and thus became a significant part of the current research [52, 56]. The inheritance relationships between the individuals are typically represented as a genealogical, coalescent or gene tree in this retrospective theory [52]. However, also the classical prospective theory is still highly relevant and its tools can be applied in the coalescent theory [56]. Most of the mathematical models used in population genetics, including the Wright-Fisher and the Kingman coalescent mentioned above, are simplified by a constant or effectively infinite population size but include mechanisms that can be used to describe sexual reproduction [56]. Further common models in the field are the Cannings model, the Moran model, the Wright-Fisher diffusion, Kimura's stepping stone model, and the Fleming-Viot process (cf. [52, 56]). Besides short-term dynamics of the gene frequencies, also long-term evolution including mutation and selection can be studied by population genetic models [40]. The selective advantage of an organism is often an a priori given quantity in these models called fitness, which depends directly on the genome but ignores the outer environment and the composition of the rest of the population [56]. Thus, by natural selection the population tries to reach the maximum of a fixed adaptive landscape. One of the problems about ignoring interaction is that it is hard to model the phenomenon of a population splitting into two lines going their separate ways. Therefore, the origin of the species is a barely understood problem of population genetics [104]. Furthermore, since selection acts on phenotypes, a knowledge of the genotype-phenotype map is required to be able to study evolution. This map is affected i.a. by the dominance between alleles and epistatic genes (which act as inhibitors of other genes) and thus, in general, extremely complicated. As a consequence, a quantitatively description of long-term evolution on the DNA-level is impractical [40]. Besides this, population genetic models of evolution usually assume that phenotypic traits are controlled by genes at a single locus (position on a chromosome), but most traits which are important for the evolutionary process are determined by several or many genes at different loci [23].

Quantitative genetics is a branch of population genetics founded by Fisher [58], which deals with evolution of phenotypic traits that vary (almost) continuously and are measured on a metric scale. (See e.g. [57] or [22] for an introduction.) These traits are usually influenced by genes at many different loci in the genome, rather than just one or two. In other words, quantitative genetics studies the inheritance of quantitative rather than qualitative traits

[57]. Examples for quantitative traits are weight and hight or, more specific, wing span in birds and milk yield in cows. In general, many morphological, physiological, or economically important traits are quantitative [22, 23]. Thus, understanding the inheritance mechanisms of these quantitative differences is important for studying evolution by natural selection or breeding. Though the Mendelian laws cannot be applied directly to quantitative traits, a basic premise of quantitative genetics inheritance is that qualitative traits depend on genes which are subject to these laws. Hence, this theory can be seen as an expansion of Mendelian genetics. The main methodical differences are that quantitative genetics studies evolution on the level of populations, not of individuals as population genetics does, and that for this study the metric measurement not only the classification of the individuals is necessary, i.e. quantitative genetics had to develop concepts for genetic properties of populations and for inheritance of metric traits [57]. Therefore, it simplifies both the parents-offspring and genotype-phenotype relation, but makes it possible to study natural selection of phenotypes independently of genetic details [40]. An important feature of most metric traits is that, on an appropriate scale, their frequency distribution is close to a normal curve [57, 23]. This can be justified by considering that quantitative traits are usually controlled by a large number of loci, whose alleles have only small contributions, and the central limit theorem [23, 22]. Thus, properties of the normal distribution and statistical techniques can be used to study the evolution of quantitative traits. Furthermore, observed metric traits can be characterized in terms of mean, variance, and covariance [57]. Of course quantitative traits are affected by the environment. Quantitative genetics incorporates besides genetic also this environmental dependence and describes the change in the distributions of the quantitative traits over time (i.e. from generation to generation) [40].

In [90], Lande introduced a simple quantitative genetic discrete time model, where phenotypic traits depend on a genetic and an environmental component and which provides a recursive equation for the evolution of the mean phenotypic trait. Selection acts in this model on the phenotypic trait, favoring the fittest. Besides the case of constant phenotypic fitnesses, also the case of frequency-dependent selection, where a phenotype's fitness depends on the frequencies of the different phenotypes present in the population modeling interaction between the individuals, is considered in this paper [90]. Once the fitness function is determined, the evolution of the main phenotypic trait value can be described in terms of the frequency distribution and this fitness function. Moreover, if the variance of the trait distribution is small, the change in the main trait value can be approximated by a deterministic recurrence independent of the frequency distribution (cf. [40]). This was done for constant fitness by Lande [91] and further developed for non-constant fitness in [80, 1, 114]. Lande's theory had a large impact on evolutionary biology because it integrates methods of quantitative genetics into evolutionary genetics, it has a simple and intuitive character, where detailed genetic information is not required, and thus it has received heuristic and predictive importance in many applications. A further reason for the success of Lande's theory was the introduction of a powerful adaptive landscape concept for phenotypic traits, which is related to Wright's concept of an evolutionary landscape for genotypes mentioned above [23]. Note that there is a close relationship between the deterministic recurrences obtained out of Lande's theory and the canonical equation of adaptive dynamics, which is a central object of this thesis (see paragraph about adaptive dynamics).

I.1.2 Evolutionary game theory

An alternative framework which ignores genes and sexual reproduction and focuses on studying phenotypic evolution in some interacting environment is *evolutionary game theory*. This

concept was introduced around 1970 by Maynard Smith and Price [99]. The simplification of asexuality allows to concentrate on the selection process as a result of interaction. While population genetics models are a good tool for studying the genetic variability of a population, game theory models are convenient for studying phenotypic evolution in a more ecological realistic manner. In Maynard Smith's words, evolutionary game theory is a way of thinking about evolution at the phenotypic level when the fitness or evolutionary advantage of an individual organism depends on its phenotype and on the frequencies of the other phenotypes currently present in the population [98], i.e. the fitness is not a given (constant) quantity, but depends on the population the individual lives in. This framework considers the individuals of a population as playing games against each other and studies the resulting population dynamics and equilibria, which may be attained by the population [98]. (Actually, this way of considering evolution was already present in a paper published in 1930 by Fisher.) The population state changes according to the rules of a game here. In this game of life, the players are individuals of a population, the strategies are their heritable phenotypes, and the payoff is their fitness in this environment. In other words, evolutionarily game theory is an expansion of the classical game theory, which was established by von Neumann, Morgenstern, and Nash [115, 107, 106] to analyze economical and social behavior. In fact, the concept of game theory turned out to be even more suitable to describe biological behavior [98]. One reason for this is that the main assumptions of the classical theory are that players behave rational and according to self-interest but it is not really reasonable to believe that humans behave rationality. In the evolutionary context, the rational behavior is replaced by the dynamics and stability of the population and self-interest by Darwinian fitness [98]. It is more reasonable to expect a population to evolve to stable states, i.e. to assume evolutionary stability, than to believe in a rational human behavior [98]. Similar as in the classical game theory, the dynamics and equilibrium states of the system resulting from playing the game are in the focus of interest here. Especially from the evolutionary point of view, the study of successive invasion strategies is very common, i.e. analyzing whether an alternative mutant strategy or rather phenotype that is initially rare can invade the current population state. Of particular importance in this context is the so-called *evolutionary stable strategy* (ESS), an extension of the usual Nash equilibrium. Once such a strategy is adopted by all individuals in the population, no initially rare mutant strategy has a higher fitness (payoff) and can thus invade the population [98]. In other word natural selection alone is sufficient to prevent that mutant strategies successfully invade such a strategy, i.e. it is stable in the evolutionary sense. In the simple model introduced in [99], the assumptions that the population is infinite and reproduces asexually and that only pairwise symmetric contests take place lead e.g. to ESS [98]. However, evolution is a process of steady change, so one can criticize that the evolutionary game theoretical approach focuses on equilibrium states [98]. Nevertheless, the idea of ESS has caused huge success in the field and evolutionary game theory has been very helpful to explain many complex and challenging aspects of biology, e.g. altruistic behaviors in the context of Darwinian evolution. Furthermore, it gains increasing importance in other fields like economics, sociology, anthropology, and philosophy.

I.1.3   Adaptive dynamics

Since the 1990s a new branch of phenotypical evolution theory, known as *adaptive dynamics*, is developing, which has its origin in the works of Hofbauer and Sigmund [74], Metz et al. [103], and Marrow et al. [95]. Though the term adaptive dynamics sometimes refers to general long-term evolutionary dynamics of quantitative traits, driven by mutations and selection, we use it for the particular theoretical framework developed in the second half of the 1990's by

Metz et al. [104], Dieckmann and Law [42], and Geritz at al. [63, 62] which integrates and expands the methods of evolutionary game theory (cf. [40]). Note that the introduction to adaptive dynamics given in [26] served as basis for parts of this paragraph.

Adaptive dynamics is a theoretical approach for modeling phenotypical evolution in various complex ecological systems and provides the basis for the work presented in this thesis. Similar to evolutionary game theory, genetic details and sexual reproduction are usually ignored to simplify the study and the fitness of an individual depends on its own phenotype and on the environment, more precisely on the composition of the population it lives in and interacts with. One advantage of the theory of adaptive dynamics is that it integrates ecological dynamics in the evolutionary process, which play an important role in natural selection, and provides powerful tools, which can be applied to many different ecological situations, e.g. to describe competitive or cooperative interactions between different individuals or species, or to describe predator-prey, immune-pathogen, host-parasite or plant-insect relationships. Moreover, standard models of adaptive dynamics can be expanded e.g. in order to study the evolution of cancer under treatment (cf. Chapter III). The ideas and the concepts of adaptive dynamics have undergone many developments and extensions over the last decades. Besides the papers we mention below, there exist many more in the context of adaptive dynamics, which focus on different biological aspects, but all of them have in common that they analyze the ecological effect on evolution. On the webpage www.mv.helsinki.fi/home/kisdi/addyn.htm Kisdi provides a huge list of adaptive dynamics references ordered according to their aim. (Apart from the probabilistic approach we consider here, Diekmann et al. proposed in [44] a corresponding deterministic approach based on partial differential equations.)

A fundamental idea of adaptive dynamics is that the current population can be assumed to be close to an equilibrium, determined by the ecological system, when a new mutant appears. This allows to introduce the notion of *invasion fitness*, which measures the selective advantage of a mutant that occurs in this environment. In the individual-based model we study in Chapter II, the selective advantage of an individual can be measured in terms of the growth rate of this individual, which depends on its phenotype and the environment it lives in (cf. [59, 25, 30]). Of course this growth rate changes if the environment changes, however, if the population forms the environment and is close to an equilibrium generated by ecological dynamics, the initial growth rate of a mutant individual appearing in this population determines the possibility whether the mutant's phenotypic trait can stabilize in the population or not. In other words, it determines if a mutant's phenotype can invade the population and thus it is called invasion fitness. Moreover, using this notion an environment-depending invasion fitness landscape can be constructed, which allows to describe the successive mutant invasions and stationary population states determined by the underlying ecological dynamics. The main biological assumptions justifying this approach are that the size of the studied population is large and that mutations during the asexual reproduction process occur only rarely. Note that this invasion fitness landscapes, despite sounding similar to the traditional fitness landscapes introduced by Wright, is conceptually different [116].

Historically, successive invasions of ecological stable strategies were first studied in the context of game theory by Hofbauer and Sigmund in [74]. In this work a dynamics to model the effect of adaptation when selection is frequency dependent is proposed and related with the stability of equilibria. The proposed ordinary differential equation (ODE) describes the evolution of dominant strategies (phenotypes) in an essentially monomorphic population. (We say that a population is *monomorphic* if all individuals have the same phenotypic trait.) To justify this continuous change in dominant strategies, the additional assumption that occurring mutants have only slightly different strategies than their predecessors, i.e. the mutational

effects are small, is necessary. The idea to deduce the evolutionary behavior of a global ecological system from the possibilities of mutant invasions and to analyze this dynamics further as the phenotypic difference between mutant and predecessor tends to zero were also exploited in a context of predator-prey interaction by Marrow et al. [95]. Further, the concept of a fitness landscape coevolving with the population was already present in this article, which was later improved in [104] and [101]. In [103], Metz et al. deal with the question how fitness should be defined for general ecological scenarios. One of the first really fundamental papers of adaptive dynamics is the paper [104] by Metz et al., which combines the approaches of [74] and [103] and, together with the papers by Dieckmann and Law [42] and Geritz at al. [63, 62], forms the foundation of this framework. In this paper the authors use more complex models, add a dynamical aspect to the approach of game theory, and specify the necessary biological assumptions. Moreover, these four papers together introduce the basic elements, methods, and graphical tools used in adaptive dynamics.

As pointed out in [104], the main assumption, large population size and rare mutations, implies that the ecological and evolutionary time scale are separated in the sense that whether a mutant's trait is selected is entirely settled before a new mutation occurs. There is enough time such that the mutant's trait either vanishes or fixates in the population before a new mutation occurs. Another common assumption in the context of studying a competitive population is called *invasion implies fixation principle* and means that if a mutant's trait fixates in the population, it replaces the previous (resident) type completely, i.e. long-term coexistence is excluded. This allows to define an evolutionary time scale where the population is monomorphic at any time t. Hence, the evolution proceeds can be modeled as a continuous time Markov process which jumps from one phenotypic trait to another fitter one. This stochastic process is usually called *trait substitution sequence* (TSS) and has been introduced by Metz et al. [103] (see also [104] and [42]) and mathematically studied in [27, 25, 30]. The TSS model is a fundamental element of adaptive dynamics and the approach leading to this model as well as this model itself provide powerful tools for understanding various evolutionary phenomena, such as polymorphism or evolutionary branching, and are the foundation for other biological concepts as the canonical equation of adaptive dynamics [25]. By polymorphism we mean the stable coexistence of different phenotypes, which arises if the invasion implies fixation principle is not assumed (cf. [104] and [30]). The phenomenon of *evolutionary branching*, meaning that a population initially concentrated around a single dominant phenotype (evolving over time) splits into two sub-populations of different dominant types going their separate ways, was vaguely already mentioned in [74] and in detail first studied by Metz et al. in [104] (see also [63, 62]). Moreover, Metz et al. identify the points in the phenotype trait space, where such a phenomenon is likely to happen, the so-called *evolutionary singular strategies*, and give a criterion for evolutionary branching depending on the derivatives of the fitness function at these points. Note that the concept of evolutionary singular strategies can be seen as a generalization of the evolutionary stable strategy concept introduced in the context of evolutionary game theory [62].

Another important concept in the theory of adaptive dynamics is the *canonical equation of adaptive dynamics* (CEAD), which was introduced by Dieckmann and Law in [42] and describes the evolution in time of the expected phenotypic trait value in a monomorphic, competitive population before an evolutionary branching. In the adaptive dynamics approach the evolutionary process of various ecological systems proceeds as a sequence of mutant invasions, where only mutants with positive invasion fitness can invade. Thus, under the additional assumption that the difference between mutant and resident trait is very small, Darwinian evolution of quantitative traits in several coevolutionary scenarios can be modeled as a grad-

ual process given by the solution of a differential equation, which has in the one dimensional case the form

$$\frac{d\,x_t}{dt} = k(x_t)\partial_1 f(x_t, x_t), \tag{I.1.1}$$

where x_t denotes the trait value of the population, $k(x_t)$ is a non-negative coefficient and $f(y, x_t)$ is the invasion fitness, i.e. the selective advantage of individuals with trait value y occurring in an environment determined in terms of the resident trait value x_t [42]. Moreover, $\partial_1 f$ denotes the partial derivative with respect to the first variable, i.e. $\partial_1 f(x_t, x_t) \equiv \frac{\partial}{\partial y} f(y, x_t)|_{y=x_t}$, and is usually called *selection gradient* in the literature. In this deterministic model, selection pushes the population to increase its fitness locally. Dynamics of this kind have been proposed and studied by many authors e.g. as a hill-climbing process on an adaptive landscape (cf. [42] and references therein). As mentioned in the paragraph about quantitative genetics, there also exists a structurally similar equation for sexually reproducing populations, which was introduced by Lande in [91] and later extended to frequency-dependent selection [80, 114, 1]. This has motivated Dieckmann and Law to consider the equation (I.1.1) as a sort of canonical equation of evolutionary models [42, 40]. Moreover, in [42] they proposed an ordinary differential equation of this kind describing long-term phenotypic evolution of quantitative traits in an asexual reproducing, competitive population, where the coefficient $k(x)$ and the invasion fitness $f(y, x)$ are given as explicit expressions taking into account the ecological processes at the level of the individuals. Dieckmann and Law called their ODE canonical equation of adaptive dynamics and showed that the coefficient $k(x)$ equals the product of the population size, mutation rate, mutation variance, and a factor $1/2$. (The distribution of the mutant trait value is assumed to be symmetric, i.e. on average half of the occurring mutants have a negative invasion fitness. See Equation (II.3.9) for the explicit equation.) Dieckmann and Law's derivation has the conceptual background of modeling the dynamics of a population as a Markov process which incorporates reproduction, mutation, and selection [27]. Furthermore, the heuristics leading to the CEAD are based on the following biological assumptions: the population size is large, mutations are rare, and mutations have a small (phenotypic) effect. In addition, an invasion implies fixation principle is assumed. One possibility to recover the CEAD, which Dieckmann and Law used in [42], is to look at small mutational effects in the TSS model. (Recall that the TSS model describes the evolution of the phenotypic trait in a monomorphic population and is already a macroscopic approximation for large populations with rare mutations.) A mathematically rigorous proof of this derivation was given later in [27] (see also [24, 30]). One problem about this approach is that it gives no clue about how the biological parameters, population size, probability of mutations, and size of mutational effects should be compared to ensure that the CEAD approximation of the microscopic model is correct. In Chapter II we show that it is also possible to apply the three limits directly to the microscopic (individual-based) model by taking them simultaneously with an explicit relation between the parameters and to recover the CEAD (see below).

The mathematical background of adaptive dynamics

While the biological theory of adaptive dynamics is based on partly heuristic derivations, various aspects of the theory have been derived rigorously over the last years in the context of *stochastic individual-based models.* Most of the models used in probability theory to describe biological evolution can be traced back either to the Galton-Watson branching process or to the Wright-Fisher model. The classical Galton-Watson process can be used e.g. to model the total population size, when individual evolve independently, and can be extended to branching random walks or branching Brownian motion if the individuals move during their lifetime

according to random walks or Brownian motion [54]. In this finite-dimensional branching models the population either goes extinct or grows without any bound. No non-trivial equilibrium can be predicted, which makes it hard to model long-term evolution [54]. On the other hand, most of the models surrounding the Wright-Fisher model, frequently used in population genetics, are simplified by a fixed constant population size (e.g. the Moran model or the Fleming-Viot process). However, in a biological reasonable model a population should be able to regulate the population size itself adapted to the environment the individuals live in. The individual-based model we study in this thesis attains this feature.

More precisely, we study a system of interacting particles modeling Darwinian evolution of an asexual population (with a large but non-constant populations size) at the individual level. Each individual in the population is characterized by its phenotype and has a natural birth rate, a density-dependent logistic death rate modeling competition, and a probability of mutation at each birth event. Thus, the three basic mechanisms of evolution, heredity, mutation, and selection, are included. This model has the conceptual background of a continuous time, branching random walk, in which the death rate of an individual with trait x depends in addition on the population density and is defined as a weighted sum of the entire population, with weights depending on the trait the individuals carry [54]. This density-dependent component can on the one hand prevent that the population grows without bound and thus give rise to non-trivial stable population size, but destroys on the other hand the convenient branching property since individuals do not evolve independent from each other anymore. This makes the study of the model more difficult because most of the beautiful mathematics in the fields of branching processes rely on the branching property [54]. The model has originally been proposed to understand stochastically driven spatial pattern formations in ecological systems by Bolker and Pacala [15], in which the dispersion of the population (plants) is described by a measure-valued Markov process. This locally regulated model has been studied in parallel by Dieckmann and Law [43] and is sometimes called BPDL model in the literature. Since the 1970s, studying measure-valued stochastic processes has become very popular in the mathematical community [38, 53, 94]. For modeling Darwinian evolution of quantitative traits, these processes have the advantage that the real metric traits, not only classifications, can be studied, i.e. the space of possible phenotypic traits in the population does not have to be finite. This allows to define mutational events where each new mutant has a new randomly chosen phenotypic trait and guarantees a steady supply of variation. In the last decades, locally regulated models have been extensively studied in many works by various authors, e.g. [54, 59, 89, 25, 13, 78, 30, 39, 67], either as a model of Darwinian evolution or as a model of dispersal in a spatially structured population. In [59], Fournier and Méléard formulate a pathwise construction of the locally regulated process in terms of a Poisson point process. The model we define in Subsection I.2.2 is based on the formalization of this work.

As mentioned before, there are mathematically rigorous papers which show that this model converges in the simultaneous limits of large population and rare mutations to the trait substitution sequence [25, 30]. Furthermore, this jump process converges, in the limit of small mutation steps, to the canonical equation of adaptive dynamics [27, 30]. In Chapter II we analyze the situation when the limits of large population size, rare mutations, and small mutation steps, are taken *simultaneously* and prove that this process, which models evolution at the individual level, convergences to the CEAD – in one step. The fact that the mutants only have an infinitesimal small evolutionary advantage in this approach slows down the dynamics of the microscopic system and leads to serious technical difficulties. However, the simultaneous limit has the advantage (amongst other things) that population size, probability of mutations, and size of mutational effects can be compared on the individual level and thus

our results can be better applied to concrete biological examples to predict their long-term behavior. (See Sections II.1 and II.3 for further details about the relation to the other works and the technical difficulties.)

I.1.4   Cancer immunotherapy

In Chapter III we propose an extension of the above mentioned individual-based process, which broadens the range of biological applications. The primary motivation is modeling *cancer immunotherapy*, i.e. we want to study the evolution of a cancer population under special treatment. During the last decades, treatment of various cancers with immunotherapies received a lot of attention in the medical as well as the mathematical modeling communities [108, 88, 50, 72, 65, 76]. The editors of *Science* even chose cancer immunotherapy as the breakthrough of the year 2013. Immunotherapy does not attack the cancer cells, as chemo- and radiotherapy do, but targets the immune system [34]. Abnormal cells, which may lead to cancer, can usually be detected and destroyed by the immune system. However, some of these cells have the ability to avoid this, e.g. by reducing the expression of tumor antigens on their surface such that they can not be recognized by the immune cells, or by suppressing the immune system's activity in their microenvironment. As a result, these cells can proliferate and generate a cancerous tumor. There are various cancer immunotherapies which help to circumvent these cancer cell's escape mechanisms such that the immune system can detect and destroy the tumor again. Current therapies can be divided into three major classes: non-specific therapies, monoclonal antibodies, and vaccines [51]. Non-specific therapies use *cytokines* or other chemicals, e.g. IL-2 and IFN-α, to stimulate the general immune response. The strength and duration of an immune response is usually limited by special checkpoint proteins to prevent overreaction and damage of normal cells. Some monoclonal antibodies, like anti-CTLA-4 and anti-PD-1, can be used as *immune checkpoint inhibitors* and thus increase the immune systems ability to destroy cancer cells. Another approach is the so-called chimeric antigen receptor therapy, an *adoptive cell transfer (ACT)*, where T-cells are taken from a patient, genetically modified to present cancer specific antigen receptors, and then infused back into the patient to target these cancer cells [111]. Also vaccines, made from patient's own tumor cells, are used to strengthen the immune response to specific cancer.

Similar as for chemo- and radiotherapy, resistance is an important issue for cancer immunotherapy. Recently, several theoretical concepts have been proposed to explain why a cancerous tumor develops resistance during an initially successful therapy, leading to a relapse. A widely accepted idea to explain relapses is that pre-existing mutants (tumor cell variants with genetic aberrations) which exhibit therapy resistance are selected in a Darwinian evolutionary process (cf. [76] and references therein). Moreover, genotypic and phenotypic heterogeneity is a general feature of advanced tumors, which is considered to be the main driving force for resistance and may be enhanced during therapy [76, 96, 65]. In contrast to genotypic heterogeneity, phenotypic plasticity is a source of tumor heterogeneity caused by, in principle reversible, phenotypic switches, i.e. the phenotypic trait of a cancer cell can change over time. Furthermore, the phenotype can depend on the microenvironment, which changes during therapy. Landsberg et al. report in [92] their experimental finding that reversible phenotypic switches due to side effects of the immunotherapy cause resistance of the tumor. Chapter III is motivated by these experiments, where melanoma (skin cancer) under ACT therapy are investigated. In [76], Hölzel et al. emphasize the importance of developing a theoretical framework that incorporates the different aspects of therapy resistance, more precisely that integrates phenotypic plasticity, clonal selection, and reciprocal interactions between tumor cells and the microenvironment. Phenotypic plasticity can be described for example by the

systems biology concept of the epigenetic landscape and gene regulatory networks, see [76] and references therein. Simple Markov models, which describe the tumor on the cell level, have been used to study the dynamics of phenotypic proportions in human breast cancer cell lines [69]. Also models based on (multi-type) branching processes have been used to describe cancer on the cellular level, e.g. to study the evolutionary dynamics of cancer in response to targeted combination therapy or the accumulation of driver and passenger mutations during tumor progression [41, 21, 3, 20, 83]. A semi-deterministic model has been used to study the influence of driver mutations on the spatial evolutionary dynamics of solid tumors, where the spatial growth of cancer clones is deterministic, while mutants arise stochastically [4]. Simple deterministic models have been successfully used to address phenotypic plasticity of the cancer stem cell and its therapeutic implications [93]. However, non of these models incorporate all the effects of cancer immunotherapy described above.

There are serval reasons why an expansion of the stochastic individual-based model of adaptive dynamics is a reasonable choice for modeling cancer immunotherapy. For example, cancer cells reproduce asexually, tumors grow from a microscopic to a macroscopic level, tumor development is microenvironment depending, tumor cells compete for resources, and they interact with immune cells during treatment. The model expansions we use in Chapter III take into account that there are different types of actors in this context (T-cells, cytokines, and cancer cells) and that each cancer cell is characterized by its genotype and its associated phenotypes. Furthermore, we include microenvironment-dependent phenotypic plasticity and the therapy effects. Apart from studying the effects on cancer immunotherapy, this model allows to study the interplay of genetic mutations and phenotypic switches on different time scales and thus can describe phenotypic *and* genotypic evolution of a population.

I.2 Stochastic individual-based models

In this thesis we study the evolution of an asexual reproducing population that is composed of a finite number of individuals, each of them characterized by a one-dimensional (phenotypic) trait, taking into account the interaction between the individuals. As mentioned above, this can be done with stochastic individual-based models, which we introduce in this section. These models build the foundation for the work presented in Chapters II and III and are based on the pathwise construction of the locally regulated process by Fournier and Méléard [59].

In the first subsection we give a simple model only describing the evolution as a result of interaction between individuals but ignoring variation. Thus, this is a model describing ecology only. In the second subsection we define the actual model describing Darwinian evolution, including the effects of interaction and variation. Variation is modeled by the possibility that at each birth event a mutation may occur with a certain probability. If this probability equals zero, one is again back in the case of the simple model.

I.2.1 Stochastic multi-type models for describing ecological dynamics

We start with a simple model which focuses on the ecological dynamics of the population (cf. [10]). To this aim let us study the behavior of a population with two different phenotypic traits in some common environment. In this example we call the phenotypic traits 1 and 2, i.e. all individuals carry either the trait 1 or 2. Thus, a population consists of two subpopulations. Let $K \in \mathbb{N}$ describe the capacity of the environment. This can be interpreted as the size of the area or the amount of available resources. We are interested in the development of the

densities of the two subpopulations, i.e. the number of individuals with trait 1 respectively trait 2 per capacity of the environment K. The dynamics of the whole population is a result of the dynamics of each individual in this population. In this simple individual-based model, all individuals with trait $i \in \{1, 2\}$ reproduce with rate $b_i \in (0, \infty)$ and die due to age with rate $d_i \in [0, \infty)$, where $b_i - d_i > 0$. Furthermore, they compete for limited area or resources: For $i, j \in \{1, 2\}$, the competitive pressure an individual with trait i feels from an individual with trait j is given by c_{ij}/K, where $c_{ij} > 0$ and $c_{11}c_{22} \geq c_{12}c_{21}$. So, if we assume that the population at time t consists of $N_1(t)$ individuals with trait 1 and $N_2(t)$ individuals with trait 2, then the current total death rate of an individual with trait $i \in \{1, 2\}$ equals

$$d_i + \frac{c_{i1}}{K}N_1(t) + \frac{c_{i2}}{K}N_2(t). \tag{I.2.1}$$

Note that the capacity of the environment K gives the magnitude of the population size, but depending on the phenotypes more or less individuals can survive in the same environment (i.e. the population size does not stay constant over time but can also not increase to any size).

In [55] (Chapter 11) Ethier and Kurtz showed that for each $K \in \mathbb{N}$ the evolution of a population with the dynamics described above can be modeled by an $(\mathbb{N}_0/K)^2$-valued Markov pure jump process x^K with the following infinitesimal generator: For all functions $f : (\mathbb{N}_0/K)^2 \to \mathbb{R}$ with compact support and $x^K \in (\mathbb{N}_0/K)^2$,

$$\mathscr{L}^K f(x^K) = \sum_{i \in \{1,2\}} \left(f\left(x^K + \frac{e_i}{K}\right) - f\left(x^K\right) \right) b_i x_i^K K \tag{I.2.2}$$

$$+ \left(f\left(x^K - \frac{e_i}{K}\right) - f\left(x^K\right) \right) (d_i + c_{i1}x_1^K + c_{i2}x_2^K) x_i^K K, \tag{I.2.3}$$

where $e_1 = (1, 0)$ and $e_2 = (0, 1)$. The first term describes birth and the second death. Note that the second term, which models the competition in the population, is non-linear in x^K and that x^K describes the development of the population densities, i.e. at time t there are $x_1^K(t)K \equiv N_1(t)$ individuals of type 1 and $x_2^K(t)K \equiv N_2(t)$ individuals of type 2 present in the population. Figure I.1 provides simulations of the Markov process x^K with three different sets of parameters.
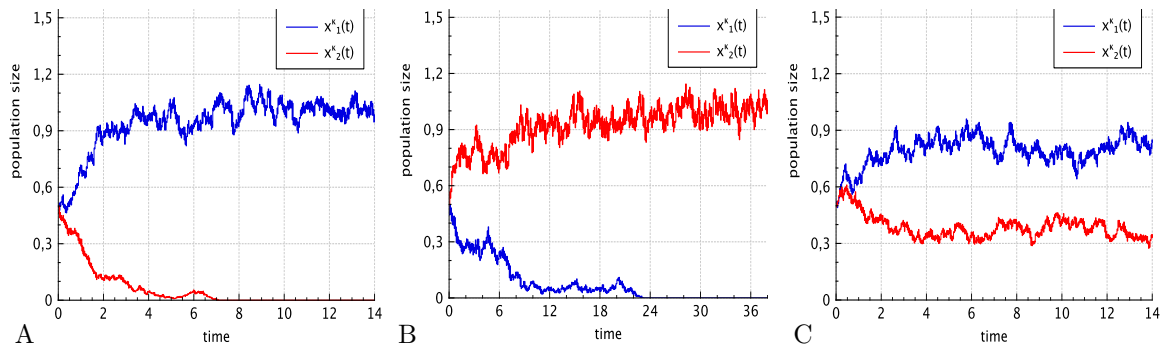
Figure I.1: Simulation of the stochastic process x^K with $K = 500$ and three different sets of parameters (cf. Table I.1). In A trait 2 (red curve) goes extinct, in B trait 1 (blue curve) goes extinct and in C both traits coexist.

The law of large numbers (cf. [55], Thm. 11.2.1) justifies that we can approximate the Markov process x^K for large K and on a finite time interval by the solution of the two-

dimensional competitive Lotka-Volterra system

$$\dot{n}_1(t) = n_1(t)(b_1 - d_1 - c_{11}n_1(t) - c_{12}n_2(t)),$$ (I.2.4)
$$\dot{n}_2(t) = n_2(t)(b_2 - d_2 - c_{21}n_1(t) - c_{22}n_2(t)).$$

In other words, the behavior of large populations can be approximated by the behavior of the deterministic system I.2.4. Figure I.2 shows the unique solutions of the deterministic systems approximating the stochastic processes simulated in Figure I.1.
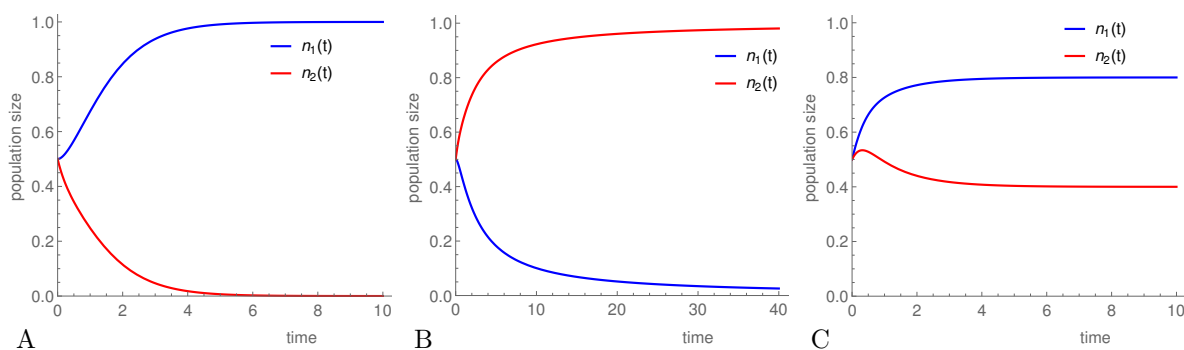
Figure I.2: Solution of the system I.2.4 with three different sets of parameters (cf. Table I.1).

Fortunately, the behavior of the two-dimensional competitive Lotka-Volterra system is well know. The fixed points of the system are: $(0,0)$, $((b_1 - d_1)/c_{11}, 0)$, $(0, (b_2 - d_2)/c_{22})$ and

$$\left(\frac{(b_1 - d_1)c_{22} - (b_2 - d_2)c_{12}}{c_{11}c_{22} - c_{12}c_{21}}, \frac{(b_2 - d_2)c_{11} - (b_1 - d_1)c_{21}}{c_{11}c_{22} - c_{12}c_{21}} \right),$$ (I.2.5)

where the latter is possible if both coordinates are positive. (Observe that if $b_1 = b_2$, $d_1 = d_2$, and $c_{11} = c_{12} = c_{12} = c_{22}$, then each point on the line, which connects $((b_1 - d_1)/c_{11}, 0)$ and $(0, (b_2 - d_2)/c_{22})$, is a fixed point. We use this fact in Chapter II.) Whether a fixed point is stable, depends on the parameters of the system. Figure I.3 shows the vector fields of the deterministic systems corresponding to Figure I.2.
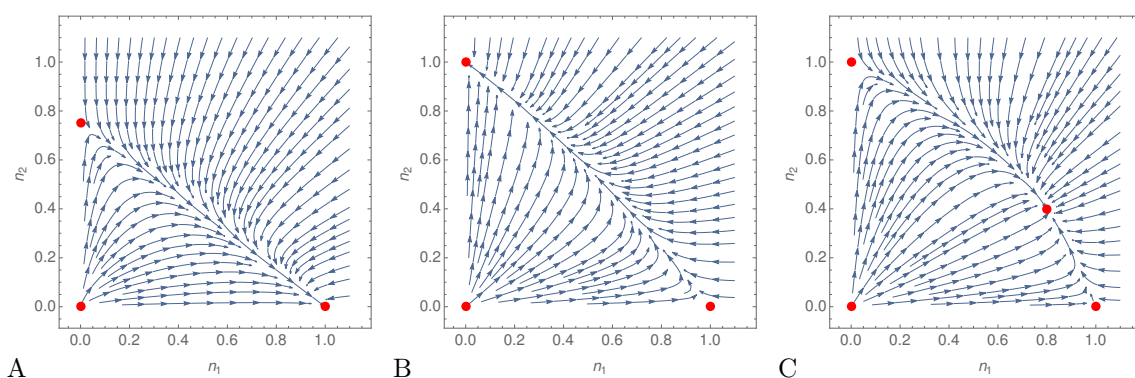
Figure I.3: Vector fields of the system I.2.4 with three different sets of parameters (cf. Table I.1). The fixed points are shown as red dots. In A, the system is attracted to the strictly stable fixed point $(1, 0)$. In B, the system is attracted to the strictly stable fixed point $(0, 1)$. In C, the system is attracted to the strictly stable fixed point $(0.8, 0.4)$. Thus, only in C both traits coexist.

Given any positive initial condition, the solution of (I.2.4) converges to a unique fixed point, describing either the fixation of a single trait or the coexistent of both. (Note that

by assuming $c_{11}c_{22} \geq c_{12}c_{21}$ we have excluded that the fixed point (I.2.5) is a saddle point.) This means that a sufficiently large population evolves until it reaches its unique ecological equilibrium, which corresponds to the strictly stable fixed point of the deterministic system. After that time it only fluctuates around this value and no real evolution takes place (cf. Figure I.1). Note that for finite K the population will go extinct if one waits long enough.

A	$b_1 = 5$	$d_1 = 1$	$b_2 = 4$	$d_2 = 1$	$c_{11} = 4$	$c_{12} = 4$	$c_{21} = 4$	$c_{22} = 4$
B	$b_1 = 5$	$d_1 = 1$	$b_2 = 5$	$d_2 = 1$	$c_{11} = 4$	$c_{12} = 4$	$c_{21} = 3$	$c_{22} = 4$
C	$b_1 = 5$	$d_1 = 1$	$b_2 = 5$	$d_2 = 1$	$c_{11} = 4$	$c_{12} = 2$	$c_{21} = 3$	$c_{22} = 4$

Table I.1: Parameters of the Figures I.1, I.2, and I.3

It is, of course, possible to study the corresponding multi-dimensional model. In this case, the large population approximation leads to the multi-dimensional competitive Lotka-Volterra system. In general, the long-term behavior of this system is very complex for higher dimensions. (Zeeman proved in [120] that three-dimensional competitive Lotka-Volterra systems admit Hopf bifurcations, which give rise to isolated periodic orbits. In [75] and [119], Hofbauer and So, and Xiao and Li give examples for systems with isolated periodic orbits. Smale showed that if $n \geq 5$, the n dimensional system can exhibit any asymptotic behavior [113]. In [73], Hirsch studied general n-dimensional competitive Lotka-Volterra systems and proved that there exists an $n-1$ dimensional manifold which attracts all non-convergent persistent trajectories, i.e. the system is essentially $n-1$ dimensional.)

However, if we assume that the competition matrix is symmetric and positive definit, the multi-dimensional competitive Lotka-Volterra system has a unique, global, asymptotically stable fixed point, i.e. if all components of $n(0)$ are positive, $n(t)$ converges as $t \to \infty$ to this fixed point (cf. e.g. Proposition 1.4 of [26] and references therein). Thus, the stochastic process is attracted to this fixed point if the population size is large.

Since we want to study the Darwinian evolution of a single species and not only the ecological interaction between different phenotypic traits, we are more interested in the case where the traits of the individuals belong to a continuum. This allows to add mutational events where each mutant has a new randomly chosen phenotypic trait, i.e. creates a new type (cf. [10]).

I.2.2   Measure-valued models for describing Darwinian evolution

In this subsection, we define the stochastic individual-based model which is the foundation of the models we study in this thesis (cf. [59, 25, 30, 10]). Note that although we study in Chapters II and III the case of one-dimensional phenotypic traits only, we give the definition of the more general model here.

The evolutionary process changes populations on a macroscopic level, but the basic mechanisms of evolution, heredity, mutation, and selection, act on the microscopic level of the individuals. Hence, we describe the evolving population by a stochastic system of interacting individuals, where each individual is characterized by a vector of phenotypic trait values.

Let $l \geq 1$ and $\mathcal{X}$ a closed subset of $\mathbb{R}^l$. Then, we call $\mathcal{X}$ the *trait space* of the population. Furthermore, let $\mathcal{M}(\mathcal{X})$ be the set of finite non-negative measures on $\mathcal{X}$, equipped with the topology of weak convergence, and $\mathcal{M}_P(\mathcal{X}) \subset \mathcal{M}(\mathcal{X})$ be the set of finite point measures on $\mathcal{X}$, i.e.

$$\mathcal{M}_P(\mathcal{X}) \equiv \left\{ \sum_{i=1}^{n} \delta_{\{x_i\}} \, : \, n \geq 0, \, x_1, \dots x_n \in \mathcal{X} \right\}, \tag{I.2.6}$$

where $\delta_{\{x\}}$ denotes the Dirac mass at $x \in \mathcal{X}$. Our aim is to study the evolution of a population as an $\mathcal{M}_P(\mathcal{X})$-valued, stochastic process $(\nu_t)_{t \geq 0}$ providing for each time t the population size as well as the phenotypic trait distribution in the population. This process should take into account the three basic mechanisms of evolution. Therefore, we introduce the following biological parameters. For any $x, y \in \mathcal{X}$,

$b(x) \in \mathbb{R}_+$ is the *rate of birth* of an individual with trait $x \in \mathcal{X}$.

$d(x) \in \mathbb{R}_+$ is the *rate of natural death* of an individual with trait $x \in \mathcal{X}$.

$c(x, y) \in \mathbb{R}_+$ is the *competition kernel* which models the competitive pressure an individual with trait $x \in \mathcal{X}$ feels from an individual with trait $y \in \mathcal{X}$.

$m(x)$ is the *probability that a mutation occurs at birth* from an individual with trait $x \in \mathcal{X}$.

$M(x, dh)$ is the *mutation law* of the mutational jump h. If a mutant is born from an individual with trait x, then the mutant trait is given by $x + h$, where h is a random variable with law $M(x, dh)$. The support of this mutation law is a subset of $\mathcal{X} - x \equiv \{h \in \mathbb{R}^l : x + h \in \mathcal{X}\}$.

At any time $t \geq 0$, we consider a finite population which consist of N_t individuals and each individual is characterized by an element of $\mathcal{X}$, its trait. Let us denote these traits by $x_1(t), ..., x_{N_t}(t)$. (Note that we allow that different individuals carry the same trait, i.e. we allow that $x_i(t) = x_j(t)$ for some $i \neq j$.) Then, we define the population state at time t as the measure

$$\nu_t = \sum_{i=1}^{N_t} \delta_{x_i(t)}. \tag{I.2.7}$$

In the following, we roughly summarize the population dynamics (cf. [59]). At time $t = 0$, we have a (possibly random) initial point measure $\nu_0 \in \mathcal{M}_P(\mathcal{X})$. Each individual with trait $x \in \mathcal{X}$, alive at time t, has three independent exponentially distributed "clocks":

(i) a *birth without mutation* clock with parameter $b(x)(1 - m(x))$,

(ii) a *birth with mutation* clock with parameter $b(x)m(x)$ and

(iii) a *death* clock due to natural death or competition with parameter $d(x) + \sum_{i=1}^{N_t} c(x, x_i(t))$.

If the *birth without mutation* clock of an individual rings, then the individual produces a new individual and this individual carries the same trait as it parent. If the *birth with mutation* clock of an individual (with trait x) rings, then the individual produces a new individual and the trait of this individual is given by $y = x + h$, where h is randomly chosen according to the mutation law $M(x, dh)$. Finally, if the *death* clock of an individual rings, then this individual disappears. Note that the parameter of this clock depends on the current population state. Whenever one of these events occurs, all clocks are reset.

In other words, we are looking for a homogenous measure-valued Markov process $(\nu_t)_{t \geq 0}$ with infinitesimal generator, $\mathscr{L}$, defined for any bounded measurable function f from $\mathcal{M}_P(\mathcal{X})$ to $\mathbb{R}$ and for all $\mu \in \mathcal{M}_P(\mathcal{X})$ by

$$\begin{aligned} \mathscr{L}f(\mu) \;=\; & \int_{\mathcal{X}} \big(f(\mu + \delta_x) - f(\mu)\big)(1 - m(x))\, b(x)\mu(dx) \\ & + \int_{\mathcal{X}} \int_{\mathbb{R}^l} (f(\mu + \delta_{x+h}) - f(\mu))\, m(x)b(x)\, M(x, dh)\, \mu(dx) \\ & + \int_{\mathcal{X}} (f(\mu - \delta_x) - f(\mu)) \Big(d(x) + \int_{\mathcal{X}} c(x, y)\mu(dy)\Big) \mu(dx). \end{aligned} \tag{I.2.8}$$

The first and the second (linear) terms describe birth with and without mutations, whereas the third (non-linear) term describes death due to age or competition. The selection process is driven by the density-dependent non-linearity of the third term modeling the competition in the population.

The following assumption allows to deduce the existence and uniqueness in law of a process on $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{X}))$ with infinitesimal generator $\mathscr{L}$, see [59], where $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{X}))$ denotes the space of càdlàg functions from $\mathbb{R}_+$ to $\mathcal{M}_P(\mathcal{X})$, endowed with the Skorokhod topology.

Assumption 1.   (i) b, d, and c are measurable functions, and there exist $\bar{b}$, $\bar{d}$, $\bar{c} < \infty$ such that $b(.) \leq \bar{b}$, $d(.) \leq \bar{d}$,   and $c(.\,,.) \leq \bar{c}$.

  (ii) Either: There exists a constant $C > 0$ and a probability density $\bar{M}$ on $\mathbb{R}^l$ such that for all $x \in \mathcal{X}$, $M(x, dh) = M(x, h)dh$ with $M(x, h) \leq C\bar{M}(h)$.

   Or: The support of the probability measure $M(x, dh)$ is finite for all $x \in \mathcal{X}$.

In fact, in [59] Méléard and Fournier give an explicit pathwise description of the population process, $(\nu_t)_{t \geq 0}$, in terms of Poisson measures. We recall this pathwise description in Sections II.5 and III.4.1 of this thesis adapted to the models we are studying there. Though we introduced this process to be able to study the case where the traits of the individuals belong to a continuum, this definition is also valid if $\mathcal{X}$ is a discrete set.

I.2.3   Examples and Simulations

We start with a simple example, where $\mathcal{X}$ is finite. So, let $\mathcal{X} = \{x_1, x_2, x_3\}$. Then, we consider an example with the following parameters

$$b(x_1) = 8, \quad b(x_2) = 9, \quad b(x_3) = 10, \tag{I.2.9}$$

$$d(x_i) = 1, \quad c(x_i, x_j) = 1, \quad \text{and} \quad m(x_i) = 0.15 \qquad \text{for all } 1 \leq i, j \leq 3. \tag{I.2.10}$$

Furthermore, we have a next neighbor mutation law i.e.

$$M(x_1, x_2 - x_1) = 1, \ M(x_2, x_1 - x_2) = 0.5, \ M(x_2, x_3 - x_2) = 0.5, \ \text{and} \ M(x_3, x_2 - x_3) = 1. \tag{I.2.11}$$

Figure I.4 shows three realizations of the process. We use in the following the shorthand notation $\nu(x)$ instead of $\nu(\{x\})$, for $x \in \mathcal{X}$.

All simulations shown in this thesis are performed with a Gillespie algorithm implemented by Boris Prochnau. Let us first explain in words how this particular example can be simulated. Afterwards we give a more general pseudocode of the algorithm. It is mainly based on elementary properties of Poisson point processes.

At time $t = 0$, we have a given initial population: $\nu_0 = 6\delta_{x_1} + \delta_{x_2} \in \mathcal{M}_P(\mathcal{X})$. Let τ_1 be the first jump time of the process. Since the minimum of finitely many exponential random variables is exponential distributed with the sum of the single parameters, τ_1 is exponential distributed with parameter

$$\text{TotalRate} \equiv \sum_{x \in Supp(\nu_0)} \nu_0(x) \left(b(x) + d(x) + \sum_{y \in Supp(\nu_0)} \nu_0(y)c(x, y) \right) \tag{I.2.12}$$

Thus, τ_1 is sampled according to $\mathcal{E}xp(138)$ in Figure I.4. Let $x^* \in Supp(\nu_0)$ be the trait of the individual which causes the jump, then we can sample x^* according to the following probability measure

$$\mathbb{P}[x^* = x] = \frac{\nu_0(x)[b(x) + d(x) + \sum_{y \in Supp(\nu_0)} c(x, y)\nu_0(y)]}{\text{TotalRate}}, \quad \text{for } x \in Supp(\nu_0) \tag{I.2.13}$$
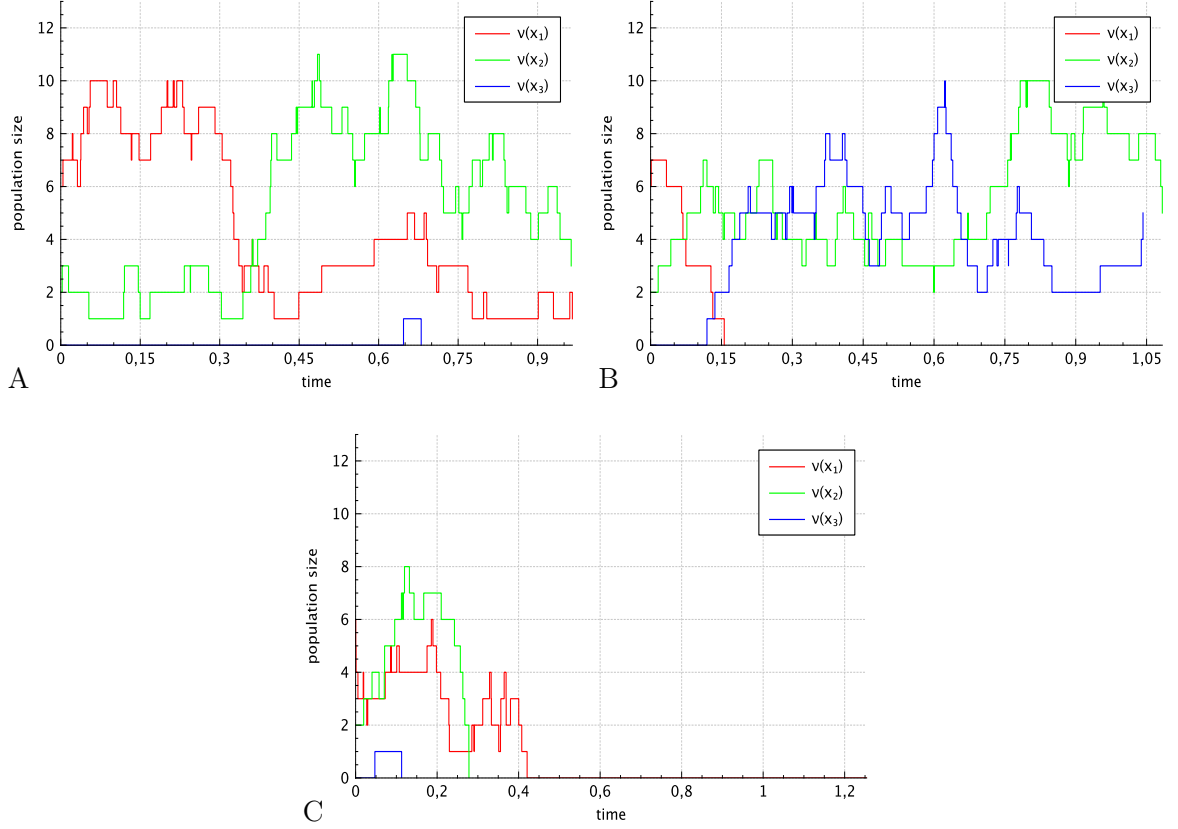
Figure I.4: Simulations of the process ν with $\nu_0 = 6\delta_{x_1} + 2\delta_{x_2}$. The biological parameters are given above. In A and C the mutant trait x_3 dies directly after it has appeared, whereas it starts growing in B. Furthermore, in C the whole population is extinct after a relatively short time.

Hence, $\mathbb{P}[x^* = x_1] = 102/138$ and $\mathbb{P}[x^* = x_2] = 36/138$ in the example. Let E be the event that the individual with trait x^* causes at time τ_1, then we can sample E according to

$$\mathbb{P}[E = \text{Birth without mutation}] \quad = \quad \frac{b(x^*)(1 - m(x^*))}{b(x^*) + d(x^*) + \sum_{y \in Supp(\nu_0)} c(x^*, y)\nu_0(y)}, \quad \text{(I.2.14)}$$

$$\mathbb{P}[E = \text{Birth with mutation}] \quad = \quad \frac{b(x^*)m(x^*)}{b(x^*) + d(x^*) + \sum_{y \in Supp(\nu_0)} c(x^*, y)\nu_0(y)}, \quad \text{(I.2.15)}$$

$$\mathbb{P}[E = \text{Death}] \quad = \quad \frac{d(x^*) + \sum_{y \in Supp(\nu_0)} c(x^*, y)\nu_0(y)}{b(x^*) + d(x^*) + \sum_{y \in Supp(\nu_0)} c(x^*, y)\nu_0(y)}. \quad \text{(I.2.16)}$$

If E = Birth without mutation, then $\nu_{\tau_1} = \nu_0 + \delta_{x^*}$. If E = Birth with mutation, then $\nu_{\tau_1} = \nu_0 + \delta_{x^*+h}$, where h is sampled according to $M(x^*, .)$, and finally if E = Death, then $\nu_{\tau_1} = \nu_0 - \delta_{x^*}$. Now we can iterate, using ν_{τ_1} instead of ν_0.

In Algorithm 1, we give the pseudocode that is the basis for all simulations we have shown in this introduction. (The algorithm for the stochastic model of cancer immunotherapy is an extension of this one and given in the appendix of Chapter III.) Below, we use the following notation: let $\mathcal{D}$ be some discrete set and X a $\mathcal{D}$-valued random variable, then X is sampled according to the weights $\{w(i), i \in \mathcal{D}\}$, which means that $\mathbb{P}(X = i) = w(i)/\sum_{i \in \mathcal{D}} w(i)$.

Algorithm 1: Pseudo-code of the Gillespie algorithm used for generating the figures.

Data: Initial conditions: $\nu_0 \in \mathcal{M}_P(\mathcal{X})$, Iterations: $N_{\text{Iterations}}$, Biological Parameters

$T_0 \leftarrow 0,\ \nu_{T_0} \leftarrow \nu_0,\ k \leftarrow 0$

while $k \leq N_{\text{Iterations}}$ **do**

    for $x \in Supp(\nu_{T_k})$ **do**

        $\mathrm{B}(x) \leftarrow \nu_{T_k}(x)b(x),$

        $\mathrm{D}(x) \leftarrow d(x) + \sum_{y \in \text{Supp}(\nu_{T_k})} c(x,y)\nu_{T_k}(y)$

    $\text{TotalEventRate} \leftarrow \sum_{x \in \text{Supp}(\nu_{T_k})} \mathrm{B}(x) + \mathrm{D}(x),$

    Sample $t \sim \mathcal{E}\mathrm{xp}(\text{TotalEventRate}),$

    $T_{k+1} \leftarrow T_k + t,$

    Sample $x^* \in \mathcal{X}$ according to the weights $\{\mathrm{B}(x) + \mathrm{D}(x),\ x \in \text{Supp}(\nu_{T_k})\},$

    Sample $E \in \{\text{Birth}, \text{Mutation}, \text{Death}\}$ according to $\{\mathrm{B}(x^*)(1 - m(x^*)), \mathrm{B}(x^*)m(x^*), \mathrm{D}(x^*)\}$

    case $E = Birth$

        $\nu_{T_{k+1}} \leftarrow \nu_{T_k} + \delta_{x^*}$

    case $E = Mutation$

        Sample h according to $M(x^*, dh),$

        $\nu_{T_{k+1}} \leftarrow \nu_{T_k} + \delta_{x^*+h}$

    case $E = Death$

        $\nu_{T_{k+1}} \leftarrow \nu_{T_k} - \delta_{x^*}$

    $k \leftarrow k + 1,$

Note that this algorithm is only reasonable if the population consists of many individuals of the same type. In this thesis we are only interested in this case since we study the limit of rare mutational events. To describe the dispersion of a plant population or, more general, to describe spatial patterns in biological populations, which was the intention in [15] and [43], it is maybe more convenient to use the algorithm provided in [59].

In [30], Champagnat and Méléard give an example with simulations where $\mathcal{X}$ is not discrete. The parameters of their example are adapted from a classical model of competition for resources (cf. references of [30]):

$$\mathcal{X} = [-2, 2],\ \ d(x) \equiv 0,\ \ m(x) \equiv 0.1,\ \ b(x) = \exp\left(\frac{-x^2}{2\sigma_b^2}\right),\ \ c(x,y) = \exp\left(\frac{-(x-y)^2}{2\sigma_c^2}\right)K^{-1},$$

and $M(x, dh)$ is the law of a normal distributed random variable Y with mean 0 and variance σ conditioned on $x + Y \in \mathcal{X}$. Numerical simulations are provided for $K = 1000$, $\sigma = 0.01$, $\sigma_b = 0.9$, $\sigma_c = 1$ and $K = 1000$, $\sigma = 0.01$, $\sigma_b = 0.9$, $\sigma_c = 0.7$. In both simulations the following is observed: In a first phase, the population trait support, $\text{Supp}(\nu_t)$, is concentrated around a single trait value that converges to zero. (The reason for this is that the growth rate is maximal at $x = 0$.) In a second phase, there are two contrary selective pressures: mutant traits close to zero have high birth rates but mutant traits far from the rest of the population compete less. If σ_c is small enough, the decrease in competitive pressure compensates the loss of reproductive efficiency and allows the appearance of new branches. Therefore, in the simulation with $\sigma_c = 0.7 < \sigma_b = 0.9$, see Figure I.5 B, an evolutionary branching is observed (the population splits into two subpopulations of different dominant types going their separate

ways), but in the simulation with $\sigma_c = 1 > \sigma_b = 0.9$ not, see Figure I.5 A. Figure I.5 is taken from [30] (cf. [30] Fig. 1).
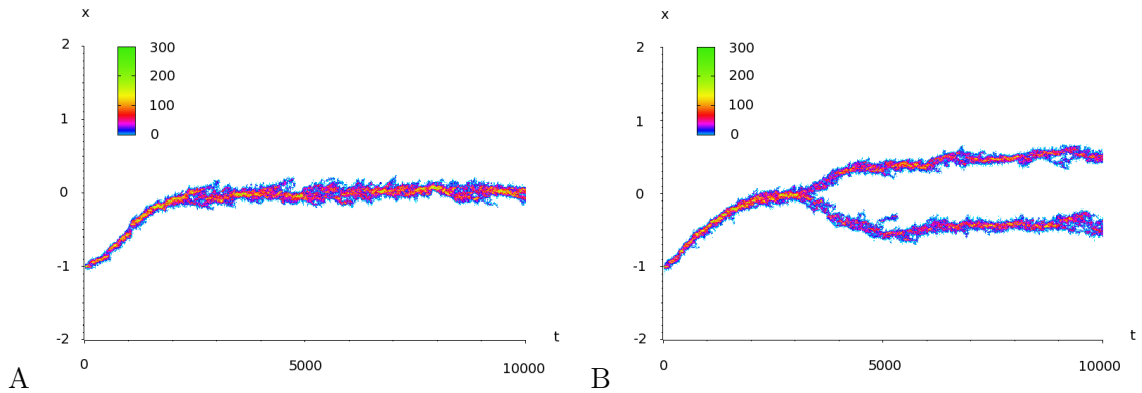
Figure I.5: Numerical simulations taken from Champagnat and Méléard, Fig. 1 of [30]. In A, where $K = 1000$, $\sigma = 0.01$, $\sigma_b = 0.9$, and $\sigma_c = 1$, there is no evolutionary branching. In B, where $K = 1000$, $\sigma = 0.01$, $\sigma_b = 0.9$, and $\sigma_c = 0.7$, there is an evolutionary branching: the population splits into two subpopulations of different dominant types going their separate ways.

I.3 The main results and outline of this thesis

This thesis is divided into two parts, which are the content of Chapters II and III. The topics are related, however, the chapters can be read independent from each other. In the first part, we study the long-term behavior of the stochastic individual-based model, defined in Section I.2.2, in the limit of large population size combined with rare mutations and small mutational effects, proving convergence to the canonical equation of adaptive dynamics (CEAD). This part will appear in Annals of Applied Probability as joint work with Anton Bovier and Nicolas Champagnat [8]. At the moment, it is available on www.imstat.org/aap/future_papers.html

> M. Baar, A Bovier, and N. Champagnat. *From stochastic, individual-based models to the canonical equation of adaptive dynamics - In one step.* Preprint (arXiv:1505.02421, 2015), to appear in Ann. Appl. Probab., 2016.

Chapter II contains this preprint, only some minor changes have been made and the layout has been adapted to the layout of this thesis.

For studying the three limits, large population size, rare mutations, and small mutational effects, we have to include the following scaling parameters in the individual-based model defined in Section I.2.2.

K is the *carry capacity* of the system meaning that the competition kernel and the point measure describing the population state are rescaled by K, i.e. we use $K^{-1}c(x, y)$ instead of $c(x, y)$ and $\nu^K \equiv \frac{1}{K}\nu$ instead of ν.

u_K is the *scaling parameter for the mutation probability*, i.e. we use $u_K m(x)$ instead of $m(x)$.

σ_K is the *scaling parameter for the mutation size*, i.e. a mutant's trait y is given by $x + \sigma_K h$ instead of $x + h$, where x is the parent's trait and $h \sim M(x, dh)$.

Note that the carrying capacity K only gives the magnitude of the population size, i.e. the actual population size changes over time, but stays of order K (cf. Section I.2.1).

In the main result of Chapter II, Theorem II.4.1, we identify a time scale on which the sequence of measure-valued Markov processes ν^K, describing the evolution of the population, converges in law to the CEAD if the three limits are taken simultaneously, i.e. u_K and σ_K will tend to zero as $K \to \infty$. On this time scale, the evolution of populations with monomorphic initial condition (meaning that all individuals have identically phenotypic traits) can be described as a succession of mutant invasions. More precisely, we use conditions on the scaling parameters that imply a separation of ecological and evolutionary time in the sense that if a mutation occurs, either the individuals with this mutant trait die out or the mutant population invades the resident population and converges to its ecological equilibrium, replacing the resident population entirely, before a new invading (successful) mutant appears. However, we allow non-invading (unsuccessful) mutation events during this time to avoid too restrictive assumptions. On a longer time scale, the single invasion steps accumulate and give rise to a macroscopic evolution of the population state, which can be described in terms of the CEAD. In other words, the population stays essentially monomorphic with a trait evolving in time according to this ordinary differential equation. The combination of the three simultaneous limits entails some considerable technical difficulties. The fact that the mutants only have a K-dependent small evolutionary advantage implies that the time of any macroscopic change in population diverges with K and makes it impossible to use a law of large numbers as in [25] to approximate the stochastic system with the corresponding deterministic system during the time of invasion. Showing that the stochastic system still follows the corresponding competition Lotka-Volterra system (with K-dependent coefficients) in an appropriate sense, requires a completely new approach. Developing this approach is the main novelty of the present thesis. The proof requires methods based on couplings with discrete time Markov chains, combined with some standard potential theoretic arguments for the exit from a domain problem in a moderate deviations regime, as well as comparison and convergence results of branching processes.

To be a bit more precise, we use the following conditions on the scaling parameters. There exists a small $\alpha > 0$ such that

$$K^{-1/2+\alpha} \ll \sigma_K \ll 1 \qquad \text{and} \qquad \text{(I.3.1)}$$

$$\exp(-K^\alpha) \ll u_K \ll \frac{\sigma_K^{1+\alpha}}{K \ln K}, \quad \text{as} \quad K \to \infty, \qquad \text{(I.3.2)}$$

where $f(K) \ll g(K)$ means that $f(K)/g(K)$ converges to zero as $K \to \infty$. The time scale, on which we control the population process, is $t/(K u_K \sigma_K^2)$ and can be explained as follows. The expected time for a mutation event to happen is of order $(K u_K)^{-1}$, the probability that a mutant invades is of order σ_K, and an order of σ_K^{-1} mutant invasions is necessary to observe a macroscopic change in the trait value of the essentially monomorphic population in the limit. Furthermore, the conditions on the scaling parameters can be explained as follows. First, we can prove that the duration of one invasion phase is of order $\sigma_K^{-1} \ln(K)$. Therefore, the condition $u_K \ll \sigma_K^{1+\alpha}/(K \ln K)$ allows mutation events during some invasion phases but ensures that there are never two invading (successful) mutational events during such a phase and thus guarantees the separation of the ecological and evolutionary time scale. Second, the random fluctuations of the population process are of order $K^{-1/2}$, therefore the condition

$K^{-1/2+\alpha} \ll \sigma_K$ ensures that the sign of a mutant's fitness (evolutionary advantage) is not influenced by the fluctuations of the population size. Finally, we prove that $\exp(K^\alpha)$ is the time the population stays within an $O(\epsilon\sigma_K)$-neighborhood of an attractive domain with high probability. This is a moderate deviation result. Therefore, the condition $\exp(-K^\alpha) \ll u_K$ ensures that the population is still in this neighborhood when a mutation occurs.

In the third chapter of the thesis we expand the stochastic individual-based models used in adaptive dynamics in order to describe cancer immunotherapy and to study the interplay between phenotypic plasticity modeled by fast phenotypic switches and rare genotypic mutations on a long evolutionary time scale. Some parts of this chapter have been published in *Scientific Reports* as the following joint work with Loren Coquille, Hannah Mayer, Michael Hölzel, Meri Rogava, Thomas Tüting, and Anton Bovier.

> M. Baar, L. Coquille, H. Mayer, M. Hölzel, M. Rogava, T. Tüting, and
> A. Bovier. *A stochastic model for immunotherapy of cancer.* Scientific
> Reports, 6: 24169, 2016.

On the arXiv webpage, there is also a more detailed preprint of this publication available (cf. arXiv:1505.00452, 2015). This arXiv version provides a basis for the third chapter of this thesis. Sections III.1, III.3, III.4.3, and III.5 are taken from this preprint, only some minor changes have been done. Section III.2 is rewritten such that the notations fit better to the rest of this thesis and to be able to focus on more theoretical aspects in Section III.4. Section III.4 contains the main mathematical work. In the publication as well as in the preprint, we give some heuristic arguments why we should obtain a generalization of the polymorphic evolution sequence in the new setup, whereas in Section III.4 of this thesis we give the proper convergence result (cf. Thm. III.4.3) and a rigorous proof. Sections 3.2 and 3.3 of the preprint are not used in this thesis. All figures and simulations in Chapter III are either new or redone with new sets of parameters, except for the figures in Section III.4.3. All simulations are based on a Gillespie algorithm implemented by Boris Prochnau.

The publication was initiated by a request from the medical scientists Hölzel and Tüting, asking whether it might be possible to describe cancer immunotherapy with mathematical models. They emphasized the importance of developing a theoretical framework that incorporates and thus may explain the different aspects of therapy resistance [76]. The primary motivation was that such a model should be able to reproduce the key phenomena of the experiments reported in Landsberg et al. [92]. In these experiments tumors (melanoma) resist immunotherapy through an inflammation-induced reversible dedifferentiation, i.e. through phenotypic plasticity. The identification of the most relevant underlying mechanisms leading to resistance of immunotherapy and the incorporation of them in a mathematical model, which on the one hand is able to describe the experiments but stays on the other hand simple enough to make numerical computations and theoretical understanding of the key phenomena feasible, was the starting point and an important part of this work. As already mentioned in Section I.1, there are several reasons for using an expansion of the stochastic individual-based model defined in Section I.2.2 to model cancer immunotherapy. For example, these type of models allow to describe several ecological situations and do not have a fixed population size as for example Wright-Fisher models. This is important since tumor development is environment depending and the size of the tumor varies over time, especially during therapy. The model we study in Chapter III is defined in Section III.2. The main expansions are:

(i) Three different classes of actors are included: T-cells, cytokines, and cancer cells.

(ii) For cancer cells, two types of transitions are allowed: genotypic mutations and phenotypic switches.

(iii) Phenotypic changes can be affected by the microenvironment which is not modeled deterministically but as particles undergoing the random dynamics as well.

(iv) For modeling the therapy effect, a predator-prey mechanism (between cancer cells and immune cells) is included.

To our knowledge it is a novel feature of our models to describe the coevolution of immune and tumor cells taking into account both, interactions and phenotypic plasticity. It is well known that in the limit of large cell populations, these models are approximated by deterministic kinetic rate models (cf. Theorem III.2.1), which are widely used in the modeling of cell populations. However, these approximations are inaccurate if the numbers of individuals in some sub-populations become small. In Section III.3 we give an example describing the experiments of [92] qualitatively. We explain why random fluctuation might be the reason for the different outcomes in the experiments. Note that it can be shown that the models are capable to reproduce the experimental data quantitatively, with biological reasonable parameters. Furthermore, the model can be used as a tool to assist the development of improved treatment protocols, for example to study the scenario with two types of T-cells. However, this is not part of this thesis, for details see either the publication or [97]. This thesis focuses more on the new theoretical aspects which arise by incorporating phenotypic plasticity in the standard model. More precisely, in Section III.4 we analyze the interplay between the fast phenotypical changes by switching and the slow genotypical changes by mutation. Without phenotypic plasticity and under conditions which separate the ecological and the evolutionary effects, the evolution of the system can be described by a succession of mutant invasions. This was studied mathematically rigorous by Champagnat and Méléard in [25, 30]. The main assumptions leading to this separation are that the population size is large and that mutational events are rare. Thus, a natural question is whether this is still true if we incorporate phenotypic plasticity in the model. Conditions such that this approximation is correct are given in Section III.4, in the sense that the individual-based process, which includes phenotypic plasticity, converges in the simultaneous limits of a large population size ($K \to \infty$) and rare mutational events ($u_K \to 0$) to a Markov jump process. More precisely, we prove by expanding the techniques of [25] that the microscopic process converges in this limit on the evolutionary time scale to a generalization of the polymorphic evolution sequences (cf. Theorem III.4.3). The main difference in the proof is that we have to couple the process with multi-type branching processes instead of normal branching processes, which also leads to a different definition of invasion fitness in this setting. Furthermore, we discuss the interplay of mutation and therapy.

