Stochastic and deterministic models for the evolution of heterogeneous populations: Multiscale approximation and applications to melanoma T-cell therapy

Dissertation

zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Anna Katharina Kraut

aus
Braunschweig

Bonn, April 2020

Angefertigt mit Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Anton Bovier

2. Gutachter: Prof. Dr. Jochen Blath

Tag der Promotion: 03. Juli 2020

Erscheinungsjahr: 2020

Summary

Stochastic modelling at the interface of mathematics and life sciences has gained great attention over the last decades. The study of many complex biological systems requires models taking random effects into account. Both sides benefit from this interdisciplinary collaboration. A structured mathematical analysis can provides a new perspective and helps to gain insights into biological problems. Vice versa, biological research inspires new mathematical questions leading to an interesting theory on their own. In this thesis, we demonstrate how mathematical modelling supports biomedical research in various ways: First, important mechanisms are identified that determine the outcome of experiments. Second, likely causes for the observed phenomena are investigated, which helps to interpret experimental data. Third, the clinical applicability of experimental scenarios is validated. Forth, predictions are made that reach beyond the experiment. Conversely, we study mathematical questions arising from biology. We approximate stochastic and deterministic models for adaptive dynamics under various parameter regimes to investigate the long-term behaviour of a population. Some of these results are again beneficial for applications in biomedicine since they have potential to improve the algorithms for simulations of the studied systems.

The thesis is divided into two more theoretical parts and one more applied part. In the theoretical parts we study individual-based Markov processes and their deterministic counterpart as models for the evolution of a heterogeneous population. We consider the limit of large populations and rare mutations. The resulting limit processes show different behaviour and are highly dependent and the scaling of mutation rates and the choice of time scales. The short-term dynamics are governed by Lotka-Volterra interactions of large subpopulations and the invasion of arising mutations can only be witnessed on a divergent time scale. In Chapter 2, we analyse the deterministic system that arises from the stochastic model in a law of large numbers. We study the limit of rare mutations. This corresponds to a scenario of relatively high mutation rates, compared to other limit regimes. It leads to multiple microscopic mutant populations that compete to invade the resident population at the same time. To determine which of the mutant traits succeeds, one has to carefully keep track of the growth of all subpopulations. The general discrete graph that we consider as a trait space induces complex dynamics of mutations between traits. To handle these, we have to introduce a new approach of inductive approximation of the population sizes of different traits, taking into account the influence of different traits at an increasing distance. Moreover, we investigate a couple of interesting special cases that relate to the scenario of adaptive walks and propose a cut-off model that mimics the simultaneous limit of large populations and rare mutations in the stochastic model. In Chapter 3, we combine the mentioned inductive procedure and couplings to branching processes to consider this simultaneous limit. To do so, we have to extend some existing limit results for branching processes to the multidimensional case. We derive a complete characterisation of the limiting

jump process in the scenario of power law mutation rates, thus extending previous results for linear trait spaces and specific parameters to general finite graphs and arbitrary fitness landscapes. In the second part of the chapter we present a collection of specific examples that represent interesting and partially counter-intuitive behaviour arising under this scaling. Chapter 4 is dedicated to an application of individual-based models in the field of oncology. We investigate the role of phenotypic and genotypic heterogeneity of melanoma cells in the development of resistance to immunotherapy. Here, we substantially extend the existing model of a previous collaboration to include effects of immunosuppression, aspects of the spatial structure of the tumour, and the possibility of spontaneous mutations. While the previous model was designed to investigate phenotypic switches, we focus on the study of genetic variants. Through simulations we analyse the effect of subclonal fitness variability on the enrichment of resistant cell types.

Acknowledgements

I would like to take this opportunity to thank the people and institutions that have supported me during the time of my PhD studies.

First and foremost, I would like to express my gratitude to my advisor Anton Bovier. Without him, I would not be in the position that I am in today. He was the one who encouraged me to pursue an academic career in the first place. It is thanks to him that I have discovered my joy in doing mathematical research and in particular in collaborating with researchers from different fields. During my time as a member of his group, he has always taken the time to share his great intuition and knowledge and to bear with me when I was hung up on technical details. I want to thank him for always giving me honest feedback, even when it was uncomfortable, and for placing his trust in me to autonomously present our results to other researchers or the public at many occasions.

I would like to appreciate all my coauthors. I want to thank Michael Hölzel and Nicole Glodde for always taking the time to discuss and answer my many biological questions, commuting between the Mathematical Institute and the Venusberg. My gratitude also belongs to Martin Rumpf and his student Kai Echelmeyer who have helped us to develop the algorithm for our simulations. I am grateful to Loren Coquille and Charline Smadi for many productive but also fun discussions on their visits to Bonn and I particularly enjoyed my week in Grenoble with them.

I want to give my appreciation to all present and former members of the probability theory group and my fellow PhD students from other groups that have made my time so enjoyable. I have to highlight Mei-Ling for being the heart and soul of the group and always taking care of us, trying to make our life as easy as possible. I will keep fond memories of my office mates Kaveh, Martina, Florian, and Luis for manoeuvring PhD life together and for many interesting exchanges on more or less scientific questions.

I would like to thank the Deutsche Forschungsgemeinschaft (DFG) for the funding under Germany's Excellence Strategy GZ 2047/1, Projekt-ID 390685813 and GZ 2151, Project-ID 390873048 and through the Priority Programme 1590 "Probabilistic Structures in Evolution". It was a great opportunity to be part of many networks that allowed me to travel and connect with fellow researches during numerous meetings, conferences, and summer schools.

I am extremely grateful to Henrik and my parents for always encouraging me and dealing with me during periods of self-doubt. They have always reassured me, even from far away, and let me know that they would support me no matter what.

I want to thank Felix and Nicole for helping me by proof reading parts of this thesis.

Finally, I want to thank the members of my committee, Anton Bovier, Jochen Blath, Martin Rumpf, and Michael Hölzel, for taking the time to review my work.

Contents

1 Introduction

Over the last years, mathematical modelling has become an important tool in analysing complex biological systems. At the same time, biological questions have given rise to many interesting mathematical problems. One prominent example is the study of evolutionary dynamics. The goal is to understand how populations adapt to their environment by a succession of changes in their traits. Both the mechanisms of how these changes arise and how certain traits are selected to be enriched in the overall population are objects of research. Similar questions arise in the context of tissue growth. In particular, they are of interest in the study of tumour genesis and evolution under the selective pressure of therapy. Here, evolutionary dynamics induce an abnormal tissue growth and can cause treatment failure. In this thesis we study models for the evolution of heterogeneous populations. We consider both aspects: On one hand, we focus on the theoretical analysis of mathematical models for evolution. On the other hand, we develop models for applications in biomedicine in the field of oncology.

There are many approaches towards mathematical modelling in biology, using both stochastic and deterministic systems. Deterministic models range from systems of ordinary differential equations to partial differential equations that take into account spatial movement. Stochastic models include diffusion processes and Markov jump processes in discrete or continuous time. Different models show the biological system at varying resolutions and focus on specific aspects of the dynamics. The choice of the correct model therefore depends on the underlying question one tries to answer. We want to study the effects that spontaneous mutations towards new traits and the interactions between individuals have on the long-term evolution of a population. Therefore, we choose a model class that captures the behaviour of single individuals and takes random effects into account. As a starting point, we consider an individual-based Markov process that has been introduced in the context of adaptive dynamics. It is quite detailed in the sense that it accounts for every individual and every birth or death event in the population.

In particular in the context of biomedical applications, it is of interest to generate realisations of this Markov process by simulations. The drawback of a stochastic individual-based model is that it is computationally heavy to simulate. The generation of singular birth and death events requires many iterations of the algorithm. Particularly in populations with many individuals, like a tumour consisting of many cells, computing times are high. An approximation of the underlying stochastic system by simpler models, that are easier to simulate, provides the opportunity to speed up simulations. Simpler models can for example be deterministic systems of differential equations or, more or less stochastic, jump processes that only show the macroscopic adaptation of the population as a succession of resident traits. We study which of these models can arise as scaling limits of individual-based Markov processes.

An important factor in determining the limit process are the various scales that are involved. As scaling parameters we consider the order of the population size at equilibrium, K, and the probability of mutation, μ_K. We study the regime of large populations ($K \to \infty$) and rare mutations ($\mu_K \to 0$). To analyse the frequencies of different subpopulations, we divide the population size by K. As a consequence, only significantly represented traits are visible in the limit and single individuals are of infinitesimal size. Depending on how quickly μ_K decays, mutations are more or less frequent. This results in mutations either to be separated in comparison to the faster ecological dynamics or to overlap and compete for invading the population. Corresponding to the different mechanisms of mutations, growth of smaller subpopulations, and the competitive interaction of larger subpopulations, the system can be considered on various time scales. Different aspects of the model are emphasised and visible in the limit, depending on the choice of the mutation probability and time scale. We thus obtain a variety of possible outcomes when approximating the stochastic individual-based model.

The easiest limit to consider is the one of $K \to \infty$ only, leaving the mutation probability constant at $\mu > 0$. This implies frequent mutations and gives rise to a deterministic system of differential equations. These equations take the form of a competitive Lotka-Volterra system with additional mutation terms. If one lets μ tend to zero in this system, it converges to a pure Lotka-Volterra system with no mutations. Only when considering another time scale that diverges as $\mu \to 0$, mutations are visible again in the macroscopic population. This scenario is considered in Chapter 2. We prove convergence to a deterministic jump process that describes the macroscopic evolution of the population in terms of a sequence of (possibly coexisting) traits at equilibrium.

The same process can be recovered in the simultaneous limit of large populations and rare mutations if μ_K decays at a certain speed. In Chapter 3 we prove this as part of a more general result, considering moderately rare mutation rates that decay as arbitrary negative powers of K. The processes that arise in this limit can be seen as interpolations between the process of Chapter 2 and the stochastic jump processes that have been derived as limits under very rare mutation rates by Champagnat, Méléard et al. They display interesting and partially unexpected behaviour, which we document by a number of specific examples.

As mentioned above, these limit approximations can be applied to improve simulation algorithms for stochastic individual-based models. We make use of this in an application in biomedical research. In cooperation with experimentalists, we use an extension of classical individual-based models to study the evolution of melanoma cell populations under immunotherapy with cytotoxic T-cells. In Chapter 4 we prove that the large population approximation result applies to this extended model. Therefore, it is reasonable to use a hybrid algorithm that combines deterministic simulations of frequent events and stochastic simulations of rare events. In simulations we compare phenotypic switches and genetic mutations as tumour escape mechanisms that cause therapy failure. We argue for a substantial heterogeneity among genetic variants that explains widespread scattering of the experimental data.

The remainder of the introduction is organised as follows: In Section 1.1 we give a brief summary of the history of the theory of evolution and of different mathematical approaches towards modelling of evolution. Section 1.2 introduces the two main models that are studied

theoretically in Chapter 2 and 3 and that also form the basis for the model in Chapter 4. We further discuss notions of fitness. In Section 1.3 we give an extensive overview of the different scales that are involved with these models and present various regimes for approximation. A biological background for cancer, its treatment, and in particular the immune system and immunotherapeutic approaches is given in Section 1.4. Moreover, we explain the benefits of mathematical modelling in life sciences and present a number of different strategies for modelling, particularly in the context of cancer. In Section 1.5 we briefly discuss open questions and future perspectives of research in this field. Finally, the main results and an outline of the thesis are presented in Section 1.6.

Note that Chapter 2 and 3 are self-contained. Depending on the biomedical knowledge of the reader, it is advised to read Section 1.4 before Chapter 4.

1.1 Mathematical models for evolution

In this section we give a brief overview of the theory of evolution and of different approaches to mathematically model the evolution of a population. This is certainly not exhaustive but sets the work in this thesis into perspective. Many of the mentioned models have been studied within the DFG Priority Programm "Probabilistic Structures in Evolution". There are various review articles of members of this programm that we refer to for further reading, for example [7, 18, 23, 105, 113]. A more detailed introduction to the models that are studied in this thesis is given in Section 1.2. Note that the introduction of [22] served as a basis for parts of this section.

1.1.1 Theory of evolution

The theory of evolution, as it is studied today, goes back to the seminal work of Charles Darwin in the 1850s. Some aspects of his theory were previously mentioned in other works, e.g. by Erasmus Darwin (his grandfather) [53, 54] and Thomas Malthus [125]. Moreover, similar theories were developed simultaneously by Alfred Wallace [52]. The term *evolution* describes the observation that populations change and adapt to their environment, i.e. the available resources and also the state of other populations, over time. This change occurs over the course of generations and is tied to the birth of new individuals, taking the place of old individuals that die. In his book *On the origins of species* [51], Darwin identifies three driving mechanisms of evolution:

> *Heredity*: at reproduction, individuals pass their traits (properties) on to their offspring,
>
> *Variation*: traits vary between individuals and heredity is not perfect, i.e. sometimes the offspring's traits vary from those of its ancestor due to *mutation*,
>
> *(Natural) Selection*: different traits display different fitness, i.e. reproductive success and survival rate.

The trait of an individual consists of two components, the genotype and the phenotype. The genotype of an individual is defined by its genetic code, the DNA, which was discovered by Francis Crick and James Watson [173], Maurice Wilkins [175], and Rosalind Franklin [78] in the 1950s. The DNA is contained in every cell of an organism and remains unchanged during its lifetime (apart from errors during the copying process). At reproduction, it gets passed on to the next generation and is thus the basis of heredity. The phenotype of an individual describes its outer appearance, i.e. its morphology and physical properties. It is determined by a combination of the individuals genotype and environmental influences, and can vary over a lifespan.

Apart from these phenotypic changes, there are two major causes of variation. At reproduction, there can be errors due to faulty copying of the DNA molecule, leading to genetic mutations. This is possible in any organism and for any form of reproduction. Additionally, in diploid organisms that obtain their genome as a combination of the genetic information of two ancestors, variation can arise due to recombination. This process of inheritance has been studied by Gregor Mendel in the 1860s [136, 137]. One gene, the part of the genome that determines a particular feature, can have several specificities, called alleles. In a diploid organism either one allele (the dominant allele) governs over the other (recessive) allele and fully determines the characteristics of a gene, or the feature is determined by a combination of the alleles (co-dominance). The procedure of recombination is summarised in the formalism of *Mendel's law of inheritance*.

According to Darwin's *survival of the fittest*, natural selection arises as a consequence of interactions with the environment and between individuals. These interactions can be competitive, due to limited resources, but also in dependence, as in predator-prey or parasite-host relations, or mutually beneficial in a symbiosis. Advantageous variations of a trait show an increase in reproductive success and survival, and are thus enriched in the population. Multiple variations can accumulate in one organism and eventually lead to speciation, i.e. the splitting into two separate species.

This theory of Darwinian evolution, based on random changes and selection, stands in contrast to the theory of Lamarck, where individuals modify their traits to adapt to their environment and these changes are inherited by the next generation [55]. Although Darwins's theory is widely accepted, recent discoveries in the field of epigenetics have shown that there is in fact the possibility to pass on adaptive changes that have been acquired during a lifetime [122]. Moreover, mechanisms like horizontal gene transfer, that can be witnessed in bacteria but also eukaryotes, allow for an active exchange of beneficial genes [149, 104].

A more extensive account of the history of the theory of evolution can be found in [6]. There are different approaches for mathematical modelling that focus on various aspects of the theory of evolution. In the following, we present a selection of them, including both deterministic and stochastic models.

1.1.2 Population dynamics

The theory of *population dynamics* focuses on the ecology, i.e. on the interaction of populations. As Malthus proposed, the exponential growth that populations would show in an interaction-free scenario is in fact restrained by limited resources, e.g. food or space [125]. Mathematically, this results in logistic growth dynamics. For a monomorphic population, i.e. where all individuals carry the same trait, this growth is described by the deterministic differential equation

$$\dot{n}(t) = rn(t) - cn(t)^2, \quad t \geq 0. \tag{1.1}$$

Here, $n(t) \geq 0$ denotes the population size at time t, r is the exponential growth rate of the population in an unrestrained environment, and $c > 0$ parametrises the competitive interaction within the population. As long as r and $n(0)$ are positive, the population size converges to its stable equilibrium r/c. If $r \leq 0$, the population size tends to zero.

In the more general multitype case, the evolution of a population with traits in the finite set $\mathcal{X}$ is given by

$$\dot{n}_x(t) = n_x(t) \left(r_x - \sum_{y \in \mathcal{X}} c_{x,y} n_y(t) \right), \quad x \in \mathcal{X}, t \geq 0. \tag{1.2}$$

This type of equation goes back to the work of Alfred Lotka [123] and Vito Volterra [170] and is hence called *(competitive) Lotka-Volterra equation*. The study of the long-term behaviour and equilibria of these competitive Lotka-Volterra systems is of great interest, as they determine the short-term dynamics of many models that include mutations. More details on this model, as well as a stochastic version in the form of logistic birth-and-death processes, are given in Section 1.2. A more extensive discussion and a modern perspective on population dynamics can be found in an essay by Hofbauer and Sigmund [96].

1.1.3 Population genetics

In contrast to population dynamics, the concept of *population genetics* is more focussed on heredity. In both theoretical and experimental approaches, the change of gene frequencies over time is studied. Instead of interactions within the population, the emphasis lies on genealogies and the transmission of genes from one generation to the next. This goes back to Mendel and his study of genes and alleles, as mentioned above.

Mathematically, the concept was introduced by Fisher [73], Haldane [90] and Wright [177] around the 1920s. The prototype of a population genetics model is the Wright-Fisher model. It considers discrete generations of N individuals. In each generation, every offspring individual selects one ancestor at random from the previous generation and adopts its trait. The probability that one individual inherits a certain trait is hence exactly this trait's frequency in the preceding generation. Since no trait is preferred, this is a neutral model and the resulting process, which describes the time evolution of the frequencies, is a martingale. This is in line with the popular Hardy-Weinberg Theorem stating that, in an infinite population with no selection, the frequencies of different genes remain constant [92, 174]. In finite

populations, the frequencies only change due to stochastic fluctuations, which is referred to as *genetic drift*. In a law of large numbers, rescaling the time as tN and letting $N \to \infty$, the Wright-Fisher model converges to the Wright-Fisher diffusion. This stochastic diffusion process was first introduced by Kimura [107] and derived as a limit by Ethier and Norman [71].

There are many variants of the Wright-Fisher model. The Moran model is a time continuous version, where generations are no longer discrete but each individual replaces another individual with its own trait at an exponential rate [141]. In the Cannings model the offspring traits of discrete generations are chosen in a more general way. The only condition is that the number of offspring per parent is exchangable, keeping the mean frequencies of traits constant. Depending on the reproduction function, the Cannings model can mimic the Wright-Fisher and the Moran model [28, 29]. Fleming-Viot processes are generalisations of the Wright-Fisher diffusion and arise as infinite population limits of Moran type models [75]. Other variants introduce selection, mutation or migration, and a recombination mechanism to such models [69, 8, 7]. Note that in this form of selection individuals do not interact. Instead, certain traits are chosen as ancestors with higher probability or individuals are replaced at a higher rate. This leads to an enrichment of traits with a selective advantage within the population. However, since there is no interaction, the population adapts in a fixed fitness landscape towards the trait of highest selective advantage. Thus, the models do not depict phenomena like coexistence and speciation due to evolutionary branching. In a deterministic setting, the spatial spread of an advantageous allele is modelled by the famous F-KKP equation due to Fisher[74] and Kolmogorov, Petrovsky, and Piscounov [111].

In population genetics, not only the forward in time evolution of gene frequencies is analysed but a big focus also lies on the study of the genealogical structure and ancestry. The genealogical tree of a sample of individuals is traced back in time, for example to find the most recent common ancestor. Providing information on when genetic leneages were seperated and where mutations occured, this genealogical tree can help to interpret genetic data. The first example for a mathematical analysis is Kingman's coalescent that describes the ancestry of the Wright-Fisher diffusion [108]. More general, ancestries are described by so-called ancestral selection or recombination graphs [145, 121]. They can be understood as coalescent processes and satisfy a mathematical duality to the forward process [8, 7]. For an overview of modern population genetics, we refer to the works of Ewens [72] and Etheridge [69].

1.1.4 Adaptive dynamics

Adaptive dynamics is a biological theory that was developed to study the interplay of both ecology and evolution. To this aim, it combines a system of interacting populations, as in population dynamics, with the study of hereditary mechanisms and the possibility of mutations, as in population genetics. The ideas go back to works of Hofbauer and Sigmund [95] on evolutionary game theory, Metz et al. [140] on fitness measures in ecological scenarios, and Marrow et al. [127] on predator-prey dynamics with small mutational effects. A more complex theoretical framework was intoduced by Metz, Geritz, Bolker, Pacala, Dieckman,

Law and coauthors in the 1990s [139, 81, 20, 21, 59, 60]. They present mainly deterministic models but also outline first stochastic versions.

A basic principle of adaptive dynamics is the separation of evolutionary and ecological time scales due to relatively rare mutation. It is often assumed that mutations occur at a low enough frequency such that the population has reached an equilibrium state before new traits arise. This assumption is certainly debatable as it is not satisfied for many biological systems. However, if one assumes that small mutant populations grow slow enough, compared to the dynamics of the larger bulk populations, it is still reasonable to assume that the fixation of a new trait mainly depends on its fitness within a population at equilibrium. A more detailed discussion of the term of fitness and the different time scales that are involved is given in Sections 1.2 and 1.3.

In adaptive dynamics, the aim is to study the consequences that microscopic events, like interactions on the individual level and changes induced by single mutant individuals, have on the macroscopic level of the whole population. This yields the analysis of a different kind of fitness landscape than the one in population genetics. Here, the fitness of a traits depends on its ecological background. The task is to identify so-called evolutionary stable states (ESS), where all mutant traits have negative fitness, and therefore the adaptation of the population comes to a halt. Those ESSs might not be unique and it is of interest to study the course of evolution and conditions under which different ESSs are attained. These thoughts are discussed extensively by Metz in his essay *Adaptive Dynamics* [138]. For a broad collection of publications on adaptive dynamics we refer to the webpage of Kisdi [109].

A very important class of mathematical models for adaptive dynamics, so-called *individual-based models*, were proposed by Bolker and Pacala in the context of the evolution of plants [20]. They were further developed by Dieckmann and Law [60] for a spatially structured population and finally rigorously constructed by Fournier and Méléard in terms of Poisson random measures [77]. Many versions of this model have been studied over the last 15 years, focussing on the limit dynamics in large populations with rare mutations (see the next sections for details). Depending on whether one assumes the so-called invasion implies fixation principle or not, the limit is a monomorphic process, e.g. the *traits substitution sequence* (TSS) [140, 31], or a polymorphic process, e.g. the *polymorphic evolution sequence* (PES) [37]. Individual-based models are a central object of this thesis and are introduced in more detail in Section 1.2. Besides large populations and rare mutations, a third limit parameter is considered for such models, namely small mutational effects. On a continuous trait space, this leads to a continuous motion of adaptation of the macroscopic population,. It is described by the canonical equation of adaptive dynamics (CEAD), which was introduced by Dieckman and Law [59]. The CEAD is derived from the TSS by Champagnat, Ferrière, and Ben Arous [32] and directly from the individual-based model by Baar, Bovier, and Champagnat [10]. The CEAD focuses on monomorphic populations. However, one can study the possibility of polymorphism due to evolutionary branching under certain conditions on the fitness landscape. First heuristics towards this were developed by Bovier and Champagnat and are stated in [22].

1.1.5 Adaptive walks

Another prominent example of a mathematical model for adaptive dynamics are *adaptive walks*. Here, the idea is to focus on the macroscopic picture and study the trajectory of the adapting population on a discrete trait space. The concept was introduced by John Maynard Smith in the 1960s [131, 132] and further developed by Kaufman and Levin more than 15 years later [102, 103]. The idea of Maynard Smith was that mutations are rare and most genetic variants are dysfunctional. As a consequence, from one genetic trait, only few fitter traits can be reached. Therefore, it makes sense to study evolution as a random walk on a discrete trait space, with a graph structure that illustrates the possibility of mutation. There are two sources of randomness in adaptive walk models. The vertices of the trait graph are usually endowed with a random but fixed fitness landscape that assigns a real value to each trait (more in the spirit of population genetics). Moreover, the walk randomly moves along the edges of the graph towards fitter neighbours according to some given transition law. Note that a common assumption in this framework is that the population stays macroscopically monomorphic, i.e. there is no branching.

Questions that are of interest in this context, are the accessibility of certain traits [161, 13, 14, 113], the distribution of final states of the adaptive walk (related to the concept of ESSs) [147], and the average length of a path before reaching such a final state [150]. The results very much depend on the choice of fitness landscape and transition laws. Since random walks can only move to fitter neighbours, final states are local maxima of the fitness landscape. An extended version of adaptive walks, where transitions are determined by balancing high fitness increase and short distance on the graph, overcomes this property. It allows for longer jumps that can skip over valleys in the fitness landscape [99, 98, 100]. More recently, so-called *adaptive flights* that travel from one local maximum to another were introduced in [144]. The topic of crossing a fitness valley has already been studied by Gillespie in the 1980s [85]. He estimated the time for double mutations to occur, where the single mutations themselves are disadvantageous. Metastable behaviour reminding of adaptive flights can also be witnessed in limits of individual-based models and is discussed in Section 1.3.

Two particular types of adaptive walks, characterised by their transition law, are of interest in this thesis. On one hand, the *natural adaptive walk* jumps to fitter neighbours with a probability that is proportional to the increase in fitness. On the other hand, the *greedy adaptive walk* always jumps to the fittest of its neighbours (as long as they are fitter than the current trait) [147]. An extensive overview of current results on adaptive walks and flight can be fond in Joachim Krug's article [113].

1.2 The models discussed in this thesis

In this section we introduce the two models that are the focus of this thesis: An individual-based Markov process and the corresponding deterministic system of Lotka-Volterra equations with mutation. We briefly discuss the existence of equilibrium states and comment on different notions of fitness that are used in the context of these models.

In this thesis, we consider the evolution of an asexually reproducing haploid organism that adapts to its environment through mutation. Individuals or subpopulations are characterised by their geno- or phenotype, which we call trait or type. Note that, since we consider trait changes through mutations, it is possibly more fitting to think of traits as genotypes. However, we focus on those mutations that have a phenotypic effect, i.e. change the fitness of an individual. We consider the case of a possibly large but finite trait space, given by a graph $\mathcal{G} = (V, E)$. The set of vertices V denotes the set of possible traits. To give an example, we could choose $V = \mathbb{H}^n = \{0, 1\}^n$, the n-dimensional hypercube. The sequences of zeros and ones can be interpreted as sequences of different genes being activated or not. The set of edges E marks the possibilities to obtain one trait from another by mutation. If mutations are not reversible, we consider a directed edge set E. In the case of the hypercube, we choose $E = \{(v, w) \in \mathbb{H}^n \times \mathbb{H}^n : \|v - w\|_1 = 1\}$, i.e. single mutations that (in)activate exactly one of the genes in the sequence.

The dynamics of both processes are driven by a number of events that happen at exponential rates. Either as discrete events in the case of the Markov process or as continuous changes in the case of the differential equation. These events are clonal reproduction, reproduction with mutation, and death. The rates depend on the current state of the population and a number of parameters:

For traits $v, w \in V$, we introduce

- $b_v \in \mathbb{R}_+$, the birth rate of an individual of trait v,
- $d_v \in \mathbb{R}_+$, the (natural) death rate of an individual of trait v,
- $c_{v,w} \in \mathbb{R}_+$, the competitive pressure imposed by an individual of trait w onto an individual of trait v,
- $\mu \in [0, 1]$, the probability of mutation at a birth event,
- $m(v, \cdot) \in \mathcal{M}_p(V)$, the law of the trait of a mutant offspring produced by an individual of trait v.

Due to the interpretation of the edge set E as possibilities of mutation, we assume that $m(v, w) > 0$ if and only if $(v, w) \in E$. Moreover, we assume that $c_{v,v} > 0$, i.e. there is always self-competition within a subpopulation of a certain trait.

1.2.1 An individual-based model

Individual-based Markov processes have been rigorously introduced by Fournier and Méléard [77], as mentioned in the previous section. There are many works that consider variations of this model. For example, some models take into account infinite trait spaces [31, 33, 37, 10], diploid organisms and sexual reproduction [41, 146, 24], or spatial structure [119]. Moreover, phenotypic switches [9], predator-prey relations [45], and seed-bank dynamics [19] are investigated. For simplicity, we only introduce the version of this class of individual-based models that we study in this thesis.

The trajectories of the Markov process, denoted by $(N(t))_{t\geq 0}$, describe the numbers of individuals of the different traits over time. Here, $N_v(t)$ gives the number of individuals of trait $v \in V$ at time t. The overall rates for the different events, in a population at state $N \in \mathbb{N}^V$, are

- $N_v b_v(1-\mu)$ for a clonal reproduction event of trait v,
- $N_v b_v \mu m(v,w)$ for a reproduction event of trait v with mutation towards trait w,
- $N_v \left(d_v + \sum_{w\in V} c_{v,w} N_w\right)$ for a death event of trait v.

The dynamics are summarised by the infinitesimal generator of the process,

$$\begin{aligned}\mathcal{L}\phi(N) = &\sum_{v\in V} (\phi(N+\delta_v) - \phi(N)) \left(N_v b_v (1-\mu) + \sum_{w\in V} N_w b_w \mu m(w,v) \right) \\ &+ \sum_{v\in V} (\phi(N-\delta_v) - \phi(N)) N_v \left(d_v + \sum_{w\in V} c_{v,w} N_w \right). \end{aligned} \tag{1.3}$$

Here, the functions $\phi : \mathbb{N}^V \to \mathbb{R}$ are bounded. As mentioned before, this process is constructed by Fournier and Méléard in [77] in terms of Poisson random measures. It can also be constructed algorithmically, following the procedure of a Gillespie algorithm [83].

The individual-based model above is based on classical birth-and-death processes, or branching processes in general. These processes consider the growth of a population with a single or multiple traits without interaction and are well studied (e.g. in [4, 5, 162, 70]). We make extensive use of them to approximate the growth of different subpopulations in the individual-based model (1.3). A classical example are Galton-Watson processes, where, in discrete generations, individuals produce offspring according to some distribution on $\mathbb{N}$ [172]. Birth-and-death processes arise as a continuous-time version, where only 0 and 2 are allowed as numbers of offspring, occuring at rates d and b, corresponding to a death or birth event. As long as $b - d > 0$, the population in a birth-and-death process grows roughly exponentially. To limit the population size as a result of limited resources, a competitive term is usually added. In this case, classical branching process results are no longer applicable since they assume independence of the different individuals and do not allow for an interaction between them. However, there exists a theory of logistic birth-and-death processes, where the exponential growth is limited to logistic growth by an interactive competition term. Here, the equilibrium size of the population is at $(b-d)/c$, where c is the parameter for the competition. Moreover, the results can be extended to (logistic) birth-and-death processes with immigration, where the latter corresponds to incoming mutations. In [38] and previous publications, the authors derive several important results on the mean behaviour of such processes in dimension one and two. In Chapter 3 we extend the results to the multidimensional case. This allows us to approximate the individual-based model during an invasion phase where more than two traits impose notable competitive pressure.

1.2.2 A system of differential equations

The previously described dynamics can also be modelled by a deterministic system of differential equations that is another important object of study in this thesis:

$$\dot{n}_v(t) = \left(b_v(1-\mu) - d_v - \sum_{w \in V} c_{v,w} n_w(t) \right) n_w(t) + \mu \sum_{w \in V} b_w m(w,v) n_w(t), \quad v \in V,\ t \geq 0. \tag{1.4}$$

Here, $n_v(t)$ describes the population size of trait $v \in V$ at time $t \geq 0$. Note that this size is neither in terms of numbers of individuals nor in terms of frequencies, i.e. normalised, but simply a value in $\mathbb{R}$. Since we assume $c_{v,v} > 0$ for each $v \in V$, the competition term ensures boundedness for any solution n. This implies Lipschitz continuity of the coefficients and ensures existence, uniqueness (for fixed initial condition), and continuity of the solution. Moreover, we are only interested in studying biologically reasonable, non-negative states of the system. The form of the equations guarantees that, for $n(0) \in \mathbb{R}_+^V$, we obtain $n(t) \in \mathbb{R}_+^V$ for any $t \geq 0$. The relation of this deterministic system to the stochastic individual-based model is discussed in Section 1.3.

Equation (1.4) describes a competitive Lotka-Volterra system with additional mutation terms. In this thesis we study regimes of rare mutations. It is essential to understand the behaviour of the mutation-free classical Lotka-Volterra system, i.e. (1.4) with $\mu = 0$ or (1.2), since it governs the short-term dynamics of the more complex processes. In particular, it is of interest to study fixed points, so-called equilibria, of the equation.