Chapter II

From stochastic, individual-based models to the canonical equation of adaptive dynamics - in one step.

The following chapter is already published online on the webpage of *Annals of Applied Probability* (www.imstat.org/aap/future_papers.html) and will appear in this journal as joint work with Anton Bovier and Nicolas Champagnat [8].

We consider a model for Darwinian evolution in an asexual population with a large but non-constant populations size characterized by a natural birth rate, a logistic death rate modeling competition and a probability of mutation at each birth event. In the present paper, we study the long-term behavior of the system in the limit of large population ($K \to \infty$) size, rare mutations ($u \to 0$), and small mutational effects ($\sigma \to 0$), proving convergence to the canonical equation of adaptive dynamics (CEAD). In contrast to earlier works, e.g. by Champagnat and Méléard, we take the three limits simultaneously, i.e., $u = u_K$ and $\sigma = \sigma_K$, tend to zero with K, subject to conditions that ensure that the time scale of birth and death events remains separated from that of successful mutational events. This slows down the dynamics of the microscopic system and leads to serious technical difficulties that require the use of completely different methods. In particular, we cannot use the law of large numbers on the diverging time needed for fixation to approximate the stochastic system with the corresponding deterministic one. To solve this problem we develop a "stochastic Euler scheme" based on coupling arguments that allows to control the time evolution of the stochastic system over time scales that diverge with K.

II.1 Introduction

In this paper we study a microscopic model for evolution in a population characterized by a birth rate with a probability of mutation at each event and a logistic death rate, which has been studied in many works before [25, 27, 28, 30, 59]. More precisely, it is a model for an asexual population in which each individual's ability to survive and to reproduce is a function of a one-dimensional phenotypic trait, such as body size, the age at maturity, or the rate of food intake. The evolution acts on the trait distribution and is the consequence of three basic mechanisms: heredity, mutation and selection. Heredity passes the traits trough generations, mutation drives the variation of the trait values in the population, and selection acts on individuals with different traits and is a consequence of competition between the individuals

for limited resources or area.

The model is a generic stochastic individual-based model and belongs to the models of adaptive dynamics. In general, adaptive dynamics models aim to study the interplay between ecology (viewed as driving selection) and evolution, more precisely, the interplay between the three basic mechanisms mentioned above. It tries to develop general tools to study the long time evolution of a wide variety of ecological scenarios [42, 46, 104]. These tools are based on the assumption of separation of ecological and evolutionary time scales and on the notion of invasion fitness [102, 103]. While the biological theory of adaptive dynamics is based on partly heuristic derivations, various aspects of the theory have been derived rigorously over the last years in the context of stochastic individual-based models [25, 27, 28, 30, 64, 74]. All of them concern the limit when the population size, K, tends to infinity. They either study the separation of ecological and evolutionary time scales based on a limit of rare mutations, $u \to 0$, combined with a limit of large population [25, 30], the limit of small mutation effects, $\sigma \to 0$, [27, 30, 64], the stationary behavior of the system [74], or the links between individual-based and infinite-population models [28]. An important concept in the theory of adaptive dynamics is the canonical equation of adaptive dynamics (CEAD), introduced by U. Dieckmann and R. Law [42]. This is an ODE that describes the evolution in time of the expected trait value in a monomorphic population. The heuristics leading to the CEAD are based on the biological assumptions of large population and rare mutations with small effects and the assumption that no two different traits can coexist. (Note that we write sometimes mutation steps instead of effects.) There are mathematically rigorous papers that show that the limit of large population combined with rare mutations leads to a jump process, the Trait Substitution Sequence, [25], and that this jump process converges, in the limit of small mutation steps, to the CEAD, [30]. Since these two limits are applied separately and on different time scales, they give no clue about how the biological parameters (population size K, probability of mutations u and size of mutation steps σ) should compare to ensure that the CEAD approximation of the individual-based model is correct.

The purpose of the present paper is to analyze the situation when the limits of large population size, $K \to \infty$, rare mutations, $u_K \to 0$, and small mutation steps, $\sigma_K \to 0$, are taken *simultaneously*. We consider populations with monomorphic initial condition, meaning that at time zero the population consists only of individuals with the same trait. Then we identify a time scale where evolution can be described as a succession of mutant invasions. To prove convergence to the CEAD, we show that, if a mutation occurs, then the individuals with this mutant trait can either die out or invade the resident population on this time scale. Here invasion means that the mutant trait supersedes the resident trait, i.e., the individuals with the resident trait become extinct after some time. This implies that the population stays essentially monomorphic with a trait that evolves in time. We will impose conditions on the mutation rates that imply a separation of ecological and evolutionary time scales, in the sense that an invading mutant population converges to its ecological equilibrium before a new invading (successful) mutant appears. In order to avoid too restrictive hypothesis on the mutation rates, we do, however, allow non-invading (unsuccessful) mutation events during this time, in contrast to all earlier works.

We will see that the combination of the three limits simultaneously, entails some considerable technical difficulties. The fact that the mutants have only a K-dependent small evolutionary advantage decelerates the dynamics of the microscopic process such that the time of any macroscopic change between resident and mutant diverges with K. This makes it impossible to use a law of large numbers as in [25] to approximate the stochastic system with the corresponding deterministic system during the time of invasion. Showing that the stochas-

tic system still follows in an appropriate sense the corresponding competition Lotka-Volterra system (with K-dependent coefficients) requires a completely new approach. Developing this approach, which can be seen as a rigorous "stochastic Euler-scheme", is the main novelty of the present paper. The proof requires methods, based on couplings with discrete time Markov chains combined with some standard potential theory arguments for the "exit from a domain problem" in a moderate deviations regime, as well as comparison and convergence results of branching processes. Note that since the result of [25] is already different from classical time scales separations results (cf. [61]), our result differs from them a fortiori. Thus, our result can be seen as a rigorous justification of the biologically motivated, heuristic assumptions which lead to CEAD.

The remainder of this paper is organized as follows. In Section II.2 and II.3 we introduce the model and give an overview on previous related results. In Section II.4 we state our results and give a detailed outline of the proof. Full details of the proof are presented in the Section II.6, II.7 and II.8. In the appendix we state and prove several elementary facts that are used throughout the proof.

II.2 The individual-based model

In this section we introduce the model we analyze. We consider a population of a single asexual species that is composed of a finite number of individuals, each of them characterized by a one-dimensional phenotypic trait. The microscopic model is an individual-based model with non-linear density-dependence, which has already been studied in ecological or evolutionary contexts by many authors [28, 25, 30, 59]. The *trait space* $\mathcal{X}$ is assumed to be a compact interval of $\mathbb{R}$. We introduce the following biological parameters:

(i) $b(x) \in \mathbb{R}_+$ is the *rate of birth* of an individual with trait $x \in \mathcal{X}$.

(ii) $d(x) \in \mathbb{R}_+$ is the *rate of natural death* of an individual with trait $x \in \mathcal{X}$.

(iii) $K \in \mathbb{N}$ is a parameter which scales the population size.

(iv) $c(x,y)K^{-1} \in \mathbb{R}_+$ is the *competition kernel* which models the competitive pressure felt by an individual with trait $x \in \mathcal{X}$ from an individual with trait $y \in \mathcal{X}$.

(v) $u_K m(x)$ with $u_K, m(x) \in [0,1]$ is the *probability that a mutation occurs at birth* from an individual with trait $x \in \mathcal{X}$, where $u_K \in [0,1]$ is a scaling parameter.

(vi) $M(x,dh)$ is the *mutation law* of the mutational jump h. If the mutant is born from an individual with trait x, then the mutant trait is given by $x + \sigma_K h \in \mathcal{X}$, where $\sigma_K \in [0,1]$ is a parameter scaling the size of mutation and h is a random variable with law $M(x,dh)$. We restrict for simplicity the setting to mutation measures with support included in $\mathbb{Z}$.

The three scaling parameters of the model are the *population size*, controlled by the scaling parameter K, the *mutation probability*, controlled by the scaling parameter u_K, the *mutation size*, controlled by the scaling parameter σ_K. The novelty of our approach is that we consider the case where all these parameters tend to their limit jointly, more precisely that both u_K and σ_K are functions of K and tend to zero as K tends to infinity (subject to certain constraints).

At any time t we consider a finite number, N_t, of individuals, each of them having a trait value $x_i(t) \in \mathcal{X}$. It is convenient to represent the population state at time t by the rescaled point measure, ν^K, which depends on K, u_K and σ_K

$$\nu_t^K = \frac{1}{K} \sum_{i=1}^{N_t} \delta_{x_i(t)}. \tag{II.2.1}$$

Let $\langle \mu, f \rangle$ denote the integral of a measurable function f with respect to the measure μ. Then $\langle \nu_t^K, \mathbb{1} \rangle = N_t K^{-1}$ and for any $x \in \mathcal{X}$, the positive number $\langle \nu_t^K, \mathbb{1}_{\{x\}} \rangle$ is called the *density of trait x at time t*. With this notation, an individual with trait x in the population ν_t^K dies due to age or competition with rate

$$d(x) + \int_{\mathcal{X}} c(x,y) \nu_t^K(dy). \tag{II.2.2}$$

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

$$\mathcal{M}^K(\mathcal{X}) \equiv \left\{ \frac{1}{K} \sum_{i=1}^n \delta_{x_i} : n \ge 0, \ x_1, ..., x_n \in \mathcal{X} \right\}. \tag{II.2.3}$$

Similar as in [59], we obtain that the population process, $(\nu_t^K)_{t \ge 0}$, is a $\mathcal{M}^K(\mathcal{X})$-valued Markov process with infinitesimal generator, $\mathscr{L}^K$, defined for any bounded measurable function f from $\mathcal{M}^K(\mathcal{X})$ to $\mathbb{R}$ and for all $\mu^K \in \mathcal{M}^K(\mathcal{X})$ by

$$\mathscr{L}^K f(\mu^K) = \int_{\mathcal{X}} \left(f\left(\mu^K + \frac{\delta_x}{K}\right) - f(\mu^K) \right) \left(1 - u_K m(x)\right) b(x) \, K\mu^K(dx) \tag{II.2.4}$$

$$+ \int_{\mathcal{X}} \int_{\mathbb{Z}} \left(f\left(\mu^K + \frac{\delta_{x+\sigma_K h}}{K}\right) - f(\mu^K) \right) u_K m(x) b(x) \, M(x,dh) \, K\mu^K(dx)$$

$$+ \int_{\mathcal{X}} \left(f\left(\mu^K - \frac{\delta_x}{K}\right) - f(\mu^K) \right) \left(d(x) + \int_{\mathcal{X}} c(x,y)\mu^K(dy) \right) K\mu^K(dx).$$

The first and second terms are linear (in μ^K) and describe the births (without and with mutation), but the third term is non-linear and describes the deaths due to age or competition. The density-dependent non-linearity of the third term models the competition in the population, and hence drives the selection process.

Assumption 2. We will use the following assumptions on the parameters of the model:

(i) b, d and c are measurable functions, and there exist $\bar{b}$, $\bar{d}, \bar{c} < \infty$ such that

$$b(.) \le \bar{b}, \quad d(.) \le \bar{d} \quad \text{and} \quad c(.,.) \le \bar{c}.$$

(ii) For all $x \in \mathcal{X}$, $b(x) - d(x) > 0$, and there exists $\underline{c} > 0$ such that $\underline{c} \le c(x,x)$.

(iii) The support of $M(x, .)$ is a subset of $\mathbb{Z} \cap \mathcal{X} - x$ and uniformly bounded for all $x \in \mathcal{X}$. This means that there exists $A \in \mathbb{N}$ such that

$$M(x,dh) = \sum_{k=-A}^A p_k(x)\delta_k(dh), \quad \text{where } \sum_{k=-A}^A p_k(x) = 1 \text{ for any } x \in \mathcal{X}.$$

(iv) $b, d, m \in C^2(\mathcal{X}, \mathbb{R})$ and $c \in C^2(\mathcal{X}^2, \mathbb{R})$.

Assumptions (i) and (iii) allow to deduce the existence and uniqueness in law of a process on $\mathbb{D}(\mathbb{R}_+, \mathcal{M}^K(\mathcal{X}))$ with infinitesimal generator $\mathscr{L}^K$ (cf. [59]). Note that Assumption (iii) differs from the assumptions in [59] because we restrict the setting to mutation measures with support included in $\mathbb{Z}$ and that it ensures that a mutant trait remains in $\mathcal{X}$. Assumption (ii) prevents the population from exploding or becoming extinct too fast. Since $\mathcal{X}$ is compact, Assumption (iv) ensures that the derivatives of the functions b, c, d and m are uniformly Lipschitz-continuous.

Before we state the main result of the paper, Theorem II.4.1, in Section II.4, it will be helpful to recall some earlier results for this class of models and to fix some more notation. These results serve as a guideline to what behavior one should expect, even though on a technical level proofs have to be changed completely.

II.3 Some notation and previous results

We start with a theorem due to N. Fournier and S. Méléard [59] which describes the behavior of the population process, for fixed u and σ, when $K \to \infty$.

Theorem II.3.1 (Theorem 5.3 in [59])**.** *Fix u and σ. Let Assumption 2 hold and assume in addition that the initial conditions ν_0^K converge for $K \to \infty$ in law and for the weak topology on $\mathcal{M}(\mathcal{X})$ to some deterministic finite measure $\xi_0 \in \mathcal{M}(\mathcal{X})$ and that $\sup_K \mathbb{E}[\langle \nu_0^K, \mathbb{1} \rangle^3] < \infty$.*

Then for all $T > 0$, the sequence ν^K, generated by $\mathscr{L}^K$, converges for $K \to \infty$ in law, in $\mathbb{D}([0, T], \mathcal{M}(\mathcal{X}))$, to a deterministic continuous function $\xi \in C([0, T], \mathcal{M}(\mathcal{X}))$. This measure-valued function ξ is the unique solution, satisfying $\sup_{t \in [0,T]} \langle \xi_t, \mathbb{1} \rangle < \infty$, of the integro-differential equation written in its weak form: for all bounded and measurable functions, $f : \mathcal{X} \to \mathbb{R}$,

$$\int_{\mathcal{X}} \xi_t(dx) f(x) = \int_{\mathcal{X}} \xi_0(dx) f(x) + \int_0^t ds \int_{\mathcal{X}} \xi_s(dx) u m(x) b(x) \int_{\mathbb{Z}} M(x, dh) f(x + \sigma h) \quad \text{(II.3.1)}$$

$$+ \int_0^t ds \int_{\mathcal{X}} \xi_s(dx) f(x) \Big((1 - u m(x)) b(x) - d(x) - \int_{\mathcal{X}} \xi_s(dy) c(x, y) \Big).$$

Without mutation one obtains convergence to the competitive system of Lotka-Volterra equations defined below (see [59]).

Corollary II.3.2 (The special case $u = 0$ and ξ_0 is n-morphic)**.** *If the same assumptions as in the theorem above with $u = 0$ hold and if in addition $\xi_0 = \sum_{i=1}^n z_i(0) \delta_{x_i}$, then ξ_t is given by $\xi_t = \sum_{i=1}^n z_i(t) \delta_{x_i}$, where z_i is the solution of the competitive system of Lotka-Volterra equations defined below.*

Definition II.3.3. For any $(x_1, ..., x_n) \in \mathcal{X}^n$, we denote by $LV(n, (x_1, ..., x_n))$ the *competitive system of Lotka-Volterra equations* defined by

$$\frac{d z_i(t)}{dt} = z_i \Big(b(x_i) - d(x_i) - \sum_{j=1}^n c(x_i, x_j) z_j \Big), \qquad 1 \le i \le n. \quad \text{(II.3.2)}$$

Next, we introduce the notation of coexisting traits and of invasion fitness (see [30]).

Definition II.3.4. We say that the distinct traits x and y *coexist* if the system $LV(2, (x, y))$ admits a unique non-trivial equilibrium, named $\bar{z}(x, y) \in (0, \infty)^2$, which is locally strictly stable in the sense that the eigenvalues of the Jacobian matrix of the system $LV(2, (x, y))$ at $\bar{z}(x, y)$ are all strictly negative.

The invasion of a single mutant trait in a monomorphic population which is close to its equilibrium is governed by its initial growth rate. Therefore, it is convenient to define the fitness of a mutant trait by its initial growth rate.

Definition II.3.5. If the resident population has the trait $x \in \mathcal{X}$, then we call the following function *invasion fitness* of the mutant trait y

$$f(y, x) = b(y) - d(y) - c(y, x) \bar{z}(x). \quad \text{(II.3.3)}$$

Remark 1. The unique strictly stable equilibrium of $LV(1, x)$ is $\bar{z}(x) = \frac{b(x) - d(x)}{c(x,x)}$, and hence $f(x, x) = 0$ for all $x \in \mathcal{X}$.

Coexistence and invasion fitness are closely related (cf. [79]).

Proposition II.3.6. *There is coexistence in the system $LV(2, (x, y))$ if and only if*

$$f(x, y) \equiv b(x) - d(x) - c(x, y)\overline{z}(y) > 0 \quad and \quad f(y, x) \equiv b(y) - d(y) - c(y, x)\overline{z}(x) > 0. \quad \text{(II.3.4)}$$

The following convergence result from [25] describes the limit behavior of the populations process, for fixed σ, when $K \to \infty$ and $u_K \to 0$. More precisely, it says that the rescaled individual-based process converges in the sense of finite dimensional distributions to the "trait substitution sequence" (TSS), if one assumes in addition to Assumption 2 the following "Invasion implies fixation" condition.

Assumption 3. Given any $x \in \mathcal{X}$, Lebesgue almost any $y \in \mathcal{X}$ satisfies one of the following conditions: (i) $f(y, x) < 0$ or   (ii) $f(y, x) > 0$ and $f(x, y) < 0$.

Note that by Proposition II.3.6, this means that either a mutant cannot invade, or it cannot coexist with the resident.

Theorem II.3.7 (Corollary 1 in [25])**.** *Let Assumption 2 and 3 hold. Fix σ and assume that*

$$\forall V > 0, \qquad \exp(-VK) \ll u_K \ll \frac{1}{K \ln(K)}, \qquad as \ K \to \infty. \quad \text{(II.3.5)}$$

Fix also $x \in \mathcal{X}$ and let $(N_0^K)_{K \geq 1}$ be a sequence of $\mathbb{N}$-valued random variables such that (N_0^K / K) converges for $K \to \infty$ in law to $\overline{z}(x)$ and is bounded in $\mathbb{L}^p$ for some $p > 1$. Consider the processes ν^K generated by $\mathscr{L}^K$ with monomorphic initial state $(N_0^K / K)\delta_{\{x\}}$.
Then the sequence of the rescaled processes ν_{t/Ku_K}^K converges in the sense of finite dimensional distributions to the measure-valued process

$$\overline{z}(X_t)\delta_{X_t}, \quad \text{(II.3.6)}$$

where the $\mathcal{X}$-valued Markov jump process X has initial state $X_0 = x$ and infinitesimal generator

$$A\phi(x) = \int_{\mathbb{Z}} (\phi(x + \sigma h) - \phi(x)) \, m(x)b(x)\overline{z}(x) \frac{[f(x + \sigma h, x)]_+}{b(x + \sigma h)} M(x, dh). \quad \text{(II.3.7)}$$

Here we write $f(K) \ll g(K)$ if $f(K)/g(K) \to 0$ when $K \to \infty$. Note that, for any $s < t$, the convergence does not hold in law for the Skorokhod topology on $\mathbb{D}([s, t], \mathcal{M}(\mathcal{X}))$, for any topology $\mathcal{M}(\mathcal{X})$ such that the total mass function $\nu \mapsto \langle \nu, \mathbb{1} \rangle$ is continuous, because the total mass of the limit process is a discontinuous function. The main part of the proof of this theorem is the study of the invasion of a mutant trait y that has just appeared in a monomorphic population with trait x. The invasion can be divided into three steps. Firstly, as long as the mutant population size $\langle \nu_t^K, \mathbb{1}_{\{y\}} \rangle$ is smaller than some $\epsilon > 0$ (independent of K), the resident population size $\langle \nu_t^K, \mathbb{1}_{\{x\}} \rangle$ stays close to $\overline{z}(x)$. Therefore, $\langle \nu_t^K, \mathbb{1}_{\{y\}} \rangle$ can be approximated by a branching process with birth rate $b(y)$ and death rate $d(y) + c(y, x)\overline{z}(x)$ until it goes extinct or reaches ϵ. Secondly, once $\langle \nu_t^K, \mathbb{1}_{\{y\}} \rangle$ has reached ϵ, for large K, ν_t^K is close to the solution of $LV(2, (x, y))$ with initial state $(\overline{z}(x), \epsilon)$, which reaches the ϵ-neighborhood of $(0, \overline{z}(y))$ in finite time. This is a consequence of Corollary II.3.2. Finally, once $\langle \nu_t^K, \mathbb{1}_{\{y\}} \rangle$ is close to $\overline{z}(y)$ and $\langle \nu_t^K, \mathbb{1}_{\{x\}} \rangle$ is small, $\langle \nu_t^K, \mathbb{1}_{\{x\}} \rangle$ can be approximated by a subcritical process, which becomes extinct a.s. . The time of the first and third step are proportional to $\ln(K)$, whereas the time of the second step is bounded. Thus, the second inequality in (II.3.5) guarantees that, with high probability, the three steps of invasion are completed before a new mutation occurs.

Without Assumption 3 it is possible to construct the "polymorphic evolution sequence" (PES) under additional assumptions on the n-morphic logistic system. This is done in [30]. Finally, in [30], the convergence of the TSS with small mutation steps scaled by σ to the "canonical equation of adaptive dynamics" (CEAD) is proved. We indicate the dependence of the TSS of the previous Theorem on σ with the notation $(X_t^\sigma)_{t \geq 0}$.

Theorem II.3.8 (Remark 4.2 in [30])**.** *If Assumption 2 is satisfied and the family of initial states of the rescaled TSS, X_0^σ, is bounded in $\mathbb{L}^2$ and converges to a random variable X_0, as $\sigma \to 0$, then, for each $T > 0$, the rescaled TSS X_{t/σ^2}^σ converges, as $\sigma \to 0$, in the Skorokhod topology on $\mathbb{D}([0,T], \mathcal{X})$ to the process $(x_t)_{t \leq T}$ with initial state X_0 and with deterministic sample path, which is the unique solution of an ordinary differential equation, known as CEAD:*

$$\frac{dx_t}{dt} = \int_\mathbb{Z} h \left[h \, m(x_t) \, \overline{z}(x_t) \, \partial_1 f(x_t, x_t) \right]_+ M(x_t, dh), \tag{II.3.8}$$

where $\partial_1 f$ denotes the partial derivative of the function $f(x,y)$ with respect to the first variable x.

Remark 2. If $M(x, \cdot)$ is a symmetric measure on $\mathbb{Z}$ for all $x \in \mathcal{X}$, then the equation (II.3.8) has the classical form, c.f. [42],

$$\frac{d\,x_t}{dt} = \frac{1}{2} \int_\mathbb{Z} h^2 \, m(x_t) \, \overline{z}(x_t) \, \partial_1 f(x_t, x_t) M(x_t, dh), \tag{II.3.9}$$

Note that this result does not imply that, applying to the individual-based model first the limits $(K, u_K) \to (\infty, 0)$ and afterwards the limit $\sigma \to 0$ yields its convergence to the CEAD. One problem of theses two successive limits is, for example, that the first convergence holds on a finite time interval, the second requires to look at the Trait Substitution Sequence on a time interval which diverges. Moreover, as already mentioned these two limits give no clue about how K, u and σ should be compared to ensure that the CEAD approximation is correct.

II.4   The main result

In this section, we present the main result of this paper, namely the convergence to the canonical equation of adaptive dynamics in one step. The time scale on which we control the population process is $t/(\sigma_K^2 u_K K)$ and corresponds to the combination of the two time scales of Theorems II.3.7 and II.3.8. Since we combine the limits we have to modify the assumptions to obtain the convergence. We use in this section the notations and definitions introduced in Section II.3.

Assumption 4. For all $x \in \mathcal{X}, \quad \partial_1 f(x,x) \neq 0$.

Assumption 4 implies that either $\forall x \in \mathcal{X}: \partial_1 f(x,x) > 0$ or $\forall x \in \mathcal{X}: \partial_1 f(x,x) < 0$. Therefore, coexistence of two traits is not possible. Without loss of generality we can assume that, $\forall x \in \mathcal{X}$, $\partial_1 f(x,x) > 0$. In fact, a weaker assumption is sufficient, see Remark 3.(iii).

Theorem II.4.1. *Assume that Assumptions 2 and 4 hold and that there exists a small $\alpha > 0$ such that*

$$K^{-1/2+\alpha} \ll \sigma_K \ll 1 \qquad \qquad and \tag{II.4.1}$$

$$\exp(-K^\alpha) \ll u_K \ll \frac{\sigma_K^{1+\alpha}}{K \ln K}, \quad as \qquad K \to \infty. \tag{II.4.2}$$

Fix $x_0 \in \mathcal{X}$ and let $(N_0^K)_{K\geq 0}$ be a sequence of $\mathbb{N}$-valued random variables such that $N_0^K K^{-1}$ converges in law, as $K \to \infty$, to the positive constant $\overline{z}(x_0)$ and is bounded in $\mathbb{L}^p$, for some $p > 1$.

For each $K \geq 0$, let ν_t^K be the process generated by $\mathscr{L}^K$ with monomorphic initial state $N_0^K K^{-1} \delta_{\{x_0\}}$. Then, for all $T > 0$, the sequence of rescaled processes, $\left(\nu_{t/(Ku_K\sigma_K^2)}^K\right)_{0\leq t\leq T}$, converges in probability, as $K \to \infty$, with respect to the Skorokhod topology on $\mathbb{D}([0,T],\mathcal{M}(\mathcal{X}))$ to the measure-valued process $\overline{z}(x_t)\delta_{x_t}$, where $(x_t)_{0\leq t\leq T}$ is given as a solution of the CEAD,

$$\frac{dx_t}{dt} = \int_{\mathbb{Z}} h\,[h\,m(x_t)\,\overline{z}(x_t)\,\partial_1 f(x_t, x_t)]_+ M(x_t, dh), \tag{II.4.3}$$

with initial condition x_0.

Remark 3.   (i) If $x_t \in \partial\mathcal{X}$ for $t > 0$, then (II.4.3) is $\frac{d\,x_t}{dt} = 0$, i.e., the process stops.

(ii) We can prove convergence in a stronger topology. Namely, let us equip $\mathcal{M}_S(\mathcal{X})$, the vector space of signed finite Borel-measures on $\mathcal{X}$, with the following Kantorovich-Rubinstein norm:

$$\|\mu_t\|_0 \equiv \sup\left\{\int_{\mathcal{X}} f d\mu_t : f \in \mathrm{Lip}_1(\mathcal{X}) \text{ with } \sup_{x\in\mathcal{X}}|f(x)| \leq 1\right\}, \tag{II.4.4}$$

where $\mathrm{Lip}_1(\mathcal{X})$ is the space of Lipschitz continuous functions from $\mathcal{X}$ to $\mathbb{R}$ with Lipschitz norm one (cf. [14] p. 191). Then, for all $\delta > 0$, we will prove that

$$\lim_{K\to\infty} \mathbb{P}\left[\sup_{0\leq t\leq T} \|\nu_{t/(Ku_K\sigma_K^2)}^K - \overline{z}(x_t)\delta_{x_t}\|_0 > \delta\right] = 0. \tag{II.4.5}$$

By Proposition II.9.1 this implies convergence in probability with respect to the Skorokhod topology.

(iii) The main result of the paper actually holds under weaker assumptions. More precisely, Assumption 4 can be replaced by

ASSUMPTION 4'. The initial state ν_0^K has a.s. (deterministic) support $\{x_0\}$ with $x_0 \in \mathcal{X}$ satisfying $\partial_1 f(x_0, x_0) \neq 0$.

The reason is that, since $x \mapsto \partial_1 f(x, x)$ is continuous, the Assumption 4 is satisfied locally and since $x \mapsto \partial_1 f(x, x)$ is Lipschitz-continuous, the CEAD never reaches an evolutionary singularity (i.e., a value $y \in \mathcal{X}$ such that $\partial_1 f(y, y) = 0$) in finite time. In particular, for a fixed $T > 0$, the CEAD only visits traits in some interval I of $\mathcal{X}$ where $\partial_1 f(x, x) \neq 0$. By modifying the parameters of the model out of I in such a way that $\partial_1 f(x, x) \neq 0$ everywhere in $\mathcal{X}$, we can apply Thm. II.4.1 to this modified process $\tilde{\nu}$ and deduce that $\tilde{\nu}_{t/Ku_K\sigma_K^2}$ has support included in I for $t \in [0, T]$ with high probability, and hence coincides $\nu_{t/Ku_K\sigma_K^2}$ on this time interval.

(iv) The condition $u_K \ll \frac{\sigma_K^{1+\alpha}}{K\ln K}$ allows mutation events during an invasion phase of a mutant trait, see below, but ensures that there is no "successful" mutational event during this phase.

(v) The fluctuations of the resident population are of order $K^{-1/2}$, thus $K^{-1/2+\alpha} \ll \sigma_K$ ensures that the sign of the initial growth rate is not influenced by the fluctuations of the population size. We will see later that if a mutant trait y appears in a monomorphic population with trait x, then its initial growth rate is $b(y) - d(y) - c(y, x)\langle\nu_t^K, \mathbb{1}\rangle = f(y, x) + o(\sigma_K) = (y-x)\partial_1 f(x, x) + o(\sigma_K)$ since $y - x = O(\sigma_K)$.

(vi) $\exp(K^\alpha)$ is the time the resident population stays with high probability in an $O(\epsilon\sigma_K)$-neighborhood of an attractive domain. This is a moderate deviation result. Thus the condition $\exp(-K^\alpha) \ll u_K$ ensures that the resident population is still in this neighborhood when a mutant occurs.

(vii) The time scale is $(Ku_K\sigma_K^2)^{-1}$ since the expected time for a mutation event is $(Ku_K)^{-1}$, the probability that a mutant invades is of order σ_K and one needs $O(\sigma_K^{-1})$ mutant invasions to see an $O(1)$ change of the resident trait value. This is consistent with the combination of Theorems II.3.7 and II.3.8.

(viii) Note that the ϵ that we use in the proof of the theorem and in the main idea below will not depend on K, but it will converge to zero in the end of the proof of Theorem II.4.1. The constant M introduced below will be fixed all the time. It depends only the parameters of the model, but not on K and ϵ.

(ix) The conditions on the initial states $N_0^K K^{-1}$ imply that $\mathbb{E}[\langle \nu_t^K, \mathbb{1}\rangle^p] < \infty$, uniformly in K and t, and therefore, since $p > 1$, the family of random variables $\{\langle \nu_t^K, \mathbb{1}\rangle\}_{K\geq 1, t\geq 0}$ is uniformly integrable (cf. [25] Lem. 1).

II.4.1 The main idea and the structure of the proof of Theorem II.4.1

Under the conditions of the theorem, the evolution of the population will be described as a succession of *mutant invasions*.

We first control a single *invasion step*. Namely, we show that there is a time scale that is long enough for exactly one mutant population to fixate and for the resident trait to die out, but sufficiently short, such that no two successful mutant populations can exist during this time. We say the mutant trait fixates in the population. Note that this does not prevent the appearance of other mutant traits that do not invade. Second, we consider a much longer time scale on which the single invasion steps aggregate and give rise to a *macroscopic evolution* that converges to the CEAD.

Study of a single invasion step: In order to analyze the invasion of a mutant, we divide the time until a mutant trait has fixated in the population into two phases (compare with Figure II.1).

Phase 1 (Section II.6) Here we fix a small $\epsilon > 0$ and prove the existence of a constant, $M < \infty$, independent of ϵ, such that, as long as all mutant densities are smaller than $\epsilon\sigma_K$, the resident density stays in an $M\epsilon\sigma_K$-neighborhood of $\overline{z}(x)$. Note that, because mutations are rare and the population size is large, the monomorphic initial population has time to stabilize in an $M\epsilon\sigma_K$-neighborhood of this equilibrium $\overline{z}(x)$ before the first mutation occurs. (The time of stabilization is of order $\ln(K)\sigma_K^{-1}$ and the time where the first mutant occurs is of order $1/Ku_K$).

This allows us to approximate the density of one mutant trait y_1 by a branching process with birth rate $b(y_1)$ and death rate $d(y_1) - c(y_1, x)\overline{z}(x)$ such that we can compute the probability that the density of the mutant trait y_1 reaches $\epsilon\sigma_K$, which is of order σ_K, as well as the time it takes to reach this level or to die out. Therefore, the process needs $O(\sigma_K^{-1})$ mutation events until there appears a mutant subpopulation which reaches a size $\epsilon\sigma_K$. Such a mutant is called *successful mutant* and its trait will be the next resident trait. (In fact, we can calculate the distribution of the successful mutant trait only on an event with probability $1 - \epsilon$, but we show that on an event of probability $1 - o(\sigma_K)$, this distribution has support in $\{x + \sigma_K h : 1 \leq h \leq A\}$. Therefore, the exact value of the mutant trait is unknown with probability ϵ, but the difference of the possible values is only of order σ_K.)

We prove in this step also that there are never too many different mutants alive at the same time. From all this we deduce that the subpopulation of the successful mutant reaches the density $\epsilon \sigma_K$, before a different successful mutant appears. Note that we cannot use large deviation results on our time scale as used in [30] to prove this step. Instead, we use some standard potential theory and coupling arguments to obtain estimates of moderate deviations needed to prove that a successful mutant will appear before the resident density exists an $M\epsilon \sigma_K$-neighborhood of its equilibrium.

Phase 2   (Section II.7) We prove that if a mutant population with trait y_s reaches the size $\epsilon \sigma_K$, it will increase to an $M\epsilon \sigma_K$-neighborhood of its equilibrium density $\bar{z}(y_s)$. Simultaneously, the density of the resident trait decreases to $\epsilon \sigma_K$ and finally dies out. Since the fitness advantage of the mutant trait is only of order σ_K, the dynamics of the population process and the corresponding deterministic system are very slow. Even if we would start at a macroscopic density ϵ, the deterministic system needs a time of order σ_K^{-1} to reach an ϵ-neighborhood of its equilibrium density.

The law of large numbers, see Theorem II.3.1 or Chap. 11 of [55], allows to control the distance between the stochastic process and its deterministic limit only on finite, K-independent time intervals. In the regime considered in [25] and [30], namely $\sigma > 0$ independent of K, this suffices to control the stochastic process during this transition phase, since the mutant population of trait y_s only needs a finite, K-independent time, to grow from size ϵ to the ϵ-neighborhood of $\bar{z}(y_s)$. In the regime we consider here, this is no longer possible and a new technique is needed. The method we develop to handle this situation can be seen as a rigorous stochastic "Euler-Scheme" and will be explained in detail in Section II.7. Nevertheless, the proof contains an idea which is strongly connected with the properties of the deterministic dynamical system. Namely, the deterministic system of equations for the case $\sigma_K = 0$ has an invariant manifold of fixed points with a vector field independent of σ_K pointing towards this manifold. Turning on a small σ_K, we therefore expect the stochastic system to stay close to this invariant manyfold and to move along it with speed of order σ_K.

With this method we are able to prove that, in fact, the mutant density reaches the $M\epsilon \sigma_K$-neighborhood of $\bar{z}(y_s)$ and the resident trait dies out. Note that it is possible that an unsuccessful mutant is alive at this time. Therefore, we prove that after the resident trait has died out, there is a time when the population consists only of one trait, namely the one that had fixed, before the next successful mutant occurs.

Note that Figure II.1 is only an artist's sketch and not a "real" simulation.

Convergence to the CEAD: (Section II.8) The proof of convergence to the CEAD uses comparison of the measure valued process ν_t^K with two families of control processes, $\mu^{1,K,\epsilon}$ and $\mu^{2,K,\epsilon}$, which will converge to the CEAD as $K \to \infty$ and then $\epsilon \to 0$. To make more precise statements, we need the following order relation $\preccurlyeq$ for random variables. Roughly speaking, $X \preccurlyeq Y$ will mean that Y is larger than X in law.

Notation.   (a) Let X and Y be $\mathbb{R}$-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We write $X \preccurlyeq Y$, if there is a random variable, $\tilde{Y}$ on Ω, such that Y and $\tilde{Y}$ have the same distribution, and that for all $\omega \in \Omega$, $X(\omega) \le \tilde{Y}(\omega)$.

   (b) For $\mu, \nu \in \mathcal{M}(\mathcal{X})$, we write $\nu \preccurlyeq \mu$, if:
   (i) $\langle \nu, \mathbb{1} \rangle \le \langle \mu, \mathbb{1} \rangle$ and
   (ii) $\sup \{ x \in \mathcal{X} : x \in \mathrm{Supp}\,(\nu) \} \le \inf \{ x \in \mathcal{X} : x \in \mathrm{Supp}\,(\mu) \}$

Note that (i) and (ii) imply that, for all monotone increasing functions $f \in \mathrm{Lip}_1(\mathcal{X}, [-1, 1])$
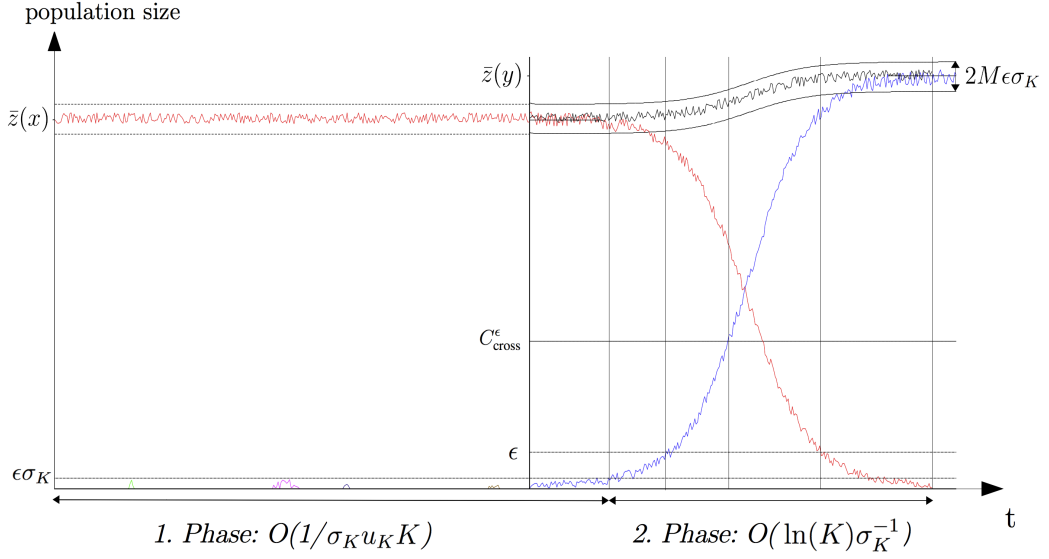
Figure II.1: Typical evolution of the population during a mutant invasion.

and for all $0 \le t \le T$,

$$\int_{\mathcal{X}} f(x) d\nu_t \le \int_{\mathcal{X}} f(x) d\mu_t. \tag{II.4.6}$$

This notion of order between measures is not very informative, except for measures which are close to Dirac masses, where it means that the masses and the supports of the measures are ordered. This is in particular the case for the measures $\mu^{1,K,\epsilon}$ and $\mu^{2,K,\epsilon}$ defined below.

Given $T > 0$, with the results of the two invasion phases, we will define for all $\epsilon > 0$ two measure-valued processes, in $\mathbb{D}([0,\infty), \mathcal{M}(\mathcal{X}))$, such that, for all $\epsilon > 0$,

$$\lim_{K \to \infty} \mathbb{P}\left[\forall\, t \le \frac{T}{K u_K \sigma_K^2} : \quad \mu_t^{1,K,\epsilon} \preccurlyeq \nu_t^K \preccurlyeq \mu_t^{2,K,\epsilon} \right] = 1, \tag{II.4.7}$$

and, for all $\epsilon > 0$ and $i \in \{1, 2\}$,

$$\lim_{K \to \infty} \mathbb{P}\left[\sup_{0 \le t \le T/(K u_K \sigma_K^2)} \left\| \mu_{t/(K u_K \sigma_K^2)}^{i,K,\epsilon} - \bar{z}(x_t)\delta_{x_t} \right\|_0 > \delta(\epsilon) \right] = 0, \tag{II.4.8}$$

for some function δ such that $\delta(\epsilon) \to 0$ when $\epsilon \to 0$. This implies (II.4.5) and therefore the theorem.

The control processes, $\mu^{1,K,\epsilon}$ and $\mu^{2,K,\epsilon}$, are constructed as follows. Let θ_i^K be the random time of the i-th invasion phase, i.e., the first time after θ_{i-1}^K such that a mutant density is larger than $\epsilon \sigma_K$, and let R_i^K be the trait of the i-th successful mutant. Knowing the random variables θ_{i-1}^K and R_{i-1}^K, we are able to approximate θ_i^K and R_i^K: After the $(i\text{-}1)$th invasion phase (of the process ν^K), we define two random times, $\theta_i^{K,1}$ and $\theta_i^{K,2}$, and two random variables $R_i^{K,1}$ and $R_i^{K,2}$ in $\mathcal{X}$, such that

$$\lim_{K \to \infty} \mathbb{P}\left[\forall i \le \sup\left\{ j \in \mathbb{N} : \theta_j^K \le \frac{T}{K u_K \sigma_K^2} \right\} : R_i^{K,1} \preccurlyeq R_i^K \preccurlyeq R_i^{K,2} \text{ and } \theta_i^{K,2} \preccurlyeq \theta_i^K \preccurlyeq \theta_i^{K,1} \right] = 1. \tag{II.4.9}$$

Thus we define $\mu^{1,K}$ and $\mu^{2,K}$ through

$$\mu_t^{1,K} \equiv z_t^1 \delta_{R_i^{K,1}}, \quad \text{for } t \in [\theta_i^{K,1}, \theta_{i+1}^{K,1}), \tag{II.4.10}$$

$$\mu_t^{2,K} \equiv z_t^2 \delta_{R_i^{K,2}}, \quad \text{for } t \in [\theta_i^{K,2}, \theta_{i+1}^{K,2}). \tag{II.4.11}$$

for some appropriate masses z_t^1 and z_t^2. In fact, z_t^1 will be approximately $\bar{z}(R_i^{K,1})$ for $t \in [\theta_i^{K,1}, \theta_{i+1}^{K,1})$, and z_t^2 approximately $\bar{z}(R_i^{K,2})$ for $t \in [\theta_i^{K,2}, \theta_{i+1}^{K,2})$. We will prove that the times $\theta_i^{K,1}$ and $\theta_i^{K,2}$ are (approximately) exponentially distributed with parameters of order $K u_K \sigma_K$, and that the difference of $R_i^K - R_{i-1}^K$ is of order σ_K. The processes $\mu^{1,K,\epsilon}$ and $\mu^{2,K,\epsilon}$ will be constructed by slightly modifying the two processes $\mu^{1,K}$ and $\mu^{K,2}$ in order to make them Markovian. This will imply by standard arguments from [55] that the processes $\mu_{t/K u_K \sigma_K^2}^{1,K}$ and $\mu_{t/K u_K \sigma_K^2}^{2,K}$ converge to $\bar{z}(x_t)\delta_{x_t}$ when $\sigma_K \to 0$, where x_t is the solution of the canonical equation of adaptive dynamics.

We have now prepared the setting to be able to perform the steps of the proof of Theorem II.4.1 as indicated in the outline given in Section II.4.1 in Sections II.6, II.7, and II.8. Before this, we need a few more notations and preparatory results that we collect in Section II.5. Four technical propositions are delegated to an appendix.

All the remaining sections are devoted to the proof of the Theorem II.4.1.

II.5   An augmented process and some elementary properties

In the proof of Theorem II.4.1 we need to construct an augmented process $(\tilde{\nu}^K, L^K)$ that keeps track of part of the history of the population, namely L_t^K is the number of mutations that occurred before t. We first describe this process, then define it by a stochastic equation from which one finds that it is a Markov process with an explicitly given generator.

Let $\mathcal{M}_F^K(\mathbb{N}_0 \times \mathcal{X}) \equiv \left\{ \frac{1}{K} \sum_{i=1}^n \delta_{\xi(i)} : n \geq 0, \xi(1), \ldots, \xi(n) \in \mathbb{N}_0 \times \mathcal{X} \right\}$ denote the set of finite non-negative point measures on $\mathbb{N}_0 \times \mathcal{X}$ rescaled by K. We write $\xi(i) = (\xi_1(i), \xi_2(i))$, where $\xi_1(i) \in \mathbb{N}_0$ and $\xi_2(i) \in \mathcal{X}$. The augmented process, $(\tilde{\nu}^K, L^K)$, is a continuous time stochastic process with state space $\mathcal{M}_F^K(\mathbb{N}_0 \times \mathcal{X}) \times \mathbb{N}_0$. The label k of an individual with trait (k, x) denotes that there were $k - 1$ mutational events in the population before the trait (k, x) appeared for the first time in the population. As in [59], we give a pathwise description of $(\tilde{\nu}^K, L^K)$.

Notation. Let $\mu^K = \frac{1}{K} \sum_{i=1}^n \delta_{\xi(i)} \in \mathcal{M}_F^K(\mathbb{N}_0 \times \mathcal{X})$ and

$$\mathfrak{M}^k(\mu^K) \equiv K \int_{\mathbb{N}_0 \times \mathcal{X}} \mathbb{1}_{\{\xi_1 = k\}} \mu^K(d\xi) \tag{II.5.1}$$

be the number of individuals holding a mutation of label k. Then we rewrite μ^K as follows,

$$\mu^K = \frac{1}{K} \sum_{k=1}^{\infty} \sum_{j=1}^{\mathfrak{M}^k(\mu^K)} \delta_{(k, x_j^k)}, \quad \text{where} \quad \sum_{k=1}^{\infty} \mathfrak{M}^k(\mu^K) = n. \tag{II.5.2}$$

In fact, the $x_1^k, \ldots, x_{\mathfrak{M}^k(\mu^K)}^k$ will be equal in our situation, because the only variation in the trait value is driven by mutational events. We need to define three functions. First, $H : \mathcal{M}_F^K(\mathbb{N}_0 \times \mathcal{X}) \mapsto (\mathbb{N}_0 \times \mathcal{X})^{\mathbb{N}_0^2}$ is defined as

$$H(\mu^K) \equiv \begin{pmatrix} (0, x_1^0) & (0, x_2^0) & \ldots & (0, x_{\mathfrak{M}^0(\mu)}^0) & (0,0) & (0,0) & \ldots \\ (1, x_1^1) & (1, x_2^1) & \ldots & \ldots & (1, x_{\mathfrak{M}^1(\mu)}^1) & (1,0) & \ldots \\ (2, x_1^2) & (2, x_2^2) & \ldots & (2, x_{\mathfrak{M}^2(\mu)}^2) & (2,0) & (2,0) & \ldots \\ \vdots & \ldots & \ldots & \ldots & \ldots & \ldots & \ddots \end{pmatrix}, \tag{II.5.3}$$

Second, $h : \mathcal{M}_F^K(\mathbb{N} \times \mathcal{X}) \mapsto (\mathcal{X})^{\mathbb{N}_0^2}$ us given in terms of H by

$$h_{ij}(\mu^K) \equiv \text{the second component of } H_{ij}(\mu^K), \tag{II.5.4}$$

i.e., if $H_{ij}(\mu^K) = (i, x)$, then $h_{ij} = x$. Third, $\tilde{H} : \mathcal{M}_F^K(\mathbb{N} \times \mathcal{X}) \mapsto \mathcal{X}^{\mathbb{N}_0}$ is defined as follows: if $\mu = \frac{1}{K} \sum_{i=1}^n \delta_{\xi(i)}$, then

$$\tilde{H}(\mu) \equiv \big(\xi_2(\sigma(1)), \xi_2(\sigma(2)), \ldots, \xi_2(\sigma(n)), 0, \ldots\big), \tag{II.5.5}$$

where $\xi_2(\sigma(1)) \leq \ldots \leq \xi_2(\sigma(n))$.

Definition II.5.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an abstract probability space. On this space, we define the following independent random elements:

(i) an $\mathcal{X}$-valued random variable X_0 (the random initial trait),

(ii) a sequence of independent Poisson point measures, $(N_k^{\mathrm{death}}(ds, di, d\theta))_{k \geq 0}$, on $\mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) dz$,

(iii) a sequence of independent Poisson point measures, $(N_k^{\mathrm{birth}}(ds, di, d\theta))_{k \geq 0}$, on $\mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) dz$,

(vi) a Poisson point measures, $N^{\mathrm{mutation}}(ds, di, d\theta, dh)$, on $\mathbb{R}_+ \times \mathbb{N} \times \mathbb{R}_+ \times \{-A, ..., A\}$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) dz \sum_{j=-A}^A \delta_j(dh)$.

Moreover, we define the *augmented process* $(\tilde{\nu}^K, L^K)$ by setting $L_0^K \equiv 0$, $\tilde{\nu}_0^K \equiv \frac{1}{K} N_0^K \delta_{X_0}$, and, for $t > 0$,

$$(\tilde{\nu}_t^K, L_t^K) = (\tilde{\nu}_0^K, L_0^K) \tag{II.5.6}$$

$$+ \sum_{k \geq 0} \Bigg(\int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{i \leq \mathfrak{M}^k(\tilde{\nu}_{s-}^K),\, \theta \leq b\left(h_{k,i}(\tilde{\nu}_{s-}^K)\right)\left(1 - u_K m\left(h_{k,i}(\tilde{\nu}_{s-}^K)\right)\right)\right\}}$$

$$\times \left(\tfrac{1}{K} \delta_{H_{k,i}(\tilde{\nu}_{s-}^K)}, 0\right) N_k^{\mathrm{birth}}(ds, di, d\theta)$$

$$- \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{i \leq \mathfrak{M}^k(\tilde{\nu}_{s-}^K),\, \theta \leq d\left(h_{k,i}(\tilde{\nu}_{s-}^K)\right) + \int_{\mathbb{N}_0 \times \mathcal{X}} c\left(h_{k,i}(\tilde{\nu}_{s-}^K), \xi_2\right)\tilde{\nu}_{s-}^K(d\xi)\right\}}$$

$$\times \left(\tfrac{1}{K} \delta_{H_{k,i}(\tilde{\nu}_{s-}^K)}, 0\right) N_k^{\mathrm{death}}(ds, di, d\theta) \Bigg)$$

$$+ \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{\{-A,...,A\}} \mathbb{1}_{\left\{i \leq K\langle \tilde{\nu}_{s-}^K, \mathbb{1}\rangle,\, \theta \leq b\left(\tilde{H}_i(\tilde{\nu}_{s-}^K)\right) u_K m\left(\tilde{H}_i(\tilde{\nu}_{s-}^K)\right) M(\tilde{H}_i(\tilde{\nu}_{s-}^K), h)\right\}}$$

$$\times \left(\tfrac{1}{K} \delta_{\left(L(s^-)+1,\, \tilde{H}_{k,i}(\tilde{\nu}_{s-}^K) + \sigma h\right)}, 1\right) N^{\mathrm{mutation}}(ds, di, d\theta, dh).$$

Note that the process $(\tilde{\nu}_t^K, L_t^K)_{t \geq 0}$ is a Markov process with generator

$$\tilde{\mathscr{L}}^K f((\tilde{\nu}, L)) \tag{II.5.7}$$

$$= \sum_{k \geq 0} \Bigg(\int_{\mathcal{X}} \left(f\left(\tilde{\nu} + \tfrac{\delta_{(k,x)}}{K}, L\right) - f(\tilde{\nu}, L) \right) \left(1 - u_K m(x)\right) b(x) \, K\tilde{\nu}((k, dx))$$

$$+ \int_{\mathcal{X}} \left(f\left(\tilde{\nu} - \tfrac{\delta_{(k,x)}}{K}, L\right) - f(\tilde{\nu}, L) \right) \left(d(x) + \int_{\mathbb{N}_0 \times \mathcal{X}} c(x, \xi_2) \tilde{\nu}(d\xi) \right) K\tilde{\nu}((k, dx)) \Bigg)$$

$$+ \int_{\mathbb{N}_0 \times \mathcal{X}} \int_{\mathbb{Z}} \left(f\left(\tilde{\nu} + \tfrac{\delta_{(L+1, x+\sigma_K h)}}{K}, L+1\right) - f(\tilde{\nu}, L) \right) u_K m(x) b(x) \, M(x, dh) \, K\tilde{\nu}(d(k, x)).$$

Naturally, the process generated by $\mathscr{L}^K$ defined in Section II.2 is a projection of the process with generator $\tilde{\mathscr{L}}^K$.

The first elementary property we give is that we there exists a rough upper bound for the total mass of the population.

Lemma II.5.2. *Under the same assumptions as in Theorem II.4.1, there exists a constant, $V > 0$, such that*

$$\lim_{K \to \infty} \mathbb{P}\Big[\inf\{t \geq 0 : \langle \tilde{\nu}_t^K, \mathbb{1} \rangle \geq 4\bar{b}/\underline{c}\} < \exp(VK)\Big] = 0. \tag{II.5.8}$$

Proof. Apply Theorem 2 (a) and then Theorem 3 (c) of [25].   □

II.6   The first phase of an invasion

Our first task is to control the trait value (other than the resident trait) where the population first attains a density $\epsilon \sigma_K$, as well as the time when this happens. Since we need to do this for $O(\sigma_K^{-1})$ steps, we need to control this with probability at least $1 - o(\sigma_K)$. Before stating the main result of this section as Theorem II.6.2 below, we need to introduce some notation. We want to analyze such a step from a monomorphic initial condition that satisfies the following assumption that is stronger than what is assumed in Theorem II.4.1.

Assumption 5. Fix $\epsilon > 0$. Let $(R^K)_{K \geq 0}$ be a sequence random variables with values in $\mathcal{X}$. Then, there exists a constant $\tilde{M} > 0$ (independent of ϵ and K) such that for all K large enough

$$L_0^K = 0 \quad \text{and} \quad \tilde{\nu}_0^K = N_{R^K}^K K^{-1} \delta_{(0,R^K)} \tag{II.6.1}$$

where $N_{R^K}^K \in \mathbb{N}$ is a sequence of random variable with $\big|\bar{z}(R^K) - N_{R^K}^K K^{-1}\big| < \tilde{M}\epsilon\sigma_K$ a.s.. We call R^K the resident trait.

The following proposition asserts that if we start with an initial condition as in Theorem II.4.1, after a short time the state of the population satisfies the stronger conditions of Assumption 5.

Proposition II.6.1. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.4.1 hold. Then, there exists a constant $\tilde{M} > 0$ (independent from ϵ and K), such that*

$$\lim_{K \to \infty} \mathbb{P}\Big[\inf\big\{t \geq 0 : |\langle \tilde{\nu}_t^K, \mathbb{1} \rangle - \bar{z}(x)| < \tilde{M}\epsilon\sigma_K\big\} < \ln(K)\sigma_K^{-1} \wedge \inf\big\{t \geq 0 : L(t) \geq 1\big\}\Big] = 1. \tag{II.6.2}$$

Since we can assume for the moment that Assumption 5 hold, we do not state the proof here. In fact, it can be proven in similar way as Lemma II.7.4 (a). We begin with several notations, which we use in the lemmata below.

Notation. Fix $\epsilon > 0$. Suppose that Assumption 2, 4, and 5 hold. Let τ_k^K be the k-th mutant time and let $Y_k^K \in \mathcal{X}$ be the trait of the k-th mutant, i.e.,

$$\tau_k^K \equiv \inf\{t \geq 0 : L_t^K = k\} \quad \text{and} \quad Y_k^K \equiv h_{k,1}(\tilde{\nu}_{\tau_k^K}^K). \tag{II.6.3}$$

We denote by $\theta_{\text{invasion}}^K$ the first time such that a mutant density is larger than $\epsilon \sigma_K$, i.e.,

$$\theta_{\text{invasion}}^K \equiv \inf\big\{t \geq 0 : \exists k \in \{1, \ldots, L_t^K\} \text{ such that } \mathfrak{M}^k(\tilde{\nu}_t^K) > \epsilon\sigma_K K\big\}, \tag{II.6.4}$$

and let R_1^K be the trait value of the mutant which is larger than $\epsilon \sigma_K K$ at time $\theta_{\text{invasion}}^K$, i.e.,

$$R_1^K \equiv h_{k_1,1}(\tilde{\nu}_{\theta_{\text{invasion}}^K}^K) \quad \text{with } k_1 = \inf\big\{k \geq 1 : \mathfrak{M}^k(\tilde{\nu}_{\theta_{\text{invasion}}^K}^K) > \epsilon\sigma_K K\big\}. \tag{II.6.5}$$

Note that k_1 is the label of the first surviving mutant, i.e., $k_1 - 1$ mutations happened before the first surviving mutant appeared. Furthermore, let $\theta_{\text{diversity}}^K$ be the first time such that $\lceil 3/\alpha \rceil$ different traits are present in the population, i.e.,

$$\theta_{\text{diversity}}^K \equiv \inf\left\{ t \geq 0 : \sum_{k=0}^{L^K(t)} \mathbb{1}_{\{\mathfrak{M}^k(\tilde{\nu}_t^K) \geq 1\}} = \lceil 3/\alpha \rceil \right\}, \tag{II.6.6}$$

and similarly let $\theta_{\text{mut. of mut.}}^K$ the first time such that a "2nd generation mutant" occurs, i.e., a mutant which was born from a mutant that in turn was born from the resident trait R^K. Note that

$$\theta_{\text{mut. of mut.}}^K \leq \inf\left\{ t \geq 0 : \exists k \in \{1, \ldots, L_t^K\} \text{ such that } |R^K - Y_k^K| > A\sigma_K \right\}. \tag{II.6.7}$$

Then, we define

$$\hat{\theta}^K \equiv \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K \wedge \theta_{\text{mut. of mut.}}^K \wedge \exp(K^\alpha). \tag{II.6.8}$$

The following theorem collects the main results of this section.

Theorem II.6.2. *Fix $\epsilon > 0$. Under the Assumptions 2, 4, and 5, there exists a constant $M > 0$ (independent of ϵ and K) such that for all K large enough*

(i) *$\tilde{\nu}_0^K = N_{R^K}^K K^{-1}\delta_{(0,R^K)}$, where $\left| \bar{z}(R^K) - N_{R^K}^K K^{-1} \right| < (M/3)\epsilon\sigma_K$ a.s..*

(ii) *We can construct on $(\Omega, \mathcal{F}, \mathbb{P})$ two random variables, $R_1^{K,1}$ and $R_1^{K,2}$, such that*

$$\mathbb{P}\left[R_1^{K,1} \leq R_1^K \leq R_1^{K,2} \text{ and } R_1^{K,2} - R_1^{K,1} \leq A\sigma_K \right] = 1 - o(\sigma_K), \quad \text{and} \tag{II.6.9}$$

$$\mathbb{P}\left[R_1^{K,1} = R_1^K = R_1^{K,2} \right] = 1 - O(\epsilon). \tag{II.6.10}$$

The distributions of $R_1^{K,1}$ and $R_1^{K,2}$ are given in Corollary II.6.10.

(iii) *We can construct on $(\Omega, \mathcal{F}, \mathbb{P})$ two exponential random variables, $E^{K,1}$ and $E^{K,2}$, with parameters of order $\sigma_K u_K K$, such that*

$$\mathbb{P}\left[E^{K,2} \leq \theta_{invasion}^K \leq E^{K,1} + \ln(K)\sigma_K^{-1-\alpha/2} \right] = 1 - o(\sigma_K). \tag{II.6.11}$$

The distributions of $E^{K,1}$ and $E^{K,2}$ are given in Lemma II.6.7.

Moreover, until the first time of invasion, $\theta_{invasion}^K$, the resident density stays in an $\epsilon M\sigma_K$-neighborhood of $\bar{z}(R^K)$, the number of different living mutant traits is bounded by $\lceil \alpha/3 \rceil$, and there is no mutant of a mutant, with probability $1 - o(\sigma_K)$. i.e.,

$$\mathbb{P}\left[\theta_{invasion}^K < \inf\left\{ t \geq 0 : \left| \mathfrak{M}^0(\tilde{\nu}_t^K) - \lceil K\bar{z}(R^K) \rceil \right| > \epsilon M\sigma_K K \right\} \wedge \theta_{diversity}^K \wedge \theta_{mut. of mut.}^K \right] \tag{II.6.12}$$

$$= 1 - o(\sigma_K).$$

Remark 4. The constant $M > 0$ depends only on α and on the functions $b(.), d(.), c(.,.)$, and $m(.)$, but not on K, R^K, and ϵ.

II.6.1   Exit time from an attractive domain

Lemma II.6.3. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold. Then, there exists a constant $M > 0$ (independent of ϵ and K) such that*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\inf\Big\{t \geq 0 : \big|\mathfrak{M}^0(\tilde{\nu}_t^K) - \big\lceil K\overline{z}(R^K)\big\rceil\big| > \epsilon M \sigma_K K\Big\} < \hat{\theta}^K\Big] = 0. \tag{II.6.13}$$

The statement is stronger than the corresponding one in [25], Thm. 3(c), since the diameter of the domain converges to zero, when K tends to infinity and since it gives control of the speed of convergence to 0 of the probabilities. Therefore, it does not follow from the classical results about the time of exit from an attractive domain (cf. [61]). Our proof is based on a coupling with a discrete Markov chain and some standard potential theoretical argument.

Proof. Define

$$X_t \equiv \big|\mathfrak{M}_t^0 - \big\lceil K\overline{z}(R^K)\big\rceil\big| \tag{II.6.14}$$

and, for all $M \geq 0$,

$$\tau_0 \equiv \inf\{t > 0 : X_t = 0\} \quad \text{and} \quad \tau_{M\epsilon\sigma_K K} \equiv \inf\{t > 0 : X_t \geq M\epsilon\sigma_K K\}. \tag{II.6.15}$$

Note that τ_0 and $\tau_{M\epsilon\sigma_K K}$ are stopping times with respect to the natural filtration of X_t, which is equal to $\sigma\big(\mathfrak{M}_s^0 ; s \leq t\big)$, and that the process $(\mathfrak{M}_t^0)_{t \geq 0}$ is not Markovian. We can associate with the continuous time process X_t a discrete time (non-Markovian) process Y_n, which records the sequence of values that X_t takes. (This can be formally defined by introducing the sequences T_k of the stopping times which record the instances when $X_t \neq X_{t-}$ and setting $Y_n = X_{T_n}$.) Now, we can compute

$$\mathbb{P}\Big[\tau_{M\epsilon\sigma_K K} < \tau_0 \wedge \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K \wedge \theta_{\text{mut. of mut.}}^K\Big] \tag{II.6.16}$$

with respect to the stopping times defined for the discrete time process Y_n and exploit the natural renewal structure on Y_n. Therefore, we prove the following claim.

CLAIM.   *For $1 \leq i \ll K$, and K large enough,*

$$\mathbb{P}\Big[Y_{n+1} = i + 1 \big| Y_n = i, T_{n+1} < \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K \wedge \theta_{\text{mut. of mut.}}^K\Big] \tag{II.6.17}$$
$$\leq \frac{1}{2} - (\underline{c}/4\overline{b})K^{-1}i + (\overline{c}/\underline{b})\epsilon\sigma_K \equiv p_+^K(i),$$

where $\underline{c}, \underline{b}, \overline{c},$ and $\overline{b}$ are the lower, respectively upper bounds for birth and competition rates. Recall from Remark 1 that the equilibrium $\overline{z}(R^K)$ is equal to $\frac{b(R^K) - d(R^K)}{c(R^K, R^K)}$ and observe that there are at most $\lceil 3/\alpha\rceil\epsilon\sigma_K K$ mutant individuals alive at any time $t < \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K \wedge \theta_{\text{mut. of mut.}}^K$. Therefore, for $1 \leq i \ll K$ and K large enough,

$$\mathbb{P}\Big[Y_{n+1} = i + 1 \big| Y_n = i, T_{n+1} < \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K \wedge \theta_{\text{mut. of mut.}}^K\Big] \tag{II.6.18}$$
$$\leq \frac{(1 - m(R^K)u_K)b(R^K)}{(1 - m(R^K)u_K)b(R^K) + d(R^K) + c(R^K, R^K)K^{-1}(\lceil K\overline{z}(R^K)\rceil + i)}$$
$$\vee \frac{d(R^K) + c(R^K, R^K)K^{-1}(\lceil K\overline{z}(R^K)\rceil - i) + \overline{c}\lceil 3/\alpha\rceil\epsilon\sigma_K K}{(1 - m(R^K)u_K)b(R^K) + d(R^K) + c(R^K, R^K)K^{-1}(\lceil K\overline{z}(R^K)\rceil - i)}$$
$$\leq \frac{b(R^K) - m(R^K)u_K b(R^K)}{2b(R^K) - m(R^K)u_K b(R^K) + c(R^K, R^K)K^{-1}i} \vee \frac{b(R^K) - c(R^K, R^K)K^{-1}(i-1) + \overline{c}\lceil 3/\alpha\rceil\epsilon\sigma_K K}{2b(R^K) - m(R^K)u_K b(R^K) - c(R^K, R^K)K^{-1}i}$$
$$\leq \frac{1}{2} - (\underline{c}/4\overline{b})K^{-1}i + (\overline{c}/\underline{b})\lceil 3/\alpha\rceil\epsilon\sigma_K.$$

This proves the claim. Next we introduce a coupling, i.e., we define a discrete time process Z_n with the following properties

(i) $Z_0 = Y_0$,

(ii) $\mathbb{P}\big[Z_{n+1} = i+1, Y_{n+1} = i+1 \big| Y_n = Z_n = i, T_{n+1} < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}\big]$
$$= \mathbb{P}\big[Y_{n+1} = i+1 \big| Y_n = i, T_{n+1} < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}\big],$$

(iii) $\mathbb{P}\big[Z_{n+1} = i+1, Y_{n+1} = i-1 \big| Y_n = Z_n = i, T_{n+1} < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}\big]$
$$= p^K_+(i) - \mathbb{P}\big[Y_{n+1} = i+1 \big| Y_n = i, T_{n+1} < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}\big],$$

(iv) $\mathbb{P}\big[Z_{n+1} = i+1 \big| Y_n < Z_n = i, T_{n+1} < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}\big] = p^K_+(i)$,

(v) $\mathbb{P}\big[Z_{n+1} = i-1 \big| Y_n < Z_n = i, T_{n+1} < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}\big] = 1 - p^K_+(i)$.

Note that by construction $Z_n \geq Y_n$ a.s. for all n such that $T_n < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}$ and the marginal distribution of Z_n is a Markov chain with transition probabilities

$$\mathbb{P}\big[Z_{n+1} = j | Z_n = i\big] = \begin{cases} 1 & \text{for } i = 0 \text{ and } j = 1 \\ p^K_+(i) & \text{for } i \geq 1 \text{ and } j = i+1 \\ 1 - p^K_+(i) & \text{for } i \geq 1 \text{ and } j = i-1 \\ 0 & \text{else.} \end{cases} \tag{II.6.19}$$

Now we define a continuous time process, $\tilde{Z}$, associated to Z_n. To do this, let $(\tilde{T}_j)_{j\in\mathbb{N}}$ be the sequence of jump times of $\tilde{Z}$, i.e., $\tilde{Z}_t \equiv Z_n$ if $t \in [\tilde{T}_n, \tilde{T}_{n+1})$, defined for all $j \in \mathbb{N}$ as follows

$$\tilde{T}_j - \tilde{T}_{j-1} = \begin{cases} T_j - T_{j-1} & \text{if } T_j < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}} \\ W_j & \text{else,} \end{cases} \tag{II.6.20}$$

where W_j are independent exponential random variables with mean $(C_{\text{total rate}}K)^{-1}$, where $C_{\text{total rate}} = 4\bar{b}\underline{c}(\bar{b} + \bar{d} + \bar{c}(4\bar{b}\underline{c}))$. By Lemma II.5.2, $C_{\text{total rate}}K$ is an upper bound for the total event rate of $\langle \tilde{\nu}^K_t, \mathbb{1} \rangle$, and therefore also for $\mathfrak{M}^0_t$.

Define $\tau^Z_{M\epsilon\sigma_K K} \equiv \inf\{n \geq 0 : Z_n \geq M\epsilon\sigma_K K\}$ and $\tau^Z_0 \equiv \inf\{n \geq 0 : Z_n = 0\}$. Then, since $\tilde{Z}_t \geq X_t$ a.s. for all $t < \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}$,

$$\mathbb{P}\big[\tau_{M\epsilon\sigma_K K} < \tau_0 \wedge \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}} \wedge \theta^K_{\text{mut. of mut.}}\big] \leq \mathbb{P}\big[\tau^Z_{M\epsilon\sigma_K K} < \tau^Z_0\big]. \tag{II.6.21}$$

Applying Proposition II.9.2 yields that, for all $M \geq 32\lceil 3/\alpha \rceil (\bar{c}\bar{b})/(\underline{b}\underline{c})$ such that $Z_0 \leq \frac{1}{3}M\epsilon\sigma_K K$ and large K large enough,

$$\mathbb{P}\big[\tau^Z_{M\epsilon\sigma_K K} < \tau^Z_0\big] \leq \exp\left(-K^{2\alpha}\right). \tag{II.6.22}$$

Next we prove that the process X_t returns many times to zero before it reaches for the first time the value $M\epsilon\sigma_K K$. More precisely, we first prove a lower bound on the number of returns to zero of the discrete time process Z_n. Then we calculate the time for a return to zero. From now on we assume that $M \geq 32\lceil 3/\alpha \rceil (\bar{c}\bar{b})/(\underline{b}\underline{c})$. We define the following stopping times with respect to the natural filtration of Z which records the number of jumps the process Z needs for m zero-returns:

$$\tau^Z_{m \text{ returns}} \equiv \inf\left\{n \geq 1 : \sum_{i=1}^n \mathbb{1}_{Z_i = 0} = m\right\}. \tag{II.6.23}$$

Let $Q^m \equiv \mathbb{P}\big[\tau^Z_{m \text{ returns}} < \tau^Z_{M\epsilon\sigma_K K} < \tau^Z_{(m+1) \text{ returns}}\big]$ be the probability that the Markov chain Z_n returns exactly m times to zero before it reaches the value $M\epsilon\sigma_K K$. We have

$$Q^0 = \mathbb{P}\big[\tau^Z_{M\epsilon\sigma_K K} < \tau^Z_0\big] \leq \exp\left(-K^{2\alpha}\right), \tag{II.6.24}$$

and, due to the Markov property, for $m \geq 1$,

$$Q^m = \mathbb{P}\left[\tau_0^Z < \tau_{M\epsilon\sigma_K K}^Z\right]\left(1 - \mathbb{P}_1\left[\tau_{M\epsilon\sigma_K K}^Z < \tau_0^Z\right]\right)^{m-1}\mathbb{P}_1\left[\tau_{M\epsilon\sigma_K K}^Z < \tau_0^Z\right], \tag{II.6.25}$$

where the last term in the product is smaller than $\exp\left(-K^{2\alpha}\right)$. Thus,

$$Q^m \leq \exp\left(-K^{2\alpha}\right) \quad \text{for all } m \geq 0. \tag{II.6.26}$$

Let B be the random variable which records the number of zero returns of Z_n before Z_n reaches $M\epsilon\sigma_K K$. In other words, $B = n$ if and only if $\tau_{n\text{ returns}}^Z < \tau_{M\epsilon\sigma_K K}^Z < \tau_{n+1\text{ returns}}^Z$, and we obtain that

$$\mathbb{P}\left[B \leq n\right] = \sum_{i=0}^n Q^i \leq (n+1)\exp\left(-K^{2\alpha}\right). \tag{II.6.27}$$

Set $I_1 \equiv \tilde{T}_{\tau_1^Z\text{ return}}$ and $I_j \equiv \tilde{T}_{\tau_j^Z\text{ returns}} - \tilde{T}_{\tau_{(j-1)\text{ returns}}^Z}$ for $j \geq 2$. For any j, I_j is the random time between the $(j-1)$th and the j-th zero return of the associated continuous time process $\tilde{Z}_t$ and

$$\sum_{i=1}^B I_i \leq \inf\{t \geq 0 : \tilde{Z}_t \geq M\epsilon\sigma_K K\} \leq \sum_{i=1}^{B+1} I_i. \tag{II.6.28}$$

We get an upper bound for the probability which we want to compute

$$\mathbb{P}\left[\inf\left\{t \geq 0 : |\mathfrak{M}^0(\tilde{\nu}_t^K) - \lceil K\overline{z}(R^K)\rceil| > \epsilon M\sigma_K K\right\} < \hat{\theta}^K\right] \tag{II.6.29}$$

$$\leq \sum_{l=n}^\infty \mathbb{P}\left[\inf\{t \geq 0 : \tilde{Z}_t \geq M\epsilon\sigma_K K\} < \exp(K^\alpha), \, B = l\right] + \mathbb{P}\left[B \leq n\right].$$

According to (II.6.28), if $B = l$ and if in addition more than $l/2$ of the l random times I_j in the sum are larger than $2l^{-1}\exp(K^\alpha)$, then $\inf\{t \geq 0 : \tilde{Z}_t \geq M_x\epsilon\sigma_K K\}$ is larger than $\exp(K^\alpha)$. Therefore, for all $l \geq n$,

$$\mathbb{P}\left[\inf\{t \geq 0 : \tilde{Z}_t \geq M\epsilon\sigma_K K\} < \exp(K^\alpha), \, B = l\right] \tag{II.6.30}$$

$$\leq \mathbb{P}\left[\sum_{i=1}^l \mathbb{1}_{\{I_j < 2l^{-1}\exp(K^\alpha)\}} > {}^l\!/_2, \, B = l\right].$$

As mentioned before, $C_{\text{total rate}}K$ is an upper bound for the total event rate of $\langle\tilde{\nu}_t^K, \mathbb{1}\rangle$. Thus we can bound the jump times by a sequence of independent, exponential random variables $(V_j)_{j\in\mathbb{N}}$ with mean $(C_{\text{total rate}}K)^{-1}$. Namely,

$$\tilde{T}_j - \tilde{T}_{j-1} \equiv T_j - T_{j-1} \succcurlyeq V_j \quad \text{if } T_j \leq \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K \wedge \theta_{\text{mut. of mut.}}^K. \tag{II.6.31}$$

Otherwise the random times $\tilde{T}_j - \tilde{T}_{j-1}$ are by definition independent and exponentially distributed with mean $(C_{\text{total rate}}K)^{-1}$. The process $\tilde{Z}$ has to make at least two jumps to return to zero. Hence,

$$I_i \succcurlyeq \tilde{W}_i, \quad \text{for all } i \in \mathbb{N}, \tag{II.6.32}$$

where $(\tilde{W}_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. exponential random variables with mean $(C_{\text{total rate}}K)^{-1}$. Thus,

$$\mathbb{P}\left[\sum_{i=1}^l \mathbb{1}_{\{I_j < 2l^{-1}\exp(K^\alpha)\}} > \frac{l}{2}, B = l\right] \leq \mathbb{P}\left[\sum_{i=1}^l \mathbb{1}_{\{\tilde{W}_i < 2l^{-1}\exp(K^\alpha)\}} > \frac{l}{2}\right]. \tag{II.6.33}$$

Since $\mathbb{P}[\tilde{W}_i < 2l^{-1}\exp(K^\alpha)] = 1 - \exp(-C_{\text{total rate}}Kl^{-1}\exp(K^\alpha))$ and $(\tilde{W}_i)_{i\geq 1}$ are independent, we obtain that $\sum_{i=1}^{l} \mathbb{1}_{\{\tilde{W}_i<2l^{-1}\exp(K^\alpha)\}}$ is binomially distributed with $n = l$ and $p = 1-\exp(-C_{\text{total rate}}K\, l^{-1}\exp(K^\alpha))$. Therefore, the right hand side of (II.6.33) is equal to

$$\sum_{i=l/2}^{l} \binom{l}{i} \left(1 - \exp\left(-C_{\text{total rate}}Kl^{-1}e^{K^\alpha}\right)\right)^i \left(\exp\left(-C_{\text{total rate}}Kl^{-1}e^{K^\alpha}\right)\right)^{l-i}. \qquad (II.6.34)$$

For the following two computations we use the elementary facts that $\binom{l}{i} < 2^l$ and $l < 2^l$, for all $l \in \mathbb{N}$ and $i \leq l$. We obtain that, for large K enough, the left hand side of (II.6.29) is bounded from above by

$$\sum_{l=n}^{\infty} \sum_{i=l/2}^{l} \binom{l}{i} \left(1 - \exp\left(-C_{\text{total rate}}Kl^{-1}e^{K^\alpha}\right)\right)^i \left(\exp\left(-C_{\text{total rate}}Kl^{-1}e^{K^\alpha}\right)\right)^{l-i} + \mathbb{P}[B \leq n]$$

$$\leq \sum_{l=n}^{\infty} \frac{l}{2} 2^l \left(1 - \exp\left(-C_{\text{total rate}}Kl^{-1}e^{K^\alpha}\right)\right)^{l/2} + \mathbb{P}[B \leq n].$$

By (II.6.27) we see that $\mathbb{P}[B \leq n] = o(\sigma_K)$ if the variable n fulfills the following condition

$$n \ll \exp\left(K^{2\alpha}\right)\sigma_K. \qquad (II.6.35)$$

Therefore, we choose $n = \lceil\exp(2K^\alpha)\rceil$ and get, for large K enough,

$$\mathbb{P}\left[\inf\left\{t \geq 0 : |\mathfrak{M}^0(\tilde{\nu}_t^K) - \lceil K\overline{z}(R^K)\rceil| > \epsilon M\sigma_K K\right\} < \hat{\theta}^K\right] \qquad (II.6.36)$$

$$\leq \sum_{l=\lceil\exp(2K^\alpha)\rceil}^{\infty} 4^l \left(1 - \exp\left(-C_{\text{total rate}}Kl^{-1}e^{K^\alpha}\right)\right)^{l/2} + o(\sigma_K)$$

$$\leq \sum_{l=\lceil\exp(2K^\alpha)\rceil}^{\infty} \left(4\left(1 - \exp\left(-C_{\text{total rate}}Ke^{-K^\alpha}\right)\right)^{1/2}\right)^l + o(\sigma_K)$$

$$\leq 2\left(4^2\left(1 - \exp\left(-C_{\text{total rate}}Ke^{-K^\alpha}\right)\right)\right)^{\frac{1}{2}\lceil\exp(2K^\alpha)\rceil} + o(\sigma_K)$$

$$\leq 2\left(4^2 C_{\text{total rate}}Ke^{-K^\alpha}\right)^{\frac{1}{2}\lceil\exp(2K^\alpha)\rceil} + o(\sigma_K)$$

$$\leq o(Ke^{-K^\alpha}) + o(\sigma_K),$$

where we used that $\exp(-x) \geq 1 - x$ for $x \geq 0$ and $K\exp(K^{-\alpha}) \ll \sigma_K$. $\qquad\square$

II.6.2 Controlling the number L_t^K of mutations by Poisson processes

Lemma II.6.4. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold and let M be the constant of Lemma II.6.3. Then,*

$$\lim_{K\to\infty} \sigma_K^{-1}\left(1 - \mathbb{P}\left[\forall\, 0 \leq t \leq \hat{\theta}^K : A^{1,K}(t) \preccurlyeq L_t^K \preccurlyeq A^{2,K}(t)\right]\right) = 0, \qquad (II.6.37)$$

where $A^{1,K}$ and $A^{2,K}$ are Poisson counting processes with parameter $a_1^K u_K K$ and $a_2^K u_K K$ with

$$a_1^K \equiv \left(\overline{z}(R^K) - \epsilon M\sigma_K\right)b(R^K)m(R^K), \qquad (II.6.38)$$

$$a_2^K \equiv \left(\overline{z}(R^K) + \epsilon\left(M + \lceil 3/\alpha\rceil\right)\sigma_K\right)\left(b(R^K)m(R^K) + C_L^{b,m,M}A\sigma_K\right), \qquad (II.6.39)$$

and $C_L^{b,m,M}$ is a constant depending only on the functions $b(.), m(.)$ and $M(., h)$ for $h \in \{-A, \ldots, A\}$.

Proof. We obtain from the last lemma that

$$\mathbb{P}\left[\forall\, 0\leq t\leq\hat{\theta}^K : \overline{z}(R^K) - \epsilon M\sigma_K \leq \langle\tilde{\nu}_t, \mathbb{1}\rangle \leq \overline{z}(R^K) + \epsilon\left(M + \left\lceil\tfrac{3}{\alpha}\right\rceil\right)\sigma_K\right] = 1 - o(\sigma_K).$$

Therefore, define

$$A^{1,K}(t) = \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{\{-A,\dots,A\}} \mathbb{1}_{\{i\leq K(\overline{z}(R^K)-\epsilon M\sigma_K),\, \theta\leq b(R^K)u_K m(R^K)M(R^K,h)\}} \tag{II.6.40}$$
$$\times\, N^{\text{mutation}}(ds, di, d\theta, dh)$$

and similarly

$$A^{2,K}(t) = \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{\{-A,\dots,A\}} \mathbb{1}_{\{i\leq K(\overline{z}(R^K)+\epsilon(M+\lceil\frac{3}{\alpha}\rceil)\sigma_K)\}} \tag{II.6.41}$$
$$\times\, \mathbb{1}_{\left\{\theta\leq u_K\left(b(R^K)m(R^K)M(R^K,h)+C_L^{b,m,M}A\sigma_K\right)\right\}}$$
$$\times\, N^{\text{mutation}}(ds, di, d\theta, dh),$$

Since $\hat{\theta}^K \leq \theta^K_{\text{mut. of mut.}}$, any mutant trait differs at most $A\sigma_K$ from the resident trait, R^K. Thus, we have that $u_K\left(b(R^K)m(R^K)M(R^K,h) + C_L^{b,m,M}A\sigma_K\right)$ is a rough upper bound for the mutation rate per individual for an appropriate choice of $C_L^{b,m,M}$. Note that $A^{i,K}$ are Poisson counting process with parameter $a_i^K u_K K$. By construction, we obtain (II.6.37). $\quad\square$

II.6.3   Controlling the number $\mathfrak{M}^k(\tilde{\nu}_t)$ of offspring of the k-th mutant by birth-death processes

Lemma II.6.5. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold and let M be the constant in Lemma II.6.3. Then,*

$$\lim_{K\to\infty} \sigma_K^{-1}\left(1 - \mathbb{P}\left[\forall\, 1\leq k\leq L^K_{\hat{\theta}^K},\, \forall t\leq\hat{\theta}^K : Z_k^{K,1}(t) \leqslant \mathfrak{M}^k(\tilde{\nu}_t) \leqslant Z_k^{K,2}(t)\right]\right) = 0, \tag{II.6.42}$$

where $Z_k^{K,1}(t)$ resp. $Z_k^{K,2}(t)$ are $\mathbb{N}_0$-valued processes, which are zero until time τ_k^K, the first time s.t. $\mathfrak{M}^k(\tilde{\nu}_t) \neq 0$, and afterwards linear, continuous time birth-death processes with initial state 1 at time τ_k^K, birth rates per individual

$$b_k^{K,1} = b_k^{K,2} = b\left(Y_k^K\right)\left(1 - u_K m(Y_k^K)\right), \tag{II.6.43}$$

and death rate per individual

$$d_k^{K,1} = d(Y_k^K) + c(Y_k^K, R^K)\left(\overline{z}(R^K) + M\epsilon\sigma_K\right) + \overline{c}\lceil 3/\alpha\rceil\epsilon\sigma_K \tag{II.6.44}$$
$$\text{resp.}\quad d_k^{K,2} = d(Y_k^K) + c(Y_k^K, R^K)\left(\overline{z}(R^K) - M\epsilon\sigma_K\right). \tag{II.6.45}$$

Furthermore, define $\tilde{Z}_k^{K,1}(t) \equiv Z_k^{K,1}(\tau_k + t)$ and $\tilde{Z}_k^{K,2}(t) \equiv Z_k^{K,2}(\tau_k + t)$, then the processes $\{(\tilde{Z}_k^{K,1}, \tilde{Z}_k^{K,2})\}_{k\geq 1}$ are independent and identically distributed.

Proof. For any $t \leq \hat{\theta}^K$, any individual of $\mathfrak{M}^k(\tilde{\nu}_t)$ gives birth to a new individual with the same trait with rate $b\left(Y_k^K\right)\left(1 - u_K m(Y_k^K)\right)$ and dies with rate $d(Y_k^K) + \int_{\mathbb{N}\times\mathcal{X}} c(Y_k^K, \xi_2)\tilde{\nu}_t^K(d\xi)$, which belongs to the following interval

$$\left[d(Y_k^K) + c(Y_k^K, R^K)(\overline{z}(R^K) - M\epsilon\sigma_K),\right. \tag{II.6.46}$$
$$\left. d(Y_k^K) + c(Y_k^K, R^K)(\overline{z}(R^K) + M\epsilon\sigma_K) + \overline{c}\lceil 3/\alpha\rceil\epsilon\sigma_K\right].$$

Thus, let us define, for $k \leq L_{\hat{\theta}^K}$,

$$
\tilde{Z}_k^{K,1}(t) \equiv \int_{\tau_k}^{\tau_k+t} \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{i \leq \tilde{Z}_k^{K,1}(s^-),\, \theta \leq b(Y_k)(1-u_K m(Y_k))\right\}} N_k^{\text{birth}}(ds, di, d\theta) \tag{II.6.47}
$$
$$
- \int_{\tau_k}^{\tau_k+t} \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{i \leq \tilde{Z}_k^{K,1}(s^-),\, \theta \leq d(Y_k^K)+c(Y_k^K, R^K)(\bar{z}(R^K)+M\epsilon\sigma_K)+\bar{c}\lceil 3/\alpha \rceil \epsilon\sigma_K\right\}}
$$
$$
\times N_k^{\text{death}}(ds, di, d\theta)
$$

and similarly

$$
\tilde{Z}_k^{K,2}(t) \equiv \int_{\tau_k}^{\tau_k+t} \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{i \leq \tilde{Z}_k^{K,1}(s^-),\, \theta \leq b(Y_k)(1-u_K m(Y_k))\right\}} N_k^{\text{birth}}(ds, di, d\theta) \tag{II.6.48}
$$
$$
- \int_{\tau_k}^{\tau_k+t} \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{i \leq \tilde{Z}_k^{K,1}(s^-),\, \theta \leq d(Y_k^K)+c(Y_k^K, R^K)(\bar{z}(R^K)-M\epsilon\sigma_K)\right\}} N_k^{\text{death}}(ds, di, d\theta),
$$

and a similar construction for $k > L_{\hat{\theta}^K}$, where the random variables Y_k^K are replaced by i.i.d. ones with distribution $f_K * M(R^K, \cdot)$, independent of all the previously introduced random variables, where f_K is the homothety of ratio σ_K. Note that, the Poisson point measures N_k^{birth} and N_k^{death} are independent of Y_k^K and τ_k, and that the processes $\tilde{Z}_k^{K,1}$ and $\tilde{Z}_k^{K,2}$ only depend on N_k^{birth}, N_k^{death}, Y_k^K and τ_k. By construction, conditionally on $Y_k^K = y$ and $\tau_k = s$, the process $\tilde{Z}_k^{K,1}$ is distributed as a linear birth-death processes with birth rate $b(y)(1-u_K m(y))$ and death rate $d(y) + c(y, R^K)(\bar{z}(R^K) + M\epsilon\sigma_K) + \bar{c}\lceil 3/\alpha \rceil \epsilon\sigma_K$, and similarly for $\tilde{Z}_k^{K,2}$. In particular, the law of $(\tilde{Z}_k^{K,1}, \tilde{Z}_k^{K,2})$ does not depend on τ_k. Therefore, defining $\mathcal{G}_k \equiv \sigma(\tilde{\nu}_t, t \leq \tau_k, Y_k^K, N_\ell^{\text{birth}}, N_\ell^{\text{death}}, 1 \leq \ell \leq k-1)$, for all bounded measurable functions $F_1, \ldots, F_k$ on $\mathbb{D}(\mathbb{R}_+, \mathbb{Z}_+^2)$,

$$
\mathbb{E}\left[F_1(\tilde{Z}_1^{K,1}, \tilde{Z}_1^{K,2}) \ldots F_k(\tilde{Z}_k^{K,1}, \tilde{Z}_k^{K,2})\right] \tag{II.6.49}
$$
$$
= \mathbb{E}\left[F_1(\tilde{Z}_1^{K,1}, \tilde{Z}_1^{K,2}) \ldots F_{k-1}(\tilde{Z}_{k-1}^{K,1}, \tilde{Z}_{k-1}^{K,2}) \mathbb{E}[F_k(\tilde{Z}_k^{K,1}, \tilde{Z}_k^{K,2}) \mid \mathcal{G}_k]\right]
$$
$$
= \mathbb{E}\left[F_1(\tilde{Z}_1^{K,1}, \tilde{Z}_1^{K,2}) \ldots F_{k-1}(\tilde{Z}_{k-1}^{K,1}, \tilde{Z}_{k-1}^{K,2}) \mathbb{E}[F_k(\tilde{Z}_k^{K,1}, \tilde{Z}_k^{K,2}) \mid Y_k^K]\right]
$$
$$
= \mathbb{E}\left[F_1(\tilde{Z}_1^{K,1}, \tilde{Z}_1^{K,2}) \ldots F_{k-1}(\tilde{Z}_{k-1}^{K,1}, \tilde{Z}_{k-1}^{K,2})\right] \mathbb{E}[F_k(\tilde{Z}_k^{K,1}, \tilde{Z}_k^{K,2})],
$$

where the last equality follows from the fact that the random variable Y_k^K is independent of $(\tilde{Z}_\ell^{K,1}, \tilde{Z}_\ell^{K,2})$ for $1 \leq \ell \leq k-1$. Actually, $(Y_k^K)_{1 \leq k \leq L_{\hat{\theta}^K}}$ are i.i.d. random variables, with law $f_K * M(R^K, \cdot)$. This implies by induction that the processes $\{(\tilde{Z}_k^{K,1}, \tilde{Z}_k^{K,2})\}_{k \geq 1}$ are i.i.d.. $\qquad \square$

II.6.4 Controlling survival of the k-th mutant population

Notation. Let us define $B_k^K \equiv \mathbb{1}_{\inf\{t \geq \tau_k \,:\, \mathfrak{M}^k(\tilde{\nu}_t) \geq \epsilon\sigma_K K\} < \inf\{t \geq \tau_K \,:\, \mathfrak{M}^k(\tilde{\nu}_t)=0\}}$.

This random variable indicates whether or not the k-th mutant population, which appeared at time τ_k, invades, i.e., reaches $\epsilon\sigma_K K$ individuals before dying out. The following lemma introduces a sequence of i.i.d. random variables $(B_k^{1,K}, B_k^{2,K})$ which are 2-tuples of Bernoulli random variables constructed from the processes $Z_k^{K,1}(t)$ and $Z_k^{K,2}(t)$ defined in Lemma II.6.5, such that $(B_k^K)_{k \geq 0}$ is stochastically dominated by the sequences $(B_k^{i,K})_{k \geq 0}$.

Lemma II.6.6. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold and let M be the constant of Lemma II.6.3. Then,*

$$
\lim_{K \to \infty} \sigma_K^{-1}\left(1 - \mathbb{P}\left[\forall\, 1 \leq k \leq L_{\hat{\theta}^K} \,:\, B_k^{1,K} \leq B_k^K \leq B_k^{2,K}\right]\right) = 0, \tag{II.6.50}
$$

where $((B_k^{1,K}, B_k^{2,K}))_{k \geq 1}$ is a sequence of i.i.d. 2-tuples of Bernoulli random variables such that $B_k^{1,K} \leq B_k^{2,K}$ a.s. Its distribution is characterized by

$$
\sigma_K q_1^K(h) \;\; \equiv \;\; \mathbb{P}\big[B_k^{1,K} = 1 \mid Y_k^K = R^K + h\sigma_K\big] \tag{II.6.51}
$$

$$
= \begin{cases} \sigma_K\Big(h\frac{\partial_1 f(R^K,R^K)}{b(R^K)} - \epsilon C_{Bernoulli}^1\Big) & \text{if } 1 \leq h \leq A, \\ 0 & \text{otherwise} \end{cases}
$$

and

$$
\sigma_K q_2^K(h) \;\; \equiv \;\; \mathbb{P}\big[B_k^{2,K} = 1 \mid Y_k^K = R^K + h\sigma_K\big] \tag{II.6.52}
$$

$$
= \begin{cases} \sigma_K\Big(h\frac{\partial_1 f(R^K,R^K)}{b(R^K)} + \epsilon C_{Bernoulli}^2\Big) & \text{if } 1 \leq h \leq A, \\ 0 & \text{otherwise,} \end{cases}
$$

where $C_{Bernoulli}^1$ and $C_{Bernoulli}^2$ depend only on α, M, and C_L (the Lipschitz constant of our parameters). Then, for $i = 1, 2$ and $k \geq 1$, $B_k^{i,K}$ is a Bernoulli random variable of parameter $\sigma_K p_i^K$, where

$$
p_i^K \equiv \sum_{h=1}^{A} q_i^K(h) M(R^K, h). \tag{II.6.53}
$$

Remark 5.   (i) For all $k \geq 1$, $\mathbb{P}[B_k^{1,K} = 0 | B_k^{2,K} = 1] = 1 - \frac{p_1^K}{p_2^K}$ and is thereby of order ϵ.

(ii) We use in here the assumption that $\partial_1 f(x,x) > 0$ for all $x \in \mathcal{X}$.

Proof. Let $Z_k^{K,1}(t)$ resp. $Z_k^{K,2}(t)$ as defined in Lemma II.6.5 and define

$$
\tilde{B}_k^{i,K} \equiv \mathbb{1}_{\inf\{t \geq \tau_k : Z_k^{K,i}(t) \geq \epsilon \sigma_K K\} < \inf\{t \geq \tau_k : Z_k^{K,i}(t) = 0\}} \quad \text{for } i = 1, 2. \tag{II.6.54}
$$

Then, due to the last lemma

$$
\mathbb{P}\Big[\forall 1 \leq k \leq L_{\hat{\theta}^K}^K : \tilde{B}_k^{1,K} \leqslant B_k^K \leqslant \tilde{B}_k^{2,K}\Big] = 1 - o(\sigma_K). \tag{II.6.55}
$$

For all $1 \leq k \leq L_{\hat{\theta}^K}^K$, we obtain with Proposition II.9.3, that

$$
\left| \mathbb{P}\Big[\inf\big\{t \geq \tau_k : Z_k^{K,i}(t) \geq \epsilon \sigma_K K\big\} < \inf\big\{t \geq \tau_k : Z_k^{K,i}(t) = 0\big\} \Big| Y_k^K\Big] - \frac{[b_k^{i,K} - d_k^{i,K}]_+}{b_k^{i,K}} \right| \tag{II.6.56}
$$

$$
= o\left(\exp(-K^\alpha)\right),
$$

where, using that $f(x,x) = 0$ for all x, we have

$$
b_k^{1,K} - d_k^{1,K} = f(Y_k^K, R^K) - (c(Y_k^K, R^K)M + \bar{c}\lceil 3/\alpha \rceil)\epsilon \sigma_K - u_K b(Y_k^K) m(Y_k^K) \tag{II.6.57}
$$

$$
= \partial_1 f(R^K, R^K)(Y_k^{K,1} - R^K) - (c(Y_k^K, R^K)M + \bar{c}\lceil 3/\alpha \rceil)\epsilon \sigma_K + O(\sigma_K^2),
$$

and similarly

$$
b_k^{2,K} - d_k^{2,K} = \partial_1 f(R^K, R^K)(Y_k^K - R^K) + c(Y_k^K, R^K)M\epsilon \sigma_K + O(\sigma_K^2). \tag{II.6.58}
$$

Recall that the sequence $(Y_k^K)_{k\geq 1}$ used to construct the processes $Z_k^{K,1}$ and $Z_k^{K,2}$ is a sequence of i.i.d. random variables with distribution $M(R^K,\cdot)$. Since $b_k^{i,K} - d_k^{i,K} < 0$ if $Y_k^K - R^K < 0$, we obtain

$$\mathbb{P}\left[\tilde{B}_k^{1,K} = 1\right] = \mathbb{E}\left[\mathbb{P}\left[\tilde{B}_k^{1,K}|Y_k^K\right] = 1\right] \tag{II.6.59}$$

$$\geq \sum_{h=1}^{A}\left(\frac{\partial_1 f(R^K,R^K)\sigma_K h - (c(Y_k^K,R^K)M + \bar{c}[3/\alpha])\epsilon\sigma_K + O(\sigma_K^2)}{b(R^K)}\right)M(R^K,h).$$

Therefore, there exists a constant $C_{\text{Bernoulli}}^1 > 0$ (which depends only on α, M, and C_L) such that the sum in the right hand side of (II.6.59) is, term by term, bounded from below by

$$\sigma_K \sum_{h=1}^{A}\left(h\frac{\partial_1 f(R^K,R^K)}{b(R^K)} - \epsilon C_{\text{Bernoulli}}^1\right)M(R^K,h) \tag{II.6.60}$$

and similarly there exists a constant $C_{\text{Bernoulli}}^2 > 0$ such that

$$\mathbb{P}[\tilde{B}_k^{2,K} = 1] \leq \sigma_K \sum_{h=1}^{A}\left(h\frac{\partial_1 f(R^K,R^K)}{b(R^K)} + \epsilon C_{\text{Bernoulli}}^2\right)M(R^K,h). \tag{II.6.61}$$

Next, we introduces two couplings, i.e.i.e., we define a sequences of i.i.d. 2-tuples of Bernoulli random variables $((B_k^{1,K}, B_k^{2,K}))_{k\geq 1}$ with the following properties

(i) $\mathbb{P}\left[B_k^{1,K} = 0, \tilde{B}_k^{1,K} = 0 \mid Y_k^K = R^K + h\sigma_K\right] = \mathbb{P}\left[\tilde{B}_k^{1,K} = 0 \mid Y_k^K = R^K + h\sigma_K\right]$ and
$\mathbb{P}\left[B_k^{1,K} = 1, \tilde{B}_k^{1,K} = 1 \mid Y_k^K = R^K + h\sigma_K\right] = q_1^K(h)\sigma_K$

(ii) $\mathbb{P}\left[B_k^{2,K} = 1, \tilde{B}_k^{2,K} = 1 \mid Y_k^K = R^K + h\sigma_K\right] = \mathbb{P}\left[\tilde{B}_k^{2,K} = 1 \mid Y_k^K = R^K + h\sigma_K\right]$ and
$\mathbb{P}\left[B_k^{2,K} = 1, \tilde{B}_k^{2,K} = 0 \mid Y_k^K = R^K + h\sigma_K\right] = 1 - q_2^K(h)\sigma_K$.

By construction, $B_k^{1,K} \leq \tilde{B}_k^{1,K}$, a.s., and $\tilde{B}_k^{2,K} \leq B_k^{2,K}$ a.s. for all $k \geq 1$ and these random variables satisfy (II.6.51) and (II.6.52). □

II.6.5 Controlling the time of the arrival of the first successful mutant

Notation. (a) For $i \in \{1,2\}$, define

$$T_k^{K,i} \equiv \inf\left\{t \geq 0 : Z_k^{K,i}(\tau_k + t) = 0 \text{ or } Z_k^{K,i}(\tau_k + t) > \epsilon\sigma_K K\right\}. \tag{II.6.62}$$

Obverse that $(T_k^{K,i})_{k\geq 1}$ are i.i.d. random variables that are independent of $A^{K,i}$.

(b) Define $I^K = k_1 \equiv \inf\{k \geq 1 : B_k^K = 1\}$ and $I^{K,i} \equiv \inf\{k \geq 1 : B_k^{K,i} = 1\}$. Then, $I^{K,i}$ are independent of $A^{K,i}$, and we have

$$\mathbb{P}\left[\left\{I^{K,2} \preccurlyeq I^K \preccurlyeq I^{K,1}\right\} \cap \left\{\tau_{I^K} \leq \hat{\theta}^K\right\}\right] = \mathbb{P}\left[\tau_{I^K} \leq \hat{\theta}^K\right] - o(\sigma_K). \tag{II.6.63}$$

(c) Define $R_1^K \equiv Y_{\inf\{k\geq 1 : B_k^K = 1\}}^K$.

In fact, we prove at the end of this section that $\mathbb{P}\left[\tau_{I^K} \leq \hat{\theta}^K\right] = 1 - o(\sigma_K)$, i.e., R_1^K is with high probability the random variable which gives the value of the next resident trait and τ_{I^K}, the first time where a successful mutant appears, is approximately exponential distributed as stated in lemma below. Note that this time is a random time, but not a stopping time.

Lemma II.6.7. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold and let M be the constant of Lemma II.6.3. Then,*

$$\lim_{K\to\infty} \sigma_K^{-1}\Big(\mathbb{P}\big[\tau_{I^K} \leq \hat{\theta}^K\big] - \mathbb{P}\big[\{E^{K,2} \leqslant \tau_{I^K} \leqslant E^{K,1}\} \cap \{\tau_{I^K} \leq \hat{\theta}^K\}\big]\Big) = 0, \qquad (II.6.64)$$

where $E^{K,1}$ and $E^{K,2}$ are exponential random variables with mean $a_1^K p_1^K \sigma_K u_K K$ respectively $a_2^K p_2^K \sigma_K u_K K$.

In other words, we have $\mathbb{P}\big[E^{K,2} \leqslant \tau_{I^K} \leqslant E^{K,1}\big|\tau_{I^K} \leq \hat{\theta}^K\big] = 1 - o(\sigma_K)$, provided that $\liminf_{K\to\infty} \mathbb{P}\big[\tau_{I^K} \leq \hat{\theta}^K\big] > 0$.

Proof. Let $A_t^{K,i}$ be defined as in Lemma II.6.4 and observe that $\tau_{I^K} = \inf\{t \geq 0 : L_t^K = I^K\}$. Then, we obtain by construction,

$$\mathbb{P}\left[\Big\{\inf\{t \geq 0 : A_t^{K,2} = I^{K,2}\} \leqslant \tau_{I^K} \leqslant \inf\{t \geq 0 : A_t^{K,1} = I^{K,1}\}\Big\} \cap \big\{\tau_{I^K} \leq \hat{\theta}^K\big\}\right] \qquad (II.6.65)$$

$$= \mathbb{P}\big[\tau_{I^K} \leq \hat{\theta}^K\big] - o(\sigma_K).$$

By definition, $I^{K,1}$ and $I^{K,2}$ are geometrically distributed with parameter $p_1^K \sigma_K$, resp. $p_2^K \sigma_K$. $A^{K,1}$ and $A^{K,2}$ are Poisson counting processes with parameter $a_1^K u_K K$, resp. $a_2^K u_K K$. Therefore, the times between each pair of successive events is exponential distributed with parameter $a_1^K u_K K$ resp. $a_2^K u_K K$. Since the random variables $I^{K,i}$ are independent of $A^{K,i}$ and the sum of a geometrically distributed number of independent exponentially distributed random variables is again exponentially distributed, we get that $\inf\{t \geq 0 : A_t^{K,1} = I^{K,1}\}$ and $\inf\{t \geq 0 : A_t^{K,2} = I^{K,2}\}$ are exponentially distributed with parameter $a_1^K u_K K p_1^K$ respectively $a_2^K u_K K p_2^K$. $\qquad\square$

II.6.6   No surprises happen before the successful mutant invades

In the next lemma we prove that a mutant invades with high probability before the resident population exits the neighborhood of this equilibrium, before too many different mutant traits are present and before a mutant of a mutant appears.

Lemma II.6.8. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold and let M be the constant of Lemma II.6.3. Then,*

$$\lim_{K\to\infty} \sigma_K^{-1}\mathbb{P}\big[\theta_{invasion}^K \geq \theta_{diversity}^K \wedge \exp(K^\alpha) \wedge \theta_{mut.\ of\ mut.}^K\big] = 0. \qquad (II.6.66)$$

Proof. We start with proving the following

$$\mathbb{P}\big[\theta_{\text{diversity}}^K < (K u_K \sigma_K^{1+\alpha})^{-1} \wedge \theta_{\text{invasion}}^K \wedge \theta_{\text{mut. of mut.}}^K\big] = o(\sigma_K). \qquad (II.6.67)$$

Define

$$\hat{Z}_k^{K,2}(s) \equiv \begin{cases} 0 & \text{for } s < \inf\{t \geq 0 : A_t^{K,2} = k\} \\ Z_k^{K,2}\Big(\tau_k + s - \inf\{t \geq 0 : A_t^{K,2} = k\}\Big) & \text{for } s \geq \inf\{t \geq 0 : A_t^{K,2} = k\}. \end{cases}$$

By construction of $A^{K,2}$ and $\hat{Z}^{K,2}$, the left hand side of (II.6.67) does not exceed

$$\mathbb{P}\left[\inf\left\{t \geq 0 : \sum_{k=1}^{A_t^{K,2}} \mathbb{1}_{\{1 \leq \hat{Z}_k^{K,2}(t) \leq \epsilon\sigma_K K\}} \geq \Big\lceil\tfrac{3}{\alpha}\Big\rceil - 1\right\} < (K u_K \sigma_K^{1+\alpha})^{-1}\right] + o(\sigma_K). \qquad (II.6.68)$$

Next, we compute an upper bound for the mutation events that happen before $(Ku_K\sigma_K^{1+\alpha})^{-1}$. Since $A^{K,2}$ is a Poisson counting process with parameter $a_2^K u_K K$, Chebyshev's inequality implies that

$$\mathbb{P}\left[A^{K,2}_{(Ku_K\sigma_K^{1+\alpha})^{-1}} \geq 2a_2^K\sigma_K^{-1-\alpha}\right] \leq \frac{\mathrm{Var}\left(A^{K,2}_{(Ku_K\sigma_K^{1+\alpha})^{-1}}\right)}{\left(2a_2^K\sigma_K^{-1-\alpha}\right)^2} = \frac{1}{a_2^K\sigma_K^{-1-\alpha}}. \tag{II.6.69}$$

Next we need an upper bound for the lifetimes of the mutant's traits, $T_k^{K,2}$. First, observe that the probability that $Z_k^{K,2}$ goes extinct after it has reached the value $\lceil\epsilon\sigma_K K\rceil$ converges to zero very fast. More precisely, Proposition II.9.3 and II.9.4 (a) imply that

$$\mathbb{P}\left[\inf\left\{t \geq 0 : Z_k^{K,2} = \lceil\epsilon\sigma_K K\rceil\right\} < \inf\left\{t \geq \tau_k : Z_k^{K,2} = 0\right\} < \infty\right] \tag{II.6.70}$$
$$= \mathbb{P}\left[\inf\left\{t \geq \tau_k : Z_k^{K,2} = 0\right\} < \infty\right] - \mathbb{P}\left[\inf\left\{t \geq 0 : Z_k^{K,2} = \lceil\epsilon\sigma_K K\rceil\right\} > \inf\left\{t \geq \tau_k : Z_k^{K,2} = 0\right\}\right]$$
$$= o(\exp(-K^\alpha)).$$

Note that, for each k, $Z_k^{K,2}$, conditioned on extinction, is a subcritical linear birth-death process (cf. [81]). Let $\check{Z}_k^{K,2}$ denote the conditioned process. If $Z_k^{K,2}$ is subcritical, then conditioning has no effect, otherwise the birth-death rates are exchanged. Denote by $\check{b}_k^{K,2}$ the birth rate and $\check{d}_k^{K,2}$ the death rate of $\check{Z}_k^{K,2}$. Then there exist uniform constants, $\check{C}_1 > 0$ and $\check{C}_2 > 0$, such that $\check{C}_1\sigma_K \leq \check{d}_k^{K,2} - \check{b}_k^{K,2} \leq \check{C}_2\sigma_K$, for all $k < I^{K,2}$. Thus, [6] p. 109 entails, for all $k < I^{K,2}$,

$$\mathbb{P}\left[T_k^{K,2} \leq t\right] \geq \frac{\check{d}_k^{K,2} - e^{(\check{d}_k^{K,2}-\check{b}_k^{K,2})t}\check{d}_k^{K,2}}{\check{b}_k^{K,2} - e^{(\check{d}_k^{K,2}-\check{b}_k^{K,2})t}\check{d}_k^{K,2}} - o(\exp(-K^{-\alpha})). \tag{II.6.71}$$

The error term $o(\exp(-K^{-\alpha}))$ appears since $Z_k^{K,2}$, for $k < I^{K,2}$, is conditioned on extinction before reaching the value $\lceil\epsilon\sigma_K K\rceil$ and not only on extinction. Choose $t = (\check{d}_k^{K,2} - \check{b}_k^{K,2})^{-1}\ln(K)$, Then,

$$\mathbb{P}\left[T_k^{K,2} \leq (\check{d}_k^{K,2} - \check{b}_k^{K,2})^{-1}\ln(K)\right] = \frac{\check{d}_k^{K,2}(1-K)}{\check{b}_k^{K,2}(1-K) - K(\check{d}_k^{K,2} - \check{b}_k^{K,2})} - o(\exp(-K^{-\alpha})) \tag{II.6.72}$$
$$= 1 + \frac{\check{d}_k^{K,2} - \check{b}_k^{K,2}}{\check{b}_k^{K,2}(1-K) - K(\check{d}_k^{K,2} - \check{b}_k^{K,2})} - o(\exp(-K^{-\alpha}))$$
$$= 1 - O(\sigma_K K^{-1})$$

and hence

$$\mathbb{P}\left[\forall 1 \leq k < I^{K,2} : T_k^{K,2} \leq (\check{C}_1\sigma_K)^{-1}\ln(K)\right] = 1 - o(\sigma_K). \tag{II.6.73}$$

Therefore, we can bound the first summand of (II.6.68) by $2a_2^K\sigma_K^{-1-\alpha}$ times the probability that more than $\lceil3/\alpha\rceil-1$ mutation events of $A^{K,2}$ take place in an interval of length $(\check{C}_1\sigma_K)^{-1}\ln(K)$. More precisely, (II.6.68) is smaller than

$$2a_2^K\sigma_K^{-1-\alpha}\mathbb{P}\left[A^{K,2}_{(\check{C}_1\sigma_K)^{-1}\ln(K)} \geq \lceil3/\alpha\rceil - 1\right] + o(\sigma_K). \tag{II.6.74}$$

Thus, for α small enough, the proof of (II.6.67) is concluded by the observation that

$$\mathbb{P}\left[A^{K,2}_{(\check{C}_1\sigma_K)^{-1}\ln(K)} \geq \lceil 3/\alpha\rceil - 1\right] \tag{II.6.75}$$

$$= e^{-a_2^K u_K K(\check{C}_1\sigma_K)^{-1}\ln(K)} \sum_{i=\lceil 3/\alpha\rceil - 1}^{\infty} \frac{(a_2^K u_K K(\check{C}_1\sigma_K)^{-1}\ln(K))^i}{i!}$$

$$\leq \left(a_2^K u_K K(\check{C}_1\sigma_K)^{-1}\ln(K)\right)^{\lceil 3/\alpha\rceil - 1}$$

$$= o(\sigma_K^{3-\alpha}),$$

where the last equality holds since $u_K K\sigma_K^{-1}\ln(K) \ll (\sigma_K)^\alpha$.

Next, we want to prove that

$$\mathbb{P}\left[\theta^K_{\text{mut. of mut.}} < (Ku_K\sigma_K^{1+\alpha})^{-1} \wedge \theta^K_{\text{invasion}} \wedge \theta^K_{\text{diversity}}\right] = o(\sigma_K). \tag{II.6.76}$$

Set, for all $\lambda \geq 0$,

$$G(\lambda) = \mathbb{E}\left[\exp\left(-\lambda \int_0^\infty Z_t\, dt\right)\bigg| Z_0 = 1\right], \tag{II.6.77}$$

where $(Z_t, t \geq 0)$ is a linear birth-death process with individual birth rate b and individual death rate d. Applying the strong Markov property and the branching property at the first jump time of Z and using the facts that
$G(\lambda)^2 = \mathbb{E}\left[\exp\left(-\lambda \int_0^\infty Z_t\, dt\right)\big| Z_0 = 2\right]$ and $\mathbb{E}\left[\exp\left(-\lambda\tau_{\text{first jump}}\right)\big| Z_0 = 1\right] = \frac{b+d}{b+d+\lambda}$, we obtain

$$bG(\lambda)^2 - (b+d+\lambda)G(\lambda) + d = 0. \tag{II.6.78}$$

Thus, since

$$\lim_{\lambda\downarrow 0} G(\lambda) = \lim_{\lambda\downarrow 0} \mathbb{E}\left[\exp\left(-\lambda \int_0^\infty Z_t\, dt\right)\mathbb{1}_{\{\tau_{\text{extinction}}<\infty\}}\bigg| Z_0 = 1\right] \tag{II.6.79}$$

$$+ \lim_{\lambda\downarrow 0} \mathbb{E}\left[\exp\left(-\lambda \int_0^\infty Z_t\, dt\right)\mathbb{1}_{\{\tau_{\text{extinction}}=\infty\}}\bigg| Z_0 = 1\right]$$

$$= \mathbb{P}\left[\tau_{\text{extinction}} < \infty\right] + 0,$$

which is 0 in the subcritical case and $1 - d/b$ in the supercritical case, it follows that

$$G(\lambda) = \frac{b+d+\lambda - \sqrt{(b+d+\lambda)^2 - 4bd}}{2b}. \tag{II.6.80}$$

Let $\tilde{Z}_k^{K,2}(t) \equiv Z_k^{K,2}(\tau_k + t)$, i.e., a linear birth-death process with birth rate $b_k^{K,2}$ and death rate $d_k^{K,2}$. Observe that $\int_0^\infty \tilde{Z}_k^{K,2}(t)\, dt$ gives an upper bound for the sum of the lifetimes of all individuals with label k. Since the mutation rate of any individual in the population is smaller than $\bar{b}u_K$, the probability that a mutant appears, which was born from an unsuccessful mutant with label k, is bounded from above by

$$1 - \mathbb{E}\left[\exp\left(-u_K\bar{b}\int_0^\infty \tilde{Z}_k^{K,2}(t)\, dt\right)\bigg|\tau_{\text{extinction}} < \inf\{t \geq 0 : \tilde{Z}_k^{K,2}(t) > \epsilon\sigma_K K\}\right] \tag{II.6.81}$$

$$\leq 1 - \mathbb{E}\left[\exp\left(-u_K\bar{b}\int_0^\infty \tilde{Z}_k^{K,2}(t)\, dt\right)\bigg|\tau_{\text{extinction}} < \infty\right] + o(\exp(-K^\alpha)).$$

Since $\tilde{Z}_k^{K,2}(t)$, conditioned on extinction, is a sub-critical linear birth-death process, the right hand side of (II.6.81) is equal to $1 - G_{\mathbb{E}[\tilde{Z}_k^{K,2}|\tau_{\text{extinction}}<\infty]}(u_K\bar{b}) + o(\exp(-K^{\alpha}))$ and

$$G_{\mathbb{E}[\tilde{Z}_k^{K,2}|\tau_{\text{extinction}}<\infty]}(u_K\bar{b}) = \begin{cases} \frac{b_k^{K,2}+d_k^{K,2}+u_K\bar{b}-\sqrt{(b_k^{K,2}+d_k^{K,2}+u_K\bar{b})^2-4b_k^{K,2}d_k^{K,2}}}{2b_k^{K,2}} & \text{if } d_k^{K,2} > b_k^{K,2} \\ \frac{d_k^{K,2}+b_k^{K,2}+u_K\bar{b}-\sqrt{(d_k^{K,2}+b_k^{K,2}+u_K\bar{b})^2-4d_k^{K,2}b_k^{K,2}}}{2d_k^{K,2}} & \text{if } b_k^{K,2} > d_k^{K,2} \end{cases}$$

$$= \begin{cases} \frac{2b_k^{K,2}+u_K\bar{b}-O(u_K\sigma_K^{-1})}{2b_k^{K,2}} & \text{if } d_k^{K,2} > b_k^{K,2} \\ \frac{2d_k^{K,2}+u_K\bar{b}-O(u_K\sigma_K^{-1})}{2d_k^{K,2}} & \text{if } b_k^{K,2} > d_k^{K,2} \end{cases}$$

$$= 1 - O(u_K\sigma_K^{-1}) = 1 - o(\sigma_k^{2+\alpha}K^{-2\alpha}).$$

Note that we used for the second equality that $|b_k^{K,2} - d_k^{K,2}| = \xi\sigma_K$ for some $\xi > 0$. By (II.6.69), the total number of unsuccessful mutations until $(Ku_K\sigma_K^{1+\alpha})^{-1} \wedge \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K$ is with probability $1 - o(\sigma_K)$ smaller or equal $2a_2^K\sigma_K^{-1-\alpha}$. Therefore, we finally obtain that the probability to have one mutant of an unsuccessful mutant during that time is $o(\sigma_K)$. On the other hand, let P_t^K be a Poisson counting process with parameter $\bar{b}u_K\epsilon\sigma_K K$ and $(\tilde{Z}_t^{K,1}, t \geq 0)$ a linear birth-death process with initial state 1 and birth rate $b^{K,1}(Y_{I_K}^K)$ and death rate $d^{K,1}(Y_{I_K}^K)$, then the probability to have one mutant of the successful mutant until the time $(Ku_K\sigma_K^{1+\alpha})^{-1} \wedge \theta_{\text{invasion}}^K \wedge \theta_{\text{diversity}}^K$ is bounded from above by

$$\mathbb{P}\left[P_{\tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}}}^K \neq 0 \Big| \tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} < \tau_0^{\tilde{Z}^{K,1}}\right] + o(\sigma_K) \tag{II.6.82}$$

$$= \mathbb{E}\left[\mathbb{1}_{\left\{P_{\tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}}}^K \neq 0\right\}}\left(\mathbb{1}_{\{\tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} \leq t_K\}} + \mathbb{1}_{\{\tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} > t_K\}}\right)\Big| \tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} < \tau_0^{\tilde{Z}^{K,1}}\right] + o(\sigma_K)$$

$$\leq (1 - \exp(-\bar{b}u_K\epsilon\sigma_K K t_K)) + \mathbb{P}\left[\tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} > t_K \Big| \tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} < \tau_0^{\tilde{Z}^{K,1}}\right] + o(\sigma_K),$$

for each t_K, because the mutation rate per individual is bounded by $\bar{b}u_K$ and there are at most $\epsilon\sigma_K K$ successful mutant individuals alive until $\theta_{\text{invasion}}^K$. If we choose $t_K = \ln(K)\sigma_K^{-1-\alpha/2}$, then by Proposition II.9.4, all terms in the last line of (II.6.82) are $o(\sigma_K)$. This implies (II.6.76).

Note that we have $\theta_{\text{invasion}}^K = \tau_{I^K} + \inf\left\{t \geq 0 : \mathfrak{M}^{I^K}(\tilde{\nu}_{\tau_{I^K}+t}) > \epsilon\sigma_K K\right\}$. Let $E^{K,1}$ be an exponential distributed random variable with mean $a_1^K p_1^K \sigma_K u_K K$. Then,

$$\mathbb{P}\left[\tau_{I^K} + \inf\left\{t \geq 0 : \mathfrak{M}^{I^K}(\tilde{\nu}_{\tau_{I^K}+t}) > \epsilon\sigma_K K\right\} \geq \theta_{\text{diversity}}^K \wedge e^{K^{\alpha}} \wedge \theta_{\text{mut. of mut.}}^K\right] \tag{II.6.83}$$

$$\geq \mathbb{P}\left[E^{K,1} + T_{I^K}^{K,1} \geq (Ku_K\sigma_K^{1+\alpha})^{-1}\right] - o(\sigma_K).$$

Let $\tilde{Z}^{K,1}$ as defined before, then again by Proposition II.9.4

$$\mathbb{P}\left[T_{I^K}^{K,1} > \ln(K)\sigma_K^{-1-\alpha/2}\right] = \mathbb{P}\left[\tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} > \ln(K)\sigma_K^{-1-\alpha/2} \Big| \tau_{\epsilon\sigma_K K}^{\tilde{Z}^{K,1}} < \tau_0^{\tilde{Z}^{K,1}}\right] = o(\sigma_K). \tag{II.6.84}$$

Since $\ln(K)\sigma_K^{-1-\alpha/2} \ll (Ku_K\sigma_K^{1+\alpha})^{-1}$, the Markov inequality for the function $f(x) = x^n$, where n is smallest even number which is larger than $2/\alpha$, yields

$$\mathbb{P}\left[E^{K,1} + T_{I^K} > (Ku_K\sigma_K^{1+\alpha})^{-1}\right] \leq \mathbb{P}\left[E^{K,1} > (2Ku_K\sigma_K^{1+\alpha})^{-1}\right] + o(\sigma_K) \tag{II.6.85}$$

$$\leq \frac{(2Ku_K\sigma_K^{1+\alpha})^n n!}{(a_1^K p_1^K u_K K \sigma_K)^n} = O(\sigma_K^2).$$

$\square$

The following lemma shows that there are no two successful mutants during the first phase of an invasion.

Lemma II.6.9. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold and let M be the constant of Lemma II.6.3. Then,*

$$\lim_{K\to\infty} \sigma_K^{-1} \, \mathbb{P}\Big[\text{ There is a successful mutation in time interval } [\tau_{I^K}, \theta^K_{invasion}]\Big] = 0. \quad \text{(II.6.86)}$$

Proof. Let $P^K_{\text{suc. mut.}}(t)$ the process which recodes the number of successful mutants born after τ_{I^K} until $\tau_{I^K} + t$. Then,

$$\mathbb{P}\Big[\text{for all } t \geq 0 \text{ such that } \tau_{I^K} + t < \hat{\theta}^K \, : \, P^K_{\text{succ. mut.}}(t) \leqslant P^K_t \Big] = 1 - o(\sigma_K), \quad \text{(II.6.87)}$$

where P^K_t is Poisson process with parameter $a_2^K p_2^K \sigma_K u_K K$. Define $Z^{K,2}_{I^K}(t)$ as in Lemma II.6.5. Then $\mathbb{P}[\forall t \leq \hat{\theta}^K : \mathfrak{M}^{I^K}(\tilde{\nu}_t) \leqslant Z^{K,2}_{I^K}(t)] \geq 1 - o(\sigma_K)$. Note that P^K_t and $Z^{K,2}$ are independent by construction. Therefore, as in the last lemma, or each t_K,

$$\mathbb{P}\Big[\text{ There is a successful mutation in } [\tau_{I^K}, \theta^K_{invasion}] \Big] \quad \text{(II.6.88)}$$

$$\leq \mathbb{P}\Big[P^K_{\tau^{Z^{K,2}}_{\epsilon\sigma_K K}} \neq 0 \Big| \tau^{Z^{K,2}}_{\epsilon\sigma_K K} < \tau^{Z^{K,2}}_0 \Big] + o(\sigma_K)$$

$$\leq \big(1 - \exp(-a_2^K p_2^K \sigma_K u_K K t_K)\big) + \mathbb{P}\Big[\tau^{Z^{K,2}}_{\epsilon\sigma_K K} > t_K \Big| \tau^{Z^{K,2}}_{\epsilon\sigma_K K} < \tau^{Z^{K,2}}_0 \Big] + o(\sigma_K).$$

With $t_K = \ln(K)\sigma_K^{-1-\alpha/2}$, by Proposition II.9.4, all terms in the last line of (II.6.88) are $o(\sigma_K)$. $\qquad\square$

II.6.7   Finishing up: control of the distribution of the next resident trait

Corollary II.6.10. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.6.2 hold and let M be the constant of Lemma II.6.3. Then, there exist two $\mathcal{X}$-valued random variables $R_1^{K,1}$ and $R_1^{K,2}$ with distribution*

$$\mathbb{P}[R_1^{K,1} = R^K + \sigma_K h] = \begin{cases} \dfrac{M(R^K,1)q_1^K(1)}{p_2^K} + 1 - \dfrac{p_1^K}{p_2^K} & \text{if } h = 1 \\[2mm] \dfrac{M(R^K,h)q_1^K(h)}{p_2^K} & \text{if } h \in \{2, ..., A\}, \end{cases} \quad \text{(II.6.89)}$$

and

$$\mathbb{P}[R_1^{K,2} = R^K + \sigma_K h] = \begin{cases} \dfrac{M(R^K,h)q_1^K(h)}{p_2^K}, & \text{if } h \in \{1, ..., A-1\} \\[2mm] \dfrac{M(R^K,A)q_1^K(A)}{p_2^K} + 1 - \dfrac{p_1^K}{p_2^K} & \text{if } h = A, \end{cases} \quad \text{(II.6.90)}$$

such that

$$\lim_{K\to\infty} \sigma_K^{-1}\left(1 - \mathbb{P}\Big[R_1^{K,1} \leqslant R_1^K \leqslant R_1^{K,2} \Big| \theta^K_{invasion} < \theta^K_{diversity} \wedge \theta^K_{mut.\ of\ mut.} \wedge e^{K^\alpha}\Big]\right) = 0. \quad \text{(II.6.91)}$$

Proof. Define

$$R_1^{K,1} \equiv \begin{cases} Y^K_{I^K}, & \text{if } I^{K,1} = I^{K,2}, \\ R^K + \sigma_K, & \text{otherwise,} \end{cases} \quad \text{and} \quad R_1^{K,2} \equiv \begin{cases} Y^K_{I^K}, & \text{if } I^{K,1} = I^{K,2}, \\ R^K + A\sigma_K, & \text{otherwise.} \end{cases}$$

By construction of $B_k^{K,i}$ and $Y_k^{K,i}$, we have that (II.6.91) holds. Next, we compute

$$\mathbb{P}\left[Y_{I^{K,2}}^K = R^K + \sigma_K h, \ I^{K,1} = I^{K,2}\right] = \mathbb{P}\left[Y_1^K = R^K + \sigma_K h, \ B_1^{K,1} = 1 \big| B_1^{K,2} = 1\right] \tag{II.6.92}$$

$$= \frac{\mathbb{P}[Y_1^K = R^K + \sigma_K h, \ B_1^{K,1} = 1]}{\mathbb{P}[B_1^{K,2} = 1]}$$

$$= \frac{M(R^K, h)q_1^K(h)}{p_2^K}$$

and $\mathbb{P}[I^{K,1} \neq I^{K,2}] = 1 - \sum_{h=1}^{A} \frac{M(R^K, h)q_1^K(h)}{p_2^K} = 1 - p_1^K/p_2^K$. Since $\mathbb{P}[R_1^{K,1} = R^K + \sigma_K h] = \mathbb{P}[Y_{I^{K,2}}^K = R^K + \sigma_K h, \ I^{K,1} = I^{K,2}] + \mathbb{1}_{\{h=1\}}\mathbb{P}[I^{K,1} \neq I^{K,2}]$ and similarly for $R_1^{K,2}$, we deduce (II.6.89) and (II.6.90). $\qquad\square$

II.7 The second phase of an invasion

Theorem II.7.1 below describes precisely how the invading mutant replaces the resident population. This section is the central piece of the entire paper.

Notation. Let us denote

$$\theta_{\text{fixation}}^K = \inf\left\{t \geq \theta_{\text{invasion}}^K : |\text{Supp}(\tilde{\nu}_t^K)| = 1 \text{ and } |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \bar{z}(R_1^K)| < (M/3)\epsilon\sigma_K\right\}$$

i.e., the first time after $\theta_{\text{invasion}}^K$ such that the population is monomorphic and in the $(M/3)\epsilon\sigma_K$-neighborhood of the corresponding equilibrium.

Theorem II.7.1. *Fix $\epsilon > 0$. Under the Assumptions 2, 4, and 5, there exists a constant, $M > 0$, such that, for all K large enough,*

(i) *$\tilde{\nu}_0^K = N_{R^K}^K K^{-1}\delta_{(0,R^K)}$, where $\left|\bar{z}(R^K) - N_{R^K}^K K^{-1}\right| < (M/3)\epsilon\sigma_K$ a.s..*

(ii) *At the first time of invasion, $\theta_{\text{invasion}}^K$, the resident density is in an $\epsilon M\sigma_K$-neighborhood of $\bar{z}(R^K)$, the number of different living mutant traits is bounded by $\lceil\alpha/3\rceil$ and there is no mutant of a mutant, with probability $1 - o(\sigma_K)$. (cf. Theorem II.6.2)*

(iii) *The time between $\theta_{\text{invasion}}^K$ and $\theta_{\text{fixation}}^K$ is smaller than $5\ln(K)\sigma_K^{-1-\alpha/2}$, with probability $1 - o(\sigma_K)$.*

(iv) *The trait of the population at time $\theta_{\text{fixation}}^K$ is the trait of the mutant whose density was larger than $\epsilon\sigma_K$ at time $\theta_{\text{invasion}}^K$, i.e., $\text{Supp}(\tilde{\nu}_{\theta_{\text{fixation}}^K}^K) = (I^K, R_1^K)$, with probability $1 - o(\sigma_K)$. The distribution of R_1^K can be approximated as in Corollary II.6.10.*

Moreover, until time $\theta_{\text{fixation}}^K$, the total mass of the population stays in the $O(\sigma_K)$-neighborhood of $\bar{z}(R^K)$, the number of different living mutant traits is bounded by $\lceil\alpha/3\rceil$, and there is no second successful mutant, with probability $1 - o(\sigma_K)$.

To prove this theorem, we divide this phase into five steps, as illustrated in Figure II.2.

Step 1 From $\theta_{\text{invasion}}^K$ to $\theta_{\text{mut. size }\epsilon}^K$, the first time when a mutant's density reaches the value ϵ. During this period we approximate the mutant density by a continuous time branching process, which is supercritical (of order σ_K). Thus we obtain that $\theta_{\text{mut. size }\epsilon}^K - \theta_{\text{invasion}}^K$ is of order $(\ln(K)\sigma_K^{-1})$.

Step 2 From $\theta^K_{\text{mut. size } \epsilon}$ to $\theta^K_{\text{mut. size } C^\epsilon_{\text{cross}}}$, the first time when the mutant density reaches a value $C^\epsilon_{\text{cross}}$ (defined in Eq. (II.7.1) below). This step can be seen as the "stochastic Euler scheme". The idea is that the total mass of the population stays close to a function which depends only on the density of the successful mutant. This allows to approximate the number of mutants by a discrete time Markov chain until the mutant density has increased by ϵ. Furthermore we control the number of jumps needed to increase by ϵ and use upper and lower bounds for one jump time of the associated continuous time process to control the time of this step. Then we recompute the parameters and start again. Iterating, we obtain that $\theta^K_{\text{mut. size } C^\epsilon_{\text{cross}}} - \theta^K_{\text{mut. size } \epsilon}$ is also of order $\ln(K)\sigma_K^{-1}$.

Step 3 From $\theta^K_{\text{mut. size } C^\epsilon_{\text{cross}}}$ until $\theta^K_{\text{res. size } \epsilon}$, the first time when the density of the resident trait R^K decreases to the value ϵ. The proof is very similar to the proof of Step 2, the only difference is that we approximate the number of resident individuals by a discrete Markov chain, which decreases slowly.

Step 4 From $\theta^K_{\text{res. size } \epsilon}$ until $\theta^K_{\text{res. size } 0}$, the first time when the resident trait R^K goes extinct. We approximate the dynamics of the resident trait by a continuous time branching process which is subcritical (of order σ_K) and therefore goes extinct, a.s., after a time of order $\ln(K)\sigma_K^{-1}$.

Step 5 From $\theta^K_{\text{res. size } 0}$ until $\theta^K_{\text{fixation}}$. Even if it is unlikely that this time period is larger than 0, we have to obtain an upper bound for this time.
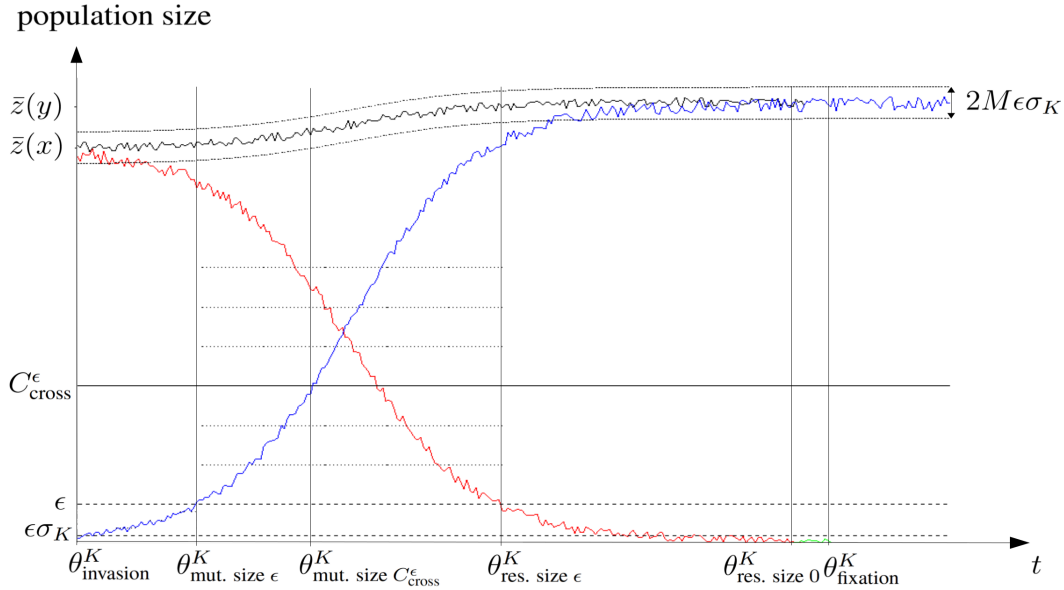
Figure II.2: Evolution of the population after the destiny of the successful mutant has reached the value ϵ_K.

Notation. Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.7.1 hold. Set

$$C^\epsilon_{\text{cross}} \equiv \left\lceil \left(\inf_{x \in \mathcal{X}} \frac{b(x) - d(x)}{c(x,x)} \right) \epsilon^{-1} \right\rceil \frac{\epsilon}{2}, \tag{II.7.1}$$

$$\text{and} \qquad \theta^K_{2 \text{ succ. mut.}} \equiv \inf \left\{ t \geq 0 : \sum_{k=0}^{\infty} \mathbb{1}_{\mathfrak{M}^k(\tilde{\nu}_t) \geq \epsilon_K K} \geq 3 \right\}. \tag{II.7.2}$$

Moreover, for any $\xi \geq 0$,

$$\theta^K_{\text{mut. size } \xi} \equiv \inf\left\{t \geq 0 : \exists k \geq 1 : \mathfrak{M}^k(\tilde{\nu}_t) = \lceil \xi K \rceil\right\}, \tag{II.7.3}$$

$$\theta^K_{\text{res. size } \xi} \equiv \inf\left\{t \geq 0 : \mathfrak{M}^0(\tilde{\nu}_t) = \lceil \xi K \rceil\right\}, \tag{II.7.4}$$

and let S_K be a sequence in K such that $1 \ll S_K \ll \epsilon\sigma_K^{-1}$.

Remark 6. Using similar arguments as in the proofs of Lemma II.6.3, II.6.8, and II.6.9, we obtain

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\left[\theta^K_{\text{invasion}} + 5\sigma_K^{-1-\alpha/2}\ln(K) > \theta^K_{\text{diversity}} \wedge \theta^K_{2 \text{ succ. mut.}} \wedge e^{K^\alpha}\right] = 0. \tag{II.7.5}$$

More precisely, until the time $\theta^K_{\text{diversity}} \wedge \theta^K_{2 \text{ succ. mut.}} \wedge \exp(K^\alpha)$ the total mass of the population stays with high probability in the $O(\sigma_K)$ neighborhood of $\bar{z}(R^K)$. This can be proved similarly as Lemma II.6.3 or II.7.2. Since we have only an approximation of order σ_K (not $\epsilon\sigma_K$), we have less precise bounds for the rates of the mutants and for their success probability. Nevertheless, we can bound the mutant subpopulations from above by linear branching processes which are slightly supercritical of order σ_K.

II.7.1 Step 1: a mutant's density reaches the value ϵ

The following lemma shows that the total mass stays from the beginning (including the first phase) until $\theta^K_{\text{mut. size } \epsilon}$ in the $M\epsilon\sigma_K$ neighborhood of $\bar{z}(x)$.

Lemma II.7.2. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.7.1 hold. Then, there exists a constant $M > 0$ (independent of ϵ and K) such that*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\inf\left\{t \geq 0 : |\langle\tilde{\nu}_t, \mathbb{1}\rangle - \bar{z}(R^K)| > M\epsilon\sigma_K\right\} \tag{II.7.6}$$

$$< \theta^K_{\text{mut. size } \epsilon} \wedge \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}} \wedge \exp(K^\alpha)\Big] = 0.$$

Proof. The proof of this lemma is very similar to the one of Lemma II.6.3, therefore we omit some details. Define

$$X_t \equiv |\langle\tilde{\nu}_t, \mathbb{1}\rangle K - \lceil K\bar{z}(R^K)\rceil|. \tag{II.7.7}$$

We associate with the continuous time process X_t a discrete time (non-Markov) process Y_n which records the sequence of values that X_t takes on.

CLAIM. *For $1 \leq i \leq \epsilon K$ and K large enough,*

$$\mathbb{P}\Big[Y_{n+1} = i+1 \big| Y_n = i, T_{n+1} < \theta^K_{\text{mut. size } \epsilon} \wedge \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}}\Big] \tag{II.7.8}$$

$$\leq \frac{1}{2} - (\underline{c}/4\overline{b})K^{-1}i + (2C_L^{b,d,c}A/\underline{b})\epsilon\sigma_K \equiv p_+^K(i),$$

where $C_L^{b,d,c}$ is the sum of the Lipschitz constants for the birth, death and competition rate.

This can be proven exactly as in Lemma II.6.3, using the facts that $b(R^K) = d(R^K) + c(R^K, R^K)\bar{z}(R^K)$ and that all mutant traits are at a distance of at most $2A\sigma_K$ from R^K, and hence, $|b(x) - b(R^K)| < C_L^b\sigma_K 2A$, $|d(x) - d(R^K)| < C_L^d\sigma_K 2A$, and $|c(x,y) - c(R^K, R^K)| < C_L^c\sigma_K 2A$ for all traits x and y alive in the population. By continuing as in Lemma II.6.3 we obtain (II.7.6). $\qquad \square$

Next we prove that $\theta_{\text{invasion}}^K - \theta_{\text{mut. size }\epsilon}^K$ is smaller than $\ln(K)\sigma_K^{-1-\alpha/2}$. We use the following notation.

Notation. $\tilde{\theta}^K \equiv \inf\left\{t \geq 0 : |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \overline{z}(R^K)| > M\epsilon\sigma_K\right\} \wedge \theta_{2 \text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K$.