For a subset $\mathbf{v} \subset V$, $\bar{n} \in \mathbb{R}_+^{\mathbf{v}}$ is an equilibrium state if

$$\left(b_v - d_v - \sum_{w \in \mathbf{v}} c_{v,w} \bar{n}_w \right) \bar{n}_v = 0, \quad \forall\ v \in \mathbf{v}. \tag{1.5}$$

Note that 0 is always a fixed point. If there exists a unique equilibrium in $\mathbb{R}_{>0}^{\mathbf{v}}$, we denote by $\bar{n}(\mathbf{v})$ its extension to $\mathbb{R}_+^V$ by zero. If, for some open subset $\bar{n}(\mathbf{v}) \in U \subset \mathbb{R}_+^{\mathbf{v}}$, the solution of the mutation-free Lotka-Volterra system converges to $\bar{n}(\mathbf{v})$, for any $n_0 \in U$, $\bar{n}(\mathbf{v})$ is called (asymptotically) stable or attractive. These points can also be characterised in terms of eigenvalues of the Jacobian matrix of the function g, where $g_v(n) = n_v(b_v - d_v - \sum_{w \in \mathbf{v}} c_{v,w} n_w)$.

There is no exhaustive study on the existence of such stable equilibria in arbitrary dimensions. Moreover, Smale has shown that many different and complex long-term behaviours can arise, ranging from fixed points to limit cycles and multiple attractors [165]. However, there exists a complete analysis for dimensions up to three and there are some results on sufficient criteria for existence in general dimensions. In the following we give a short overview.

In the case of just one trait $\mathbf{v} = \{v\}$, the situation is easy. If $b_v - d_v > 0$, the unique stable equilibrium is given by $\bar{n}_v(v) = (b_v - d_v)/c_{v,v}$. If $b_v - d_v \leq 0$, then 0 is the only fixed point and stable.

In the two-dimensional case, where $\mathbf{v} = \{v_1, v_2\}$, there are four potential fixed points, namely $(0,0)$, $(\bar{n}_{v_1}(v_1), 0)$, $(0, \bar{n}_{v_2}(v_2))$, and

$$\left(\frac{(b_{v_1} - d_{v_1})c_{v_2,v_2} - (b_{v_2} - d_{v_2})c_{v_1,v_2}}{c_{v_1,v_1}c_{v_2,v_2} - c_{v_1,v_2}c_{v_2,v_1}}, \frac{(b_{v_2} - d_{v_2})c_{v_1,v_1} - (b_{v_1} - d_{v_1})c_{v_2,v_1}}{c_{v_1,v_1}c_{v_2,v_2} - c_{v_1,v_2}c_{v_2,v_1}} \right). \tag{1.6}$$

Depending on whether they are non-negative and on the relation of the parameters, different points are stable equilibria. Only in the case where $b_{v_1} - d_{v_1} = b_{v_2} - d_{v_2}$ and $c_{v_1,v_1} = c_{v_1,v_2} = c_{v_2,v_1} = c_{v_2,v_2}$, every state n such that $n_{v_1} + n_{v_2} = (b_{v_1} - d_{v_1})/c_{v_1,v_1}$ is a fixed point.

In dimension three, already many different scenarios are possible, including coexistence of up to three traits and cyclic behaviour. A complete classification with conditions on the parameters b, d, and c is given in a paper from Zeeman, see [180].

For higher dimensions, there exists no full classification yet. However, there are results on sufficient criteria for the existence of stable equilibria, as for example by Champagnat, Jabin, and Raoul [35]. Under certain symmetry and positive definiteness assumptions on the matrix $C = (c_{v,w})_{v,w \in \mathbf{v}}$, they show existence and uniqueness of stable equilibria with the help of Lyapunov functions.

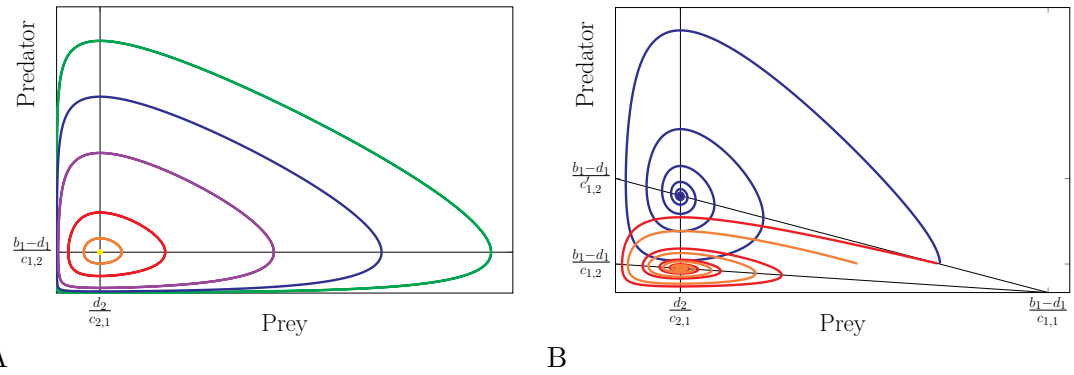
It is also possible to relax the assumption of $c_{v,w} \geq 0$ and allow for beneficial interactions between individuals. To give an example, one can consider predator-prey relations. Here, the presence of a prey trait increases the growth rate of a predator trait, while the prey trait experiences an increased death rate in the presence of a predator. Similar scenarios arise for parasite-host relations. Moreover, there is also the possibility of mutually beneficial interactions in the context of symbiosis. Allowing for more general parameters also increases the number of different behaviours, where, for example in a predator-prey setting, periodic cycles already arise in a two-dimensional system.

As an example, we briefly consider a specific case of predator-prey dynamics. It demonstrates behaviour that is referenced for the interactions of melanoma cells and T-cells in Chapter 4. We study the system given by

$$\begin{aligned} \dot{n}_1(t) &= (b_1 - d_1 - c_{1,1}n_1(t) - c_{1,2}n_2(t))n_1(t), \\ \dot{n}_2(t) &= (c_{2,1}n_1(t) - d_2)n_2(t). \end{aligned} \tag{1.7}$$

All parameters are assumed to be positive, except that $c_{1,1}$ might be zero, and $b_1 - d_1 > 0$. Trait 1 symbolises the prey and trait 2 the predator. Note that $c_{2,1}$ appears in a beneficiary term for the predator, as it feeds off the prey. Depending on whether we allow for self-competition of the prey, i.e. $c_{1,1} > 0$, different behaviour arises. As seen in Figure 1.1, if $c_{1,1} = 0$, solutions are periodic. There are limit cycles that depend on the initial condition and center around the one interior fixed point $(d_2/c_{2,1}, (b_1 - d_1)/c_{1,2})$. If self-competition is included, we observe fluctuating convergence towards a single stable equilibrium, independent of the initial condition. This stable equilibrium can be shifted, e.g. by changing the parameter $c_{1,2}$.

In the theoretical part of this thesis, we concentrate on purely competitive systems. However, the existence of equilibria is not the main focus of this work. Therefore, in many cases we assume the existence of unique stable equilibria and terminate the construction otherwise.



A B

Figure 1.1: Solution of (1.7) for parameters $b_1 = 3$, $d_1 = 1$, $c_{1,2} = 2$, $c_{2,1} = 0.5$, and $d_2 = 1$. Black lines indicate a change in signs of $\dot{n}_1(t)$ or $\dot{n}_2(t)$, solutions evolve counter-clockwise. (A) Solutions for $c_{1,1} = 0$ and different initial conditions. (B) Solutions for $c_{1,1} = 0.15$ and different initial conditions (orange and red) or $c'_{1,2} = 0.5$ (blue).

1.2.3 Notions of fitness

We now look into the term of *fitness landscapes*. There are at least two notions of fitness that are used in the context of mathematical models for evolution. In this thesis, we call them *individual fitness* and *context-dependent* or *invasion fitness*. Both terms quantify how well-adapted a trait is to its environment. They are related to the exponential growth rate of a subpopulation of a certain trait.

The individual fitness is fixed over time and assigns a value in $\mathbb{R}$ to each trait. It does not reflect an interaction with other individuals. For the individual-based model that we consider, the individual fitness r_v of a trait $v \in V$ is defined by $r_v := b_v - d_v$. This value describes the average growth rate in the competition-free pure birth-and-death process. It also quantifies the exponential growth in the corresponding competition-free Lotka-Volterra system.

The system we consider is not competition-free and the competitive interaction with other individuals limits the actual growth rates of the subpopulations. Therefore, the context-dependent fitness $f_v(N)$ of a trait $v \in V$ in a population at state N is defined through $f_v(N) := b_v - d_v - \sum_{w \in V} c_{v,w} N_w$. If there is just one individual of a trait $w \in V$ within a bulk population of traits $\mathbf{v} \subset V$ close to their equilibrium $\bar{n}(\mathbf{v})$, this context-dependent fitness is approximately equal to $f_{w,\mathbf{v}} := b_w - d_w - \sum_{v \in \mathbf{v}} c_{w,v} \bar{n}_v(\mathbf{v})$. Since it describes the initial growth rate of a new mutant w invading an equilibrium population of $\mathbf{v}$, $f_{w,\mathbf{v}}$ is called invasion fitness. It was first introduced in a more abstract way by Metz et al. in [139] and formalised for individual-based models as above by Champagnat and Méléard in [31, 37]. If the context-dependent fitness of a trait is negative, its population can be approximated by a sub-critical branching process and is likely to die out. If it is positive, the population behaves approximately like a super-critical branching process and has a positive probability of growing and fixating in the population before extinction. Similar implications are valid for the deterministic system. The long-term behaviour of competitive Lotka-Volterra systems can be characterised in terms of the invasion fitnesses of the different traits.

These two notions of fitness are related to the concept of fixed fitness landscapes in the context of adaptive walks and flights in the following way: If we assume that the competition between each two traits is constant, i.e. $c_{v,w} \equiv c$, for all $v, w \in V$, then the invasion fitness of a trait w in a monomorphic equilibrium population of trait v simplifies to $f_{w,v} = r_w - r_v$. In this case, w can invade the population of trait v if and only if $r_w > r_v$. This is in line with the transition probabilities of an adaptive walk, where (in most cases) the random walk can jump to a neighbour of the current trait if and only if the new trait has higher fitness.

1.3 Approximation at different scales

Stochastic individual-based models are good in the sense that they are quite realistic representations of certain biological systems (e.g. bacteria). However, due to the stochasticity, there are many possible outcomes for single realisations of the process. Moreover, simulations of individual-based models are computationally heavy, which is impractical for applications. We are interested in the long-term evolution of the system. The goal is therefore to approximate the mean behaviour of the individual-based model by simpler, partially deterministic processes and to characterise the adaptive evolutionary path that the population takes on the trait space.

There are two quantities that shape the long-term behaviour of the system, namely the mutation probability and the fitness landscape. The mutation probability decides how fast or often mutations occur and whether the system has time to reach a new equilibrium between mutations. Once a mutation occurs, the fitness landscape determines whether the new trait can fixate in the population or dies out immediately, and if a new equilibrium state is reached.

In this section we first present a deterministic approximation of the individual-based model for large populations with frequent mutations. It relates the two models that were mentioned in Section 1.2 and is the basis for all following approximations, as it determines the short-term dynamics of the individual-based model. Second, we give an extensive overview of the different scales that are involved when considering the simultaneous limit of large populations and rare mutation and provide previous results that have been derived for the different regimes. Third, we discuss the limit of rare mutations in the deterministic system. Finally, we consider several quantities that have to be studied in order to characterise the limiting processes in these different scenarios.

1.3.1 Large population approximation

A first approach to approximate the stochastic system (1.3) is a law of large numbers. However, since the population size is not fixed, we need to introduce a parameter to scale the average population size. As discussed above, the equilibrium size of a logistic birth-and-death process with parameters b, d, and c is $(b-d)/c$. We therefore scale the competition by a parameter K as $c^K_{v,w} := c_{v,w}/K$ and denote the corresponding stochastic process by N^K. As a result, the equilibrium size of any coexisting population of N^K is of order K. Note that

this does not change the notion of invasion fitness that was discussed in the previous section. If we substitute $c^K_{w,v}$ for $c_{w,v}$ and $\bar{n}(\mathbf{v})K$ for $\bar{n}(\mathbf{v})$, the Ks cancel out and $f_{w,\mathbf{v}}$ remains the same.

The parameter K is sometimes called *carrying capacity* and can be interpreted as the capacity of the environment to support living organisms in terms of space, nutrients etc. Individuals compete for these resources and, as the capacity increases, the competitive pressure is reduced.

The classical result by Ethier and Kurtz [70, Chap.11, Thm.2.1] now yields convergence of the rescaled process N^K/K, as $K \to \infty$. Under the assumption that $N^K(0)/K$ converges to $n_0 \in \mathbb{R}^V_+$, N^K/K almost surely converges uniformly to the solution n of (1.4) with initial condition n_0, on any finite time interval $[0, T]$. As a result, for large populations, the short-term dynamics of the stochastic process are governed by the deterministic system. Note that a similar result for continuous trait spaces is proved by Fournier and Méléard in [77].

For the rescaled process N^K/K, we can distinguish traits with population sizes that vanish in the limit of $K \to \infty$ and those that do not. According to this, at time $t \geq 0$, we call traits $v \in V$ with $N^K(t)/K \to 0$ *microscopic*, and *macroscopic* otherwise. The macroscopic traits that are close to their (possibly coexisting) equilibrium size are called *resident*.

1.3.2 Multiscale approximation in the large population-rare mutation limit

Many genetic mutations that occur do not have a detectable effect on the phenotypic level. This is because they only slightly vary the genome and thus not influence the overall translation of DNA enough to cause a change in fitness. Other mutations are lethal since they damage a part of the genome that is essential for survival. The frequency of effective but non-lethal mutations is a topic of ongoing discussion and dependent on the considered species. However, with a finite trait space we focus on a projection of the full genome onto a collection of alleles that determine phenotypic characteristics of interest. For this collection of specific genes, it is reasonable to assume that effective mutations that produce a new phenotype are rare. In the individual-based model for evolution that we consider, we thus let the mutation probability μ_K depend on K and again denote the corresponding process by N^K. In the limit of $K \to \infty$, we let μ_K either converge to zero or to some small $\mu > 0$.

The resulting individual-based Markov process N^K is a multiscale model. There are several phenomena that occur on different time scales. Depending on the choice of mutation rate and rescaling of time, multiple different stochastic or deterministic processes can arise in the limit of $K \to \infty$. In the following, we give an overview of the different scales involved.

For simplicity we first consider a single invasion and study the different phases. They are visualized in Figure 1.2.

Phase 1: Arrival of a mutant
If the resident population has a size of order K and the probability of mutation at birth is of order μ_K, then the occurrences of mutations behave like a Poisson point process with intensity of order $K\mu_K$. The waiting time until a single mutant individual arrives is hence of order $1/K\mu_K$.

Phase 2: Growth to a macroscopic level
If a mutation occurs, the mutant trait has positive invasion fitness, and it does not go extinct due to random fluctuations, its population initially grows roughly exponentially. This is because the self-competition is not relevant yet and the resident traits have roughly constant size. The exponential growth rate equals the invasion fitness of the mutant trait. In a resident population of traits $\mathbf{v} \subset V$, the time that a mutant population of trait w takes to grow from a level of K^γ individuals, $\gamma \in (0,1)$, to a macroscopic population size of order K is roughly $(1-\gamma)\log K / f_{w,\mathbf{v}}$, i.e. of order $\log K$.

Phase 3: Invasion/Re-equilibration
Once the mutant trait reaches a macroscopic population size, the populations of the resident traits and the mutant trait behave according to the corresponding mutation-free Lotka-Volterra system. Since the initial conditions for the Lotka-Volterra dynamics are of order K, i.e. non-vanishing when rescaled by K, the time until the system is close to its new equilibrium (if existent) can be bounded uniformly by a time of order 1. In the stochastic system this can be argued since individual birth and death events occur at frequency 1 times the population size. In a population with size of order K, changes of order K can therefore ba achieved in a time of order 1.

Phase 4: Extinction
After the new equilibrium is reached, the former resident traits that are no longer resident have a negative fitness and thus their population size declines. Equivalently to the second phase, their populations take a time of order $\log K$ to go extinct.

Depending on the choice of the mutation probability μ_K and the time scale, some of these phases are visible or not.

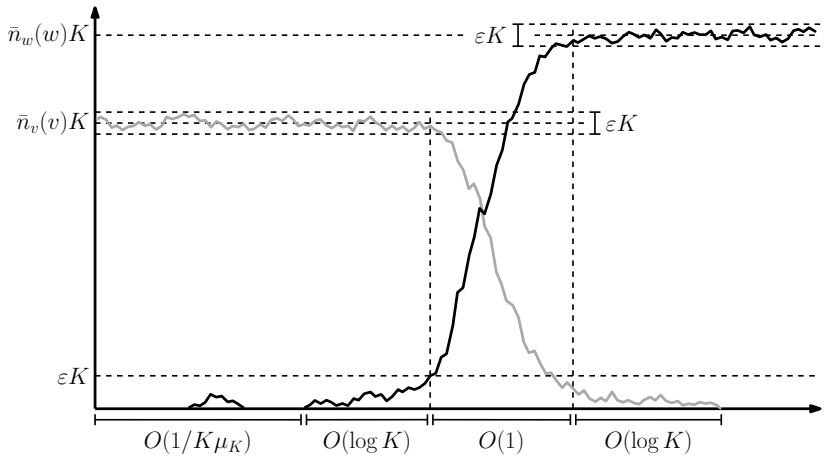


Figure 1.2: Four phases of an invasion and their duration, for the case of one resident trait (grey) and one mutant trait (black).

The general strategy to prove convergence for the individual-based model is to approximate the microscopic mutant populations by branching-processes during the phase of growth. This is possible since one can use the stability of the resident population at equilibrium,

e.g. due to large deviation results from [63, 79], and assume the competitive pressure to be roughly constant. Once a mutant reaches a macroscopic population size, and does hence no longer vanish in the limit, the invasion of the resident population can be approximated by the corresponding Lotka-Volterra dynamics. The occurrence of mutations is estimated by Poisson processes. A similar strategy is applied for the convergence of the deterministic system in the limit of rare mutations, as mentioned below. Here, the growth is approximated by sums of pure exponential functions and stability of the resident population is shown via stability of the equilibrium state under slight perturbations.

The above considerations lead to the distinction between three time scales: First, the short time scale of order 1, on which the Lotka-Volterra dynamics between macroscopic traits play out. Second, the *evolutionary time scale* of order $1/K\mu_K$, on which mutations arise. Third, the *ecological time scale* of order $\log K$, on which microscopic populations grow according to their invasion fitness. We can distinguish four different regimes of mutation rates μ_K that we can study on these multiple time scales. In the following, we summarise what is known about the different cases. Most of the results are also assembled and discussed in a little more detail in a review article by Bovier [23]. All convergences, orders, and notions of macro- or microscopic are meant with respect to the limit of $K \to \infty$.

Regime 1: $\mathrm{e}^{-VK} \ll \mu_K \ll \frac{1}{K \log K}$

In this scenario, we obtain $\log K \ll 1/K\mu_K \ll \mathrm{e}^{V'K}$. As a result, mutations occur before the resident populations can go extinct due to random fluctuations [79]. However, mutations are extremely rare and the waiting time between single mutations is much larger than the time that a mutant population needs to grow from one individual to a macroscopic level (if it successfully does so). Therefore, mutation events are separated in the way that a new trait either fixates or goes extinct before the next mutation occurs.

If time is not rescaled, microscopic traits grow too slowly and new mutations are too rare. Therefore, N^K/K converges almost surely to the solution n of the mutation-free Lotka-Volterra system ((1.4) with $\mu = 0$) that involves the traits $v \in V$ with $N_v^K(0)/K \to n_{0,v} > 0$. After reaching an equilibrium state between those traits (if existent), there is no further evolution.

If we consider the time scale $\log K$, mutations are still too rare to be seen but traits with microscopic initial condition such that $N_v^K(0) \sim K^\gamma$, $\gamma \in (0,1)$, can grow and hence reach a macroscopic population size. In this case, the time to obtain a new equilibrium with the former resident traits vanishes. Thus, N^K/K converges in probability to a deterministic jump process that jumps between coexistence equilibria involving only traits with the mentioned initial conditions. The characterisation of this jump process is a topic of this thesis and is discussed in Section 1.6 and Chapter 3.

The most interesting time scale for this regime is $1/K\mu_K$. Here, the waiting time between mutations is finite but non-vanishing. Assuming that the initial living traits reach a final state according to the previous paragraph, all traits that are not part of this final coexistence equilibrium die out within a time of order $\log K$, i.e. before the first mutation occurs. The time that a mutant population needs to grow from a single individual to a macroscopic

level vanishes. The occurrence of a mutant, the trait of the mutant, and the choice whether it fixates or not are random. Consequently, N^K/K converges for finite dimensional distributions to a random jump process that jumps between (coexistence) equilibria between different subsets of traits. It is called *trait substitution sequence* (TSS) [31] or *polymorphic evolution sequence* (PES) [37], depending on whether coexistence is allowed or not. In the monomorphic case, jumps from v to w occur at a rate proportional to $m(v,w)[f_{w,v}]_+/b_v$. This is similar to a natural adaptive walk, which jumps to direct neighbours with a probability proportional to $[f_{w,v}]_+$. Both processes move along edges of the graph towards higher fitnesses until a local maximum is reached, i.e. a trait $v \in V$ such that, for every neighbour w, $f_{w,v} < 0$.

Regime 2: $\mu_K \sim \frac{1}{K \log K}$

In this setting, mutations occur at the same speed as the mutant populations grow. The ecological and evolutionary time scale are therefore not separated. Multiple, but not necessarily all possible, mutants may interact at the same time. If time is not rescaled, mutations are too rare and, as in the previous scenario, N^K/K converges to the solutions of the mutation-free Lotka-Volterra system involving all initially macroscopic traits.

On the time scale $1/K\mu_K \sim \log K$, many different outcomes are possible due to random events. So far, there exists no general study of the limit process (if existent). However, for small trait spaces with only two or three possible mutants, this scaling has recently been considered by Billiard and Smadi in [15, 16]. They derive conditions for coexistence and fixation, and study the invasion time.

Regime 3: $\mu_K \gg \frac{1}{K \log K}$, $\mu_K \to 0$

This is the scenario that we focus on in this thesis. Since $1/K\mu_K \ll \log K$ in this case, mutations occur much more frequent and are no longer separated. However, since $\mu_K \to 0$, mutant traits can not reach a population size of order K through incoming mutations alone and the $\log K$ growth phase governs the dynamics.

If time is not rescaled, either mutations occur too rarely ($1/K\mu_K \to \infty$) or mutations in fact arise but the mutant population stays at a microscopic level since the time to reach a macroscopic population size is of order $\log K$. As a result, N^K/K converges almost surely to the solution n of the mutation-free Lotka-Volterra system ((1.4) with $\mu = 0$) that involves all the traits $v \in V$ with $N_v^K(0)/K \to n_{0,v} > 0$. If existent, the process reaches an equilibrium between those traits within a time of order 1 and remains in this state.

Similarly, on the mutational time scale $1/K\mu_K$, mutations frequently occur but the mutant populations can not grow to a macroscopically visible size. Depending on how μ_K decreases compared to $1/K$, more or less of the Lotka-Volterra dynamics mentioned above are visible. If $\mu_K \gg 1/K$, e.g. $\mu_K \sim 1/\log K$, changes in the population due to birth and death events are of an order that is macroscopically not visible and N^K/K converges to the constant process at state $\lim_{K\to\infty} N^K(0)/K$. In the case of $\mu_K \sim 1/K$, time is not rescaled and the result is as in the previous paragraph. If $1/K \gg \mu_K \gg 1/(K \log K)$, the Lotka-Volterra

dynamics are faster than the occurrence of mutations and therefore N^K/K converges to the constant process that takes the value of the Lotka-Volterra equilibrium of all traits $v \in V$ with $N_v^K(0)/K \to n_{0,v} > 0$ (if existent).

The time scale that is mostly considered in this thesis is the one of $\log K$. On this scale mutations occur infinitely fast and the growth of microscopic populations is visible. Since mutations are much more frequent than in the previous scenarios, every possible mutation of the resident traits arises infinitely often and fixates, as long as it has a positive fitness. As a result, in the macroscopic picture, mutations are no longer random and the limiting process of N^K/K is a deterministic jump processes that moves between Lotka-Volterra equilibria of different subsets of V. The dynamics of this process are rather complicated since multiple growing mutant populations compete to invade the resident population at the same time. They are one of the main issues of this thesis and are further discussed in Section 1.6 and Chapter 3. We consider the case of power law mutation rates, $\mu_K = K^{-1/\alpha}$, $\alpha > 0$. Note that, for $\alpha < 1$, this falls under regime 1. This scaling of negative powers of K was previously discussed in [25, 38]. However, the authors consider a more restrictive setting, where the trait graph consists of a subset of $\mathbb{N}$ and a specific fitness landscape is assumed. Depending on the choice of α, the limiting jump processes interpolate between adaptive walks and adaptive flights. They can cross fitness valleys of width smaller than α and possible final states are local maxima such that no fitter trait exists in a radius of α.

Regime 4: $\mu_K \to \mu > 0$

In this case, mutations occur on the same time scale as birth and death events. The mutant populations therefore reach a size of order K simply due to incoming mutations (and not their own population growth) within any time of order 1. Consequently, the $\log K$ growth phase becomes obsolete for those traits. Since we consider a finite trait space, every trait that can be reached through a chain of mutations from an initial resident trait is macroscopically present within any time of order 1.

If we do not rescale time, N^K/K converges almost surely to the solution n of (1.4). All traits that are accessible through mutations from a trait $v \in V$ with $N_v^K(0)/K \to n_{v,0} > 0$, have a positive population size $n_v(t)$ for any $t > 0$. Under certain assumptions on the parameters for competition and mutation, $c_{v,w}$ and $\mu m(v,w)$, which allow to construct a Lyapunov functional, Coville and Fabre show convergence to a stable equilibrium in such systems [47].

If we consider the process on the $\log K$ time scale, the traits that are accessible from the initial resident traits immediately reach an equilibrium state of (1.4), if existent. Since we do not assume that the trait graph is connected, there might however be other components, that do not contain any of the initial resident traits, but some traits $v \in V$ with $N_v^K(0)/K \sim K^\gamma$, $\gamma \in (0,1)$. If those traits have positive fitness in the mentioned equilibrium state, they can grow on the $\log K$ time scale and invade the resident population to change the equilibrium. Note that, once a new trait reaches a population size of order K, immediately every trait that is accessible through mutation from this trait does so, too. Hence, the limiting process of

N^K/K jumps between equilibria of (1.4) (if existent), taking into account increasing subsets of V.

One could also consider the somewhat artificial mutational time scale $1/K\mu_K \sim 1/K$. This is the time scale on which single mutation events but also birth and death events occur. Since these change the state of the system by single individuals, in the rescaled process N^K/K, changes are of size $1/K$ and not visible in the limit. However, if we consider $N^K(t/K) - Kn(t/K)$, it is possible to study the Poisson variation around the mean state of the population.

A forth time scale for metastability

The three time scales that are considered above (Lotka-Volterra dynamics, occurrence of mutations, and growth of mutant populations) are derived from the study of mutants that directly stem of a resident trait. They occur at rate $K\mu_K$. However, these mutants, having an initial population size of order $\mu_K K$, are also able to produce second order mutations at rate $K\mu_K^2$, and similar for traits with higher distance to the initial resident traits. As long as $K\mu_K^\ell \gg 1$, the higher order mutants with distance ℓ to the resident traits still have a positive population size and can grow on the $\log K$ time scale. If however $K\mu_K^\ell \ll 1$, these traits only occur as mutations extremely rarely. They are only visible on the time scale $1/K\mu_K^\ell$. On this time scale, assuming that $1/K\mu_K^\ell \gg \log K$, we can observe a metastable behaviour.

Regime 1 constitutes the trivial case with $\ell = 1$, where $K\mu_K \ll 1$ and mutations are visible on the time scale $1/K\mu_K$. In regime 3, according to the dynamics described above, the process first attains a locally stable equilibrium on the $\log K$ time scale which is surrounded by a fitness valley of respective width. This step is immediate on the new, more accelerated time scale. The process then randomly gains access to further parts of the trait graph and, once it does so, promptly attains a new local equilibrium.

In the case of the trait space $[0, \ell] \cap \mathbb{N}$ and power law mutation rates, this metastable time scale is rigorously studied by Bovier, Coquille, and Smadi in [25]. Two interesting scenarios of non-linear trait spaces are also heuristically discussed in this thesis as a part of [42].

1.3.3 Rare mutations in the deterministic system

Another approach is to consider the limits of large populations and rare mutations separately. This is done letting the mutation probability μ tend to zero in the deterministic system (1.4). This limit has first been proposed by Bovier and Wang in [26] and is further considered in [25] and in this thesis as part of [112]. Here, the terms microscopic and macroscopic are to be interpreted as vanishing and non-vanishing in the limit of $\mu \to 0$. All orders are given with respect to this limit.

In this case, due to the continuous flow of mutation, every trait that is attainable through mutation from a trait with positive initial size has a positive population size after any time $t_0 > 0$. This population size is of some order μ^γ, $\gamma \geq 0$. For example initial resident traits

satisfy $\gamma = 0$, neighbours of those $\gamma = 1$, and so on. The time to grow from a microscopic population size μ^γ, $\gamma > 0$, to a macroscopic size of order μ^0 with a roughly exponential rate is of order $\log 1/\mu$.

Let us denote the solution to (1.4) with parameter μ by n^μ. If time is not rescaled, only the traits with an initial population size of order 1 are visible in the limit since the other traits do not reach a macroscopic level within a finite time. As a result, n^μ converges to the mutation-free Lotka-Volterra system ((1.4) with $\mu = 0$) that involves all traits with $n^\mu(0) \to \tilde{n}_0 > 0$.

We do not have a mutational time scale since there is no waiting time before the first mutant appears. However, it is interesting to consider the time scale $\log 1/\mu$ on which microscopic traits can grow to a macroscopic level. As in regime 3 above, we see multiple mutants competing to invade the population. The first mutant trait to reach a macroscopic level and change the equilibrium according to the Lotka-Volterra dynamics is found optimising the ratio between the initial population size of the trait and and its speed of growth, i.e. its invasion fitness. Since the microscopic populations are not visible in the limit and the Lotka-Volterra phase vanishes on the accelerated time scale, n^μ converges pointwise almost everywhere to a deterministic jump process that moves between Lotka-Volterra equilibria involving different subsets of V. The derivation of this jump process is one of the main topics of this thesis and is further discussed in Section 1.6 and Chapter 2.

This scenario is a good first step towards studying limiting processes for vanishing but not too rare mutation rates. It is related to the simultaneous limit in the following way: All populations of positive size, no matter how small, are able to grow. The minimal population size of a trait, which is obtained by mutations from a resident trait, is of order $\mu^{\ell_\mathcal{G}}$, where $\ell_\mathcal{G}$ is the diameter of the trait space in terms of (directed) paths. In the stochastic system, this implies that $K\mu_K^{\ell_\mathcal{G}} \gg 1$, which is equivalent to $\mu_K \gg K^{-1/\ell_\mathcal{G}}$. This scenario falls into regime 3. Indeed, under this assumption, the limits of $N^K(t \log K)/K$, as $K \to 0$, and $n^\mu(t \log 1/\mu)$, as $\mu \to 0$, coincide (up to linear rescaling of time). In this thesis we study the connection between the simultaneous limit for mutation rates of the form $\mu_K = K^{-1/\alpha}$ and a slightly modified version of the deterministic system, where only traits with a population size above a certain threshold can grow. This version is introduced in [112].

1.3.4 Two processes to study as $K \to \infty$

There are two ways of studying the population in the limit of $K \to \infty$ or $\mu \to 0$ that capture different aspects. Depending on the time scale, N^K/K converges to a jump process that resembles an adaptive walk between (coexistence) equilibria and, more or less deterministically, describes the macroscopic adaptation of the population. In addition to the succeeding sets of resident traits, it also reflects the precise states of the Lotka-Volterra equilibria. For the rare mutation limit in the deterministic system, one alternatively studies the limit of n^μ.

Particularly in the case where multiple different mutants are microscopically present and can grow at the same time, it is of interest to consider

$$\beta_v^K(t) := \frac{\log(1 + N_v^K(t \log K))}{\log K}, \tag{1.8}$$

which is equivalent to $N_v^K(t \log K) = K^{\beta_v^K(t)} - 1$. This exponent, which asymptotically ranges in $[0, 1]$, gives more details on the growth of microscopic populations.

In the case of the rare mutation limit in the deterministic system there is no K. β^K is therefore substituted by

$$\rho_v^\mu(t) := \frac{\log(n_v^\mu(t \log(1/\mu)))}{\log \mu}, \tag{1.9}$$

which is equivalent to $n_v^\mu(t \log(1/\mu)) = \mu^{\rho_v^\mu(t)}$. In the limit of $\mu \to 0$, this exponent ranges in $[0, \infty)$ and macroscopic traits satisfy $\rho_v^K(t) = 0$. When comparing N^K/K to n^μ for $\mu = \mu_K = K^{-1/\alpha}$, this yields the relation of $\beta_v^K(t) \approx 1 - \rho_v^{\mu_K}(\alpha t)/\alpha$.