Lemma II.7.3. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.7.1 hold. Let M be the constant from Lemma II.7.2. Then,*

$$\lim_{K\to\infty} \sigma_K^{-1}\, \mathbb{P}\left[\theta_{\text{mut. size }\epsilon}^K > \left(\theta_{\text{invasion}}^K + \ln(K)\sigma_K^{-1-\alpha/2}\right) \wedge \tilde{\theta}^K\right] = 0. \tag{II.7.9}$$

Proof. To prove this lemma we use a coupling with a linear continuous time birth-death process. From the results on Phase 1 and Lemma II.7.3, we know that $\theta_{\text{invasion}}^K$ is, with probability $1 - o(\sigma_K)$, smaller than $\tilde{\theta}^K$. Recall $I^K \equiv k_1$, the label of the first successful mutation (see (II.6.5)). For any $t \in (\theta_{\text{invasion}}^K, \tilde{\theta}^K]$, any individual of $\mathfrak{M}^{k_1}(\tilde{\nu}_t)$ gives birth to a new individual with the same trait, R_1^K, with rate

$$\left(1 - u_K\, m(R_1^K)\right)b(R_1^K) \in \left[b(R_1^K) - u_K\,\overline{b}\,, b(R_1^K)\right], \tag{II.7.10}$$

and dies with rate

$$d(R_1^K) + \int_{\mathcal{X}\times\mathbb{N}_0} c(R_1^K, \xi)d\tilde{\nu}_t(\xi), \tag{II.7.11}$$

which is smaller than $d_Z \equiv d(R_1^K) + c(R_1^K, R^K)(\overline{z}(R^K) + M\epsilon\sigma_K) + \overline{c}(\epsilon + \lceil 3/\alpha\rceil\sigma_K)A\sigma_K$. Similarly as in Lemma II.6.5 we construct, by using a standard coupling argument, a processes Z_t such that

$$Z_t \leq \mathfrak{M}^{k_1}\left(\tilde{\nu}_{\theta_{\text{invasion}}^K + t}\right) \tag{II.7.12}$$

for all t such that $\theta_{\text{invasion}}^K + t \leq \tilde{\theta}^K \wedge \inf\{t \geq 0 : \mathfrak{M}^{k_1}(\tilde{\nu}_t) \geq \epsilon K\}$. The processes Z_t is a branching process starting at $\lceil\epsilon\sigma_K K\rceil$, with birth rate per individual $b_Z = b(R_1^K) - \overline{b}u_K$ and with death rate per individual d_Z. For all $\epsilon < \inf_{x\in\mathcal{X}} \frac{\partial_1 f(x,x)}{2(M+A+1)}$, we have

$$b_Z - d_Z \geq f(R_1^K, R^K) - \overline{c}\sigma_K(M\epsilon + A(\epsilon + \lceil 3/\alpha\rceil\sigma_K)) \geq \sigma_K \inf_{x\in\mathcal{X}} \frac{\partial_1 f(x,x)}{2}. \tag{II.7.13}$$

Thus Z_t is super-critical of order σ_K. Let τ_i^Z be the first hitting time of level i by Z_t, then by Proposition II.9.4

$$\mathbb{P}[\tau_{\lceil\epsilon K\rceil}^Z > \tau_0^Z] \leq \exp(-K^\alpha). \tag{II.7.14}$$

Furthermore, we have the following exponential tail bound, see [2] page 41,

$$\mathbb{P}\left[\tau_{\lceil\epsilon K\rceil}^Z \geq \ln(K)\sigma_K^{-1-\alpha/2}\big|\tau_{\lceil\epsilon K\rceil}^Z < \tau_0^Z\right] \leq \exp\left(-\left\lfloor \frac{\ln(K)\sigma_K^{-1-\alpha/2}}{e\max_{n\leq\lceil\epsilon K\rceil} \mathbb{E}_n[\tau_{\lceil\epsilon K\rceil}^Z|\tau_{\lceil\epsilon K\rceil}^Z < \tau_0^Z]}\right\rfloor\right), \tag{II.7.15}$$

and $\max_{n\leq\lceil\epsilon K\rceil} \mathbb{E}_n[\tau_{\lceil\epsilon K\rceil}^Z|\tau_{\lceil\epsilon K\rceil}^Z < \tau_0^Z] \leq O(\ln(K)\sigma_K)$ (compare with Proposition II.9.3). Therefore,

$$\mathbb{P}\left[\tau_{\lceil\epsilon K\rceil}^Z < \ln(K)\sigma_K^{-1-\alpha/2}\right] \geq \left(1 - e^{-\sigma_K^{-\alpha/3}}\right)\left(1 - e^{-K^\alpha}\right) = 1 - o(\sigma_K), \tag{II.7.16}$$

which implies the claim. $\qquad\square$

II.7.2 Step 2: the mutant density reaches a value $C_{\text{cross}}^{\epsilon}$ (Stochastic Euler scheme)

Recall that the trait of the successful mutant is $R^K + \sigma_K h$ where $h \in \{1, \ldots, A\}$. Due to the regularity assumptions (iv) in Assumption 2, we have the following estimates:

$$
\begin{aligned}
b(R^K + \sigma_K h) &= b(R^K) + b'(R^K)\sigma_K h + O((\sigma_K h)^2) & \text{(II.7.17)} \\
d(R^K + \sigma_K h) &= d(R^K) + d'(R^K)\sigma_K h + O((\sigma_K h)^2) \\
r(R^K + \sigma_K h) &= r(R^K) + r'(R^K)\sigma_K h + O((\sigma_K h)^2) \\
c(R^K + \sigma_K h, R^K) &= c(R^K, R^K) + \partial_1 c(R^K, R^K)\sigma_K h + O((\sigma_K h)^2) \\
c(R^K, R^K + \sigma_K h) &= c(R^K, R^K) + \partial_2 c(R^K, R^K)\sigma_K h + O((\sigma_K h)^2) \\
c(R^K + \sigma_K h, R^K + \sigma_K h) &= c(R^K, R^K) + \big(\partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K)\big)\sigma_K h + O((\sigma_K h)^2).
\end{aligned}
$$

The deterministic system: Although we cannot use a law of large numbers, to understand the behavior of the stochastic system it is useful to look at the properties of the corresponding deterministic Lotka-Volterra system. The limiting system when $K \to \infty$, with $\sigma_K = 0$, takes the simple form

$$
\frac{dm_t^0}{dt} = m_t^0 \big(r(R^K) - c(R^K, R^K)(m_t^0 + m_t^{k_1}) \big), \tag{II.7.18}
$$

$$
\frac{dm_t^{k_1}}{dt} = m_t^{k_1} \big(r(R^K) - c(R^K, R^K)(m_t^0 + m_t^{k_1}) \big). \tag{II.7.19}
$$

The corresponding vector field is depicted in Figure II.3. This system has an invariant manifold made of fixed points given by the roots of the equation

$$
m^0 + m^{k_1} = r(R^K)/c(R^K, R^K) = \bar{z}(R^K), \tag{II.7.20}
$$

with $m^0, m^{k_1} \geq 0$. This manifold connects the fixed points of the monomorphic equations, $(\bar{z}(R^K), 0)$ and $(0, \bar{z}(R^K))$. Note that $\bar{z}(R^K)$ has the interpretation of the total mass of the population in equilibrium. A simple computation shows that the Hessian matrix on the invariant manifold is given by

$$
H(m^0, m^{k_1}) = -c(R^K, R^K) \begin{pmatrix} m^0 & m^0 \\ m^{k_1} & m^{k_1} \end{pmatrix}. \tag{II.7.21}
$$

The corresponding eigenvectors are $(1, -1)$ with eigenvalue 0, and $(m^0, \bar{z}(R^K) - m^0)$ with eigenvalue $-c(R^K, R^K)\bar{z}(R^K)$.

It follows that the perturbed system

$$
\frac{dm_t^0}{dt} = m_t^0 \big(r(R^K) - c(R^K, R^K)m_t^0 - c(R^K, R^K + \sigma_K h)m_t^{k_1} \big), \tag{II.7.22}
$$

$$
\frac{dm_t^{k_1}}{dt} = m_t^{k_1} \big(r(R^K + \sigma_K h) - c(R^K + \sigma_K h, R^K)m_t^0 - c(R^K + \sigma_K h, R^K + \sigma_K h)m_t^{k_1} \big), \tag{II.7.23}
$$

has an invariant manifold connecting its fixed points $(\bar{z}(R^K), 0)$ and $(0, \bar{z}(R^K + \sigma_K h))$, where $\bar{z}(R^K + \sigma_K h) = r(R^K + \sigma_K h)/c(R^K + \sigma_K h, R^K + \sigma_K h)$ in a σ_K-neighborhood of the unperturbed invariant manifold (see Figure II.3). Thus the perturbed deterministic system will move quickly towards a small neighborhood of this invariant manifold and then move slowly with speed $O(\sigma_K)$ along it. Since the invariant manifold is close to the curve $m^0 + m^{k_1} = \bar{z}(R^K)$,
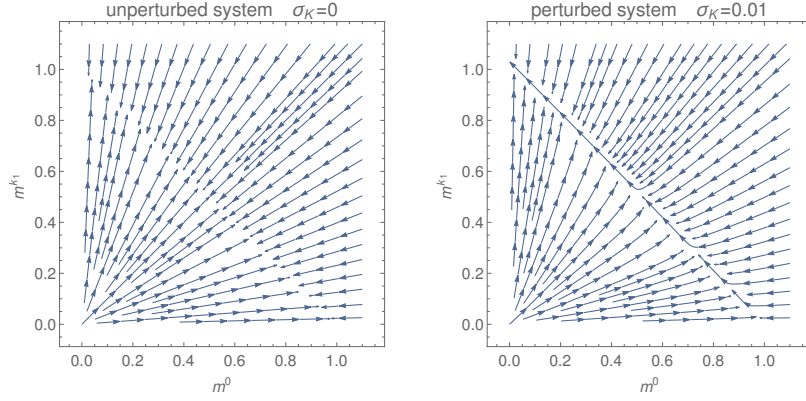
Figure II.3: Right: Vector field of the unperturbed system ($\sigma_K = 0$), Left: Vector field of the perturbed system ($\sigma_K = 0.01$). Parameters are given in Table II.1

Table II.1: Parameters of the Figures II.3

$b(R^K) = 2$	$d(R^K) = 1$	$c(R^K, R^K) = 1$	$c(R^K, R^K + \sigma_K h) = 1 - 2\sigma_K$
$b(R^K + \sigma_K h) = 2 + \sigma_K$	$d(R^K + \sigma_K h) = 1 - \sigma_K$	$c(R^K + \sigma_K h, R^K) = 1 - 2\sigma_K$	$c(R^K + \sigma_K h, R^K + \sigma_K h) = 1 - \sigma_K$

it is reasonable to choose as variables $M_t = m_t^0 + m_t^{k_1}$. The motion of the system will then be close to the curve $\tilde{\phi}(m_t^{k_1})$ defined by the condition that the derivative of M_t vanishes for $M_t = \tilde{\phi}(m_t^{k_1})$.

Since

$$\frac{dM_t}{dt} = M_t \left(r(R^K) - c(R^K, R^K) M_t \right) \tag{II.7.24}$$
$$- \left[\left(\partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K) \right) M_t - r'(R^K) \right] \sigma_K h m_t^{k_1} + O(\sigma_K^2).$$

Setting the right hand side to zero yields the leading orders in σ_K

$$\tilde{\phi}(m_t^{k_1}) = \overline{z}(R^K) + \sigma_K h m_t^{k_1} \left(\frac{r'(R^K)}{r(R^K)} - \frac{\partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K)}{c(R^K, R^K)} \right) + O(\sigma_K^2). \tag{II.7.25}$$

We expect that the stochastic system also evolves along this curve. I.e., we will show that m^{k_1} increases while the total mass stays close to the curve defined in (II.7.25).

Define the function

$$\phi(y) \equiv \overline{z}(R^K) + \sigma_K h y \left(\frac{r'(R^K)}{r(R^K)} - \frac{\partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K)}{c(R^K, R^K)} \right), \tag{II.7.26}$$

and the stopping time

$$\theta^K_{\text{near } \phi(i\frac{\epsilon}{2})} \equiv \inf \left\{ t \geq \theta^K_{\text{mut. size } i(\epsilon/2)} : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \phi(i(\epsilon/2))| < (M/3)\epsilon\sigma_K \right\}. \tag{II.7.27}$$

The dependence of ϕ with respect to the mutant density allows us to decompose the increase of the mutant density into successive steps during which the total mass does not move more than $M\epsilon\sigma_K$.

Lemma II.7.4. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.7.1 hold. Then, there exists a constant $M > 0$ (independent of ϵ, K, and i) such that and for all $2 \leq i \leq 2\epsilon^{-1} C^\epsilon_{cross}$,*

(a) Soon after $\theta^K_{\text{mut. size } i(\epsilon/2)}$, the total population size is close to $\phi(i\frac{\epsilon}{2})$:

$$\lim_{K \to \infty} \sigma_K^{-1} \, \mathbb{P}\Big[\theta_{near \ \phi(i\frac{\epsilon}{2})}^K > \Big(\theta_{mut. \ size \ i(\epsilon/2)}^K + S_K\Big) \wedge \theta_{2 \ succ. \ mut.}^K \wedge \theta_{diversity}^K$$

$$\wedge \inf\Big\{t \geq \theta_{mut. \ size \ i(\epsilon/2)}^K : \exists k \geq 1 : \mathfrak{M}^k(\tilde{\nu}_t) = \lceil (i \pm \tfrac{1}{2})(\epsilon/2)K \rceil\Big\}\Big] = 0.$$

(b) A change of order ϵ for the mutant density takes more than $o(\sigma_K^{-1})$ time:

$$\lim_{K \to \infty} \sigma_K^{-1} \, \mathbb{P}\Big[\inf\Big\{t \geq \theta_{mut. \ size \ i(\epsilon/2)}^K : \exists k \geq 1 : \mathfrak{M}^k(\tilde{\nu}_t) = \lceil (i \pm \tfrac{1}{2})(\epsilon/2)K \rceil\Big\}$$

$$< \Big(\theta_{mut. \ size \ i(\epsilon/2)}^K + S_K\Big) \wedge \theta_{near \ \phi(i\frac{\epsilon}{2})}^K \wedge \theta_{2 \ succ. \ mut.}^K \wedge \theta_{diversity}^K\Big] = 0.$$

(c) At the time when the mutant density has changed of order ϵ the total population size is still close to $\phi(i\frac{\epsilon}{2})$:

$$\lim_{K \to \infty} \sigma_K^{-1} \, \mathbb{P}\Big[\inf\Big\{t \geq \theta_{near \ \phi(i\frac{\epsilon}{2})}^K : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \phi(i\tfrac{\epsilon}{2})| > M\epsilon\sigma_K\Big\} < \theta_{2 \ succ. \ mut.}^K \wedge \theta_{diversity}^K$$

$$\wedge \inf\Big\{t \geq \theta_{mut. \ size \ i(\epsilon/2)}^K : \exists k \geq 1 : \mathfrak{M}^k(\tilde{\nu}_t) = \lceil (i \pm 1)(\epsilon/2)K \rceil\Big\}\Big] = 0.$$

(d) A change of order ϵ for the mutant density takes no more than $(i\sigma_K)^{-1-\alpha/2}$ time:

$$\lim_{K \to \infty} \sigma_K^{-1} \, \mathbb{P}\Big[\theta_{mut. \ size \ (i+1)(\epsilon/2)}^K > \Big(\theta_{near \ \phi(i\frac{\epsilon}{2})}^K + (i\sigma_K)^{-1-\alpha/2}\Big) \wedge \theta_{2 \ succ. \ mut.}^K$$

$$\wedge \theta_{diversity}^K \wedge \inf\Big\{t \geq \theta_{near \ \phi(i\frac{\epsilon}{2})}^K : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \phi(i(\epsilon/2))| > M\epsilon\sigma_K\Big\}\Big] = 0.$$

Remark 7. For each $\epsilon > 0$, Lemma II.7.4 implies that the mutant density reaches the value $C_{\text{cross}}^\epsilon$ with high probability, since ϵ is independent of K. Moreover, for all $\epsilon > 0$,

$$\mathbb{P}\Big[\theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K > \Big(\theta_{\text{mut. size } \epsilon}^K + \frac{\ln(K)}{\sigma_K^{1+\alpha/2}}\Big) \wedge \theta_{2 \text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K\Big] = o(\sigma_K) \qquad \text{(II.7.28)}$$

$$\text{and} \qquad \mathbb{P}\Big[\big|\langle \tilde{\nu}_{\theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K}, \mathbb{1} \rangle - \phi(C_{\text{cross}}^\epsilon)\big| > M\epsilon\sigma_K\Big] = o(\sigma_K). \qquad \text{(II.7.29)}$$

Proof. We will prove the lemma by induction over i. Base clause: Compare with Lemma II.7.2 and II.7.3 that there exists a constant $M > 0$ such that $|\langle \tilde{\nu}_{\theta_{\text{mut. size } \epsilon}^K}, \mathbb{1} \rangle - \phi(0)|$ is smaller than $M\epsilon\sigma_K$ and that $\theta_{\text{mut. size } \epsilon}^K < \theta_{2 \text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K$ both with probability $1 - o(\sigma_K)$.

Induction step from $i-1$ to i: Assume that the lemma holds true for $i-1$, then be prove separately that (a)-(d) are true for i, as long as $i < 2\epsilon^{-1}C_{\text{cross}}^\epsilon$

Proof. of (a) for i by assuming that the lemma holds for $i-1$. In the proof we use the following notation

$$\tilde{\theta}_i^K \equiv \theta_{2 \text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K \wedge \inf\Big\{t \geq \theta_{\text{mut. size } i(\epsilon/2)}^K : \exists k \geq 1 : \mathfrak{M}^k(\tilde{\nu}_t) = \lceil (i \pm \tfrac{1}{2})\tfrac{\epsilon}{2}K \rceil\Big\}. \quad \text{(II.7.30)}$$

Note that $\tilde{\theta}_i^K$ differs from $\tilde{\theta}^K$ defined in Lemma II.7.3. We will prove (a) provided it happens before $\tilde{\theta}_i^K$ and we use the estimates of step (b) for i to prove that it indeed happens before $\tilde{\theta}_i^K$ with high probability.

If the Lemma is true for $i-1$, we know that (with (d))

$$\mathbb{P}\Big[\big|\langle \tilde{\nu}_{\theta_{\text{mut. size } i(\epsilon/2)}^K}, \mathbb{1} \rangle - \phi((i-1)\tfrac{\epsilon}{2})\big| < M\epsilon\sigma_K\Big] = 1 - o(\sigma_K). \qquad \text{(II.7.31)}$$

Since $\phi(x) - \phi(y) = O(h(x-y)\sigma_K)$, we have with probability $1 - o(\sigma_K)$ either

$$\inf\Big\{t \geq \theta_{\text{mut. size } i(\epsilon/2)}^K : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \phi(i\tfrac{\epsilon}{2})| < (M/3)\epsilon\sigma_K\Big\} = \theta_{\text{mut. size } i(\epsilon/2)}^K, \qquad \text{(II.7.32)}$$

which implies (a) for i, or at least

$$\left|\langle\tilde{\nu}_{\theta^K_{\text{mut. size }i(\epsilon/2)}},\mathbb{1}\rangle-\phi(i\tfrac{\epsilon}{2})\right|<\left(M+\left|h\left(\frac{r'(R^K)}{r(R^K)}-\frac{(\partial_1 c(R^K,R^K)+\partial_2 c(R^K,R^K))}{c(R^K,R^K)}\right)\right|\right)\epsilon\sigma_K. \tag{II.7.33}$$

Similarly as in many previous lemmata we want to couple $K\langle\tilde{\nu}_t,\mathbb{1}\rangle$ with a discrete time Markov chain. Therefore, let

$$X^i_t=|K\langle\tilde{\nu}_t,\mathbb{1}\rangle-\lceil\phi(i\tfrac{\epsilon}{2})K\rceil|, \tag{II.7.34}$$

and $T^i_0=\theta^K_{\text{mut. size }i(\epsilon/2)}$ and $(T^i_k)_{k\geq1}$ be the sequences of the jump times of $\langle\tilde{\nu}_t,\mathbb{1}\rangle$ after $\theta^K_{\text{mut. size }i(\epsilon/2)}$. Then let Y^i_k be the associated discrete time process which records the values that X^i_t takes after time $\theta^K_{\text{mut. size }i(\epsilon/2)}$.

CLAIM. *There exists a constant, $C^{b,d,c}_{derivative}>0$, such that for all $\lceil C^{b,d,c}_{derivative}\epsilon\sigma_K K\rceil\leq j<\lceil\epsilon K\rceil$ and K large enough,*

$$\mathbb{P}\left[Y^i_{n+1}=j+1|Y^i_n=j,T_{n+1}<\tilde{\theta}^K_i\right]\leq\tfrac{1}{2}-\epsilon\sigma_K=:p^K_+. \tag{II.7.35}$$

Moreover, we can choose

$$C^{b,d,c}_{derivative}=\sup_{x\in\mathcal{X}}\frac{1}{c(x,x)}\left(4b(x)+A\left|\frac{r'(x)c(x,x)}{r(x)}-\partial_1 c(x,x)-\partial_2 c(x,x)\right|\right). \tag{II.7.36}$$

If $\langle\tilde{\nu}_t,\mathbb{1}\rangle K>\lceil\phi(i(\epsilon/2))K\rceil$ at time $t=T^i_n$, then $\langle\tilde{\nu}_{T^i_n},\mathbb{1}\rangle K=\lceil\phi(i(\epsilon/2))K\rceil+Y^i_n$ and, conditionally on $\mathcal{F}_{T^i_n}$, the left hand side of (II.7.35) is equal to the probability that the next event is a birth. Namely,

$$\frac{\sum_{k\geq0}b(h_{k,1}(\tilde{\nu}_{T^i_n}))\mathfrak{M}^k(\tilde{\nu}_{T^i_n})}{\sum_{k\geq0}\left(b(h_{k,1}(\tilde{\nu}_{T^i_n}))+d(h_{k,1}(\tilde{\nu}_{T^i_n}))+\int_{\mathbb{N}\times\mathcal{X}}c(h_{k,1}(\tilde{\nu}_{T^i_n}),\xi)d\tilde{\nu}_{T^i_n}(\xi)\right)\mathfrak{M}^k(\tilde{\nu}_{T^i_n})} \tag{II.7.37}$$

$$\leq\left(b(R^K)\sum_{k\geq0}\mathfrak{M}^k(\tilde{\nu}_{T^i_n})+\sigma_K hb'(R^K)\mathfrak{M}^{k_1}(\tilde{\nu}_{T^i_n})+C^b_L 2A\sigma_K\lceil\tfrac{3}{\alpha}\rceil\sigma_K\epsilon K+O(\sigma^2_K K)\right)$$

$$\times\left(\sum_{k\geq0}\left(b(R^K)+d(R^K)+\sum_{k\geq0}\frac{c(R^K,R^K)}{K}\mathfrak{M}^k(\tilde{\nu}_{T^i_n})\right)\mathfrak{M}^k(\tilde{\nu}_{T^i_n})\right.$$

$$+\sigma_K h\mathfrak{M}^{k_1}(\tilde{\nu}_{T^i_n})\left(b'(R^K)+d'(R^K)+\frac{\partial_1 c(R^K,R^K)+\partial_2 c(R^K,R^K)}{K}\left(\mathfrak{M}^0(\tilde{\nu}_{T^i_n})+\mathfrak{M}^{k_1}(\tilde{\nu}_{T^i_n})\right)\right)$$

$$\left.-(C^{b,d,c}_L)2A\sigma_K\lceil3/\alpha\rceil\sigma_K\epsilon K-O(\sigma^2_K K)\right)^{-1}.$$

For the inequality we have used the fact that, conditioned on $T_n<\tilde{\theta}^K_i$, there at most $\sigma_K\epsilon\lceil3/\alpha\rceil$ many unsuccessful mutant individuals which differ at most $2A\sigma_K$ from the resident trait R^K. Since $\sum_{k\geq0}\mathfrak{M}^k(\tilde{\nu}_{T^i_n})=\langle\tilde{\nu}_{T^i_n},\mathbb{1}\rangle K$ which equals $\lceil\phi(i(\epsilon/2))K\rceil+j$ conditioned on $j=Y^i_n$, the right hand side of the last inequality is smaller or equals

$$\left(b(R^K)+\sigma_K hb'(R^K)\frac{\mathfrak{M}^{k_1}(\tilde{\nu}_{T^i_n})}{\lceil\phi(i(\epsilon/2))K\rceil+j}+O(\sigma^2_K)\right) \tag{II.7.38}$$

$$\times\left(b(R^K)+d(R^K)+c(R^K,R^K)\frac{\lceil\phi(i(\epsilon/2))K\rceil+j}{K}+\sigma_K\frac{h\mathfrak{M}^{k_1}(\tilde{\nu}_{T^i_n})}{\lceil\phi(i(\epsilon/2))K\rceil+j}\right.$$

$$\times\left.\left(b'(R^K)+d'(R^K)+\frac{\partial_1 c(R^K,R^K)+\partial_2 c(R^K,R^K)}{K}\left(\mathfrak{M}^0(\tilde{\nu}_{T^i_n})+\mathfrak{M}^{k_1}(\tilde{\nu}_{T^i_n})\right)\right)-O(\sigma^2_K)\right)^{-1}$$

and by definition of ϕ the denominator equals

$$2b(R^K)\sigma_K + 2\sigma_K h b'(R^K)\frac{\mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})}{\lceil \phi(i(\epsilon/2))K \rceil + j} + c(R^K, R^K)\frac{j}{K} - O(\sigma_K^2) \tag{II.7.39}$$

$$+ \sigma_K h \Big[i\frac{\epsilon}{2}\Big(\frac{r'(R^K)}{\overline{z}(R^K)} + \partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K)\Big) + \frac{\mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})}{\lceil \phi(i(\epsilon/2))K \rceil + j}$$

$$\times \Big(d'(R^K) - b'(R^K) + \frac{\partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K)}{K}\big(\mathfrak{M}^0(\tilde{\nu}_{T_n^i}) + \mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})\big)\Big)\Big].$$

Thus, we obtain that the right hand side of (II.7.37) is bounded from above by

$$\frac{1}{2} - \frac{c(R^K, R^K)}{3b(R^K)}jK^{-1} - \frac{\sigma_K h}{4b(R^K)}\Big[i\frac{\epsilon}{2}\Big(\frac{r'(R^K)}{\overline{z}(R^K)} - \partial_1 c(R^K, R^K) - \partial_2 c(R^K, R^K)\Big) \tag{II.7.40}$$

$$+ \frac{\mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})}{\lceil \phi(i(\epsilon/2))K \rceil + j}\Big(-r'(R^K) + \frac{\partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K)}{K}\big(\mathfrak{M}^0(\tilde{\nu}_{T_n^i}) + \mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})\big)\Big)\Big] + O(\sigma_K^2).$$

In the case where $\langle \tilde{\nu}_t, \mathbb{1}\rangle K < \lceil \phi(i(\epsilon/2))K \rceil$ at time $t = T_n^i$, we obtain the same inequality but with an opposite sign in front of the third term. Since

$$\Big| \frac{i\epsilon}{2}\frac{r'(R^K)}{\overline{z}(R^K)} - \frac{\mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})}{K}\frac{r'(R^K)K}{\lceil \phi((\epsilon/2))K \rceil \pm j} - \big(\partial_1 c(R^K, R^K) + \partial_2 c(R^K, R^K)\big)\Big(\frac{i\epsilon}{2} - \frac{\mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})}{K}\Big)\Big| \tag{II.7.41}$$

$$< (\epsilon/2)\Big|\frac{r'(R^K)}{\overline{z}(R^K)} - \partial_1 c(R^K, R^K) - \partial_2 c(R^K, R^K)\Big|,$$

we deduce the claim. Since we choose M such that $M \geq 3C_{\text{derivative}}^{b,d,c}$, we can construct a Markov chain Z_n^i such that $Z_n^i \geq Y_n^i$, a.s., for all n such that $T_n^i < \tilde{\theta}_i^K \wedge \inf\{t \geq \theta_{\text{mut. size } i(\epsilon/2)}^K : |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \phi(i(\epsilon/2))| < \frac{1}{3}M\epsilon\sigma_K\}$ and the marginal distribution of Z_n is a Markov chain with $Z_0^i = Y_0^i$ and transition probabilities

$$\mathbb{P}\big[Z_{n+1}^i = j_2 | Z_n^i = j_1\big] = \begin{cases} p_+^K & \text{for } j_1 \geq 1 \text{ and } j_2 = j_1 + 1, \\ 1 - p_+^K & \text{for } j_1 \geq 1 \text{ and } j_2 = j_1 + 1, \\ 0 & \text{else.} \end{cases} \tag{II.7.42}$$

Let $C_{\text{exit}} = \sup_{x \in \mathcal{X}} 2A\Big|\frac{r'(x)}{r(x)} - \frac{(\partial_1 c(x,x) + \partial_2 c(x,x))}{c(x,x)}\Big|$. Then, by applying Proposition II.9.5 (b), we obtain, for all $a \leq (M + C_{\text{exit}})\epsilon\sigma_K K$ and K large enough,

$$\mathbb{P}_a\Big[\inf\{n \geq 0 : Z_n^i \geq 2(M + C_{\text{exit}})\epsilon\sigma_K K\} < \inf\{n \geq 0 : Z_n^i \leq (\tfrac{M}{3})\epsilon\sigma_K K\}\Big] \tag{II.7.43}$$

$$\leq \exp\big(-K^\alpha\big).$$

Next define $B^i \equiv \inf\{n \geq 0 : Z_n^i \leq \frac{1}{3}M\epsilon\sigma_K K\}$. This is the random variable, which counts the number of jumps Z^i makes until it is smaller than $\epsilon\sigma_K K$. Note that $(T_{n+1}^i - T_n^i)$, the times between two jumps of X_t^i, are exponential distributed with a parameter $(b(R^K) + d(R^K) + c(R^K, R^K)\overline{z}(R^K))\overline{z}(R^K)K + O(\sigma_K K)$ if T_{n+1}^i is smaller than $\tilde{\theta}_i^K$. Thus,

$$(T_{l+1}^i - T_l^i) \preccurlyeq E_l^i, \tag{II.7.44}$$

where $(E_l^i)_{l \geq 0}$ is a sequence of independent exponential random variables with parameter $\inf_{x \in \mathcal{X}} b(x)\overline{z}(x)K$. Therefore,

$$\mathbb{P}\Big[\theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K > \theta_{\text{mut. size } i(\epsilon/2)}^K + S_K \wedge \tilde{\theta}_i^K\Big] \tag{II.7.45}$$

$$\leq \mathbb{P}\Big[\sum_{l=0}^{B^i} E_l^i > S_K\Big] + \mathbb{P}\Big[\tilde{\theta}_i^K < \theta_{\text{mut. size } i(\epsilon/2)}^K + S_K \wedge \theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K\Big].$$

Our next goal is to find a number, n_i, such that $\mathbb{P}[B^i > n_i]$ is $o(\sigma_K)$. Since the transition probabilities of Z^i do not depend on the present state, we have that $Z_n^i - Z_0^i$ has the same law as $\sum_{k=1}^n V_k^i$, where $(V_k^i)_{k \in \mathbb{N}}$ is a sequence of i.i.d. random variables with

$$\mathbb{P}[V_k^i = 1] = p_+^K \quad \text{and} \quad \mathbb{P}[V_k^i = -1] = 1 - p_+^K \tag{II.7.46}$$

and $\mathbb{E}[V_k^i] = -2\epsilon\sigma_K$ and $|V_k^i| = 1$. Furthermore, we get

$$\mathbb{P}[B^i \le n_i] \ge \mathbb{P}[\inf\{j \ge 0 : Z_j - Z_0 \le -\lceil(\tfrac{3}{2}M + C_{\text{exit}})\epsilon\sigma_K K\rceil\} \le n_i] \tag{II.7.47}$$
$$\ge \mathbb{P}[\textstyle\sum_{k=1}^{n_i} V_k^i \le -\lceil(\tfrac{3}{2}M + C_{\text{exit}})\epsilon\sigma_K K\rceil]$$

and by applying the

HOEFFDING'S INEQUALITY. (Appendix 2 in [110])   *Let* $Y_1, \ldots, Y_n$ *be independent random variables such that, for all* $j \in \mathbb{N}$, $a_j \le Y_j - \mathbb{E}[Y_j] \le b_j$ *for some real constants* a_j, b_j. *Then, for* $x > 0$,

$$\mathbb{P}[\textstyle\sum_{j=1}^n Y_j - \mathbb{E}[Y_j] \ge x] \le \exp\left(-2x^2\left(\textstyle\sum_{j=1}^n (a_j - b_j)^2\right)^{-1}\right). \tag{II.7.48}$$

we obtain

$$\mathbb{P}[\textstyle\sum_{k=1}^{n_i} V_k^i \ge -2\epsilon\sigma_K n_i + (n_i)^{1/2+\alpha/2}] \le 2\exp(-(n_i)^\alpha). \tag{II.7.49}$$

With $n_i \equiv \lceil K(\tfrac{3}{2}M + C_{\text{exit}})\rceil$, we get $-2\epsilon\sigma_K n_i + (n_i)^{1/2+\alpha/2} \le -\lceil(\tfrac{3}{2}M + C_{\text{exit}})\epsilon\sigma_K K\rceil$, since $K^{-\frac{1}{2}+\alpha} \ll \sigma_K$. Applying the exponential Chebyshev's inequality (with $\lambda = K^\alpha$)

$$\mathbb{P}\left[\textstyle\sum_{l=0}^{\lceil K(\frac{3}{2}M+C_{\text{exit}})\rceil} E_l^i > S_K\right] \le \exp(-\lambda S_K)\mathbb{E}\left[\exp\left(\lambda \textstyle\sum_{l=0}^{\lceil K(\frac{3}{2}M+C_{\text{exit}})\rceil} E_l^i\right)\right] \tag{II.7.50}$$

$$\le \exp(-\lambda S_K)\left(\frac{\inf_{x\in\mathcal{X}} b(x)\bar{z}(x)K}{\inf_{x\in\mathcal{X}} b(x)\bar{z}(x)K - \lambda}\right)^{\lceil K(\frac{3}{2}M+C_{\text{exit}})\rceil+1}$$

$$\le \exp\left(-\lambda S_K + (\lceil K(\tfrac{3}{2}M + C_{\text{exit}})\rceil + 1)\ln\left(1 + \tfrac{\lambda}{\inf_{x\in\mathcal{X}} b(x)\bar{z}(x)K-\lambda}\right)\right)$$

$$\le \exp\left(-\lambda S_K + \lambda\frac{\tfrac{3}{2}M + C_{\text{exit}} + 1}{\inf_{x\in\mathcal{X}} b(x)\bar{z}(x)} + O(\lambda^2 K^{-1})\right) \le \exp\left(-K^\alpha\right).$$

Hence, the left hand side of (II.7.45) is bounded from above by

$$\exp\left(-K^\alpha\right) + 2\exp(-(K(\tfrac{3}{2}M + C_{\text{exit}}))^\alpha) + \mathbb{P}\left[\tilde{\theta}_i^K < (\theta_{\text{mut. size } i(\epsilon/2)}^K + S_K) \wedge \theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K\right]. \tag{II.7.51}$$

This proves the lemma if we can show that

$$\mathbb{P}\left[\tilde{\theta}_i^K < (\theta_{\text{mut. size } i(\epsilon/2)}^K + S_K) \wedge \theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K\right] = o(\sigma_K). \tag{II.7.52}$$

According to Remark 6 and Lemma II.7.3, we have that

$$\mathbb{P}\left[\theta_{2 \text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K < \theta_{\text{mut. size } i(\epsilon/2)}^K + S_K\right] = o(\sigma_K). \tag{II.7.53}$$

Therefore, the following proof of (b) for i implies (a) for i. $\square$

Proof. of (b) for i by assuming that the lemma holds for $i-1$. Note that the random elements B^i, T^i, V^i, W^i, X^i, Y^i, and Z^i are not the ones of the last proof. They will be defined during this proof. In fact, the structure of the proof is similar to the one of (a), except that

we prove a lower bound for the time of a change of order ϵ for the mutant density instead of upper bound for the time of a change of order $\epsilon\sigma_K$ of the total mass. We couple $\mathfrak{M}_t^{k_1}$, for $t \geq \theta_{\text{mut. size } i(\epsilon/2)}^K$, with a discrete time Markov chain (depending on i). Therefore, let $T_0^i = \theta_{\text{mut. size } i(\epsilon/2)}^K$ and $(T_k^i)_{k\geq 1}$ be the sequences of jump times of $\mathfrak{M}_t^{k_1}$ after $\theta_{\text{mut. size } i(\epsilon/2)}^K$. Furthermore, let $(Y_n^i)_{n\geq 0}$ be the discrete time process which records the values that $\mathfrak{M}_t^{k_1}$ takes i.e., $Y_0^i = \mathfrak{M}^{k_1}(\tilde{\nu}_{T_0^i}) = \lceil Ki(\epsilon/2)\rceil$ and $Y_n^i = \mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})$. Observe that if

$$\tilde{\theta}_i^K > \theta_{\text{near }\phi(i\frac{\epsilon}{2})}^K \wedge \inf\left\{t \geq \theta_{\text{mut. size } i(\epsilon/2)}^K : |\langle\tilde{\nu}_t, \mathbb{1}\rangle - \phi(i\tfrac{\epsilon}{2})| \geq 2(M + C_{\text{exit}})\epsilon\sigma_K K\right\}, \quad \text{(II.7.54)}$$

we know from the inequality (II.7.43) that the probability that $\theta_{\text{near }\phi(i\frac{\epsilon}{2})}^K$ is larger than $\inf\{t \geq \theta_{\text{mut. size } i(\epsilon/2)}^K : |\langle\tilde{\nu}_t, \mathbb{1}\rangle - \phi(i(\epsilon/2))| \geq 2(M + C_{\text{exit}})\epsilon\sigma_K K\}$ is smaller than $\exp(-K^\alpha)$. Define

$$\hat{\theta}_i^K \equiv \inf\left\{t \geq \theta_{\text{mut. size } i(\epsilon/2)}^K : |\langle\tilde{\nu}_t, \mathbb{1}\rangle - \phi(i\tfrac{\epsilon}{2})| \geq 2(M + C_{\text{exit}})\epsilon\sigma_K K\right\} \quad \text{(II.7.55)}$$
$$\wedge\, \theta_{\text{near }\phi(i\frac{\epsilon}{2})}^K \wedge \theta_{2\text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K$$

and

$$\tilde{C}_{\text{fitness}} \equiv \inf_{x\in\mathcal{X}} \partial_1 f(x,x)/\overline{b}. \quad \text{(II.7.56)}$$

Note that $\hat{\theta}_i^K \neq \hat{\theta}^K$. Then, for all $-\lceil\frac{\epsilon}{4}K\rceil \leq j \leq \lceil\frac{\epsilon}{4}K\rceil$, for K large enough and for ϵ small enough, we have that

$$\mathbb{P}\left[Y_{n+1}^i = \lceil i\tfrac{\epsilon}{2}K\rceil + j + 1 \big| Y_n^i = \lceil i\tfrac{\epsilon}{2}K\rceil + j, T_{n+1}^i < \hat{\theta}_i^K\right] \quad \text{(II.7.57)}$$
$$\in \left[\tfrac{1}{2} + \tfrac{1}{2}\tilde{C}_{\text{fitness}}\,\sigma_K,\ \tfrac{1}{2} + 2A\tilde{C}_{\text{fitness}}\,\sigma_K\right],$$

since the left hand side of (II.7.57) is equal to the expectation of the probability that the next event is a birth without mutation conditioned on $\mathcal{F}_{T_n^i}$. Namely,

$$\frac{b(R^K + \sigma_K h)(1 - u_K m(R^K - \sigma_K h))}{\left(b(R^K + \sigma_K h) + d(R^K + \sigma_K h) + \int_{\mathbb{N}\times\mathcal{X}} c(R^K + \sigma_K h, \xi)d\tilde{\nu}_{T_n}(\xi)\right)} \quad \text{(II.7.58)}$$

$$= b(R^K + \sigma_K h)\Big[b(R^K + \sigma_K h) + d(R^K + \sigma_K h) + c(R^K + \sigma_K h, R^K)\Big(\phi(i\tfrac{\epsilon}{2}) - \tfrac{\lceil i\frac{\epsilon}{2}K\rceil + j}{K}\Big)$$
$$+ c(R^K + \sigma_K h, R^K + h\sigma_K)\Big(\tfrac{\lceil i\frac{\epsilon}{2}K\rceil + j}{K}\Big) + \xi_1\big(\epsilon\sigma_K C_L^c(\lceil\tfrac{3}{\alpha}\rceil + 2(M + C_{\text{exit}}))\big)\Big]^{-1} + O(u_K)$$

$$= b(R^K + \sigma_K h)\Big[2b(R^K + \sigma_K h) - f(R^K + \sigma_K h, R^K) + c(R^K + \sigma_K h, R^K)\Big(\phi(i\tfrac{\epsilon}{2}) - \tfrac{r(R^K)}{c(R^K, R^K)}\Big)$$
$$+ \sigma_K h \partial_2 c(R^K, R^K)\Big(\tfrac{\lceil i\frac{\epsilon}{2}K\rceil + j}{K}\Big) + \xi_1\big(\epsilon\sigma_K C_L^c(\lceil\tfrac{3}{\alpha}\rceil + 2(M + C_{\text{exit}}))\big)\Big]^{-1} + O(u_K).$$

for some $\xi_1 \in (-1, 1)$. By definition of ϕ of (II.7.58) is equal to

$$b(R^K + \sigma_K h)\Big[2b(R^K + \sigma_K h) - \partial_1 f(R^K, R^K)\sigma_K h + c(R^K + \sigma_K h, R^K)\sigma_K h \quad \text{(II.7.59)}$$
$$\times (i\tfrac{\epsilon}{2})\Big(\tfrac{r'(R^K)}{r(R^K)} - \tfrac{\partial_1 c(R^K, R^K)}{c(R^K, R^K)}\Big) + \xi_1\big(\epsilon\sigma_K C_L^c(\lceil\tfrac{3}{\alpha}\rceil + 2(M + C_{\text{exit}}))\big)\Big]^{-1}$$
$$+ O\big(\tfrac{\sigma_K j}{K} + \sigma_K^2 + u_K\big)$$

$$= b(R^K + \sigma_K h)\Big[2b(R^K + \sigma_K h) - \sigma_K h\Big(1 - i\tfrac{\epsilon}{2}\tfrac{c(R^K, R^K)}{r(R^K)}\Big)\partial_1 f(R^K, R^K)$$
$$+ \xi_1\big(\epsilon\sigma_K C_L^c(\lceil\tfrac{3}{\alpha}\rceil + 2(M + C_{\text{exit}}))\big)\Big]^{-1} + O\big(\tfrac{\sigma_K j}{K} + \sigma_K^2 + u_K\big)$$

$$= \tfrac{1}{2} + \sigma_K h\Big(1 - i\tfrac{\epsilon}{2}\tfrac{c(R^K, R^K)}{r(R^K)}\Big)\tfrac{\partial_1 f(R^K, R^K)}{b(R^K)} + \epsilon\sigma_K \xi_1 \tfrac{(C_L^c(\lceil\frac{3}{\alpha}\rceil + 2(M + C_{\text{exit}}))}{b(R^K)} + O\big(\tfrac{\sigma_K j}{K} + \sigma_K^2 + u_K\big).$$

Then, because $i < 2\epsilon^{-1}C_{\text{cross}}^{\epsilon}$ implies that $1 - i\frac{\epsilon}{2}\frac{c(R^K, R^K)}{r(R^K)} > 0$, we obtain (II.7.57). Thus we can construct a Markov chain Z_n^i such that $Z_n^i \geq Y_n^i$, a.s., for all n such that $T_n^i < \hat{\theta}^K$ and such that the marginal distribution of Z_n^i is a Markov chain with transition probabilities

$$\mathbb{P}\big[Z_{n+1}^i = j_2 | Z_n^i = j_1\big] = \begin{cases} \frac{1}{2} + 2A\tilde{C}_{\text{fitness}}\sigma_K & \text{for } j_2 = j_1 + 1, \\ \frac{1}{2} - 2A\tilde{C}_{\text{fitness}}\sigma_K & \text{for } j_2 = j_1 - 1, \\ 0 & \text{else.} \end{cases} \qquad \text{(II.7.60)}$$

We define a continuous time process, $\tilde{Z}^i$, associate to Z_n^i. To do this, we define first $(\tilde{T}_j^i)_{j\in\mathbb{N}}$, the sequence of jump times, by $\tilde{T}_0^i = 0$ and

$$\tilde{T}_j^i - \tilde{T}_{j-1}^i = \begin{cases} T_j^i - T_{j-1}^i & \text{if } T_j^i < \tilde{\theta}^K, \\ W_j^i & \text{else,} \end{cases} \qquad \text{(II.7.61)}$$

where W_j^i are exponential random variables with mean $(\lceil K(i+\frac{1}{2})(\epsilon/2)\rceil(\bar{b}+\bar{d}+\bar{c}(4\bar{b}/\underline{c})))^{-1}$. We set $\tilde{Z}_t^i = Z_n^i$ if $t \in [\tilde{T}_n^i, \tilde{T}_{n+1}^i)$. Obverse that we obtain by construction $\tilde{Z}_t^i \geq \mathfrak{M}^{k_1}(\tilde{\nu}_{\theta_{\text{mut. size } i(\epsilon/2)}^K + t})$, for all t such that $\theta_{\text{mut. size } i(\epsilon/2)}^K + t \leq \hat{\theta}_i^K$. Next we want to show that

$$\mathbb{P}\Big[\inf\big\{t \geq 0 : \tilde{Z}_t^i \geq \lceil K(i+\tfrac{1}{2})(\epsilon/2)\rceil\big\} > S_K\Big] = 1 - o(\sigma_K). \qquad \text{(II.7.62)}$$

Therefore, let $B_i^Z = \inf\{n \geq 0 : Z_n^i = \lceil K(i+\frac{1}{2})(\epsilon/2)\rceil\}$. We can construct $(X_j^i)_{j\geq 1}$ a sequence of independent, exponential random variables with parameter $x_i^K \equiv \lceil K(i+\frac{1}{2})(\epsilon/2)\rceil(\bar{b}+\bar{d}+\bar{c}(4\bar{b}/\underline{c}))$ such that

$$(\tilde{T}_{j+1}^i - \tilde{T}_j^i) \succcurlyeq X_j^i \quad \text{for all } 1 \leq j \leq B_i^Z. \qquad \text{(II.7.63)}$$

Our next goal is to find a barrier, n_i, such that B_i^Z is smaller than n_i only with very small probability. Since the transition probabilities of Z^i do not depend on the present state, $Z_{B_i^Z}^i - Z_0$ is stochastically equivalent to $\sum_{k=1}^j V_k^i$, where $(V_k^i)_{k\in\mathbb{N}}$ are i.i.d. random variables taking values ± 1 with probabilities

$$\mathbb{P}[V_k^i = 1] = \tfrac{1}{2} + 2A\tilde{C}_{\text{fitness}}\sigma_K \quad \text{and} \quad \mathbb{P}[V_k^i = -1] = \tfrac{1}{2} - 2A\tilde{C}_{\text{fitness}}\sigma_K. \qquad \text{(II.7.64)}$$

Note that $\mathbb{E}[V_k^i] = 4A\tilde{C}_{\text{fitness}}\sigma_K$ and $|V_k^i| = 1$. Furthermore, we get

$$\mathbb{P}\big[B_i^Z \leq n_i\big] = \mathbb{P}\Big[\exists \lceil(\epsilon/4)K\rceil \leq j \leq n_i : \textstyle\sum_{k=1}^j V_k^i \geq \lceil(\epsilon/4)K\rceil\Big]. \qquad \text{(II.7.65)}$$

Hoeffding's inequality implies that, for $j \geq \lceil(\epsilon/4)K\rceil$,

$$\mathbb{P}\Big[\textstyle\sum_{k=1}^j V_k^i \geq 4A\tilde{C}_{\text{fitness}}\sigma_K j + j^{1/2+\alpha/2}\Big] \leq 2\exp(-j^\alpha). \qquad \text{(II.7.66)}$$

We take $n_i \equiv \epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}$ and get for all $\lceil(\epsilon/4)K\rceil \leq j \leq n_i$,

$$4A\tilde{C}_{\text{fitness}}\sigma_K j + j^{1/2+\alpha/2} \leq \lceil(\epsilon/4)K\rceil, \qquad \text{(II.7.67)}$$

since $K^{-\frac{1}{2}+\alpha} \ll \sigma_K$. Then, the probability that $B_i^Z \leq \epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}$ is bounded from above by $2\exp(-K^\alpha)$. Therefore, the left hand side of equation (II.7.62) is larger than

$$\mathbb{P}\Big[\textstyle\sum_{j=1}^{\epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}} X_j^i > S_K\Big] - 2\exp(-K^\alpha), \qquad \text{(II.7.68)}$$

By applying the exponential Chebyshev's inequality we get, similarly as in (a),

$$\mathbb{P}\left[\sum_{j=1}^{\epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}} X_j^i \le S_K\right] \tag{II.7.69}$$

$$= \mathbb{P}\left[-\sum_{j=1}^{\epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}} X_j^i \ge -S_K\right]$$

$$\le \exp(K^{\alpha}S_K)\mathbb{E}\left[\exp(-K^{\alpha}X_j^i)\right]^{\epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}}$$

$$\le \exp(K^{\alpha}S_K)\exp\left(\epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}\ln\left(\frac{x_i^K}{x_i^{K+K^{\alpha}}}\right)\right)$$

$$\le \exp(K^{\alpha}S_K - \epsilon K(8A\tilde{C}_{\text{fitness}}\sigma_K)^{-1}CK^{-1+\alpha})), \quad \text{for some small } C > 0,$$

$$\le \exp\left(-K^{\alpha}\right).$$

This proves that $\mathbb{P}\left[\inf\{t \ge 0 : \tilde{Z}_t^i \ge \lceil K(i+\frac{1}{2})(\epsilon/2)\rceil\} > S_K\right] \ge 1 - 3\exp\left(-K^{\alpha}\right)$, and therefore (b) and (a) for i, provided that the lemma holds for $i-1$. $\qquad\square$

Proof. of (c) for i by assuming that the lemma holds for $i-1$. Note that the random elements T^i, X^i, and Y^i are not the ones of the last proof. As in (a) we couple $K\langle\tilde{\nu}_t, \mathbb{1}\rangle$ with a discrete time Markov chain. Therefore, let

$$X_t^i = |K\langle\tilde{\nu}_t, \mathbb{1}\rangle - \lceil\phi(i(\epsilon/2))K\rceil| \tag{II.7.70}$$

and $T_0^i = \theta_{\text{mut. size } i(\epsilon/2)}^K$ and $(T_k^i)_{k\ge1}$ be the sequences of the jump times of $\langle\tilde{\nu}_t, \mathbb{1}\rangle$ after $\theta_{\text{mut. size } i(\epsilon/2)}^K$. Then, let Y_k^i be the associated discrete time process which records the values that X_t^i takes after time $\theta_{\text{mut. size } i(\epsilon/2)}^K$.

CLAIM. *There exists a constant $\tilde{C}_{\text{derivative}}^{b,d,c}$ such that for all $j < \lceil\epsilon K\rceil$ and K large enough,*

$$\mathbb{P}\left[Y_{n+1}^i = j+1 \big| Y_n^i = j, T_{n+1} < \tilde{\theta}_i^K\right] \le \frac{1}{2} - \frac{c}{3\bar{b}}jK^{-1} + \epsilon\sigma_K\tilde{C}_{\text{derivative}}^{b,d,c} \equiv p_+^K(j), \tag{II.7.71}$$

Moreover, we can choose $\tilde{C}_{\text{derivative}}^{b,d,c} \equiv \sup_{x\in\mathcal{X}}\frac{A}{4b(x)}|\frac{r'(x)}{\bar{z}(x)} - \partial_1 c(x,x) - \partial_2 c(x,x)|$.

From (a) we know that the left hand side of (II.7.71) is smaller or equals

$$\frac{1}{2} - \frac{c(R^K, R^K)}{3b(R^K)}jK^{-1} + \frac{\epsilon\sigma_K h}{8b(R^K)}\left|\frac{r'(R^K)}{\bar{z}(R^K)} - \partial_1 c(R^K, R^K) - \partial_2 c(R^K, R^K)\right| + O(\sigma_K^2). \tag{II.7.72}$$

This proves the Claim. Note that $p_+^K(j)$ depends on j. Since we can choose $M \ge 8\tilde{C}_{\text{derivative}}^{b,d,c}\frac{3\bar{b}}{c}$, continuing as in Lemma II.6.3 implies that (c) is true for i, provided that the lemma holds for $i-1$. $\qquad\square$

Proof. of (d) for i by assuming that the lemma holds for $i-1$. Again we couple $\mathfrak{M}_t^{k_1}$, for $t \ge \theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K$, with a discrete time Markov chain. Let $T_0^i = \theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K$ and $(T_k^i)_{k\ge1}$ be the sequences of the jump times of $\mathfrak{M}_t^{k_1}$ after $\theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K$. Then, let $(Y_n^i)_{n\ge0}$ be the discrete time process which records the values that $\mathfrak{M}_t^{k_1}$, i.e.,

$$Y_0^i = \mathfrak{M}^{k_1}(\tilde{\nu}_{T_0^i}) \in \left[K(\tfrac{i\epsilon}{2} - \tfrac{\epsilon}{4}) - 1, K(\tfrac{i\epsilon}{2} + \tfrac{\epsilon}{4}) + 1\right], \tag{II.7.73}$$

and $Y_n^i = \mathfrak{M}^{k_1}(\tilde{\nu}_{T_n^i})$. Define

$$\hat{\theta}_i^K \equiv \inf\left\{t \ge \theta_{\text{near } \phi(i\frac{\epsilon}{2})}^K : |\langle\tilde{\nu}_t, \mathbb{1}\rangle - \phi(i\tfrac{\epsilon}{2})| > M\epsilon\sigma_K\right\} \wedge \theta_{2\text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K. \tag{II.7.74}$$

Note that this $\hat{\theta}_i^K$ differs only a bit from the one defined in (b). From the proof of (b), we know that the density of the mutant trait has the tendency to increase. More precisely, since $i \leq C_{\text{cross}}^\epsilon(2/\epsilon)$, we have, for all $-\lceil \frac{\epsilon}{4}K \rceil \leq j \leq \lceil \frac{\epsilon}{2}K \rceil$, for K large enough and ϵ small enough,

$$\mathbb{P}\Big[Y_{n+1}^i = \lceil i\frac{\epsilon}{2}K \rceil + j + 1 \Big| Y_n^i = \lceil i\frac{\epsilon}{2}K \rceil + j, T_{n+1}^i < \hat{\theta}_i^K \Big] \geq \frac{1}{2} + \sigma_K \frac{\inf_{x \in \chi} \partial_1 f(x,x)}{2\bar{b}} \qquad (\text{II.7.75})$$

By Continuing in a similar way as in (b) with bounding the random variables in the in the other direction (as in (a)), implies that (d) is true for i, provided that the lemma holds for $i - 1$. $\qquad\qquad\square$

II.7.3 Step 3: the density of the resident trait R^K decreases to ϵ

Similarly as in Step 2 we define a function which allows us to approximate the total mass of the population for a given density of the resident trait.

Notation. Let us define

$$\psi(x) \equiv \overline{z}(R^K) + \sigma_K h(\overline{z}(R^K) - x)\left(\frac{r'(R^K)}{r(R^K)} + \frac{\partial_1 c(R^K,R^K) + \partial_2 c(R^K,R^K)}{c(R^K,R^K)}\right). \qquad (\text{II.7.76})$$

Note that $\phi(y) = \psi(\phi(y) - y) + O(\sigma_K^2)$. Therefore, since $|\langle \tilde{\nu}_{\theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K}, \mathbb{1}\rangle - \phi(C_{\text{cross}}^\epsilon)| < M\epsilon\sigma_K$ with probability $1 - o(\sigma_K)$, we get that at time $\theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K$ the density of the resident population belongs to an interval centered at $\phi(C_{\text{cross}}^\epsilon) - C_{\text{cross}}^\epsilon$ with diameter $2(M + \lceil 3/\alpha \rceil)\epsilon\sigma_K$ with probability $1 - o(\sigma_K)$, and hence

$$\psi\Big(\mathfrak{M}^0\Big(\tilde{\nu}_{\theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K}\Big)K^{-1}\Big) = \psi(\phi(C_{\text{cross}}^\epsilon) - C_{\text{cross}}^\epsilon) + O(\epsilon\sigma_K^2) \qquad (\text{II.7.77})$$

$$= \phi(C_{\text{cross}}^\epsilon) + O(\sigma_K^2)$$

with probability $1 - o(\sigma_K)$. Thus, the total mass of the population also belongs to an interval centered at $\psi(\phi(C_{\text{cross}}^\epsilon) - C_{\text{cross}}^\epsilon)$ with diameter $2(M\epsilon\sigma_K + O(\sigma_K^2)) < 2(M + 1)\epsilon\sigma_K$.

Notation. Let us define

$$\tilde{C}_{\text{cross}}^K \equiv \lceil (\phi(C_{\text{cross}}^\epsilon) - C_{\text{cross}}^\epsilon - \epsilon)2/\epsilon \rceil (\epsilon/2) \qquad\qquad \text{and} \qquad\qquad (\text{II.7.78})$$

$$\theta_{\text{near } \psi(\tilde{C}_{\text{cross}}^\epsilon - \frac{\epsilon}{2})} \equiv \inf\Big\{t \geq \theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K : |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \psi(\tilde{C}_{\text{cross}}^\epsilon - \frac{\epsilon}{2})| < (M/3)\epsilon\sigma_K\Big\}. \qquad (\text{II.7.79})$$

Note that the term $-\epsilon$ in the definition of $\tilde{C}_{\text{cross}}^K$ ensures that resident population is larger than $\tilde{C}_{\text{cross}}^K$ at time $\theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K$.

First, we need a lemma to connect Step 2 and Step 3.

Lemma II.7.5. *Fix $\epsilon > 0$. Suppose that the assumptions of Theorem II.7.1 hold. Then, there exists a constant $M > 0$ (independent of ϵ and K) such that,*

(a) Soon after $\theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K$, the total population size is close to $\psi(\tilde{C}_{\text{cross}}^\epsilon - \frac{\epsilon}{2})$:

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\theta_{\text{near } \psi(\tilde{C}_{\text{cross}}^\epsilon - \frac{\epsilon}{2})} > \theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K + S_K \wedge \theta_{2 \text{ succ. mut.}}^K \wedge \theta_{\text{diversity}}^K$$

$$\wedge \inf\Big\{t \geq \theta_{\text{mut. size } C_{\text{cross}}^\epsilon}^K : \mathfrak{M}^0(\tilde{\nu}_t) = \lceil (\tilde{C}_{\text{cross}}^\epsilon \pm 3\epsilon/4)K \rceil\Big\}\Big] = 0.$$

(b) A change of order ϵ for the resident density takes more than $o(\sigma_K^{-1})$ time:

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\inf \big\{ t \geq \theta^K_{mut.\ size\ C^{\epsilon}_{cross}} : \mathfrak{M}^0(\tilde{\nu}_t) = \big\lceil (\tilde{C}^{\epsilon}_{cross} \pm 3\epsilon/4)K \big\rceil \big\}$$

$$< \theta^K_{mut.\ size\ C^{\epsilon}_{cross}} + S_K \wedge \theta_{near\ \psi(\tilde{C}^{\epsilon}_{cross} - \frac{\epsilon}{2})} \wedge \theta^K_{2\ succ.\ mut.} \wedge \theta^K_{diversity} \Big] = 0.$$

(c) *At the time when the resident density has changed of order ϵ the total population size is still close to $\psi(\tilde{C}^{\epsilon}_{cross} - \frac{\epsilon}{2})$:*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\inf \big\{ t \geq \theta^K_{near\ \psi(\tilde{C}^{\epsilon}_{cross} - \frac{\epsilon}{2})} : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \psi(\tilde{C}^{\epsilon}_{cross} - \tfrac{\epsilon}{2})| > M\epsilon\sigma_K \big\} < \theta^K_{2\ succ.\ mut.}$$

$$\wedge \theta^K_{diversity} \wedge \inf \big\{ t \geq \theta^K_{mut.\ size\ C^{\epsilon}_{cross}} : \mathfrak{M}^0(\tilde{\nu}_t) = \big\lceil (\tilde{C}^{\epsilon}_{cross} \pm \epsilon)K \big\rceil \big\} \Big] = 0.$$

(d) *A change of order ϵ for the resident density takes no more than $(i\sigma_K)^{-1-\alpha/2}$ time:*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\theta^K_{res.\ size\ \tilde{C}^{\epsilon}_{cross} - \epsilon} > \theta^K_{mut.\ size\ C^{\epsilon}_{cross}} + (i\sigma_K)^{-1-\alpha/2}) \wedge \theta^K_{2\ succ.\ mut.} \wedge \theta^K_{diversity}$$

$$\wedge \inf \big\{ t \geq \theta^K_{near\ \psi(\tilde{C}^{\epsilon}_{cross} - \frac{\epsilon}{2})} : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \psi(\tilde{C}^{\epsilon}_{cross} - \tfrac{\epsilon}{2})| > M\epsilon\sigma_K \big\} \Big] = 0.$$

Proof. Apply the methods of (a) to (d) from Lemma II.7.4. $\qquad\square$

Next, we have the following similar lemmata as in Step 2, for them let us define

$$\theta^K_{near\ \psi(i\frac{\epsilon}{2})} \equiv \inf\{ t \geq \theta^K_{res.\ size\ i(\epsilon/2)} : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \psi(i(\epsilon/2))| < (M/3)\epsilon\sigma_K \}. \tag{II.7.80}$$

Lemma II.7.6. *Suppose that the assumptions of Theorem II.7.1 hold. Then, there exists a constant $M > 0$ (independent of ϵ, K, and i) such that, for all $\epsilon > 0$ and for all $(\tilde{C}^{\epsilon}_{cross} - \epsilon)(2/\epsilon) \geq i \geq 2$,*

(a) *Soon after $\theta^K_{res.\ size\ i(\epsilon/2)}$, the total population size is close to $\psi(i\frac{\epsilon}{2})$:*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\theta^K_{near\ \psi(i\frac{\epsilon}{2})} > \theta^K_{res.\ size\ i(\epsilon/2)} + S_K \wedge \theta^K_{2\ succ.\ mut.} \wedge \theta^K_{diversity}$$

$$\wedge \inf \big\{ t \geq \theta^K_{res.\ size\ i(\epsilon/2)} : \mathfrak{M}^0(\tilde{\nu}_t) = \big\lceil (i \pm \tfrac{1}{2})(\epsilon/2)K \big\rceil \big\} \Big] = 0.$$

(b) *A change of order ϵ for the resident density takes more than $o(\sigma_K^{-1})$ time:*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\inf \big\{ t \geq \theta^K_{res.\ size\ i(\epsilon/2)} : \mathfrak{M}^0(\tilde{\nu}_t) = \big\lceil (i \pm \tfrac{1}{2})(\epsilon/2)K \big\rceil \big\}$$

$$< \theta^K_{res.\ size\ i(\epsilon/2)} + S_K \wedge \theta^K_{near\ \psi(i\frac{\epsilon}{2})} \wedge \theta^K_{2\ succ.\ mut.} \wedge \theta^K_{diversity} \Big] = 0.$$

(c) *At the time when the resident density has changed of order ϵ the total population size is still close to $\psi(i\frac{\epsilon}{2})$:*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\inf \big\{ t \geq \theta^K_{near\ \psi(i\frac{\epsilon}{2})} : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \psi(i(\epsilon/2))| > M\epsilon\sigma_K \big\} < \theta^K_{2\ succ.\ mut.}$$

$$\wedge \theta^K_{diversity} \wedge \inf \big\{ t \geq \theta^K_{res.\ size\ i(\epsilon/2)} : \mathfrak{M}^0(\tilde{\nu}_t) = \big\lceil (i \pm 1)(\epsilon/2)K \big\rceil \big\} \Big] = 0.$$

(d) *A change of order ϵ for the resident density takes no more than $(i\sigma_K)^{-1-\alpha/2}$ time:*

$$\lim_{K \to \infty} \sigma_K^{-1} \mathbb{P}\Big[\theta^K_{res.\ size\ (i-1)(\epsilon/2)} > \big(\theta^K_{near\ \psi(i\frac{\epsilon}{2})} + (i\sigma_K)^{-1-\alpha/2} \big) \wedge \theta^K_{2\ succ.\ mut.} \wedge \theta^K_{diversity}$$

$$\wedge \inf \big\{ t \geq \theta^K_{near\ \psi(i\frac{\epsilon}{2})} : |\langle \tilde{\nu}_t, \mathbb{1} \rangle - \psi(i(\epsilon/2))| > M\epsilon\sigma_K \big\} \Big] = 0.$$

Proof. Apply the methods of (a) to (d) from Lemma II.7.4.   □

Remark 8. Lemma II.7.5 and II.7.6 imply that the density of the resident trait decreases to the value ϵ. Moreover,

$$\mathbb{P}\left[\theta^K_{\text{res. size } \epsilon} > \theta^K_{\text{mut. size } C^\epsilon_{\text{cross}}} + \frac{\ln(K)}{\sigma^{1+\alpha/2}_K} \wedge \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}}\right] = o(\sigma_K) \qquad \text{(II.7.81)}$$

$$\text{and} \qquad \mathbb{P}\left[|\langle \tilde{\nu}_{\theta^K_{\text{res. size } \epsilon}}, \mathbb{1}\rangle - \psi(\epsilon)| > M\epsilon\sigma_K\right] = o(\sigma_K). \qquad \text{(II.7.82)}$$

II.7.4   Step 4: the resident trait R^K goes extinct

After the time $\theta^K_{\text{res. size } \epsilon}$ we have to wait less than $\ln(K)\sigma^{1+\alpha/2}_K$ time to know that the resident trait is extinct with high probability.

Notation. Define $\theta^K_{\text{near } \psi(0)} \equiv \inf\{t \geq \theta^K_{\text{res. size } \epsilon} : |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \psi(0)| < (M/3)\epsilon\sigma_K\}$.

Lemma II.7.7. *Suppose that the assumptions of Theorem II.7.1 hold. Then, there exists a constant $M > 0$ (independent of ϵ and K) such that, for all $\epsilon > 0$*

(a) Soon after $\theta^K_{\text{res. size } \epsilon}$, the total population size is close to $\psi(0)$:

$$\lim_{K \to \infty} \sigma^{-1}_K \mathbb{P}\left[\theta^K_{\text{near } \psi(0)} > \theta^K_{\text{res. size } \epsilon} + S_K \wedge \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}}\right.$$
$$\left. \wedge \inf\left\{t \geq \theta^K_{\text{res. size } \epsilon)} : \mathfrak{M}^0(\tilde{\nu}_t) = \left\lceil(1 \pm \tfrac{1}{4})\epsilon K\right\rceil\right\}\right] = 0.$$

(b) A change of order ϵ for the resident density takes more than $o(\sigma^{-1}_K)$ time:

$$\lim_{K \to \infty} \sigma^{-1}_K \mathbb{P}\left[\inf\left\{t \geq \theta^K_{\text{res. size } \epsilon} : \mathfrak{M}^0(\tilde{\nu}_t) = \left\lceil(1 \pm \tfrac{1}{4})\epsilon K\right\rceil\right\}\right.$$
$$\left. < \theta^K_{\text{res. size } \epsilon} + S_K \wedge \theta_{\text{near } \psi(0)} \wedge \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}}\right] = 0.$$

Proof. See proof of Lemma II.7.4   □

Lemma II.7.8. *Suppose that the assumptions of Theorem II.7.1 hold. Then, there exists a constant $M > 0$ (independent of ϵ and K) such that, for all $\epsilon > 0$*

$$\lim_{K \to \infty} \sigma^{-1}_K \mathbb{P}\left[\theta^K_{\text{res. size } 0} > \left(\theta^K_{\text{near } \psi(0)} + \ln(K)\sigma^{-1-\alpha/2}_K\right) \wedge \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}}\right. \qquad \text{(II.7.83)}$$
$$\left. \wedge \inf\left\{t \geq \theta^K_{\text{near } \psi(0)} : |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \psi(0))| > M\epsilon\sigma_K\right\}\right] = 0.$$

Proof. To prove this lemma we use a coupling with a continuous time branching process as in the proof of lemma II.7.3. For any $\theta^K_{\text{near } \psi(0)} \leq t \leq \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}} \wedge \inf\{t \geq \theta^K_{\text{near } \phi(0)} : |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \psi(0))| > M\epsilon\sigma_K\}$, any individual of $\mathfrak{M}^0(\tilde{\nu}_t)$ gives birth to a new individual with trait R^K with rate

$$\left(1 - u_K\, m(R^K)\right)b(R^K) \in \left[b(R^K) - u_K\, \overline{b}, b(R^K)\right], \qquad \text{(II.7.84)}$$

and dies with rate

$$d(R^K) + c(R^K, R^K)\mathfrak{M}^0(\tilde{\nu}_t) + \int_{\mathcal{X} \times \mathbb{N}} c(R^K, \xi)\tilde{d}\nu_t(\xi), \qquad \text{(II.7.85)}$$

which is larger than $d_Z \equiv d(R^K) + c(R^K, R^K + \sigma_K h)\bar{z}(R^K + \sigma_K h) - C^M_{\text{total death}}\epsilon\sigma_K$ where $C^M_{\text{total death}} \equiv M + \bar{c}\lceil 3/\alpha \rceil - 2h\partial_2 c(R^K, R^K)$. Therefore, we construct, by using a standard coupling argument, a process Z_t such that

$$Z_t \geq \mathfrak{M}^0(\tilde{\nu}_t) \tag{II.7.86}$$

for all $\theta^K_{\text{near } \psi(0)} \leq t \leq \theta^K_{2 \text{ succ. mut.}} \wedge \theta^K_{\text{diversity}} \wedge \inf\{t \geq \theta^K_{\text{near } \phi(0)} : |\langle\tilde{\nu}_t, \mathbb{1}\rangle - \psi(0))| > M\epsilon\sigma_K\}$. The process Z_t is a linear birth-death process starting at $\lceil \frac{5}{4}\epsilon K \rceil$, with birth rate per individual $b_Z = b(R^K)$ and with death rate per individual d_Z. Since

$$b_Z - d_Z = f(R^K, R^K + \sigma_K h) + C^M_{\text{total death}}\epsilon\sigma_K \tag{II.7.87}$$
$$= -\sigma_K h\partial_1 f(R^K + \sigma_K h, R^K + \sigma_K h) + C^M_{\text{total death}}\epsilon\sigma_K + O((\sigma_K h)^2)$$
$$\equiv -\sigma_K \xi_K$$

is negative and of order σ_K, the process Z_t is sub-critical.