In the scenarios of regime 3 and the limit in the deterministic system, it is necessary to know β^K or ρ^μ (or better their limit) to construct the limiting adaptive walk. However, they only reflect the sets of resident traits (as traits with $\beta^K = 1$ or $\rho^\mu = 0$) and do not quantify the equilibrium states. The full information is therefore only captured in a combination of the limits of these two quantities, N^K/K and β^K or n^μ and ρ^μ, respectively.

1.4 Immunotherapy of cancer

In this section we give a brief summary of the biology of cancer, treatment strategies, the immune system, and immunotherapy of cancer for a reader with less biomedical background. For more details we refer to classic text books for internal medicine [101] and immunobiology [143]. Cancer appears in many different forms and is one of the deadliest diseases for humanity. There are many different approaches to treat cancer and just as many resistance mechanisms that suppress the effectiveness of therapy. Most of the underlying effects are not yet well understood. Therefore, we argue that mathematical modelling is a beneficial tool in studying the dynamics of the tumour tissue, but also other complex biological systems. We present a number of different strategies for mathematical modelling and simulations of stochastic processes, and reason why stochastic individual-based models of evolution under selective pressure are suitable for studying the particular scenario of melanomas under immunotherapy.

1.4.1 Cancer and treatment strategies

Tumours in general are masses of abnormally growing cell tissue (neoplasms). There are two major classes of tumours, namely benign and malignant tumours. The former do not invade neighbouring tissue or metastasise, i.e. cells of the primary tumour spread throughout the body, and they usually do not grow back when removed surgically. Malignant tumours, also

called cancer, show a much higher growth rate than benign tumours. They often metastasise and invade organs and tissues of the body, thus impairing their function and causing major symptoms or even death [91]. There are many different types of cancers. Some of the major causes are smoking, unhealthy diets, infections, excessive radiation, inherited genetic defects, or simply pure chance. They induce a change in the genome of tissue cells, resulting in increased cell growth. As opposed to their benign form, malignant tumours display a variety of escape mechanisms from treatment and often relapse after an initial remission.

For a long time, there have been three major approaches for cancer therapy: surgery, radiation, and chemotherapy. *Surgery* can remove solid tumours that are isolated from the surrounding tissue. It has the advantage that it allows for a biopsy, i.e. an analysis of the removed tissue. However, surgery cannot easily be applied to tumours that have invaded important organ tissue or metastasised throughout the body. *Radiation* therapy targets the tumour tissue with concentrated X-ray beams to damage the DNA of the cells and thus induce cell death. It can be applied to cancer tissue that is hard to reach by surgery but always damages the neighbouring tissue in the process. *Chemotherapy* is a drug-based treatment that targets rapidly dividing cells all over the body. It is directed at the cell division program and induces cytotoxic functions. A drawback of chemotherapy is its toxicity to other tissues of the body which often limits its effectiveness. Often, two or more of these three methods are combined. For example the bulk of the primary tumour is removed surgically and residual tissue and metastases are treated with chemotherapy.

All of the approaches above are fairly undirected and do not only harm the tumour tissue but also the surrounding healthy body tissue. Therefore, it is of great interest to develop so-called *targeted therapies* that specifically treat the cancerous cells. A major challenge in this context is the identification of suitable molecular targets that are characteristic for the tumour tissue alone. This is particularly difficult as tumours are usually highly heterogeneous and do not consist of a single cell type.

There exists a variety of targeted approaches that use *small molecules* to target surface receptors or intracellular proteins of cancer cells. The small molecules bind to their target and thus impair its function, blocking tumour cell growth and metastasis or inducing apoptosis. In this thesis we focus on a new method that has come into use since the 1990s, namely immunotherapy of cancer. It utilises the bodies own defence mechanisms to fight the cancerous cells.

1.4.2 Immune system and immunotherapy

The immune system consists of many biological structures and mechanisms that defend the body against disease. To do so, it needs to detect a variety of pathogens, such as harmful bacteria, viruses, and mutated cells. In this process it not only needs to differentiate between self and non-self structures, but also between healthy and diseased tissues. The immune system in vertebrates is categorised into two major components: The innate and the adaptive immune system [128].

The *innate immune system* is evolutionary much older and already found in plants, insects and other multicellular organisms. It targets pathogens in an unspecific way. The innate

immune system is comprised of physical barriers such as the skin, the complement system, inflammatory cytokines, and leukocytes (white blood cells) such as phagocytes, neutrophiles, and natural killer cells. Apart from an assortment of inflammatory cytokines, in particular TNF-α and IFN-γ, and some effects that are mediated by neutrophiles, we do not study the innate immune response in this thesis.

The *adaptive immune system* is only present in vertebrates and develops within the life-time of an individual, through confrontation with different pathogens. Therefore, it is often referred to as the acquired immune system. It is responsible for a specific pathogen recognition and suppression, as well as the establishment of memory. There are two important types of adaptive immune cells, namely B- and T-lymphocytes. They are part of the leukocytes and are comprised of different subgroups of cells themselves. B-lymphocytes, or B-cells for short, produce antibodies that circulate the blood and lymph systems and are part of the humoral response. They can specifically bind to certain peptides (antigens) and are responsible for agglutination and neutralisation of pathogens, for activation of the complement system, and enhancement of phagocytosis and innate cytotoxicity.

The focus of this thesis lies on the T-lymphocytes, more specifically on cytotoxic T-cells. With their T-cell receptors (TCRs), which resemble antibodies bound to the cell surface, they specifically recognise certain peptides bound to MHC molecules. Upon activation through presentation of an antigen in the lymph node, naive T-cells are activated and proliferate. Directed by chemokine gradients, they migrate to the infected tissue. When exposed to the specific antigen, they secrete cytotoxins that perforate the cell membrane of the dysfunctional or infected cell and induce apoptosis (cell death).

There are many different methods to improve the immune system's ability to fight cancer [46]. Besides a treatment with certain cytokines, which is mentioned in the context of chemotherapy above, three major approaches of immunotherapy are cancer vaccines, checkpoint inhibition, and adoptive cell transfer (ACT). Cancer vaccines can be used both as prevention and as treatment. Tumour specific antigens are injected into the patient to boost the immune systems activity and thus either form a memory for future diseases or enhance the response to acute cancer.

The immune system possesses a number of negative feedback mechanisms that regulate and inhibit excessive immune responses, preventing for example autoimmunity. Tumour cells can evade immunosurveillance, amongst others, by up-regulation of negative immune checkpoints that suppress antitumour responses. Immune checkpoint therapy is part of so-called passive therapy approaches that aim to enhance the existing tumour-specific immune responses by blocking these inhibitory immune checkpoints [135]. For example, binding of programmed cell death (PD-1) receptors expressed on T-cells to their ligand (PD-L1) expressed on cancer cells induces an inhibitory signal, limiting T-cell efficiency. Monoclonal antibodies, such as aPD-L1, are used to block this interaction and maintain T-cell responses. For their discovery of cancer therapy through inhibition of negative immune regulation by PD-1 and another receptor CTLA-4, Tasuku Honjo and James P. Allison received the Nobel Prize in Physiology or Medicine in 2018.

ACT therapy is a form of active immunotherapy. Tumour specific autologous T-cells are extracted from the patient, activated and clonally expanded *in vitro*, and re-infused into the

patient. Chimeric Antigen Receptor (CAR) T-cell therapy represents a further development in which autologous T-cells are genetically modified to express a CAR specific for a tumour antigen. This procedure is able to enhance an otherwise sub-critical immune response.

In this thesis we focus on malignant melanomas, which are an aggressive type of skin cancer and known to grow quickly and metastasise to vital organs such as the lung, liver, and brain. We study ACT therapy with cytotoxic T-cells that specifically target a melanocytic antigen, combined with injections of a small molecule inhibitor targeting the HGF/c-MET signalling pathway (METi) [117, 87].

Besides an immunosuppressive environment, other resistance mechanisms of malignant tumours against targeted immunotherapy have been observed. Most prominently, genotypic and phenotypic plasticity of the cells comprising the tumour is a driving force for resistance [129]. Due to their high reproduction rate, cancer cell populations are prone to harbour pre-existing and acquired mutations that cause great genotypic heterogeneity among the cells [148]. On top, a changing environment, for example in an inflammatory milieu during treatment, can result in phenotypic changes in the cells. This can for example cause reversible dedifferentiation and down-regulation of antigen expression [44]. Both mechanisms can lead to resistance by abrogating T-cell recognition. Under the increased selective pressure during treatment, cell variants with pre-existing or acquired genetic and non-genetic aberrations have a fitness advantage and are enriched in the tumour tissue. This can induce a relapse of the tumour, even after initial remission [86, 97].

Landsberg et al. [117] and Baar et al. [11] have studied inflammation-induced phenotypic dedifferentiation of melanoma cells (down-regulation of the melanocytic antigen) as a resistance mechanism to ACT therapy. Small molecule inhibitors (METi) partially prevent this mechanisms [87]. In this thesis and in [88], we compare these phenotypic switches to genetic antigen loss as mechanisms to escape a combined $\mathrm{ACT}^{\mathrm{METi}}$ therapy.

1.4.3 Mathematical modelling in life sciences

The immune system, like many other biological systems, is a very complicated network of many different cell and molecule types that interact in multiple ways to regulate the body's response to disease. Modern experimental techniques make it possible to study many aspects of these kind of complex systems. However, the options to manipulate biological processes in living creatures or even full ecosystems are limited. Experiments under laboratory conditions, like *in vitro* experiments in cell cultures, often over-simplify and do not reflect the actual situation.

In these cases it is often helpful to employ mathematical modelling to further study the dynamics of the system. In contrast to many experimental models, mathematical models and simulations are less time and money consuming and easily manipulable. Even though theoretical models may not always be able to make quantitative predictions and can of course not replace clinical studies, they can still be beneficial. Through modelling, one can test hypotheses, identify important mechanisms, i.e. those that determine the outcome of an experiment, and predict general trends for scenarios beyond the experiments. Thus, mathematical

modelling can guide experimental research towards promising ideas, validate experimental results, and reduce the amount of necessary long-term and animal experiments.

The general approach of mathematical modelling is to start out from some input data, and most of the time some hypotheses on which players (cell types, molecules, organisms, etc.) are involved and how they interact. From this information, a model is built. It is validated by comparing simulation results, which are generated based on this model, with experimental data. Optimally, this validation data should be different to the input data. If the simulation and experimental results coincide, it is a strong indication, although not a proof, that the underlying hypothesis is true.

Depending on the pre-existing knowledge and amount and quality of available data, there are different approaches to building a suitable mathematical model. If the data is high in quantity and quality, it makes sense to apply statistical and computational methods to choose the best model out of a pool of hypotheses and determine the parameters to fit the experimental data. In [50] Costa and coauthors present a Bayesian approach for model selection and parameter estimation in tumour growth models. Recently, Fröhlich et al. have introduced a computational framework for the parameterisation of large-scale mechanistic models. They apply these methods to a large network of cancer related signalling pathways to predict responses to combined drug treatments [80]. Computational methods like this allow to integrate large sets of diverse data to investigate complex biological systems. Particularly in the parametrisation of large systems it is however important to conduct an uncertainly analysis. This determines the reliability of model predictions, accounting for various sources of uncertainty in model input and design.

Particularly if the available data is not that extensive, statistical methods can give a false confidence into the attained model and the predictions generated by simulations. In those cases, it is often useful to employ a more simplistic modelling approach. Here, the goal is to keep the number of different cell or molecule types and interactions or mechanisms that are incorporated in the model as low as possible. This produces low dimensional models that allow for a more theoretical analysis of the dynamics. Such an approach aims more towards a general understanding of the structure of the system than an exact quantitative fitting. It looks for common phenomena that arise across a wider set of parameter choices. In case the simulations do not fit the experimental data, one retains a basic understanding of causalities within the mathematical model and can suggest which mechanisms might be depicted in a wrong way.

As an example, Kuznetsoz and coauthors have proposed a simple ODE model for the interaction between tumour and immune cells and did parameter estimations [116]. A review on more of such models can be found in [67]. In [1], Altrock, Liu, and Michor review a number of stochastic and deterministic models for different aspects of cancer, like tumour initiation, progression, metastasis, and treatment resistance. Kimmel et al. consider a hybrid model that combines a deterministic ODE model with stochastic birth-and-death processes for small tumour cell populations to study CAR T-cell therapy of B-cell lymphomas [106].

On a more theoretical level, with no connection to specific data, Mayer and Bovier study the activation of T-Cells as a statistical test problem, using large deviation techniques [130]. In [64, 76], Foo, Leder, and coauthors investigate tumour genesis, where several mutations

have to be acquired to gain a fitness advantage, in a spatial setting similar to a voter model. Gunnarsson et al. consider multitype branching processes to study the stabilisation of reversible phenotypic switches that lead to drug-resistance, for various treatment approaches [89].

With the experiments on ACT therapy of melanoma in mouse models from Landsberg et al. and Glodde et al. [117, 87, 88], we are in the situation of relatively sparse data. The available measurements give information about the variation of the total number of cells (comprised of many different cell types) over time and the genetic composition of the tumour at inoculation and harvesting. Therefore, we employ a rather simple model that only involves the most important cell or molecule types and mechanisms. Among other reasons, the study of spontaneously occurring mutations makes it necessary to consider a stochastic model. Since we consider evolutionary dynamics within a growing tumour tissue that has not yet reached an equilibrium size, the model should not be restricted to a fixed population size and depict the competitive interaction between different cell types. All of these reasons make an extension of the previously mentioned class of individual-based Markov processes a suitable choice of model. In this thesis we extend the model of Baar at al. [11] that is used to study the experiments in [117]. Parameter estimation for this model by an SAEM algorithm is proposed by Diabaté et al. in [58]. More details are given in Section 1.6 and Chapter 4.

There are a couple of different approaches to simulate sample paths of such generalised individual-based Markov processes, where a number of different events or reactions occur at exponential rates, depending on some parameters and the current state of the population. A stochastic simulation algorithm to produce an exact realisation of such a Markov process was introduced by Gillespie in the context of chemical reactions [83]. There are various ways to improve this algorithm, e.g. by reducing the number of required random variables. For example the next reaction method [82] reuses random variables by rescaling, which is further improved by efficient binning of events in [160]. Both approaches are particularly suited for systems with many possible events with rates that only depend on few cell or molecule types. However, all of these algorithms separately generate single events. Particularly in large populations, where there are many frequently occurring events like proliferation and death, they are computationally heavy and need many iterations to simulate the evolution over a time span that is of interest.

To overcome this problem, there are different procedures to approximate the number of occurrences of a certain event within some time interval. The simplest one is to consider the corresponding deterministic system (according to the large population approximation as described above). This way, the mean dynamics of the system can be simulated with classical numerical techniques for solving differential equations, as for example Runge-Kutta methods. The shortcoming of this approach is that random effects are no longer included. Therefore, the idea of hybrid algorithms is to combine deterministic and stochastic methods in some way to speed up simulation but also maintain stochastic fluctuations. To do so, either subpopulations are sorted by large and small population sizes or events are sorted by frequent and rare occurrence, i.e. the size of their event rate. The evolution of large populations or frequent events is handled deterministically, according to the mean dynamics, while small populations or rare events are treated stochastically. There are also approaches

that introduce an intermediate step of stochastic diffusion approximations. Examples for approximative hybrid simulations of individual-based models can be found in [159, 126, 62].

Another approach is the one of so-called tau-leaping, where the number of events within a certain time interval is generated as a Poisson random variable [84]. The approximation in this case is that the rates of the different events are assumed to be constant for this time interval. Therefore, the length of the interval is dynamically chosen based on the sensitivity of the rates, i.e. on how much the rates vary with a changing population state. For linear, quadratic, and cubic rates, there is a good theory in place to choose the interval length [30]. However, general rates are not treated yet.

In the simulations in this thesis, we apply a hybrid algorithm that combines deterministic Runge-Kutta methods and a stochastic Gillespie algorithm, differentiating between frequent and rare events. This is justified by applying the large population approximation result from [70]. More details on this are given in Chapter 4.

1.5 Outlook and open questions

There are multiple open questions related to the work in this thesis. They are promising for future research, both of more theoretical nature and related to applications. In the following we present a few of those topics.

As mentioned above, Bovier, Coquille, and Smadi introduce a fourth time scale for the study of individual-based models with power law mutation rates in [25]. On this time scale the limiting object is no longer fully deterministic but metastability phenomena can arise with random jumps between local maxima of the fitness landscape (with respect to invasion fitness). We heuristically study these phenomena for more general trait spaces in Chapter 3. A promising goal for future work is to fully characterise the resulting limit for general finite trait spaces on this faster time scale.

Moreover, so far we only characterise the limit as long as there is a unique stable equilibrium to the Lotka-Volterra system involving the resident traits and the new mutant. It is of great interest to relax these assumptions and study for example the behaviour of new mutants in a regularly oscillating population, i.e. when the Lotka-Volterra system shows limit cycles. This ties in with a scenario arising in tumour immunotherapy. The occurrence and fixation of spontaneous mutations in a small tumour of oscillating melanoma cell and T-cell populations is of interest as it has potential to cause late relapses of previously shrinking tumours.

A similar idea is to study a system with time dependent parameters. It is reasonable to consider parameters that change with respect to the system time t, corresponding to a changing environment. Here, it is again a question of time scales. It is important how fast the changes of the environment are with respect to the time scale on which mutations occur, populations grow, and the Lotka-Volterra dynamics change the resident population. A quickly fluctuating environment, as for example seasons that change multiple times during the life time of an individual, lead to a mean field behaviour. Conversely, slow changes, like global hot and cold phases of the span of hundreds of years, allow the population to adapt to the changing environment. In the latter scenario, it is then a question of whether

subpopulations that are unfit with respect to one parameter set survive long enough to reach a phase where they posses a positive fitness. In [49], Cvijović and coauthors study the fixation of a mutation in a population of constant size, depending on the lengths of beneficial and deleterious epochs as well as the size of fitness change during these phases.

Motivated by medical applications, the study of parameters that change with respect to the age of an individual is particularly interesting. As we observe that T-cell dynamics play an important role in determining the outcome of immunotherapy, we want to gain a better understanding of T-cell activation, inhibition, and exhaustion. The latter is a phenomenon of ongoing discussion in biomedical research, see [17] where a large number of researchers present their viewpoint on the topic. Although it is not settled how and why exactly exhaustion occurs, it is observed that effector T-cells loose their function over time, for example in chronic infections. With this perspective, the question ties in nicely with the topic of age-dependent parameters. Here, the effects of T-cell exhaustion can be modelled independently of the underlying mechanism as a decreased efficiency with age. This is different to the approach of parameters depending on the system time since new unexhausted T-cells can be recruited to the tissue. To reduce the complexity of the mathematical model, in a first step one should study the dynamics that determine the T-cell activity in a more simple context. This could be the scenario of chronic infections, where intricate pheno- and genotype dynamics, like the ones of melanoma cells, do not play a role.

In addition to T-cell activity, we see that spatial aspects of the tumour are important and need to be incorporated more into the model, which currently assumes a well-mixed population. One possibility is to consider a partially spatial model that introduces several compartments that differentiate between inner and outer layers of the tumour. It is promising to combine this with a change in parameters, where the outside of the tumour is easily reachable by T-cells but the core shows a hypoxic and immunosuppressive environment.

Computation wise, we observe that stochastic fluctuations play an important role not only in small populations but in general at "critical points" (e.g. bifurcations) of the deterministic system. In the future, we want to improve how to dynamically identify and deal with these interior critical points and how to choose the threshold for deterministic approximations in the hybrid algorithm. Another promising way to go is a tau-leaping approach to bundle multiple events, as long as the rates don't change too much and can be approximated as constant for a certain time interval [84, 30]. However, also with this method, there remains work to be done to derive an appropriate choice of time steps in the case of rates that are more complicated than linear, quadratic, or cubic functions of the sizes of subpopulations.

1.6 Outline and main results of the thesis

The main part of this thesis is divided into three chapters. All chapters can be read independently, however, their contents are related. Chapter 2 and 3 are more theoretic, where the former introduces some of the approximation techniques that are used in the latter. Chapter 4 presents an application of the type of model that is analysed in the first two chapters in the context of oncological research. In the following we summarise the content and main results of the three chapters.

1.6.1 Rare mutations in competitive Lotka-Volterra systems with mutation

Chapter 2 studies the deterministic system (1.4) in the limit of rare mutations, i.e. as the mutation probability μ tends to zero. We derive an algorithmic description of the limiting deterministic jump process, which is simplified in the case of equal competition. Finally, we consider a cut-off model, where proliferation is only possible above a certain threshold population size. This chapter was published in the Journal of Mathematical Biology as joint work with Anton Bovier [112],

> A. Kraut and A. Bovier, *From adaptive dynamics to adaptive walks.* Journal of Mathematical Biology, Volume 79 , Number 5, pp. 1699–1747, 2019.

Chapter 2 contains the published version, with only minor changes to correct some typing errors and to adapt the layout to the format of this thesis.

In the main result of Chapter 2 we derive a full description of the limiting jump process of (1.4), as $\mu \to 0$, see Theorem 2.12. As discussed above, we rescale the time as $t \log 1/\mu$. This is the time scale on which microscopic mutant populations can grow. The same limit is also considered in [26, 25], but in a much more restrictive setting. There, the trait graph is a subset of $\mathbb{N}$ with nearest neighbour mutations and very specific assumptions on the (invasion) fitness landscape. We consider the n-dimensional hypercube $\mathbb{H}^n = \{0,1\}^n$, due to its nice interpretation as sequences of (in)active genes and since the length of a shortest path between two traits $v, w \in \mathbb{H}^n$ is exactly the 1-norm $\|v - w\|_1$. However, the results can easily be transferred to any finite, possibly directed graph.

The main difficulties of the proof, compared to previous results like [26], come from two sources. First, we allow for coexistence of multiple resident traits. This leads to more complicated calculations for the stability of the resident equilibrium state during the growth phase of the mutants. Instead of approximating the derivative of the population size of one resident trait from above and below, we prove a contraction close to the equilibrium state. To do so, we compare the Euclidean norm and a norm that is related to the positive definiteness of the competition kernel. The latter is an assumption that we make to also guarantee the existence of unique stable equilibria according to [35].

The second complication stems from the fact that we consider a trait graph with circles. As a consequence, two traits are connected by multiple different paths and mutant traits can arise due to different chains of consecutive single mutations. Since we consider a regime of very large mutation rates, this means that each subpopulation has a continuous influx of mutants from different sources. To approximate the growth rate of a population of a certain trait or better its population size, we carefully take into account all these different influences. Here, a population can grow according to its own fitness or due to incoming mutants from a neighbouring trait. To derive the correct rates or sizes, we introduce a particular induction procedure. For all traits in parallel, we take into account the incoming mutants arising due to an increasing number of mutations, i.e. stemming from traits of increasing distance to the considered trait. This is a fundamentally different approach to the previous papers, where mutation was only going in one direction and one could therefore approximate the population sizes trait by trait, already knowing the size of the direct neighbour.

The first mutant trait to reach a macroscopic population size invades the resident population according to the Lotka-Volterra dynamics. It is determined by optimising the quotient of initial population size (as exponent of μ) and invasion fitness of the different traits. This is due to the fact that traits reaching a population size of order 1 cannot have gotten there purely due to incoming mutations (those result at most in a population size of order μ), but only due to their own growth with the rate of their invasion fitness. On the time scale $t \log 1/\mu$, it then takes a time of order $t = \gamma/f_{w,\mathbf{v}}$ to grow from μ^γ to μ^0 at rate $f_{w,\mathbf{v}}$. Consequently, minimising this quotient not only gives the next trait to invade the population but also the time step between invasions. The exact bookkeeping of the population sizes of all other traits is necessary to determine the initial conditions for the next invasion step. This is important because the fitness landscape changes according to the new resident population and formerly unfit traits might become fit.

In Theorem 2.14, we present a simplified characterisation of the limiting process for the case of constant competition $c_{v,w} \equiv c$. In this case, a trait that is once unfit stays unfit indefinitely. As a result, we do not need to keep track of the population sizes of all microscopic populations, which simplifies the algorithmic description. Note that this is not a direct corollary from Theorem 2.12 since the assumption on the positive definiteness of the competition kernel can no longer be satisfied. However, we assume that the individual fitness r is different for each trait, which prevents coexistence and allows us to argue differently in the few places where the positive definiteness is used.

Finally, we propose a cut-off version of the deterministic system (1.4), where proliferation is only possible above a certain threshold for population size, depending on μ. This model is intended to mimic the simultaneous limit of large populations and rare mutations, where reproduction is only possible in populations with at least one individual. Depending on the choice of the threshold, populations of size μ^ℓ are able to reproduce, which corresponds to $K\mu_K^\ell$ being of order at least one and hence $\mu_K \geq K^{-1/\ell}$. In the case of $\ell = 1$, we again characterise the limiting process in Theorem 2.18. Moreover, we give some results on the accessibility of traits in relation to fitness valleys.

The limiting processes of the deterministic system, corresponding to the different scenarios, behave similar to adaptive walks or flights in the sense of [132, 102, 144, 147]. In the case of Theorem 2.12, the fitness landscape is changing after each invasion. The limiting jump process can take arbitrary large steps and reaches a final state only if there exists a set of coexisting resident traits, for which every invasion fitness is non-positive. In the cases of constant competition, the fitness landscape defined by the individual fitnesses r_v corresponds to the setting of adaptive walks. For non-restricted reproduction, as in Theorem 2.14, the limiting process moves along traits of increasing fitness r and eventually reaches the global maximum, where it remains. In Theorem 2.18, we only allow for nearest neighbour jumps. Thus, the process always jumps to the fittest neighbour of a resident trait, which corresponds to the greedy adaptive walk in [147].

Note that in Chapter 2 we employ a slightly different notation. We consider the trait space $\mathcal{X} = \mathbb{H}^n$ with elements x, y, z. The rescaled stochastic process is denoted by $\nu_t^{\mu,K}$ instead of $N^K(t)/K$ and the solution of the Lotka-Volterra system with mutation is denoted by ξ_t^μ instead of $n^\mu(t)$. Finally, the parameters for competition are $\alpha(x, y)$ instead of $c_{v,w}$.

1.6.2 Simultaneous large population-rare mutation limit for moderate power law mutation

In Chapter 3 we study the simultaneous limit of large populations and rare mutations in the individual-based model (1.3), i.e. $K \to \infty$ and $\mu_K \to 0$. For finite trait spaces and on the time scale $\log K$, we consider moderate mutation probabilities that decay like $K^{-1/\alpha}$, $\alpha > 0$. This corresponds to regime 3, or regime 1 in the special case of $\alpha < 1$. We derive a full description of the limiting jump process, thus extending the results from Chapter 2 to the simultaneous limit and the results from [25] to a more general trait space. Moreover, we present multiple specific examples of interesting evolutionary scenarios that occur for this choice of mutation rate. This chapter is available as a preprint as joint work with Loren Coquille and Charline Smadi [42],

> L. Coquille, A. Kraut, and C. Smadi, *Stochastic individual-based models with power law mutation rates on a general finite trait space.* Preprint, arXiv:2003.03452, 2020.

Chapter 3 contains the preprint, with only minor changes to correct some typing errors and adapt the layout to the format of this thesis.

The content of Chapter 3 is divided into two parts. In the first part, in Theorem 3.3 and Proposition 3.6, we give a full characterisation of the limiting process, both in terms of the exponents β_v^K, as defined in (1.8), and the rescaled population size N^K/K. This result applies to general finite, possibly directed graphs and provides an algorithmic construction of the limit, as long as there exist unique stable equilibria to the Lotka-Volterra dynamics for the resident and invading mutant traits. For α larger than the longest distance within the trait graph (in terms of shortest path length), the limiting process coincides with the one in Theorem 2.12. For smaller α, the description is more intricate since not all mutant traits are present in the beginning. Therefore, we have to introduce intermediate time steps (between invasions) when new subpopulations arise due to mutation from other growing subpopulations.

Both results are proved simultaneously and the proof relies on the induction approach from Chapter 2, as well as an approximation by (logistic) birth-and-death processes with and without migration. We make extensive use of the limit results for such processes from Champagnat, Méléard, and Tran's paper [38]. However, we have to generalise some of their results, for example to deal with coexisting resident traits.

In the second part we consider several specific graphs and parameter choices to demonstrate interesting and partially counter-intuitive behaviours that arise in the limiting process under this particular scaling of mutation rates. To name some examples, mutational paths can be longer or shorter than expected, taking seemingly unnecessary detours or skipping traits on the graph. Adding an edge in the graph, i.e. a possibility for mutation, can prolong the time to reach a fitness maximum. Moreover, the process can take arbitrarily large steps, in particular farther than the radius α, in which a resident trait can produce mutants. In this regime, the limiting process can get stuck in a local maximum of the (invasion) fitness landscape, surrounded by unfit traits in a radius of α.

Moreover, we consider two cases where metastable stochastic behaviour arises in the otherwise deterministic limit, when considering an even more accelerated time scale in the spirit of [25]. In those cases, we see an effective random walk that jumps between clusters of traits, on which a metastable equilibrium is reached according to the deterministic dynamics. These clusters are separated by fitness valleys wider than α. In Chapter 3 we give these metastability results without proof but provide a heuristic derivation.

1.6.3 Modelling of genetic variation as an escape mechanisms from cancer immunotherapy

In Chapter 4 we present an extension of the individual-based Markov process in (1.3) that models the evolution of tumour cell populations in melanomas under ACT immunotherapy with cytotoxic T-cells. We identify several important mechanisms that are essential to reproducing the experimental data and derive likely causes for high variation witnessed in the measurements of tumour composition. Moreover, we validate the clinical relevance of the experimental results by going beyond the experimental setup and simulating scenarios of spontaneously occurring mutations. Larger parts of this chapter are available as a preprint as joint work with Nicole Glodde, Thomas Tüting, Anton Bovier, Michael Hölzel et al. [88],

> N. Glodde, A. Kraut, D. van den Boorn-Konijnenberg, S. Vadder, F. Kreten, J. Schmid-Burgk, P. Aymans, K. Echelmeyer, M. Rumpf, J. Landsberg, T. Bald, T. Tüting, A. Bovier, and M. Hölzel, *Experimental and stochastic models of melanoma T-cell therapy define impact of subclone fitness on selection of antigen loss variants.* Preprint, bioRxiv:10.1101/860023, 2019.

This is an interdisciplinary project with experimentalists from the Faculty of Medicine. For Chapter 4, the preprint is adapted to a purely mathematical audience, presenting the experimental results in a more condensed way and giving more details on the mathematical model and the simulation algorithm. Moreover, a couple of simulation results that were not part of the preprint, but are essential to the derivation of the model or present interesting mathematical scenarios, are included. More details on this are given in Chapter 4.

In the beginning of Chapter 4 we briefly summarise the medical background, setup, and results of the mouse experiments done by our collaborators in the group of Prof. Michael Hölzel at the Institute for Experimental Oncology. The two most important observations can be summed up as follows: First, a combination of the ACT immunotherapy protocol presented in [117] with small molecule inhibitors (METi) results in reduced tumour growth, less cytokine-induced dedifferentiation (down-regulation of presentation of the melanocytic antigen), and increased T-cell efficiency [87]. Second, experiments with genetic antigen loss variants (KO) show a reduced fitness with respect to the wild type (WT) melanoma cells in an untreated scenario. However, under ACT^{METi} therapy, the KO cells are enriched since they are not recognised by the T-cells due to the missing presentation of melanocytic antigen. Surprisingly, the enrichment of KO cells was highly variable between individual mice and a notable amount of WT cells survived, i.e. the KO cells did not fully invade the tumour tissue. This high variability is one of the main motivations to apply mathematical

modelling to better understand the evolutionary dynamics and interactions between different melanoma cell types.

In the second part of Chapter 4, we introduce the mathematical model used to simulate the tumour cell dynamics. We give a description of the stochastic process and and discuss the derivation of the parameters. Moreover, this section includes a rigorous justification of the applicability of the deterministic limit result from [70]. This justifies an approximative hybrid algorithm that we describe along with the corresponding pseudo code.

The model is an individual-based Markov process that is an extension of the model for ACT therapy introduced in [11]. In their work, Baar et al. model two types of melanoma cells (differentiated and dedifferentiated) that interact according to the competitive dynamics in (1.3), with the difference that mutations are substituted by phenotypic switches between the two states. These switches are more frequent than mutations and reach an equilibrium after a short time. In addition to the melanoma cells, the authors consider T-cells, which interact with the differentiated cells according to standard predator-prey dynamics, and TNF-α cytokines that are secreted by active T-cells and induce additional dedifferentiation. This model is now extended by adding the effects of METi injections and the additional cell type of KO melanoma cells. Moreover, the rates for T-cell proliferation and differentiated WT cell killing are more intricate as they now include the effects of cytokine-mediated T-cell inhibition and spatial shielding of WT cells from T-cells by KO melanoma cells. The model is characterised by its generator that takes the form of

$$\mathcal{L}\phi(\nu) = \sum_{e \in E} (\phi(\nu + \nu_e) - \phi(\nu)) r_e(\nu), \tag{1.10}$$

where E is the set of possible events, like T-cell proliferation, dedifferentiation of WT cells, or mutation to a KO cell at WT cell reproduction. ν_e is the change in the population at such an event and $r_e(\nu)$ is the exponential rate at which the event occurs.

The third part of Chapter 4 provides the results of the simulations. We comment on the necessity of including T-cell inhibition into the model and discuss a critical threshold for the tumour size at treatment onset that determines the therapy success. This threshold represents a critical point in the corresponding deterministic system. It further underlines the need to consider a stochastic model, as the course of therapy close to this threshold depends on random fluctuations. We analyse the changing fitness of KO melanoma cells and the shielding effect that explains the remaining portion of WT cells under therapy. The most likely causes for the high variability in measured KO cell enrichment are identified as a varying time point of harvesting and, most importantly, a subclonal fitness variability. The latter is confirmed by experiments. Finally, we study the spontaneous occurrence of KO mutations, validating that effects similar to the ones witnessed in experiments with artificially introduced KO cells also arise in clinically relevant scenarios.

In the last part, we discuss the clinical relevance and implications of these results. We argue that competitive interactions and evolutionary dynamics within the melanoma cell population have to be taken into account when analysing heterogeneous tumour samples. Moreover, we propose that immunotherapy is best applied to small tumours and could be enhanced through a combination with apoptosis-inducing drugs.

2 From Adaptive Dynamics to Adaptive Walks

We consider an asexually reproducing population on a finite type space whose evolution is driven by exponential birth, death and competition rates, as well as the possibility of mutation at a birth event. On the individual-based level this population can be modelled as a measure-valued Markov process. Multiple variations of this system have been studied in the simultaneous limit of large populations and rare mutations, where the regime is chosen such that mutations are separated. We consider the deterministic system, resulting from the large population limit, and then let the mutation probability tend to zero. This corresponds to a much higher frequency of mutations, where multiple microscopic types are present at the same time. The limiting process resembles an adaptive walk or flight and jumps between different equilibria of coexisting types. The graph structure on the type space, determined by the possibilities to mutate, plays an important role in defining this jump process. In a variation of the above model, where the radius in which mutants can be spread is limited, we study the possibility of crossing valleys in the fitness landscape and derive different kinds of limiting walks.

2.1 Introduction

The concept of *adaptive dynamics* is a heuristic biological theory for the evolution of a population made up of different types that has been developed in the 1990s, see [139, 59, 20, 21, 60]. It assumes asexual, clonal reproduction with the possibility of mutation. These mutations are rare and new types can initially be neglected, but selection acts fast and the population is assumed to always be at equilibrium. This implies a separation of the fast ecological and slow evolutionary time scale. Fixation or extinction of a mutant are determined by its *invasion fitness* that describes its exponential growth rate in a population at equilibrium. This notion of fitness is dependent on the current resident population and therefore changes over time. The equilibria do not need to be monomorphic and allow for coexistence and evolutionary branching. Eventually, so-called *evolutionary stable states* can be reached, where all possible mutants have negative invasion fitness and therefore the state of the population is final.

A special case of adaptive dynamics are so-called *adaptive walks* or *adaptive flights*. The concept of adaptive walks was introduced by Maynard Smith [131, 132] and further developed by Kauffmann, Levin, and Orr [102, 103, 150]. Here, evolution is modelled as a random walk on the type space that moves towards higher fitness as the population adapts to its environment. More precisely, a discrete state space is equipped with a graph structure

that marks the possibility of mutation between neighbours. A fixed, but possibly random, fitness landscape is imposed on the type space. In contrast to the above, this *individual fitness* is not dependent on the current state of the population. Adaptive walks move along neighbours of increasing fitness, according to some transition law, towards a local or global optimum. Adaptive flights, a term that has been introduced by Neidhart and Krug [144], can take larger steps and jump between local fitness maxima to eventually attain a global maximum. Quantities of interest are, among others, the typical length of an adaptive walk before reaching a local fitness maximum and the distribution of maxima, see [147], as well as the number of accessible paths, see [161, 13, 14]. They have been studied under various assumptions on the correlations of the fitness landscape and the transition law of the walk. Examples, mentioned by Nowak and Krug [147], are the *natural adaptive walk*, where the transition probabilities are proportional to the increase in fitness, or the *greedy adaptive walk*, which always jumps to the fittest available neighbour.

Over the last years, stochastic individual-based models have been introduced to study different aspects of evolution. They start out with a model that considers a collection of individuals. Each individual is characterised by a type, for example its genotype. The population evolves in time under the mechanisms of birth, death, and mutation, where the parameters depend on the types. The population size is not fixed but the resources of the environment, represented by the *carrying capacity* K, are limited. This results in a competitive interaction between the individuals, which limits the population size to the order of K. The dynamics are modelled as a continuous time Markov process, as shown by Fournier and Méléard [77]. It is of particular interest to study the convergence of this process in the limits of large populations, rare mutations, and small mutation steps.

For a finite type space, Ethier and Kurtz [70] have shown that, rescaling the population by K, the process converges to the deterministic solution of a system of differential equations in the limit of large populations, i.e. as K tends to infinity. The differential equations are of Lotka-Volterra type with additional terms for the effects of mutation. This result was generalised for types in $\mathbb{R}^d$ in [77]. For finite times, in the limit of rare mutations, this deterministic system converges to the corresponding mutation-free Lotka-Volterra system. Under certain conditions, these converge in time to unique equilibrium configurations, see [96, 35].

Champagnat, Ferrière, Méléard and others have considered the simultaneous limit of large populations and rare mutations [31, 33, 37]. Here, the mutation probability μ_K tends to zero as K tends to infinity. They make strong assumptions on the scaling of μ_K, where only very small mutation probabilities $\mu_K \ll 1/(K \log K)$ are considered. This ensures the separation of different mutation events. With high probability, a mutant either dies out or fixates in the resident population before the next mutation occurs. To balance the rare mutations, time is rescaled by $1/(K\mu_K)$, which corresponds to the average time until a mutation occurs. The limiting process is a Markov jump process called *trait substitution sequence* (TSS) or *polymorphic evolution sequence* (PES), depending on whether the population stays monomorphic or branches into several coexisting types. In the framework of adaptive walks, these sequences correspond to the natural walk, mentioned above.

Similar convergence results have been shown for many variations of the original individual-based model under the same scaling, including small mutational effects, fast phenotypic

switches, spatial aspects, and also diploid organisms, see, e.g. [10, 9, 36, 169, 119, 41, 146, 24].

The drawback of all these results is the strong assumption on the mutation rate. The separation of mutations which results in small mutational effects and slow evolution has been criticised by Barton and Polechová [12]. We therefore consider a scenario where the mutation rate is much higher, although decreasing, and the mutation events are no longer separated. This allows for several mutations to accumulate before a new type fully invades the population. To study the extreme case, as first done by Bovier and Wang [26] and recently by Bovier, Coquille, and Smadi [25], we consider the two limits separately. We take the deterministic model, arising from the limit of large populations, and let the mutation rate μ tend to zero while rescaling the time by $\ln 1/\mu$. This corresponds to the time that a mutant takes to reach a macroscopic population size of order 1, rather than the time until a mutant appears, as before. The time that the system takes to re-equilibrate is negligible on the chosen time scale and hence the resulting limit is a jump process between metastable equilibrium states.

We consider a finite type space with a graph structure representing the possibility of mutation. First, we prove that, under certain assumptions, the deterministic model converges pointwise to a deterministic jump process in the rare mutation limit. This process jumps between Lotka-Volterra equilibria of the current macroscopic types. For a (possibly polymorphic) resident population, we have to carefully track the growth of the different microscopic mutants that compete to invade the population. The first mutant to reach a macroscopically visible population size solves an optimisation problem and balances high invasion fitness and large initial conditions, where the latter is determined by the graph distance to the resident types. The limiting process can be fully described by its jump times and jump chain, which are closely related to this optimisation problem. It can make arbitrarily large jumps and may reach an evolutionary stable state.

Second, we show how we can derive different limiting processes by changing the parameters of the system. On one hand, assuming equal competition between all individuals and monomorphic initial conditions, the description of the jump process can be simplified. In this case, the invasion fitness of a type is just the difference between its own individual fitness, defined by its birth and death rate, and that of the resident type. Hence, we can relate back to the classical notion of fixed fitness landscapes in the context of adaptive walks. The limiting process resembles an adaptive flight since it always jumps to types of higher individual fitness, eventually reaching a global fitness maximum. A similar scenario was studied in the context of adaptive walks and flights in [114, 99, 100, 98]. Here the fitness is also assumed to be fixed but time steps are discrete. As in our case, the transitions between macroscopic types are determined by balancing high initial conditions, depending on the distance in the type space, and high fitness.

On the other hand, we modify the deterministic system such that the subpopulations can only reproduce when their size lies above a certain threshold. This limits the radius in which a resident population can foster mutants. A threshold of μ^ℓ mimics the scaling of $\mu_K \approx K^{-1/\ell}$ in the simultaneous limit, where resident types can produce mutants in a

radius of ℓ. Bovier, Coquille, and Smadi [25] as well as Champagnat, Méléard, and Tran [38] recently studied this scaling for the type space of a discrete line. A similar scaling has also been applied to a Moran-type model by Durrett and Mayberry [65] and an adaptive walk-type model with restricted mutation radius has been studied by Jain and Krug [100]. The resulting limit processes of the modified deterministic system are similar to the previously mentioned greedy adaptive walk. However, they are not all restricted to jumping to direct neighbours only, and thus can cross valleys in the fitness landscape and reach a global fitness maximum. Only when we choose the extreme case $\ell = 1$, the resulting limit is exactly the greedy adaptive walk.

The remainder of this paper is organised as follows. In Section 2.2, we introduce the deterministic system and the corresponding mutation free Lotka-Volterra system and present the main theorems, stating the convergence to different jump processes in the limit of rare mutation for different scenarios. We relate the deterministic system to the individual-based stochastic model and present a modification that mimics the simultaneous limit of large populations and rare, but still overlapping, mutations. Moreover, we give a short outline of the strategy of the proofs. Sections 2.3 and 2.4 are devoted to the proof of the first convergence result. The proof is split into three parts. The analysis of the exponential growth phase of the mutants, which follows ideas from Bovier and Wang [26], is given in Section 2.3. The following Lotka-Volterra invasion phase has been studied in detail by Champagnat, Jabin, and Raoul [35]. In Section 2.4, we show how to combine the two phases to prove the main result. Next, in Section 2.5, we consider the special case of equal competition, where we can simplify the description of the limiting jump process. Since the assumptions of the result from [35] are no longer satisfied, we have to slightly change the proof. In Section 2.6, we finally present an extension of the original deterministic system, where we limit the range of mutation to mimic the scaling of $\mu_K \approx K^{-1/\ell}$ in the simultaneous limit. In the extreme case, where only resident types can foster mutants, the greedy adaptive walk arises in the limit. For the intermediate cases, we present some first results on accessibility of types.

2.2 Model introduction and main results

In this section we introduce the deterministic model for evolution that is the focus of our studies. Similar models have been studied by Hofbauer and Sigmund [96], who give an extensive overview of models of population dynamics in their book. We present the main result of convergence of this deterministic process in the limit of rare mutations on a divergent time scale. The limiting process is a deterministic jump process that jumps between Lotka-Volterra equilibria, involving different types. In the special case of equal competition, we derive a simplified description of this limiting process. Moreover, we relate the model to the stochastic individual-based model introduced by Fournier and Méléard [77] and present a modification of the deterministic system that mimics the simultaneous limit of large populations and rare, but not too rare, mutations. In the case where only neighbouring types of the current resident type can arise as mutants, the limiting object is a true adaptive walk. At the end of the section we outline the proofs that are given in the following sections.

2.2.1 The deterministic system and relations to Lotka-Volterra systems

The model we consider is a classical Lotka-Volterra system with additional mutation terms. We consider a population consisting of subpopulations that are characterised by their types (e.g. geno- or phenotypes). In this paper we choose the n-dimensional hypercube $\mathbb{H}^n := \{0,1\}^n$ as our *type space*. The sequences of ones and zeros can, for example, be interpreted as sequences of loci with different alleles. The type $(0,...,0)$ can be seen as the wild type while all other types have accumulated mutations on some loci. However, we will not assume to always start out with a monomorphic population of this type.

The choice of $\mathbb{H}^n$ can easily be generalised to any finite set. We comment on this in Remark 2.11.

The state of the system is described by $\xi_t^\mu = (\xi_t^\mu(x))_{x\in\mathbb{H}^n}$, where $\xi_t^\mu(x)$ denotes the size of the subpopulation of type x at time t. ξ_t^μ can be seen as a non-negative vector or (not necessarily normalised) measure on $\mathbb{H}^n$.

The dynamics of $(\xi_t^\mu)_{t\geq 0}$ are determined by the system of differential equations

$$\begin{aligned}\tfrac{d}{dt}\xi_t^\mu(x) = &\left[b(x) - d(x) - \sum_{y\in\mathbb{H}^n}\alpha(x,y)\xi_t^\mu(y)\right]\xi_t^\mu(x)\\ &+ \mu\sum_{y\in\mathbb{H}^n}\xi_t^\mu(y)b(y)m(y,x) - \mu\xi_t^\mu(x)b(x)\sum_{y\in\mathbb{H}^n}m(x,y), \quad x\in\mathbb{H}^n, \end{aligned} \tag{2.1}$$

where the parameters are chosen as follows.

Definition 2.1. For $x, y \in \mathbb{H}^n$, we define

- $b(x) \in \mathbb{R}_+$, the *birth rate* of an individual with type x,
- $d(x) \in \mathbb{R}_+$, the *(natural) death rate* of an individual with type x,
- $\alpha(x,y) \in \mathbb{R}_+$, the *competitive pressure* that is imposed upon an individual with type x by an individual with type y,
- $\mu \in [0,1]$, the *probability of mutation* at a birth event,
- $m(x,\cdot) \in \mathcal{M}_p(\mathbb{H}^n)$, the *law of the mutant*.

Here $\mathcal{M}_p(\mathbb{H}^n)$ is the set of probability measures on $\mathbb{H}^n$. We assume that $m(x,x) = 0$, for every $x \in \mathbb{H}^n$. For each $x \in \mathbb{H}^n$, we define $r(x) := b(x) - d(x)$, its *individual fitness*.

Abiotic factors like temperature, chemical milieu, or other environmental properties enter through b and d, while biotic factors such as competition due to limited food supplies, segregated toxins, or predator-prey relationships are reflected in the competition kernel α.

We could also let the probability of mutation depend on $x \in \mathbb{H}^n$ in a way such that it is still proportional to some μ, i.e. $\mu M(x)$. However, this would not change the limiting process, therefore we stick with a constant μ for simplicity of notation.

Note that the competition term ensures that solutions are always bounded. This implies Lipschitz continuity for the coefficients, and hence the classical theory for ordinary differential equations ensures existence, uniqueness, and continuity in t of such solutions ξ_t^μ. Moreover, for non-negative initial condition ξ_0^μ, ξ_t^μ is non-negative at all times.

Definition 2.2. For $x \in \mathbb{H}^n$, we denote by $|x| := \sum_{i=1}^n x_i$ the 1-norm. We write $x \sim y$ if x and y are direct neighbours on the hypercube, i.e. if $|x - y| = 1$. Else, we write $x \nsim y$. We denote the standard Euclidean norm by $\|\cdot\|$.

To ensure that the mutants which a type $x \in \mathbb{H}^n$ can produce are exactly its direct neighbours, we introduce the following assumption. It corresponds to only allowing single mutations.

(A) For every $x, y \in \mathbb{H}^n$, $m(x, y) > 0$ if and only if $x \sim y$.

Again, this assumption is not necessary and can easily be relaxed. However, it simplifies notation and does not change the method of the proofs. We comment on the case of general finite (directed) graphs as type spaces in Remark 2.11.

Under the above assumption, (2.1) reduces to

$$\frac{d}{dt}\xi_t^\mu(x) = \left[r(x) - \sum_{y\in\mathbb{H}^n} \alpha(x,y)\xi_t^\mu(y)\right]\xi_t^\mu(x) + \mu\sum_{y\sim x} b(y)m(y,x)\xi_t^\mu(y) - \mu b(x)\xi_t^\mu(x). \quad (2.2)$$

In the mutation-free case, where $\mu = 0$, the equations take the form of a competitive Lotka-Volterra system

$$\frac{d}{dt}\xi_t^0(x) = \left[r(x) - \sum_{y\in\mathbb{H}^n} \alpha(x,y)\xi_t^0(y)\right]\xi_t^0(x). \quad (2.3)$$

Understanding this system is essential since it determines the short term dynamics of the system with mutation as $\mu \to 0$. For a subset of types we study the stable states of the Lotka-Volterra system involving these types.

Definition 2.3. For a subset $\mathbf{x} \subset \mathbb{H}^n$ we define the set of *Lotka-Volterra equilibria* by

$$\mathrm{LVE}(\mathbf{x}) := \left\{\xi \in (\mathbb{R}_{\geq 0})^{\mathbf{x}} : \forall\, x \in \mathbf{x} :\ \Big[r(x) - \sum_{y\in\mathbf{x}} \alpha(x,y)\xi(y)\Big]\xi(x) = 0\right\}. \quad (2.4)$$

Moreover, we let $\mathrm{LVE}_+(\mathbf{x}) := \mathrm{LVE}(\mathbf{x}) \cap (\mathbb{R}_{>0})^{\mathbf{x}}$. If $\mathrm{LVE}_+(\mathbf{x})$ contains exactly one element, we denote it by $\bar{\xi}_{\mathbf{x}}$, the *equilibrium size* of a population of coexisting types $\mathbf{x}$.

Remark 2.4. If $\mathrm{LVE}_+(\mathbf{x}) = \{\bar{\xi}_{\mathbf{x}}\}$, this implies $r(x) > 0$ for all $x \in \mathbf{x}$. In the case where $\mathbf{x} = \{x\}$, we obtain $\bar{\xi}_x(x) := \bar{\xi}_{\mathbf{x}}(x) = \frac{r(x)}{\alpha(x,x)}$.

The following assumption ensures that for a subset $\mathbf{x} \subset \mathbb{H}^n$, such that $r(x) > 0$ for all $x \in \mathbf{x}$, there exists a unique asymptotically stable equilibrium of the Lotka-Volterra system involving types $\mathbf{x}$.

($\mathbf{B_x}$) There exist $\theta_x > 0$, $x \in \mathbf{x}$, such that

$$\forall\ x, y \in \mathbf{x} : \theta_x \alpha(x,y) = \theta_y \alpha(y,x), \tag{2.5}$$

$$\forall\ u \in \mathbb{R}^{\mathbf{x}} \backslash \{0\} : \sum_{x,y \in \mathbf{x}} \theta_x \alpha(x,y) u(x) u(y) > 0. \tag{2.6}$$

This is, for example, trivially satisfied by any symmetric, positive definite matrix $(\alpha(x,y))_{x,y\in\mathbf{x}}$. Under this condition, Champagnat, Jabin, and Raoul have proven convergence to a unique stable equilibrium.

Theorem 2.5 ([35], Prop.1)**.** *Assume ($B_{\boldsymbol{x}}$) for a subset $\boldsymbol{x} \subset \mathbb{H}^n$ such that $r(x) > 0$, for all $x \in \boldsymbol{x}$. Then there exists a unique $\bar{\xi}_{\boldsymbol{x}} \in (\mathbb{R}_+)^{\boldsymbol{x}} \backslash \{0\}$ such that for any solution ξ_t^0 to (2.3) with initial condition $\xi_0^0 \in (\mathbb{R}_{>0})^{\boldsymbol{x}} \times \{0\}^{\mathbb{H}^n \backslash \boldsymbol{x}}$,*

$$\xi_t^0\Big|_{\boldsymbol{x}} \to \bar{\xi}_{\boldsymbol{x}} \ \text{ as } t \to \infty. \tag{2.7}$$

The proof of this theorem uses the Lyapunov functional

$$L(\xi) = \frac{1}{2} \sum_{x,y \in \mathbf{x}} \theta_x \alpha(x,y) \xi(x) \xi(y) - \sum_{x \in \mathbf{x}} \theta_x r(x) \xi(x), \quad \xi \in \mathbb{R}^{\mathbf{x}}. \tag{2.8}$$

(2.5) ensures that

$$\frac{d}{dt} L(\xi_t^0\Big|_{\mathbf{x}}) = (\nabla L)(\xi_t^0\Big|_{\mathbf{x}}) \cdot \tfrac{d}{dt}\, \xi_t^0\Big|_{\mathbf{x}} = -\sum_{x \in \mathbf{x}} \theta_x \left[r(x) - \sum_{y \in \mathbf{x}} \alpha(x,y) \xi_t^0(y) \right]^2 \xi_t^0(x) \le 0, \tag{2.9}$$

while (2.6) gives convexity of L.

Remark 2.6. Note that 2.6 implies

$$\forall\ u \in \mathbb{R}^{\mathbf{x}} \backslash \{0\} : \sum_{x,y \in \mathbf{x}} \theta_x \alpha(x,y) u(x) u(y) \ge \kappa_{\mathbf{x}} \|u\|^2, \tag{2.10}$$

where

$$\kappa_{\mathbf{x}} := \min_{u : \|u\| = 1} \sum_{x,y \in \mathbf{x}} \theta_x \alpha(x,y) u(x) u(y) > 0. \tag{2.11}$$

We set $\kappa := \min_{\mathbf{x} \subset \mathbb{H}^n} \kappa_{\mathbf{x}}$.

Connected to this positive definiteness property and the Lotka-Volterra equilibria, we define a norm that is used to measure the distance between the current state of the population and the equilibrium size. Since the θ_x, $x \in \mathbf{x}$, in ($\mathrm{B}_{\mathbf{x}}$) are not unique, we fix an arbitrary choice of such parameters. In the case where $(\alpha(x,y))_{x,y\in\mathbf{x}}$ is irreducible, we can choose the unique normalised version where $\sum_{x\in\mathbf{x}} \theta_x = 1$.

Definition 2.7. For $\mathbf{x} \subset \mathbb{H}^n$ such that $\mathrm{LVE}_+(\mathbf{x}) = \{\bar{\xi}_{\mathbf{x}}\}$ and ($\mathrm{B}_{\mathbf{x}}$) is satisfied, we define a scalar product on $\mathbb{R}^{\mathbf{x}}$ (or $\mathcal{M}(\mathbf{x})$, the set of non-negative measures on $\mathbf{x}$) by

$$\langle u, v \rangle_{\mathbf{x}} := \sum_{x \in \mathbf{x}} \frac{\theta_x}{\bar{\xi}_{\mathbf{x}}(x)} u(x) v(x), \quad u, v \in \mathbb{R}^{\mathbf{x}}. \tag{2.12}$$

The corresponding norm is defined by $\|u\|_{\mathbf{x}} := \sqrt{\langle u, u \rangle_{\mathbf{x}}}$.

This scalar product is chosen exactly in a way such that we can use the positive definiteness (2.10) and the properties of $\bar{\xi}_{\mathbf{x}}$. Moreover, we notice that

$$c_{\mathbf{x}}^2 \|u\|^2 := \left(\min_{x \in \mathbf{x}} \frac{\theta_x}{\bar{\xi}_{\mathbf{x}}(x)} \right) \|u\|^2 \leq \|u\|_{\mathbf{x}}^2 \leq \left(\max_{x \in \mathbf{x}} \frac{\theta_x}{\bar{\xi}_{\mathbf{x}}(x)} \right) \|u\|^2 =: C_{\mathbf{x}}^2 \|u\|^2 . \tag{2.13}$$

Remark 2.8. Throughout the paper, constants labelled c and C have varying values. Specific constants, as $c_{\mathbf{x}}$ and $C_{\mathbf{x}}$ above, are labelled differently and referenced when used repetitively.

While some types $\mathbf{x}$ coexist at their equilibrium size $\bar{\xi}_{\mathbf{x}}$, other types $y \in \mathbb{H}^n \backslash \mathbf{x}$, which only have a small population size, grow in their presence. Considering the rate of exponential growth in (2.3), we formulate a notion of invasion fitness.

Definition 2.9. For $\mathbf{x} \subset \mathbb{H}^n$ such that $\mathrm{LVE}_+(\mathbf{x}) = \{\bar{\xi}_{\mathbf{x}}\}$ and $y \in \mathbb{H}^n$, we define the *invasion fitness* of an individual with type y in a population of coexisting types $\mathbf{x}$ at equilibrium by $f_{y,\mathbf{x}} := r(y) - \sum_{x \in \mathbf{x}} \alpha(y,x) \bar{\xi}_{\mathbf{x}}(x)$.

Notice that $f_{x,\mathbf{x}} = 0$ for all $x \in \mathbf{x}$. In contrast to the individual fitness r, which is fixed, this notion of fitness varies over time and depends on the current resident types.

2.2.2 Convergence to a deterministic jump process

We now come back to the system (2.2), involving mutation. We assume that the system starts out close to the equilibrium size of some subset of types $\mathbf{x} \subset \mathbb{H}^n$ and study its evolution over time. We distinguish between macroscopic resident types that coexist at their equilibrium size and microscopic mutant types that have a population size that tends to 0 as $\mu \to 0$. The initial conditions are specified as follows.

Definition 2.10. A collection of measures $\xi_0^\mu \in \mathcal{M}(\mathbb{H}^n)$, depending on μ, *satisfies the initial conditions for resident types* $\boldsymbol{x} \subset \mathbb{H}^n$, $\eta > 0$, *and* $\bar{c} > 0$ if $\mathrm{LVE}_+(\mathbf{x}) = \{\bar{\xi}_{\mathbf{x}}\}$ and there exists a $\mu_0 \in (0,1]$ and constants $0 \leq c_y \leq C_y < \infty$ and $\lambda_y \geq 0$, for each $y \in \mathbb{H}^n$, such that, for every $\mu \in (0, \mu_0]$,

$$\xi_0^\mu(y) \in [c_y \mu^{\lambda_y}, C_y \mu^{\lambda_y}], \tag{2.14}$$

where

$$\forall\ y \in \mathbf{x} : \lambda_y = 0,\ \bar{\xi}_{\mathbf{x}}(y) - \eta \frac{\bar{c}}{\sqrt{|\mathbf{x}|}} \leq c_y, C_y \leq \bar{\xi}_{\mathbf{x}}(y) + \eta \frac{\bar{c}}{\sqrt{|\mathbf{x}|}}, \tag{2.15}$$

$$\forall\ y \in \mathbb{H}^n \backslash \mathbf{x} : \lambda_y > 0,\ 0 \leq c_y, C_y < \infty \quad \text{or} \tag{2.16}$$

$$\lambda_y = 0,\ 0 \leq c_y, C_y \leq \frac{\eta}{3},\ f_{y,\mathbf{x}} < 0. \tag{2.17}$$

If $\xi_0^\mu(y) \equiv 0$, we choose any $\lambda_y > \max_{z \in \mathbb{H}^n : \xi_0^\mu(z) > 0} \lambda_z + n$.
We write $\xi_0^\mu \in \mathrm{IC}(\mathbf{x}, \eta, \bar{c})$.

This definition is very technical. We could choose more simple initial conditions for our main theorem, for example a monomorphic macroscopic type and no microscopic types of positive population size. However, after the first invasion step, this is exactly what the system looks like, and we want to be able to iterate our procedure. The definition roughly implies that all macroscopic types are close to their coexistence equilibrium (within an attractive domain) and all microscopic types y are of order μ^{λ_y} as $\mu \to 0$. The types that are not part of the resident types but of order μ^0 are assumed to be unfit and of small enough size. This ensures that they do not "trigger" the stopping time that marks the beginning of the next Lotka-Volterra phase, i.e. the time when the first fit mutant reaches a macroscopic level. This stopping time is defined in (2.39).

Let $\mathbf{x}^0 \subset \mathbb{H}^n$ be the initial set of coexisting types, i.e. $\xi_0^\mu \in \mathrm{IC}(\mathbf{x}^0, \eta, \bar{c})$, and set $T_0 := 0$. During a time of order 1, each type $y \in \mathbb{H}^n$ grows to a size of order $\mu^{\rho_y^0}$, where

$$\rho_y^0 := \min_{z \in \mathbb{H}^n} [\lambda_z + |z - y|], \tag{2.18}$$

due to incoming mutants from other types. This can be argued as follows. The population of type y collects incoming mutants from all other types z of order $\xi_0^\mu(z)\mu^{|z-y|}$. These influences are summed up but in the limit of $\mu \to 0$, the asymptotically largest summand, i.e. the smallest exponent of μ, dominates all other terms.

Assume now that, after the $(i-1)^{\text{st}}$ invasion, at time $T_{i-1} \ln 1/\mu$, we have coexisting resident types $\mathbf{x}^{i-1}$ and all types $y \in \mathbb{H}^n$ have population size of order $\mu_y^{\rho_y^{i-1}}$, where $\rho_y^{i-1} = \min_{z \in \mathbb{H}^n} [\rho_z^{i-1} + |z - y|]$ is satisfied. During a time of order $\ln 1/\mu$, microscopic types grow until the first type reaches a population size of order 1. The population sizes during growth can be approximated as

$$\xi^\mu_{t \ln \frac{1}{\mu}}(y) \approx \mu^{\min_{z \in \mathbb{H}^n} [\rho_z^{i-1} + |z-y| - (t - T_{i-1}) f_{z, \mathbf{x}^{i-1}}]}. \tag{2.19}$$

This is a little tricky and takes into account that there are three possible sources that could dominate the growth of type y: First, the population at y could just grow at its own exponential growth rate $f_{y,\mathbf{x}^{i-1}}$. This gives $\mu^{\rho_y^{i-1} - (t-T_{i-1}) f_{y,\mathbf{x}^{i-1}}}$. Second, it could come from mutants from the large populations in $x \in \mathbf{x}^{i-1}$. This gives $\mu^{|x-y|}$ since x has to mutate $|x - y|$ times to reach y. Finally, it could come from the mutants that have grown at any other site z over the last period. This gives $\mu^{\rho_z^{i-1} + |z-y| - (t-T_{i-1}) f_{z,\mathbf{x}^{i-1}}}$.

Since mutants from another type can never increase the population size past μ^1, the first microscopic type y to reach a size of order 1 must have grown at its own rate $f_{y,\mathbf{x}^{i-1}} > 0$. The time to reach this macroscopic size (after the last invasion) is of order $(\rho_y^{i-1}/f_{y,\mathbf{x}^{i-1}}) \ln 1/\mu$.

Summarising these thoughts, we inductively define

$$y_*^i := \operatorname*{arg\,min}_{\substack{y \in \mathbb{H}^n: \\ f_{y,\mathbf{x}^{i-1}} > 0}} \frac{\rho_y^{i-1}}{f_{y,\mathbf{x}^{i-1}}}, \tag{2.20}$$

the i^{th} invading type (if the minimiser is unique),

$$T_i := T_{i-1} + \min_{\substack{y\in\mathbb{H}^n:\\ f_{y,\mathbf{x}^{i-1}}>0}} \frac{\rho_y^{i-1}}{f_{y,\mathbf{x}^{i-1}}}, \tag{2.21}$$

the time of the i^{th} invasion on the time scale $\ln 1/\mu$, and

$$\rho_y^i := \min_{z\in\mathbb{H}^n}[\rho_z^{i-1} + |z-y| - (T_i - T_{i-1})f_{z,\mathbf{x}^{i-1}}] \tag{2.22}$$

the μ-exponent of the population size of type y at the time of the i^{th} invasion. If there is no $y \in \mathbb{H}^n$ such that $f_{y,\mathbf{x}^{i-1}} > 0$, we set $T_i := \infty$.

At time $T_i \ln 1/\mu$, the types y_*^i and $\mathbf{x}^{i-1}$ re-equilibrate according to the mutation-free Lotka-Volterra dynamics. If it is unique, we denote the support of the new equilibrium, i.e. the new coexisting resident types, by $\mathbf{x}^i$.

Remark 2.11. The results in this paper can easily be generalised to finite, possibly directed graphs as type spaces, where (directed) edges mark the possibility of mutation. In these cases the Hamming distance on the hypercube (e.g. $|z-y|$ in (2.22)) is replaced by a "directed" distance, corresponding to lengths of directed paths (e.g. by the length of the shortest path from z to y). Note that this directed distance is not a distance in the classical sense since it might not be symmetric. For ease of notation and due to the nice applicability to genetic sequences, we stick with the hypercube in this paper.