Note that $\xi_K \geq \inf_{x\in\mathcal{X}} \frac{\partial_1 f(x,x)}{2} > 0$. Let τ^Z_i be the first hitting time of level i by Z_t, then we have

$$\mathbb{P}[\tau^Z_{\lceil 2\epsilon K\rceil} < \tau^Z_0] \leq \exp(-K^\alpha) \tag{II.7.88}$$

compare with the proof of Proposition II.9.5. Since $Z_t \geq \mathfrak{M}^0(\tilde{\nu}_t)$, we obtain also that, with high probability, $\mathfrak{M}^0(\tilde{\nu}_t)$ stays smaller than $\lceil 2\epsilon K\rceil$ before it dies out. For any $t \geq 0$ and $n \in \mathbb{N}$, the distribution of the extinction time of Z_t for $b_Z \neq d_Z$ is given by:

$$\mathbb{P}_n(\tau^Z_0 \leq t) = \left(\frac{d_Z - d_Z\exp((d_Z - b_Z)t)}{b_Z - d_Z\exp((d_Z - b_Z)t)}\right)^n. \tag{II.7.89}$$

(cf. [6] p. 109 and [25]). Therefore, we can compute in our case where $d_Z - b_Z = \sigma_K\xi_K$ with ξ_K uniformly positive

$$\mathbb{P}\left[\tau^Z_0 \leq \ln(K)\sigma_K^{-1-\alpha/2}\right] = \left(\frac{d_Z - d_Z\exp\left((d_Z - b_Z)\ln(K)\sigma_K^{-1-\alpha/2}\right)}{b_Z - d_Z\exp\left((d_Z - b_Z)\ln(K)\sigma_K^{-1-\alpha/2}\right)}\right)^{\frac{5}{4}\epsilon K} \tag{II.7.90}$$

$$= \left(\frac{d_Z - d_Z K^{\xi_K\sigma_K^{-\alpha/2}}}{d_Z - \sigma_K\xi_K - d_Z K^{\xi_K\sigma_K^{-\alpha/2}}}\right)^{\frac{5}{4}\epsilon K}$$

$$= \left(1 - \frac{\xi_K\sigma_K}{d_Z\left(K^{\xi_K\sigma_K^{-\alpha/2}} - 1\right) + \sigma_K\xi_K}\right)^{\frac{5}{4}\epsilon K}$$

$$\geq \left(1 - \sigma_K(\tfrac{5}{4}\epsilon K)^{-1}K^{-1}\right)^{\frac{5}{4}\epsilon K}$$

$$\geq 1 - O(\sigma_K K^{-1}) \geq 1 - o(\sigma_K),$$

which proves the lemma. $\qquad\square$

II.7.5   Step 5: the population becomes monomorphic and stays close to its equilibrium

After the extinction time of the resident trait, we have to wait at most $\ln(K)\sigma_K^{-1-\alpha/2}$ time until the population is monomorphic with trait $R^K + \sigma_K h$.

Lemma II.7.9. *Suppose that the assumptions of Theorem II.7.1 hold. Then, there exists a constant $M > 0$ (independent of ϵ and K) such that, for all $\epsilon > 0$*

$$\lim_{K \to \infty} \sigma_K^{-1} \, \mathbb{P}\Big[\theta_{fixation}^K > (\theta_{res.\ size\ 0}^K + \ln(K)\sigma_K^{-1-\alpha/2}) \wedge \theta_{2\ succ.\ mut.}^K \wedge \theta_{diversity}^K \tag{II.7.91}$$

$$\wedge \inf\Big\{t \geq \theta_{near\ \phi(0)}^K : |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \psi(0)| > M\epsilon\sigma_K\Big\}\Big] = 0.$$

Proof. By the last lemmata, we have $\theta_{fixation}^K = \inf\{t \geq \theta_{res.\ size\ 0}^K : |\mathrm{Supp}(\tilde{\nu}_t^K)| = 1, |\langle \tilde{\nu}_t, \mathbb{1}\rangle - \psi(0)| < (M/3)\epsilon\sigma_K\}$ with probability $1 - o(\sigma_K)$. Set $D \equiv \{k \in \mathbb{N} : 1 \leq \mathfrak{M}^k(\tilde{\nu}_{\theta_{res.\ size\ 0}^K}) < \epsilon\sigma_K K\}$. Then $|D| \leq \lceil 3/\alpha \rceil$, and none of these traits are successful since we have seen that $\theta_{res.\ size\ 0}^K$ is smaller than $\theta_{2\ succ.\ mut.}^K$ and $\theta_{diversity}^K$ with probability of order $1 - o(\sigma_K)$. By applying Proposition II.9.3 and using the Markov inequality, we obtain that the life time of each of these subpopulations is with probability $1 - o(\sigma_K)$ smaller than $\ln(K)\sigma_K^{-1-\alpha/4}$. Therefore, if no new mutant is born between $\theta_{res.\ size\ 0}^K$ and $\theta_{res.\ size\ 0}^K + \ln(K)\sigma_K^{-1-\alpha/4}$, we obtain the claim. On the other hand, as in Lemma II.6.4, the number of mutants born in the time interval $[\theta_{res.\ size\ 0}^K, \theta_{res.\ size\ 0}^K + \ln(K)\sigma_K^{-1-\alpha/2}]$ is stochastically dominated by a Poisson point process, $A^K(t)$, with parameter $a \, u_K K$, where $a \equiv \sup_{x \in \mathcal{X}} \overline{z}(x)b(x)m(x) + 1$. Hence, the probability to have no new mutant in this interval is

$$\mathbb{P}\big[A^K(\ln(K)\sigma_K^{-1-\alpha/2}) = 0\big] = \exp(-\ln(K)\sigma_K^{-1-\alpha/2}a \, u_K K) \tag{II.7.92}$$

$$\geq \exp(-\sigma_K^{\alpha/2}) \geq 1 - o(1).$$

Because the probability that a mutant is successful is of order σ_K, the probability that a successful mutant is born between times $\theta_{res.\ size\ 0}^K$ and $\theta_{res.\ size\ 0}^K + \ln(K)\sigma_K^{-1-\alpha/2}$ is $o(\sigma_K)$. Since

$$\mathbb{P}\big[A^K(\ln(K)\sigma_K^{-1-\alpha/2}) \leq \lceil 3/\alpha \rceil\big] \tag{II.7.93}$$

$$= \exp\Big(-\ln(K)\sigma_K^{-1-\alpha/2} \, au_K K\Big) \sum_{i=0}^{\lceil 3/\alpha \rceil} \frac{\ln(K)\sigma_K^{-1-\alpha/2}a \, u_K K}{i}$$

$$\geq 1 - (\ln(K)\sigma_K^{-1-\alpha/2}a \, u_K K)^{\lceil 3/\alpha \rceil + 1}$$

$$\geq 1 - \sigma_K^{3/2} = 1 - o(\sigma_K),$$

there are maximal $\lceil 3/\alpha \rceil$ unsuccessful mutations in this interval. With the same argument as before the life time of each of these subpopulations is with probability $1 - o(\sigma_K)$ smaller than $\ln(K)\sigma_K^{-1-\alpha/4}$. Therefore, with probability $1 - o(\sigma_K)$ the maximal possible time interval where at least one mutant individual is alive is smaller or equal $\ln(K)\sigma_K^{-1-\alpha/4} + \lceil 3/\alpha \rceil \ln(K)\sigma_K^{-1-\alpha/4} \ll \ln(K)\sigma_K^{-1-\alpha/2}$. Recall from Lemma II.7.7 that if $|\langle \tilde{\nu}_t, \mathbb{1}\rangle - \psi(0)| > (M/3)\epsilon\sigma_K$ at the first time when the population is again monomorphic, then the time the process needs to enter the $(M/3)\epsilon\sigma_K$-neighborhood of $\psi(0)$ is smaller than S_K, which can be chosen smaller than $\sigma_K^{1+\alpha}/(Ku_K)$. This proves the lemma. $\qquad\square$

This ends Step 5 and the second invasion phase. Note that the estimates of the two phases do not depend on the exact trait value of the resident trait, especially the a priori different constants M. In fact, we can use in all lemmata the same constant M, namely the largest. Therefore, we can apply our results for the successful mutant trait $R_1^K = R^K + \sigma_K h$, which is the next resident trait by using the strong Markov property for $(\tilde{\nu}, L)$ at the stopping time $\theta_{fixation}^K$.

II.8 Convergence to the CEAD

Our goal is to find $T_0 > 0$ and to construct, for all $\epsilon > 0$, two measure valued processes, $(\mu_t^{1,K,\epsilon}, t \geq 0)$ and $(\mu_t^{2,K,\epsilon}, t \geq 0)$, in $\mathbb{D}([0,\infty), \mathcal{M}(\mathcal{X}))$ such that

$$\lim_{K\to\infty} \mathbb{P}\left[\forall\, t \leq \frac{T_0}{K u_K \sigma_K^2} : \quad \mu_t^{1,K,\epsilon} \preccurlyeq \nu_t^K \preccurlyeq \mu_t^{2,K,\epsilon} \right] = 1, \tag{II.8.1}$$

and for $j \in \{1,2\}$

$$\lim_{K\to\infty} \mathbb{P}\left[\sup_{0 \leq t \leq T_0} \left\| \mu_{t/(K u_K \sigma_K^2)}^{i,K,\epsilon} - \overline{z}(x_t)\delta_{x_t} \right\|_0 > \delta(\epsilon) \right] = 0, \tag{II.8.2}$$

for some function δ independent of x, K such that $\delta(\epsilon) \to 0$ when $\epsilon \to 0$. This easily implies (II.4.5) for all $T \leq T_0$.

The result for all $T > 0$ then follows from the strong Markov property. Indeed, the construction below implies that there exists a stopping time

$$\tau \in [T_0/2K u_K \sigma_K^2, T_0/K u_K \sigma_K^2] \tag{II.8.3}$$

(a fixation time) such that, with probability converging to 1, ν_τ^K has a unique (random) point Y as support and a total mass belonging to $[\overline{z}(Y) - M\sigma_K, \overline{z}(Y) + M\sigma_K]$. Hence (II.8.1) and (II.8.2) also hold for the process $(\nu_{\tau+t}^K, t \geq 0)$, and (II.4.5) is thus true for all $T \leq 3T_0/2$. We obtain (II.4.5) for any fixed $T > 0$ by induction.

II.8.1 Construction of two processes $\mu^{K,1}$ and $\mu^{K,2}$ such that $\mu_t^{1,K} \preccurlyeq \nu_t^K \preccurlyeq \mu_t^{2,K}$

Fix $T > 0$. Let θ_i^K denote the random time of i-th invasion (i.e., $\theta_i^K = \theta_{i,\text{invasion}}^K$), $\theta_{i,\text{fixation}}^K$ the time of i-th fixation and R_i^K the trait of the i-th successful mutant. Let us fix the following initial conditions $R_0^{K,1} = R_0^K - A\sigma_K$, $R_0^{K,2} = R_0^K + A\sigma_K$ and $\theta_0^{K,1} = \theta_0^{K,2} = 0$. Assume that we have constructed $\theta_i^{K,1}$ and $\theta_i^{K,2}$, and $R_i^{K,1}$ and $R_i^{K,2}$. By Theorem II.6.2 and the Markov property, we can construct two random variables $R_{i+1}^{K,1}$ and $R_{i+1}^{K,2}$ such that

$$R_{i+1}^{K,1} - R_i^{K,1} \leq R_{i+1}^K - R_i^K \leq R_{i+1}^{K,2} - R_i^{K,2} \tag{II.8.4}$$

with probability $1 - o(\sigma_K)$. Moreover, $R_{i+1}^{K,1} - R_i^{K,1} = R_{i+1}^K - R_i^K = R_{i+1}^{K,2} - R_i^{K,2}$ with probability $1 - O(\epsilon)$ and $R_{i+1}^{K,2} - R_{i+1}^{K,1} \leq A\sigma_K$. The distributions of $R_{i+1}^{K,1} - R_i^{K,1}$ and $R_{i+1}^{K,2} - R_i^{K,2}$ are (cf. Corollary II.6.10)

$$
\begin{aligned}
r_1^\epsilon(R_i^K, h) &\equiv \mathbb{P}[R_{i+1}^{K,1} = R_i^K + \sigma_k h] \\
&= \begin{cases} \dfrac{M(R_i^K,1)q_1^\epsilon(R_i^K,1)}{p_2^\epsilon(R_i^K)} + 1 - \dfrac{p_1^\epsilon(R_i^K)}{p_2^\epsilon(R_i^K)} & \text{if } h = 1 \\[3mm] \dfrac{M(R_i^K,h)q_1^\epsilon(R_i^K,h)}{p_2^\epsilon(R_i^K)} & \text{if } h \in \{2, ..., A\} \end{cases}
\end{aligned}
\tag{II.8.5}
$$

and

$$
\begin{aligned}
r_2^\epsilon(R_i^K, h) &\equiv \mathbb{P}[R_{i+1}^{K,2} = R_i^K + \sigma_k h] \\
&= \begin{cases} \dfrac{M(R_i^K,h)q_1^\epsilon(R_i^K,h)}{p_2^\epsilon(R_i^K)} & \text{if } h \in \{1, ..., A-1\} \\[3mm] \dfrac{M(R_i^K,A)q_1^\epsilon(R_i^K,A)}{p_2^\epsilon(R_i^K)} + 1 - \dfrac{p_1^\epsilon(R_i^K)}{p_2^\epsilon(R_i^K)} & \text{if } h = A, \end{cases}
\end{aligned}
\tag{II.8.6}
$$

where

$$q_1^\epsilon(x,h) = h\frac{\partial_1 f(x,x)}{b(x)} - C_{\text{Bernoulli}}^1\epsilon, \quad q_2^\epsilon(x,h) = h\frac{\partial_1 f(x,x)}{b(x)} + C_{\text{Bernoulli}}^2\epsilon \qquad (\text{II}.8.7)$$

and $p_j^\epsilon(x) = \sum_{h=1}^A q_j^\epsilon(x,h)M(x,h)$ for $j = 1,2$. (Note that we changed a bit the notations of Corollary II.6.10 to make explicit the dependence on ϵ and R_i^K.) Since we assumed that the fitness gradient $\partial_1 f(x,x)$ is positive and uniformly lower bounded on $\mathcal{X}$, the transition probabilities $r_j^\epsilon(x,h)$, $j = 1,2$ are uniformly Lipschitz-continuous functions of x with some Lipschitz constant C_{Lip}^r. By Theorem II.6.2 and Lemmata II.6.7 and II.6.4, we can construct two exponential random variables, $E_{i+1}^{K,1}$ and $E_{i+1}^{K,2}$, with parameters $a_1^{K,\epsilon}(R_i^K)p_1^\epsilon(R_i^K)\sigma_K u_K K$ and $a_2^{K,\epsilon}(R_i^K)p_2^\epsilon(R_i^K)\sigma_K u_K K$ given by

$$a_1^{K,\epsilon}(x) = (\bar{z}(x) - \epsilon\sigma_K M)b(x)m(x) \qquad (\text{II}.8.8)$$

$$a_2^{K,\epsilon}(x) = (\bar{z}(x) + \epsilon\sigma_K(M + \lceil 3/\alpha\rceil))(b(x)m(x) + C_L^{b,m,M}A\sigma_K), \qquad (\text{II}.8.9)$$

such that

$$\mathbb{P}\left(E_{i+1}^{K,2} \le \theta_{i+1}^K - \theta_{i,\text{fixation}}^K \le E_{i+1}^{K,1} + \ln(K)\sigma_K^{-1-\alpha/2}\right) = 1 - o(\sigma_K). \qquad (\text{II}.8.10)$$

Note that this inequality involves $\theta_{i,\text{fixation}}^K$ instead of θ_i^K since we apply the Markov property at the fixation time of Lemma II.7.9 before we can apply Theorem II.6.2. However, Lemma II.7.9 entails that we also have

$$\mathbb{P}\left(E_{i+1}^{K,2} \le \theta_{i+1}^K - \theta_i^K \le E_{i+1}^{K,1} + 6\ln(K)\sigma_K^{-1-\alpha/2}\right) = 1 - o(\sigma_K). \qquad (\text{II}.8.11)$$

We then define

$$\theta_{i+1}^{K,1} - \theta_i^{K,1} \equiv E_{i+1}^{K,1} + 6\ln(K)\sigma_K^{-1-\alpha/2} \quad \text{and} \quad \theta_{i+1}^{K,2} - \theta_i^{K,2} \equiv E_{i+1}^{K,2}. \qquad (\text{II}.8.12)$$

In addition, by their construction in Section II.6, it is clear that the random vectors $\{(E_{i+1}^{K,1}, E_{i+1}^{K,2}, R_{i+1}^{K,1} - R_i^{K,1}, R_{i+1}^{K,2} - R_i^{K,2})\}_{i\ge0}$ are independent conditionally on $(R_j^K)_{j\ge0}$.

Lemma II.8.1. *With the previous notations, the processes $\mu^{K,1}$ and $\mu^{K,2}$ in $\mathbb{D}([0,\infty),\mathcal{M}(\mathcal{X}))$ defined for all $t \ge 0$ by*

$$\mu_t^{1,K} = (\bar{z}(R_j^K) - (M\epsilon + \overline{C})\sigma_K)\delta_{R_i^{K,1}}, \qquad \text{for } t\in[\theta_i^{K,1},\theta_{i+1}^{K,1})\cap[\theta_j^K,\theta_{j+1}^K), \qquad (\text{II}.8.13)$$

$$\mu_t^{2,K} = \left(\bar{z}(R_j^K) + \left(M\epsilon + \left\lceil\frac{3}{\alpha}\right\rceil\epsilon + \overline{C}\right)\sigma_K\right)\delta_{R_i^{K,2}}, \qquad \text{for } t\in[\theta_i^{K,2},\theta_{i+1}^{K,2})\cap[\theta_j^K,\theta_{j+1}^K), \qquad (\text{II}.8.14)$$

for some constant $\overline{C}$ independent of K,x,ϵ, satisfy for all $T > 0$

$$\lim_{K\to\infty}\mathbb{P}\left[\forall\, t \le \frac{T}{Ku_K\sigma_K^2}: \quad \mu_t^{1,K} \preccurlyeq \nu_t^K \preccurlyeq \mu_t^{2,K}\right] = 1. \qquad (\text{II}.8.15)$$

Note that the support of $\mu^{j,K}$, $j = 1,2$, is defined from the sequences $(R_i^{K,j})_{i\ge0}$ and $(\theta_i^{K,j})_{i\ge1}$ but the mass of $\mu^{j,K}$ is defined from the sequences $(R_i^K)_{i\ge0}$ and $(\theta_i^K)_{i\ge1}$.

Proof. Let us fix $T > 0$ and $\Gamma > 0$. Since each of the steps previously described holds with probability $1 - o(\sigma_K)$, we deduce that the above construction can be done on a *good event* of

probability $1 - o(1)$, for all integers $i \leq \Gamma/\sigma_K$. Since in addition, on $\mathcal{X}$, $a_2^{K,\epsilon}(x)p_2^\epsilon(x)$ is uniformly bounded from below by a positive constant $\underline{a}$, the random variables $E_i^{K,2}$ can be coupled with i.i.d. exponential ones of parameter $\underline{a}Ku_K\sigma_K$, and hence $\mathbb{P}\big[\theta_{\lfloor\Gamma/\sigma_K\rfloor}^{K,2} < T/(Ku_K\sigma_K^2)\big]$ is smaller than the probability that a Poisson process with parameter $\underline{a}Ku_K\sigma_K$ is larger that $\lfloor\Gamma/\sigma_K\rfloor$ at time $T/(Ku_K\sigma_K^2)$. By the law of large numbers for Poisson processes, we deduce that if $\Gamma > T\underline{a}$ (which we assume true in the sequel),

$$\lim_{K\to\infty} \mathbb{P}\left[\theta_{\lfloor\Gamma/\sigma_K\rfloor}^{K,2} < \frac{T}{Ku_K\sigma_K^2}\right] = 0. \tag{II.8.16}$$

Let us recall that, on the previous good event of probability $1 - o(1)$, the number, the trait, and the size of the living mutant populations and the size of the resident population are controlled at any time in the i-th first phase (Lemmata II.6.3 and II.6.9). In addition, during the i-th second phase, the number, trait, and size of living mutant populations are controlled (see all the Lemmas of Section II.7), the total mass of the population stays within the $M\epsilon\sigma_K$-neighborhood of $\phi(y)$ or $\psi(y)$ for some $y \in [0, \bar{z}(R_i^K)]$ (Lemmata II.7.4 and II.7.6). Since $|\phi(y) - \bar{z}(R_i^K)| \leq \overline{C}\sigma_K$ and $|\psi(y) - \bar{z}(R_i^K)| \leq \overline{C}\sigma_K$ for some constant $\overline{C}$, as seen in (II.7.26) and (II.7.76), and since the sequences $(R_i^{j,K})_{i\geq 0}$ for $j = 1, 2$ and $(R_i^K)_{i\geq 0}$ are all increasing on the good event, we deduce the required comparison between the supports of $\mu_t^{1,K}$, ν_t^K, and $\mu_t^{2,K}$ for $t \leq \frac{T}{Ku_K\sigma_K^2}$, on the good event. Since we used $\bar{z}(R_j^K)$ to define the masses of $\mu_t^{1,K}$ and $\mu_t^{2,K}$, the required comparison between the masses is also clear. $\qquad\square$

Note that, since the function $\bar{z}$ may not be non-decreasing, replacing $\bar{z}(R_j^K)$ by $\bar{z}(R_j^{K,1})$ in the definition of $\mu_t^{1,K}$ may not imply the required comparison between the masses of $\mu_t^{1,K}$, ν_t^K, and $\mu_t^{2,K}$.

The next goal is now to prove the convergence of both processes $\mu_{t/Ku_K\sigma_K^2}^{K,j}$ for $j = 1, 2$ to $\bar{z}(x_t)\delta_{x_t}$ in probability in $L^\infty(\mathcal{M}(\mathcal{X}), \|\cdot\|_0)$. For this, we will use standard convergence results of Markov jump processes. However, the two processes $\mu^{K,j}$, $j = 1, 2$ are not Markov because the i-th jump rates and transition probabilities defined above depend on R_i^K which is close, but different from $R_i^{K,j}$. Therefore, we introduce a small parameter, $\eta > 0$, and we construct two Markov processes $\mu^{K,j,\epsilon,\eta}$, $j = 1, 2$ in $\mathbb{D}([0,\infty), \mathcal{M}(\mathcal{X}))$ such that

$$\lim_{K\to+\infty} \mathbb{P}\left[\mu_{(t-1/(Ku_K\sigma_K))\vee 0}^{K,1,\epsilon,\eta} \preccurlyeq \mu_t^{1,K} \preccurlyeq \nu_t^K \preccurlyeq \mu_t^{2,K} \preccurlyeq \mu_t^{K,2,\epsilon,\eta}, \ \forall t \leq \frac{T}{Ku_K\sigma_K^2} \wedge S_\eta^K\right] = 1, \tag{II.8.17}$$

where S_η^K is the first time where the distance between the support of $\mu_t^{K,1,\epsilon,\eta}$, and $\mu_t^{K,2,\epsilon,\eta}$ is larger than η. The last equation will be proved below in Section II.8.2. The time-shift of $-1/(Ku_K\sigma_K)$ in $\mu^{K,1,\epsilon,\eta}$ is due to the terms $6\ln(K)\sigma_K^{-1-\alpha/2}$ in (II.8.12). We will next study the convergence of these two Markov processes when $K \to \infty$ and prove in Section II.8.3 that, for a convenient choice of η, there exists some $T_0 > 0$ independent of K, x, ϵ, η such that

$$\lim_{K\to+\infty} \mathbb{P}\left[S_\eta^K < \frac{T_0}{Ku_K\sigma_K^2}\right] = 0. \tag{II.8.18}$$

II.8.2 Proof of (II.8.17)

For all $x \in \mathcal{X}$, we define $(\bar{r}_1^{\epsilon,\eta}(x,h), 1 \leq h \leq A)$, and $(\bar{r}_2^{\epsilon,\eta}(x,h), 1 \leq h \leq A)$ by, for all $1 \leq \ell \leq A$,

$$\textstyle\sum_{h=1}^\ell \bar{r}_1^{\epsilon,\eta}(x,h) \equiv \left[\sum_{h=1}^\ell (r_1^\epsilon(x,h) + C_{\mathrm{Lip}}^r \eta)\right] \wedge 1 \geq \sup_{y\in[x,x+\eta]} \textstyle\sum_{h=1}^\ell r_1^\epsilon(y,h) \tag{II.8.19}$$

and

$$\sum_{h=1}^{\ell} \bar{r}_2^{\epsilon,\eta}(x,h) \equiv \left[\sum_{h=1}^{\ell}(r_2^{\epsilon}(x,h) - C_{\mathrm{Lip}}^r \eta)\right] \vee 0 \leq \inf_{y \in [x,x+\eta]} \sum_{h=1}^{\ell} r_1^{\epsilon}(y,h). \tag{II.8.20}$$

Note that $\bar{r}_1^{\epsilon,\eta}(x,\cdot)$ and $\bar{r}_2^{\epsilon,\eta}(x,\cdot)$ are probability distributions on $\{1,\ldots,A\}$ for all $x \in \mathcal{X}$ and that, by standard coupling arguments, for all $x < y$ such that $y - x \leq \eta$, the distribution $\bar{r}_1^{\epsilon,\eta}(x,\cdot)$ is stochastically dominated by the distribution $r_1^{\epsilon}(y,\cdot)$ and the distribution $r_2^{\epsilon}(x,\cdot)$ is stochastically dominated by the distribution $\bar{r}_2^{\epsilon,\eta}(y,\cdot)$. We define similarly

$$\bar{a}_1^{K,\epsilon,\eta}(x) \equiv a_1^{K,\epsilon}(x)p_1^{\epsilon}(x) - C_{\mathrm{Lip}}^a \eta \leq \inf_{y \in [x,x+\eta] \cap \mathcal{X}} a_1^{K,\epsilon}(y)p_1^{\epsilon}(y), \tag{II.8.21}$$

and

$$\bar{a}_2^{K,\epsilon,\eta}(x) \equiv a_2^{K,\epsilon}(x)p_2^{\epsilon}(x) + C_{\mathrm{Lip}}^a \eta \geq \sup_{y \in [x-\eta,x] \cap \mathcal{X}} a_2^{K,\epsilon}(y)p_2(y), \tag{II.8.22}$$

where C_{Lip}^a is a uniform Lipschitz constant for the functions $a_j^{K,\epsilon}p_j^{\epsilon}$, $j = 1,2$. Note that $a_1^{K,\epsilon,\eta}(x) > 0$ for all $x \in \mathcal{X}$ if η is small enough.

It is then clear that there exist two Markov chains $(\bar{R}_i^{K,j,\eta})_{i\geq 0}$, $j = 1,2$, with initial condition $\bar{R}_0^{K,j,\eta} = R_0^{K,j}$ and with transition probabilities $\bar{r}_j^{K,\epsilon,\eta}(x,h)$ from x to $x+h$, such that, for all $i \geq 0$ satisfying $\bar{R}_i^{K,2,\eta} - \bar{R}_i^{K,1,\eta} \leq \eta$,

$$\bar{R}_{i+1}^{K,1,\eta} - \bar{R}_i^{K,1,\eta} \leq R_{i+1}^{K,1} - R_i^{K,1} \text{ and } R_{i+1}^{K,2} - R_i^{K,2} \leq \bar{R}_{i+1}^{K,2,\eta} - \bar{R}_i^{K,2,\eta} \leq R_{i+1}^{K,1} - R_i^{K,1}. \tag{II.8.23}$$

Similarly, there are random variables $\bar{E}_{i+1}^{K,j,\eta}$, $j = 1,2$, independent and exponentially distributed with parameters $\bar{a}_j^{K,\epsilon,\eta}(\bar{R}_i^{K,j,\eta})$ conditionally on $(\bar{R}_i^{K,j,\eta})_{i\geq 0}$, such that $\bar{E}_{i+1}^{K,2,\eta} \leq E_{i+1}^{K,2}$ and $E_{i+1}^{K,1} \leq \bar{E}_{i+1}^{K,1,\eta}$. We then define $\bar{\theta}_{i+1}^{K,j,\eta} - \bar{\theta}_i^{K,j,\eta} = E_{i+1}^{K,j,\theta}$ with $\bar{\theta}_0^{K,j,\eta} = 0$.

Since the function $\bar{z}$ is $C_{\mathrm{Lip}}^{\bar{z}}$-Lipschitz, it is clear that (II.8.17) is satisfied for the processes

$$\bar{\mu}_t^{K,1,\epsilon,\eta} = (\bar{z}(\bar{X}_t^{K,1,\eta}) - (M\epsilon + \bar{C})\sigma_K - C_{\mathrm{Lip}}^{\bar{z}}\eta)\delta_{\bar{X}_t^{K,1,\eta}} \tag{II.8.24}$$

$$\text{and} \qquad \mu_t^{K,2,\epsilon,\eta} = (\bar{z}(X_t^{K,2,\eta}) + (M\epsilon + \lceil 3/\alpha \rceil \epsilon + \bar{C})\sigma_K + C_{\mathrm{Lip}}^{\bar{z}}\eta)\delta_{X_t^{K,2,\eta}}, \tag{II.8.25}$$

where

$$\bar{X}_t^{K,1,\eta} = \bar{R}_i^{K,1,\eta}, \text{ for } t \in \left[\bar{\theta}_i^{K,1,\eta} + 6i\ln(K)\sigma_K^{-1-\alpha/2}, \bar{\theta}_{i+1}^{K,1,\eta} + 6(i+1)\ln(K)\sigma_K^{-1-\alpha/2}\right), \tag{II.8.26}$$

and

$$X_t^{K,2,\eta} = \bar{R}_i^{K,2,\eta}, \text{ for } t \in [\bar{\theta}_i^{K,2,\eta}, \bar{\theta}_{i+1}^{K,2,\eta}). \tag{II.8.27}$$

By construction, the processes $X^{K,2,\eta}$ and $\mu^{K,2,\eta}$ are Markov jump processes, but the process $\bar{X}^{K,1,\eta}$ is not because of the terms $6\ln(K)\sigma_K^{-1-\alpha/2}$ involved in its definition. However, the process $\mu_t^{K,1,\epsilon,\eta} = (\bar{z}(X_t^{K,1,\eta}) - \epsilon\sigma_K M - C_{\mathrm{Lip}}^{\bar{z}}\eta)\delta_{X_t^{K,1,\eta}}$ is Markov, where

$$X_t^{K,1,\eta} = \bar{R}_i^{K,1,\eta}, \qquad \text{for } t \in [\bar{\theta}_i^{K,1,\eta}, \bar{\theta}_{i+1}^{K,1,\eta}). \tag{II.8.28}$$

The proof of (II.8.16) above also applies to the processes $\mu^{K,1,\epsilon,\eta}$, and $\bar{\mu}^{K,1,\epsilon,\eta}$. Since in addition the support of $\mu_t^{K,1,\epsilon,\eta}$ is non-decreasing, it follows that

$$\mu_{(t-6\Gamma\ln(K)\sigma_K^{-2-\alpha/2})\vee 0}^{K,1,\epsilon,\eta} \preccurlyeq \bar{\mu}_t^{K,1,\epsilon,\eta} \qquad \text{for all } t \leq T/(Ku_K\sigma_K^2) \tag{II.8.29}$$

with probability $1 + o(1)$. Our assumption (II.4.2) entails (II.8.17).

II.8.3 Convergence of $X^{K,j,\eta}$ when $K \to +\infty$ and proof of (II.8.18)

The two Markov processes $X^{K,1,\eta}_{t/(Ku_K\sigma_K)}$, and $X^{K,2,\eta}_{t/(Ku_K\sigma_K)}$ fit exactly to the framework and assumptions of Theorem 2.1 of Chapter 11 of [55]: their state spaces are (up to a translation) a subset of $\sigma_K\mathbb{Z}$, and their transition rates from z to $z+h\sigma_K$ have the form $\sigma_K^{-1}[\beta_h(z)+O(\sigma_K)]$ for some Lipschitz functions β_h. For such a process X, provided X_0 converges a.s. to x_0, the process $(X_{t/\sigma_K}, t \geq 0)$ converges when $\sigma_K \to 0$ almost surely in $L^\infty([0,T])$ for all $T > 0$ to the unique deterministic solution of the ODE $dx(t)/dt = \sum_h h\beta_h(x)$ with $x(0) = x_0$. In our situation, we obtain, for $j = 1, 2$, that

$$\lim_{K\to+\infty} \sup_{t\in[0,T]} \left| X^{K,j,\eta}_{t/(Ku_K\sigma_K^2)} - x_j(t) \right| = 0 \quad \text{a.s.,} \tag{II.8.30}$$

where x_1 and x_2 are the unique solutions such that $x_1(0) = x_2(0) = x$ of the ODEs

$$\frac{dx_1(t)}{dt} = \left(\bar{z}(x_1(t))b(x_1(t))m(x_1(t))p_1^\epsilon(x_1(t)) - C_{\text{Lip}}^a\eta\right)\sum_{h=1}^A h\bar{r}_1^{\epsilon,\eta}(x_1(t),h) \tag{II.8.31}$$

and

$$\frac{dx_2(t)}{dt} = \left(\bar{z}(x_2(t))b(x_2(t))m(x_2(t))p_2^\epsilon(x_2(t)) + C_{\text{Lip}}^a\eta\right)\sum_{h=1}^A h\bar{r}_2^{\epsilon,\eta}(x_2(t),h). \tag{II.8.32}$$

Lemma II.8.2. *For all $T > 0$, and for $j = 1, 2$,*

$$\sup_{t\in[0,T]} |x_j(t) - x_t| \leq CTe^{CT}(\eta + \epsilon), \tag{II.8.33}$$

for a constant C independent of x, T, ϵ, and η, where x_t is the solution of the CEAD (II.4.3) with initial condition $x_0 = x$.

Proof. We only write the proof for $j = 1$, the case $j = 2$ being similar. Since the functions $\bar{r}_j^{\epsilon,\eta}$, $j = 1, 2$, $\bar{z}$, b, m and p_1 are bounded by constants independent of K, ϵ, η, we have for all $t \in [0,T]$ and for a constant $C > 0$ that may change from line to line,

$$|x_t - x_1(t)| \leq CC_{\text{Lip}}^a\eta T + \int_0^t \left| (\bar{z}bmp_1^\epsilon)(x_1(s)) \sum_{h=1}^A h\bar{r}_1^{\epsilon,\eta}(x_1(s),h) \right. \tag{II.8.34}$$

$$\left. - (\bar{z}bmp_1^\epsilon)(x_s) \sum_{h=1}^A \frac{h^2 M(x_s,h)\partial_1 f(x_s,x_s)}{b(x_s)p_1^\epsilon(x_s)} \right| ds$$

$$\leq C(C_{\text{Lip}}^a + AC_{\text{Lip}}^r)T\eta + C\int_0^t |x_s - x_1(s)|ds$$

$$+ C\int_0^t \sum_{h=1}^A \left| r_1^\epsilon(x_s,h) - \frac{hM(x_s,h)\partial_1 f(x_s,x_s)}{b(x_s)p_1^\epsilon(x_s)} \right| ds,$$

where the last inequality follows from the uniform Lipschitz-continuity of all functions involved in the computation. Now, $|p_2^\epsilon(x) - p_1^\epsilon(x)| \leq C\epsilon$ and $p_j^\epsilon(x) \geq c > 0$ for $j = 1, 2$, for some constants $C, c > 0$ independent of ϵ and x. Hence, there exists a constant C such that

$$|x_t - x_1(t)| \leq CT(\eta + \epsilon) + C\int_0^t |x_s - x_1(s)|ds \tag{II.8.35}$$

$$+ C\int_0^t \sum_{h=1}^A \left| q_1^\epsilon(x_s,h) - h\frac{\partial_1 f(x_s,x_s)}{b(x_s)} \right| M(x_s,h)ds.$$

In view of (II.8.7), we obtain $|x_t - x_1(t)| \leq CT(\eta + \epsilon) + C\int_0^t |x_s - x_1(s)|ds$. Gronwall's lemma ends the proof of Lemma II.8.2. $\qquad\square$

In view of Lemma II.8.2, there exists $T_0 > 0$ independent of x, ϵ, η such that, for all $\eta \geq \epsilon$, $\sup_{t \in [0, T_0]} |x_j(t) - x_t| \leq \eta/4$. Let us fix $\eta = \epsilon$. Combining (II.8.30) with the last inequality entails (II.8.18).

II.8.4   End of the proof

Proof of Theorem II.4.1. Defining $\bar{\mu}^{K,1,\epsilon} = \mu^{K,1,\epsilon,\epsilon}$ and $\mu^{K,2,\epsilon} = \mu^{K,2,\epsilon,\epsilon}$ and combining (II.8.17) and (II.8.18), we see that we have defined a constant, $T_0 > 0$, such that

$$\lim_{K \to +\infty} \mathbb{P}\left[\bar{\mu}^{K,1,\epsilon}_{(t-1/(K u_K \sigma_K)) \vee 0} \preccurlyeq \mu^{1,K}_t \preccurlyeq \nu^K_t \preccurlyeq \mu^{2,K}_t \preccurlyeq \mu^{K,2,\epsilon}_t, \ \forall t \leq \tfrac{T_0}{K u_K \sigma_K^2} \right] = 1, \tag{II.8.36}$$

This is (II.8.1) with $\mu^{K,1,\epsilon}_t = \bar{\mu}^{K,1,\epsilon}_{(t-1/(K u_K \sigma_K)) \vee 0}$. It only remains to check (II.8.2).

Using that $\eta = \epsilon$, we get

$$\left\| \mu^{K,1,\epsilon}_{t/K u_K \sigma_K^2} - \bar{z}(x(t)) \delta_{x(t)} \right\|_0 \tag{II.8.37}$$

$$\leq C \left(\epsilon + \sigma_K + \left| \bar{z}(x_t) - \bar{z}\left(X^{K,1,\eta}_{(t-\sigma_K) \vee 0 / K u_K \sigma_K^2} \right) \right| + \left| x_t - x_1\big((t - \sigma_K) \vee 0\big) \right| \right.$$

$$\left. + \left| X^{K,1,\eta}_{(t-\sigma_K) \vee 0 / K u_K \sigma_K^2} - x_1\big((t - \sigma_K) \vee 0\big) \right| \right)$$

$$\leq C' \left(\epsilon + \sigma_K + \sup_{t \in [0,T]} \left(|x_{(t-\sigma_K) \vee 0} - x_t| + |x_t - x_1(t)| + \left| X^{K,1,\eta}_{t/K u_K \sigma_K^2} - x_1(t) \right| \right) \right),$$

for some finite constants $C, C' > 0$. The analogous estimate holds for for $\mu^{2,K,\eta}_{t/K u_K \sigma_K^2}$. Setting for example $\delta(\epsilon) = \sqrt{\epsilon}$, (II.8.2) follows from (II.8.30), Lemma II.8.2 and the uniform continuity of x_t. This ends the proof of Theorem II.4.1. $\square$

II.9   Appendix

In this section, we state and prove several elementary results, which we used in the proof of our main theorem. Recall that $\| . \|_0$ is the Kantorovich-Rubinstein norm on the vector space of finite, signed measures on $\mathcal{X}$, i.e.,

$$\|\mu_t\|_0 \equiv \sup \left\{ \int_\mathcal{X} f d\mu_t : f \in \mathrm{Lip}_1(\mathcal{X}) \ \text{with} \ \sup_{x \in \mathcal{X}} |f(x)| \leq 1 \right\}, \tag{II.9.1}$$

where $\mathrm{Lip}_1(\mathcal{X})$ is the space of Lipschitz continuous functions from $\mathcal{X}$ to $\mathbb{R}$. Let $\mathcal{M}_F(\mathcal{X})$ be the set of non-negative finite Borel-measures on $\mathcal{X}$.

Proposition II.9.1. *Let $\{\nu^K, K \geq 0\}$ and μ be random elements in $\mathbb{D}([0,T], \mathcal{M}_F(\mathcal{X}))$. If, for all $\delta > 0$,*

$$\lim_{K \to \infty} \mathbb{P}\left[\sup_{0 \leq t \leq T} \|\nu^K_t - \mu_t\|_0 > \delta \right] = 0, \tag{II.9.2}$$

then ν^K converges in probability, as $K \to \infty$, with respect to the Skorokhod topology on $\mathbb{D}([0,T], \mathcal{M}(\mathcal{X}))$ to μ.

Proof. Let us equip $\mathcal{M}_F(\mathcal{X})$ with the topology of weak convergence. Obverse that this topology is metrizable with the Kantorovich-Rubinstein norm, see [14] Vol. II, p. 193. Let Λ be

the class of strictly increasing, continuous mapping of $[0, T]$ onto itself. If $\lambda \in \Lambda$, then $\lambda(0) = 0$ and $\lambda(T) = T$. The Skorokhod topology on $\mathbb{D}\left([0, T], (\mathcal{M}_F(\mathcal{X}), \| . \|_0)\right)$ is generated by the distance

$$d(\mu, \nu) = \inf_{\lambda \in \Lambda} \left\{ \max \left\{ \sup_{t \in [0,T]} |\lambda(t) - t|, \sup_{t \in [0,T]} \|\mu_t - \nu_{\lambda t}\|_0 \right\} \right\}, \tag{II.9.3}$$

on $\mathbb{D}\left([0, T], (\mathcal{M}_F(\mathcal{X}), \| . \|_0)\right)$, see e.g. [12], Chap. 3. Since the identity lies in Λ it is clear that $d(\mu, \nu) \leq \sup_{t \in [0,T]} \|\mu_t - \nu_t\|_0$. Therefore, if a sequence of random elements with state space $\mathbb{D}([0, T], \mathcal{M}_F(\mathcal{X}))$ equipped with the metric induced by the norm $\sup_{t \in [0,T]} \|\mu_t\|_0$ convergences in probability to μ, it also convergences in probability to μ if $\mathbb{D}([0, T], \mathcal{M}_F(\mathcal{X}))$ is equipped with the metric d. $\qquad\square$

Proposition II.9.2. *Fix $\epsilon > 0$ and let σ_K a sequence in K with $K^{-1/2+\alpha} \ll \sigma_K \ll 1$. Let Z_n be a Markov chain with state space $\mathbb{N}_0$ and with the following transition probabilities*

$$\mathbb{P}[Z_{n+1} = j | Z_n = i] = p(i,j) = \begin{cases} 1, & \text{for } i = 0 \text{ and } j = 1, \\ \frac{1}{2} - C_1 i K^{-1} + C_2 \epsilon \sigma_K, & \text{for } i \geq 1 \text{ and } j = i + 1, \\ \frac{1}{2} + C_1 i K^{-1} - C_2 \epsilon \sigma_K, & \text{for } i \geq 1 \text{ and } j = i - 1, \end{cases} \tag{II.9.4}$$

for some constants $C_1 > 0$ and $C_2 \geq 0$. Let τ_i be the first hitting time of level i by Z and let $\mathbb{P}_a$ denote the law of Z conditioned on $Z_0 = a$. Then, for all $M \geq \frac{8C_2}{C_1}$ and for all $a \leq \frac{1}{3} M \epsilon \sigma_K K$

$$\lim_{K \to \infty} e^{K^{2\alpha}} \mathbb{P}_a\left[\tau_{\lceil M \epsilon \sigma_K K \rceil} < \tau_0\right] = 0. \tag{II.9.5}$$

Remark 9. The proposition can be seen as a moderate deviation result for this particular Markov chain. More precisely, we can prove that there exist two constants $M > 0$ and $C_3 > 0$ which depend only on C_1 and C_2 such that for $a < \frac{1}{3} M \epsilon \sigma_K K$

$$\mathbb{P}_a\left[\tau_{\lceil M \epsilon \sigma_K K \rceil} < \tau_0\right] \leq \exp\left(-C_3 K^{-1}\left(\left(\tfrac{1}{3} M \epsilon \sigma_K K\right)^2 - a^2\right)\right), \tag{II.9.6}$$

for all K large enough.

Proof. We calculate this probability with some standard potential theory arguments (cf. [17]). Let $h_{\lceil M \epsilon \sigma_K K \rceil, 0}(a)$ be the solution of the Dirichlet problem with $\lambda = 0$, i.e.,

$$\begin{aligned} \mathscr{L} h_{\lceil M \epsilon \sigma_K K \rceil, 0}(x) &= 0, & \text{for } 0 < x < \lceil M \epsilon \sigma_K K \rceil \\ h_{\lceil M \epsilon \sigma_K K \rceil, 0}(x) &= 1, & \text{for } x \geq \lceil M \epsilon \sigma_K K \rceil \\ h_{\lceil M \epsilon \sigma_K K \rceil, 0}(x) &= 0, & \text{for } x = 0. \end{aligned} \tag{II.9.7}$$

Therefore, we obtain for $0 < a < \lceil M \epsilon \sigma_K K \rceil$

$$\mathbb{P}_a\left[\tau_{\lceil M \epsilon \sigma_K K \rceil} < \tau_0\right] = h_{\lceil M \epsilon \sigma_K K \rceil, 0}(a) = \frac{\sum_{i=1}^{a} \frac{1}{\pi(i)} \frac{1}{p(i,i-1)}}{\sum_{i=1}^{\lceil M \epsilon \sigma_K K \rceil} \frac{1}{\pi(i)} \frac{1}{p(i,i-1)}}, \tag{II.9.8}$$

where $\pi = (\pi(0), \pi(1), \pi(2), \dots)$ is an invariant measure of the one-dimensional Markov chain Z_n. In our case any invariant measure π has to satisfy, for all $i \geq 1$,

$$\pi(0) = p(1,0)\pi(1) \quad \text{and} \quad \pi(i) = p(i-1,i)\pi(i-1) + p(i+1,i)\pi(i+1). \tag{II.9.9}$$

Therefore, π with $\pi(0) = 1, \pi(1) = \frac{1}{p(1,0)}$ and $\pi(i) = \prod_{j=1}^{i-1} \frac{p(j,j+1)}{p(j,j-1)} \frac{1}{p(i,i-1)}$ is the unique invariant measure for the Markov chain Z_n. Thus we get from (II.9.8) that

$$
\begin{aligned}
h_{\lceil M\epsilon\sigma_K K \rceil, 0}(a) &= \frac{\sum_{i=1}^{a} \prod_{j=1}^{i-1} \frac{p(j,j-1)}{p(j,j+1)}}{\sum_{i=1}^{\lceil M\epsilon\sigma_K K \rceil} \prod_{j=1}^{i-1} \frac{p(j,j-1)}{p(j,j+1)}} \\
&= \frac{\sum_{i=1}^{a} \exp\left(\sum_{j=1}^{i-1} \ln\left(\frac{1+2C_1 K^{-1}j - 2C_2\epsilon\sigma_K}{1-2C_1 K^{-1}j + 2C_2\epsilon\sigma_K}\right)\right)}{\sum_{i=1}^{\lceil M\epsilon\sigma_K K \rceil} \exp\left(\sum_{j=1}^{i-1} \underbrace{\ln\left(\frac{1+2C_1 K^{-1}j - 2C_2\epsilon\sigma_K}{1-2C_1 K^{-1}j + 2C_2\epsilon\sigma_K}\right)}_{=:f(j)}\right)} .
\end{aligned}
\tag{II.9.10}
$$

For all $j \leq M\epsilon\sigma_K K$ we can approximate $f(j)$ as follows

$$
\begin{aligned}
f(j) &= \ln\left(1 + \frac{4C_1 K^{-1}j - 4C_2\epsilon\sigma_K}{1 - 2C_1 K^{-1}j + 2C_2\epsilon\sigma_K}\right) \\
&= \frac{4C_1 K^{-1}j - 4C_2\epsilon\sigma_K}{1 - 2C_1 K^{-1}j + 2C_2\epsilon\sigma_K} - O\left(\left(\frac{4C_1 K^{-1}j - 4C_2\epsilon\sigma_K}{1 - 2C_1 K^{-1}j + 2C_2\epsilon\sigma_K}\right)^2\right) \\
&= 4C_1 \frac{j}{K} - 4C_2\epsilon\sigma_K + O\left(\left(\frac{j}{K}\right)^2 + \epsilon\sigma_K \frac{j}{K} + \epsilon^2\sigma_K^2\right) \\
&= 4C_1 \frac{j}{K} - 4C_2\epsilon\sigma_K + O\left((M\epsilon\sigma_K)^2\right)
\end{aligned}
\tag{II.9.11}
$$

Therefore,

$$
\begin{aligned}
h_{\lceil M\epsilon\sigma_K K \rceil, 0}(a) &\leq \frac{\sum_{i=1}^{a} \exp\left(\sum_{j=1}^{i-1} 4C_1 \frac{j}{K} + O\left((M\epsilon\sigma_K)^2\right)\right)}{\sum_{i=1}^{\lceil M\epsilon\sigma_K K \rceil} \exp\left(\sum_{j=1}^{i-1} 4C_1 \frac{j}{K} - 4C_2\epsilon\sigma_K - O\left((M\epsilon\sigma_K)^2\right)\right)} \\
&\leq \frac{a \exp\left(2C_1 a^2 K^{-1} + O\left(a(M\epsilon\sigma_K)^2\right)\right)}{\sum_{i=1}^{\lceil M\epsilon\sigma_K K \rceil} \exp\left(2C_1 K^{-1}(i^2 - i) - 4C_2\epsilon\sigma_K i - O\left((i-1)(M\epsilon\sigma_K)^2\right)\right)} \\
&\leq \frac{a \exp\left(2C_1 a^2 K^{-1} + O\left(a(M\epsilon\sigma_K)^2\right)\right)}{\sum_{i=\frac{1}{2}\lceil M\epsilon\sigma_K K \rceil}^{\lceil M\epsilon\sigma_K K \rceil} \exp\left(2C_1 K^{-1} i^2 - (2C_1 K^{-1} + 4C_2\epsilon\sigma_K)i - O\left(i(M\epsilon\sigma_K)^2\right)\right)} .
\end{aligned}
\tag{II.9.12}
$$

Choosing $M \geq \frac{8C_2}{C_1}$, if $a < \frac{M\epsilon\sigma_K K}{3}$, then

$$
\begin{aligned}
h_{\lceil M\epsilon\sigma_K K \rceil, 0}(a) &\leq \frac{a \exp\left(2C_1 a^2 K^{-1} + O\left(a(M\epsilon\sigma_K)^2\right)\right)}{\frac{1}{2}\lceil M\epsilon\sigma_K K \rceil \exp\left(\left(\frac{1}{2}C_1 M - 2C_2\right)M\epsilon^2\sigma_K^2 K - O\left((\epsilon\sigma M)^3 K + \epsilon\sigma_K M\right)\right)} \\
&\leq 2a(\lceil M\epsilon\sigma_K K \rceil)^{-1} \exp\left(C_1 K^{-1}\left(2a^2 - \frac{1}{4}(\lceil M\epsilon\sigma_K K \rceil)^2\right)\right) \\
&\leq \exp\left(-C_3 K^{-1}\left((\frac{1}{3}\lceil M\epsilon\sigma_K K \rceil)^2 - a^2\right)\right).
\end{aligned}
\tag{II.9.13}
$$

Since $K^{-1/2+\alpha} \ll \sigma_K$ when K tends to infinity, (II.9.5) follows. $\qquad\square$

Proposition II.9.3. *Let $(Z_t)_{t\geq 0}$ be a branching process with birth rate per individual b and death rate per individual d. Let τ_i be the first hitting time of level i by Z and let $\mathbb{P}_j$ denote the law of Z conditioned on $Z_0 = j$, and $\mathbb{E}_j$ the corresponding expectation. Then*

$$
\mathbb{P}_j\left[\tau_k < \tau_0\right] = \frac{(d/b)^j - 1}{(d/b)^k - 1} \qquad \text{for all } 1 \leq j \leq k-1, \tag{II.9.14}
$$

$$
\left|\mathbb{P}_1\left[\tau_k < \tau_0\right] - \frac{[b-d]_+}{b}\right| \leq k^{-1} \qquad \text{and} \tag{II.9.15}
$$

$$
\mathbb{E}_1\left[\tau_k \wedge \tau_0\right] \leq \frac{1 + \ln(k)}{b}, \tag{II.9.16}
$$

where $[b-d]_+ \equiv \max\{b-d,0\}$. *Moreover, if Z_t is slightly super-critical, i.e., $b = d + \epsilon$, then*

$$\max_{n \le k} \frac{\mathbb{E}_n[\tau_k \wedge \tau_0]}{\mathbb{P}_n[\tau_k < \tau_0]} \le \frac{1 + \ln(k)}{\epsilon} \tag{II.9.17}$$

Proof. Let $p_j \equiv \mathbb{P}_j[\tau_k < \tau_0]$. Then $p_0 = 0$, $p_k = 1$, and $p_j = \frac{b}{b+d} p_{j+1} + \frac{d}{b+d} p_{j-1}$ for all $1 \le j \le k-1$ by the Markov property. From this recursion, we obtain the characteristic polynomial

$$P(x) = bx^2 - (b+d)x + d. \tag{II.9.18}$$

With its roots 1 and d/b, we obtain the following general solution for the recursion

$$p_n = \kappa_0 \cdot 1^n + \kappa_1 \left(\frac{d}{b}\right)^n, \tag{II.9.19}$$

where κ_0 and κ_1 are constants. From the initial condition $p_0 = 0$ and $p_k = 1$, we obtain $\kappa_0 = -\left(\left(\frac{d}{b}\right)^k - 1\right)^{-1}$ and $\kappa_1 = \left(\left(\frac{d}{b}\right)^k - 1\right)^{-1}$. Therefore,

$$p_n = \frac{\left(\frac{d}{b}\right)^n - 1}{\left(\frac{d}{b}\right)^k - 1} \quad \text{and} \quad p_1 = \frac{\frac{d}{b} - 1}{\left(\frac{d}{b}\right)^k - 1} = \frac{1}{1 + \frac{d}{b} + \ldots + \left(\frac{d}{b}\right)^{k-1}}. \tag{II.9.20}$$

If $d \ge b$, this computation implies that $p_1 \equiv \mathbb{P}_1[\tau_k < \tau_0] \le 1/k$ and $[b-d]_+ = 0$. If $d < b$,

$$\mathbb{P}_1[\tau_k < \tau_0] - \frac{b-d}{b} = \frac{\frac{d}{b} - 1}{\left(\frac{d}{b}\right)^k - 1} - \left(1 - \frac{d}{b}\right) \frac{\left(\frac{d}{b}\right)^k - 1}{\left(\frac{d}{b}\right)^k - 1} = \frac{\left(\frac{d}{b} - 1\right)\left(\frac{d}{b}\right)^k}{\left(\frac{d}{b}\right)^k - 1} = \frac{\frac{d}{b} - 1}{1 - \left(\frac{b}{d}\right)^k} \tag{II.9.21}$$

$$= \frac{\frac{d}{b}\left(1 - \frac{b}{d}\right)}{1 - \left(\frac{b}{d}\right)^k} = \frac{1}{\frac{b}{d}\left(1 + \frac{b}{d} + \ldots + \left(\frac{b}{d}\right)^{k-1}\right)} = \frac{1}{\frac{b}{d} + \ldots + \left(\frac{b}{d}\right)^k}$$

$$\le \frac{1}{k}.$$

Similarly, if $e_n \equiv \mathbb{E}_n[\tau_k \wedge \tau_0]$, then e_n is the solution of the following non-homogeneous Dirichlet problem:

$$\mathscr{L} e_n = -1, \qquad \text{for } n \in \{1, .., k-1\} \tag{II.9.22}$$
$$e_n = 0, \qquad \text{for } n \in \mathbb{N}_0 \smallsetminus \{1, .., k-1\},$$

where $(\mathscr{L}f)(x) = x\big(b[f(x+1) - f(x)] + d[f(x-1) - f(x)]\big)$ is the generator of the branching process Z. Therefore, we have to solve the following non-homogeneous recurrence

$$e_{n+2} - \frac{b+d}{b} e_{n+1} + \frac{d}{b} e_n = \frac{-1}{b(n+1)} \quad \text{and} \quad e_0 = e_k = 0 \tag{II.9.23}$$

We solve this by variation of parameters. Thus, we first solve the associated linear homogeneous recurrence relation:

$$h_{n+2} - \frac{b+d}{b} h_{n+1} + \frac{d}{b} h_n = 0 \tag{II.9.24}$$

As we have seen before $h_n = \kappa_2 1 + \kappa_3 \left(\frac{d}{b}\right)^j$ for any $\kappa_2, \kappa_3 \in \mathbb{R}$ solves the equation. Obverse that this functions are the harmonic functions of $\mathscr{L}$. Second, we have to find a particular solution. Let (x_{1j}, x_{2j}) the solution of the system of linear equations

$$x_{1j} + \left(\frac{d}{b}\right)^{j+1} x_{2j} = 0 \tag{II.9.25}$$
$$x_{1j} + \left(\frac{d}{b}\right)^{j+2} x_{2j} = -\frac{1}{b(j+1)}, \tag{II.9.26}$$

then

$$e_n^p = \sum_{j=0}^{n-1} x_{1j} 1^n + \sum_{j=0}^{n-1} x_{2j} \left(\frac{d}{b}\right)^n = \frac{-1}{b-d} \sum_{j=1}^{n} \frac{1}{j} + \frac{1}{b-d} \sum_{j=1}^{n} \frac{1}{j} \left(\frac{b}{d}\right)^j \left(\frac{d}{b}\right)^n \tag{II.9.27}$$

$$= \frac{1}{b-d} \sum_{j=1}^{n} \frac{1}{j} \left(\left(\frac{d}{b}\right)^{n-j} - 1\right)$$

is a particular solution. Now, we obtain we obtain the following general solution for the recurrence:

$$e_n = h_n + e_n^p = \kappa_2 + \kappa_3 \left(\frac{d}{b}\right)^n + \frac{1}{b-d} \sum_{j=1}^{n} \frac{1}{j} \left(\left(\frac{d}{b}\right)^{n-j} - 1\right). \tag{II.9.28}$$

We have the boundary condition $e_0 = e_k = 0$, therefore κ_2 and κ_3 are given by the solution of the following system of linear equations

$$\kappa_2 + \kappa_3 \left(\frac{d}{b}\right)^0 + \frac{1}{b-d} \sum_{j=1}^{0} \frac{1}{j} \left(\left(\frac{d}{b}\right)^{0-j} - 1\right) = 0, \tag{II.9.29}$$

$$\kappa_2 + \kappa_3 \left(\frac{d}{b}\right)^k + \frac{1}{b-d} \sum_{j=1}^{k} \frac{1}{j} \left(\left(\frac{d}{b}\right)^{k-j} - 1\right) = 0, \tag{II.9.30}$$

and we obtain that

$$e_n = \frac{1}{b-d} \sum_{j=1}^{k} \frac{1}{j} \frac{\left(\frac{d}{b}\right)^{k-j} - 1}{\left(\frac{d}{b}\right)^k - 1} - \frac{1}{b-d} \sum_{j=1}^{k} \frac{1}{j} \frac{\left(\frac{d}{b}\right)^{k-j} - 1}{\left(\frac{d}{b}\right)^k - 1} \left(\frac{d}{b}\right)^n + \frac{1}{b-d} \sum_{j=1}^{n} \frac{1}{j} \left(\left(\frac{d}{b}\right)^{n-j} - 1\right) \tag{II.9.31}$$

$$= \frac{1}{b-d} \sum_{j=1}^{k} \frac{1}{j} \frac{\left(\left(\frac{d}{b}\right)^{k-j} - 1\right)\left(1 - \left(\frac{d}{b}\right)^n\right)}{\left(\frac{d}{b}\right)^k - 1} + \frac{1}{b-d} \sum_{j=1}^{n} \frac{1}{j} \left(\left(\frac{d}{b}\right)^{n-j} - 1\right).$$

With this formula we can easily prove the second inequality of the proposition,

$$e_1 = \frac{1}{b-d} \sum_{n=1}^{k} \frac{1}{n} \frac{\left(\frac{d}{b}\right)^{k-n} - 1}{\left(\frac{d}{b}\right)^k - 1} \left(1 - \frac{d}{b}\right) + 0 \le \frac{1}{b} \sum_{n=1}^{k} \frac{1}{n} \le \frac{1 + \ln(k)}{b}. \tag{II.9.32}$$

Finally, we obtain for slightly super-critical Z_t, i.e., with $b = d + \epsilon$,

$$\frac{\mathbb{E}_n[\tau_k \wedge \tau_0]}{\mathbb{P}_n[\tau_k < \tau_0]} = \frac{e_n}{p_n} = \frac{1}{b-d} \sum_{j=1}^{k} \frac{1}{j} \underbrace{\left(\left(\frac{d}{b}\right)^{k-j} - 1\right)(-1)}_{\le 1} + \frac{1}{b-d} \sum_{j=1}^{n} \frac{1}{j} \underbrace{\frac{\left(\left(\frac{d}{b}\right)^{n-j} - 1\right)\left(1 - \left(\frac{d}{b}\right)^k\right)}{1 - \left(\frac{d}{b}\right)^n}}_{\le 0} \tag{II.9.33}$$

$$\le \frac{1}{\epsilon} \sum_{j=1}^{k} \frac{1}{j} \le \frac{1 + \ln(k)}{\epsilon},$$

which proves (II.9.17). $\square$

Proposition II.9.4. *Let $(Z_t^K)_{t\ge0}$ be a sequence branching process with birth rate per individual $b \ge 0$ and death rate per individual $d \ge 0$ and $|b - d| = O(\sigma_K)$, where $K^{-1/2+\alpha} \ll \sigma_K \ll 1$. Let τ_i be the first hitting time of level i by Z and let $\mathbb{P}_j$ denote the law of Z conditioned on $Z_0 = j$.*

(a) The invasion probability can be approximated up to an error of order $\exp(-K^\alpha)$:

$$\lim_{K \to \infty} \exp(K^\alpha) \left| \mathbb{P}_1 \left[\tau_{\lceil \epsilon \sigma_K K \rceil} < \tau_0 \right] - \frac{[b-d]_+}{b} \right| = 0. \tag{II.9.34}$$

(b) If $b > d$ (super-critical case), we have exponential tails, i.e.,

$$\lim_{K \to \infty} \exp(\sigma_K^{-\alpha/3}) \mathbb{P}_1 \left[\tau_{\lceil \epsilon \sigma_K K \rceil} > \ln(K) \sigma_K^{-1-\alpha/2} \Big| \tau_{\lceil \epsilon \sigma_K K \rceil} < \tau_0 \right] = 0 \tag{II.9.35}$$

and

$$\lim_{K \to \infty} \exp(K^\alpha) \mathbb{P}_{\lceil \epsilon \sigma_K K \rceil} \left[\tau_{\lceil \epsilon K \rceil} > \tau_0 \right] = 0 \tag{II.9.36}$$

Proof. (a) Compare with (II.9.14) that

$$\mathbb{P}_1 \left[\tau_{\lceil \epsilon \sigma_K K \rceil} < \tau_0 \right] = \frac{(d/b) - 1}{(d/b)^{\lceil \epsilon \sigma_K K \rceil} - 1}. \tag{II.9.37}$$

If $b > d$ (sub-critical case), there exist two constants $\underline{C}^{\mathrm{sub}} > 0$ and $\bar{C}^{\mathrm{sub}} > 0$ such that $1 + \underline{C}^{\mathrm{sub}} \sigma_K \le d/b \le 1 + \bar{C}^{\mathrm{sub}} \sigma_K$. Therefore, the left hand side of (II.9.37) does not exceed

$$\frac{\bar{C}^{\mathrm{sub}} \sigma_K}{(1 + \underline{C}^{\mathrm{sub}} \sigma_K)^{\lceil \epsilon \sigma_K K \rceil} - 1} \le \frac{\bar{C}^{\mathrm{sub}} \sigma_K}{\exp(\underline{C}^{\mathrm{sub}} \sigma_K \lceil \epsilon \sigma_K K \rceil - O(\sigma_K^3 \epsilon K)) - 1} = o(e^{-K^\alpha}). \tag{II.9.38}$$

The last equality holds, since $K^{2\alpha} \ll \sigma_K^2 K$. If $b > d$ (super-critical case), we obtain similarly

$$\left| \mathbb{P}_1 [\tau_k < \tau_0] - \frac{b - d}{b} \right| = \left| \frac{\frac{d}{b} - 1}{1 - (\frac{b}{d})^k} \right| = o(\exp(-K^\alpha)). \tag{II.9.39}$$

(b) Compare with [2] page 41, that

$$\mathbb{P}_1 \left[\tau_{\lceil \epsilon \sigma_K K \rceil} > \ln(K) \sigma_K^{-1-\alpha/2} \Big| \tau_{\lceil \epsilon \sigma_K K \rceil} < \tau_0 \right] \tag{II.9.40}$$

$$\le \exp\left(- \left\lfloor \frac{\ln(K) \sigma_K^{-1-\alpha/2}}{e \max_{n \le \lceil \epsilon \sigma_K K \rceil} \mathbb{E}_n \left[\tau_{\lceil \epsilon \sigma_K K \rceil} \big| \tau_{\lceil \epsilon \sigma_K K \rceil} < \tau_0 \right]} \right\rfloor \right) \le \exp\left(-\sigma_K^{-\alpha/3} \right),$$

where the last inequality holds, because we can apply Proposition II.9.3

$$\max_{n \le \lceil \epsilon \sigma_K K \rceil} \mathbb{E}_n \left[\tau_{\lceil \epsilon \sigma_K K \rceil} \big| \tau_{\lceil \epsilon \sigma_K K \rceil} < \tau_0 \right] = \max_{n \le \lceil \epsilon \sigma_K K \rceil} \frac{\mathbb{E}_n \left[\tau_{\lceil \epsilon \sigma_K K \rceil} \wedge \tau_0 \mathbb{1}_{\tau_0 > \tau_{\lceil \epsilon \sigma_K K \rceil}} \right]}{\mathbb{P}_n \left[\tau_0 > \tau_{\lceil \epsilon \sigma_K K \rceil} \right]} \tag{II.9.41}$$

$$\le O(\ln(K) \sigma_K^{-1}).$$

On the other hand, we have

$$\mathbb{P}_{\lceil \epsilon \sigma_K K \rceil} \left[\tau_{\lceil \epsilon K \rceil} > \tau_0 \right] = 1 - \frac{(d/b)^{\lceil \epsilon \sigma_K K \rceil} - 1}{(d/b)^{\lceil \epsilon K \rceil} - 1} \le \exp(-K^{2\alpha}) \tag{II.9.42}$$

since $d/b = 1 - O(\sigma_K)$ and $K^{2\alpha} \ll \sigma_K \epsilon K$. $\qquad \square$

Proposition II.9.5. *Let $(Z_n^K)_{n \ge 0}$ a sequence of discrete time Markov chain with state space $\mathbb{Z}$ and with transition probabilities*

$$\mathbb{P}[Z_{n+1}^K = j | Z_n^K = i] = p(i, j) = \begin{cases} \frac{1}{2} + C\sigma_K, & \text{if } j = i + 1, \\ \frac{1}{2} - C\sigma_K, & \text{if } j = i - 1, \\ 0, & \text{else}, \end{cases} \tag{II.9.43}$$

for some constant $C \ne 0$. Let τ_i be the first hitting time of level i by Z^K and let $\mathbb{P}_j$ denote the law of Z^K conditioned on $Z_0^K = j$ and let σ_K a zero sequence such that $K^{-\frac{1}{2}+\alpha} \ll \sigma_K \ll 1$.