With the above notation, we can now characterise the limiting process as follows.

Theorem 2.12. *Consider the system of differential equations (2.2) and let $\xi_0^\mu \in IC(\boldsymbol{x}^0, \eta, \bar{c})$, for η small enough. Assume (A) and ($B_{\boldsymbol{x}^{i-1}\cup y_*^i}$), for every $1 \le i < I$, where we set $I := i$ for the smallest $i \in \mathbb{N}$ where either*

(a) the minimiser in (2.20) is not unique, or

(b) there is a $y \in (\boldsymbol{x}^{i-1} \cup y_*^i)\backslash \boldsymbol{x}^i$ such that $f_{y,\boldsymbol{x}^i} \ge 0$,

and $I := \infty$ if none of these occur. In the latter case, we set $T_\infty := \infty$.

Then, for every $t \in [0, T_I)\backslash\{T_i, 0 \le i \le I\}$,

$$\lim_{\mu\to 0} \xi^\mu_{t\ln\frac{1}{\mu}} = \sum_{i=0}^{I-1} \mathbb{1}_{T_i\le t<T_{i+1}} \sum_{x\in\boldsymbol{x}^i} \delta_x \bar{\xi}_{\boldsymbol{x}^i}(x). \tag{2.23}$$

Remark 2.13. (i) Note that ($B_{\mathbf{x}^{i-1}\cup y_*^i}$) implies ($B_{\mathbf{x}^{i-1}}$) and ($B_{\mathbf{x}^i}$), with the same constants θ_x.

(ii) Case (a) is very unlikely if the parameters of the model are chosen in a random fashion since it requires a very particular equality. Case (b) guarantees that we terminate the procedure as soon as the conditions of $\mathrm{IC}(\mathbf{x}^i, \eta, \bar{c})$ are not satisfied after the i^{th} invasion. For every $y \in (\mathbf{x}^{i-1} \cup y_*^i)\backslash\mathbf{x}^i$, $f_{y,\mathbf{x}^i} \le 0$ is ensured (going through the proof of Theorem 2.5), so the only problem can arise from equality.

(iii) Note that the theorem implies that, in the case of $T_i = \infty$, even if there was a mutant type $y \in \mathbb{H}^n\backslash\mathbf{x}^{i-1}$ such that $f_{y,\mathbf{x}^{i-1}} = 0$, it would not be able to invade the resident population.

The proof of this result is given in Sections 2.3 and 2.4.

2.2.3 Convergence in the case of equal competition

In the context of adaptive walks and flights, the fitness landscape on the type space is possibly random, but usually fixed over time. For a monomorphic resident type, the current fitness of any type, corresponding to its invasion fitness, is determined by the difference between its individual fitness and the fitness of the resident type.

As a special case of our model, we consider equal competition between all types on the hypercube. In this case, one can simplify the description of the limit process and derive some interesting properties.

We introduce the additional assumption

(C) For every $x, y \in \mathbb{H}^n$, $\alpha(x, y) \equiv \alpha > 0$.

This leads to a couple of nice properties of the invasion fitness $f_{y,\{x\}}$. As in the adaptive walks framework, we obtain

$$f_{y,\{x\}} = r(y) - \alpha(y, x)\bar{\xi}_{\{x\}}(x) = r(y) - r(x), \tag{2.24}$$

which yields

$$f_{y,\{x\}} = -f_{x,\{y\}} \quad \text{and} \quad f_{z,\{y\}} + f_{y,\{x\}} = f_{z,\{x\}}. \tag{2.25}$$

As a consequence, there is some kind of transitivity of invasion fitness. A type z that is unfit relative to some other type y, i.e. $f_{z,\{y\}} < 0$, is unfit relative to all types that are fitter than y. This ensures that types which are once suppressed by resident types stay at a microscopic level forever. In particular, case (b) in Theorem 2.12 is automatically excluded by assumption (C).

As before, we terminate the procedure as soon as case (a) in Theorem 2.12 occurs to ensure that there is always a unique mutant that reaches the threshold of order 1 first after an invasion. Starting out with only a single type at its equilibrium size, i.e. $\mathbf{x}^0 = \{x^0\}$, this also implies that we avoid coexistence and always maintain a monomorphic resident population. This is due to the fact that an invading type has to have higher rate r than the current resident type, which prevents a polymorphic Lotka-Volterra equilibrium.

Assumption ($\mathrm{B}_{\mathbf{x}}$) can no longer be satisfied for constant α, as soon as $|\mathbf{x}| \geq 2$. However, it is no longer needed since the resident types are monomorphic, i.e. $|\mathbf{x}^i| = 1$, and have a lower rate r than the invading types, which implies a unique stable equilibrium of the Lotka-Volterra system involving $\mathbf{x}^i \cup y^i_*$. We comment on this in more detail in Section 2.5, where we adapt the proof of Theorem 2.12 to this situation.

In the case of a monomorphic resident population $\mathbf{x} = \{x\}$, we use the shorthand notation $\bar{\xi}_x := \bar{\xi}_{\{x\}}$, $f_{y,x} := f_{y,\{x\}}$. For types x^i, x^j, we write $f_{i,j} := f_{x^i,x^j}$.

The limiting jump process can now be described in a simple way.

Theorem 2.14. *Consider the system of differential equations (2.2) and let $\xi_0^\mu \in IC(\{x^0\}, \eta, \bar{c})$ such that $\lambda_y \geq |y - x^0|$, for all $y \in \mathbb{H}^n$, and η small enough. Assume (A) and (C) and set $I := i$ for the smallest $i \in \mathbb{N}$ such that the minimiser in (2.20) is not unique, and $I := \infty$ if this does not occur. In the latter case, we set $T_\infty := \infty$.*

Then, for every $t \in [0, T_I) \backslash \{T_i, 0 \leq i \leq I\}$,

$$\lim_{\mu \to 0} \xi^\mu_{t \ln \frac{1}{\mu}} = \sum_{i=0}^{I-1} \mathbb{1}_{T_i \leq t < T_{i+1}} \delta_{x^i} \bar{\xi}_{x^i}(x^i). \tag{2.26}$$

Moreover, the following identities hold:

$$x^i = \operatorname*{arg\,min}_{y \in \mathbb{H}^n : f_{y,x^{i-1}} > 0} \frac{|y - x^0| - |x^{i-1} - x^0|}{f_{y,x^{i-1}}}, \tag{2.27}$$

$$T_i = \frac{|x^i - x^0| - |x^{i-1} - x^0|}{f_{i,i-1}}. \tag{2.28}$$

Remark 2.15. (i) $\lambda_y \geq |y - x^0|$ ensures that the initial population size of all microscopic types is not larger than what they gain due to incoming mutants from x^0 within a time of order 1. This is necessary to obtain the identities for x^i and T_i.

(ii) Uniqueness of the minimiser in (2.20) is equivalent to uniqueness of the minimiser in (2.27) and hence x^i is well-defined.

(iii) In the case where $I = \infty$, the jump process in Theorem 2.14 continues as long as there is a type with higher individual fitness, i.e. higher rate r. As a result, it can cross arbitrarily large valleys in the fitness landscape (defined by r) and eventually reaches a global fitness maximum, where it remains. Note that this global maximum does not have to be unique. The jump process reaches the maximum that is closest to x^0 in $\mathbb{H}^n$, which is unique if $I = \infty$, and equally fit types cannot invade as mentioned in Remark 5(iii). With these properties, the jump process resembles an adaptive flight. However, it does not quite fit into that framework since it is not only jumping to local fitness maxima.

(iv) Every invasion step increases the distance on $\mathbb{H}^n$ between the resident type and x_0. This can be seen inductively as follows. Consider the $(i+1)^{\text{st}}$ invasion. x^i was a minimiser of $(|y - x^0| - |x^{i-1} - x^0|)/f_{y,x^{i-1}}$. If now y satisfies $f_{y,x^i} > 0$, then

$$\frac{|y - x^0| - |x^{i-1} - x^0|}{f_{y,x^{i-1}}} \geq \frac{|x^i - x^0| - |x^{i-1} - x^0|}{f_{i,i-1}}, \tag{2.29}$$

and since $f_{y,x^{i-1}} = f_{y,x^i} + f_{i,i-1} > f_{i,i-1}$ and $|x^i - x^0| > |x^{i-1} - x^0|$ (by assumption), $|y - x^0| - |x^{i-1} - x^0| > |x^i - x^0| - |x^{i-1} - x^0|$, and hence $|y - x^0| > |x^i - x^0|$.

The proof of Theorem 2.14 is found in Section 2.5.

2.2.4 Derivation from the individual-based stochastic model in the large population limit

The deterministic system, that is studied above, can be obtained as the large population limit of an individual-based Markov process. At time t, we consider a population of finite size $N(t) \in \mathbb{N}$. Each living individual is represented by its type $x_1(t), ..., x_{N(t)}(t) \in \mathbb{H}^n$ and the state of the population is described by the finite point measure

$$\nu_t^\mu = \sum_{i=1}^{N(t)} \delta_{x_i(t)}. \tag{2.30}$$

$\nu_t^\mu(x)$ describes the number of individuals of type $x \in \mathbb{H}^n$ at time t. The dynamics of the Markov process are determined by the same parameters b, d, α, μ, and m as for the deterministic system ξ_t^μ.

To let the size of the population tend to infinity, we introduce the *carrying capacity* of the environment, denoted by $K \in \mathbb{N}$. This can for example be interpreted as the amount of available space or resources. As K increases, the competitive pressure between individuals decreases and we scale $\alpha_K(x,y) \equiv \frac{\alpha(x,y)}{K}$. This leads to an equilibrium population size of order K. To derive a finite limit for large populations, i.e. as $K \to \infty$, we consider the rescaled measure

$$\nu_t^{\mu,K} := \frac{\nu_t}{K}. \tag{2.31}$$

This measure-valued Markov process can be constructed similar to [77, Ch. 2], with infinitesimal generator

$$\begin{aligned}
\mathcal{L}^K\phi(\nu) = &\sum_{x\in\mathbb{H}^n} K\nu(x)\left(\phi\left(\nu + \frac{\delta_x}{K}\right) - \phi(\nu)\right) b(x)(1-\mu) \\
&+ \sum_{x\in\mathbb{H}^n} K\nu(x) \sum_{y\sim x}\left(\phi\left(\nu + \frac{\delta_y}{K}\right) - \phi(\nu)\right) b(x)\mu m(x,y) \\
&+ \sum_{x\in\mathbb{H}^n} K\nu(x)\left(\phi\left(\nu - \frac{\delta_x}{K}\right) - \phi(\nu)\right)\left(d(x) + \sum_{y\in\mathbb{H}^n}\frac{\alpha(x,y)}{K}K\nu(y)\right),
\end{aligned} \tag{2.32}$$

where $\nu \in \mathcal{M}(\mathbb{H}^n)$ is a non-negative measure on $\mathbb{H}^n$ and ϕ a measurable bounded function from $\mathcal{M}(\mathbb{H}^n)$ to $\mathbb{R}$.

Ethier and Kurtz have shown convergence of this process to ξ^μ as K tends to infinity.

Theorem 2.16 ([70], Ch. 11, Thm. 2.1)**.** *Assume that the initial conditions converge almost surely to a deterministic limit, i.e. $\nu_0^{\mu,K} \to \xi_0^\mu$, as $K \to \infty$. Then, for every $T \geq 0$, $(\nu_t^{\mu,k})_{0\leq t\leq T}$ almost surely converges uniformly to the deterministic process $(\xi_t^\mu)_{0\leq t\leq T}$, which is the unique solution to the system of differential equations 2.2 with initial condition ξ_0^μ.*

2.2.5 Convergence for a limited radius of mutation

The limiting process in Theorem 2.14 already looks similar to the greedy adaptive walk of Nowak and Krug [147], mentioned in the introduction. It is a monomorphic jump process on the type space that always jumps to types of higher individual fitness r. However, it can take larger steps than just to neighbouring types and we have seen that the initial type x^0 plays an important role in determining the jump chain. This is due to the fact that, already after an arbitrarily small time, mutation has induced a positive population size for every possible type. These mutant populations have size of order μ to the power of the distance to x^0 on $\mathbb{H}^n$. The next invading type is then found balancing low initial μ-power and high (invasion) fitness.

In all our previous considerations, arbitrarily small populations were able to reproduce and foster mutants, which can lead to population sizes as small as μ^n. This might not always fit reality well.

If we consider the stochastic model, introduced in the previous subsection, and allow for the mutation probability μ_K to decrease as K increases, we can study the simultaneous limit of large populations and rare mutations. To be able to reproduce within a time of order 1 in a population of size μ_K^n implies that

$$\lim_{K\to\infty} \mu_K^n \cdot K \geq 1, \tag{2.33}$$

or equivalently

$$\lim_{K\to\infty} \frac{\mu_K}{K^{-\frac{1}{n}}} \geq 1. \tag{2.34}$$

In this case, we would recover the deterministic system (2.2) in the limit of $K \to \infty$.

If now μ_K was of order $K^{-\frac{1}{\ell}}$ for some $\ell < n$, this implies that populations with a size of order μ_K^λ, for $\lambda > \ell$ are vanishing as $K \to \infty$ and hence cannot reproduce. If we consider a monomorphic resident type x, it spreads mutants y of population size $\mu_K^{|y-x|}$. This means that it can initially only foster mutant populations in a radius of ℓ.

This regime has already been studied by Bovier, Coquille, and Smadi [25] as well as Champagnat, Méléard, and Tran [38]. It is shown that, on the type space $\mathbb{N}$ (with neighbours having difference exactly 1) and on the usual time scale of $\ln 1/\mu_K$, a fitness valley of width $\leq \ell$, but no further, can be crossed. However, crossing a wider valley is possible on a faster diverging time scale.

In the following, we want to mimic this parameter regime of the stochastic system in our deterministic model. To do so, we introduce a cut-off that freezes the dynamics of a population below the threshold of $\bar{\xi}\mu^\ell$, where $\bar{\xi} := \min\{\bar{\xi}_x(x)/2 : x \in \mathbb{H}^n, r(x) > 0\} > 0$ is chosen such that every resident type will eventually surpass this value (which is relevant in the case

$\ell = 1$). The new system of differential equations then reads as follows.

$$\begin{aligned}\tfrac{d}{dt}\xi_t^\mu(x) = &\left[b(x)\mathbb{1}_{\xi_t^\mu(x)\geq\bar{\xi}\mu^\ell} - d(x) - \sum_{y\in\mathbb{H}^n}\alpha(x,y)\xi_t^\mu(y)\mathbb{1}_{\xi_t^\mu(y)\geq\bar{\xi}\mu^\ell}\right]\xi_t^\mu(x)\\ &+\mu\sum_{y\sim x}\xi_t^\mu(y)\mathbb{1}_{\xi_t^\mu(y)\geq\bar{\xi}\mu^\ell}b(y)m(y,x) - \mu\xi_t^\mu(x)\mathbb{1}_{\xi_t^\mu(x)\geq\bar{\xi}\mu^\ell}b(x). \qquad (2.35)\end{aligned}$$

Remark 2.17. Reproduction (clonal and non-clonal) is stopped for types below the threshold of $\bar{\xi}\mu^\ell$. As a result, those types are in a kind of dormant state and can only grow due to the mutational influence of other, larger types. It does not affect the system that these dormant types remain at a low level since they do not influence the dynamics of other types and only become active again if they gain a larger amount due to incoming mutants.

The death rate of populations below $\bar{\xi}\mu^\ell$ is not set to zero. This is necessary to actually drop below the threshold if a population declines due to negative fitness. Otherwise, the population would remain at exactly $\bar{\xi}\mu^\ell$ and could immediately start growing again when its fitness becomes positive due to a change of resident types. This is however not what we want to achieve since populations that drop to the threshold are supposed to go extinct and only reappear due to incoming new mutants.

We cannot simply set the population size of a type to zero below the threshold. In that case, a zero-type would never become active since every gain due to mutation would immediately be cancelled.

As mentioned above, for $\ell \geq n$, we just recover the original scenario of Theorem 2.12. This is due to the fact that, as long as there is at least one macroscopic type, every other type has population size of at least μ^n, due to mutants from this macroscopic type.

For $\ell = 1$, if we keep assumption (C) of constant competition and a monomorphic initial condition, we obtain the greedy adaptive walk of Nowak and Krug [147], where the process always jumps to the fittest direct neighbour of the current resident type. We re-define

$$x^i := \arg\max_{y\sim x^{i-1}} r(y), \qquad (2.36)$$

$$T_i := T_{i-1} + \frac{1}{f_{i,i-1}}, \qquad (2.37)$$

and set $T_i := \infty$, as soon as there exists no $y \sim x^{i-1}$ such that $r(y) > r(x^{i-1})$.

The convergence can be stated as follows.

Theorem 2.18. *Consider the system of differential equations (2.35) for $\ell = 1$ and let $\xi_0^\mu \in IC(\{x^0\}, \eta, \bar{c})$ such that $\lambda_y \geq 1$, for all $y \sim x^0$, $\lambda_y \geq 2$, for $|y - x^0| \geq 2$, and η small enough. Assume (A) and (C) and set $I := i$ for the smallest $i \in \mathbb{N}$ such that the maximiser in (2.36) is not unique, and $I := \infty$ if this does not occur. In the latter case, we set $T_\infty := \infty$.*

Then, for every $t \in [0, T_I)\backslash\{T_i, 0 \leq i \leq I\}$,

$$\lim_{\mu\to 0}\xi^\mu_{t\ln\frac{1}{\mu}} = \sum_{i=0}^{I-1}\mathbb{1}_{T_i\leq t<T_{i+1}}\delta_{x^i}\bar{\xi}_{x^i}(x^i). \qquad (2.38)$$

Remark 2.19. (i) $\lambda_y \geq 2$, for $|y - x^0| \geq 2$, ensures that no microscopic type has a larger initial population that what it gains due to the first incoming mutants from other types.

(ii) The adaptive walk in Theorem 2.18 stops as soon as it reaches a local maximum of the individual fitness r since only direct neighbours of the resident type can be reached. Local maxima do not need to be strict. However, as in the previous cases, mutants with invasion fitness 0 cannot invade the resident population.

(iii) It is no longer the case that every step increases the distance to x^0. The walk could return to a type close to x^0, which just could not be reached before because one had to go around a valley in the fitness landscape defined by r.

In Section 2.6, we discuss the proof of Theorem 2.18, as well as the intermediate cases of $1 < \ell < n$.

2.2.6 Structure of the proofs

The general strategy of the proofs of all three theorems is to split the analysis of the evolution into two parts. First, the microscopic mutants grow in the presence of the coexisting resident types until one of them reaches a macroscopic population size of order 1, i.e. that does not vanish as $\mu \to 0$. Second, this macroscopic mutant and the resident types attain a new equilibrium according to the Lotka-Volterra dynamics. The two phases are visualised in Figure 2.2, found in Section 2.4, prior to the proof of Theorem 2.12.

The first phase is studied in detail in Section 2.3. Theorem 2.21 gives upper and lower bounds for the exponential growth of the non-resident types. The growth can be due to a type's own (invasion) fitness or due to mutants from a growing neighbour. To get the correct approximation, the influences of all existing types have to be summed up. Meanwhile, the resident types stay close to their equilibrium. Corollary 2.22 considers the $\ln 1/\mu$-time scale and derives an approximation for the first time that a mutant reaches the macroscopic threshold.

After the threshold is reached, for the second phase, we can apply Theorem 2.5 to the Lotka-Volterra system involving the macroscopic mutant type and the resident types to derive the convergence to a new equilibrium state. This is possible since we now have a non-negative initial condition that does not vanish as $\mu \to 0$.

In Section 2.4, this theorem is combined with Theorem 2.21, or rather Corollary 2.22, to analyse the full evolution of our system (2.2). First, in Lemma 2.23, the dependence of solutions on the initial condition and the size of μ is studied to be able to approximate the full system by the Lotka-Volterra system only involving the macroscopic types. Second, in Lemma 2.25, continuity of the duration of the Lotka-Volterra phase in the initial condition is shown. From this, a uniform bound on the time to reach the initial conditions of Theorem 2.21 again is derived. All of this is then combined to show the convergence in Theorem 2.12, one invasion step at a time, and recursively describe the limiting process.

To prove Theorem 2.14, only slight changes have to be made. Since assumption ($\mathrm{B}_{\mathbf{x}}$) is not satisfied, Theorem 2.5 can no longer be applied directly. However, the assumption is mainly

needed to show uniqueness of the limiting equilibrium, which is, in this case, already implied by the structure of the individual fitness landscape. The rest of the proof, found in Section 2.5, is then devoted to simplifying the expressions for y_*^i, T_i, and ρ_y^i.

In Section 2.6, Theorem 2.18 is proved. Here, the bounds from Theorem 2.21 have to be revised. The rest of the argument follows the previous proofs.

2.3 Invasion Analysis

In this section, we prove an exponential approximation of the growth of the non-resident subpopulations until the first type reaches a macroscopic threshold of order 1. We choose this threshold to be at $\eta > 0$, independent of μ, and pick η small enough for our purposes in the end.

Definition 2.20. For a resident population of $\mathbf{x} \subset \mathbb{H}^n$, the time when the first mutant type reaches $\eta > 0$ is defined as

$$\tilde{T}_\eta^\mu := \inf\{s \geq 0 : \exists\ y \in \mathbb{H}^n \backslash \mathbf{x} : \xi_s^\mu(y) > \eta\}. \tag{2.39}$$

To consider the evolutionary time scale $\ln 1/\mu$, we define T_η^μ through $\tilde{T}_\eta^\mu = T_\eta^\mu \ln 1/\mu$.

We can now state the first result that describes the evolution of the system until $\tilde{T}_\eta^\mu$.

Theorem 2.21. *Consider the system of differential equations (2.2) and assume (A). Then there exist $\tilde{\eta} > 0$ and $0 < \bar{c} \leq \bar{C}$, uniform in all $\boldsymbol{x} \subset \mathbb{H}^n$ for which $LVE_+(\boldsymbol{x}) = \{\bar{\xi}_{\boldsymbol{x}}\}$ and $(B_{\boldsymbol{x}})$ is satisfied, such that for $\eta \leq \tilde{\eta}$ and $\mu < \eta$ the following holds:*
If $\xi_0^\mu \in IC(\boldsymbol{x}, \eta, \bar{c})$, then, for every $0 < t_0 \leq t < \tilde{T}_\eta^\mu$ and every $y \in \mathbb{H}^n$,

$$\check{c} \sum_{z \in \mathbb{H}^n} \mathrm{e}^{t(f_{z,\boldsymbol{x}} - \eta\check{C})} \mu^{\rho_z + |z-y|} \leq \xi_t^\mu(y) \leq \hat{c} \sum_{z \in \mathbb{H}^n} \mathrm{e}^{t(f_{z,\boldsymbol{x}} + \eta\hat{C})} \mu^{\rho_z + |z-y|} (1+t)^m, \tag{2.40}$$

where $\rho_y := \min_{z \in \mathbb{H}^n}(\lambda_z + |z-y|)$, $m \in \mathbb{N}$, and $0 < \check{c}, \check{C}, \hat{c}, \hat{C} < \infty$ are independent of μ and η (but dependent on t_0).
Moreover, for all $x \in \boldsymbol{x}$,

$$\xi_t^\mu(x) \in [\bar{\xi}_{\boldsymbol{x}}(x) - \eta\bar{C}, \bar{\xi}_{\boldsymbol{x}}(x) + \eta\bar{C}]. \tag{2.41}$$

As a corollary, we estimate the growth of the different subpopulations on the time scale $\ln 1/\mu$ and derive the asymptotics of T_η^μ as $\mu \to 0$.

Corollary 2.22. *Under the same assumptions as in Theorem 2.21 and with the same constants, we obtain that, for every $y \in \mathbb{H}^n$ and every $t_0 \leq t \ln 1/\mu \leq \tilde{T}_\eta^\mu$,*

$$\check{c}\mu^{\min_{z \in \mathbb{H}^n}[\rho_z + |z-y| - t(f_{z,\boldsymbol{x}} - \eta\check{C})]} \leq \xi^\mu_{t \ln \frac{1}{\mu}}(y) \leq 2^n \hat{c} \mu^{\min_{z \in \mathbb{H}^n}[\rho_z + |z-y| - t(f_{z,\boldsymbol{x}} + \eta\hat{C})]} \left(1 + t \ln \frac{1}{\mu}\right)^m. \tag{2.42}$$

Moreover, as long as there is a $y \in \mathbb{H}^n$ for which $f_{y,\boldsymbol{x}} > 0$, there is an $\bar{\eta} \leq \tilde{\eta}$ such that for every $\eta \leq \bar{\eta}$

$$\min_{\substack{y\in\mathbb{H}^n\\ \lambda_y>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\boldsymbol{x}}>0}} \frac{\rho_z + |z-y|}{f_{z,\boldsymbol{x}} + \eta\hat{C}} \leq \liminf_{\mu\to 0} T_\eta^\mu \leq \limsup_{\mu\to 0} T_\eta^\mu \leq \min_{\substack{y\in\mathbb{H}^n\\ \lambda_y>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\boldsymbol{x}}>0}} \frac{\rho_z + |z-y|}{f_{z,\boldsymbol{x}} - \eta\check{C}}. \tag{2.43}$$

Proof of Theorem 2.21. The proof consists of two steps. We only derive the existence of $\tilde{\eta}$ for a specific set $\mathbf{x}$. To get a uniform parameter, we just have to minimise over the finite set of all such sets $\mathbf{x}$.

First, we show that (2.41) holds up to time $\tilde{T}_\eta^\mu$. Second, we inductively prove the upper bound in (2.40). The lower bound can derived analogously.

Step 1: $\xi_t^\mu(x) \in [\bar{\xi}_{\mathbf{x}}(x) - \eta\bar{C}, \bar{\xi}_{\mathbf{x}}(x) + \eta\bar{C}]$.
To prove our first claim, we analyse the distance of $\xi_t^\mu|_{\mathbf{x}} := (\xi_t^\mu(x))_{x\in\mathbf{x}}$ from $\bar{\xi}_{\mathbf{x}}$ with respect to the norm $\|\cdot\|_{\mathbf{x}}$, defined in (2.12). We prove that, in an annulus with respect to the norm $\|\cdot\|_{\mathbf{x}}$, this distance declines. Hence, starting inside the annulus, $\xi_t^\mu|_{\mathbf{x}}$ will remain there. This argument is depicted in Figure 2.1.

To approximate

$$\frac{d}{dt}\frac{\left\|\xi_t^\mu|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}^2}{2} = \left\langle \xi_t^\mu|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}, \frac{d}{dt}(\xi_t^\mu|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}})\right\rangle_{\mathbf{x}} \tag{2.44}$$

from above, we split the right hand side of (2.2) into two parts.

We define $F, V : \mathcal{M}(\mathbb{H}^n) \to \mathbb{R}^{\mathbf{x}}$,

$$F_x(\xi) = \left[r(x) - \sum_{y\in\mathbf{x}} \alpha(x,y)\xi(y)\right]\xi(x), \quad x \in \mathbf{x}, \tag{2.45}$$

the Lotka-Volterra part, and

$$V_x(\xi) = -\sum_{y\in\mathbb{H}^n\setminus\mathbf{x}} \alpha(x,y)\xi(y)\xi(x) + \mu\sum_{y\sim x} b(y)m(y,x)\xi(y) - \mu b(x)\xi(x), \quad x \in \mathbf{x}, \tag{2.46}$$

the error part of the differential equation.

With this,

$$\left\langle \xi_t^\mu|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}, \frac{d}{dt}(\xi_t^\mu|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}})\right\rangle_{\mathbf{x}} = \langle \xi_t^\mu|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}, F(\xi_t^\mu)\rangle_{\mathbf{x}} + \langle \xi_t^\mu|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}, V(\xi_t^\mu)\rangle_{\mathbf{x}}. \tag{2.47}$$

We first approximate the norm of the error part, using that $|\xi_t^\mu(y)| \leq \eta$ for $y \in \mathbb{H}^n\setminus\mathbf{x}$. In addition, we assume that, for every $x \in \mathbf{x}$, $\xi_t^\mu(x) \geq \eta$. We choose η such that this is always implied by (2.41) at the end of Step 1.

We estimate

$$|V_x(\xi_t^\mu)| \le \eta 2^n \max_{x\in\mathbf{x},y\in\mathbb{H}^n} \alpha(x,y)|\xi_t^\mu(x)| + \mu n \max_{y\in\mathbb{H}^n} b(y) \max_{y\in\mathbb{H}^n} |\xi_t^\mu(y)| + \mu \max_{y\in\mathbb{H}^n} b(y)|\xi_t^\mu(x)| \tag{2.48}$$

and hence, using that $\max_{y\in\mathbb{H}^n} |\xi_t^\mu(y)| \le \|\,\xi_t^\mu|_{\mathbf{x}}\| \le c_{\mathbf{x}}^{-1} \|\,\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}}$,

$$\begin{aligned}
\|V(\xi_t^\mu)\|_{\mathbf{x}} &\le \eta 2^n \max_{x\in\mathbf{x},y\in\mathbb{H}^n} \alpha(x,y) \|\,\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}} \\
&\quad + \mu \max_{y\in\mathbb{H}^n} b(y) \left(n\sqrt{|\mathbf{x}| \max_{x\in\mathbf{x}} \frac{\theta_x}{\bar\xi_{\mathbf{x}}(x)} \max_{y\in\mathbb{H}^n} |\xi_t^\mu(y)|^2} + \|\,\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}} \right) \\
&\le \eta 2^n \max_{x\in\mathbf{x},y\in\mathbb{H}^n} \alpha(x,y) \|\,\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}} + \mu \max_{y\in\mathbb{H}^n} b(y) \left(n\sqrt{|\mathbf{x}|} C_{\mathbf{x}} c_{\mathbf{x}}^{-1} + 1 \right) \|\,\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}} \\
&\le \eta \|\,\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}} C,
\end{aligned} \tag{2.49}$$

for some $C < \infty$ independent of η and μ.

Next, we approximate the Lotka-Volterra part. To do so, we show that a slight perturbation of the positive definite matrix $(\theta_x\alpha(x,y))_{x,y\in\mathbf{x}}$ is still positive definite. Let $\zeta \in \mathbb{R}^{\mathbf{x}}$ such that, for $x \in \mathbf{x}$, $|\zeta(x) - 1| \le \tilde\varepsilon_{\mathbf{x}}$. Then

$$\begin{aligned}
\sum_{x,y\in\mathbf{x}} \zeta(x)\theta_x\alpha(x,y)u(x)u(y) &= \sum_{x,y\in\mathbf{x}} \theta_x\alpha(x,y)u(x)u(y) + \sum_{x,y\in\mathbf{x}} (\zeta(x)-1)\theta_x\alpha(x,y)u(x)u(y) \\
&\ge \kappa \|u\|^2 - \max_{x\in\mathbf{x}} |\zeta(x)-1| \max_{x,y\in\mathbf{x}} (\theta_x\alpha(x,y)) \sum_{x,y\in\mathbf{x}} |u(x)||u(y)| \\
&\ge \|u\|^2 \left[\kappa - \tilde\varepsilon_{\mathbf{x}} |\mathbf{x}|^2 \max_{x,y\in\mathbf{x}} \theta_x\alpha(x,y)\right] \ge \frac{\kappa}{2} \|u\|^2,
\end{aligned} \tag{2.50}$$

as long as $\tilde\varepsilon_{\mathbf{x}} \le \kappa(2|\mathbf{x}|^2 \max_{x,y\in\mathbf{x}} \theta_x\alpha(x,y))^{-1}$.

We now apply this to $\zeta(x) = \xi_t^\mu(x)/\bar\xi_{\mathbf{x}}(x)$. The condition $|\zeta(x)-1| \le \tilde\varepsilon_{\mathbf{x}}$ is satisfied whenever

$$|\xi_t^\mu(x) - \bar\xi_{\mathbf{x}}(x)| \le \tilde\varepsilon_{\mathbf{x}} \bar\xi_{\mathbf{x}}(x), \tag{2.51}$$

which is the case if

$$\left\| \xi_t^\mu|_{\mathbf{x}} - \bar\xi_{\mathbf{x}} \right\|_{\mathbf{x}} \le c_{\mathbf{x}} \tilde\varepsilon_{\mathbf{x}} \min_{x\in\mathbf{x}} \bar\xi_{\mathbf{x}}(x) =: \varepsilon_{\mathbf{x}}. \tag{2.52}$$

Using the fact that $\bar\xi_{\mathbf{x}}$ is an equilibrium of (2.4) for which $\bar\xi_{\mathbf{x}}(x) > 0$ holds for all $x \in \mathbf{x}$, we derive

$$\begin{aligned}
\langle \xi_t^\mu|_{\mathbf{x}} - \bar\xi_{\mathbf{x}}, F(\xi_t^\mu)\rangle_{\mathbf{x}} &= \sum_{x\in\mathbf{x}} \frac{\theta_x}{\bar\xi_{\mathbf{x}}(x)} (\xi_t^\mu(x) - \bar\xi_{\mathbf{x}}(x)) \left[r(x) - \sum_{y\in\mathbf{x}} \alpha(x,y)\xi_t^\mu(y) \right] \xi_t^\mu(x) \\
&= -\sum_{x\in\mathbf{x}} \frac{\theta_x}{\bar\xi_{\mathbf{x}}(x)} (\xi_t^\mu(x) - \bar\xi_{\mathbf{x}}(x)) \left[\sum_{y\in\mathbf{x}} \alpha(x,y)(\xi_t^\mu(y) - \bar\xi_{\mathbf{x}}(y)) \right] \xi_t^\mu(x) \\
&= -\sum_{x,y\in\mathbf{x}} \frac{\xi_t^\mu(x)}{\bar\xi_{\mathbf{x}}(x)} \theta_x\alpha(x,y)(\xi_t^\mu(x) - \bar\xi_{\mathbf{x}}(x))(\xi_t^\mu(y) - \bar\xi_{\mathbf{x}}(y)) \\
&\le -\frac{\kappa}{2} \left\| \xi_t^\mu|_{\mathbf{x}} - \bar\xi_{\mathbf{x}} \right\|^2.
\end{aligned} \tag{2.53}$$

Combining estimates (2.49) and (2.53), we get

$$\begin{aligned}\frac{d}{dt}\frac{\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}^2}{2} &= \langle \xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}, F(\xi_t^\mu)\rangle_{\mathbf{x}} + \langle \xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}, V(\xi_t^\mu)\rangle_{\mathbf{x}} \\ &\leq -\frac{\kappa}{2}\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|^2 + \left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} \|V(\xi_t^\mu)\|_{\mathbf{x}} \\ &\leq -\frac{\kappa}{2}\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|^2 + \left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} \eta \,\|\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}} C \\ &\leq -\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}^2 \left(\frac{\kappa}{2C_{\mathbf{x}}^2} - \eta\frac{C\,\|\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}}}{\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}}\right) \\ &\leq -\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}^2 \frac{\kappa}{4C_{\mathbf{x}}^2} < 0, \end{aligned} \tag{2.54}$$

whenever

$$\varepsilon_{\mathbf{x}} \geq \left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} \geq \eta C\,\|\xi_t^\mu|_{\mathbf{x}}\|_{\mathbf{x}}\frac{4C_{\mathbf{x}}^2}{\kappa} \geq \eta C(\left\|\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}-\varepsilon_{\mathbf{x}})\frac{4C_{\mathbf{x}}^2}{\kappa} =: \eta\underline{c}. \tag{2.55}$$

Finally, we choose $\tilde{\eta}$ small enough such that $\tilde{\eta} < \varepsilon_{\mathbf{x}}/\underline{c}$.

Now we can follow the argument that was outlined in the beginning and is supported by Figure 2.1. As long as $\eta \leq \tilde{\eta}$ and

$$\left\|\xi_0^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\| \leq \eta\underline{c}C_{\mathbf{x}}^{-1} =: \eta\bar{c}_{\mathbf{x}}, \tag{2.56}$$

we obtain that $\left\|\xi_0^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} \leq \eta c$. Because of (2.54), we obtain that $\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} \leq \eta\underline{c}$, for every $0 \leq t \leq \tilde{T}_\eta^\mu$, and hence

$$\left\|\xi_t^\mu|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\| \leq \eta\underline{c}c_{\mathbf{x}}^{-1} =: \eta\bar{C}_{\mathbf{x}}. \tag{2.57}$$

For the single types, this implies, for every $0 \leq t \leq \tilde{T}_\eta^\mu$, that

$$\xi_t^\mu(x) \in [\bar{\xi}_{\mathbf{x}}(x) - \eta\bar{C}_{\mathbf{x}}, \bar{\xi}_{\mathbf{x}}(x) + \eta\bar{C}_{\mathbf{x}}], \quad x \in \mathbf{x}, \tag{2.58}$$

whenever

$$\xi_0^\mu(x) \in \left[\bar{\xi}_{\mathbf{x}}(x) - \eta\frac{\bar{c}_{\mathbf{x}}}{\sqrt{|\mathbf{x}|}}, \bar{\xi}_{\mathbf{x}}(x) + \eta\frac{\bar{c}_{\mathbf{x}}}{\sqrt{|\mathbf{x}|}}\right], \quad x \in \mathbf{x}. \tag{2.59}$$

Setting $\bar{c} := \min_{\mathbf{y}\subset\mathbb{H}^n} \bar{c}_{\mathbf{y}}$ and $\bar{C} := \max_{\mathbf{y}\subset\mathbb{H}^n} \bar{C}_{\mathbf{y}}$, and choosing $\tilde{\eta} \leq \min_{x\in\mathbf{x}} \bar{\xi}_{\mathbf{x}}(x)/(2\bar{C}+2)$ to ensure that $\xi_t^\mu(x) > \eta$, for every $x \in \mathbf{x}$, we arrive at the claim.

Step 2: Inductive exponential bounds.
We derive the upper bound for $\xi_t^\mu(y)$ in (2.40) in full length. At the end of the proof, we comment on how the same strategy can be adapted to the lower bound.

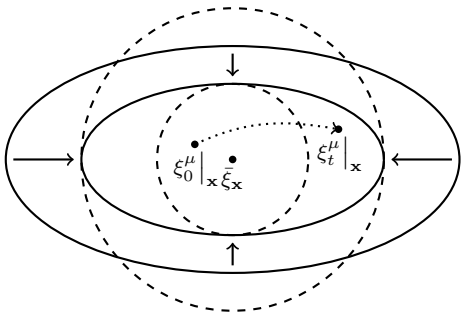


Figure 2.1: Scheme for the argument in Step 1. Dashed lines indicate balls $B(\bar{\xi}_{\mathbf{x}}, \eta\bar{c}_{\mathbf{x}})$ and $B(\bar{\xi}_{\mathbf{x}}, \eta\bar{C}_{\mathbf{x}})$ with respect to the standard Euclidean norm, while solid lines correspond to balls $B_{\mathbf{x}}(\bar{\xi}_{\mathbf{x}}, \eta c)$ and $B_{\mathbf{x}}(\bar{\xi}_{\mathbf{x}}, \varepsilon_{\mathbf{x}})$ with respect to the $\|\cdot\|_{\mathbf{x}}$ norm.

To begin, we establish an upper bound on $\frac{d}{dt}\xi_t^\mu$.

$$\begin{aligned}\tfrac{d}{dt}\xi_t^\mu(y) &\le \Big[r(y) - \sum_{x\in\mathbf{x}} \alpha(y,x)(\bar{\xi}_{\mathbf{x}}(x) - \eta\bar{C})\Big]\xi_t^\mu(y) + \mu \sum_{z\sim y} \underbrace{b(z)m(z,y)}_{\le \tilde{C}_y \forall z\sim y}\xi_t^\mu(z)\\ &\le \Big[r(y) - \sum_{x\in\mathbf{x}} \alpha(y,x)\bar{\xi}_{\mathbf{x}}(x) + \eta\bar{C}\underbrace{\sum_{x\in\mathbf{x}} \alpha(y,x)}_{=:\hat{C}_y}\Big]\xi_t^\mu(y) + \mu\tilde{C}_y \sum_{z\sim y}\xi_t^\mu(z)\\ &\le [f_{y,\mathbf{x}} + \eta\hat{C}]\xi_t^\mu(y) + \mu\tilde{C}\sum_{z\sim y}\xi_t^\mu(z), \end{aligned} \tag{2.60}$$

where $\hat{C} := \max_{y\in\mathbb{H}^n} \hat{C}_y < \infty$ and $\tilde{C} := \max_{y\in\mathbb{H}^n} \tilde{C}_y < \infty$.

We prove by induction that, for every $m \ge 0$, there exists a constant $C_m < \infty$, independent of μ, η, and y, such that, for every $0 \le t \le \tilde{T}_\eta^\mu$,

$$\xi_t^\mu(y) \le C_m\Bigg[\sum_{\substack{z\in\mathbb{H}^n\\ |z-y|\le m}} \mathrm{e}^{t(f_{z,\mathbf{x}}+\eta\hat{C})}\Big(\mu^{\rho_z+|z-y|} + \frac{1}{\eta}\mu^{m+1}\Big)(1+t)^m + \mu^{m+1}\Bigg]. \tag{2.61}$$

For the case $m = 0$, we approximate

$$\tfrac{d}{dt}\xi_t^\mu(y) \le [f_{y,\mathbf{x}} + \eta\hat{C}]\xi_t^\mu(y) + \mu\tilde{C}\underbrace{\sum_{z\sim y} \mathbb{1}_{z\in\mathbf{x}}(\bar{\xi}_{\mathbf{x}}(z) + \eta\bar{C}) + \mathbb{1}_{z\in\mathbb{H}^n\setminus\mathbf{x}}\eta}_{\le C \text{ uniformly in } y,z}, \tag{2.62}$$

and hence

$$\begin{aligned}\xi_t^\mu(y) &\le \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}\xi_0^\mu(y) + \mu\tilde{C}C\int_0^t \mathrm{e}^{(t-s)(f_{y,\mathbf{x}}+\eta\hat{C})}ds\\ &\le \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}C_y\mu^{\lambda_y} + \mu\tilde{C}C\frac{1}{f_{y,\mathbf{x}}+\eta\hat{C}}(\mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})} - 1). \end{aligned} \tag{2.63}$$

Choose $\tilde{\eta} > 0$ small enough such that $f_{y,\mathbf{x}} + \tilde{\eta}\hat{C} < 0$, for every $y \in \mathbb{H}^n$ for which $f_{y,\mathbf{x}} < 0$. Then, for $\eta \le \tilde{\eta}$ and a different constant $C < \infty$, the second summand can be bounded

from above by $C\mu$, for $f_{y,\mathbf{x}} < 0$, and by $C/\eta \cdot \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}\mu$, for $f_{y,\mathbf{x}} \geq 0$. C can be chosen independent of y, μ, $\eta \leq \tilde{\eta}$, and $0 \leq t \leq \tilde{T}^{\mu}_{\eta}$. Overall, using $\lambda_y \geq \rho_y$, we get

$$\xi^{\mu}_t(y) \leq \underbrace{((\max_{y\in\mathbb{H}^n} C_y) \vee C)}_{=:C_0<\infty} \Big[\mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}\Big(\mu^{\rho_y} + \frac{1}{\eta}\mu\Big) + \mu\Big], \tag{2.64}$$

which is the desired bound.
Assuming that the hypothesis holds for $m-1$ and using (2.60), we approximate

$$\begin{aligned}\tfrac{d}{dt}\xi^{\mu}_t(y) &\leq [f_{y,\mathbf{x}} + \eta\hat{C}]\xi^{\mu}_t(y) \\ &\quad + \mu\tilde{C}\sum_{z\sim y} C_{m-1}\Bigg[\sum_{\substack{u\in\mathbb{H}^n \\ |u-z|\leq m-1}} \mathrm{e}^{t(f_{u,\mathbf{x}}+\eta\hat{C})}\Big(\mu^{\rho_u+|u-z|} + \frac{1}{\eta}\mu^m\Big)(1+t)^{m-1} + \mu^m\Bigg].\end{aligned} \tag{2.65}$$

Splitting up the second summand, Gronwall's inequality yields

$$\begin{aligned}\xi^{\mu}_t(y) &\leq \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}\xi^{\mu}_0(y) + \tilde{C}C_{m-1}n\mu^{m+1}\int_0^t \mathrm{e}^{(t-s)(f_{y,\mathbf{x}}+\eta\hat{C})}ds \\ &\quad + \tilde{C}C_{m-1}\sum_{z\sim y}\sum_{\substack{u\in\mathbb{H}^n \\ |u-z|\leq m-1}} \Big(\mu^{\rho_u+|u-z|+1} + \frac{1}{\eta}\mu^{m+1}\Big) \\ &\quad\quad \cdot \int_0^t (1+s)^{m-1}\mathrm{e}^{s(f_{u,\mathbf{x}}+\eta\hat{C})}\mathrm{e}^{(t-s)(f_{y,\mathbf{x}}+\eta\hat{C})}ds \\ &\leq \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}C_y\mu^{\lambda_y} + C\mu^{m+1}\Big(1 + \frac{1}{\eta}\mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}\Big) \\ &\quad + \tilde{C}C_{m-1}\sum_{z\sim y}\sum_{\substack{u\in\mathbb{H}^n \\ |u-z|\leq m-1}} \Big(\mu^{\rho_u+|u-z|+1} + \frac{1}{\eta}\mu^{m+1}\Big)(1+t)^{m-1} \\ &\quad\quad \cdot \int_0^t \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}\mathrm{e}^{s(f_{u,\mathbf{x}}-f_{y,\mathbf{x}})}ds,\end{aligned} \tag{2.66}$$

where we bound the first integral just as before in the base case.

We distinguish two cases to approximate the second integral. If $f_{u,\mathbf{x}} \neq f_{y,\mathbf{x}}$, then

$$\begin{aligned}\int_0^t \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\check{C})}\mathrm{e}^{s(f_{u,\mathbf{x}}-f_{y,\mathbf{x}})}ds &= \frac{1}{f_{u,\mathbf{x}} - f_{y,\mathbf{x}}}(\mathrm{e}^{t(f_{u,\mathbf{x}}+\eta\check{C})} - \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\check{C})}) \\ &= \frac{1}{|f_{u,\mathbf{x}} - f_{y,\mathbf{x}}|}|\mathrm{e}^{t(f_{u,\mathbf{x}}+\eta\check{C})} - \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}| \\ &\leq C'(\mathrm{e}^{t(f_{u,\mathbf{x}}+\eta\check{C})} + \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\check{C})}),\end{aligned} \tag{2.67}$$

for some $C' < \infty$ large enough, uniformly in y and u.
If $f_{u,\mathbf{x}} = f_{y,\mathbf{x}}$, then

$$\int_0^t \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\check{C})}\mathrm{e}^{s(f_{u,\mathbf{x}}-f_{y,\mathbf{x}})}ds = t\mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\check{C})}. \tag{2.68}$$

Plugging this back into (2.66) we get

$$\begin{aligned}\xi_t^\mu(y) \le& \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}C_y\mu^{\lambda_y} + C\mu^{m+1}\Big(1+\frac{1}{\eta}\mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}\Big)\\ &+\tilde{C}C_{m-1}\sum_{z\sim y}\sum_{\substack{u\in\mathbb{H}^n\\ |u-z|\le m-1}}\Big(\mu^{\rho_u+|u-z|+1}+\frac{1}{\eta}\mu^{m+1}\Big)(1+t)^{m-1}\\ &\cdot C'(1+t)(\mathrm{e}^{t(f_{u,\mathbf{x}}+\eta\hat{C})}+\mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})})\\ \le& \underbrace{(n+o(1))((\max_{y\in\mathbb{H}^n}C_y)\vee C\vee\tilde{C}C_{m-1}C')}_{\le C_m \text{ for } \mu<\tilde{\eta}}\\ &\cdot\Bigg[\sum_{\substack{z\in\mathbb{H}^n\\ |z-y|\le m}}\mathrm{e}^{t(f_{z,\mathbf{x}}+\eta\hat{C})}\Big(\mu^{\rho_z+|z-y|}+\frac{1}{\eta}\mu^{m+1}\Big)(1+t)^m+\mu^{m+1}\Bigg],\end{aligned}\tag{2.69}$$

where we used that $\rho_y \le (\rho_u + |u-z|+1)\wedge\lambda_y$ for all $z\sim y$ and $|u-z|\le m-1$, and gathered all the higher μ-powers in the $o(1)$ with respect to the limit $\mu\to 0$. This concludes the proof of (2.61).
Finally, we can choose $m \ge \max_{y\in\mathbb{H}^n}\max_{z\in\mathbb{H}^n}\rho_z + |z-y| \ge n$ and, since $f_{z,\mathbf{x}} = 0$ for all $z\in\mathbf{x}$, we get

$$\begin{aligned}\xi_t^\mu(y) &\le C_m\Bigg[\sum_{z\in\mathbb{H}^n}\mathrm{e}^{t(f_{z,\mathbf{x}}+\eta\hat{C})}\Big(\mu^{\rho_z+|z-y|}+\frac{1}{\eta}\mu^{m+1}\Big)(1+t)^m+\mu^{m+1}\Bigg]\\ &\le C_m\Bigg[\sum_{z\in\mathbb{H}^n}\mathrm{e}^{t(f_{z,\mathbf{x}}+\eta\hat{C})}(\mu^{\rho_z+|z-y|}+\mu^m)(1+t)^m+\sum_{z\in\mathbf{x}}\mathrm{e}^{t(f_{z,\mathbf{x}}+\eta\hat{C})}\mu^{m+1}\Bigg]\\ &\le 3C_m\sum_{z\in\mathbb{H}^n}\mathrm{e}^{t(f_{z,\mathbf{x}}+\eta\hat{C})}\mu^{\rho_z+|z-y|}(1+t)^m.\end{aligned}\tag{2.70}$$

With $\hat{c} := 3C_m$ and choosing $\tilde{\eta}$ uniform over all subsets $\mathbf{x}\subset\mathbb{H}^n$ of coexisting resident types, this yields the desired upper bound.

The proof of the lower bound is very similar. We approximate, for every $y\in\mathbb{H}^n$,

$$\tfrac{d}{dt}\xi_t^\mu(y) \ge [f_{y,\mathbf{x}} - \eta\check{C}]\xi_t^\mu(y) + \mu\tilde{c}\sum_{z\sim y}\xi_t^\mu(z),\tag{2.71}$$

and then use the inductive application of Gronwall's inequality twice.

First, to prove that, for an arbitrarily small $t_0 > 0$, $\xi_{t_0/2}^\mu(y) \ge c_{t_0}\mu^{\rho_y}$, where $c_{t_0} > 0$ can be chosen uniformly in μ, η, and y. This corresponds to mutation producing a positive population size for every type within a time of order 1.

Second, we show that, for every $0\le m\le n$, there exists a constant $c_m > 0$, independent of μ, η, and y, such that, for $(n+m)t_0/(2n) \le t \le \tilde{T}_\eta^\mu$,

$$\xi_t^\mu(y) \ge c_m\sum_{\substack{z\in\mathbb{H}^n\\ |z-y|\le m}}\mu^{\rho_z+|z-y|}\mathrm{e}^{t(f_{z,\mathbf{x}}-\eta\check{C})}.\tag{2.72}$$

Setting $\check{c} := c_n$ yields the lower bound in (2.40), for $t_0 \le t \le \tilde{T}_\eta^\mu$. □

Proof of Corollary 2.22. The inequalities in (2.42) follow directly from (2.40) by inserting the new time scale. For the lower bound, only the asymptotically largest summand, corresponding to the smallest μ-power, is kept. For the upper bound, every one of the 2^n summands is estimated against this largest one.

To prove the second part of the corollary, we first show that, for μ small enough, the first non-resident type y that reaches the η-threshold, i.e. the type that determines the stopping time T_η^μ, satisfies $\lambda_y > 0$ and hence $\rho_z + |z-y| > 0$, for every $z \in \mathbb{H}^n$.
Let $y \in \mathbb{H}^n \backslash \mathbf{x}$ be a non-resident type for which $\lambda_y = 0$. This implies $\xi_0^\mu(y) \leq \eta/3$ and $f_{y,\mathbf{x}} < 0$. Going back into the proof of (2.61) and using that $\tilde{\eta}$ is chosen such that $f_{y,\mathbf{x}} + \tilde{\eta}\hat{C} < 0$, this yields

$$\begin{aligned}
\xi_t^\mu(y) &\leq \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})} C_y \mu^{\lambda_y} + \mu\tilde{C}C \frac{1}{f_{y,\mathbf{x}}+\eta\hat{C}} (\mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})} - 1) \\
&\leq \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})} \frac{\eta}{3} + \mu\tilde{C}C \frac{1}{|f_{y,\mathbf{x}}+\eta\hat{C}|} (1 - \mathrm{e}^{t(f_{y,\mathbf{x}}+\eta\hat{C})}) \\
&\leq \frac{\eta}{3} + \frac{\mu\tilde{C}C}{|f_{y,\mathbf{x}}+\tilde{\eta}\hat{C}|} \leq \frac{2}{3}\eta,
\end{aligned} \tag{2.73}$$

whenever $\mu \leq \eta |f_{y,\mathbf{x}} + \tilde{\eta}\hat{C}|/3\tilde{C}C$. As a consequence, as $\mu \to 0$, y stays strictly below η and does not determine T_η^μ.

Now we assume that T_η^μ is determined by a non-resident type $y \in \mathbb{H}^n$ for which $\lambda_y > 0$, i.e. y is the first mutant to reach the η-threshold. Let $\bar{\eta} \leq \tilde{\eta} \wedge 1 \wedge \check{c}$. Then, assuming that $0 < \mu \leq \eta \leq \bar{\eta}$, the lower bound in (2.42) yields

$$\check{c}\mu^{\min_{z\in\mathbb{H}^n}[\rho_z+|z-y|-T_\eta^\mu(f_{z,\mathbf{x}}-\eta\check{C})]} \leq \xi_{T_\eta^\mu}^\mu(y) = \eta, \tag{2.74}$$

and hence

$$\ln(\mu) \min_{z\in\mathbb{H}^n} [\rho_z + |z-y| - T_\eta^\mu(f_{z,\mathbf{x}} - \eta\check{C})] \leq \ln\left(\frac{\eta}{\check{c}}\right) \leq 0. \tag{2.75}$$

Since $\ln(\mu) < 0$, we obtain, for every $z \in \mathbb{H}^n$, that

$$\rho_z + |z-y| \geq T_\eta^\mu(f_{z,\mathbf{x}} - \eta\check{C}), \tag{2.76}$$

and therefore, if we choose $\bar{\eta}$ small enough such that, for every $\eta \leq \bar{\eta}$ and every $z \in \mathbb{H}^n$ for which $f_{z,\mathbf{x}} > 0$, also $f_{z,\mathbf{x}} - \eta\check{C} > 0$,

$$T_\eta^\mu \leq \min_{\substack{z\in\mathbb{H}^n \\ f_{z,\mathbf{x}}>0}} \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}} - \eta\check{C}}. \tag{2.77}$$

To get a lower bound for T_η^μ, (2.42) implies

$$\eta = \xi_{T_\eta^\mu}^\mu(y) \leq 2^n \hat{c}\mu^{\min_{z\in\mathbb{H}^n}[\rho_z+|z-y|-T_\eta^\mu(f_{z,\mathbf{x}}+\eta\hat{C})]} \left(1 + \tilde{T}_\eta^\mu\right)^m, \tag{2.78}$$

which yields

$$\ln(\mu) \min_{z\in\mathbb{H}^n}[\rho_z + |z-y| - T_\eta^\mu(f_{z,\mathbf{x}} + \eta\hat{C})] \geq \ln\left(\frac{\eta}{2^n\hat{c}(1+\tilde{T}_\eta^\mu)^m}\right), \tag{2.79}$$

and therefore there exists a $z \in \mathbb{H}^n$ such that

$$\rho_z + |z-y| \leq T_\eta^\mu(f_{z,\mathbf{x}} + \eta\hat{C}) + \frac{\ln\left(\frac{2^n\hat{c}}{\eta}\right) + m\ln(1+\tilde{T}_\eta^\mu)}{\ln\frac{1}{\mu}}. \tag{2.80}$$

The second summand on the right hand side is positive and, with (2.77), converges to zero as $\mu \to 0$. Since the left hand side is positive this implies that $f_{z,\mathbf{x}} + \eta\hat{C} > 0$ and by our choice of $\tilde{\eta}$ in the proof of (2.61) we obtain $f_{z,\mathbf{x}} \geq 0$.
Consequently, for every fixed $0 < \eta \leq \bar{\eta}$, it follows that

$$\liminf_{\mu\to 0} T_\eta^\mu \geq \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}} + \eta\check{C}} \geq \min_{\substack{z\in\mathbb{H}^n \\ f_{z,\mathbf{x}}\geq 0}} \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}} + \eta\check{C}}. \tag{2.81}$$

Overall, for every fixed $0 < \eta \leq \bar{\eta}$, we obtain

$$\min_{\substack{z\in\mathbb{H}^n \\ f_{z,\mathbf{x}}\geq 0}} \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}} + \eta\check{C}} \leq \liminf_{\mu\to 0} T_\eta^\mu \leq \limsup_{\mu\to 0} T_\eta^\mu \leq \min_{\substack{z\in\mathbb{H}^n \\ f_{z,\mathbf{x}}>0}} \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}} - \eta\check{C}}. \tag{2.82}$$

If we now pick $\bar{\eta}$ small enough, both minima are realised by the same $z \in \mathbb{H}^n$ for which $f_{z,\mathbf{x}} > 0$, that also minimise

$$\min_{\substack{z\in\mathbb{H}^n \\ f_{z,\mathbf{x}}>0}} \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}}}, \tag{2.83}$$

and we can reduce to only considering $z \in \mathbb{H}^n$ such that $f_{z,\mathbf{x}} > 0$ in the lower bound.

All the above considerations apply to a single y for which $\lambda_y > 0$. Considering all such $y \in \mathbb{H}^n$ we get that asymptotically

$$\min_{\substack{y\in\mathbb{H}^n \\ \lambda_y>0}} \min_{\substack{z\in\mathbb{H}^n \\ f_{z,\mathbf{x}}>0}} \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}} + \eta\check{C}} \leq \liminf_{\mu\to 0} T_\eta^\mu \leq \limsup_{\mu\to 0} T_\eta^\mu \leq \min_{\substack{y\in\mathbb{H}^n \\ \lambda_y>0}} \min_{\substack{z\in\mathbb{H}^n \\ f_{z,\mathbf{x}}>0}} \frac{\rho_z + |z-y|}{f_{z,\mathbf{x}} - \eta\check{C}}. \tag{2.84}$$

For the upper bound, the minimum can be used since, if T_η^μ was larger than this minimum, the minimiser would reach the η-level before $\tilde{T}_\eta^\mu$, which would be a contradiction.
This finishes the proof of the corollary. □

2.4 Construction of the Jump Process

In this section we combine the results of Theorem 2.21, or rather Corollary 2.22, and Theorem 2.5 to derive the convergence of ξ^μ as $\mu \to 0$ to a jump process that moves between Lotka-Volterra equilibria of coexistence. We prove the convergence by an induction over the invasion steps and show that after each invasion the criteria for the initial conditions in Theorem 2.21 are again satisfied.

Before we get to the actual proof, we derive two lemmas. The first lemma treats the boundedness of solutions of (2.2), the continuity in the initial condition, and the perturbation through the mutation rate μ.

Lemma 2.23. *Let*

$$\Omega := \left\{ \xi \in \mathcal{M}(\mathbb{H}^n) : \forall x \in \mathbb{H}^n : \xi(x) \in \left[0, 2\frac{|r(x)|}{\alpha(x,x)}\right] \right\}. \tag{2.85}$$

There is a $\mu_0 > 0$ such that, for every $0 \le \mu < \mu_0$, for every $\xi_0^\mu \in \Omega$, and for every $t \ge 0$, we obtain $\xi_t^\mu \in \Omega$, where ξ_t^μ is the solution of (2.2).

Moreover, there are positive, finite constants A, B such that, for every $0 \le \mu_1, \mu_2 < \mu_0$, for every $\xi_0^{\mu_1}, \xi_0^{\mu_2} \in \Omega$, and every $t \ge s \ge 0$,

$$\|\xi_t^{\mu_1} - \xi_t^{\mu_2}\| \le \mathrm{e}^{(t-s)A} \left(\|\xi_s^{\mu_1} - \xi_s^{\mu_2}\| + \sqrt{(\mu_1+\mu_2)\frac{B}{A}} \right). \tag{2.86}$$

Proof. To prove the first claim, assume that $\xi_t^\mu \in \Omega$ and $\xi_t^\mu(x) = 2|r(x)|/\alpha(x,x)$, for some $x \in \mathbb{H}^n$. Then

$$\begin{aligned} \tfrac{d}{dt}\xi_t^\mu(x) &\le [r(x) - \alpha(x,x)\xi_t^\mu(x)]\xi_t^\mu(x) + \mu \sum_{y\sim x} b(y)m(y,x)\xi_t^\mu(y) \\ &\le \frac{-2|r(x)|^2}{\alpha(x,x)} + \mu 2n \max_{y\in\mathbb{H}^n} \frac{b(y)|r(y)|}{\alpha(y,y)} < 0, \end{aligned} \tag{2.87}$$

for

$$\mu < \mu_0 := \min_{y\in\mathbb{H}^n} \frac{2|r(y)|^2}{\alpha(y,y)} \left(2n \max_{y\in\mathbb{H}^n} \frac{b(y)|r(y)|}{\alpha(y,y)} \right)^{-1}. \tag{2.88}$$

Hence, ξ_t^μ cannot leave Ω.

For the second claim, we approximate

$$\begin{aligned}
&\frac{d}{dt}\frac{\|\xi_t^{\mu_1}-\xi_t^{\mu_2}\|^2}{2}\\
&\quad=\sum_{x\in\mathbb{H}^n}(\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x))r(x)(\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x))\\
&\qquad-\sum_{x\in\mathbb{H}^n}(\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x))\sum_{y\in\mathbb{H}^n}\alpha(x,y)(\xi_t^{\mu_1}(x)\xi_t^{\mu_1}(y)-\xi_t^{\mu_2}(x)\xi_t^{\mu_2}(y))\\
&\qquad+\sum_{x\in\mathbb{H}^n}(\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x))\mu_1\left(\sum_{y\sim x}b(y)m(y,x)\xi_t^{\mu_1}(y)-b(x)\xi_t^{\mu_1}(x)\right)\\
&\qquad-\sum_{x\in\mathbb{H}^n}(\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x))\mu_2\left(\sum_{y\sim x}b(y)m(y,x)\xi_t^{\mu_2}(y)-b(x)\xi_t^{\mu_2}(x)\right)\\
&\quad\le\max_{x\in\mathbb{H}^n}|r(x)|\sum_{x\in\mathbb{H}^n}(\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x))^2\\
&\qquad-\sum_{x\in\mathbb{H}^n}\sum_{y\in\mathbb{H}^n}\alpha(x,y)(\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x))^2\xi_t^{\mu_1}(y)\\
&\qquad+\sum_{x\in\mathbb{H}^n}\sum_{y\in\mathbb{H}^n}\alpha(x,y)|\xi_t^{\mu_1}(x)-\xi_t^{\mu_2}(x)|\cdot|\xi_t^{\mu_1}(y)-\xi_t^{\mu_2}(y)|\cdot|\xi_t^{\mu_2}(x)|\\
&\qquad+\mu_1\max_{x\in\mathbb{H}^n}b(x)\sum_{x\in\mathbb{H}^n}\max_{x\in\mathbb{H}^n}(|\xi_t^{\mu_1}(x)|+|\xi_t^{\mu_2}(x)|)\max_{x\in\mathbb{H}^n}|\xi_t^{\mu_1}(x)|\left(\sum_{y\sim x}m(y,x)+1\right)\\
&\qquad+\mu_2\max_{x\in\mathbb{H}^n}b(x)\sum_{x\in\mathbb{H}^n}\max_{x\in\mathbb{H}^n}(|\xi_t^{\mu_1}(x)|+|\xi_t^{\mu_2}(x)|)\max_{x\in\mathbb{H}^n}|\xi_t^{\mu_2}(x)|\left(\sum_{y\sim x}m(y,x)+1\right),
\end{aligned}\tag{2.89}$$

which implies

$$\begin{aligned}
\frac{d}{dt}\frac{\|\xi_t^{\mu_1}-\xi_t^{\mu_2}\|^2}{2}&\le\|\xi_t^{\mu_1}-\xi_t^{\mu_2}\|^2\Big[\max_{x\in\mathbb{H}^n}|r(x)|+2^{2n}\max_{x,y\in\mathbb{H}^n}\alpha(x,y)\,\|\xi_t^{\mu_2}\|\Big]\\
&\qquad+(\mu_1+\mu_2)(2^n\cdot 2)\max_{x\in\mathbb{H}^n}b(x)(\|\xi_t^{\mu_1}\|+\|\xi_t^{\mu_2}\|)^2\\
&=:\|\xi_t^{\mu_1}-\xi_t^{\mu_2}\|^2A+(\mu_1+\mu_2)B,
\end{aligned}\tag{2.90}$$

where A and B depend on b, r, α, and can be chosen uniformly in $t\ge 0$, $0\le\mu_i<\mu_0$, and initial values $\xi_0^{\mu_i}\in\Omega$ since $\|\xi_t^{\mu_i}\|\le\max_{\xi\in\Omega}\|\xi\|<\infty$. Applying Gronwall's inequality and taking the square root implies the claim. □

Theorem 2.21 and Corollary 2.22 provide us with approximations for ξ_t^μ during the exponential growth phase and Theorem 2.5 guarantees convergence to a new equilibrium during the invasion phase. To show that this second phase vanishes on the time scale $\ln 1/\mu$, we need to bound its duration uniformly in the approximate state of the system at its beginning.