(a) *If Z^K is slightly supercritical, i.e., $C > 0$, then, for all $i \geq 1$*

$$\lim_{K \to \infty} \exp(K^\alpha) \, \mathbb{P}_{i \lceil (\epsilon/2)\sigma_K K \rceil} \left[\tau_{(i-1)\lceil (\epsilon/2)\sigma_K K \rceil} < \tau_{(i+1)\lceil (\epsilon/2)\sigma_K K \rceil} \right] = 0. \qquad \text{(II.9.44)}$$

(b) *If Z^K is slightly subcritical, i.e., $C < 0$, then, for all constants $C_1, C_2, C_3 > 0$*

$$\lim_{K \to \infty} \exp(K^\alpha) \, \mathbb{P}_{(C_1+C_2)\lceil \epsilon\sigma_K K \rceil} \left[\tau_{(C_1+C_2+C_3)\lceil \epsilon\sigma_K K \rceil} < \tau_{C_1 \lceil \epsilon\sigma_K K \rceil} \right] = 0. \qquad \text{(II.9.45)}$$

Proof. Since the transition probabilities of Z^K do not depend on the state of Z^K, we have that

$$\mathbb{P}_{i \lceil (\epsilon/2)\sigma_K K \rceil} \left[\tau_{(i-1)\lceil (\epsilon/2)\sigma_K K \rceil} > \tau_{(i+1)\lceil (\epsilon/2)\sigma_K K \rceil} \right] = \mathbb{P}_{\lceil (\epsilon/2)\sigma_K K \rceil} \left[\tau_0 > \tau_{2\lceil (\epsilon/2)\sigma_K K \rceil} \right] \qquad \text{(II.9.46)}$$

By (II.9.14) the left side of (II.9.46) is equal

$$\frac{1 - (1 - 2C\sigma_K + O(\sigma_K^2))^{\lceil (\epsilon/2)\sigma_K K \rceil}}{1 - (1 - 2C\sigma_K + O(\sigma_K^2))^{2\lceil (\epsilon/2)\sigma_K K \rceil}} \geq 1 - \exp(-K^2\alpha), \qquad \text{(II.9.47)}$$

since $\sigma_K^2 K \gg K^{2\alpha}$. With the same arguments, we obtain also (II.9.45). $\qquad \square$

Chapter III

A stochastic model for immunotherapy of cancer and the polymorphic evolution sequence for populations with phenotypic plasticity

In this chapter we propose an extension of the individual-based model in population dynamics introduced in Section I.2, which broadens the range of biological applications. The primary motivation is modeling of immunotherapy of malignant tumors. The main expansions are that we have three different actors in this context (T-cells, cytokines, and cancer cells), that we distinguish cancer cells by phenotype and genotype, that we include environment-dependent phenotypic plasticity, and that we take into account the therapy effects. We illustrate the new setup by using it to model various phenomena arising in immunotherapy and we argue why stochastic models may help to understand the resistance of tumors to therapeutic approaches and thus may have non-trivial consequences on tumor treatment protocols. Furthermore, we show that the interplay of genetic mutations and phenotypic switches on different time scales as well as the occurrence of metastability phenomena raise new mathematical challenges. In the present thesis we focuses more on these theoretical aspects which arise by including phenotypic plasticity in the standard individual-based model describing the evolution of an asexual reproducing, competitive population. More precisely, we study the behavior of this process on a large (evolutionary) time scale and in the simultaneous limits of large population size ($K \to \infty$) and rare mutations ($u_K \to 0$), proving convergence to a Markov pure jump process, which can be seen as a generalization of the polymorphic evolution sequence (cf. [25, 30]).

Parts of the presented results were previously published in Scientific Reports [9] as a joint work with L. Coquille, H. Mayer, M. Hölzel, M. Rogava, T. Tüting, and A. Bovier (cf. Section I.3 for details).

III.1   Introduction

The treatment of various cancers with immunotherapies received a lot of attention in the medical as well as the mathematical modeling communities during the last decades [108, 88, 50, 72, 65, 76]. Many different therapeutic approaches were developed and tested experimentally. As for the classical therapies such as chemo- and radiotherapy, *resistance* is an important issue also for immunotherapy: although a therapy leads to an initial phase of remission, very often

a relapse occurs. The main driving forces for resistance are considered to be the genotypic and phenotypic heterogeneity of tumors, which may be enhanced during therapy, see [76, 96, 65] and references therein. A tumor is a complex tissue which evolves in mutual influence with its environment [32]. In this chapter, we consider the example of melanoma (tumors associated to skin cancer) under T-cell therapy. Our work is motivated by the experiments of Landsberg et al. [92], which investigate melanoma in mice under *adoptive cell transfer* (ACT) therapy. This therapeutic approach involves the injection of T-cells which recognize a melanocyte-specific antigen and are able to kill differentiated types of melanoma cells. The therapy induces an inflammation and the melanoma cells react to this environmental change by switching their phenotype, i.e. by passing from a differentiated phenotype to a dedifferentiated one (special markers on the cell surface disappear). The T-cells recognize the cancerous cells through the markers which are down regulated in the dedifferentiated types. Thus, they are not capable of killing the dedifferentiated cancer cells anymore and a relapse is often observed. The phenotype switch is enhanced, if pro-inflammatory cytokines, called TNF-α (Tumor Necrosis Factor), are present. A second reason for the appearance of a relapse is that the T-cells become exhausted and are not working efficiently anymore. This problem was addressed by re-stimulation of the T-cells, but this led only to a delay in the occurrence of the relapse. Of course, other immune cells and cytokines are also present. However, according to the careful control experiments, their influence can be neglected in the context of the phenomena considered here. Cell division is not required for switching, and switching is reversible. This means that the melanoma cells can recover their initial (differentiated) phenotype [92]. The switch is thus a purely phenotypic change which is not induced by a mutation. Figure III.1 is a graphical representation of the relevant underlying mechanisms, reported in [92].
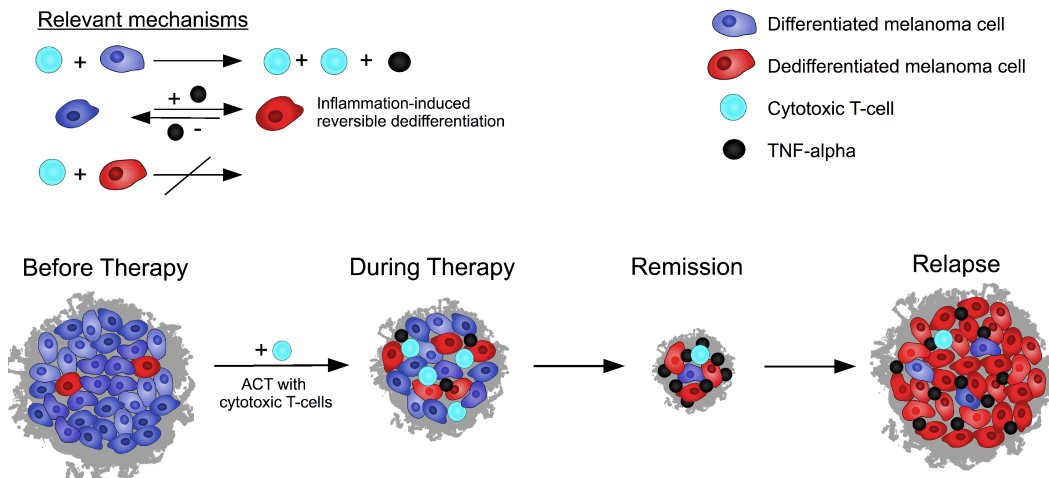
Figure III.1: Dynamics of the experiments described in [92].

In this chapter, we propose a quantitative mathematical model that can reproduce the phenomena observed in the experiments of [92], and which allows to simulate different therapy protocols, including some where several types of T-cells are used. It is an extension of the individual-based stochastic models of adaptive dynamics, introduced in Metz et al. [104] and developed and analyzed by many authors in recent years (see e.g. [15, 16, 43, 27, 25, 28, 19, 29, 33]), to the setting of tumor growth under immunotherapy. More precisely, the main expansions are:

(i) Three different classes of actors are included: T-cells, cytokines, and cancer cells.

(ii) For cancer cells two types of transitions are allowed: genotypic mutations and phenotypic switches.

(iii) Phenotypic changes can be affected by the environment which is not modeled deterministically as in [29] but as particles undergoing the random dynamics as well.

(iv) For modeling the therapy effect, a predator-prey mechanism (between cancer cells and immune cells) is included.

(v) A birth-reducing competition term is included which takes into account that competition may also affect the reproduction behavior.

In general, these class of stochastic models describe the evolution of interacting cell populations, in which the relevant events for each individual (e.g. birth and death) occur randomly. It is well known that in the limit of large cell-populations, these models are approximated by deterministic kinetic rate models (cf. Theorem III.2.1), which are widely used in the modeling of cell populations. However, these approximations are inaccurate and fail to account for important phenomena if the numbers of individuals in some sub-populations become small. In such situations, random fluctuation may become highly significant and completely alter the long-term behavior of the system. For example, in a phase of remission during therapy, the cancer and the T-cell populations drop to a low level and may die out due to fluctuations. A number of (mostly deterministic) models have been proposed that describe the development of a tumor under treatment, focusing on different aspects. For example, a deterministic model for ACT therapy is presented in [50]. Stochastic approaches were used to understand certain aspects of tumor development, for example rate models [69] or multi-type branching processes; see the book by Durrett [49] or [20, 3, 48]. To our knowledge, however, it is a novel feature of our models to describe the coevolution of immune- and tumor cells taking into account both interactions and phenotypic plasticity. Our models can help to understanding the interplay of therapy and resistance, in particular in the case of immunotherapy, and may be used to predict successful therapy protocols.

Besides being able to describe the experiments and making predictions about therapy protocols, we are also interested in more theoretical aspects which arise by including switching rates in the standard model, more precisely, in the interplay between the fast phenotypical changes by switching and the slow genotypical changes by mutation. In this context, the typical questions of adaptive dynamics arise again. E.g., can we describe the evolution of the system by successive mutant invasions, or rather, under which conditions does the microscopic process which incorporates fast phenotypical switches converge in the limit of a large population size in combination with only rare mutational events to a Markov jump process and how does this jump process look like. In fact, we prove by expanding the techniques of [30] that the microscopic process converges in this limit on the evolutionary time scale to a generalization of the Polymorphic Evolution Sequences (PES) introduced in [30] (cf. Theorem III.4.3). The main difference in the proof is that we have to couple the process with multi-type branching processes instead of normal branching processes, which leads also to a different definition of invasion fitness in this setting.

The remainder of this chapter is structured as follows. In Section III.2 we define the model and state the convergence towards a quadratic system of ODEs in the large population limit. In Section III.3, we present an example which qualitatively models the therapy carried out in Landsberg et al. [92]. We point out a phenomenon of relapse caused by random fluctuations. In Section III.4 we consider the case of rare mutations. We start with giving a pathwise definition of the individual-based model which is only extended by phenotypic plasticity (cf. Subsection III.4.1). In Subsection III.4.2 we state and prove the convergence of

this process to the Polymorphic Evolution Sequences with phenotypic Plasticity (PESP). In Subsection III.4.3, we study the interplay of mutation and therapy. We consider birth-reducing competition between tumor cells and show that the appearance of a mutant genotype may be enhanced under treatment.

III.2   The microscopic model

Let us introduce the model we analyze. Since we want to be able to model the evolution of the tumor during immunotherapy, we consider a tumor system that is composed of a finite number of cancer cells, T-cells, and cytokines, where each cancer cell is characterized by a tuple consisting of genotype and phenotype and each T-cell is characterized by its specificity. More precisely, we introduce a general model which contains three types of actors:

Cancer cells: each cell is characterized by a genotype and a phenotype. These cells can divide (with or without mutation), die (due to age, competition or therapy), and switch their phenotype. We assume that the switch is inherited by the descendants of the switched cells.

T-cells: each cell can divide, die, and produce cytokines.

Cytokines: each messenger can vanish and influence the switching of cancer cells.

Thus, the trait space, $\mathcal{X}$, of the population is in this chapter a finite set of the form

$$\mathcal{X} = \mathcal{G} \times \mathcal{P} \uplus \mathcal{Z} \uplus \mathcal{W} = \left\{g_1, \ldots, g_{|\mathcal{G}|}\right\} \times \left\{p_1, \ldots, p_{|\mathcal{P}|}\right\} \uplus \left\{z_1, \ldots, z_{|\mathcal{Z}|}\right\} \uplus \left\{w_1, \ldots, w_{|\mathcal{W}|}\right\} \quad \text{(III.2.1)}$$

where $\mathcal{G}$ is the set of cancer genotypes, $\mathcal{P}$ is the set of cancer phenotypes, $\mathcal{Z}$ is the set of T-cell types, and $\mathcal{W}$ the set cytokine types.

We write $|\cdot|$ for the number of elements of a set and $\uplus$ for disjoint unions of sets. The relation between $\mathcal{G}$ and $\mathcal{P}$ is encoded in the switch kernels (see below). They specify which phenotypes are expressed by a given genotype and influence the proportions of the different phenotypes in a (dynamic) environment.

In the following, we introduce the biological parameters (separately for each type of actors, i.e $\mathcal{G} \times \mathcal{P}$, $\mathcal{Z}$, and $\mathcal{W}$) which determine the dynamics of the population:

$K \in \mathbb{N}$ is a parameter scaling the population size and the resources. It is usually called *carrying-capacity* of the environment.

Cancer cells: For any $(g, p) \in \mathcal{G} \times \mathcal{P}$,

$b(p) \in \mathbb{R}_+$ is the *rate of birth* of a cancer cell with phenotype p.

$d(p) \in \mathbb{R}_+$ is the *rate of natural death* of a cancer cell with phenotype p.

$c(p, \tilde{p}) K^{-1} \in \mathbb{R}_+$ and $c_b(p, \tilde{p}) K^{-1} \in \mathbb{R}_+$ are the *competition kernels* which model the competitive pressure felt by a cancer cell with phenotype p from one with phenotype $\tilde{p}$.

The first term results in a higher death rate and the second term, called birth-reducing competition, in a lower birth rate (it inhibits cell division). If the total birth rate is already at a level zero, then $c_b(p, \tilde{p}) K^{-1} \in \mathbb{R}_+$ acts as an additional death rate.

$t(z, p) K^{-1} \in \mathbb{R}_+$ is the *therapy kernel* which models the effect of immunotherapy. It is a death rate of a cancer cell of phenotype p due to the presence of a T-cell of type z.

In addition, $\ell_w^{\text{kill}}(z, p) \in \mathbb{N}_0$ cytokines of type w are deterministically produced at each killing event (see also below).

$s^g(p, \tilde{p}) \in \mathbb{R}_+$ and $s^g_w(p, \tilde{p})K^{-1} \in \mathbb{R}_+$ are the *natural* and *cytokine-induced switch kernels* which model the switching from phenotype p to $\tilde{p}$ of cancer cell with genotype g.

$u_K m(g)$ is the *probability that a mutation occurs* at birth from a cancer cell with genotype g, where $u_K \in [0,1]$ is a scaling parameter.

$M((g,p),(\tilde{g},\tilde{p}))$ is the *mutation law*, i.e. if a mutant is born from a cancer cell with trait (g,p), then the mutant's trait is $(\tilde{g},\tilde{p})$ with probability $M((g,p),(\tilde{g},\tilde{p}))$. By definition $M((g,p),(g,p)) = 0$ and $\sum_{\tilde{g},\tilde{p}} M((g,p),(\tilde{g},\tilde{p})) = 1$.

T-cells: For any $z \in \mathcal{Z}$,

$b(z) \in \mathbb{R}_+$ is the *rate of birth* of a T-cell with type z.

$d(z) \in \mathbb{R}_+$ is the *rate of natural death* of a T-cell with type z.

$b(z,p)K^{-1} \in \mathbb{R}_+$ is the *reproduction kernel*. It models the rate of reproduction of a T-cell with trait z in presence of a cancer cell of phenotype p. In addition, $\ell^{\mathrm{prod}}_w(z,p) \in \mathbb{N}$ cytokines of type w are deterministically produced at each reproduction event.

Cytokines: For any $w \in \mathcal{W}$,

$d(w) \in \mathbb{R}_+$ is the *natural death rate* of a cytokines with type z.

$\ell^{\mathbf{kill}}_w(z,p) \in \mathbb{N}_0$ and $\ell^{\mathbf{prod}}_w(z,p) \in \mathbb{N}$ are the amounts of cytokines of type w, which are deterministically produced at each killing respectively reproduction event.

Note that the cytokines are produced when a cancer cell dies due to therapy or a T-cell reproduces and have no own birth rate.

At any time t, we consider a finite number N_t of individuals, each of them having a trait value $x_i(t) \in \mathcal{X}$. As in the last chapter, we represent the population state at time t by the rescaled point measure on $\mathcal{X}$, which depends on K

$$\nu^K_t = \frac{1}{K} \sum_{i=1}^{N_t} \delta_{x_i(t)}. \tag{III.2.2}$$

Let $\nu^K_t(x) = \nu^K_t(\{x\})$ for $x \in \mathcal{X}$ and $\nu^K_t(p) = \sum_{g \in \mathcal{G}} \nu^K_t(g,p)$ for $p \in \mathcal{P}$. (Note that we used in the last chapter the notation $\langle \nu^K_t, \mathbb{1}_{\{x\}} \rangle$ for $\nu^K_t(\{x\})$). Furthermore, let $\lfloor \cdot \rfloor_\pm$ denote the positive/negative part of the argument. With this notation, a cancer cell with trait (g,p) in the population ν^K_t reproduces an offspring with rate

$$\left\lfloor b(p) - \sum_{\tilde{p} \in \mathcal{P}} c_b(p,\tilde{p}) \nu^K_t(\tilde{p}) \right\rfloor_+, \tag{III.2.3}$$

which is with probability $u_K m(p)$ a mutant. It dies due to age or competition or therapy with rate

$$d(p) + \sum_{\tilde{p} \in \mathcal{P}} c(\tilde{p},p) \nu^K_t(\tilde{p}) + \left\lfloor b(p) - \sum_{\tilde{p} \in \mathcal{P}} c_b(p,\tilde{p}) \nu^K_t(\tilde{p}) \right\rfloor_- + \sum_{z \in \mathcal{Z}} t(z,p) \nu^K_t(z) \tag{III.2.4}$$

(if the death is caused by a T-cell of type z, then $\ell^{\mathrm{kill}}_w(z,p)$ cytokines of type w appear) and it switches its phenotype with rate

$$\sum_{\tilde{p} \in \mathcal{P}} \left(s^g(p,\tilde{p}) + \sum_{w \in \mathcal{W}} s^g_w(p,\tilde{p}) \nu^K_t(w) \right). \tag{III.2.5}$$

A T-cell with trait z in the population ν_t^K reproduces an offspring with rate

$$b(z) + \sum_{p \in \mathcal{P}} b(z,p)\nu_t^K(p) \tag{III.2.6}$$

If the birth event is caused due to the presence of cancer cells, then $\ell_w^{\text{prod}}(z,p)$ cytokines of type w appear. T-cells and cytokines die with their natural death rate independent of the current population state. Note that all other rates depend on the current population state.

Thus, for each $K \geq 1$ the population process, $(\nu_t^K)_{t \geq 0}$, is a Markov process with state space $\mathcal{M}^K(\mathcal{X})$, the set of finite point measures on $\mathcal{X}$ rescaled by K as in the last chapter. The infinitesimal generator of the process can be found in the appendix of this chapter. (For the explicit construction of this Markov process in terms of independent Poisson processes or Poisson point measures, see [55] Chap. 11 or [59].)

The figure below (Fig. III.2) provides a schematic representation of the transitions for a population with trait space $\mathcal{X} = \{(g,p),(g,\tilde{p})\} \cup \{z\} \cup \{w\}$, which constitute our model for the ACT therapy described in Landsberg et al. [92], see Subsection III.3.
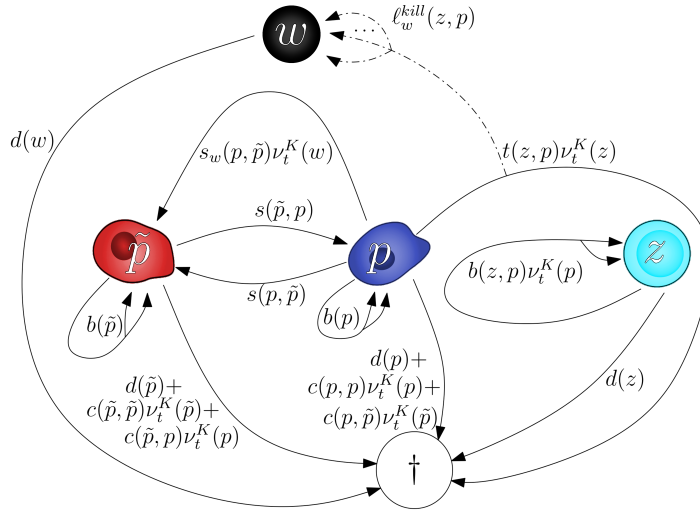

Figure III.2: Dynamics of the process (without mutations) modeling the experiments described in [92]. Here, p denotes differentiated melanoma cells, $\tilde{p}$ dedifferentiated melanoma cells, z T-cells, and w TNF-α. At each arrow the rate for occurrence of the corresponding event is indicated (e.g. birth is illustrated with two arrowheads and death with an arrow directed to †).

Remark 10.   (i) Since $\mathcal{X}$ is finite in this chapter we could also represent the population state as an $|\mathcal{X}|$-dimensional vector. More precisely, let E be a subset of $\mathbb{R}^{|\mathcal{X}|}$ and $E^K \equiv E \cup \{n/K : n \in \mathbb{N}_0\}$, then for fixed $K \geq 1$, the population process can be constructed as Markov process with state space E^K by using independent standard Poisson processes (cf. [55] Chap. 11).

(ii) For an extension to a non finite trait space, e.g. if $\mathcal{G}, \mathcal{P}, \mathcal{Z}$, and $\mathcal{W}$ are compact subsets of $\mathbb{R}^k$ for some $k \geq 1$, the modeling of switching the phenotype and production of cytokines have to be changed in the following way: Each cancer cell with trait $(g,p) \in \mathcal{G}, \mathcal{P}$ has instead of the natural switch kernel $s^g(p,\tilde{p})$ a natural switch rate $\hat{s}((g,p))$ combined with a probability measure $S_{(g,p)}(d\tilde{p})$ on $\mathcal{P}$ and instead of the cytokine-induced switch kernel $s_w^g(p,\tilde{p})K^{-1}$ a cytokine-induced switch kernel $\hat{s}((g,p),w)K^{-1}$ combined with a family probability measure $\{S_{(w,(g,p))}(d\tilde{p})\}_{w \in \mathcal{W}}$ on $\mathcal{P}$. Furthermore, the numbers $\ell_w^{\text{kill}}(z,p)$ have

to be changed in a family $\{\ell^{\text{kill}}_{z,(g,p)}(dw)\}_{z\in\mathcal{Z}}$ of finite point measures on $\mathcal{W}$. Similar, for each T-cell of type $z \in \mathcal{Z}$, the numbers $\ell^{\text{prod}}_w(z,p)$ have to be changed in a family $\{\ell^{\text{prod}}_{z,p}(dw)\}_{p\in\mathcal{P}}$ of finite point measures on $\mathcal{W}$. The infinitesimal generator belonging to this process is also given in the appendix.

III.2.1   The Law of Large Numbers

Suppose the mutation rate is fixed, i.e. $u_K \equiv 1$, and the initial conditions converge, then the sequence of rescaled processes $(\nu^K)_{K\geq 1}$ converges almost surely as $K \to \infty$ to the solution of a quadratic system of ODEs, as stated below. In fact, it follows directly from the law of large numbers for density depending processes of Ethier and Kurtz (cf. [55] Chap. 11). The deterministic system, which provides (partial) information on the stochastic system, consists of a logistic part, a predator-prey relation between T-cells and cancer cells, a mutation and a switch part. Theorem II.3.1 is the corresponding result for the model of the last chapter.

Theorem III.2.1. *Fix $u_K \equiv 1$. Suppose that the initial conditions converge almost surely to a deterministic limit, i.e. $\lim_{K\to\infty} \nu^K_0 = \nu_0$, where ν_0 is a finite measure on $\mathcal{X}$. Then, for every $T > 0$, exists a deterministic function $\xi \in C([0,T], \mathcal{M}_F(\mathcal{X}))$ such that*

$$\lim_{K\to\infty} \sup_{t\in[0,T]} \left\| \nu^K_t - \xi_t \right\|_{TV} = 0 \qquad a.s., \tag{III.2.7}$$

where $\|.\|_{TV}$ is the total variation norm. Moreover, let $\mathfrak{n}$ be the unique solution to the following quadratic dynamical system:

For all $(g,p) \in \mathcal{G} \times \mathcal{P}$,

$$\dot{\mathfrak{n}}_{(g,p)}(t) = \mathfrak{n}_{(g,p)}(t)\left((1-m(g))\left[b(p) - \textstyle\sum_{(\tilde{g},\tilde{p})} c_b(p,\tilde{p})\mathfrak{n}_{(\tilde{g},\tilde{p})}(t)\right]_+ - d(p) \right.$$
$$- \textstyle\sum_{(\tilde{g},\tilde{p})} c(p,\tilde{p})\mathfrak{n}_{(\tilde{g},\tilde{p})}(t) - \left[b(p) - \textstyle\sum_{(\tilde{g},\tilde{p})} c_b(p,\tilde{p})\mathfrak{n}_{(\tilde{g},\tilde{p})}(t)\right]_-$$
$$\left. - \textstyle\sum_z t(z,p)\mathfrak{n}_z(t) - \textstyle\sum_{\tilde{p}}\left(s^g(p,\tilde{p}) + \textstyle\sum_w s^g_w(p,\tilde{p})\mathfrak{n}_w(t)\right)\right)$$
$$+ \textstyle\sum_{\tilde{p}} \mathfrak{n}_{(g,\tilde{p})}(t)\left(s^g(\tilde{p},p) + \textstyle\sum_w s^g_w(\tilde{p},p)\mathfrak{n}_w(t)\right)$$
$$+ \textstyle\sum_{(\tilde{g},\tilde{p})} \mathfrak{n}_{(\tilde{g},\tilde{p})}(t)m(\tilde{g})\left[b(\tilde{p}) - \textstyle\sum_{(g',p')} c_b(\tilde{p},p')\mathfrak{n}_{(g',p')}(t)\right]_+ M((\tilde{g},\tilde{p}),(g,p)),$$

for all $z \in \mathcal{Z}$,

$$\dot{\mathfrak{n}}_z(t) = \mathfrak{n}_z(t)\left(b(z) - d(z) + \textstyle\sum_{(g,p)} b(z,p)\mathfrak{n}_{(g,p)}(t)\right), \tag{III.2.8}$$

for all $w \in \mathcal{W}$,

$$\dot{\mathfrak{n}}_w(t) = -\mathfrak{n}_w(t)d(w) + \textstyle\sum_{(g,p)} \mathfrak{n}_{(g,p)}(t)\left(\textstyle\sum_z \left(\ell^{\text{kill}}_w(z,p)\,t(z,p) + \ell^{\text{prod}}_w(z,p)\,b(z,p)\right)\mathfrak{n}_z(t)\right),$$

and with initial condition: $\mathfrak{n}_x(0) = \nu_0(x)$ for all $x \in \mathcal{X}$.

Then, ξ is given as $\xi_t = \sum_{x\in\mathcal{X}} \mathfrak{n}_x(t)\delta_x$.

Proof. This result follows from Theorem 2.1 in Chapter 11 of [55], since we can construct the process as described in Remark 10 (i). For more details see [97]. $\qquad\square$

It is an important feature of stochastic models opposed to deterministic ones that populations can die out. There are two main reasons for the extinction of a population for finite K:

first, the trajectory of the population size is transient and passes during therapy typically through a low minimum. In this case, random fluctuations can lead to extinction before the population reaches its equilibrium. Second, fluctuations around a finite equilibrium cause extinction of a population after a long enough time. The second case happens at much longer times scales than the first one. (We will see later that time of exit from an attractive domain is of order $\exp(VK)$ for some constant $V > 0$ as in the model before (cf. Thm II.3.7).) In both cases, the value of K plays a crucial role, since it determines the amplitude of the fluctuations and thus the probability of extinction. The relevant mutations in the setup of cancer therapy are driver mutations and appear only rarely. In this case, more precisely, when the mutation probabilities, $u_K m(g)$, tend to zero as $K \to \infty$, mutations are invisible in the deterministic limit. Due to the presence of the switches the analysis of the system is difficult. Unlike in the last chapter, it is *not* a generalized Lotka-Volterra system of the form $\dot{\mathfrak{n}} = \mathfrak{n} f(\mathfrak{n})$, where f is linear in $\mathfrak{n}$. In Section III.4 we show how to deal mathematically with rare mutations and their interactions with fast phenotypic switches or therapy.

III.3   Immunotherapy: Relapse due to random fluctuations

In this section we present an example which qualitatively models the experiment of Landsberg et al. [92], where melanoma escape ACT therapy by phenotypic plasticity in presence of TNF-α. Mutations and birth-reduction competition are not considered (i.e. $c_b \equiv 0$ and $u_K \equiv 0$), since this was not investigated in the experiments. Let us denote by p the differentiated cancer cells, by $\tilde{p}$ the dedifferentiated cancer cells, by z the T-cells of the experiments which can only recognize (are specific for) the differentiated cancer cells p, and by w the TNF-α proteins. We start with describing the deterministic system and denote by $\mathfrak{n}$ the solution to the following system of four differential equations:

$$\begin{aligned}
\dot{\mathfrak{n}}_p &= \mathfrak{n}_p\big(b(p) - d(p) - c(p,p)\mathfrak{n}_p - c(p,\tilde{p})\mathfrak{n}_{\tilde{p}} - s(p,\tilde{p}) - s_w(p,\tilde{p})\mathfrak{n}_w - t(z,p)\mathfrak{n}_z\big) + s(\tilde{p},p)\mathfrak{n}_{\tilde{p}} \\
\dot{\mathfrak{n}}_{\tilde{p}} &= \mathfrak{n}_{\tilde{p}}\big(b(\tilde{p}) - d(\tilde{p}) - c(\tilde{p},\tilde{p})\mathfrak{n}_{\tilde{p}} - c(\tilde{p},p)\mathfrak{n}_p - s(\tilde{p},p)\big) + s(p,\tilde{p})\mathfrak{n}_p + s_w(p,\tilde{p})\mathfrak{n}_w\mathfrak{n}_p \\
\dot{\mathfrak{n}}_z &= \mathfrak{n}_z(b(z,p)\mathfrak{n}_p - d(z)) \\
\dot{\mathfrak{n}}_w &= -\mathfrak{n}_w d(w) + \big(\ell_w^{\text{kill}}(z,p)\, t(z,p) + \ell_w^{\text{prod}}(z,p)\, b(z,p)\big)\mathfrak{n}_p\mathfrak{n}_z.
\end{aligned} \tag{III.3.1}$$

The solution to the deterministic system (III.3.1) with parameters Table (III.1) can be seen on Figure III.3.
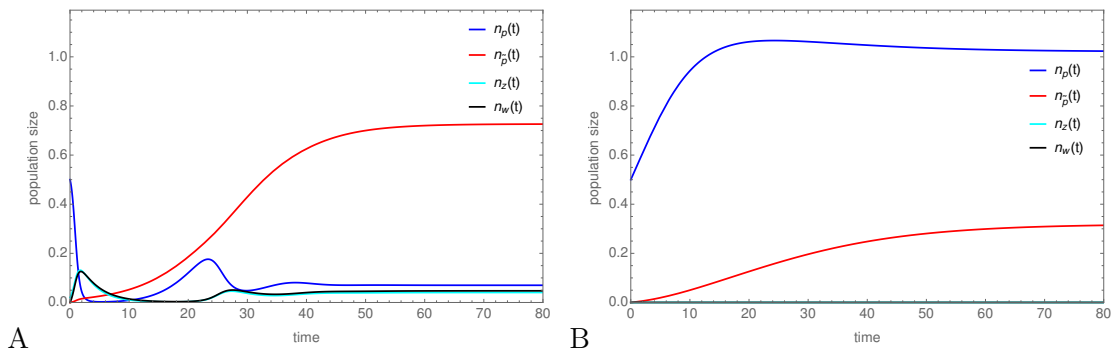
Figure III.3: Solutions to the deterministic system (III.3.1) with parameters (III.1). In A, with immunotherapy: With initial conditions $(\mathfrak{n}_p, \mathfrak{n}_{\tilde{p}}, \mathfrak{n}_z, \mathfrak{n}_w)(0) = (0.5, 0, 0.02, 0)$ the system is attracted to the fixed point $P_{p\tilde{p}zw}$. In B, without immunotherapy: With initial conditions $(\mathfrak{n}_p, \mathfrak{n}_{\tilde{p}}, \mathfrak{n}_z, \mathfrak{n}_w)(0) = (0.5, 0, 0, 0)$ the system is attracted to the fixed point $P_{p\tilde{p}00}$.

There are three fixed points in this example: P_{0000} where all populations sizes are zero, $P_{p\tilde{p}00}$ where the T-cells and TNF-α are absent and both melanoma populations are present, and $P_{p\tilde{p}zw}$ where all populations are present. $P_{p\tilde{p}zw}$ is the only stable fixed point. (The deterministic system is attracted to this fixed point, see Figure III.3 A.) $P_{p\tilde{p}00}$ is stable in the invariant subspace $\{\mathfrak{n}_z = 0\}$ (i.e. if the T-cell population is zero, see Figures III.3 B and III.4 A) and P_{0000} is stable in the invariant subspace $\{\mathfrak{n}_p = 0, \mathfrak{n}_{\tilde{p}} = 0\}$ (i.e. if the tumor is eradicated, see Figure III.4 B). To highlight the qualitative features of the system, we choose parameters such that the minimum of the T-cell population during remission is low, and such that the equilibrium value of melanoma of type p in presence of T-cells is low, whereas equilibrium values of both melanoma types in absence of T-cells are high. For initial conditions such that the number of differentiated melanoma cells, $\mathfrak{n}_p(0)$, is large, the number of injected T-cells, $\mathfrak{n}_z(0)$, is small, and the numbers of dedifferentiated melanoma cells, $\mathfrak{n}_{\tilde{p}}(0)$, and TNF-$\alpha$ molecules, $\mathfrak{n}_w(0)$, are small or equal to zero, the *deterministic* system is attracted to $P_{p\tilde{p}zw}$: the T-cell population, $\mathfrak{n}_z$, increases in presence of its target p, TNF-α is secreted, and the population of differentiated melanoma cells, $\mathfrak{n}_p$, shrinks due to killing and TNF-α induced switching, whereas the population of dedifferentiated melanoma cells, $\mathfrak{n}_{\tilde{p}}$, grows.
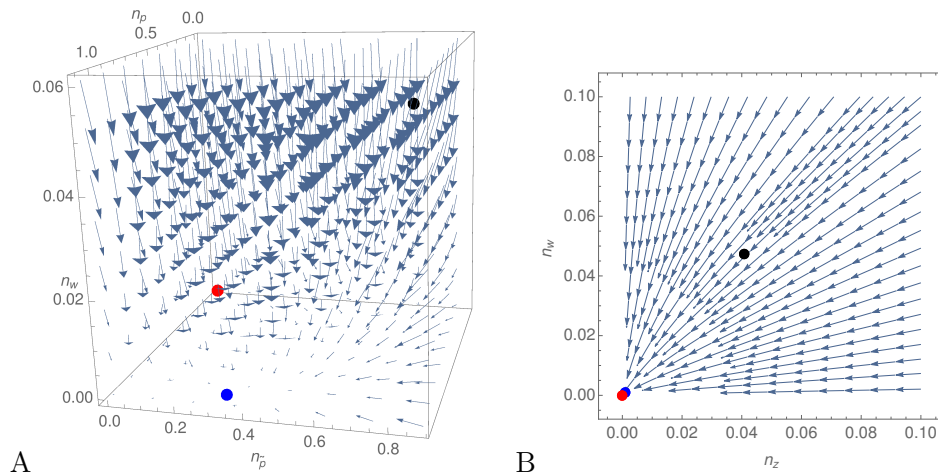

Figure III.4: Vector fields of the deterministic system. The black dots show $P_{p\tilde{p}zw}$, the blue dots $P_{p\tilde{p}00}$, and the red dots P_{0000}. (A) For the invariant subspace $\{\mathfrak{n}_z = 0\}$. (B) For the invariant subspace $\{\mathfrak{n}_p = 0, \mathfrak{n}_{\tilde{p}} = 0\}$.

For the *stochastic system*, several types of behavior can occur with certain probabilities: either the trajectory stays close to that of the deterministic system and the system reaches a neighborhood of the fixed point $P_{p\tilde{p}zw}$ (Fig. III.5 A) or the T-cell population, $\nu^K(z)$, dies out and the system reaches a neighborhood of $P_{p\tilde{p}00}$ (Fig. III.5 B) or the tumor is eradicated, i.e $\nu^K(p)$ and $\nu^K(\tilde{p})$ die out, and the system reaches P_{0000} (Fig. III.5 C). In the second case the TNF-α population, $\nu^K(w)$, becomes extinct shortly after the extinction of the T-cells, $\nu^K(z)$, and the population of differentiated melanoma cells, $\nu^K(p)$, can grow again. Moreover, TNF-α inducing the switch from p to $\tilde{p}$ vanishes and we observe a relapse which consists mainly of differentiated cells. This case was often observed in the experiments. Depending on the choice of parameters (in particular switching, therapy or cross-competition), a variety of different behavior is possible.

A therapy can only be called successful if the whole tumor is eradicated or kept small for a long time. Thus, a natural idea to obtain this is to inject two types of T-cells in future
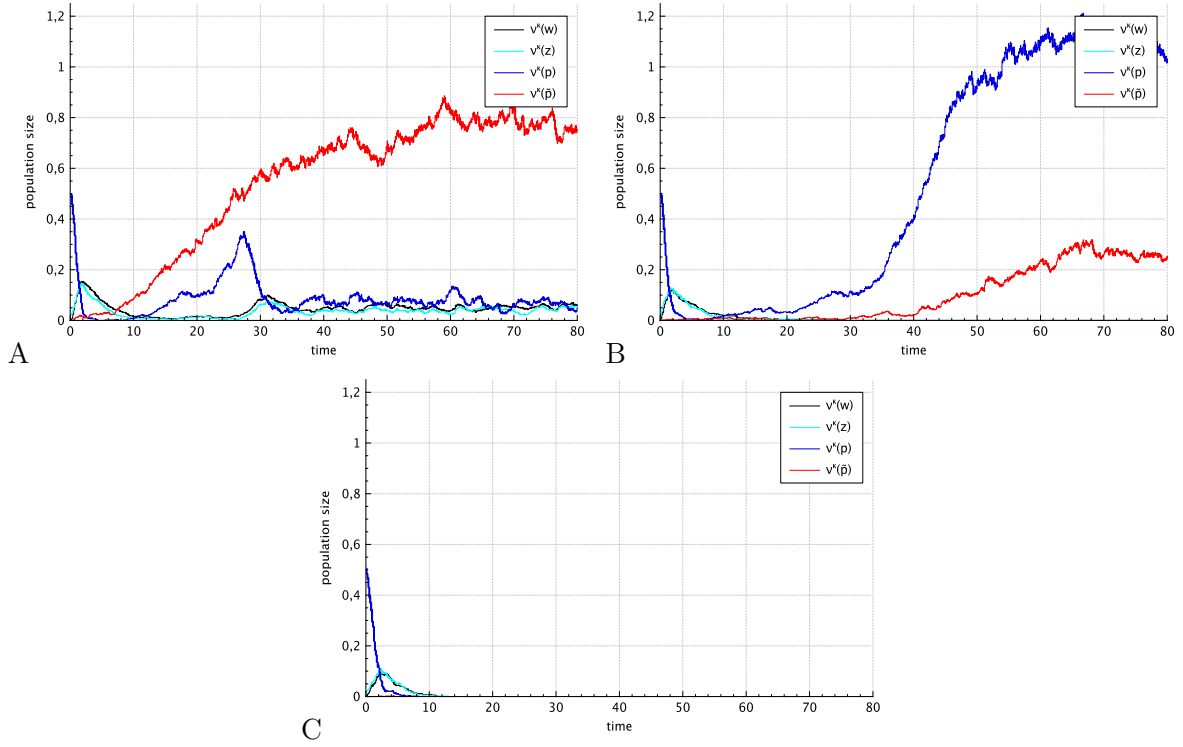
Figure III.5: Simulations of the evolution of melanoma under T-cell therapy for parameters (III.1), initial conditions $\nu_0^K = 0.5\delta_p + 0.02\delta_z$ and $K = 500$. The graphs show the number of individuals divided by K versus time. Three scenarios are possible for therapy with T-cells of one specificity: (A) T-cells z survive and the system is attracted to $P_{p\tilde{p}zw}$ (B) T-cells z die out and the system is attracted to $P_{p\tilde{p}00}$ (C) the tumor is eradicated and the system reaches P_{0000}.

therapies as suggested in [92]. To model this scenario, one needs to add T-cells attacking the dedifferentiated cells as new actors to the setting described above. The corresponding deterministic system contains one extra predator-prey term. The introduction of a second T-cell type which recognizes and kills only the dedifferentiated cancer cells (cells with trait $\tilde{p}$) leads to a more complex system with two new fixed points. One of these, namely the one where all populations are non-zero, is the new stable fixed point of the system. Starting from our choice of initial conditions, the deterministic system converges to $P_{p\tilde{p}z_p z_{\tilde{p}}w}$, but the stochastic system can hit one of the invariant hyperplanes due to fluctuations similar as in the one T-cell case. The scenario with two type of T-cells is studied in detail in [9] and [97].

$b(p) = 0.5$	$d(p) = 0.3$	$c(p,p) = 0.175$	$c(p,\tilde{p}) = 0.1$	$s(p,\tilde{p}) = 0.005$	$s_w(p,\tilde{p}) = 0.4$
$b(\tilde{p}) = 0.5$	$d(\tilde{p}) = 0.3$	$c(\tilde{p},p) = 0.1$	$c(\tilde{p},\tilde{p}) = 0.2$	$s(\tilde{p},p) = 0.05$	-
$b(z,p) = 5$	$d(z) = 0.35$	$t(z,p) = 15$	-	-	-
-	$d(w) = 0.3$	$\ell_w^{\mathrm{kill}}(z,p) = 0$	$\ell_w^{\mathrm{prod}}(z,p) = 1$	-	-

Table III.1: Parameters of the Figures III.3, III.4 and III.5

The parameters of this section are chosen ad hoc to highlight the influence of randomness and the possible behavior of the system. However, it can be shown that our models are capable to reproduce the experimental data of Landsberg et al. [92] *quantitatively*, with biological reasonable parameters. For more details see either [9] or [97].
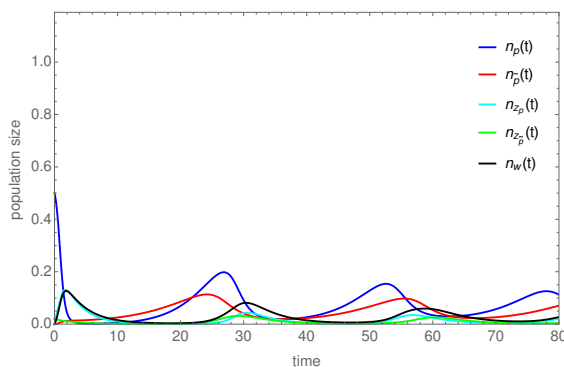
Figure III.6: Solution of the deterministic system for two T-cell typewith initial conditions $(\mathfrak{n}_p, \mathfrak{n}_{\tilde{p}}, \mathfrak{n}_{z_p}, \mathfrak{n}_{z_{\tilde{p}}}, \mathfrak{n}_w)(0) = (0.5, 0, 0.02, 0.02, 0)$. The system is attracted to the new fixed point (cf. [9] and [97]).

III.4 The interplay between rare mutations and fast switches

In this section we study more theoretical aspects which arise by including switching rates in the standard model. More precisely, we are interested in the interplay between the fast phenotypical changes by switching and the slow genotypical changes by mutation.

The model we defined in Section III.2 can be seen as a generalization of one of the standard individual-based models of adaptive dynamics, usually called BPDL-Process, which was introduced in publications of Bolker, Pacala, Dieckmann, and Law [15, 16, 43]. (See Section I.1 for more background information.) The simultaneous limits of large populations ($K \to \infty$) and rare mutations ($u_K \to 0$), under conditions which separate the ecological and the evolutionary effects, were studied mathematically rigorous by Champagnat and Méléard [25, 30]. At this scale the system has time to equilibrate between two mutational events. On the mutation time scale (evolutionary time scale) the population can be described as a Markov jump process along a sequence of equilibria of, in general, polymorphic populations. An important (and in some sense generic) special case occurs when the mutant population fixates while the resident population dies out in each step. The corresponding jump process is called the *Trait Substitution Sequence* (TSS) in adaptive dynamics. Champagnat [25] derived criteria in the context of individual-based models under which convergence to the TSS can be proven (cf. Thm. II.3.7). The general process is called the *Polymorphic Evolution Sequence* (PES). It is described partly implicitly in [30], as it involves the identification of attractive fixed points in a sequence of Lotka-Volterra equations that are in general not tractable analytically. Costa et al. study an extension of the model with a predator-prey relation [33]. The predator-prey kernel is an explicit function of parameters describing defense strategies for preys, together with the ability of predators to circumvent the defense mechanism. In the simultaneous limits of large populations and rare mutations convergence to a Markov jump process that generalizes the PES is derived. Furthermore, by taking the subsequent limit of small mutational effects ($\sigma \to 0$), convergence to an extended version of the canonical equation of adaptive dynamics (CEAD) is obtained in the case of monomorphic prey and predator populations. In [29], Champagnat et al. construct a stochastic multi-resource chemostat model (based on the standard one) which couples deterministic and stochastic dynamics. In the simultaneous limits of large populations and rare mutations, convergence to an extended PES is observed for this process, too. Collet et al. develop and analyze an individual-based model of the adaptive dynamics for sexual reproducing populations, i.e. for Mendelian diploids [31]. By taking the simultaneous limits of large populations and rare mutations convergence to a Markov jump

process generalizing the TSS is derived (cf. also [18]) and by the taking the subsequent limit of small mutational effects a canonical equation for sexual reproducing population is obtain.

The goal of this section is to extend the techniques of adaptive dynamics used in the articles mentioned above to prove that our process (ignoring therapy and birth-reducing competition) converges, in the simultaneous limits of large populations and rare mutations, also to an extended PES. Furthermore, we discuss in Subsection III.4.3 the relation between therapy and mutations.

III.4.1   Explicit construction of the population process with phenotypic plasticity

To derive a generalization of the PES in the presence of fast switches in the phenotypic space, space in the absence of therapy (i.e. no predator-prey term present), we start with the definition and construction of the process we want to analyze. Since T-cells are not present, the trait space $\mathcal{X}$ always equals $\mathcal{G} \times \mathcal{P}$ in this section, except in Subsection III.4.3.

For each $K \geq 1$, the population process ν^K is a $\mathcal{M}^K(\mathcal{X})$-valued Markov processes with infinitesimal generator $\mathscr{L}^K$, defined, for any bounded measurable function $\phi : \mathcal{M}^K(\mathcal{X}) \to \mathbb{R}$ and for all $\mu^K \in \mathcal{M}^K(\mathcal{X})$ by

$$\mathscr{L}^K \phi(\mu^K) \tag{III.4.1}$$
$$= \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \left(\phi\left(\mu^K + \frac{\delta_{(g,p)}}{K} \right) - \phi(\mu^K) \right) (1 - u_K m(g)) b(p) K \mu^K(g,p)$$
$$+ \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \sum_{(\tilde{g},\tilde{p}) \in \mathcal{G} \times \mathcal{P}} \left(\phi\left(\mu^K + \frac{\delta_{(\tilde{g},\tilde{p})}}{K} \right) - \phi(\mu^K) \right) u_K m(g) M\big((g,p),(\tilde{g},\tilde{p})\big) b(p) K \mu^K(g,p)$$
$$+ \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \left(\phi\left(\mu^K - \frac{\delta_{(g,p)}}{K} \right) - \phi(\mu^K) \right) \left(d(p) + \sum_{\tilde{p} \in \mathcal{P}} c(p,\tilde{p}) \mu^K(\tilde{p}) \right) K \mu^K(g,p)$$
$$+ \sum_{(g,p) \in \mathcal{G} \times \mathcal{P}} \sum_{\tilde{p} \in \mathcal{P}} \left(\phi\left(\mu^K + \frac{\delta_{(g,\tilde{p})}}{K} - \frac{\delta_{(g,p)}}{K} \right) - \phi(\mu^K) \right) s^g(p,\tilde{p}) K \mu^K(g,p).$$

The biological parameters $(m(.), b(.), \text{etc.})$ are defined in Section III.2. The first and second terms describe the births (without and with mutation), the third term describes the deaths due to age or competition, and the last term describes the phenotypic plasticity. Note that the first, second, and last terms are linear (in μ^K), but the third term is non-linear. The only difference to the standard model is the presence of the fourth term that correspond to the phenotypic switches. However, this term changes the dynamics sustainably. The system of differential equations which arises in the large population limit without mutation ($u_K = 0$) is not a generalized Lotka-Volterra system anymore, i.e. has *not* the form $\dot{\mathfrak{n}} = \mathfrak{n} f(\mathfrak{n})$, where f is linear in $\mathfrak{n}$.

Since we want to approximate the process later by multi-type branching processes, it is useful to give a pathwise description of ν^K in terms of Poisson point measures (cf. [59]). Let us recall this construction. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an abstract probability space. On this space, we define the following independent random elements:

(i) a convergent sequence $(\nu_0^K)_{K \geq 1}$ of $\mathcal{M}^K(\mathcal{X})$-valued random measures (the random initial population),

(ii) $|\mathcal{X}|$ independent Poisson point measures $(N_{(g,p)}^{\text{birth}}(ds, di, d\theta))_{(g,p) \in \mathcal{X}}$ on $[0, \infty) \times \mathbb{N} \times \mathbb{R}_+$ with intensity measure $ds \sum_{n \geq 0} \delta_n(di) d\theta$,

(iii) $|\mathcal{X}|$ independent Poisson point measures $(\,N_{(g,p)}^{\text{mut.}}(ds,di,d\theta,dx)\,)_{(g,p)\in\mathcal{X}}$ on $[0,\infty)\times\mathbb{N}\times\mathbb{R}_+\times\mathcal{X}$ with intensity measure $ds\sum_{n\geq0}\delta_n(di)d\theta\sum_{\tilde{x}\in\mathcal{X}}\delta_{\tilde{x}}(dx)$.

(iv) $|\mathcal{X}|$ independent Poisson point measures $(\,N_{(g,p)}^{\text{death}}(ds,di,d\theta)\,)_{(g,p)\in\mathcal{X}}$ on $[0,\infty)\times\mathbb{N}\times\mathbb{R}_+$ with intensity measure $ds\sum_{n\geq0}\delta_n(di)d\theta$,

(v) $|\mathcal{X}|$ independent Poisson point measures $(\,N_{(g,p)}^{\text{switch}}(ds,di,d\theta,dp)\,)_{(g,p)\in\mathcal{X}}$ on $[0,\infty)\times\mathbb{N}\times\mathbb{R}_+\times\mathcal{P}$ with intensity measure $ds\sum_{n\geq0}\delta_n(di)d\theta\sum_{\tilde{p}\in\mathcal{P}}\delta_{\tilde{p}}(dp)$,

Then, ν^K is given by the following equation

$$\nu_t^K = \nu_0^K + \sum_{(g,p)\in\mathcal{X}}\int_0^t\int_{\mathbb{N}_0}\int_{\mathbb{R}_+}\mathbb{1}_{\left\{i\leq K\nu_{s-}^K(g,p),\,\theta\leq b(p)(1-u_K m(g))\right\}}\frac{1}{K}\delta_{(g,p)}N_{(g,p)}^{\text{birth}}(ds,di,d\theta) \quad\text{(III.4.2)}$$

$$+ \sum_{(g,p)\in\mathcal{X}}\int_0^t\int_{\mathbb{N}_0}\int_{\mathbb{R}_+}\int_{\mathcal{X}}\mathbb{1}_{\left\{i\leq K\nu_{s-}^K(g,p),\,\theta\leq b(p)u_K m(g)M((g,p),x)\right\}}\frac{1}{K}\delta_x N_{(g,p)}^{\text{mut.}}(ds,di,d\theta,dx)$$

$$- \sum_{(g,p)\in\mathcal{X}}\int_0^t\int_{\mathbb{N}_0}\int_{\mathbb{R}_+}\mathbb{1}_{\left\{i\leq K\nu_{s-}^K(g,p),\,\theta\leq d(p)+\sum_{\tilde{p}\in\mathcal{P}}c(p,\tilde{p})\nu_{s-}^K(\tilde{p})\right\}}\frac{1}{K}\delta_{(g,p)}N_{(g,p)}^{\text{death}}(ds,di,d\theta)$$

$$+ \sum_{(g,p)\in\mathcal{X}}\int_0^t\int_{\mathbb{N}_0}\int_{\mathbb{R}_+}\int_{\mathcal{P}}\mathbb{1}_{\left\{i\leq K\nu_{s-}^K(g,p),\,\theta\leq s^g(p,\tilde{p})\right\}}\frac{1}{K}\left(\delta_{(g,\tilde{p})}-\delta_{(g,p)}\right)N_{(g,p)}^{\text{switch}}(ds,di,d\theta,d\tilde{p}).$$

Remark 11. This construction uses that $\mathcal{X}$ is a finite set and is in some sense closer to the definition given in [55] (p. 455). For non finite trait spaces the process can be constructed as in [59] or Section II.5.

III.4.2   The generalized Polymorphic Evolution Sequence.

In this subsection we consider the case of rare mutations in large populations on a time scale such that a population reaches equilibrium before a new mutant appears:

$$\forall V>0,\qquad \exp(-VK)\ll u_K\ll\frac{1}{K\ln K},\qquad\text{as }K\to\infty \qquad\text{(III.4.3)}$$

and prove that the individual-based process with phenotypic plasticity convergences to a generalization of the PES. Let us start with describing the techniques of adaptive dynamics used in [30] to prove that the standard individual-based process (without phenotypic plasticity) convergences to the PES.

The key element in the proof of the convergence to the PES used in Champagnat and Méléard [30] is a precise analysis of how a mutant population fixates, which we now describe (cf. also Thm. II.3.7). Note that a crucial assumption in [30] is that the large population limit is a competitive Lotka-Volterra system with a unique stable fixed point $\bar{\mathbf{n}}$. Thus, the main task is to *study the invasion of a mutant* that has just appeared in a population close to equilibrium. The invasion can be divided into three steps:

First, as long as the mutant population size is smaller than $K\epsilon$, for a fixed small $\epsilon>0$, the resident population stays close to its equilibrium. Therefore, the mutant population can be approximated by a branching process. Second, once the mutant population reaches the level $K\epsilon$, the whole system is close to the solution of the corresponding deterministic system (this is a consequence of Thm. II.3.1) and reaches an ϵ-neighborhood of $\bar{\mathbf{n}}$ in finite time. Third, the subpopulations which have a zero coordinate in $\bar{\mathbf{n}}$ can be approximated by subcritical branching processes.

The durations of the first and third steps are proportional to $\ln(K)$, whereas that of the second step is independent of K. Thus, the second inequality in (III.4.3) guarantees that, with high probability, the three steps of invasion are completed before a new mutation occurs.

In the first invasion step the invasion fitness of a mutant plays a crucial role. Given a population in a stable equilibrium that populates a certain set of traits, say $M \subset \mathcal{X}$, the invasion fitness $f(x, M)$ is the growth rate of a population consisting of a single individual with trait $x \notin M$ in the presence of the equilibrium population $\bar{\mathfrak{n}}$ on M. In the case of the standard model, it is given by

$$f(x, M) = b(x) - d(x) - \sum_{y \in M} c(x, y) \bar{\mathfrak{n}}_y \tag{III.4.4}$$

(cf. Def. II.3.5). Positive $f(x, M)$ implies that a mutant appearing with trait x from the equilibrium population on M has a positive probability (uniformly in K) to grow to a population of size of order K; negative invasion fitness implies that such a mutant population will die out with probability tending to one (as $K \to \infty$) before it can reach a size of order K. The reason for this is that the branching process (birth-death process) which approximates the mutant population in Champagnat and Méléard's proof is supercritical if $f(x, M)$ is positive and subcritical if $f(x, M)$ is negative.

In order to describe the dynamics of a phenotypic heterogeneous population on the evolutionary time scale, we have to generalize (among other things) the notion of invasion fitness to the case where fast phenotypic switches are present. The issue here is to analyze the probability that a mutant with a new genotype will fixate. Since switches between phenotypes associated to the same genotype happen at times of order one, the growth rate of the initial mutant phenotype does not determine the probability of fixation. See [31] for a similar issue in a simple sexual reproducing model. In the proof of Theorem III.4.3, we will see that we can approximate the mutant's dynamics by a multi-type branching process until it is macroscopic, i.e. until the density reaches the level $K\epsilon$. It is well known that a continuous time multi-type branching process is supercritical if and only if the largest eigenvalue of the infinitesimal generator of its mean matrix is larger than zero (cf. [6, 112]). Therefore, this eigenvalue will be an appropriate generalization of the invasion fitness in our case.

The competitive Lotka-Volterra system with phenotypic plasticity.

Let us begin with a recurrence assumption about the phenotypic plasticity. For all $g \in \mathcal{G}$, let X^g be the Markov chain, which describes the phenotypic plasticity, i.e. a stationary Markov chain with state space $\mathcal{P}$ and with transition probability

$$\mathbb{P}[X_i^g = \tilde{p} \mid X_{i-1}^g = p] = \frac{s^g(p, \tilde{p})}{\sum_{\hat{p} \in \mathcal{P}} s^g(p, \hat{p})}, \quad \text{if } \sum_{\hat{p} \in \mathcal{P}} s^g(p, \hat{p}) > 0 \tag{III.4.5}$$

and

$$\mathbb{P}[X_i^g = p \mid X_{i-1}^g = p] = 1, \quad \text{if } \sum_{\hat{p} \in \mathcal{P}} s^g(p, \hat{p}) = 0. \tag{III.4.6}$$

Then, we assume the following:

Assumption 6. For all $g \in \mathcal{G}$, all communication classes of X^g are recurrent.

In other words, X^g can be decomposed in finitely many irreducible Markov chains, which are recurrent. For every $(g, p) \in \mathcal{G} \times \mathcal{P}$, let us denote the communication class associated with (g, p) by $[p]_g$. This is the communication class of X^g which contains p, i.e. p can be
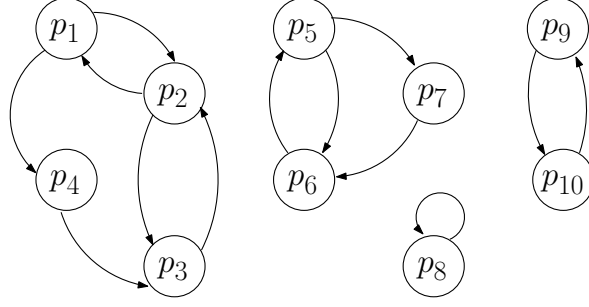
Figure III.7: Example of a Markov chain X^g. Here, $\mathcal{P} = \{p_1, \ldots, p_{10}\}$ and X^g has four communication classes: $\{p_1, p_2, p_3, p_4\}, \{p_5, p_6, p_7\}, \{p_8\}, \{p_9, p_{10}\}$. The class $\{p_8\}$ has only one element, i.e. $\sum_{i=1}^{10} s^g(p_8, p_i) = 0$ in this example.

seen as a representative of the class, which has an equivalence relation depending on g. Since we have seen in the experiments that the phenotypic change is reversible, this assumption is reasonable. It ensures that if we start with a large enough population consisting only of individuals carrying the same trait (g, p), then after a short time all phenotypes of $[p]_g$ will be present in the population, but non of the other classes.

Next, let us have a look at the large population limit of the individual-based process without mutation ($u_K \equiv 0$). Fix d traits in the population: $(\mathbf{g}, \mathbf{p}) = ((g_1, p_1), \ldots, (g_d, p_d)) \in (\mathcal{G} \times \mathcal{P})^d$, suppose that the support of the initial conditions is given by these traits and that the sequence of the initial conditions converges almost surely to a deterministic limit, i.e.

$$\lim_{K \to \infty} \nu_0^K = \sum_{i=1}^{d} n_i(0) \delta_{(g_i, p_i)} \quad \text{a.s.,} \qquad \text{where } n_i(0) > 0 \text{ for all } i \in \{1 \ldots d\}. \qquad \text{(III.4.7)}$$

Then, compare with Theorem III.2.1, for every $T > 0$, the sequences of processes $\nu^K \in \mathbb{D}([0, T], \mathcal{M}^K(\mathcal{X}))$ generated by $\mathcal{L}^K$ with initial state ν_0^K converges almost surely to a deterministic function $\xi \in C([0, T], \mathcal{M}(\mathcal{X}))$. Moreover, let $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ be the set of traits which can be reached by switching, i.e.

$$\mathcal{X}_{(\mathbf{g}, \mathbf{p})} \equiv \bigcup_{i=1}^{d} \{g_i\} \times [p_i]_{g_i}. \qquad \text{(III.4.8)}$$

With this notation, ξ is given by $\xi(t) = \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathfrak{n}_x(t) \delta_x$, where $\mathfrak{n}$ is the solution of the competitive Lotka-Volterra system with phenotypic plasticity defined below. (The initial condition of $\mathfrak{n}$ is given by $n_i(0)$ for all $i \in \{1 \ldots d\}$.)

Definition III.4.1. For any $(\mathbf{g}, \mathbf{p}) \in (\mathcal{G} \times \mathcal{P})^d$, we denote by $LVS(d, (\mathbf{g}, \mathbf{p}))$ the *competitive Lotka-Volterra system with phenotypic plasticity*. This is an $|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|$-dimensional system of ODEs defined for all $(g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ by

$$\dot{\mathfrak{n}}_{(g,p)} = \mathfrak{n}_{(g,p)} \left(b(p) - d(p) - \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(p, \tilde{p}) \mathfrak{n}_{(\tilde{g}, \tilde{p})} - \sum_{\tilde{p} \in [p]_g} s^g(p, \tilde{p}) \right) + \sum_{\tilde{p} \in [p]_g} s^g(\tilde{p}, p) \mathfrak{n}_{(g, \tilde{p})}. \quad \text{(III.4.9)}$$

We choose the name competitive Lotka-Volterra system with phenotypic plasticity to emphasize that we add phenotypic plasticity (induced by switching rates) in the usual competitive Lotka-Volterra system. However, the system LVS is *not* a system of generalized Lotka-Volterra equations, which could be misleading by using this name.

Next, we introduce the notation of coexisting traits in this context (cf. [30] and Definition II.3.4 for the standard model).

Definition III.4.2. For any $d \geq 2$, we say that the distinct traits $(g_1, p_1), \ldots, (g_d, p_d)$ *coexist* if the system $LVS(d, (\mathbf{g}, \mathbf{p}))$ admits a unique non-trivial equilibrium $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}) \in (0, \infty)^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$ which is locally strictly stable meaning that all eigenvalues of the Jacobian matrix of the system $LVS(d, (\mathbf{g}, \mathbf{p}))$ at $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ have strictly negative real parts.

Note that if $(g_1, p_1), \ldots, (g_d, p_d)$ coexist, then all traits of $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ coexist and the equilibrium $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ is asymptotically stable. We will prove later that if the traits $(g_1, p_1), \ldots, (g_d, p_d)$ coexist, then the *invasion probability* of a mutant trait $(\tilde{g}, \tilde{p})$ which appears in the resident population $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ close to $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ is given by the function

$$1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}), \tag{III.4.10}$$

where $q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p})$ is given as follows: Let us denote the elements of $[\tilde{p}]_{\tilde{g}}$ by $\tilde{p} = \tilde{p}_1, \ldots, \tilde{p}_{|[\tilde{p}]_{\tilde{g}}|}$. Then, $q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p})$ is the first component of the smallest root of

$$\mathbf{u}(\mathbf{y}) = \mathbf{0}, \tag{III.4.11}$$

where $\mathbf{u}$ is a map from $\mathbb{R}^{|[\tilde{p}]_{\tilde{g}}|}$ to $\mathbb{R}^{|[\tilde{p}]_{\tilde{g}}|}$ defined for all $i \in \{1, \ldots, |[\tilde{p}]_{\tilde{g}}|$ by

$$
\begin{aligned}
u_i(\mathbf{y}) \;\equiv\; & b(\tilde{p}_i)\, y_i^2 + \sum_{j=1}^{|[\tilde{p}]_{\tilde{g}}|} s^g(\tilde{p}_i, \tilde{p}_j)\, y_j + d(\tilde{p}_i) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_i, p)\bar{\mathfrak{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) \quad \text{(III.4.12)} \\
& - \left(b(\tilde{p}_i) + \sum_{j=1}^{|[\tilde{p}]_{\tilde{g}}|} s^g(\tilde{p}_i, \tilde{p}_j) + d(\tilde{p}_i) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_i, p)\bar{\mathfrak{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) \right) y_i.
\end{aligned}
$$

In fact, $(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}))$ is the probability that a single mutant survives in a resident population with traits $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$. We obtain this by approximating the mutant population with multi-type branching processes (cf. proof of Theorem III.4.6). The function $(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}))$ plays the same role as the function $[f(y; \mathbf{x})]_+/b(y)$ in the standard case (cf. Theorem II.3.7).

To obtain that the process jumps on the evolutionary time scale from one equilibrium to the next, we need an assumption to prevent cycles, unstable equilibria or chaotic dynamics in the deterministic system (cf. [30] Ass. B).

Assumption 7. For any given traits $(g_1, p_1), \ldots, (g_d, p_d) \in \mathcal{G} \times \mathcal{P}$ that coexist and for any mutant trait $(\tilde{g}, \tilde{p}) \in \mathcal{X} \setminus \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ such that $1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}) > 0$, there exists a neighborhood $U \subset \mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}| + |[\tilde{p}]_{\tilde{g}}|}$ of $(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}), 0, \ldots, 0)$ such that all solutions of $LVS(d+1, ((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})))$ with initial condition in $U \cap (0, \infty)^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}| + |[\tilde{p}]_{\tilde{g}}|}$ converge as $t \to \infty$ to a unique locally strictly stable equilibrium in $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}| + |[\tilde{p}]_{\tilde{g}}|}$ denoted by $\mathfrak{n}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$.

We write $\mathfrak{n}^*$ and not $\bar{\mathfrak{n}}$ to emphasize that some coordinates of $\mathfrak{n}^*$ can be zero. Furthermore, we use the shorthand notation $((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ for $((g_1, p_1), \ldots, (g_d, p_d), (\tilde{g}, \tilde{p}))$. Assumption 7 does not have to hold for all traits in $\mathcal{X} \setminus \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ only for those traits $(\tilde{g}, \tilde{p})$ which can appear in the resident population by mutation, i.e. only if $\sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} m(g)M((g,p), (\tilde{g}, \tilde{p}))$ is positive.

Note that it is possible to extend the definitions and assumptions for the study of rare mutations and fast switches in populations with *non discrete trait* space if one assumes that an individual can change its phenotype only to finitely many other phenotypes. This must be encoded in the switching kernels. More precisely, for all $(g, p) \in \mathcal{G} \times \mathcal{P}$ the communication class $[p]_g$ should contain finitely many elements.

Convergence to the generalized Polymorphic Evolution Sequence.

In this subsection we state the main theorem of this chapter and give the general idea of the proof illustrated by an example.

Theorem III.4.3. *Suppose that the Assumptions 6 and 7 hold. Fix $(g_1, p_1), \ldots, (g_d, p_d) \in \mathcal{G} \times \mathcal{P}$ coexisting traits and assume that the initial conditions have support $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ and converge almost surely to $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$, i.e. $\lim_{K \to \infty} \nu_0^K = \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{\mathfrak{n}}_x(\mathbf{g}, \mathbf{p}) \delta_x$ a.s.. Furthermore, assume that*

$$\forall V > 0, \qquad \exp(-VK) \ll u_K \ll \frac{1}{K \ln(K)}, \qquad as \ K \to \infty. \tag{III.4.13}$$

Then, the sequence of the rescaled processes $(\nu_{t/Ku_K}^K)_{t \geq 0}$, generated by $\mathscr{L}^K$ with initial state ν_0^K, converges in the sense of finite dimensional distributions to the measure-valued pure jump process Λ defined as follows: $\Lambda_0 = \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{\mathfrak{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) \delta_{(g,p)}$ and the process Λ jumps for all $(\hat{g}, \hat{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ from

$$\sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{\mathfrak{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) \delta_{(g,p)} \quad to \quad \sum_{(g,p) \in \mathcal{X}_{((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))}} \mathfrak{n}_{(g,p)}^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) \delta_{(g,p)} \tag{III.4.14}$$

with infinitesimal rate

$$m(\hat{g}) b(\hat{p}) \bar{\mathfrak{n}}_{(\hat{g}, \hat{p})}(\mathbf{g}, \mathbf{p})(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p})) M((\hat{g}, \hat{p}), (\tilde{g}, \tilde{p})). \tag{III.4.15}$$

Remark 12.   (i) The convergence cannot hold in law for the Skorokhod topology (cf. [25]). It holds only in the sense of finite dimensional distributions on $\mathcal{M}_F(\mathcal{X})$, the set of finite positive measures on $\mathcal{X}$ equipped with the topology of the total variation norm.

  (ii) The process Λ is a generalized version of the usual PES. Therefore, we call Λ Polymorphic Evolution Sequence with phenotypic Plasticity (PESP).

  (iii) Assumption 7 is essential for this statement. In the case when the dynamical system has multiple attractors and different points near the initial state lie in different basins of attraction, it is not clear and may be random which attractor the system approaches. The characterization of the asymptotic behavior of the dynamical system is needed to describe the final state of the stochastic process. This is in general a difficult and complex problem, which is not doable analytically and requires numerical analysis. Thus, we restrict ourselves to the Assumption 7.