We introduce the following notation for the time until the initial conditions for the next growth phase are reached.

Definition 2.24.

$$\tilde{\tau}^{\mu}_{\eta}(\xi,\mathbf{x}) := \inf\{t\geq 0 : \forall\, x\in\mathbf{x} : |\xi^{\mu}_t(x)-\bar{\xi}_{\mathbf{x}}(x)|\leq \eta\frac{\bar{c}}{\sqrt{|\mathbf{x}|}},\\ \forall\, y\in\mathbb{H}^n\backslash\mathbf{x} : \xi^{\mu}_t(y)\leq \frac{\eta}{3}; \xi^{\mu}_0=\xi\}, \tag{2.91}$$

In the proof of Theorem 2.12, we approximate the true system, solving (2.2), by the mutation-free Lotka-Volterra system during the invasion. The second lemma proves continuity in the initial condition for a slight variation of $\tilde{\tau}^{\mu}_{\eta}(\xi,\mathbf{x})$, corresponding to the case of $\mu=0$.

Lemma 2.25. *Let $\boldsymbol{y}\subset\mathbb{H}^n$ such that $r(y)>0$, for all $y\in\boldsymbol{y}$, and ($B_{\boldsymbol{y}}$) is satisfied. Let $\boldsymbol{x}\subset\boldsymbol{y}$ such that the equilibrium state of the Lotka-Volterra system involving types $\boldsymbol{y}$ is supported on $\boldsymbol{x}$ and assume $f_{y,\boldsymbol{x}}<0$, for every $y\in\boldsymbol{y}\backslash\boldsymbol{x}$. Define*

$$\bar{\tau}^{0}_{\eta}(\xi,\boldsymbol{x},\boldsymbol{y}) := \inf\{t\geq 0 : \left\|\xi^0_t\right|_{\boldsymbol{x}}-\bar{\xi}_{\boldsymbol{x}}\right\|_{\boldsymbol{x}}\leq \frac{\eta\bar{c}c_{\boldsymbol{x}}}{2\sqrt{|\boldsymbol{x}|}},\\ \forall\, y\in\boldsymbol{y}\backslash\boldsymbol{x} : \xi^0_t(y)\leq \frac{\eta}{6}\wedge\hat{\eta}; \xi^0_0=\xi\}, \tag{2.92}$$

where $\|\cdot\|_{\boldsymbol{x}}$ is the norm defined in (2.12), corresponding to $\bar{\xi}_{\boldsymbol{x}}$, and $\hat{\eta}:=\eta\bar{c}c_{\boldsymbol{x}}/(2\sqrt{|\boldsymbol{x}|}\underline{c})$. Then, for η small enough, $\bar{\tau}^0_{\eta}(\xi,\boldsymbol{x},\boldsymbol{y})$ is continuous in $\xi\in(\mathbb{R}_{>0})^{\boldsymbol{y}}\times\{0\}^{\mathbb{H}^n\backslash\boldsymbol{y}}$.

Remark 2.26. Theorem 2.5 ensures that the Lotka-Volterra system involving the types $\mathbf{y}$ converges to a unique equilibrium and hence $\mathbf{x}$ in Lemma 2.25 is uniquely determined.

Proof. Since we are considering the case of $\mu=0$, we obtain $\xi^0_t\in(\mathbb{R}_{>0})^{\mathbf{y}}\times\{0\}^{\mathbb{H}^n\backslash\mathbf{y}}$, for all $t\geq 0$ and $\xi^0_0\in(\mathbb{R}_{>0})^{\mathbf{y}}\times\{0\}^{\mathbb{H}^n\backslash\mathbf{y}}$. As in Step 1 of the proof of Theorem 2.21, it follows that, as long as $\xi^0_t(y)\leq\hat{\eta}$, for $y\in\mathbf{y}\backslash\mathbf{x}$, and

$$\hat{\eta}\underline{c}\leq\left\|\xi^0_t\right|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}\leq\varepsilon_{\mathbf{x}}, \tag{2.93}$$

we obtain

$$\frac{d}{dt}\frac{\left\|\xi^0_t|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|^2_{\mathbf{x}}}{2}\leq-\left\|\xi^0_t\right|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|^2_{\mathbf{x}}\frac{\kappa}{4C^2_{\mathbf{x}}}=:-\tilde{\kappa}\left\|\xi^0_t\right|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|^2_{\mathbf{x}}. \tag{2.94}$$

Hence

$$\left\|\xi^0_t\right|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}\leq \mathrm{e}^{-\tilde{\kappa}(t-t_0)}\left\|\xi^0_{t_0}\right|_{\mathbf{x}}-\bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}}. \tag{2.95}$$

Moreover, (2.93) implies, for every $x\in\mathbf{x}$,

$$|\xi^0_t(x)-\bar{\xi}_{\mathbf{x}}(x)|\leq\frac{\varepsilon_{\mathbf{x}}}{c_{\mathbf{x}}}. \tag{2.96}$$

Since $f_{y,\mathbf{x}} < 0$ for every $y \in \mathbf{y}\backslash\mathbf{x}$, we can choose $\varepsilon_{\mathbf{x}}$ small enough such that

$$\begin{aligned}\tfrac{d}{dt}\xi_t^0(y) &= [r(y) - \sum_{z\in\mathbb{H}^n} \alpha(y,z)\xi_t^0(z)]\xi_t^0(y) \\ &\leq \left[f_{y,\mathbf{x}} + \sum_{x\in\mathbf{x}} \alpha(y,x)\frac{\varepsilon_{\mathbf{x}}}{c_{\mathbf{x}}}\right]\xi_t^0(y) \leq -C\xi_t^0(y), \end{aligned} \tag{2.97}$$

for some $C > 0$. Hence,

$$\xi_t^0(y) \leq \mathrm{e}^{-C(t-t_0)}\xi_{t_0}^0(y). \tag{2.98}$$

We have now found an attractive domain around the limiting equilibrium of the Lotka-Volterra system.

Next, we can derive the continuity of $\bar{\tau}_\eta^0(\xi,\mathbf{x},\mathbf{y})$. Let $\gamma > 0$ arbitrarily small such that $\mathrm{e}^{\tilde{\kappa}\gamma}, \mathrm{e}^{C\gamma} \leq 2$. Let $\xi^{0,1}$ and $\xi^{0,2}$ be two versions of the process with different initial values $\xi_0^{0,1}$ and $\xi_0^{0,2}$. By Lemma 2.23,

$$\left\|\left|\xi_t^{0,1}\right|_{\mathbf{x}} - \left|\xi_t^{0,2}\right|_{\mathbf{x}}\right\|_{\mathbf{x}} \leq C_{\mathbf{x}}\left\|\left|\xi_t^{0,1}\right|_{\mathbf{x}} - \left|\xi_t^{0,2}\right|_{\mathbf{x}}\right\| \leq \mathrm{e}^{(t-t_0)A}C_{\mathbf{x}}\left\|\left|\xi_{t_0}^{0,1}\right|_{\mathbf{x}} - \left|\xi_{t_0}^{0,2}\right|_{\mathbf{x}}\right\|, \tag{2.99}$$

$$|\xi_t^{0,1}(y) - \xi_t^{0,2}(y)| \leq \left\|\xi_t^{0,1} - \xi_t^{0,2}\right\| \leq \mathrm{e}^{(t-t_0)A}\left\|\xi_{t_0}^{0,1} - \xi_{t_0}^{0,2}\right\|. \tag{2.100}$$

Now, if we pick initial conditions that are very similar, namely that satisfy

$$\left\|\xi_0^{0,1} - \xi_0^{0,2}\right\| \leq \mathrm{e}^{-(\bar{\tau}_\eta^0(\xi_{t_0}^{0,1},\mathbf{x},\mathbf{y})+\gamma)A}\left[(\mathrm{e}^{\tilde{\kappa}\gamma}-1)\frac{\eta\bar{c}c_{\mathbf{x}}}{2\sqrt{|\mathbf{x}|}C_{\mathbf{x}}} \wedge (\mathrm{e}^{C\gamma}-1)\left(\frac{\eta}{6}\wedge\hat{\eta}\right)\right], \tag{2.101}$$

we can apply (2.99) and (2.100) and use the definition of $\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})$ to derive

$$\begin{aligned}\left\|\left|\xi^{0,2}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}\right|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} &\leq \left\|\left|\xi^{0,2}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}\right|_{\mathbf{x}} - \left|\xi^{0,1}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}\right|_{\mathbf{x}}\right\|_{\mathbf{x}} + \left\|\left|\xi^{0,1}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}\right|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} \\ &\leq \mathrm{e}^{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})A}C_{\mathbf{x}}\left\|\left|\xi_0^{0,2}\right|_{\mathbf{x}} - \left|\xi_0^{0,1}\right|_{\mathbf{x}}\right\| + \frac{\eta\bar{c}c_{\mathbf{x}}}{2\sqrt{|\mathbf{x}|}} \leq \mathrm{e}^{\tilde{\kappa}\gamma}\frac{\eta\bar{c}c_{\mathbf{x}}}{2\sqrt{|\mathbf{x}|}}, \end{aligned} \tag{2.102}$$

and for $y \in \mathbf{y}\backslash\mathbf{x}$,

$$\begin{aligned}\xi^{0,2}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}(y) &\leq |\xi^{0,2}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}(y) - \xi^{0,1}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}(y)| + \xi^{0,1}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}(y) \\ &\leq \mathrm{e}^{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})A}\left\|\xi_0^{0,2} - \xi_0^{0,1}\right\| + \left(\frac{\eta}{6}\wedge\hat{\eta}\right) \leq \mathrm{e}^{C\gamma}\left(\frac{\eta}{6}\wedge\hat{\eta}\right). \end{aligned} \tag{2.103}$$

For all $\eta > 0$ such that

$$\hat{\eta}\underline{c} = \frac{\eta\bar{c}c_{\mathbf{x}}}{2\sqrt{|\mathbf{x}|}} \leq \frac{\varepsilon_{\mathbf{x}}}{2}, \tag{2.104}$$

we obtain

$$\left\|\left|\xi^{0,2}_{\bar{\tau}_\eta^0(\xi_0^{0,1},\mathbf{x},\mathbf{y})}\right|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}}\right\|_{\mathbf{x}} \leq \varepsilon_{\mathbf{x}} \tag{2.105}$$

and hence (2.95) and (2.98) can be applied to $\xi^{0,2}$ with $t = \bar{\tau}_\eta^0(\xi_0^{0,1}, \mathbf{x}, \mathbf{y}) + \gamma$ and $t_0 = \bar{\tau}_\eta^0(\xi_0^{0,1}, \mathbf{x}, \mathbf{y})$ to obtain $\bar{\tau}_\eta^0(\xi_0^{0,2}, \mathbf{x}, \mathbf{y}) \leq \bar{\tau}_\eta^0(\xi_0^{0,1}, \mathbf{x}, \mathbf{y}) + \gamma$.

Repeating the same calculation switching 1 and 2 and using this bound for $\bar{\tau}_\eta^0(\xi_0^{0,2}, \mathbf{x}, \mathbf{y})$ to apply (2.101), it follows that

$$\left\| \xi^{0,1}_{\bar{\tau}_\eta^0(\xi_0^{0,2},\mathbf{x},\mathbf{y})}\Big|_{\mathbf{x}} - \bar{\xi}_{\mathbf{x}} \right\|_{\mathbf{x}} \leq \mathrm{e}^{\tilde{\kappa}\gamma} \frac{\eta \bar{c} c_{\mathbf{x}}}{2\sqrt{|\mathbf{x}|}}, \tag{2.106}$$

$$\xi^{0,1}_{\bar{\tau}_\eta^0(\xi_0^{0,2},\mathbf{x},\mathbf{y})}(y) \leq \mathrm{e}^{C\gamma} \left(\frac{\eta}{6} \wedge \hat{\eta} \right), \tag{2.107}$$

and therefore $\bar{\tau}_\eta^0(\xi_0^{0,1}, \mathbf{x}, \mathbf{y}) \leq \bar{\tau}_\eta^0(\xi_0^{0,2}, \mathbf{x}, \mathbf{y}) + \gamma$. Hence, $|\bar{\tau}_\eta^0(\xi_0^{0,1}, \mathbf{x}, \mathbf{y}) - \bar{\tau}_\eta^0(\xi_0^{0,2}, \mathbf{x}, \mathbf{y})| \leq \gamma$, which proves the continuity. □

To mark the transition between the exponential growth phase and the Lotka-Volterra invasion phase, we extend the definition of $\tilde{T}_\eta^\mu$ in (2.39) to the i^{th} invasion.

Definition 2.27. For $i \geq 1$, the time when the first mutant type reaches $\eta > 0$ after the $(i-1)^{\text{st}}$ invasion is defined as

$$\tilde{T}_{\eta,i}^\mu := \inf\{s \geq \tilde{T}_{\eta,i-1}^\mu : \exists\, y \in \mathbb{H}^n \backslash (\mathbf{x}^{i-2} \cup \mathbf{x}^{i-1}) : \xi_s^\mu(y) > \eta\}. \tag{2.108}$$

We set $\tilde{T}_{\eta,0}^\mu := 0$ and $\mathbf{x}^{-1} := \emptyset$.

To consider the evolutionary time scale $\ln 1/\mu$, we define $T_{\eta,i}^\mu$ through $\tilde{T}_{\eta,i}^\mu = T_{\eta,i}^\mu \ln 1/\mu$.

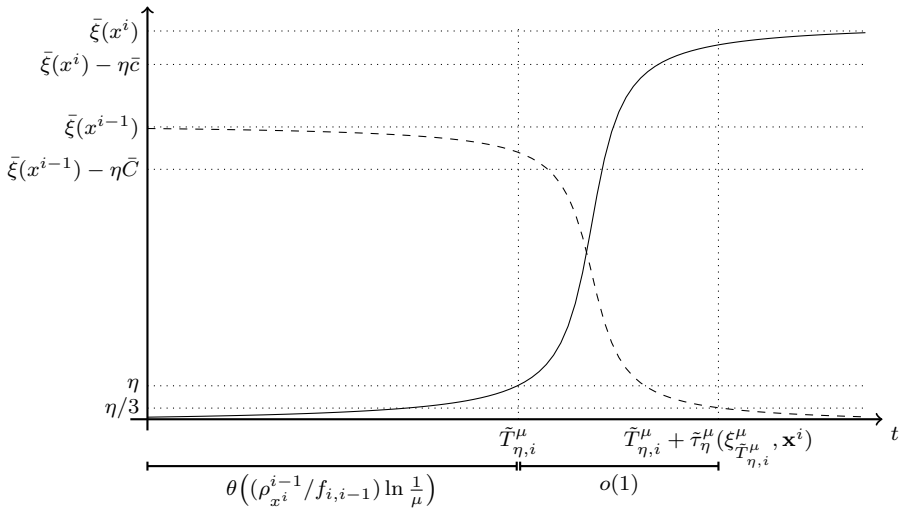


Figure 2.2: The two phases of $y_*^i = x^i$ invading x^{i-1}, in the case where there is no coexistence. The dashed line corresponds to $\xi_t^\mu(x^{i-1})$, the solid line depicts $\xi_t^\mu(x^i)$.

We can now turn to the proof of Theorem 2.12 and inductively derive the convergence of $\xi^\mu_{t \ln 1/\mu}$ to a jump process as $\mu \to 0$. The two phases of an invasion (exponential growth and Lotka-Volterra) are depicted in Figure 2.2.

Proof of Theorem 2.12. The proof is split into several parts. The main goal is to inductively approximate $T^\mu_{\eta,i}$ and $\xi^\mu_{t\ln 1/\mu}$, similar to Corollary 2.22. We claim that, for each $1 \le i \le I$ such that $T_i < \infty$,

$$\min_{\substack{y\in\mathbb{H}^n\\ \rho_y^{i-1}>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\rho_z^{i-1}+|z-y|-\eta\hat{C}_{i-1}}{f_{z,\mathbf{x}^{i-1}}+\eta\hat{C}} \le \liminf_{\mu\to 0} T^\mu_{\eta,i} - T_{i-1}$$

$$\le \limsup_{\mu\to 0} T^\mu_{\eta,i} - T_{i-1} \le \min_{\substack{y\in\mathbb{H}^n\\ \rho_y^{i-1}>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\rho_z^{i-1}+|z-y|+\eta\check{C}_{i-1}}{f_{z,\mathbf{x}^{i-1}}-\eta\check{C}}. \tag{2.109}$$

Moreover, for each $0 \le i < I$ such that $T_i < \infty$, $T_i < t < T_{i+1}$, there are positive constants $\check{c}_i$, $\check{C}_i$, $\hat{c}_i$, $\hat{C}_i$, and m, such that, for every $y \in \mathbb{H}^n$,

$$\begin{aligned}\check{c}_i\mu^{\min_{z\in\mathbb{H}^n}[\rho_z^i+|z-y|-(t-T_i)(f_{z,\mathbf{x}^i}-\eta\check{C})]+\eta\check{C}_i} &\le \xi^\mu_{t\ln 1/\mu}(y)\\ \le \hat{c}_i\mu^{\min_{z\in\mathbb{H}^n}[\rho_z^i+|z-y|-(t-T_i)(f_{z,\mathbf{x}^i}+\eta\hat{C})]-\eta\hat{C}_i}&\left(1+t\ln\frac{1}{\mu}\right)^{(i+1)m},\end{aligned} \tag{2.110}$$

while, for each $x \in \mathbf{x}^i$, $\xi^\mu_{t\ln 1/\mu}(x) \in [\bar{\xi}_{\mathbf{x}^i}(x)-\eta\bar{C}, \bar{\xi}_{\mathbf{x}^i}(x)+\eta\bar{C}]$.

In the first step, we approximate $|T^\mu_{\eta,i} - T_i| \le \eta C$, assuming that the claim holds true. Second, we derive a uniform bound on the duration of the i^{th} invasion phase, using Lemma 2.25. In Step 3, we prove the bounds that are claimed above. Finally, we use these bounds to derive the convergence as $\mu \to 0$.

Step 1: $|T^\mu_{\eta,i} - T_i| \le \eta C$.
In the case where there exists a $y \in \mathbb{H}^n$ such that $f_{y,x^{i-1}} > 0$, we want to relate T_i, as defined in (2.21), to $T^\mu_{\eta,i}$.

First, we prove a different identity for T_i that is similar to (2.109), namely the second equality of

$$T_i - T_{i-1} = \min_{\substack{y\in\mathbb{H}^n:\\ f_{y,\mathbf{x}^{i-1}}>0}} \frac{\rho_y^{i-1}}{f_{y,\mathbf{x}^{i-1}}} = \min_{\substack{y\in\mathbb{H}^n\\ \rho_y^{i-1}>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\rho_z^{i-1}+|z-y|}{f_{z,\mathbf{x}^{i-1}}}. \tag{2.111}$$

On one hand, $f_{y,\mathbf{x}^{i-1}} > 0$ implies $\rho_y^{i-1} > 0$. The only cases in which $\rho_y^{i-1} = 0$ are if $y \in \mathbf{x}^{i-1}$, then $f_{y,\mathbf{x}^{i-1}} = 0$, or if $y \in \mathbf{x}^{i-2}\backslash\mathbf{x}^{i-1}$, which implies $f_{y,\mathbf{x}^{i-1}} < 0$ (else we would have terminated the procedure after the $(i-1)^{\text{st}}$ invasion due to case (b) in Theorem 2.12). Hence

$$\min_{\substack{y\in\mathbb{H}^n\\ \rho_y^{i-1}>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\rho_z^{i-1}+|z-y|}{f_{z,\mathbf{x}^{i-1}}} \le \min_{\substack{y\in\mathbb{H}^n\\ f_{y,\mathbf{x}^{i-1}}>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\rho_z^{i-1}+|z-y|}{f_{z,\mathbf{x}^{i-1}}} \le \min_{\substack{y\in\mathbb{H}^n\\ f_{y,\mathbf{x}^{i-1}}>0}} \frac{\rho_y^{i-1}}{f_{y,\mathbf{x}^{i-1}}}, \tag{2.112}$$

where we inserted $z = y$ in the second step.

On the other hand, if we assume that $\bar{y}$ and $\bar{z}$ realise the minima, which implies that $f_{\bar{z},\mathbf{x}^{i-1}} > 0$, we obtain

$$\min_{\substack{y\in\mathbb{H}^n\\ \rho_y^{i-1}>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\rho_z^{i-1}+|z-y|}{f_{z,\mathbf{x}^{i-1}}} = \frac{\rho_{\bar{z}}^{i-1}+|\bar{z}-\bar{y}|}{f_{\bar{z},\mathbf{x}^{i-1}}} \geq \frac{\rho_{\bar{z}}^{i-1}}{f_{\bar{z},\mathbf{x}^{i-1}}} \geq \min_{\substack{y\in\mathbb{H}^n\\ f_{y,\mathbf{x}^{i-1}}>0}} \frac{\rho_y^{i-1}}{f_{y,\mathbf{x}^{i-1}}}. \tag{2.113}$$

Now, under the assumption that (2.109) holds true, we approximate

$$\begin{aligned}
\liminf_{\mu\to 0} T_{\eta,i}^{\mu} - T_{i-1} \geq & \left(\min_{\substack{y\in\mathbb{H}^n\\ \rho_y^{i-1}>0}} \min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\rho_z^{i-1}+|z-y|}{f_{z,\mathbf{x}^{i-1}}} \right) \left(\min_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{f_{z,\mathbf{x}^{i-1}}}{f_{z,\mathbf{x}^{i-1}}+\eta\hat{C}} \right) \\
& - \eta\hat{C}_{i-1} \max_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{1}{f_{z,\mathbf{x}^{i-1}}+\eta\hat{C}} \\
= & (T_i - T_{i-1}) \left(1 - \max_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{\eta\hat{C}}{f_{z,\mathbf{x}^{i-1}}+\eta\hat{C}} \right) - \eta\hat{C}_{i-1} \max_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{1}{f_{z,\mathbf{x}^{i-1}}+\eta\hat{C}} \\
= & (T_i - T_{i-1}) - \eta((T_i - T_{i-1})\hat{C} + \hat{C}_{i-1}) \max_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{1}{f_{z,\mathbf{x}^{i-1}}+\eta\hat{C}}
\end{aligned} \tag{2.114}$$

and, analogously,

$$\limsup_{\mu\to 0} T_{\eta,i}^{\mu} - T_{i-1} \leq (T_i - T_{i-1}) + \eta((T_i - T_{i-1})\check{C} + \check{C}_{i-1}) \max_{\substack{z\in\mathbb{H}^n\\ f_{z,\mathbf{x}^{i-1}}>0}} \frac{1}{f_{z,\mathbf{x}^{i-1}}-\eta\check{C}}. \tag{2.115}$$

As a result there is a constant $C > 0$ such that, for η and μ small enough,

$$|T_{\eta,i}^{\mu} - T_i| \leq \eta C. \tag{2.116}$$

Step 2: Uniform time bound on the Lotka-Volterra phase.
We show that, for η small enough,

$$\begin{aligned}
\tilde{\tau}_{\eta}^{\mu}(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}}, \mathbf{x}^i) = \inf \Big\{ t \geq 0 : \forall\, x \in \mathbf{x}^i : |\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}+t}(x) - \bar{\xi}_{\mathbf{x}^i}(x)| \leq \eta \frac{\bar{c}}{\sqrt{|\mathbf{x}^i|}}, \\
\forall\, y \in \mathbb{H}^n \backslash \mathbf{x}^i : \xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}+t}(y) \leq \frac{\eta}{3} \Big\}
\end{aligned} \tag{2.117}$$

is bounded by some constant $\bar{T}_{\eta}$.

Since $\mathrm{LVE}_+(\mathbf{x}^{i-1}) = \{\bar{\xi}_{\mathbf{x}^{i-1}}\}$ and $f_{y_*^i,\mathbf{x}^{i-1}} > 0$, we obtain $r(y) > 0$, for every $y \in (\mathbf{x}^{i-1} \cup y_*^i)$. $(\mathrm{B}_{\mathbf{x}^{i-1}\cup y_*^i})$ holds by assumption and hence Lemma 2.25 can be applied to $\mathbf{y} = \mathbf{x}^{i-1} \cup y_*^i$ and $\mathbf{x} = \mathbf{x}^i$.

Let

$$\Omega_{\eta}^i := \{\xi : \xi(y_*^i) = \eta, \xi(x) \in [\bar{\xi}_{\mathbf{x}^{i-1}}(x) - \eta\bar{C}, \bar{\xi}_{\mathbf{x}^{i-1}}(x) + \eta\bar{C}] \ \forall\, x \in \mathbf{x}^{i-1}, \xi(y) = 0 \text{ else}\}, \tag{2.118}$$

then, by continuity of $\bar{\tau}_\eta^0(\xi, \mathbf{x}^i, \mathbf{x}^{i-1} \cup y_*^i)$ in ξ (Lemma 2.25) and the compactness of Ω_η^i,

$$\sup_{\xi \in \Omega_\eta^i} \bar{\tau}_\eta^0(\xi, \mathbf{x}^i, \mathbf{x}^{i-1} \cup y_*^i) =: \bar{T}_\eta < \infty. \tag{2.119}$$

Using Lemma 2.23, for

$$\xi := \begin{cases} \xi^\mu_{\tilde{T}^\mu_{\eta,i}}(x) & x \in \mathbf{x}^{i-1} \cup y_*^i \\ 0 & \text{else} \end{cases} \in \Omega_\eta^i, \qquad \bar{\tau} := \bar{\tau}_\eta^0(\xi, \mathbf{x}^i, \mathbf{x}^{i-1} \cup y_*^i), \tag{2.120}$$

we obtain, for $x \in \mathbf{x}^i$, $y \in \mathbf{x}^{i-1} \cup y_*^i \backslash \mathbf{x}^i$, $\xi_0^0 = \xi$, and μ small enough, that

$$\begin{aligned} |\xi^\mu_{\tilde{T}^\mu_{\eta,i}+\bar{\tau}}(x) - \bar{\xi}_{\mathbf{x}^i}(x)| &\le \left\| \xi^\mu_{\tilde{T}^\mu_{\eta,i}+\bar{\tau}} - \xi^0_{\bar{\tau}} \right\| + c_{\mathbf{x}^i}^{-1} \left\| \xi^0_{\bar{\tau}}\big|_{\mathbf{x}^i} - \bar{\xi}_{\mathbf{x}^i} \right\|_{\mathbf{x}^i} \\ &\le \mathrm{e}^{\bar{\tau}A} \left(\left\| \xi^\mu_{\tilde{T}^\mu_{\eta,i}} - \xi \right\| + \sqrt{\mu \frac{B}{A}} \right) + \frac{\eta \bar{c}}{2\sqrt{|\mathbf{x}^i|}} \le \frac{\eta \bar{c}}{\sqrt{|\mathbf{x}^i|}}, \end{aligned} \tag{2.121}$$

$$\begin{aligned} \xi^\mu_{\tilde{T}^\mu_{\eta,i}+\bar{\tau}}(y) &\le \left\| \xi^\mu_{\tilde{T}^\mu_{\eta,i}+\bar{\tau}} - \xi^0_{\bar{\tau}} \right\| + \xi^0_{\bar{\tau}}(y) \\ &\le \mathrm{e}^{\bar{\tau}A} \left(\left\| \xi^\mu_{\tilde{T}^\mu_{\eta,i}} - \xi \right\| + \sqrt{\mu \frac{B}{A}} \right) + \frac{\eta}{6} \le \frac{\eta}{3}. \end{aligned} \tag{2.122}$$

Here we used that, for η small enough, $\left\| \xi^\mu_{\tilde{T}^\mu_{\eta,i}} - \xi \right\| \le 2^n \max_{y \in \mathbb{H}^n \backslash (\mathbf{x}^{i-1} \cup y_*^i)} \xi^\mu_{\tilde{T}^\mu_{\eta,i}}(y)$ tends to zero as $\mu \to 0$. A more precise approximation for this is given in Step 3 and 4.

Overall, $\tilde{\tau}_\eta^\mu(\xi^\mu_{\tilde{T}^\mu_{\eta,i}}, \mathbf{x}^i) \le \bar{\tau} \le \bar{T}_\eta$.

Step 3: Approximation of $\xi^\mu_{t \ln 1/\mu}$ and $T^\mu_{\eta,i}$.
We now turn to the proof of (2.109) and (2.110).

(2.110) in the case of $i = 0$ is given by Theorem 2.21 and Corollary 2.22, setting $\check{c}_0 := \check{c}$, $\check{C}_i := 0$, $\hat{c}_0 := 2^n \hat{c}$, and $\hat{C}_i := 0$ and using that by Step 1, for every $t < T_1$, there are η and μ small enough such that $t < T^\mu_{\eta,1}$. Corollary 2.22 also gives(2.109) for $i = 1$.

Assuming that the claims holds for $0 \le i-1 < I$, $T_i < \infty$ implies that there is some $y' \in \mathbb{H}^n$ for which $f_{y', \mathbf{x}^{i-1}} > 0$, and hence, for every $y \in \mathbb{H}^n$,

$$\begin{aligned} &\check{c}_{i-1} \mu^{\min_{z \in \mathbb{H}^n} [\rho_z^{i-1} + |z-y| - (T^\mu_{\eta,i} - T_{i-1})(f_{z,\mathbf{x}^{i-1}} - \eta \check{C})] + \eta \check{C}_{i-1}} \le \xi^\mu_{\tilde{T}^\mu_{\eta,i}}(y) \\ &\quad \le \hat{c}_{i-1} \mu^{\min_{z \in \mathbb{H}^n} [\rho_z^{i-1} + |z-y| - (T^\mu_{\eta,i} - T_{i-1})(f_{z,\mathbf{x}^{i-1}} + \eta \hat{C})] - \eta \hat{C}_{i-1}} \left(1 + \tilde{T}^\mu_{\eta,i} \right)^{im}. \end{aligned} \tag{2.123}$$

Moreover, $\xi^\mu_{\tilde{T}^\mu_{\eta,i}}(y_*^i) = \eta$ and, for every $x \in \mathbf{x}^{i-1}$, $\xi^\mu_{\tilde{T}^\mu_{\eta,i}}(x) \in [\bar{\xi}_{\mathbf{x}^{i-1}}(x) - \eta \bar{C}, \bar{\xi}_{\mathbf{x}^{i-1}}(x) + \eta \bar{C}]$. Similar to Corollary 2.22, we obtain (2.109).

Next, we estimate the evolution of the different types during the Lotka-Volterra phase. Lemma 2.23 gives $\xi_t^\mu(z) \le 2|r(z)|/\alpha(z,z)$, for all $z \in \mathbb{H}^n$ and $t \ge 0$, and therefore

$$\tfrac{d}{dt}\xi_t^\mu(y) \ge \left[r(y) - \sum_{z\in\mathbb{H}^n} \alpha(y,z)\frac{2|r(z)|}{\alpha(z,z)} - \mu b(y) \right] \xi_t^\mu(y) \ge -K\xi_t^\mu(y), \tag{2.124}$$

for some $K > 0$.

By Step 2, we know that $\tilde\tau(\xi^\mu_{\tilde T^\mu_{\eta,i}}, \mathbf{x}^i) \le \bar T_\eta$ and hence (2.123) yields

$$\xi^\mu_{\tilde T^\mu_{\eta,i}+\tilde\tau(\xi^\mu_{\tilde T^\mu_{\eta,i}},\mathbf{x}^i)}(y) \ge \mathrm{e}^{-K\bar T_\eta}\check c_{i-1}\mu^{\min_{z\in\mathbb{H}^n}[\rho_z^{i-1}+|z-y|-(T^\mu_{\eta,i}-T_{i-1})(f_{z,\mathbf{x}^{i-1}}-\eta\check C)]+\eta\check C_{i-1}} \tag{2.125}$$

Using Step 1, we can approximate

$$\begin{aligned}
&\min_{z\in\mathbb{H}^n}[\rho_z^{i-1} + |z-y| - (T^\mu_{\eta,i} - T_{i-1})(f_{z,\mathbf{x}^{i-1}} - \eta\check C)] + \eta\check C_{i-1} \\
&\quad = \min_{z\in\mathbb{H}^n}[\rho_z^{i-1} + |z-y| - (T^\mu_{\eta,i} - T_{i-1})f_{z,\mathbf{x}^{i-1}}] + \eta(\check C_{i-1} + (T^\mu_{\eta,i} - T_{i-1})\check C) \\
&\quad \le \rho_y^i + \eta(\check C_{i-1} + (T^\mu_{\eta,i} - T_{i-1})\check C + C\max_{z\in\mathbb{H}^n} f_{z,\mathbf{x}^{i-1}}).
\end{aligned} \tag{2.126}$$

We now plug this back in as the exponent and set $\check c_i' := \mathrm{e}^{-K\bar T_\eta}\check c_{i-1}$ as well as $\check C_i' \ge \check C_{i-1} + (T^\mu_{\eta,i} - T_{i-1})\check C + C\max_{z\in\mathbb{H}^n} f_{z,\mathbf{x}^{i-1}}$ to derive

$$\xi^\mu_{\tilde T^\mu_{\eta,i}+\tilde\tau(\xi^\mu_{\tilde T^\mu_{\eta,i}},\mathbf{x}^i)}(y) \ge \check c_i'\mu^{\rho_y^i+\eta\check C_i'}. \tag{2.127}$$

Note that $\check C_i'$ can be chosen uniformly in η since $T^\mu_{\eta,i} \le T_i + \eta C$ by Step 1, while $\check c_i'$ may depend on η.