We describe in the following the general idea of the proof, which is quite similar to the one given in [30]. As in the last chapter the population is either in a stable phase or in an invasion phase. Until the first mutant appears the population is in a stable phase, i.e. the population stays close to a given equilibrium. From the first mutational event until the population reaches again a stable state, the population is in an invasion phase. In fact, the mutant either survives and the population reaches fast a new stable state (where the mutant trait is present) or the mutant goes extinct and the population is again in the old stable state. After this the populations is again in a stable phase until the next mutation, etc..

Note that we prove in the following that the invasion phases are relatively short $(O(\ln(K)))$ compared to the stable phase $(O(1/u_K K))$. Since we study the process on the time scale $1/Ku_K$, the limit process proceeds as a pure jump process which jumps from one stable state to another.

The stable phase: Fix $\epsilon > 0$. Let $\mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ be the support of the initial conditions, For large

K, the population process ν^K is, with high probability, still in a small neighborhood of the equilibrium $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ when the first mutant appears. In fact, using large deviation results on the problem of exit from a domain (cf. [61]), we obtain that there exists a constant $M > 0$ such that the first time ν^K leave the $M\epsilon$-neighborhood of $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ is bigger than $\exp(VK)$ for some $V > 0$ with high probability. Thus, until this stopping time, mutations born from individuals with trait $x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$ appear with a rate which is close to

$$u_K m(x) b(x) K \bar{\mathfrak{n}}_x(\mathbf{g}, \mathbf{p}).$$

The condition (III.4.13), more precisely $1/(K u_K) \ll \exp(VK)$ for all $V > 0$, ensures that the first mutation appears before this exit time. Note that we expand here the arguments of the corresponding results of Champagnat (cf. Thm. 3 of [25]) to the multi-type case with phenotypic plasticity. We prove this in Theorem III.4.4.

The invasion phase: We divide the invasion of a given mutant trait $(\tilde{g}, \tilde{p})$ into *three steps*, in a similar way as done in [25] and [30] (cf. Figure III.8).

In the first step, from a mutational event until the mutant population goes extinct or the mutant density reaches the value ϵ, the number of mutant individuals is small (cf. Fig. III.8, $[0, t_1]$). Thus, applying a perturbed version of the large deviation result we used in the first phase, we obtain that the resident population stays close to its equilibrium density $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ during this step. Using similar arguments as Champagnat et al. [25, 30], we prove that the mutant population is well approximated by a $|[\tilde{p}]_{\tilde{g}}|$-type branching process Z, as long as the mutant population has less than ϵK individuals. More precisely, let us denote the elements of $[\tilde{p}]_{\tilde{g}}$ by $\tilde{p}_1, \ldots, \tilde{p}_{|[\tilde{p}]_{\tilde{g}}|}$, then, for each $1 \le i \le |[\tilde{p}]_{\tilde{g}}|$, each individual in Z (carrying trait $(\tilde{g}, \tilde{p}_i)$) undergoes

(i) birth (without mutation) with rate $b(\tilde{p}_i)$,

(ii) death with rate $d(\tilde{p}_i) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_i, p) \bar{\mathfrak{n}}_{(g,p)}(\mathbf{g}, \mathbf{p})$ and

(iii) switching to phenotype $\tilde{p}_j$ with rate $s^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j)$ for all $1 \le j \le |[\tilde{p}]_{\tilde{g}}|$.

This continuous time multi-type branching process is supercritical if and only if the largest eigenvalue of the infinitesimal generator of its mean matrix, which we denote by $\lambda_{\max}$, is larger than zero. Hence, the mutant invades with positive probability if and only if $\lambda_{\max} > 0$. Moreover, the probability that the density of the mutant's genotype, $\nu^K(\tilde{g})$, reaches ϵ at some time t_1 is close to the probability that the multi-type branching process reaches the total mass ϵK, which converges as $K \to \infty$ to $(1 - q_{(\mathbf{g}, \mathbf{p})}(\tilde{g}, \tilde{p}))$.

In the second step, we obtain as a consequence of Theorem III.2.1 that once the mutant density has reached ϵ, for large K, the stochastic process ν^K can be approximated on any finite time interval by the solution of $LVS(d+1, ((g_1, p_1), \ldots, (g_d, p_d), (\tilde{g}, \tilde{p})))$ with a given initial state. By Assumption 7, this solution reaches the ϵ-neighborhood of its new equilibrium $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ in finite time. Therefore, for large K, the stochastic process ν^K also reaches with high probability the ϵ-neighborhood of $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ at some bounded (K independent) time t_2.

In the third step, we can use similar arguments as in the first. Since $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ is a strongly locally stable equilibrium (Assumption 7), the stochastic process ν_t^K stays close $n^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))$ and we can approximate the densities of the traits $(g, p) \in \mathcal{X}_{((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))}$ with $n^*_{(g,p)}((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})) = 0$ by $|[p]_g|$-type branching processes which are subcritical and therefore become extinct a.s..

The time of the first and third step are proportional to $\ln(K)$, whereas the time of the second step is bounded. Thus, the second inequality in (III.4.13) guarantees that, with high probability, the three steps of invasion are completed before a new mutation occurs. After the last step the process comes back to a stable phase, but with a possible different resident population, until the next mutation happens.
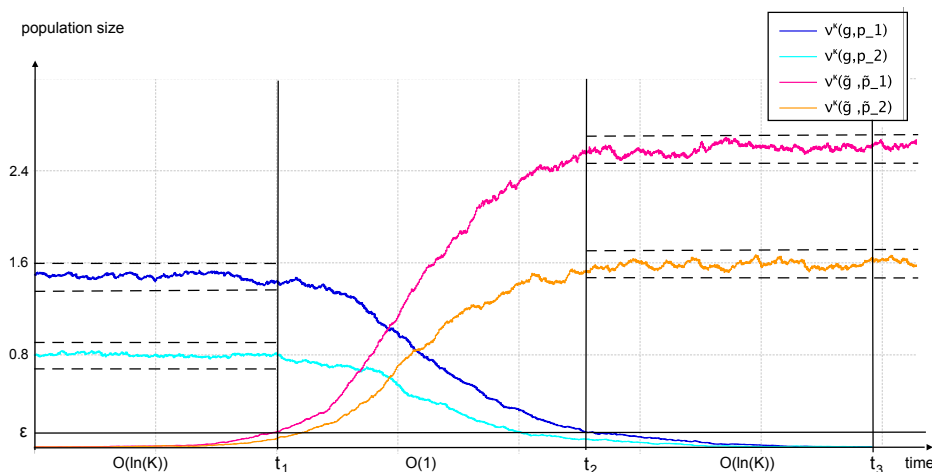

Figure III.8: The three steps of one invasion phase.

An example: Figure III.8 shows the invasion phase of a single mutant with trait $(\tilde{g}, \tilde{p}_1)$, which appeared (at time 0) in a population close to $\bar{n}(\mathbf{g}, \mathbf{p})$ (indicated by the dashed lines). In this example the resident population consists of two coexisting traits (g, p_1) and (g, p_2) and the mutant individuals can switch to one other the phenotype only, i.e. $[\tilde{p}_1]_{\tilde{g}} = \{\tilde{p}_1, \tilde{p}_2\}$. The parameters of the simulation III.8 are given in Table III.2. The stable fixed point of

$b(p_1) = 3$	$d(p_1) = 1$	$c(p_1, p_1) = 1$	$c(p_1, p_2) = 0.7$	$c(p_1, \tilde{p}_1) = 0.7$	$c(p_1, \tilde{p}_2) = 0.7$	$s^g(p_1, p_2) = 1$
$b(p_2) = 3$	$d(p_2) = 1$	$c(p_2, p_1) = 0.7$	$c(p_2, p_2) = 1$	$c(p_2, \tilde{p}_1) = 0.7$	$c(p_2, \tilde{p}_2) = 0.7$	$s^g(p_2, p_1) = 2$
$b(\tilde{p}_1) = 5$	$d(\tilde{p}_1) = 1$	$c(\tilde{p}_1, p_1) = 0.7$	$c(\tilde{p}_1, p_2) = 0.7$	$c(\tilde{p}_1, \tilde{p}_1) = 1$	$c(\tilde{p}_1, \tilde{p}_2) = 0.7$	$s^{\tilde{g}}(\tilde{p}_1, \tilde{p}_2) = 1.5$
$b(\tilde{p}_2) = 4$	$d(\tilde{p}_2) = 1$	$c(\tilde{p}_2, p_1) = 0.7$	$c(\tilde{p}_2, p_2) = 0.7$	$c(\tilde{p}_2, \tilde{p}_1) = 0.7$	$c(\tilde{p}_2, \tilde{p}_2) = 1$	$s^{\tilde{g}}(\tilde{p}_2, \tilde{p}_1) = 2$
$K = 2000$	$u_K = 0$	$\nu_0^K(g, p_1) = 1.5$	$\nu_0^K(g, p_2) = 0.8$	$\nu_0^K(\tilde{g}, \tilde{p}_1) = 1/K$	$\nu_0^K(\tilde{g}, \tilde{p}_1) = 0$	-

Table III.2: Parameters of Figure III.8

the system $LVS(2, ((g, p_1), (g, p_2)))$ is $\bar{n}((g, p_1), (g, p_2)) \approx (1.507, 0.809)$. The infinitesimal generator of the mean matrix of the multi-type branching process that approximates the mutant population in the first step is approximately

$$\begin{pmatrix} 0.879 & 1.5 \\ 2 & -0.621 \end{pmatrix}. \tag{III.4.16}$$

Since the largest eigenvalue of this matrix is positive (≈ 2.016), the mutant population reaches with positive probability the second invasion step (cf. Figure III.8). Furthermore, $n^* \approx (0, 0, 2.608, 1.608)$ is the unique locally strictly stable fixed point of the dynamical system $LVS(4, ((g, p_1), (g, p_2), (\tilde{g}, \tilde{p}_1), (\tilde{g}, \tilde{p}_2)))$. The dynamical system and hence also the stochastic process reach in finite time the ϵ-neighborhood of this value. The infinitesimal generator of the mean matrix of the multi-type branching process that approximates the resident population

in the third step is approximately

$$\begin{pmatrix} -1.951 & 2 \\ 1 & -2.951 \end{pmatrix}. \tag{III.4.17}$$

The largest eigenvalue of this matrix is negative (≈ -0.951) meaning that the process is subcritical and goes extinct a.s.. Therefore, there exists a time t_3 such that all individuals which carry trait (g, p_1) or (g, p_2) are a.s. dead at time t_3.

The proof of Theorem III.4.3.

In this paragraph we prove the convergence to the PESP. (The proof uses the same arguments and techniques as [30], which were developed in [25]. However, some extension are necessary, if fast phenotypic switches are included in the process, which we state and prove in this subsection.) We start with an expansion of Theorem 3 of [25]. Part (i) of the following theorem strengthens Theorem III.2.1, and part (ii) studies the problem of exit from an attractive domain in the polymorphic case with phenotypic plasticity.

Theorem III.4.4.   *(i) Assume that the initial conditions have support $\{(g_1, p_1), \dots (g_d, p_d)\}$ and are uniformly bounded, i.e. , for all $1 \leq i \leq d$, $\nu_0^K(g_i, p_i) \in A$, where A is a compact subset of $\mathbb{R}_{>0}$. Then, for all $T > 0$*

$$\lim_{K \to \infty} \sup_{t \in [0,T]} \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathfrak{n}_x(t, \nu_0^K) \delta_x \right\|_{TV} = 0 \quad a.s., \tag{III.4.18}$$

where $\mathfrak{n}(t, \nu_0^K) \in \mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$ denotes the value of the solution of $LVS(d, (\mathbf{g}, \mathbf{p}))$ at time t with initial condition $\mathfrak{n}_x(0, \nu_0^K) = \nu_0^K(x)$ for all $x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$. Note that the measure $\sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathfrak{n}_x(t, \nu_0^K) \delta_x$ depends on K, since the initial condition and hence the solution of $LVS(d, (\mathbf{g}, \mathbf{p}))$ depend on K.

(ii) Let $(g_1, p_1), \dots, (g_d, p_d) \in \mathcal{X}$ coexist. Assume that, for any $K \geq 1$, $Supp(\nu_0^K) = \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$. Let $\tau_{mut.}$ be the first mutation time. Furthermore, let us define the first exit time from the ξ-neighborhood of $\bar{\mathfrak{n}}_x(\mathbf{g}, \mathbf{p})$ by

$$\theta_{exit}^{K, \xi} \equiv \inf \left\{ t \geq 0 : \exists x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})} \text{ such that } \left| \nu_t^K(x) - \bar{\mathfrak{n}}_x(\mathbf{g}, \mathbf{p}) \right| > \xi \right\}. \tag{III.4.19}$$

Then, there exist $\epsilon_0 > 0$ and $M > 0$ such that for all $\epsilon < \epsilon_0$, there exists $V > 0$ such that if the initial states of ν^K belong to the ϵ-neighborhood of $\bar{\mathfrak{n}}_x(\mathbf{g}, \mathbf{p})$, the probability that $\theta_{exit}^{K, M\epsilon}$ is larger than $e^{KV} \wedge \tau_{mut.}$ is converging to one, i.e.

$$\lim_{K \to \infty} \sup_{\mathbf{n}^K \in (\mathbb{N}/K)^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|} \cap B_\epsilon(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}))} \mathbb{P}\left[\theta_{exit}^{K, M\epsilon} < e^{KV} \wedge \tau_{mut.} \ \middle| \ \nu_0^K(x) = n_x^K \text{ for all } x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})} \right] = 0,$$

$$\tag{III.4.20}$$

where $\mathbf{n}^K \equiv (n_x^K)_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}}$ and $B_\epsilon(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}))$ denotes the ϵ-neighborhood of $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$.

Moreover, (III.4.20) also holds if, for all $(g, p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$, the total death rate of an individual with trait (g, p)

$$d(p) + \sum_{(\tilde{g}, \tilde{p}) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(p, \tilde{p}) \nu_t^K(\tilde{g}, \tilde{p}) \tag{III.4.21}$$

is perturbed by an additional random process that is uniformly bounded by $|\mathcal{X}| c_{max} \epsilon$, where $c_{max} = \max_{x, y \in \mathcal{X}} c(x, y)$.

Remark 13. (i) One consequence of the second part of (ii) is that, with high probability, the process stays in the $M\epsilon$-neighborhood of $\bar{\mathfrak{n}}_x(\mathbf{g}, \mathbf{p})$ until the first time that a mutant's density reaches the value ϵ. In other words, let $\theta^K_{\text{Invasion}}$ denote the first time that a mutant's density reaches the value ϵ, i.e

$$\theta^K_{\text{Invasion}} \equiv \left\{ t \geq 0 : \exists (g, p) \notin \mathcal{X}_{(\mathbf{g}, \mathbf{p})} : \sum_{\tilde{p} \in [p]_g} \nu^K_t(g, \tilde{p}) \geq \epsilon \right\}. \tag{III.4.22}$$

Then, the probability that $\theta^{K, M\epsilon}_{\text{exit}}$ is larger than $e^{KV} \wedge \theta^K_{\text{Invasion}}$ is converging to one. We use this result also for the third invasion step.

(ii) Since $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ is a locally strictly stable fixed point of the system $LVS(d, (\mathbf{g}, \mathbf{p}))$, there exists a constant $M > 0$ such that for all $\epsilon > 0$ small enough

$$\text{for all trajectories } \mathfrak{n}(t) \text{ with } \|\mathfrak{n}(0) - \bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})\| < \epsilon : \qquad \sup_{t \geq 0} \|\mathfrak{n}(t) - \bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})\| < M\epsilon.$$

Proof. The proof is an expansion of the arguments of the proof of Theorem 3 of [25]. The main task of (i) is to show that a large deviation principle on $[0, T]$ holds if we modify the process a bit. The main task of (ii) is to show that we can use the classical estimates for time of exit from a domain (cf. [61]) for the jump process ν^K. Note that Freidlin and Wentzell study in the book [61] mainly small white noise perturbations of dynamical systems. However, there also are some comments on the generalization to dynamical systems with small jump-like perturbations (cf. [61], Sec. 5.4).

Proof of (i). In Chapter 10 of [45], Dupuis and Ellis prove the Laplace principle for continuous time Markov Processes with continuous statistics including diffusions and jump processes. Such a Laplace principle is equivalent to a large deviation principle with the same rate function (cf. Section 1.2 of [45]). (The first general large deviation results for these processes are due to Freidlin and Wentzell [61].) In fact, the process, we are studying, does not full fill the conditions of the Laplace principle of Dupuis and Ellis, but we show that a modification does. To this aim, observe that any solution of $LVS(d, (\mathbf{g}, \mathbf{p}))$ with uniformly bounded initial condition is uniformly bounded, too. (This is true because we can bound the total mass of the population from above by a one dimensional competitive Lotka-Volterra system (cf. Def. II.3.3).) More precisely, for all $x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}$

$$\sup_{t \geq 0} \mathfrak{n}_x(t, \nu^K_0) \leq \max_{p, \tilde{p}, \hat{p} \in \mathcal{P}} \left(2 \frac{b(p)}{c(\tilde{p}, \hat{p})} \right) \vee \max(A) \equiv C \tag{III.4.23}$$

We begin with some notation. First, fix $\delta > 0$, let χ be a map from $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}_{\geq 0}$ to $[0, C + \delta]^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$ defined by

$$\chi(y) \equiv (\chi_{(g,p)}(y))_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \quad \text{with} \quad \chi_{(g,p)}(y) = \begin{cases} y_{(g,p)} & \text{if } y_{(g,p)} \in [0, C + \delta] \\ C + \delta & \text{else.} \end{cases} \tag{III.4.24}$$

Let $(Y^K)_{K \geq 1}$ be a sequence of continuous time, $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$-valued Markov processes with infinitesimal generator $\mathbf{L}^K$, defined on continuous differential functions with compact support, $f \in C^1_c(\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}, \mathbb{R})$, by

$$\mathbf{L}^K f(y) \equiv \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} a_{(g,p)}(y) \frac{\partial f(y)}{\partial y_{(g,p)}} \tag{III.4.25}$$

$$+ K \int_{\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}} \left(f\left(y + K^{-1}z\right) - f(y) - K^{-1} \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} z_{(g,p)} \frac{\partial f(y)}{\partial y_{(g,p)}} \right) \mu_y(dz),$$

where

$$
\begin{aligned}
a_{(g,p)}(y) \;\equiv\; & \chi_{(g,p)}(y)\left(b(p) - d(p) - \sum_{(\tilde{g},\tilde{p})\in\mathcal{X}_{(\mathbf{g},\mathbf{p})}} c(p,\tilde{p})\chi_{(\tilde{g},\tilde{p})}(y) - \sum_{\tilde{p}\in[p]_g} s^g(p,\tilde{p}) \right) \quad (\text{III.4.26}) \\
& + \sum_{\tilde{p}\in[p]_g} s^g(\tilde{p},p)\chi_{(g,\tilde{p})}(y)
\end{aligned}
$$

and μ_y is a finite point measure defined as follows. Let $\mathbf{e}_{(g,p)}$ denote the "(g,p)"-th unit vector in $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|}$, then

$$
Supp(\mu_y) \subseteq \bigcup_{(g,p)\in\mathcal{X}_{(\mathbf{g},\mathbf{p})},\,\tilde{p}\in[g]_p} \left\{ \mathbf{e}_{(g,p)},\; -\mathbf{e}_{(g,p)},\; \mathbf{e}_{(g,p)} - \mathbf{e}_{(g,\tilde{p})} \right\} \subset \mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|} \qquad (\text{III.4.27})
$$

and

$$
\mu_y(\{\mathbf{e}_{(g,p)}\}) \;\equiv\; b(p)\chi_{(g,p)}(y), \qquad\qquad\qquad (\text{III.4.28})
$$

$$
\mu_y(\{-\mathbf{e}_{(g,p)}\}) \;\equiv\; \chi_{(g,p)}(y)\left(d(p) + \sum_{(\tilde{g},\tilde{p})\in\mathcal{X}_{(\mathbf{g},\mathbf{p})}} c(p,\tilde{p})\chi_{(\tilde{g},\tilde{p})} \right), \qquad (\text{III.4.29})
$$

$$
\mu_y(\{\mathbf{e}_{(g,p)} - \mathbf{e}_{(g,\tilde{p})}\}) \;\equiv\; s^g(p,\tilde{p})\chi_{(g,\tilde{p})}(y). \qquad\qquad (\text{III.4.30})
$$

Furthermore, define $\tau \equiv \inf\{t \geq 0 : \exists x \in \mathcal{X}_{(\mathbf{g},\mathbf{p})} \text{ s.t. } \nu^K(x) > C + \delta\} \wedge \tau_{\text{mut.}}$, then the processes Y^K (considered as measure-valued process) and ν^K have the same law on the time interval $[0,\tau]$. In contrast to ν^K, Y^K satisfy the conditions assumed in Theorem 10.2.6 of [45]. Thus, for each $T > 0$, the sequence of $(Y^K)_{K\geq 0}$ with $Y^K(0) = Y_0$ satisfies the Laplace principle on $\mathbb{D}([0,T],\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|})$ with rate function

$$
I_{0T}(\phi) = \begin{cases} \int_0^T L(\phi(t),\dot{\phi}(t))dt & \text{if } \phi \text{ is absolute continuous on } [0,T] \\ \infty & \text{else,} \end{cases} \qquad (\text{III.4.31})
$$

where

$$
L(\phi(t),\dot{\phi}(t)) \equiv \sup_{\alpha\in\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|}} \left(\sum_{i=1}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|} \alpha_i\dot{\phi}_i(t) - \int_{\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|}} \left(e^{\sum_{i=1}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|}\alpha_i y_i} - 1 \right)\mu_{\phi(t)}(dy) \right). \quad (\text{III.4.32})
$$

Note that $L(\phi(t),\dot{\phi}(t))$ is zero if and only if ϕ is absolute continuous and $\dot{\phi}_{(g,p)}(t) = a_{(g,p)}(\phi(t))$ for all $(g,p) \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}$. Moreover, the Laplace principle holds uniformly on compacts.

(Freidlin and Wentzell state in [61] a comparable result in terms of action functionals (cf. Thm 5.2.1 of [61]). Namely, $KI_{0T}(\phi)$ is the action functional for the sequence (Y_t^K, P_y^K) in the metric $\rho_{0T}(\phi,\psi) = \sup_{0\leq t\leq T}|\phi_t - \psi_t|$ uniformly in the initial point as $K \to \infty$. P_y^K is used to emphasize that the initial conditions depend on K.)

As already mentioned the Laplace principle is equivalent to the large derivation principle. Thus, for each closed subset F of $\mathbb{D}([0,T],\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|})$

$$
\limsup_{K\to\infty} \frac{1}{K}\ln\left(\sup_{Y_0^K\in A} \mathbb{P}_{Y_0^K}[Y^K \in F] \right) \leq -\inf_{\phi\in F} I_{0T}(\phi). \qquad (\text{III.4.33})
$$

If we choose

$$
F^\delta \equiv \left\{ \phi \in \mathbb{D}([0,T],\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|}) : \phi(0) \in A \text{ and } \sup_{t\in[0,T]}|\phi(t) - \mathfrak{n}(t,\phi(0))| \geq \delta \right\}, \qquad (\text{III.4.34})
$$

then inequality (III.4.33) becomes

$$\lim_{K \to \infty} \frac{1}{K} \ln \left(\sup_{Y_0^K \in A} \mathbb{P}_{Y_0^K} \left[\sup_{0 \le t \le T} |Y^K - \mathfrak{n}(t, Y_0^K)| \ge \delta \right] \right) \le - \inf_{\phi \in F^\delta} I_{0T}(\phi). \tag{III.4.35}$$

Thus, on the event $\{\tau_{\text{mut.}} < T\}$, we have

$$\lim_{K \to \infty} \frac{1}{K} \ln \left(\sup_{\nu_0^K \in A} \mathbb{P}_{\nu_0^K} \left[\sup_{0 \le t \le T} \left\| \nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \mathfrak{n}_x(t, \nu_0^K) \delta_x \right\|_{TV} \ge \delta \right] \right) \le - \inf_{\phi \in F^\delta} I_{0T}(\phi). \tag{III.4.36}$$

Since any solution of $LVS(d, (\mathbf{g}, \mathbf{p}))$ that starts in A does not leave $[0, C]^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$, the set F^δ does not contain such a solution. Therefore, $\inf_{\phi \in F^\delta} I_T(\phi)$ is non-zero. Using the Borel-Cantelli Lemma, this implies (III.4.18) provided that $\lim_{K \to \infty} \mathbb{P}[\tau_{\text{mut.}} < T] = 0$.

Proof of (ii). In Section 5.4 of [61], Freidlin and Wentzell explain how some of their results on the problem of exit from a domain for diffusion processes can be generalized to dynamical systems with small jump-like perturbations under some conditions. In fact, we need a generalization of Theorem 4.4.2 of [61]. As in the proof of (i) the jump processes ν^K do not belong to the class of processes Freidlin and Wentzell consider in Section 5.4.. However, we define similar as in (i) processes, Y^K, which will belong to this class and have the same law until $\theta_{\text{exit}}^{K, \epsilon} \wedge \tau_{\text{mut.}}$: Fix $\epsilon > 0$, let $\tilde{\chi}$ be a map from $\mathbb{R}_{\ge 0}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$ to $\mathbb{R}_{\ge 0}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$ defined by

$$\tilde{\chi}(y) \equiv \begin{cases} y & \text{if} \quad y \in B_{M\epsilon}(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})) \\ \bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}) + M\epsilon \frac{y - \bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})}{\|y - \bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}))\|} & \text{else.} \end{cases} \tag{III.4.37}$$

Let $(Y^K)_{K \ge 1}$ be a sequence of continuous time, $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}$-valued Markov processes with infinitesimal generator $\tilde{\mathbf{L}}^K$ defined as above except that χ is replaced by $\tilde{\chi}$. Let $\theta_{\text{exit}}^{K, M\epsilon, Y}$ denote the first exit time from the $M\epsilon$-neighborhood of $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ of the process Y^K, i.e.

$$\theta_{\text{exit}}^{K, M\epsilon, Y} \equiv \inf \left\{ t \ge 0 : Y_t^K \notin B_{M\epsilon}(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})) \right\}. \tag{III.4.38}$$

Then, by applying the generalized version of Theorem 4.4.2 of [61] (cf. [61], p.138) to Y^K we obtain that for all $\epsilon > 0$ small enough there exists V_ϵ such that for every $\alpha > 0$

$$\lim_{K \to \infty} \sup_{y \in B_\epsilon(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}))} \mathbb{P}_y \left[\theta_{\text{exit}}^{K, M\epsilon, Y} > \exp(K(V_\epsilon + \alpha)) \right] = 0. \tag{III.4.39}$$

Moreover, let $V(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}), y)$ denote the quasipotential of the dynamical system with respect to $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$, i.e.

$$V(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}), y) \equiv \inf \left\{ I_{T_1 T_2}(\phi) : \phi \in C([T_1, T_2], \mathbb{R}^{|\mathcal{X}_{(\mathbf{g}, \mathbf{p})}|}), \phi_{T_1} = \bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}), \phi_{T_2} = y, T_1 \le T_2 \right\}. \tag{III.4.40}$$

The endpoints of the interval $[T_1, T_2]$ are not fixed and $I_{T_1 T_2}(\phi)$ is defined as above. The constant V_ϵ is given by $V_\epsilon \equiv \min_{y \in B_{M\epsilon}(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}))} V((\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})), y)$. Since $\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})$ is a locally strictly stable fixed point there exists $M > 0$ such that for all $\epsilon > 0$ small enough all trajectories of $LVS(d, (\mathbf{g}, \mathbf{p}))$ which start in $B_\epsilon(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}))$ stay in $B_{M\epsilon/2}(\bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p}))$. On the other hand, $L(\phi(t), \dot{\phi}(t)) = 0$ if and only if $\phi(t)$ is a solution of $LVS(d, (\mathbf{g}, \mathbf{p}))$. Thus $V_\epsilon > 0$, which implies (III.4.20).

If the total death rate of an individual is perturbed by an additional random process that is uniformly bounded by $|\mathcal{X}| c_{\text{max}} \epsilon$, we can construct (by using the pathwise construction via

Poisson point measures) two sequences of continuous time, $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|}$-valued Markov processes $Y^{K,1,\epsilon}$ and $Y^{K,2,\epsilon}$ such that

$$Y_t^{K,1,\epsilon} \le \nu_t^K \le Y_t^{K,2,\epsilon} \qquad \text{for all } t \in [0,\tau], \tag{III.4.41}$$

where $Y_t^{K,1,\epsilon}$ and $Y_t^{K,1,\epsilon}$ have the same generators as Y^K except that the death rate $d(p)$ is replaced by $d(p) + |\mathcal{X}|c_{\max}\epsilon$ respectively $d(p) - |\mathcal{X}|c_{\max}\epsilon$. The corresponding dynamical systems have unique locally strictly stable fixed point which lie in $B_{\hat{M}\epsilon}(\bar{\mathfrak{n}}(\mathbf{g},\mathbf{p}))$ for some constant $\hat{M} > 0$. (This can be obtained by applying the implicit function theorem. For the stability of the fixed point see e.g. [68]. Moreover, see [31] for a similar consideration.) Using the same arguments for $Y^{K,1,\epsilon}$ and $Y^{K,2,\epsilon}$ as above ends the proof. $\qquad\square$

The following Lemma describes the asymptotic behavior of $\tau_{\text{mut.}}$ and can be seen as an extension of Lemma 2 of [25] or Lemma A.3 of [30].

Lemma III.4.5. *Let $(g_1,p_1),...,(g_d,p_d) \in \mathcal{X}$ coexist. Assume that, for any $K \ge 1$, $Supp(\nu_0^K) = \mathcal{X}_{(\mathbf{g},\mathbf{p})}$. Let $\tau_{mut.}$ denote the first mutation time. Then, there exists $\epsilon_0 > 0$ such that if the initial states of ν^K belong to the ϵ_0-neighborhood of $\bar{\mathfrak{n}}_x(\mathbf{g},\mathbf{p})$, then, for all $\epsilon \in (0,\epsilon_0)$,*

$$\lim_{K\to\infty} \mathbb{P}\left[\tau_{mut.} > \ln(K), \sup_{t\in[\ln(K),\tau_{mut.}]} \left\| \nu_t^K - \sum_{x\in\mathcal{X}_{(\mathbf{g},\mathbf{p})}} \bar{\mathfrak{n}}_x(\mathbf{g},\mathbf{p})\delta_x \right\|_{TV} < \epsilon \right] = 1, \tag{III.4.42}$$

Furthermore, $(\tau_{mut.} u_K K)_{K\ge 1}$ converges in law to an exponential distributed random variable with parameter $\sum_{(g,p)\in\mathcal{X}_{(\mathbf{g},\mathbf{p})}} m(g)b(p)\bar{\mathfrak{n}}_{(g,p)}(\mathbf{g},\mathbf{p})$ and the probability that the mutant, which appears at time $\tau_{mut.}$, is born from an individual with trait $(g,p) \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}$ converges to

$$\frac{m(g)b(p)\bar{\mathfrak{n}}_{(g,p)}(\mathbf{g},\mathbf{p})}{\sum_{(\tilde{g},\tilde{p})\in\mathcal{X}_{(\mathbf{g},\mathbf{p})}} m(\tilde{g})b(\tilde{p})\bar{\mathfrak{n}}_{(\tilde{g},\tilde{p})}(\mathbf{g},\mathbf{p})} \tag{III.4.43}$$

as $K \to \infty$.

Proof. There exist constants $C > 0$ and $V > 0$, such that on the time interval $[0, \exp(KV)]$ the total mass of the population, $\nu_t^K(\mathcal{X})$, is bounded from above by C (cf. proof of Thm. III.4.4 (i)). Therefore, we can construct an exponential random variable A with parameter $C'Ku_K$, where $C' = C \max_{g\in\mathcal{G},p\in\mathcal{P}} m(g)b(p)$, such that

$$A \le \tau_{\text{mut.}} \qquad \text{on the event} \quad \{\tau_{\text{mut.}} < \exp(KV)\}. \tag{III.4.44}$$

Thus, $P[\tau_{\text{mut.}} > \ln(K)] \ge \mathbb{P}[A > \ln(K)] = e^{-C'\ln(K)Ku_K}$. Since $(III.4.13)$ implies that $\ln(K)Ku_K$ converges to zero as $K \to \infty$, we get $\lim_{K\to\infty} \mathbb{P}[\tau_{\text{mut.}} > \ln(K)] = 1$.

The fixed point $\bar{\mathfrak{n}}(\mathbf{g},\mathbf{p})$ is asymptotic stable. Thus, $\exists\epsilon_0 > 0 : \forall\tilde{\epsilon} \in (0,\epsilon_0) \exists T(\tilde{\epsilon})$:

$$\text{whenever } \|\mathfrak{n}(\mathbf{g},\mathbf{p})(0) - \bar{\mathfrak{n}}(\mathbf{g},\mathbf{p})\| < \epsilon_0, \qquad \text{then } \sup_{t\ge T(\tilde{\epsilon})} |\mathfrak{n}(\mathbf{g},\mathbf{p})(t) - \bar{\mathfrak{n}}(\mathbf{g},\mathbf{p})| < \tilde{\epsilon}/2. \tag{III.4.45}$$

In words, there exists a finite time $T(\tilde{\epsilon})$ such that all trajectories, which start in the ϵ_0 neighborhood of the fixed point, stay after $T(\tilde{\epsilon})$ in the $\tilde{\epsilon}/2$-neighborhood of the fixed point.

Next, we apply the last theorem: By (i), for all $\tilde{\epsilon} \in (0,\epsilon_0) \exists T(\tilde{\epsilon})$ such that, for K large enough,

$$\left\| \nu_{T(\tilde{\epsilon})}^K - \sum_{x\in\mathcal{X}_{(\mathbf{g},\mathbf{p})}} \bar{\mathfrak{n}}_x(\mathbf{g},\mathbf{p})\delta_x \right\|_{TV} < \tilde{\epsilon}. \tag{III.4.46}$$

Then, by (ii), there exist $\epsilon_0 > 0$ and $M > 0$: for all $\tilde{\epsilon} \in (0, \epsilon_0)$ there exists $V > 0$ such that

$$\lim_{K \to \infty} \mathbb{P}\left[\sup_{t \in [T(\tilde{\epsilon}),\, e^{KV} \wedge \tau_{\text{mut.}})} \left\|\nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}} \bar{\mathfrak{n}}_x(\mathbf{g},\mathbf{p})\delta_x\right\|_{TV} < M\tilde{\epsilon}\right] = 1. \qquad (\text{III.4.47})$$

Furthermore, for all $\tilde{\epsilon} \in (0, \epsilon_0)$ there exists $K_0 \in \mathbb{N}$ such that $T(\tilde{\epsilon}) < \ln(K)$ for all $K \geq K_0$. Thus, setting $\epsilon = M\tilde{\epsilon}$, ends the proof of (III.4.42), provided that $\lim_{K \to \infty}[\tau_{\text{mut.}} < e^{KV}] = 1$.

Again, we can construct for all $\epsilon > 0$ two exponential random variables $A^{1,K,\epsilon}$ and $A^{2,K,\epsilon}$ with parameters

$$a_1 u_K K \equiv \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}} u_K m(g) b(p)(\bar{\mathfrak{n}}_{(g,p)}(\mathbf{g},\mathbf{p}) + \epsilon) K \qquad (\text{III.4.48})$$

and

$$a_2 u_K K \equiv \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}} u_K m(g) b(p)(\bar{\mathfrak{n}}_{(g,p)}(\mathbf{g},\mathbf{p}) - \epsilon) K \qquad (\text{III.4.49})$$

such that

$$A^{1,K,\epsilon} \leq \tau_{\text{mut.}} \leq A^{2,K,\epsilon} \qquad \text{on the event } \{T(\tilde{\epsilon}) < \tau_{\text{mut.}} < e^{KV}\}, \qquad (\text{III.4.50})$$

where $T(\tilde{\epsilon})$ is the time defined in equation (III.4.46) and $\tilde{\epsilon} = \epsilon/M$. Moreover, we have

$$\lim_{K \to \infty} \mathbb{P}[\tau_{\text{mut.}} < \ln(K)] = 0 \qquad \text{and} \qquad \lim_{K \to \infty} \mathbb{P}[A^{2,K,\epsilon} > e^{KV}] = 0, \qquad (\text{III.4.51})$$

because $u_K K e^{KV} \to \infty$ as $K \to \infty$. Therefore, for all $\epsilon > 0$, the probability of the event $\{T(\tilde{\epsilon}) < \tau_{\text{mut.}} < e^{KV}\}$ converges to one as K goes to infinity. Furthermore, the random variables $A^{1,K,\epsilon} u_K K$ and $A^{2,K,\epsilon} u_K K$ converge both in law to the same exponential distributed random variable with parameter

$$\sum_{(g,p) \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}} m(g) b(p) \bar{\mathfrak{n}}_{(g,p)}(\mathbf{g},\mathbf{p}) \qquad (\text{III.4.52})$$

as first $K \to \infty$ and then $\epsilon \to 0$. The random variables A, $A^{1,K,\epsilon}$ and $A^{2,K,\epsilon}$ can easily be constructed by using the pathwise description of ν^K (cf. Lemma II.6.4 or [29]). $\qquad \square$

Theorem III.4.6 (The three steps of invasion). *Let $(g_1, p_1), \ldots, (g_d, p_d) \in \mathcal{X}$ coexist. Assume that, for any $K \geq 1$, $Supp(\nu_0^K) = \mathcal{X}_{(\mathbf{g},\mathbf{p})} \cup \{(\tilde{g}, \tilde{p})\}$. Let $\tau_{mut.}$ denote the next mutation time (after time zero) and define*

$$\theta_{No\ Jump}^{K,\xi} \equiv \inf\left\{t \geq 0 : \nu_t^K(\tilde{g}) = 0 \text{ and } \left\|\nu_t^K - \sum_{x \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}} \bar{\mathfrak{n}}_x(\mathbf{g},\mathbf{p})\delta_x\right\|_{TV} < \xi\right\}$$

$$\theta_{Jump}^{K,\xi} \equiv \inf\left\{t \geq 0 : \left\|\nu_t^K - \sum_{x \in \mathcal{X}_{((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))}} \mathfrak{n}_x^*((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))\delta_x\right\|_{TV} < \xi\right.$$

$$\left. \text{and } \forall \hat{x} \notin \{x \in \mathcal{X} : \mathfrak{n}_x^*((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p})) > 0\} : \nu_t^K(\hat{x}) = 0\right\}.$$

Assume that we have a single initial mutant, i.e. $\nu_0^K(\tilde{g}, \tilde{p}) = 1/K$. Then, there exist $\epsilon_0 > 0, C > 0$, and $M > 0$ such that for all $\epsilon \in (0, \epsilon_0)$ if $\|\nu_0^K - \sum_{x \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}} \bar{\mathfrak{n}}_x(\mathbf{g},\mathbf{p})\delta_x\|_{TV} < \epsilon$,

$$\lim_{K \to \infty} \mathbb{P}\left[\theta_{No\ Jump}^{K,M\epsilon} < \theta_{Jump}^{K,M\epsilon}\right] \geq q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}) - C\epsilon, \qquad (\text{III.4.53})$$

$$\lim_{K \to \infty} \mathbb{P}\left[\theta_{Jump}^{K,M\epsilon} < \theta_{No\ Jump}^{K,M\epsilon}\right] \geq 1 - q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}) - C\epsilon, \qquad (\text{III.4.54})$$

where $1 - q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p})$ is the invasion probability defined in (III.4.10) and

$$\forall \eta > 0, \quad \lim_{K \to \infty} \mathbb{P}\left[\theta_{Jump}^{K,M\epsilon} \wedge \theta_{No\ Jump}^{K,M\epsilon} \geq \frac{\eta}{u_K K} \wedge \tau_{mut.}\right] \leq C\epsilon. \qquad (\text{III.4.55})$$

The structure of the proof is similar to the one of Lemma 3 in [25]. However, we have to extend the theory to multi-type branching processes. Thus, the proof is not a simple copy the arguments in [25]. Before we prove the theorem, let us collect *some properties about multi-type continuous time branching processes*. Most of the properties, can be found either in [6] or [112]. The limit theorems, we need in the following, were first obtained by Kesten and Stigum [85, 84, 86] in the discrete time case and by Athreya [5] in the continuous time case.

Let $Z(t)$ be a k-dimensional continuous time branching process. Assume that $Z(t)$ is nonsingular and that the first moments exist. (Note that a process is *singular* if and only if each individual has exactly one offspring and that the existence of the first moments is sufficient for the non-exposition hypothesis.) Then, the mean matrix of $Z(t)$ is a $k \times k$ matrix defined by

$$\mathbf{M}(t) = \{m_{ij}(t),\ i,j = 1,\ldots,k\}, \qquad \text{where} \quad m_{ij}(t) \equiv \mathbb{E}[Z_j(t)|Z(0) = \mathbf{e}_i] \qquad \text{(III.4.56)}$$

and $\mathbf{e}_i$ is the i-th unit vector in $\mathbb{R}^k$. It is well known (cf. [6] p. 202) that there exists a matrix $\mathbf{A}$, called the infinitesimal generator of the semigroup $\{\mathbf{M}(t), t \geq 0\}$, such that

$$\mathbf{M}(t) \equiv \exp(\mathbf{A}t) = \sum_{n=0}^{\infty} \frac{t^n (\mathbf{A})^n}{n!}. \qquad \text{(III.4.57)}$$

Furthermore, let $\mathbf{r} = (r_1,\ldots,r_k)$ be the vector of the branching rates, meaning that every individual of type i has an exponentially distributed lifetime of parameter r_i and let $\mathbf{M}$ be the mean matrix of the corresponding discrete time process, i.e. $\mathbf{M} \equiv \{m_{ij},\ i,j = 1,\ldots,k\}$, where m_{ij} is the expected number of type j offsprings of a single type-i-particle in one generation. Then, we can identify the infinitesimal generator $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{R}(\mathbf{M} - \mathbf{I}), \qquad \text{(III.4.58)}$$

where $\mathbf{R} = diag(r_1,\ldots,r_k)$, i.e. $r_{ij} = r_i \delta_i(j)$ and $\mathbf{I}$ is the identity matrix of size k.

Under the basic assumption of *positive regularity*, namely that there exists a time t_0 such that $\mathbf{M}(t_0)$ is strictly positive, the Perron-Frobenius theory can be used to deduce that

(i) the largest eigenvalue of $\mathbf{M}(t_0)$ is real-valued and strictly positive,

(ii) the algebraic and geometric multiplicities of this eigenvalue are both unity and

(iii) there exists an eigenvector with this eigenvalue, which is strictly positive.

By (III.4.57), all eigenvalues of $\mathbf{M}(t)$ are given by $e^{\lambda_i t}$, where $\{\lambda_i; i = 1,\ldots,k\}$ are the eigenvalues of the infinitesimal generator $\mathbf{A}$ and both matrixes have the same eigenvectors, which implies that we can determine the left and right eigenvectors $\mathbf{u}$ and $\mathbf{v}$ of $\lambda_{\max}(\mathbf{A})$ with all coordinates strictly positive, and such that

$$\sum_{i=1}^{k} v_i u_i = 1 \quad \text{and} \quad \sum_{i=1}^{k} u_i = 1. \qquad \text{(III.4.59)}$$

Thus, the behavior of a Z can be classified in terms of the largest eigenvalue of its infinitesimal generator, $\lambda_{\max}(\mathbf{A})$. More precisely, the process Z is called supercritical, critical, or subcritical according as $\lambda_{\max}(\mathbf{A})$ is larger, equal, or smaller than zero.

Observe that the following properties are equivalent (cf. [112] p. 95-99 and [109]):

Z is irreducible $\Leftrightarrow$ $\mathbf{M}$ is irreducible $\Leftrightarrow$ $\mathbf{A}$ is irreducible $\Leftrightarrow$ $\mathbf{M}(t)$ is irreducible for all

$t > 0 \quad \Leftrightarrow \quad \mathbf{M}(t) > 0$ for all $t > 0$.

In particular, irreducible implies positive regular. Note that a matrix is irreducible if it is not similar via a permutation to a block upper triangular matrix and that a Markov chain is irreducible if and only if the transition matrix is irreducible.

The following lemma is an expansion of Theorem 4 of [25] for multi-type branching processes.

Lemma III.4.7. *Let $(Z(t))_{t \geq 0}$ be a non-singular and irreducible, k-dimensional continuous time Markov branching process and $\mathbf{q}$ the extinction vector of Z, i.e.*

$$q_i \equiv \mathbb{P}[Z(t) = 0 \text{ for some } t \geq 0 | Z(0) = \mathbf{e}_i] \quad \text{for } 1 \leq i \leq k. \tag{III.4.60}$$

Furthermore, let $(t_K)_{K \geq 1}$ be a sequence of positive numbers such that $\ln(K) \ll t_K$, define $T_\rho \equiv \inf\{t \geq 0 : \sum_{i=1}^k Z_i(t) = \rho\}$ and assume that for all $i, j \in \{1, \ldots, k\}$ and $t \in [0, \infty)$

$$\mathbb{E}[Z_j(t) \ln(Z_j(t)) | Z(0) = \mathbf{e}_i] < \infty. \tag{III.4.61}$$

*(i) If Z is **subcritical**, i.e. $\lambda_{max}(\mathbf{A}) < 0$, then for any $\epsilon > 0$*

$$\lim_{K \to \infty} \mathbb{P}\left[T_0 \leq t_K \wedge T_{\lceil \epsilon K \rceil} \,\Big|\, Z(0) = \mathbf{e}_i\right] = 1 \qquad \text{for all } i \in \{1, \ldots, k\} \tag{III.4.62}$$

and

$$\lim_{K \to \infty} \inf_{\mathbf{x} \in \partial B_{\epsilon K}} \mathbb{P}[T_0 \leq t_K | Z(0) = \mathbf{x}] = 1, \quad \text{where } \partial B_{\epsilon K} \equiv \{\mathbf{x} \in \mathbb{N}_0^k : \sum_{i=1}^k x_i = \lceil \epsilon K \rceil\}. \tag{III.4.63}$$

Moreover, for $\bar{u} = \frac{\max_{1 \leq i \leq k} u_i}{\min_{1 \leq j \leq k} u_j}$ and for any $\epsilon > 0$,

$$\lim_{K \to \infty} \sup_{\mathbf{x} \in B_{\epsilon^2 K}} \mathbb{P}\left[T_{\lceil \epsilon K \rceil} \leq T_0 \,\Big|\, Z(0) = \mathbf{x}\right] \leq \bar{u}\epsilon, \quad \text{where } B_{\epsilon^2 K} \equiv \{\mathbf{x} \in \mathbb{N}_0^k : \sum_{i=1}^k x_i \leq \lceil \epsilon^2 K \rceil\}. \tag{III.4.64}$$

*(ii) If Z is **supercritical**, i.e. $\lambda_{max}(\mathbf{A}) > 0$, then for any $\epsilon > 0$ (small enough)*

$$\lim_{K \to \infty} \mathbb{P}\left[T_0 \leq t_K \wedge T_{\lceil \epsilon K \rceil} \,\Big|\, Z(0) = \mathbf{e}_i\right] = q_i \qquad \text{for all } i \in \{1, \ldots, k\} \tag{III.4.65}$$

and

$$\lim_{K \to \infty} \mathbb{P}\left[T_{\lceil \epsilon K \rceil} \leq t_K \,\Big|\, Z(0) = \mathbf{e}_i\right] = 1 - q_i \qquad \text{for all } i \in \{1, \ldots, k\}. \tag{III.4.66}$$

Moreover, conditionally on survival, the proportions of the different types present in the population converge almost surely as $t \to \infty$ to the corresponding ratios of the components of the eigenvector: for all $i = 1, \ldots, k$,

$$\lim_{t \to \infty} \frac{Z_i(t)}{\sum_{j=1}^k Z_j(t)} = \frac{v_i}{\sum_{j=1}^k v_j}, \quad \text{a.s. on } \{T_0 = \infty\}. \tag{III.4.67}$$

Proof. We start with the proof of (i). Since $Z(t)$ is in this case a subcritical irreducible continuous time branching process and $\mathbb{E}[Z_j(t) \ln(Z_j(t)) | Z(0) = \mathbf{e}_i] < \infty$, we obtain by applying Satz 6.2.7 of [112] the existence of a constant $C > 0$ such that

$$\lim_{t \to \infty} \frac{1 - q_i(t)}{e^{\lambda_{\max}(\mathbf{A})t}} = C u_i, \tag{III.4.68}$$

where $q_i(t) \equiv \mathbb{P}[Z(t) = 0 \mid Z(0) = \mathbf{e}_i]$. Moreover, we have a non-explosion condition. Thus, for all $\epsilon > 0$, either $T_{\lceil \epsilon K \rceil}$ equals infinity or it converges to infinity as $K \to \infty$. Putting both together, there exists a sequence s_K with $\lim_{K \to \infty} s_K = +\infty$ such that

$$\lim_{K \to \infty} \mathbb{P}\Big[T_0 \le t_K \wedge T_{\lceil \epsilon K \rceil}\Big| Z(0) = \mathbf{e}_i\Big] \ge \lim_{K \to \infty} \mathbb{P}[T_0 \le s_K \mid Z(0) = \mathbf{e}_i] = \lim_{K \to \infty} q_i(s_K) = 1. \quad \text{(III.4.69)}$$

The branching property implies that for all $\mathbf{x} \in \mathbb{N}^k$, $\mathbb{P}[Z(t) = 0 \mid Z(0) = \mathbf{x}] = \prod_{i=1}^{k}(q_i(t))^{x_i}$ (cf. [109] p. 25). So, we get

$$\inf_{\mathbf{x} \in \partial B_{\epsilon K}} \mathbb{P}[T_0 \le t_K \mid Z(0) = \mathbf{x}] = \inf_{\mathbf{x} \in \partial B_{\epsilon K}} \mathbb{P}[Z(t_K) = 0 \mid Z(0) = \mathbf{x}] = \inf_{\mathbf{x} \in \partial B_{\epsilon K}} \prod_{i=1}^{k}(q_i(t_K))^{x_i}. \quad \text{(III.4.70)}$$

For all $i \in \{1, \ldots, k\}$, $1 \ge (q_i(t_K))^{x_i} \ge (q_i(t_K))^{\lceil \epsilon K \rceil}$ and by (III.4.68) we have $1 - q_i(t_K) = O(e^{\lambda_{\max}(\mathbf{A})t_K})$. Furthermore,

$$\lim_{K \to \infty}\Big(1 + \frac{w_K}{K}\Big)^K = \exp(0) = 1 \qquad \text{for any sequence } (w_K)_{K \ge 1} \text{ with } \lim_{K \to \infty} w_K = 0 \quad \text{(III.4.71)}$$

implies that, for all t_K with $t_K \gg \ln(K)$ and $C > 0$, since $\lim_{K \to \infty} Ce^{\lambda_{\max}(\mathbf{A})t_k}\lceil \epsilon K \rceil = 0$,

$$\lim_{K \to \infty}(1 - Ce^{\lambda_{\max}(\mathbf{A})t_k})^{\lceil \epsilon K \rceil} = 1. \quad \text{(III.4.72)}$$

Thus, taking the limit $K \to \infty$ in (III.4.70), we archive the desired equation (III.4.63). To prove inequality (III.4.64) we use the fact that $(\sum_{i=1}^{k} u_i Z_i(t))e^{-\lambda_{\max}t}$ is a martingale (cf. [5], Prop. 2). By applying Doob's stopping theorem to the stopping time $T_{\lceil \epsilon K \rceil} \wedge T_0$ we obtain for all $\mathbf{x} \in B_{\epsilon^2 K}$ that

$$\mathbb{E}\left[\left(\sum_{i=1}^{k} u_i Z_i(T_{\lceil \epsilon K \rceil})\right)e^{-\lambda_{\max}(\mathbf{A})T_{\lceil \epsilon K \rceil}}\mathbb{1}_{\{T_{\lceil \epsilon K \rceil} < T_0\}}\Bigg| Z(0) = \mathbf{x}\right] = \sum_{i=1}^{k} u_i x_i. \quad \text{(III.4.73)}$$

Therefore, since $\lambda_{\max}(\mathbf{A}) < 0$ in the subcritical case,

$$\mathbb{E}\left[\min_{1 \le i \le k} u_i \lceil \epsilon K \rceil \mathbb{1}_{\{T_{\lceil \epsilon K \rceil} < T_0\}}\Bigg| Z(0) = \mathbf{x}\right] \le \max_{1 \le i \le k} u_i \lceil \epsilon^2 K \rceil, \quad \text{for all } \mathbf{x} \in B_{\epsilon^2 K}, \quad \text{(III.4.74)}$$

which implies (III.4.64).

Let us continue with proving (ii). Since $Z(t)$ is supercritical in this case, applying Theorem 5.7.2 of [6] yields that

$$\lim_{t \to \infty} Z(t)(\omega)e^{-\lambda_{\max}(\mathbf{A})t} = W(\omega)\mathbf{v} \quad \text{(III.4.75)}$$

exists a.s., where W is a nonnegative random variable. Furthermore, since we assumed that $\mathbb{E}[Z_j(t)\ln(Z_j(t))|Z(0) = \mathbf{e}_i] < \infty$, for all $i \in \{1, \ldots, k\}$,

$$\mathbb{P}[W = 0|Z(0) = \mathbf{e}_i] = q_i, \quad \mathbb{E}[W|Z(0) = \mathbf{e}_i] = u_i \quad \text{(III.4.76)}$$

and W has an absolutely continuous distribution on $(0, \infty)$. All coordinates of $\mathbf{v}$ are strictly positive and $W > 0$ a.s. on the event $\{\omega : T_0(\omega) = \infty\}$. Hence, we have

$$Z(t) = O\left(e^{\lambda_{\max}(\mathbf{A})t}\right) \quad \text{a.s.} \quad \text{on } \{T_0 = \infty\}. \quad \text{(III.4.77)}$$

This implies, for K large enough, $\mathbb{P}[Z(t_K) < \lceil \epsilon K \rceil, T_0 = \infty] = 0$ and thus

$$\lim_{K \to \infty} \mathbb{P}[T_0 = \infty, T_{\lceil \epsilon K \rceil} \ge t_K] = 0. \quad \text{(III.4.78)}$$

Note that we used that $t_K \gg \ln(K)$. Since $\mathbb{P}[T_0 = \infty | Z(0) = \mathbf{e}_i] = 1 - q_i$, we deduce (III.4.66). On the other hand, there exist two sequences s_K^1 and s_K^2, which converge to infinity as $K \to \infty$, such that, for K large enough, $s_K^1 \leq t_K \wedge T_{\lceil \epsilon K \rceil} \leq s_K^2$ a.s.. This implies (III.4.65), because for all $i \in \{1, \ldots k\}$ and $l = 1, 2$, hold $\lim_{K \to \infty} \mathbb{P}[T_0 < s_K^l | Z_0 = \mathbf{e}_i] = q_i$. Note that equation (III.4.67) is a simple consequence of (III.4.75). $\qquad\square$

Using these properties about multi-type branching processes we can now prove the theorem about the three steps of invasion.

Proof of Theorem III.4.6. The first invasion step. Let us introduce the following stopping times

$$
\theta_{\text{exit}}^{K, M\epsilon} \;=\; \inf\left\{ t \geq 0 : \|\nu_t^K - \textstyle\sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p}) \delta_x\|_{TV} > M\epsilon \right\} \tag{III.4.79}
$$

$$
\tilde{\theta}_\epsilon^K \;=\; \inf\left\{ t \geq 0 : \nu_t^K(\tilde{g}) \geq \epsilon \right\} \tag{III.4.80}
$$

$$
\tilde{\theta}_0^K \;=\; \inf\left\{ t \geq 0 : \nu_t^K(\tilde{g}) = 0 \right\} \tag{III.4.81}
$$

Until $\tilde{\theta}_\epsilon^K$ the mutant population $\nu_t^K(\tilde{g})$ influences only the death rate of the resident population and this perturbation is uniformly bounded by $c_{\max}\epsilon$, where $c_{\max} \equiv \max_{x, \hat{x} \in \mathcal{X}} c(x, \hat{x})$. Thus, by applying Theorem III.4.4 (ii), we obtain

$$
\lim_{K \to \infty} \mathbb{P}\left[\theta_{\text{exit}}^{K, M\epsilon} < e^{KV} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_\epsilon^K \right] = 0. \tag{III.4.82}
$$

On the time interval $[0, \theta_{\text{exit}}^{K, M\epsilon} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_\epsilon^K]$, the resident population can be approximated by $\sum_{x \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} \bar{\mathbf{n}}_x(\mathbf{g}, \mathbf{p}) \delta_x$ and no further mutant appears. This allows us to approximate $\nu_t^K(\tilde{g})$ by multi-type branching processes.

Let $k \equiv |[\tilde{p}]_{\tilde{g}}|$. Similar as in the controlling results of Section II.6 we construct two $(\mathbb{N}_0)^k$-valued processes $X^{1,\epsilon}(t)$ and $X^{2,\epsilon}(t)$, using the pathwise definition in terms of Poisson point measures of ν_t^K, which control the mutant population $\nu_t^K(\tilde{g})$. To this aim let us denote the elements of $[\tilde{p}]_{\tilde{g}}$ by $\tilde{p} = \tilde{p}_1, \ldots, \tilde{p}_k$. Then, we define $X^{1,\epsilon}$ by

$$
X^{1,\epsilon}(t) \tag{III.4.83}
$$

$$
\equiv X^{1,\epsilon}(0) + \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{ i \leq X_j^{1,\epsilon}(s-), \, \theta \leq b(\tilde{p}_i) - \epsilon \right\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{birth}}(ds, di, d\theta)
$$

$$
- \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{ i \leq X_j^{1,\epsilon}(s-), \, \theta \leq d(\tilde{p}_j) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_j, p) \bar{\mathbf{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) + c_{\max} M\epsilon \right\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{death}}(ds, di, d\theta)
$$

$$
+ \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{[\tilde{p}]_{\tilde{g}}} \mathbb{1}_{\left\{ i \leq X_j^{1,\epsilon}(s-), \, \theta \leq s^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l) \right\}} (\mathbf{e}_l - \mathbf{e}_j) N_{(\tilde{g}, \tilde{p}_j)}^{\text{switch}}(ds, di, d\theta, d\tilde{p}_l),
$$

and similar $X^{2,\epsilon}$ by

$$
X^{2,\epsilon}(t) \tag{III.4.84}
$$

$$
\equiv X^{2,\epsilon}(0) + \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{ i \leq X_j^{2,\epsilon}(s-), \, \theta \leq b(\tilde{p}_i) + \epsilon \right\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{birth}}(ds, di, d\theta)
$$

$$
- \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \mathbb{1}_{\left\{ i \leq X_j^{2,\epsilon}(s-), \, \theta \leq d(\tilde{p}_j) + \sum_{(g,p) \in \mathcal{X}_{(\mathbf{g}, \mathbf{p})}} c(\tilde{p}_j, p) \bar{\mathbf{n}}_{(g,p)}(\mathbf{g}, \mathbf{p}) - c_{\max} M\epsilon \right\}} \mathbf{e}_j N_{(\tilde{g}, \tilde{p}_j)}^{\text{death}}(ds, di, d\theta)
$$

$$
+ \sum_{j=1}^k \int_0^t \int_{\mathbb{N}_0} \int_{\mathbb{R}_+} \int_{[\tilde{p}]_{\tilde{g}}} \mathbb{1}_{\left\{ i \leq X_j^{2,\epsilon}(s-), \, \theta \leq s^{\tilde{g}}(\tilde{p}_j, \tilde{p}_l) \right\}} (\mathbf{e}_l - \mathbf{e}_j) N_{(\tilde{g}, \tilde{p}_j)}^{\text{switch}}(ds, di, d\theta, d\tilde{p}_l),
$$

where $\mathbf{e}_j$ is the j-th unit vector in $\mathbb{R}^k$ and N^{birth}, N^{death}, and N^{switch} are the collections of Poisson point measures defined in Subsection III.4.1. Note that $X^{1,\epsilon}(t)$ and $X^{2,\epsilon}(t)$ are k-type branching processes with the following dynamics: For each $1 \le i \le k$, each individual in $X^{1,\epsilon}(t)$, respectively $X^{2,\epsilon}(t)$, with trait $(\tilde{g}, \tilde{p}_i)$ undergoes

 (i) birth (without mutation) with rate $b(\tilde{p}_i) - \epsilon$, respectively $b(\tilde{p}_i) + \epsilon$

 (ii) death with rate $D_{(\mathbf{g},\mathbf{p})}(\tilde{p}_i) + c_{\max} M \epsilon$, respectively $D_{(\mathbf{g},\mathbf{p})}(\tilde{p}_i) - c_{\max} M \epsilon$,
     where $D_{(\mathbf{g},\mathbf{p})}(\tilde{p}_i) = d(\tilde{p}_i) + \sum_{(g,p)\in\mathcal{X}_{(\mathbf{g},\mathbf{p})}} c(\tilde{p}_i, p) \bar{\mathfrak{n}}_{(g,p)}(\mathbf{g}, \mathbf{p})$

(iii) switching to phenotype $\tilde{p}_j$ with rate $s^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j)$ for all $1 \le j \le k$ (in both processes).

Furthermore, the processes $X^{1,\epsilon}(t)$ and $X^{2,\epsilon}(t)$ have the following property: There exists a $K_0 > 1$ such that for all $\tilde{p}_i \in [\tilde{p}]_{\tilde{g}}$ and for all $K \ge K_0$

$$\forall\, 0 \le t \le \theta_{\text{exit}}^{K,\epsilon} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_\epsilon^K : \qquad X_i^{1,\epsilon}(t) \le \nu_t^K(\tilde{g}, \tilde{p}_i) K \le X_i^{2,\epsilon}(t). \tag{III.4.85}$$

Hence, if $\tilde{\theta}_\epsilon^K \le \theta_{\text{exit}}^{K,\epsilon} \wedge \tau_{\text{mut.}}$, then

$$\inf\{t \ge 0 : X^{2,\epsilon}(t) = \lceil \epsilon K \rceil\} \le \tilde{\theta}_\epsilon^K \le \inf\{t \ge 0 : X^{1,\epsilon}(t) = \lceil \epsilon K \rceil\}. \tag{III.4.86}$$

On the other hand, if $\inf\{t \ge 0 : X^{2,\epsilon}(t) = 0\} \le \inf\{t \ge 0 : X^{2,\epsilon}(t) = \lceil \epsilon K \rceil\} \wedge \theta_{\text{exit}}^{K,\epsilon} \wedge \tau_{\text{mut.}}$, then

$$\tilde{\theta}_0^K \le \inf\{t \ge 0 : X^{2,\epsilon}(t) = 0\}. \tag{III.4.87}$$

Next, let us identify the infinitesimal generator of the control processes $X^{1,\epsilon}$ and $X^{2,\epsilon}$. Therefore, define, for $i = 1, \ldots, k$,

$$f_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}_i) \equiv b(\tilde{p}_i) - D_{(\mathbf{g},\mathbf{p})}(\tilde{p}_i) - \sum_{j=1}^k s^{\tilde{g}}(\tilde{p}_i, \tilde{p}_j). \tag{III.4.88}$$

(Note that $f_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}_i)$ would be the invasion fitness of phenotype $\tilde{p}_i$ if there was no switch back from the other phenotypes to $\tilde{p}_i$.) Then, by Equation (III.4.58), the infinitesimal generators are given by the following matrixes

$$\mathbf{A}(X^l) = \begin{pmatrix} f_{(\mathbf{g},\mathbf{p})}^{l,\epsilon}(\tilde{g}, \tilde{p}_1) & s^{\tilde{g}}(\tilde{p}_1, \tilde{p}_2) & \ldots & s^{\tilde{g}}(\tilde{p}_1, \tilde{p}_k) \\ s^{\tilde{g}}(\tilde{p}_2, \tilde{p}_1) & f_{(\mathbf{g},\mathbf{p})}^{l,\epsilon}(\tilde{g}, \tilde{p}_2) & & \\ \vdots & & \ddots & \vdots \\ s^{\tilde{g}}(\tilde{p}_k, \tilde{p}_1) & & \ldots & f_{(\mathbf{g},\mathbf{p})}^{l,\epsilon}(\tilde{g}, \tilde{p}_k) \end{pmatrix} \quad \text{for } l = 1, 2, \tag{III.4.89}$$

where $f_{(\mathbf{g},\mathbf{p})}^{1,\epsilon}(\tilde{g}, \tilde{p}_i) \equiv f_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}_i) - \epsilon(1 + c_{\max} M)$ and $f_{(\mathbf{g},\mathbf{p})}^{2,\epsilon}(\tilde{g}, \tilde{p}_i) \equiv f_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}_i) + \epsilon(1 + c_{\max} M)$.

We prove in the following that the number of mutant individuals grow with positive probability to ϵK before dying out if and only if $\lambda_{\max}$ of $\mathbf{A}_{(\tilde{g},\tilde{p})} \equiv \lim_{\epsilon \to 0} \mathbf{A}(X^{1,\epsilon})$ is strictly positive. Thus, $\lambda_{\max}(\mathbf{A}_{(\tilde{g},\tilde{p})})$ is an appropriate generalization of the invasion fitness of the class $[\tilde{p}]_{\tilde{g}}$:

$$F_{[\tilde{p}]_{\tilde{g}}}(\mathbf{g}, \mathbf{p}) \equiv \lambda_{\max}(\mathbf{A}_{(\tilde{g},\tilde{p})}). \tag{III.4.90}$$

Since the birth and death rates of $X^{1,\epsilon}$ and $X^{2,\epsilon}$ are positive and since Assumption 6 implies that $\mathbf{M}(X^{1,\epsilon})$ and $\mathbf{M}(X^{2,\epsilon})$ are irreducible, we obtain that the processes $X^{1,\epsilon}$ and $X^{2,\epsilon}$ are

non-singular and irreducible. Thus, $X^{1,\epsilon}$ and $X^{2,\epsilon}$ full fill the conditions of Lemma III.4.7. For $l = 1, 2$, let $\mathbf{q}(X^{l,\epsilon})$ denote the extinction probability vector of $X^{l,\epsilon}$, i.e.

$$\mathbf{q}(X^{l,\epsilon}) \equiv (q_1(X^{l,\epsilon}), \ldots, q_k(X^{l,\epsilon})), \quad \text{where } q_i(X^{l,\epsilon})) \equiv \mathbb{P}\left[X^{l,\epsilon}(t) = 0 \text{ for some } t \,\big|\, X^{l,\epsilon}(0) = \mathbf{e}_i\right].$$

Note that $\mathbf{q}(X^{l,\epsilon}) = (1, \ldots, 1)$ if $X^{l,\epsilon}$ is not supercritical. To characterize $\mathbf{q}(X^{l,\epsilon})$ in the supercritical case, let us introduce the following function

$$\mathbf{u} : [0,1]^k \times (-\eta, \eta) \to \mathbb{R}^k, \quad \text{where } \eta \text{ is some small enough constant,} \qquad \text{(III.4.91)}$$

defined, for all $1 \le i \le k$, by

$$
\begin{aligned}
\mathbf{u}_i(\mathbf{y}, \epsilon) \quad \equiv \quad & (b(\tilde{p}_i) - \epsilon)\, y_i^2 + \sum_{j=1}^k s^g(\tilde{p}_i, \tilde{p}_j)\, y_j + D_{(\mathbf{g},\mathbf{p})}(\tilde{p}_i) + c_{\max} M \epsilon \qquad \text{(III.4.92)} \\
& - \Big(b(\tilde{p}_i) - \epsilon + \sum_{j=1}^k s^g(\tilde{p}_i, \tilde{p}_j) + D_{(\mathbf{g},\mathbf{p})}(\tilde{p}_i) + c_{\max} M \epsilon \Big)\, y_i.
\end{aligned}
$$

Observe that $\mathbf{u}(\mathbf{y}, \epsilon)$ is the infinitesimal generating function of $X^{1,\epsilon}$ and $\mathbf{u}(\mathbf{y}, -\epsilon)$ the one of $X^{2,\epsilon}$. Furthermore, the extinction vector of a multi-type branching process is given as the smallest solution such that the generating function is zero (cf. [6] p. 205 or [112] Chap. 5). Thus, in the supercritical case $\mathbf{q}(X^{1,\epsilon})$ is the unique solution of

$$\mathbf{u}(\mathbf{y}, \epsilon) = 0 \qquad \text{for } \mathbf{y} \in [0,1)^k \qquad \text{(III.4.93)}$$

and $\mathbf{q}(X^{2,\epsilon})$ is the unique solution of

$$\mathbf{u}(\mathbf{y}, -\epsilon) = 0 \qquad \text{for } \mathbf{y} \in [0,1)^k. \qquad \text{(III.4.94)}$$

These solutions are in general not analytic. Applying Lemma III.4.7 to $X^{1,\epsilon}$ and $X^{2,\epsilon}$ we obtain that there exists $C_1 > 0$ such that for all $\eta > 0$, $\epsilon > 0$ sufficiently small and K large enough

$$
\begin{aligned}
\mathbb{P}\left[\theta_{\text{No Jump}}^{K,M\epsilon} < \tfrac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_\epsilon^K\right] \quad &\ge \quad \mathbb{P}\left[\inf\{t \ge 0 : X^{2,\epsilon}(t) = 0\} < \tfrac{\eta}{Ku_K}\right] \\
&\ge \quad q_1(X^{2,\epsilon}) - C_1\epsilon \qquad \text{(III.4.95)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}\left[\tilde{\theta}_\epsilon^K < \tfrac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}} \wedge \tilde{\theta}_0^K\right] \quad &\ge \quad \mathbb{P}\left[\inf\{t \ge 0 : X^{1,\epsilon}(t) = 0\} < \tfrac{\eta}{Ku_K}\right] \\
&\ge \quad 1 - q_1(X^{1,\epsilon}) - C_1\epsilon. \qquad \text{(III.4.96)}
\end{aligned}
$$

If $X^{2,\epsilon}$ is sub- or critical for ϵ small enough, then $\lim_{\epsilon \downarrow 0} q_1(X^{2,\epsilon}) = \lim_{\epsilon \downarrow 0} q_1(X^{1,\epsilon}) = 1$. In the supercritical case, let $\mathbf{q} \in [0,1)^k$ be the solution of $\mathbf{u}(\mathbf{y}, 0) = 0$. Then, by applying the implicit function theorem there exists an open set $U \subset \mathbb{R}$ containing 0, an open set $V \subset \mathbb{R}^k$ containing $\mathbf{q}$, and a unique continuously differentiable function $g : U \to V$ such that

$$\{(\epsilon, g(\epsilon)) | \epsilon \in U\} = \{(\epsilon, \mathbf{y}) \in U \times V | \mathbf{u}(\mathbf{y}, \epsilon) = 0\}. \qquad \text{(III.4.97)}$$

By definition, $g(0) = \mathbf{q}$ and $q_1 = q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p})$. We can linearize and obtain that there exists a constant $C_2 > 0$ such that

$$q_1(X^{1,\epsilon}) \le q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}) + C_2\epsilon \quad \text{and} \quad q_1(X^{2,\epsilon}) \ge q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}) - C_2\epsilon \qquad \text{(III.4.98)}$$

Therefore,

$$\lim_{K \to \infty} \mathbb{P}\left[\theta_{\text{No Jump}}^{K,M\epsilon} \wedge \tilde{\theta}_\epsilon^K < \frac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}}\right] \geq 1 - 2(C_1 + C_2)\epsilon. \tag{III.4.99}$$

Conditionally on survival, the proportions of the different phenotypes in $X^{1,\epsilon}$ converge almost surely as $t \to \infty$ to the corresponding ratios of the components of the eigenvector, which are all strictly positive (cf. Lemma III.4.7, Eq. (III.4.67)). Furthermore, there exists a constant $C_3 > 0$ such that for all ϵ small enough

$$\lim_{K \to \infty} \mathbb{P}\left[\left\{\tilde{\theta}_\epsilon^K < \frac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}}\right\} \cap \left\{\inf\{t \geq 0 : X^{1,\epsilon}(t) = 0\} < \infty\right\}\right] < C_3\epsilon \tag{III.4.100}$$

and $\tilde{\theta}_\epsilon^K$ converges to infinity as $K \to \infty$. Thus, conditionally on $\{\tilde{\theta}_\epsilon^K < \frac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}}\}$, there exists a (small) constant $C_4 > 0$ such that the probability that the densities of the phenotypes $\{\tilde{p}_1, \ldots, \tilde{p}_k\}$, are all larger than $C_4\epsilon$ at time $\tilde{\theta}_\epsilon^K$ convergences to one as first $K \to \infty$ and then $\epsilon \to 0$. More precisely, there exists constants $C_4 > 0$ and $C_5 > 0$ such that for all ϵ small enough,

$$\lim_{K \to \infty} \mathbb{P}\left[\tilde{\theta}_\epsilon^K < \frac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}}, \exists i \in \{1, \ldots k\} : \nu_{\tilde{\theta}_\epsilon^K}^K(\tilde{p}_i) \leq C_4\epsilon\right] \leq C_5\epsilon. \tag{III.4.101}$$

The second invasion step. By Assumption 7, any solution of $LVS(d+1, ((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p})))$ with initial state in the compact set

$$A \equiv \{x \in \mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|} : |x - \bar{\mathfrak{n}}(\mathbf{g}, \mathbf{p})| \leq M\epsilon\} \times [C_4\epsilon, \epsilon]^k \tag{III.4.102}$$

converge as $t \to \infty$ to the unique locally strictly stable equilibrium $\mathfrak{n}^*((\mathbf{g}, \mathbf{p})), (\tilde{g}, \tilde{p}))$. Therefore, for all $\epsilon > 0$ there exists $T(\epsilon) \in \mathbb{R}$ such that any of these trajectories do not leave the set

$$\{x \in \mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|+k} : |x - \mathfrak{n}^*((\mathbf{g}, \mathbf{p})), (\tilde{g}, \tilde{p}))| \leq \epsilon^2/2\} \tag{III.4.103}$$

after time $T(\epsilon)$. Back to the stochastic system, let us introduce on the event $\{\tilde{\theta}_\epsilon^K < \frac{\eta}{Ku_K} \wedge \theta_{\text{exit}}^{K,M\epsilon} \wedge \tau_{\text{mut.}}\}$ the following stopping time

$$\theta_{\text{near } \mathfrak{n}^*}^{K,\epsilon} = \inf\left\{t \geq \tilde{\theta}_\epsilon^K : \left\|\nu_t^K - \sum_{x \in \mathcal{X}_{((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))}} \mathfrak{n}_x^*((\mathbf{g}, \mathbf{p}), (\tilde{g}, \tilde{p}))\delta_x\right\|_{TV} < \epsilon^2\right\}. \tag{III.4.104}$$

Then, we conclude by using the strong Markov property at $\tilde{\theta}_\epsilon^K$ and Theorem III.4.4 (i) on $[0, T(\epsilon)]$ that there exists a constant $C_6 > 0$ such that, for all ϵ small enough,

$$\lim_{K \to \infty} \mathbb{P}\left[\tilde{\theta}_\epsilon^K < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K} \text{ and } \sup_{s \in [\tilde{\theta}_\epsilon^K, \tilde{\theta}_\epsilon^K + T(\epsilon)]} \left\|\nu_s^K - \sum_{x \in \mathcal{X}_{(\mathbf{g},\mathbf{p})}} \mathfrak{n}_x(s, \nu_0^K)\delta_x\right\|_{TV} \leq \epsilon^2\right] \tag{III.4.105}$$

$$\geq 1 - q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}) - C_6\epsilon,$$

which implies

$$\lim_{K \to \infty} \mathbb{P}\left[\tilde{\theta}_\epsilon^K < \theta_{\text{near } \mathfrak{n}^*}^{K,\epsilon} < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K}\right] \geq 1 - q_{(\mathbf{g},\mathbf{p})}(\tilde{g}, \tilde{p}) - C_6\epsilon. \tag{III.4.106}$$

We used that, at time $\tilde{\theta}_\epsilon^K$, the stochastic process ν^K (considered as element of $\mathbb{R}^{|\mathcal{X}_{(\mathbf{g},\mathbf{p})}|+k}$) lies in the compact set A, where A is defined in (III.4.102).

The third invasion step. After time $\theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*}$ we use again comparisons with multi-type branching processes to show that all individuals carrying a trait which is not present in the new equilibrium $\mathfrak{n}^*$ die out. To this aim let us define

$$\mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}} = \{(g,p) \in \mathcal{X}_{((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))} : \mathfrak{n}^*_{(g,p)}((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p})) = 0\} \qquad (\text{III.4.107})$$

For proving that the populations with traits in $\mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}$ stay small after $\theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*}$ and that the populations with traits not in $\mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}$ stay close to its equilibrium value after $\theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*}$, let us define

$$\theta^{K,\epsilon}_{\text{not small}} = \inf\left\{t \geq \theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*} : \exists (g,p) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}} \text{ such that } \nu^K_t(g,p) > \epsilon\right\} \qquad (\text{III.4.108})$$

and

$$\theta^{K,M\epsilon}_{\text{exit } \mathfrak{n}^*} \equiv \inf\left\{t \geq \theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*} : \left\|\nu^K_t - \sum_{x \in \mathcal{X}_{((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))}} \mathfrak{n}^*_x((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))\delta_x\right\|_{TV} > M\epsilon\right\}. \qquad (\text{III.4.109})$$

By using first the strong Markov property at $\theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*}$, we can apply Theorem III.4.4 (ii) and obtain that there exist constants $M > 0$ and $C_7 > 0$ such that for all ϵ small enough

$$\lim_{K \to \infty} \mathbb{P}\left[\tilde{\theta}^K_\epsilon < \theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*} < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K} \text{ and } \theta^{K,M\epsilon}_{\text{exit } \mathfrak{n}^*} < e^{KV} \wedge \tau_{\text{mut.}} \wedge \theta^{K,\epsilon}_{\text{not small}}\right] < C_7\epsilon \qquad (\text{III.4.110})$$

This is obtained in a similar way as Equation (III.4.82) in the first step. Note that $(g,p) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}$ implies that $(g,p_i) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}$ for all $p_i \in [p]_g$, which is a consequence of Assumption 6.

Using the same arguments as in the first step, we can construct, for all $(g,p) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}$, a $|[p]_g|$-type continuous time branching process $Y^{\epsilon,(g,p)}(s)$ with initial condition

$$Y^{\epsilon,(g,p)}_i(0) = \nu^K_{\theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*}}(g,p_i)K \qquad \text{for all } p_i \in [p]_g \qquad (\text{III.4.111})$$

such that for all K large enough and for all $t \in [\theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*}, \theta^{K,M\epsilon}_{\text{exit } \mathfrak{n}^*} \wedge \theta^{K,\epsilon}_{\text{not small}} \wedge \tau_{\text{mut.}}]$

$$\nu^K_t(g,p_i)K \leq Y^{\epsilon,(g,p)}_i(t - \theta^{K,\epsilon}_{\text{near } \mathfrak{n}^*}) \qquad \text{for all } p_i \in [p]_g. \qquad (\text{III.4.112})$$

Furthermore, $Y^{\epsilon,(g,p)}(t)$ is characterized as follows: For each $p_i \in [p]_g$, each individual in $Y^{\epsilon,(g,p)}(t)$ with trait (g,p_i) undergoes

(i) birth (without mutation) with rate $b(p_i)$

(ii) death with rate $d(p_i) + \sum_{(\hat{g},\hat{p}) \in \mathcal{X}_{(\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p})}} c(p_i,\hat{p})\mathfrak{n}^*_{(g,p)}((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p})) - c_{\max}(M + |\mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}|)\epsilon$

(iii) for all $1 \leq j \leq |[p]_g|$ switch from phenotype p_i to p_j with rate $s^g(p_i,p_j)$.

Let $\mathbf{A}(Y^{\epsilon,(g,p)})$ denote the infinitesimal generator of the process $Y^{\epsilon,(g,p)}$. Since the equilibrium $\mathfrak{n}^*((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))$ is locally strictly stable (cf. Assumption 7), the eigenvalues of the Jacobian matrix of the dynamical system at $\mathfrak{n}^*((\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p}))$ are all strictly negative. If ϵ is small enough, this implies that all eigenvalues of $\{\mathbf{A}(Y^{\epsilon,(g,p)}), (g,p) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}\}$ are strictly negative. (There exists an order of the elements of $\mathcal{X}_{(\mathbf{g},\mathbf{p}),(\tilde{g},\tilde{p})}$ such that the Jacobian matrix is an upper-block-triangular matrix and $\{\mathbf{A}(Y^{0,(g,p)}), (g,p) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}\}$ are on the diagonal.) Thus, for all ϵ small enough, the branching processes $\{Y^{\epsilon,(g,p)}, (g,p) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}\}$ are all subcritical. Moreover, we can apply Lemma III.4.7 and get, for all ϵ small enough and $(g,p) \in \mathcal{X}^{\mathfrak{n}^*}_{\text{extinct}}$

$$\lim_{K \to \infty} \mathbb{P}\left[\inf\{t \geq 0 : Y^{\epsilon,(g,p)}(t) = 0\} \leq \frac{\eta}{Ku_K}\right] = 1, \qquad (\text{III.4.113})$$

and there exists a constant C_8 such that for all ϵ small enough and $(g,p) \in \mathcal{X}_{\text{extinct}}^{\mathfrak{n}^*}$

$$\lim_{K \to \infty} \mathbb{P}\left[\inf\{t \geq 0 : Y^{\epsilon,(g,p)}(t) = \lceil \epsilon K \rceil\} \leq \inf\{t \geq 0 : Y^{\epsilon,(g,p)}(t) = 0\}\right] \leq C_8 \epsilon. \qquad \text{(III.4.114)}$$

Hence, there exists a constant $M > 0$ and $C_9 > 0$ such that, for all $\eta > 0$ and ϵ small enough,

$$\lim_{K \to \infty} \mathbb{P}\left[\tilde{\theta}_\epsilon^K < \theta_{\text{Jump}}^{K,M\epsilon} < \tau_{\text{mut.}} \wedge \frac{\eta}{Ku_K} \wedge \theta_{\text{not small}}^{K,\epsilon}\right] \geq 1 - q_{(\mathbf{g},\mathbf{p})}(\tilde{g},\tilde{p}) - C_9 \epsilon, \qquad \text{(III.4.115)}$$

which finishes the proof of the theorem. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Combining all the previous results, we can prove similar as in [25] that for all $\epsilon > 0, t > 0$ and $\Gamma \subset \mathcal{X}$,

$$\lim_{K \to \infty} \mathbb{P}\Big[\text{Supp}(\nu_{t/Ku_K}^K) = \Gamma, \text{ all traits of } \Gamma \text{ coexist in } LVS(|\Gamma|,\Gamma), \qquad\qquad\qquad \text{(III.4.116)}$$

$$\text{and } \|\nu_{t/Ku_K}^K - \sum_{x \in \Gamma} \bar{\mathbf{n}}_x(\Gamma)\delta_x\|_{TV} < \epsilon\Big] = \mathbb{P}[\text{Supp}(\Lambda_t) = \Gamma]$$

where Λ is the PES with phenotypic plasticity defined in Theorem III.4.3. Finally, generalizing this to any sequence of times $0 < t_1 < \ldots < t_n$, implies that $(\nu_{t/Ku_K}^K)_{t \geq 0}$ converges in the sense of finite dimensional distributions to $(\Lambda_t)_{t \geq 0}$ (cf. [25] Corollary 1 and Lemma 1), which ends the proof of Theorem III.4.3.

Examples.

Figure III.9 shows examples, where in a population consisting only of type (g,p) and being close to $\mathfrak{n}(g,p)$ a mutation to genotype $\tilde{g}$ occurs. In this example, $\tilde{g}$ is associated with two possible phenotypes $\tilde{p}_1$ and $\tilde{p}_2$.
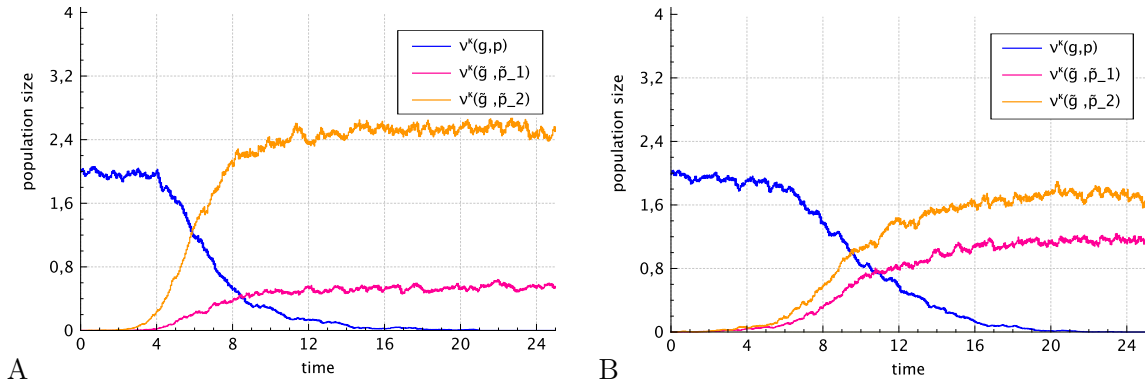

Figure III.9: Simulations of the invasion phase with $K = 1000$. (A) The mutant phenotype $\tilde{p}_1$ has a negative initial growth rate but can switch to $\tilde{p}_2$ which has a positive one. The fitness of the genotype $\tilde{g}$ is positive. (B) The fitness of the mutant genotype $\tilde{g}$ is positive, although each phenotype has a negative initial growth rate. This is possible because an outgoing switch is a loss of a cell for a phenotype, but not for the whole genotype.

In example (A), we start with a single mutant carrying trait $(\tilde{g},\tilde{p}_1)$ and which can switch to $\tilde{p}_2$ but the back-switch is relative weak (cf. Table III.3). According to definition (III.4.88) we have $f_{(g,p)}(\tilde{g},\tilde{p}_1) < 0$ and $f_{(g,p)}(\tilde{g},\tilde{p}_2) > 0$. However, the global fitness of the genotype $\tilde{g}$

is positive. More precisely, it is given by the largest eigenvalue of $\left(\begin{smallmatrix} -3 & 2 \\ 0.6 & 1 \end{smallmatrix}\right)$, which equals approximatively 1.280. Therefore, the multi-type branching process approximating the mutant population in the first step is supercritical. This does not depend on the ion the phenotypic trait of the first mutant, i.e. we would have the same if we would have started with a single mutant carrying trait $(\tilde{g}, \tilde{p}_2)$). However, the probability of invasion depends this. In this example, the invasion probability is given by the solution of

$$2y_1^2 + 2y_2 + 3 - 7y_1 = 0, \qquad (\text{III.4.117})$$

$$4y_2^2 + 0.6y_1 + 2.4 - 7y_2 = 0. \qquad (\text{III.4.118})$$

Thus, if we start with the trait $(\tilde{g}, \tilde{p}_1)$, the invasion probability is approximately 0.199. Whereas it is 0.338 if the first one has trait $(\tilde{g}, \tilde{p}_2)$. In Figure III.9 (A), the mutant population with genotype $\tilde{g}$ survives and the stochastic process is attracted to the new equilibrium $\mathfrak{n}^*((g,p),(\tilde{g},\tilde{p}_1),(\tilde{g},\tilde{p}_2)) \approx (0, 0.543, 2.554)$, which is a strictly stable.

$b(p) = 3$	$d(p) = 1$	$c(p,p) = 1$	$c(p,\tilde{p}_1) = 1$	$c(p,\tilde{p}_2) = 0.7$	-	$\nu_0^K(g,p) = 2$
$b(\tilde{p}_1) = 2$	$d(\tilde{p}_1) = 1$	$c(\tilde{p}_1,p) = 1$	$c(\tilde{p}_1,\tilde{p}_1) = 1$	$c(\tilde{p}_1,\tilde{p}_2) = 0.5$	$s^{\tilde{g}}(\tilde{p}_1,\tilde{p}_2) = 2$	$\nu_0^K(\tilde{g},\tilde{p}_1) = 1/K$
$b(\tilde{p}_2) = 4$	$d(\tilde{p}_2) = 1$	$c(\tilde{p}_2,p) = 0.7$	$c(\tilde{p}_2,\tilde{p}_1) = 0.5$	$c(\tilde{p}_2,\tilde{p}_2) = 1$	$s^{\tilde{g}}(\tilde{p}_2,\tilde{p}_1) = 0.6$	$\nu_0^K(\tilde{g},\tilde{p}_2) = 0$

Table III.3: Parameters of Figure III.9 (A)

In example (B), $f_{(g,p)}(\tilde{g},\tilde{p}_1)$ and $f_{(g,p)}(\tilde{g},\tilde{p}_2)$ are both negative. Nevertheless, the fitness of the genotype is positive and thus the mutant invades with positive probability. (It is given by the largest eigenvalue of $\left(\begin{smallmatrix} -3 & 2 \\ 2 & -0.4 \end{smallmatrix}\right)$, which equals approximatively 0.685.) However, the invasion probability is smaller in this example. It is approximately 0.127 if we start with the trait $(\tilde{g}, \tilde{p}_1)$ and 0.207 else. In Figure III.9 (B), the mutant population survives and the process is attracted to the stable fixed point $\mathfrak{n}^*((g,p),(\tilde{g},\tilde{p}_1),(\tilde{g},\tilde{p}_2)) \approx (0, 1.153, 1.745)$. Hence, this examples illustrate that the usual definition of invasion fitness fails for populations with phenotypic plasticity.

$b(p) = 3$	$d(p) = 1$	$c(p,p) = 1$	$c(p,\tilde{p}_1) = 1$	$c(p,\tilde{p}_2) = 0.7$	-	$\nu_0^K(g,p) = 2$
$b(\tilde{p}_1) = 2$	$d(\tilde{p}_1) = 1$	$c(\tilde{p}_1,p) = 1$	$c(\tilde{p}_1,\tilde{p}_1) = 1$	$c(\tilde{p}_1,\tilde{p}_2) = 0.5$	$s^{\tilde{g}}(\tilde{p}_1,\tilde{p}_2) = 2$	$\nu_0^K(\tilde{g},\tilde{p}_1) = 1/K$
$b(\tilde{p}_2) = 4$	$d(\tilde{p}_2) = 1$	$c(\tilde{p}_2,p) = 0.7$	$c(\tilde{p}_2,\tilde{p}_1) = 0.5$	$c(\tilde{p}_2,\tilde{p}_2) = 1$	$s^{\tilde{g}}(\tilde{p}_2,\tilde{p}_1) = 2$	$\nu_0^K(\tilde{g},\tilde{p}_2) = 0$

Table III.4: Parameters of Figure III.9 (B)

III.4.3 Interplay of mutation and therapy

In the previous subsection we considered the probability of invasion of a mutant when the resident population is at an equilibrium given by a stable fixed point. In the context of therapy, there are phases when populations shrink and regrow due to treatment and relapse phenomena. In the medical literature, there are frequent allusions to the possibility that such growth cycles may induce fixation of a "super-resistant mutant", see e.g. [60, 65, 66]. It is important to understand whether and under what circumstances such effects may happen.

In this subsection we show an example where the appearance of a mutant genotype may indeed be enhanced under treatment. We consider *birth-reducing competition* (BRC) between tumor cells. In such a case, a large population at equilibrium may encounter fewer births and hence mutations, than a smaller population growing towards its equilibrium size. Let us discuss in more detail how the birth-reducing competition can have a crucial effect on

the mutation time scale. For the sake of simplicity we ignore in this subsection the effect of switching the phenotype and the usual competition, i.e. we consider an example where all switching rates and the usual competition kernel are set to zero. Since the presence of TNF-α only influences the switch between phenotypes, it does not play any role. So, let us consider a melanoma population of type (g, p) which can be attacked by T-cells but cannot change its phenotype, i.e. it cannot escape therapy by switching. However, on a longer time scale it is able to mutate to a fitter type of melanoma $(\tilde{g}, \tilde{p})$, which cannot be attacked by the T-cells. Now we are interested whether therapy can lead to earlier mutations. This is the simplest scenario where the effect of therapy in this context can be explained.

As in the last subsection we are interested in the case of a large population size ($K \to \infty$) in combination with rare mutations ($u_K \to 0$). The interesting scaling of the mutation rate is $K \mu_g^K \to \alpha > 0$ as $K \to \infty$. Thus, mutational events happen slightly faster than in the last subsection. In [97], Mayer gives a more detailed analysis of the same scenario when mutational events happen as fast as in the last subsection, i.e. $K \mu_g^K \to 0$ as $K \to \infty$.

Observe that the total mutation rate of the population of type (g, p) at time t is

$$u_K m(g) \left\lfloor b(p) - c_b(p, p) \nu_t^K(g, p) \right\rfloor_+ \nu_t^K(g, p) K. \tag{III.4.119}$$

This is a positive and concave function of $\nu_t^K(g, p)$ on the interval $[0, b(p)/c_b(p, p)]$, see Figure III.10. Without therapy and before the first mutant appears, the melanoma population can
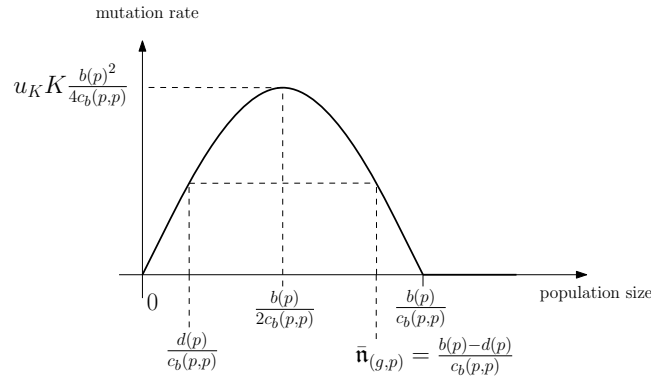

Figure III.10: Initial total mutation rate of the population (g, p).

be approximated by the solution of

$$\dot{\mathfrak{n}}_{(g,p)} = \mathfrak{n}_{(g,p)} \left([b(p) - c_b(p, p) \mathfrak{n}_{(g,p)}]_+ - d(p) \right) \tag{III.4.120}$$

Thus, if the melanoma population is close to its equilibrium $\bar{\mathfrak{n}}_{(g,p)} = (b(p) - d(p))/c_b(p, p)$ at the beginning, we can approximate the first mutation time by an exponential random variable with parameter

$$u_K K m(g) d(p) \bar{\mathfrak{n}}_{(g,p)}. \tag{III.4.121}$$

Since the mutation rate is maximal at $\nu_t^K(g, p) = \frac{b(p)}{2c_b(p,p)}$, we conclude that if $\bar{\mathfrak{n}}_{(g,p)} > \frac{b(p)}{2c_b(p,p)}$, then smaller populations (more precisely populations which stay a long enough time between $d(p)/c_b(p, p)$ and $\bar{\mathfrak{n}}_{(g,p)}$) have a higher total mutation rates. This effect is illustrated in Figure III.11. If the population starts below its equilibrium (cf. Figure III.11 (A)), the mutant occurs in the simulations in average earlier than when the population starts at equilibrium (cf. Figure III.11 (B)).
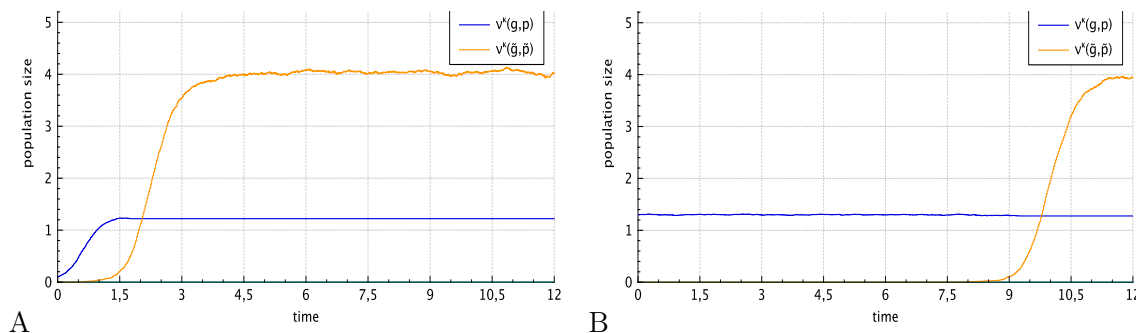
Figure III.11: Simulations of mutation events in a population, where competition is acting via birth reduction ($K = 10^3$ and $u_K = 1/K$). The number of individuals divided by 1000 is plotted versus time. Effect for an initial population which is small (A), or at equilibrium (B).

$b(p) = 4$	$d(p) = 0.1$	$c_b(p,p) = 3$	$c_b(p,\tilde{p}) = 0.8$	$m(g) = 1$	$\nu_0^K(g,p) = 0.1$ resp. 1.3
$b(\tilde{p}) = 6$	$d(\tilde{p}) = 1$	$c_b(\tilde{p},p) = 0.8$	$c_b(\tilde{p},\tilde{p}) = 1$		$\nu_0^K(\tilde{g},\tilde{p}) = 0$

Table III.5: Parameters of Figure III.11

During therapy, a tumor which is close to equilibrium can shrink to a small size: the injection of T-cells in the system lowers the population size of melanoma, and thus the total mutation rate in the tumor population of type (g, p) can be larger during treatment. This means that treatment could lead to earlier mutations and thereby accelerate the evolution towards more aggressive tumor variants.

The following example provides an interesting situation of interplay between therapy and mutation. By lowering the melanoma population, the T-cell therapy actually increases the probability for it to mutate to a potentially fitter and pathogenic genotype, which is not affected by the T-cells. Under treatment (Figure III.12), the fitter mutant occurs on average much earlier than without therapy (Figure III.11 (B)).
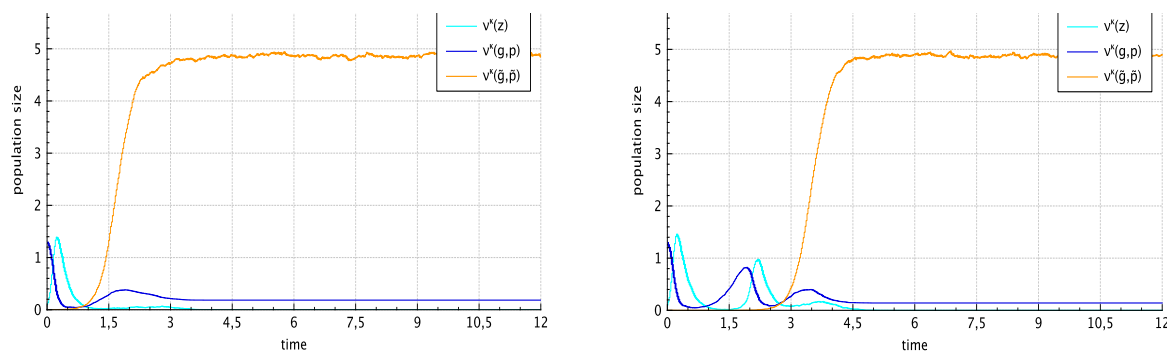

Figure III.12: Simulations of mutation events in a melanoma population under therapy, where competition is acting via birth reduction ($K = 10^3$ and $u_K = 1/K$). The mutant occurs earlier than in the simulation without therapy (cf. Figure III.11 (B)).

Another effect of therapy, which we don't discuss in detail, is that a mutant has a higher invasion probability if it appears in a smaller melanoma population, since the mutant competes for resources with less other melanoma cells. Furthermore, the proliferation of the mutant population is may be faster under therapy, too.

$b(p) = 4$	$d(p) = 0.1$	$c_b(p,p) = 3$	$c_b(p,\tilde{p}) = 0.8$	$m(g) = 1$	$\nu_0^K(g,p) = 1.3$
$b(\tilde{p}) = 6$	$d(\tilde{p}) = 1$	$c_b(\tilde{p},p) = 0.8$	$c_b(\tilde{p},\tilde{p}) = 1$		$\nu_0^K(\tilde{g},\tilde{p}) = 0$
$b(z,p) = 20$	$d(z) = 6$	$t(z,p) = 10$			$\nu_0^K(z) = 0.1$

Table III.6: Parameters of Figure III.12

III.5 Discussion and clinical relevance

Therapy resistance is a major issue in the treatment of advanced stages of cancer. We have proposed a stochastic mathematical model that allows to simulate treatment scenarios and applied it to the specific case of immunotherapy of melanomas. Comparison to experimental data is so far promising (cf. [9] and [97]). The models pose challenging new mathematical questions, in particular due to the interplay of fast phenotypic switches and rare driver mutations as we have seen e.g. in Section III.4. First numerical results point to a significant effect of stochastic fluctuations in the success of therapies. More precise experimental data will be needed in the future to fit crucial model parameters. While our models describe the actions of individual cells and cytokines, they do not by far resolve the full complexity of the biological system. In particular, they do not reflect the spatial structure of the tumor and its microenvironment. Also, the distinction of only two phenotypes of the tumor cells is a simplification. The same is true for the interaction with other immune cells and cytokines. This reflects on the one hand the limitation due to available experimental data, on the other hand the use of a model of reduced complexity also makes numerical computations and theoretical understanding of the key phenomena feasible. The rates entering as model parameters therefore have to be understood as *effective* parameters, e.g. the death rate of T-cells accounts for their natural death as well as the exhaustion phenomenon. In principle it is possible to increase the resolution of the model; this, however, increases the number of parameters that need to be determined experimentally which would pose a major challenge. Already at the present stage, the model parameters are not known well enough and are adjusted to reproduce the experimental findings. Some parameters that would be very useful to see measured precisely are:

(i) birth and death rates of tumor cells, both in differentiated and dedifferentiated types. Currently these are estimated from the growth rate of the tumor, but this yields only the difference of these rates;

(ii) killing rates of T-cells, both of the differentiated and the dedifferentiated tumor cells;

(iii) rates of phenotypic switches, both in the absence and the presence of TNF-α;

(iv) death rates of T-cells and their expansion rates when interacting with tumor cells.

Nevertheless, we see the proposed model as a promising tool to assist the development of improved treatment protocols. Simulations may guide the choice of experiments such that the number of necessary experiments can be reduced. The obvious strength of our approach is to model reciprocal interactions and phenotypic state transitions of tumor and immune cells in a heterogeneous and dynamic microenvironment in the context of therapeutic perturbations.

The clinical importance of phenotypic coevolution in response to therapy has been recently documented in patients' samples from melanomas acquiring resistance to MAPK inhibitors [77]. Of note and similar to our previous study, dedifferentiation of melanoma cells was identified as a major mechanism of escape from MAPK inhibitors [105, 87]. Hence, malignant melanoma is a paradigm for a phenotypic heterogeneous tumor and a future goal is to in-

corporate this increasing knowledge of melanoma cell plasticity into our method to refine its capability to model complex interactions with immune cells.

Importantly, phenotypic plasticity in response to therapy is a widespread phenomenon and non-small cell lung cancer represents a prominent example. Our mathematical approach could represent a valuable tool to support this research, too. Finally, our results suggest that stochastic events play an unanticipated critical role in the dynamic evolution of tumors and the emergence of therapy resistance that requires further experimental and clinical investigation.

III.6   Appendix

III.6.1   Infinitesimal generator

The measure-valued process $(\nu_t^K)_{t\geq 0}$ described in section III.2 is a Markov process whose law is characterized by its infinitesimal generator L^K (cf. [55] Chapter 11 and [59]). The generator acts on bounded measurable functions ϕ from $\mathcal{M}^K(\mathcal{X})$ into $\mathbb{R}$, for all $\mu^K \in \mathcal{M}^K(\mathcal{X})$ by

$$\left(L^K\phi\right)\left(\mu^K\right) \tag{III.6.1}$$

$$= \sum_{(g,p)\in\mathcal{G}\times\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_{(g,p)}}{K}\right) - \phi(\mu^K)\right)\left(1 - u_K m(g)\right)\left\lfloor b(p) - \sum_{\tilde{p}\in\mathcal{P}} c_b(p,\tilde{p})\mu^K(\tilde{p})\right\rfloor_+ K\mu^K(g,p)$$

$$+ \sum_{(g,p)\in\mathcal{G}\times\mathcal{P}} \sum_{(\tilde{g},\tilde{p})\in\mathcal{G}\times\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_{(\tilde{g},\tilde{p})}}{K}\right) - \phi(\mu^K)\right) u_K m(g) M((g,p),(\tilde{g},\tilde{p}))\left\lfloor b(p) - \sum_{p'\in\mathcal{P}} c_b(p,p')\mu^K(p')\right\rfloor_+ K\mu^K(g,p)$$

$$+ \sum_{(g,p)\in\mathcal{G}\times\mathcal{P}} \left(\phi\left(\mu^K - \tfrac{\delta_{(g,p)}}{K}\right) - \phi(\mu^K)\right)\left(d(p) + \sum_{\tilde{p}\in\mathcal{P}} c(p,\tilde{p})\mu^K(\tilde{p}) + \left\lfloor b(p) - \sum_{\tilde{p}\in\mathcal{P}} c_b(p,\tilde{p})\mu^K(\tilde{p})\right\rfloor_-\right)K\mu^K(g,p)$$

$$+ \sum_{(g,p)\in\mathcal{G}\times\mathcal{P}} \sum_{z\in\mathcal{Z}} \left(\phi\left(\mu^K - \tfrac{\delta_{(g,p)}}{K} + \textstyle\sum_{w\in\mathcal{W}} \ell_w^{\text{kill}}(z,p)\tfrac{\delta_w}{K}\right) - \phi(\mu^K)\right) t(z,p)\mu^K(z)K\mu^K(g,p)$$

$$+ \sum_{(g,p)\in\mathcal{G}\times\mathcal{P}} \sum_{\tilde{p}\in\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_{(g,\tilde{p})}}{K} - \tfrac{\delta_{(g,p)}}{K}\right) - \phi(\mu^K)\right)\left(s^g(p,\tilde{p}) + \textstyle\sum_{w\in\mathcal{W}} s_w^g(p,\tilde{p})\mu^K(w)\right)K\mu^K(g,p)$$

$$+ \sum_{z\in\mathcal{Z}} \left(\phi\left(\mu^K + \tfrac{\delta_z}{K}\right) - \phi(\mu^K)\right) b(z)K\mu^K(z)$$

$$+ \sum_{z\in\mathcal{Z}} \left(\phi\left(\mu^K - \tfrac{\delta_z}{K}\right) - \phi(\mu^K)\right) d(z)K\mu^K(z)$$

$$+ \sum_{z\in\mathcal{Z}} \sum_{p\in\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_z}{K} + \textstyle\sum_{w\in\mathcal{W}} \ell_w^{\text{prod}}(z,p)\tfrac{\delta_w}{K}\right) - \phi(\mu^K)\right) b(z,p)\mu^K(p)K\mu^K(z)$$

$$+ \sum_{w\in\mathcal{W}} \left(\phi\left(\mu^K - \tfrac{\delta_w}{K}\right) - \phi(\mu^K)\right) d(w)K\mu^K(w)$$

and for the extension to a non-finite trait space the infinitesimal generator is given by

$$\left(L^K\phi\right)\left(\mu^K\right) \tag{III.6.2}$$

$$= \int_{\mathcal{G}\times\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_{(g,p)}}{K}\right) - \phi(\mu^K)\right) \times \left(1 - u_K m(g)\right)\left\lfloor b(p) - \int_{\mathcal{P}} c_b(p,\tilde{p})\mu^K(\mathrm{d}\tilde{p})\right\rfloor_+ K\mu^K(\mathrm{d}(g,p))$$

$$+ \int_{\mathcal{G}\times\mathcal{P}} \int_{\mathcal{G}\times\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_{(\tilde{g},\tilde{p})}}{K}\right) - \phi(\mu^K)\right) u_K m(g) M((g,p),\mathrm{d}(\tilde{g},\tilde{p}))\left\lfloor b(p) - \int_{\mathcal{P}} c_b(p,p')\mu^K(\mathrm{d}p')\right\rfloor_+ K\mu^K(\mathrm{d}(g,p))$$

$$+ \int_{\mathcal{G}\times\mathcal{P}} \left(\phi\left(\mu^K - \tfrac{\delta_{(g,p)}}{K}\right) - \phi(\mu^K)\right)\left(d(p) + \int_{\mathcal{P}} c(p,\tilde{p})\mu^K(\mathrm{d}\tilde{p}) + \left\lfloor b(p) - \int_{\mathcal{P}} c_b(p,\tilde{p})\mu^K(\mathrm{d}\tilde{p})\right\rfloor_-\right)K\mu^K(\mathrm{d}(g,p))$$

$$+ \int_{\mathcal{G}\times\mathcal{P}} \int_{\mathcal{Z}} \left(\phi\left(\mu^K - \tfrac{\delta_{(g,p)}}{K} + \int_{\mathcal{W}} \tfrac{\delta_w}{K}\ell_{z,(g,p)}^{\text{kill}}(\mathrm{d}w)\right) - \phi(\mu^K)\right) t(z,p)\mu^K(\mathrm{d}z)K\mu^K(\mathrm{d}(g,p))$$

$$+ \int_{\mathcal{G}\times\mathcal{P}} \int_{\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_{(g,\tilde{p})}}{K} - \tfrac{\delta_{(g,p)}}{K}\right) - \phi(\mu^K)\right) S_{(g,p)}(\mathrm{d}\tilde{p})\,\hat{s}(g,p)K\mu^K(\mathrm{d}(g,p))$$

$$+ \int_{\mathcal{G}\times\mathcal{P}} \int_{\mathcal{W}} \int_{\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_{(g,\tilde{p})}}{K} - \tfrac{\delta_{(g,p)}}{K}\right) - \phi(\mu^K)\right) S_{w,(g,p)}(\mathrm{d}\tilde{p})\,\hat{s}((g,p),w)\mu^K(\mathrm{d}w)\,K\mu^K(\mathrm{d}(g,p))$$

$$+ \int_{\mathcal{Z}} \left(\phi\left(\mu^K + \tfrac{\delta_z}{K}\right) - \phi(\mu^K)\right) b(z)K\mu^K(\mathrm{d}z)$$

$$+ \int_{\mathcal{Z}} \left(\phi\left(\mu^K - \tfrac{\delta_z}{K}\right) - \phi(\mu^K)\right) d(z)K\mu^K(\mathrm{d}z)$$

$$+ \int_{\mathcal{Z}} \int_{\mathcal{P}} \left(\phi\left(\mu^K + \tfrac{\delta_z}{K} + \int_{\mathcal{W}} \tfrac{\delta_w}{K}\ell_{z,p}^{\text{prod}}(\mathrm{d}w)\right) - \phi(\mu^K)\right) b(z,p)\mu^K(\mathrm{d}p)K\mu^K(\mathrm{d}z)$$

$$+ \int_{\mathcal{W}} \left(\phi\left(\mu^K - \tfrac{\delta_w}{K}\right) - \phi(\mu^K)\right) d(w)K\mu^K(\mathrm{d}w).$$

III.6.2 Pseudo-code

Algorithm 2: Pseudo-code of the Gillespie algorithm used for generating the figures.

Data: Initial conditions: $\nu_0^K \in \mathcal{M}^K(\mathcal{X})$, Iterations: $N_{\text{Iterations}}$, Parameters of Section III.1.

$T_0 \leftarrow 0$, $\nu_{T_0}^K \leftarrow \nu_0^K$, $k \leftarrow 0$

while $k \le N_{\text{Iterations}}$ **do**

  for $x \in Supp(\nu_{T_k}^K)$ **do**

    if $x = (g,p) \in \mathcal{G} \times \mathcal{P}$ **then**

$$B(x) \leftarrow K\nu_{T_k}^K(g,p) \left\lfloor b(p) - \sum_{(\tilde{g},\tilde{p}) \in \text{Supp}(\nu_{T_k}^K)} c_b(p,\tilde{p})\nu_{T_k}^K(\tilde{g},\tilde{p}) \right\rfloor_+$$

$$D(x) \leftarrow K\nu_{T_k}^K(g,p) \left(d(p) + \left\lfloor b(p) - \sum_{(\tilde{g},\tilde{p}) \in \text{Supp}(\nu_{T_k}^K)} c_b(p,\tilde{p})\nu_{T_k}^K(\tilde{g},\tilde{p}) \right\rfloor_- \right.$$
$$\left. + \sum_{(\tilde{g},\tilde{p}) \in \text{Supp}(\nu_{T_k}^K)} c(p,\tilde{p})\nu_{T_k}^K(\tilde{g},\tilde{p}) \right),$$

$$T(x) \leftarrow K\nu_{T_k}^K(g,p) \sum_{z \in \text{Supp}(\nu_{T_k}^K)} t(z,p)\nu_{T_k}^K(z),$$

$$S(x) \leftarrow K\nu_{T_k}^K(g,p) \sum_{\tilde{p} \in \mathcal{P}} \left(s^g(p,\tilde{p}) + \sum_{w \in \text{Supp}(\nu_{T_k}^K)} s_w^g(p,\tilde{p})\nu_{T_k}^K(w) \right),$$

$$P(x) \leftarrow 0,$$

    if $x = z \in \mathcal{Z}$ **then**

$$B(x) \leftarrow K\nu_{T_k}^K(z)\,b(z), \quad D(x) \leftarrow K\nu_{T_k}^K(z)\,d(z), \quad T(x) \leftarrow 0, \quad S(x) \leftarrow 0$$
$$P(x) \leftarrow K\nu_{T_k}^K(z) \sum_{(g,p) \in \text{Supp}(\nu_{T_k}^K)} b(z,p)\nu_{T_k}^K(g,p),$$

    if $x = w \in \mathcal{W}$ **then**

$$B(x) \leftarrow 0, \quad D(x) \leftarrow K\nu_{T_k}^K(w)\,d(w), \quad T(x) \leftarrow 0, \quad S(x) \leftarrow 0, \quad P(x) \leftarrow 0,$$

    $\text{TotalTraitRate}(x) \leftarrow B(x) + D(x) + T(x) + P(x) + S(x)$

  $\text{TotalEventRate} \leftarrow \sum_{x \in \text{Supp}(\nu_{T_k}^K)} \text{TotalTraitRate}(x),$

  Sample $t \sim \mathcal{E}\text{xp}(\text{TotalEventRate}), \quad T_{k+1} \leftarrow T_k + t$

  Sample $x_{\text{chosen}} \in \mathcal{X}$ according to $\{\text{TotalTraitRate}(x), x \in \text{Supp}(\nu_{T_k}^K)\}$,

  Sample $E \in \{\text{Birth}, \text{Death}, \text{Therapy}, \text{Production}, \text{Switch}\}$ according to

  $\{B(x_{\text{chosen}}), D(x_{\text{chosen}}), T(x_{\text{chosen}}), P(x_{\text{chosen}}), S(x_{\text{chosen}})\},$

  case $E = \boldsymbol{Birth}$

    if $x_{chosen} = (g,p) \in \mathcal{G} \times \mathcal{P}$ **then**

      Sample $B \in \{\text{No Mutation}, \text{Mutation}\}$ according to $\{1 - u_K m(g), u_K m(g)\}$

      case $B = \boldsymbol{No\ Mutation}$

        $\nu_{T_{k+1}}^K \leftarrow \nu_{T_k}^K + \frac{1}{K}\delta_{x_{\text{chosen}}}$

      case $B = \boldsymbol{Mutation}$

        Sample $(\tilde{g}, \tilde{p})$ according to $M((g,p), (\tilde{g}, \tilde{p}))$

        $\nu_{T_{k+1}}^K \leftarrow \nu_{T_k}^K + \frac{1}{K}\delta_{(\tilde{g},\tilde{p})}$

    else

      $\nu_{T_{k+1}}^K \leftarrow \nu_{T_k}^K + \frac{1}{K}\delta_{x_{\text{chosen}}}$

  case $E = \boldsymbol{Death}$

    $\nu_{T_{k+1}}^K \leftarrow \nu_{T_k}^K - \frac{1}{K}\delta_{x_{\text{chosen}}}$

  case $E = \boldsymbol{Therapy}$ (Note that $x_{\text{chosen}} = (g,p)$ for some $(g,p) \in \mathcal{G} \times \mathcal{P}$ in this case.)

    Sample $z \in \mathcal{Z}$ according to $\{t(z,p)\nu_{T_k}^K(z), z \in \text{Supp}(\nu_{T_k}^K) \cap \mathcal{Z}\}$

    $\nu_{T_{k+1}}^K \leftarrow \nu_{T_k}^K - \frac{1}{K}\delta_{(g,p)} + \sum_{w \in \mathcal{W}} \ell_w^{\text{kill}}(z,p)\frac{1}{K}\delta_w$

  case $E = \boldsymbol{Production}$

    Sample $(g,p) \in \text{Supp}(\nu_{T_k}^K)$ according to $\{b(x_{\text{chosen}}, p)\nu_{T_k}^K(g,p), (g,p) \in \text{Supp}(\nu_{T_k}^K)\}$

    $\nu_{T_{k+1}}^K \leftarrow \nu_{T_k}^K + \frac{1}{K}\delta_{x_{\text{chosen}}} + \sum_{w \in \mathcal{W}} \ell_w^{\text{prod}}(x_{\text{chosen}}, p)\frac{1}{K}\delta_w$

  case $E = \boldsymbol{Switch}$ (Note that $x_{\text{chosen}} = (g,p)$ for some $(g,p) \in \mathcal{G} \times \mathcal{P}$ in this case.)

    Sample $\tilde{p} \in \mathcal{P}$ according to $\{s^g(p,\tilde{p}) + \sum_{w \in \text{Supp}(\nu_{T_k}^K)} s_w^g(p,\tilde{p})\nu_{T_k}^K(w), \tilde{p} \in \mathcal{P}\}$

    $\nu_{T_{k+1}}^K \leftarrow \nu_{T_k}^K - \frac{1}{K}\delta_{(g,p)} + \frac{1}{K}\delta_{(g,\tilde{p})}$

  $k = k + 1$

We use the following notations: let $\mathcal{D}$ be some discrete set and X a $\mathcal{D}$-valued random variable, then X sampled according to the weights $\{w(i), i \in \mathcal{D}\}$ means that $\mathbb{P}(X = i) = w(i)/\sum_{i \in \mathcal{D}} w(i)$.

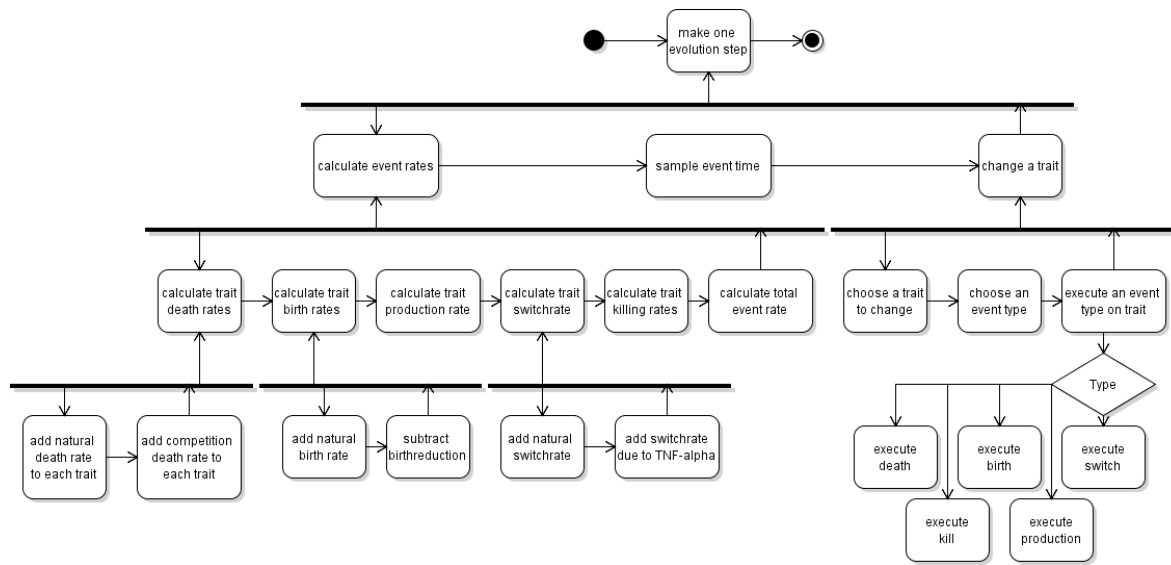The depth diagram of the algorithm we used to generate the simulations in this thesis is given in Figure III.13.

Figure III.13: Depth diagram of the program.

Bibliography

[1] P. A. Abrams, Y. Harada, and H. Matsuda. On the relationship between quantitative genetic and ESS models. *Evolution*, 47(3):982–985, 1993.

[2] D. Aldous and J. Fill. *Reversible Markov chains and random walks on graphs*. In progress. Manuscript available at https://www.stat.berkeley.edu/ aldous/RWG/book.pdf.

[3] T. Antal and P. Krapivsky. Exact solution of a two-type branching process: models of tumor progression. *J. Stat. Mech.*, 2011(08):P08018, 2011.

[4] T. Antal, P. Krapivsky, and M. Nowak. Spatial evolution of tumors with successive driver mutations. *Phys. Rev. E*, 92(2):022705, 2015.

[5] K. B. Athreya. Some results on multitype continuous time Markov branching processes. *Ann. Math. Stat.*, 39:347–357, 1968.

[6] K. B. Athreya and P. E. Ney. *Branching processes*. Die Grundlehren der mathematischen Wissenschaften, Vol. 196. Springer-Verlag Berlin Heidelberg, 1972.

[7] E. Baake. Aims and scope of the DGF SPP Priority Programme 1590, retrieved from www.dfg-spp1590.de, last up-date December 2015.

[8] M. Baar, A. Bovier, and N. Champagnat. From stochastic, individual-based models to the canonical equation of adaptive dynamics - in one step. *Preprint* (arXiv:1505.02421, 2015), *to appear in Ann. Appl. Probab.*, 2016.

[9] M. Baar, L. Coquille, H. Mayer, M. Hölzel, M. Rogava, T. Tüting, and A. Bovier. A stochastic model for immunotherapy of cancer. *Scientific Reports*, 6:24169, 2016.

[10] V. Bansaye and S. Méléard. *Stochastic models for structured populations. Scaling limits and long time behavior*, volume 1.4 of *Mathematical Biosciences Institute Lecture Series: Stochastics in Biological Systems*. Springer International Publishing, 2015.

[11] N. H. Barton. Sewall Wright on evolution in Mendelian populations and the "Shifting Balance". *Genetics*, 202(1):3–4, 2016.

[12] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 2nd edition, 1999.

[13] J. Blath, A. Etheridge, and M. Meredith. Coexistence in locally regulated competing populations and survival of branching annihilating random walk. *Ann. Appl. Probab.*, 17(5-6):1474–1507, 2007.

[14] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[15] B. Bolker and S. W. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.*, 52(3):179 – 197, 1997.

[16] B. M. Bolker and S. W. Pacala. Spatial moment equations for plant competition: understanding spatial strategies and the advantages of short dispersal. *Am. Nat.*, 153(6):575–602, 1999.

[17] A. Bovier and F. den Hollander. *Metastability. A potential-theoretic approach*, volume 351 of *Grundlehren der Mathematischen Wissenschaften*. Springer International Publishing, 2015.

[18] A. Bovier and R. Neukirch. Survival of the recessive allele in the mendelian diploid model. *Preprint*, arXiv: 1505.02109, 2015.

[19] A. Bovier and S.-D. Wang. Trait substitution trees on two time scales analysis. *Markov Proc. Rel.*, 19(4):607–642, 2013.

[20] I. Bozic, T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein, and M. A. Nowak. Accumulation of driver and passenger mutations during tumor progression. *Proc. Nat. Acad. Sci. USA*, 107(43):18545–18550, 2010.

[21] I. Bozic, J. G. Reiter, B. Allen, T. Antal, K. Chatterjee, P. Shah, Y. S. Moon, A. Yaqubie, N. Kelly, D. T. Le, E. J. Lipson, P. B. Chapman, L. A. Diaz Jr, B. Vogelstein, and M. A. Nowak. Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife*, 2:e00747, 2013.

[22] M. G. Bulmer. *The mathematical theory of quantitative genetics.* The Clarendon Press, Oxford University Press, New York, 1980.

[23] R. Bürger. Some mathematical models in evolutionary genetics. In *The mathematics of Darwin's legacy*, Mathematics and Biosciences in Interaction, pages 67–89. Birkhäuser Basel/Springer Basel AG, 2011.

[24] N. Champagnat. *Etude mathématique de modèles stochastiques d'évolution issus de la théorie écologique des dynamiques adaptives.* PhD thesis, Université Paris X, Paris (France), 2004.

[25] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.*, 116(8):1127–1160, 2006.

[26] N. Champagnat. *Approches stochastiques et déterministes en biologie: dynamique adaptative, modélisation pour l'écologie, génétique des populations et dynamique moléculaire; caractère bien posé d'équations différentielles ordinaires et stochastiques.* Habilitation thesis, Université de Lorraine, 2015.

[27] N. Champagnat, R. Ferrière, and G. Ben Arous. The canonical equation of adaptive dynamics: a mathematical view. *Selection*, 2:73–83, 2001.

[28] N. Champagnat, R. Ferrière, and S. Méléard. From individual stochastic processes to macroscopic models in adaptive evolution. *Stoch. Models*, 24:2–44, 2008.

[29] N. Champagnat, P.-E. Jabin, and S. Méléard. Adaptation in a stochastic multi-resources chemostat model. *J. Math. Pures Appl.*, 101(6):755–788, 2014.

[30] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Prob. Theory Rel.*, 151(1-2):45–94, 2011.

[31] P. Collet, S. Méléard, and J. A. J. Metz. A rigorous model study of the adaptive dynamics of Mendelian diploids. *J. Math. Biol.*, 67(3):569–607, 2013.

[32] A. L. Correia and M. J. Bissell. The tumor microenvironment is a dominant force in multidrug resistance. *Drug Resist. Update*, 15(1):39–49, 2012.

[33] M. Costa, C. Hauzy, N. Loeuille, and S. Méléard. Stochastic eco-evolutionary model of a prey-predator community. *J. Math. Biol.*, 72(3):573–622, 2016.

[34] J. Couzin-Frankel. Cancer immunotherapy. *Science*, 342(6165):1432–1433, 2013.

[35] J. F. Crow and M. Kimura. *An introduction to population genetics theory.* Harper & Row, New York, 1970.

[36] C. Darwin. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* John Murray, London, 1859.

[37] C. Darwin and A. Wallace. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zool. J. Linn. Soc.-Lond.*, 3:45–62, 1858.

[38] D. A. Dawson. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Math.*, pages 1–260. Springer, Berlin, 1993.

[39] D. A. Dawson and A. Greven. Multiscale analysis: Fisher-Wright diffusions with rare mutations and selection, logistic branching system. In *Probability in complex physical systems*, volume 11 of *Springer Proc. Math.*, pages 373–408. Springer, Heidelberg, 2012.

[40] F. Dercole and S. Rinaldi. *Analysis of evolutionary processes. The adaptive dynamics approach and its applications.* Princeton Series in Theoretical and Computational Biology. Princeton University Press, Princeton, NJ, 2008.

[41] L. A. Diaz Jr, R. T. Williams, J. Wu, I. Kinde, J. R. Hecht, J. Berlin, B. Allen, I. Bozic, J. G. Reiter, M. A. Nowak, K. W. Kinzler, K. S. Oliner, and B. Vogelstein. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*, 486(7404):537–540, 06 2012.

[42] U. Dieckmann and R. Law. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.*, 34(5-6):579–612, 1996.

[43] U. Dieckmann and R. Law. Moment approximations of individual-based models. In U. Dieckmann, R. Law, and J. A. J. Metz, editors, *The geometry of ecological interactions: simplifying spatial complexity*, pages 252–270. Cambridge University Press, 2000.

[44] O. Diekmann, P.-E. Jabin, S. Mischler, and B. Perthame. The dynamics of adaptation: an illuminating example and a Hamilton–Jacobi approach. *Theor. Popul. Biol.*, 67(4):257–271, 2005.

[45] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, 1997.

[46] M. Durinx, J. A. J. Metz, and G. Meszéna. Adaptive dynamics for physiologically structured population models. *J. Math. Biol.*, 56(5):673–742, 2008.

[47] R. Durrett. *Probability models for DNA sequence evolution.* Springer-Verlag, New York, 2nd edition, 2008.

[48] R. Durrett. Cancer modeling: a personal perspective. *Not. Am. Math. Soc.*, 60(3):304–309, 2013.

[49] R. Durrett. *Branching Process Models of Cancer*, volume 1.1 of *Stochastics in Biological Systems.* Springer International Publishing, 2015.

[50] R. Eftimie, J. Bramson, and D. Earn. Interactions between the immune system and cancer: A brief review of non-spatial mathematical models. *B. Math. Biol.*, 73(1):2–32, 2011.

[51] E. Elert. Calling cells to arms. *Nature*, 504(7480):S2–S3, 2013.

[52] A. Etheridge. *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics.* Springer, Heidelberg, 2011.

[53] A. M. Etheridge. *An introduction to superprocesses*, volume 20 of *University Lecture Series.* American Mathematical Society, Providence, RI, 2000.

[54] A. M. Etheridge. Survival and extinction in a locally regulated population. *Ann. Appl. Probab.*, 14(1):188–214, 2004.

[55] S. N. Ethier and T. G. Kurtz. *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[56] W. J. Ewens. *Mathematical population genetics. I. Theoretical introduction*, volume 27 of *Interdisciplinary Applied Mathematics.* Springer-Verlag, New York, 2nd edition, 2004.

[57] D. S. Falconer and T. F. Mackay. *Introduction to quantitative genetics.* Longman Scientific and Technical, Harlow, England, 3rd edition, 1989.

[58] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *T. Roy. Soc. Edin. Earth*, 52(02):399–433, 1918.

[59] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004.

[60] S. A. Frank and M. R. Rosner. Nonheritable cellular variability accelerates the evolutionary processes of cancer. *PLoS Biol.*, 10(4):e1001296, 2012.

[61] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften.* Springer, Heidelberg, 3rd edition, 2012.

[62] S. A. Geritz, G. Mesze, J. A. Metz, et al. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.*, 12(1):35–57, 1998.

[63] S. A. Geritz, J. A. Metz, É. Kisdi, and G. Meszéna. Dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.*, 78(10):2024, 1997.

[64] S. A. H. Geritz. Resident-invader dynamics and the coexistence of similar strategies. *J. Math. Biol.*, 50(1):67–82, 2005.

[65] R. J. Gillies, D. Verduzco, and R. A. Gatenby. Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat. Rev. Cancer*, 12(7):487–493, 2012.

[66] M. Greaves and C. C. Maley. Clonal evolution in cancer. *Nature*, 481:306–313, 2012.

[67] A. Greven, A. Sturm, A. Winter, and I. Zähle. Multi-type spatial branching models for local self-regulation I: Construction and an exponential duality. *Preprint*, arXiv:1509.04023, 2015.

[68] J. Guckenheimer and P. Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

[69] P. B. Gupta, C. M. Fillmore, G. Jiang, S. D. Shapira, K. Tao, C. Kuperwasser, and E. S. Lander. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4):633–644, 2011.

[70] J. B. S. Haldane. A mathematical theory of natural and artificial selection. A series of ten scientific papers published between 1924 and 1934. *Trans. Cambridge Philos. Soc., Proc. Cambridge Philos. Soc. and Genetics*, 1924–1934.

[71] J. B. S. Haldane. *The Causes of Evolution*. Longmans, Green & Co., New York, 1932.

[72] D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.

[73] M. W. Hirsch. Systems of differential equations which are competitive or cooperative. III. Competing species. *Nonlinearity*, 1(1):51–71, 1988.

[74] J. Hofbauer and K. Sigmund. Adaptive dynamics and evolutionary stability. *Appl. Math. Lett.*, 3(4):75–79, 1990.

[75] J. Hofbauer and J. W.-H. So. Multiple limit cycles for three-dimensional Lotka-Volterra equations. *Appl. Math. Lett.*, 7(6):65–70, 1994.

[76] M. Hölzel, A. Bovier, and T. Tüting. Plasticity of tumour and immune cells: a source of heterogeneity and a cause for therapy resistance? *Nat. Rev. Cancer*, 13(5):365–376, 2013.

[77] W. Hugo, H. Shi, L. Sun, M. Piva, C. Song, X. Kong, G. Moriceau, A. Hong, K. B. Dahlman, D. B. Johnson, J. A. Sosman, A. Ribas, and R. S. Lo. Non-genomic and immune evolution of melanoma acquiring MAPKi resistance. *Cell*, 162(6):1271 – 1285, 2015.

[78] M. Hutzenthaler and A. Wakolbinger. Ergodic behavior of locally regulated branching populations. *Ann. Appl. Probab.*, 17(2):474–501, 2007.

[79] J. Istas. *Mathematical modeling for the life sciences*. Universitext. Springer-Verlag, Berlin, 2005.

[80] Y. Iwasa, A. Pomiankowski, and S. Nee. The evolution of costly mate preferences II. The "handicap" principle. *Evolution*, 45(6):1431–1442, 1991.

[81] P. Jagers and A. N. Lageras. General branching processes conditioned on extinction are still branching processes. *Electron. Commun. Prob.*, 13:540–547, 2008.

[82] H. C. F. Jenkin. Review of "The origin of species". *North British Review*, 46:277–318, 1867.

[83] P. Keller and T. Antal. Mutant number distribution in an exponentially growing population. *J. Stat. Mech. Theory Exp.*, (1):P01011, 29, 2015.

[84] H. Kesten and B. P. Stigum. Additional limit theorems for indecomposable multidimensional Galton-Watson processes. *Ann. Math. Stat.*, 37:1463–1481, 1966.

[85] H. Kesten and B. P. Stigum. A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Stat.*, 37:1211–1223, 1966.

[86] H. Kesten and B. P. Stigum. Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.*, 17:309–338, 1967.

[87] D. J. Konieczkowski, C. Johannessen, O. Abudayyeh, J. Kim, Z. Cooper, A. Piris, D. Frederick, M. Barzily-Rokni, R. Straussman, R. Haq, D. Fisher, J. Mesirov, W. Hahn, K. Flaherty, J. Wargo, P. Tamayo, and L. Garraway. A melanoma cell state distinction influences sensitivity to MAPK pathway inhibitors. *Cancer Discov.*, 4:816–827, 2014.

[88] V. Kuznetsov, I. Makalkin, M. Taylor, and A. Perelson. Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis. *B. Math. Biol.*, 56(2):295–321, 1994.

[89] A. Lambert. The branching process with logistic growth. *Ann. Appl. Probab.*, 15(2):1506–1535, 2005.

[90] R. Lande. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30(2):314–334, 1976.

[91] R. Lande. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*, 33(1):402–416, 1979.

[92] J. Landsberg, J. Kohlmeyer, M. Renn, T. Bald, M. Rogava, M. Cron, M. Fatho, V. Lennerz, T. Wölfel, M. Hölzel, and T. Tüting. Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. *Nature*, 490(7420):412–416, 10 2012.

[93] K. Leder, E. C. Holland, and F. Michor. The therapeutic implications of plasticity of the cancer stem cell phenotype. *PloS ONE*, 5(12):e14366, 2010.

[94] Z. Li. *Measure-valued branching Markov processes*. Probability and Its Applications. Springer-Verlag Berlin Heidelberg, 2011.

[95] P. Marrow, R. Law, and C. Cannings. The coevolution of predator–prey interactions: ESSs and Red Queen dynamics. *P. Roy. Soc. Lond. B Bio.*, 250(1328):133–141, 1992.

[96] A. Marusyk, V. Almendro, and K. Polyak. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, 12(5):323–334, 2012.

[97] H. Mayer. *Contributions to stochastic modelling of the immune system.* PhD thesis, Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn (Germany), 2016.

[98] J. Maynard Smith. *Evolution and the theory of games.* Cambridge University Press, 1982.

[99] J. Maynard Smith and G. R. Price. The logic of animal conflict. *Nature*, 246(5427):15–18, 1973.

[100] G. Mendel. Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, Bd. IV:3–47, 1866.

[101] G. Meszéna, M. Gyllenberg, F. J. Jacobs, and J. A. Metz. Link between population dynamics and dynamics of Darwinian evolution. *Phys. Rev. Lett.*, 95(7):078105, 2005.

[102] J. Metz. Fitness. In S. E. Jørgensen and B. D. Fath, editors, *Encyclopedia of Ecology*, pages 1599 – 1612. Academic Press, Oxford, 2008.

[103] J. Metz, R. Nisbet, and S. Geritz. How should we define 'fitness' for general ecological scenarios? *Trends Ecol. Evol.*, 7(6):198 – 202, 1992.

[104] J. A. J. Metz, S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs, and J. S. van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. In *Stochastic and spatial structures of dynamical systems (Amsterdam, 1995)*, Konink. Nederl. Akad. Wetensch. Verh. Afd. Natuurk. Eerste Reeks, 45, pages 183–231. North-Holland, Amsterdam, 1996.

[105] J. Müller, O. Krijgsman, J. Tsoi, L. Robert, W. Hugo, C. Song, X. Kong, P. A. Possik, P. D. Cornelissen-Steijger, M. H. G. Foppen, K. Kemper, C. R. Goding, U. McDermott, C. Blank, J. Haanen, T. G. Graeber, A. Ribas, R. S. Lo, and D. S. Peeper. Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat. Commun.*, 5(5712), 2014.

[106] J. Nash. Non-cooperative games. *Ann. Math.*, 54:286–295, 1951.

[107] J. F. Nash. Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. USA*, 36:48–49, 1950.

[108] P. C. Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28, 1976.

[109] S. Pénisson. *Conditional limit theorems for multitype branching processes and illustration in epidemiological risk analysis.* PhD thesis, Universität Potsdam, Potsdam (Germany), 2010.

[110] D. Pollard. *Convergence of stochastic processe.* Springer Series in Statistics. Springer-Verlag, New York, 1984.

[111] S. A. Rosenberg and N. P. Restifo. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science*, 348(6230):62–68, 2015.

[112] B. A. Sewastjanow. *Verzweigungsprozesse*. R. Oldenbourg Verlag, Munich-Vienna, 1975.

[113] S. Smale. On the differential equations of species in competition. *J. Math. Biol.*, 3(1):5–7, 1976.

[114] T. L. Vincent, Y. Cohen, and J. S. Brown. Evolution via strategy dynamics. *Theor. Popul. Biol.*, 44(2):149–176, 1993.

[115] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey, 1944.

[116] D. Waxman and S. Gavrilets. 20 questions on adaptive dynamics. *J. Evolution. Biol.*, 18(5):1139–1154, 2005.

[117] S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97–159, 1931.

[118] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 6th Int. Congr. Genet.*, 1:356–366, 1932.

[119] D. Xiao and W. Li. Limit cycles for the competitive three dimensional Lotka-Volterra system. *J. Differ. Equations*, 164(1):1–15, 2000.

[120] M. L. Zeeman. Hopf bifurcations in competitive three-dimensional Lotka-Volterra systems. *Dynam. Stabil. Syst.*, 8(3):189–217, 1993.