On the other hand,

$$\tfrac{d}{dt}\xi_t^\mu(y) \le r(y)\xi_t^\mu(y) + \mu\tilde C\sum_{z\sim y}\xi_t^\mu(z). \tag{2.128}$$

Following the same argument as for the upper bound in (2.40) (compare Step 2 of the proof of Theorem 2.21, with $t = \tilde\tau(\xi^\mu_{\tilde T^\mu_{\eta,i}}, \mathbf{x}^i)$ and $\xi^\mu_{\tilde T^\mu_{\eta,i}}$ instead of ξ_0^μ), we obtain

$$\xi^\mu_{\tilde T^\mu_{\eta,i}+\tilde\tau(\xi^\mu_{\tilde T^\mu_{\eta,i}},\mathbf{x}^i)}(y) \le \hat c\mathrm{e}^{\tilde\tau(\xi^\mu_{\tilde T^\mu_{\eta,i}},\mathbf{x}^i)\max_{z\in\mathbb{H}^n} r(z)}(1+\tilde\tau(\xi^\mu_{\tilde T^\mu_{\eta,i}},\mathbf{x}^i))^m \sum_{z\in\mathbb{H}^n}\xi^\mu_{\tilde T^\mu_{\eta,i}}(z)\mu^{|z-y|}. \tag{2.129}$$

By Step 1,

$$\begin{aligned}
&\min_{z'\in\mathbb{H}^n}[\rho_{z'}^{i-1} + |z'-z| - (T^\mu_{\eta,i} - T_{i-1})(f_{z',\mathbf{x}^{i-1}} + \eta\hat C)] - \eta\hat C_{i-1} + |z-y| \\
&\quad \ge \min_{z'\in\mathbb{H}^n}[\rho_{z'}^{i-1} + |z'-y| - (T^\mu_{\eta,i} - T_{i-1})f_{z',\mathbf{x}^{i-1}}] - \eta(\hat C_{i-1} + (T^\mu_{\eta,i} - T_{i-1})\hat C) \\
&\quad \ge \rho_y^i - \eta(\hat C_{i-1} + (T^\mu_{\eta,i} - T_{i-1})\hat C + C\max_{z\in\mathbb{H}^n} f_{z,\mathbf{x}^{i-1}}).
\end{aligned} \tag{2.130}$$

Using this and Step 2, we derive

$$\begin{aligned}\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}+\tilde{\tau}(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i)}(y) &\le \hat{c}\mathrm{e}^{\bar{T}_\eta \max_{z\in\mathbb{H}^n} r(z)}(1+\bar{T}_\eta)^m \\ &\quad\cdot \sum_{z\in\mathbb{H}^n} \hat{c}_{i-1}\mu^{\rho^i_y-\eta(\hat{C}_{i-1}+(T^{\mu}_{\eta,i}-T_{i-1})\hat{C}+C\max_{z\in\mathbb{H}^n} f_{z,\mathbf{x}^{i-1}})}\left(1+\tilde{T}^{\mu}_{\eta,i}\right)^{im} \\ &\le \hat{c}'_i\left(1+\tilde{T}^{\mu}_{\eta,i}\right)^{im}\mu^{\rho^i_y-\eta\hat{C}'_i} \end{aligned} \tag{2.131}$$

where $\hat{c}'_i := 2^n\hat{c}\mathrm{e}^{\bar{T}_\eta\max_{z\in\mathbb{H}^n} r(z)}(1+\bar{T}_\eta)^m\hat{c}_{i-1}$ and $\hat{C}'_i \ge \hat{C}_{i-1}+(T^{\mu}_{\eta,i}-T_{i-1})\hat{C}+C\max_{z\in\mathbb{H}^n} f_{z,\mathbf{x}^{i-1}}$. As above, $\hat{C}'_i$ can be chosen uniformly in η since $T^{\mu}_{\eta,i}\le T_i+\eta C$ by Step 1, while $\hat{c}'_i$ may depend on η.

For $\tilde{\tau}(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i)=\tau(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i)\ln\frac{1}{\mu}$ and μ small enough, Step 1 implies

$$|T^{\mu}_{\eta,i}+\tau(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i)-T_i| \le \eta C+\frac{\bar{T}_\eta}{\ln\frac{1}{\mu}} \le 2\eta C. \tag{2.132}$$

For $T_i<t<T_{i+1}$, we can now pick η small enough such that $T_i+2\eta C<t<T_{i+1}-\eta C$, and hence

$$\limsup_{\mu\to 0} T^{\mu}_{\eta,i}+\tau(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i) < t < \liminf_{\mu\to 0} T^{\mu}_{\eta,i+1}. \tag{2.133}$$

As in Corollary 2.22, with the above bounds on $\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}+\tilde{\tau}(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i)}$, we derive

$$\begin{aligned}\xi^{\mu}_{t\ln 1/\mu}(y) &\ge \check{c}\check{c}'_i\mu^{\min_{z\in\mathbb{H}^n}[\rho^i_z+\eta\check{C}'_i+|z-y|-(t-(T^{\mu}_{\eta,i}+\tau(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i)))(f_{z,\mathbf{x}^i}-\eta\check{C})]} \\ &\ge \check{c}\check{c}'_i\mu^{\min_{z\in\mathbb{H}^n}[\rho^i_z+|z-y|-(t-T_i)(f_{z,\mathbf{x}^i}-\eta\check{C})]+\eta(\check{C}'_i+2C\max_{z\in\mathbb{H}^n}(f_{z,\mathbf{x}^i}-\eta\check{C}))} \\ &= \check{c}_i\mu^{\min_{z\in\mathbb{H}^n}[\rho^i_z+|z-y|-(t-T_i)(f_{z,\mathbf{x}^i}-\eta\check{C})]+\eta\check{C}_i}, \end{aligned} \tag{2.134}$$

defining $\check{c}_i := \check{c}\check{c}'_i$ and $\check{C}_i := \check{C}'_i+2C\max_{z\in\mathbb{H}^n}(f_{z,\mathbf{x}^i}-\eta\check{C})$.

Similarly, the upper bound is derived as

$$\begin{aligned}\xi^{\mu}_{t\ln 1/\mu}(y) &\le 2^n\hat{c}\hat{c}'_i\mu^{\min_{z\in\mathbb{H}^n}[\rho^i_z-\eta\hat{C}'_i+|z-y|-(t-(T^{\mu}_{\eta,i}+\tau(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i)))(f_{z,\mathbf{x}^i}+\eta\hat{C})]} \\ &\quad\cdot(1+\tilde{T}^{\mu}_{\eta,i})^{im}\left(1+\left(t\ln\frac{1}{\mu}-(\tilde{T}^{\mu}_{\eta,i}+\tilde{\tau}(\xi^{\mu}_{\tilde{T}^{\mu}_{\eta,i}},\mathbf{x}^i))\right)\right)^m \\ &\le 2^n\hat{c}\hat{c}'_i\mu^{\min_{z\in\mathbb{H}^n}[\rho^i_z+|z-y|-(t-T_i)(f_{z,\mathbf{x}^i}+\eta\hat{C})]-\eta(\hat{C}'_i+2C\max_{z\in\mathbb{H}^n}(f_{z,\mathbf{x}^i}+\eta\hat{C}))}\left(1+t\ln\frac{1}{\mu}\right)^{(i+1)m} \\ &= \hat{c}_i\mu^{\min_{z\in\mathbb{H}^n}[\rho^i_z+|z-y|-(t-T_i)(f_{z,\mathbf{x}^i}+\eta\hat{C})]-\eta\hat{C}_i}\left(1+t\ln\frac{1}{\mu}\right)^{(i+1)m}, \end{aligned} \tag{2.135}$$

with $\hat{c}_i := 2^n\hat{c}\hat{c}'_i$ and $\hat{C}_i := \hat{C}'_i+2C\max_{z\in\mathbb{H}^n}(f_{z,\mathbf{x}^i}+\eta\hat{C})$. This concludes the proof of (2.110).

Notice, that, although $\check{c}_i$ and $\hat{c}_i$ may vary for different η, $\check{C}_i$ and $\hat{C}_i$ can be chosen uniformly in η.

For every $x \in \mathbf{x}^i$, we obtain $\xi^\mu_{t\ln 1/\mu}(x) \in [\bar{\xi}_{\mathbf{x}^i}(x) - \eta\bar{C}, \bar{\xi}_{\mathbf{x}^i}(x) + \eta\bar{C}]$, as in Theorem 2.21.

Step 4: Convergence for $T_i < t < T_{i+1}$.
We now want to prove the actual convergence. We already know that the resident types are staying close to their equilibrium between T_i and T_{i+1} and therefore mainly have to show that the population sizes of the non-resident types vanish as $\mu \to 0$.

We claim that, for each $i \geq 0$, $T_i < t < T_{i+1}$, and $y \in \mathbb{H}^n \backslash \mathbf{x}^i$,

$$\min_{z\in\mathbb{H}^n} [\rho^i_z + |z-y| - (t-T_i)(f_{z,\mathbf{x}^i} + \eta\hat{C})] - \eta\hat{C}_i \geq \gamma, \tag{2.136}$$

for some $\gamma > 0$ and all η small enough, and hence

$$0 \leq \lim_{\mu\to 0} \xi^\mu_{t\ln 1/\mu}(y) \leq \lim_{\mu\to 0} \hat{c}_i \mu^\gamma \left(1 + t\ln\frac{1}{\mu}\right)^{(i+1)m} = 0. \tag{2.137}$$

We distinguish several cases. If $z \in \mathbf{x}^i$, this implies $f_{z,\mathbf{x}^i} = 0$, $\rho^i_z = 0$, and $|z-y| \geq 1$. Hence

$$\rho^i_z + |z-y| - (t-T_i)(f_{z,\mathbf{x}^i} + \eta\hat{C}) - \eta\hat{C}_i \geq 1 - \eta((t-T_i)\hat{C} + \hat{C}_i). \tag{2.138}$$

If $z \in \mathbb{H}^n \backslash \mathbf{x}^i$ and $\rho^i_z = 0$, this implies $f_{z,\mathbf{x}^i} < 0$ and

$$\rho^i_z + |z-y| - (t-T_i)(f_{z,\mathbf{x}^i} + \eta\hat{C}) - \eta\hat{C}_i \geq -(t-T_i)f_{z,\mathbf{x}^i} - \eta((t-T_i)\hat{C} + \hat{C}_i). \tag{2.139}$$

If $z \in \mathbb{H}^n \backslash \mathbf{x}^i$, $\rho^i_z > 0$, and $f_{z,\mathbf{x}^i} \leq 0$, we get

$$\rho^i_z + |z-y| - (t-T_i)(f_{z,\mathbf{x}^i} + \eta\hat{C}) - \eta\hat{C}_i \geq \rho^i_z - \eta((t-T_i)\hat{C} + \hat{C}_i). \tag{2.140}$$

Since $\check{C}_i$ does not depend on η, all these expressions can be bounded from below by a positive constant γ if η is small enough.

Finally, if $z \in \mathbb{H}^n \backslash \mathbf{x}^i$, $\rho^i_z > 0$, and $f_{z,\mathbf{x}^i} > 0$, we obtain $t < T_{i+1} \leq \rho^i_z / f_{z,\mathbf{x}^i} + T_i$ and, for η and γ small enough, $t - T_i < (\rho^i_z - \eta\hat{C}_i - \gamma)/(f_{z,\mathbf{x}^i} + \eta\hat{C})$. Therefore,

$$\rho^i_z + |z-y| - (t-T_i)(f_{z,\mathbf{x}^i} + \eta\hat{C}) - \eta\hat{C}_i > \rho^i_z - \eta\hat{C}_i - (\rho^i_z - \eta\check{C}_i - \gamma) = \gamma. \tag{2.141}$$

This proves the claim, in particular in the case where $T_{i+1} = \infty$ and there is no $y \in \mathbb{H}^n$ such that $f_{y,\mathbf{x}^i} > 0$.

Last, we consider the $x \in \mathbf{x}^i$. For every η small enough,

$$\lim_{\mu\to 0} \xi^\mu_{t\ln 1/\mu}(x) \in [\bar{\xi}_{\mathbf{x}^i}(x) - \eta\bar{C}, \bar{\xi}_{\mathbf{x}^i}(x) + \eta\bar{C}]. \tag{2.142}$$

As a result, $\lim_{\mu\to 0} \xi^\mu_{t\ln 1/\mu}(x) = \bar{\xi}_{\mathbf{x}^i}(x)$ and

$$\lim_{\mu\to 0} \xi^\mu_{t\ln 1/\mu} = \sum_{x\in\mathbf{x}^i} \delta_x \bar{\xi}_{\mathbf{x}^i}(x). \tag{2.143}$$

□

2.5 Special Case of Equal Competition

In this section we turn to the proof of Theorem 2.14, the special case of equal competition between types. We go through the proof of Theorem 2.12 to make changes where assumptions are no longer satisfied and check the identities for x^i and T_i.

Proof of Theorem 2.14. Unfortunately, assumption ($B_{\mathbf{x}}$) is not satisfied since there are no constants θ_x such that $(\theta_x \alpha)_{x,y\in\mathbf{x}}$ is positive definite for $|\mathbf{x}| \geq 2$. To still be able to apply the results of Theorem 2.12, we have to carefully go through all the points, where assumption ($B_{\mathbf{x}}$) was used.

In the proof of Theorem 2.21, this property is only used for the resident types $\mathbf{x}$. In the case where $\mathbf{x}$ consists of a single type, the positive definiteness is trivially satisfied since $\alpha > 0$.

In the case of Theorem 2.5, we have to argue differently in a few places. Champagnat, Jabin, and Raoul derive Proposition 1 from a more general theorem [35]. If one adapts the proof of this theorem to our situation, one sees that assumption ($B_{\mathbf{x}}$) is first used to prove that there are only finitely many equilibrium points. In our special case, we are only considering Lotka-Volterra systems involving the old resident type x^{i-1} and the minimizing mutant $y^i_* = x^i$. An equilibrium point $\xi^* \in (\mathbb{R}_{\geq 0})^{\{x^{i-1},x^i\}}$ has to satisfy

$$\begin{aligned} \xi^*(x^{i-1}) = 0 \text{ or } r(x^{i-1}) = \alpha(\xi^*(x^{i-1}) + \xi^*(x^i)), \\ \text{and } \xi^*(x^i) = 0 \text{ or } r(x^i) = \alpha(\xi^*(x^{i-1}) + \xi^*(x^i)). \end{aligned} \tag{2.144}$$

Since $f_{x^i,x^{i-1}} > 0$, we obtain $r(x^i) > r(x^{i-1})$ and there are only three equilibrium points, namely $(0,0)$, $(r(x^{i-1})/\alpha, 0)$, and $(0, r(x^i)/\alpha)$.

Moreover, assumption ($B_{\mathbf{x}}$) is used to prove that the evolutionary stable state (if existent) is unique. An evolutionary stable state $\bar{\xi} \in (\mathbb{R}_{\geq 0})^{\{x^{i-1},x^i\}}$ is characterised by

$$\begin{cases} r(x^j) - \alpha(\bar{\xi}(x^{i-1}) + \bar{\xi}(x^i)) \leq 0, \text{if } \bar{\xi}(x^j) = 0, \\ r(x^j) - \alpha(\bar{\xi}(x^{i-1}) + \bar{\xi}(x^i)) = 0, \text{if } \bar{\xi}(x^j) > 0, \end{cases} \tag{2.145}$$

for $j \in \{i-1, i\}$. Since $f_{i,i+1} > 0$, only the last of the three equilibrium points satisfies these assumptions,

$$r(x^{i-1}) - \alpha(\bar{\xi}(x^{i-1}) + \bar{\xi}(x^i)) = r(x^{i-1}) - \alpha\left(0 + \frac{r(x^i)}{\alpha}\right) = -f_{i,i-1} \leq 0, \tag{2.146}$$

$$r(x^i) - \alpha(\bar{\xi}(x^{i-1}) + \bar{\xi}(x^i)) = r(x^i) - \alpha\left(0 + \frac{r(x^i)}{\alpha}\right) = 0. \tag{2.147}$$

Finally, in Lemma 2.25, we are again in the situation where $\mathbf{x}$ consists of only one type and hence the positive definiteness is trivial.

The only thing left is to show the identities for x^i and T_i. We claim that, for $i \geq 0$,

$$\rho_y^{i+1} = \min_{z_{i+1}\in\mathbb{H}^n} \cdots \min_{z_1\in\mathbb{H}^n} \Bigg[|y - z_{i+1}| + \sum_{j=1}^{i} |z_{j+1} - z_j| + |z_1 - x^0| \\ - f_{z_1,x^0} T_1 - \sum_{j=1}^{i} f_{z_{j+1},x_j}(T_{j+1} - T_j) \Bigg]. \tag{2.148}$$

From the initial condition we obtain $\rho_y^0 = \min_{z\in\mathbb{H}^n}[\lambda_z + |z - y|] = |y - x^0|$. Hence,

$$y_*^1 = \operatorname*{arg\,min}_{y\in\mathbb{H}^n : f_{y,x^0}>0} \frac{|y - x^0|}{f_{y,x^0}} \tag{2.149}$$

and

$$T_1 = \min_{\substack{y\in\mathbb{H}^n: \\ f_{y,x^0}>0}} \frac{|y - x^0|}{f_{y,x^0}}. \tag{2.150}$$

Since $f_{y_*^1,x^0} = r(y_*^1) - r(x^0) > 0$, the new equilibrium is monomorphic of type $x^1 = y_*^1$ and $T_1 = |x^1 - x^0|/f_{1,0}$. Moreover,

$$\rho_y^1 = \min_{z\in\mathbb{H}^n}[\rho_z^0 + |z - y| - T_1 f_{z,x^0}] = \min_{z\in\mathbb{H}^n} \Big[|y - z| + |z - x^0| - f_{z,x^0} T_1 \Big]. \tag{2.151}$$

Assume that x^i, T_i, and ρ_y^i are of the proposed form. Then there is a unique

$$\begin{aligned} x^{i+1} = y_*^{i+1} &= \operatorname*{arg\,min}_{y\in\mathbb{H}^n : f_{y,x^i}>0} \frac{\rho_y^i}{f_{y,x^i}} \\ &= \operatorname*{arg\,min}_{y\in\mathbb{H}^n : f_{y,x^i}>0} \frac{\min_{z_i\in\mathbb{H}^n} \left[|y - z_i| + \rho_{z_i}^{i-1} - f_{z_i,x^{i-1}}(T_i - T_{i-1}) \right]}{f_{y,x^i}} \\ &= \operatorname*{arg\,min}_{y\in\mathbb{H}^n : f_{y,x^i}>0} \min_{z_i\in\mathbb{H}^n} F(y, z_i), \end{aligned} \tag{2.152}$$

where the last equality serves as the definition of the function $F : \mathbb{H}^n \times \mathbb{H}^n \to \mathbb{R}_+$.

Assume that the minimum over z_i is only realised by some $\bar{z} \neq y_*^{i+1}$, i.e.

$$\min_{\substack{y\in\mathbb{H}^n \\ f_{y,x^i}>0}} \min_{z_i\in\mathbb{H}^n} F(y, z_i) = \min_{z_i\in\mathbb{H}^n} F(y_*^{i+1}, z_i) = F(y_*^{i+1}, \bar{z}) < F(y_*^{i+1}, y_*^{i+1}). \tag{2.153}$$

Looking back at the definition of F and using that

$$\begin{aligned} \rho_{y_*^{i+1}}^{i-1} &= \min_{z\in\mathbb{H}^n}[\rho_z^{i-2} + |z - y_*^{i+1}| - (T_{i-1} - T_{i-2}) f_{z,\mathbf{x}^{i-2}}] \\ &\leq \min_{z\in\mathbb{H}^n}[\rho_z^{i-2} + |z - \bar{z}| - (T_{i-1} - T_{i-2}) f_{z,\mathbf{x}^{i-2}}] + |\bar{z} - y_*^{i+1}| \\ &= \rho_{\bar{z}}^{i-1} + |y_*^{i+1} - \bar{z}|, \end{aligned} \tag{2.154}$$

this yields

$$0 \leq |y_*^{i+1} - \bar{z}| + \rho_{\bar{z}}^{i-1} - \rho_{y_*^{i+1}}^{i-1} < (f_{\bar{z},x^{i-1}} - f_{y_*^{i+1},x^{i-1}})(T_i - T_{i-1}) \tag{2.155}$$

and, since $T_i > T_{i-1}$, we obtain $f_{\bar{z},x^{i-1}} > f_{y_*^{i+1},x^{i-1}} > 0$. But this would imply

$$\min_{z_i \in \mathbb{H}^n} F(\bar{z}, z_i) \leq F(\bar{z}, \bar{z}) < F(y_*^{i+1}, \bar{z}) = \min_{\substack{y \in \mathbb{H}^n \\ f_{y,x^i} > 0}} \min_{z_i \in \mathbb{H}^n} F(y, z_i), \tag{2.156}$$

which is a contradiction. Hence, $\bar{z}$ can be chosen equal to y_*^{i+1}.

Repeating the previous argument shows that the minimum over $z_1, ..., z_{i+1}$ is achieved at $z_1 = ... = z_{i+1} = y$ and hence

$$\begin{aligned}
x^{i+1} &= \operatorname*{arg\,min}_{y \in \mathbb{H}^n : f_{y,x^i} > 0} \frac{\rho_y^{i-1} - f_{y,x^{i-1}}(Ti - T_{i-1})}{f_{y,x^i}} = \dots \\
&= \operatorname*{arg\,min}_{y \in \mathbb{H}^n : f_{y,x^i} > 0} \frac{|y - x^0| - f_{y,x^0} \frac{|x^1 - x^0|}{f_{1,0}} - \sum_{j=1}^{i-1} f_{y,x^j}(T_{j+1} - T_j)}{f_{y,x^i}} \\
&= \operatorname*{arg\,min}_{y \in \mathbb{H}^n : f_{y,x^i} > 0} \frac{|y - x^0|}{f_{y,x^i}} - \frac{f_{y,x^{i-1}}}{f_{y,x^i}} T_i - \sum_{j=1}^{i-1} \frac{f_{y,x^{j-1}} - f_{y,x^j}}{f_{y,x^i}} T_j \\
&= \operatorname*{arg\,min}_{y \in \mathbb{H}^n : f_{y,x^i} > 0} \frac{|y - x^0|}{f_{y,x^i}} - \frac{f_{y,x^{i-1}}(|x^i - x^0| - |x^{i-1} - x^0|)}{f_{y,x^i} f_{i,i-1}} \\
&\qquad - \sum_{j=1}^{i-1} \frac{|x^j - x^0| - |x^{j-1} - x^0|}{f_{j,j-1}} \frac{f_{y,x^{j-1}} - f_{y,x^j}}{f_{y,x^i}} \\
&= \operatorname*{arg\,min}_{y \in \mathbb{H}^n : f_{y,x^i} > 0} \frac{|y - x^0|}{f_{y,x^i}} - (|x^i - x^0| - |x^{i-1} - x^0|) \left(\frac{1}{f_{i,i-1}} + \frac{1}{f_{y,x^i}} \right) \\
&\qquad - \frac{|x^{i-1} - x^0| - |x^0 - x^0|}{f_{y,x^i}} \\
&= \operatorname*{arg\,min}_{y \in \mathbb{H}^n : f_{y,x^i} > 0} \frac{|y - x^0| - |x^i - x^0|}{f_{y,x^i}} - T_i,
\end{aligned} \tag{2.157}$$

where we use (2.25) several times. Analogously,

$$T_{i+1} = T_i + \min_{\substack{y \in \mathbb{H}^n : \\ f_{y,x^i} > 0}} \frac{\rho_y^i}{f_{y,x^i}} = T_i + \left(\frac{|x^{i+1} - x^0| - |x^i - x^0|}{f_{i+1,i}} - T_i \right) = \frac{|x^{i+1} - x^0| - |x^i - x^0|}{f_{i+1,i}}. \tag{2.158}$$

Finally,

$$\rho_y^{i+1} = \min_{z_{i+1} \in \mathbb{H}^n} [\rho_{z_{i+1}}^i + |z_{i+1} - y| - (T_{i+1} - T_i) f_{z_{i+1},x^i}], \tag{2.159}$$

which is of the desired form. This proves the claim and hence the theorem. □

2.6 A First Look at Limited Range of Mutation

In this section we present the proof of Theorem 2.18, where $\ell = 1$, and take a first look at the intermediate cases of $1 < \ell < n$.

2.6.1 Proof for the case $\ell = 1$

We again go over the previous proofs and make alterations where necessary.

Proof of Theorem 2.18. We only consider the first invasion step. We can assume that $\eta < \bar{\xi}$. Consequently, up to time $\tilde{T}^{\mu}_{\eta,1} \wedge \inf\{t \geq 0 : \exists\, z \in \mathbb{H}^n, |z - x^0| > 1, \xi^{\mu}_t(z) \geq \bar{\xi}\mu\}$, the neighbours of type x^0 are the only active mutants. As before,

$$\xi^{\mu}_t(x^0) \in [\bar{\xi}_{x^0}(x^0) - \eta\bar{C}, \bar{\xi}_{x^0}(x^0) + \eta\bar{C}]. \tag{2.160}$$

Moreover, as in (2.60) and (2.71), we obtain

$$[f_{x^0,x^0} - \eta\check{C}]\xi^{\mu}_t(x^0) \leq \tfrac{d}{dt}\xi^{\mu}_t(x^0) \leq [f_{x^0,x^0} + \eta\hat{C}]\xi^{\mu}_t(x^0), \tag{2.161}$$

and with $f_{x^0,x^0} = 0$, $c := \bar{\xi}_{x^0}(x^0) - \bar{c}\bar{\xi}$, and $C := \bar{\xi}_{x^0}(x^0) + \bar{c}\bar{\xi}$,

$$c\mathrm{e}^{-t\eta\check{C}} \leq \xi^{\mu}_t(x^0) \leq C\mathrm{e}^{t\eta\hat{C}}. \tag{2.162}$$

Considering the neighbours $y \sim x^0$ of the resident type, we derive

$$[f_{y,x^0} - \eta\check{C}]\xi^{\mu}_t(y) + \mu\tilde{c}\xi^{\mu}_t(x^0) \leq \tfrac{d}{dt}\xi^{\mu}_t(y) \leq [f_{y,x^0} + \eta\hat{C}]\xi^{\mu}_t(y) + \mu\tilde{C}\xi^{\mu}_t(x^0), \tag{2.163}$$

and hence the upper bound,

$$\begin{aligned}
\xi^{\mu}_t(y) &\leq \mathrm{e}^{t(f_{y,x^0}+\eta\hat{C})}C_y\mu^{\lambda_y} + \mu\tilde{C}C\int_0^t \mathrm{e}^{s\eta\hat{C}}\mathrm{e}^{(t-s)(f_{y,x^0}+\eta\hat{C})}ds \\
&\leq \mu\mathrm{e}^{t(f_{y,x^0}+\eta\hat{C})}\left(C_y\mu^{\lambda_y-1} + \tilde{C}C\int_0^t \mathrm{e}^{-sf_{y,x^0}}ds\right) \\
&\leq \hat{c}'\mu\mathrm{e}^{t\eta\hat{C}}\left((1+t)\mathrm{e}^{tf_{y,x^0}} + 1\right),
\end{aligned} \tag{2.164}$$

for some $\hat{c}' < \infty$, uniformly in $y \sim x^0$, $\eta < \bar{\xi}$, and μ.

A similar lower bound can be shown and, on the $\ln 1/\mu$-time scale, we obtain

$$\check{c}'\mu^{((1-tf_{y,x^0})\wedge 1)+t\eta\check{C}} \leq \xi^{\mu}_{t\ln\frac{1}{\mu}}(y) \leq \hat{c}'\mu^{((1-tf_{y,x^0})\wedge 1)-t\eta\hat{C}}\left(1 + t\ln\frac{1}{\mu}\right). \tag{2.165}$$

Using this bound, all types z such that $|z - x^0| = 2$ can be bounded from above using the same type of calculation to derive

$$\xi^{\mu}_{t\ln\frac{1}{\mu}}(z) \leq C\mu^{2-t\eta\hat{C}}\left(\left(1 + t\ln\frac{1}{\mu}\right)^2\mu^{-t\max_{y\sim x^0} f_{y,x^0}} + 1\right). \tag{2.166}$$

Hence, for η small enough, $\tilde{T}^{\mu}_{\eta,1} \approx \inf\{t \geq 0 : \exists\, z \in \mathbb{H}^n, |z - x^0| > 1, \xi^{\mu}_t(z) \geq \bar{\xi}\mu\}$.

As in Corollary 2.22, we can now argue that

$$\min_{\substack{y \sim x^0 \\ f_{y,x^0} > 0}} \frac{1}{f_{y,x^0} + \eta\check{C}} \leq \liminf_{\mu \to 0} T^{\mu}_{\eta,1} \leq \limsup_{\mu \to 0} T^{\mu}_{\eta,1} \leq \min_{\substack{y \sim x^0 \\ f_{y,x^0} > 0}} \frac{1}{f_{y,x^0} - \eta\check{C}}. \tag{2.167}$$

The first mutant y^1_* to reach the η-level is the neighbour of x^0 minimising $1/f_{y,x^0}$ (given $f_{y,x^0} > 0$), hence maximising $r(y)$, which is unique (or else we set $I := i$ and terminate the procedure). This yields $T^{\mu}_{\eta,1} \approx T_1 = 1/f_{y^1_*,x^0}$.

The Lotka-Volterra phase can be analysed just as before. Since y^1_* satisfies $r(y^1_*) > r(x^0)$, the new equilibrium has $x^1 = y^1_*$ as the only resident type.

Since, for every other $y \sim x^0$, $r(y) < r(x^1)$, these types always stay unfit, do not foster mutants above the threshold, and we do not need to consider them any further.

During the Lotka-Volterra phase, once the $\xi^{\mu}_t(z)$, $z \sim x^1$ have surpassed $\bar{\xi}\mu$, they start to grow. However, since the duration of the Lotka-Volterra phase can be bounded uniformly as before, this only results in mutant populations of order μ^1, which fits the initial conditions for the next invasion step. □

2.6.2 The intermediate cases

For now, we stick with the assumption of constant competition. In the case of $\ell \geq n$, arbitrarily large steps can be taken. In particular, arbitrarily large valleys in the fitness landscape (defined by r) can be crossed. A (strict) global fitness maximum is reached eventually and is the only stable point. If $\ell = 1$, the limiting walk always jumps to the fittest nearest neighbour and (strict) local fitness maxima are stable points. In both cases, the microscopic types do not have to be tracked to characterise the jump process. The next step is determined only by the previous and possibly the initial resident type.

The cases $2 \leq \ell \leq n-1$ interpolate between the two extreme scenarios. To study accessibility of different types, we again need to keep track of the microscopic populations. To this extent, we define some new quantities.

Definition 2.28. The *first appearance time* of a type y (on the $\ln 1/\mu$-time scale) is denoted by

$$\tau^{\mu}_y := \inf\{s \geq 0 : \xi^{\mu}_{s \ln \frac{1}{\mu}}(y) > 0\}. \tag{2.168}$$

The μ-power the population size of type y would have at time $t \ln 1/\mu$ due to its own growth rate (neglecting mutation from neighbours after τ^{μ}_y) is

$$\lambda_t(y) := \mathbb{1}_{t \geq \tau^{\mu}_y} \Big(\underbrace{\ell \wedge |y - x^0|}_{\text{initial size}} - \sum_{i=0}^{\infty} \underbrace{f_{y,x^i}(t \wedge T^{\mu}_{\eta,i+1} - \tau^{\mu}_y \vee T^{\mu}_{\eta,i})_+}_{\substack{\text{growth between} \\ i^{\text{th}} \text{ and } (i+1)^{\text{st}} \text{ invasion}}} \Big) + \mathbb{1}_{t < \tau^{\mu}_y} \infty, \tag{2.169}$$

where x^i and $T^\mu_{\eta,i}$ are just as before.

All types under the *mutational influence* of type y are denoted by

$$\Lambda_t(y) := \{z \in \mathbb{H}^n : |z - y| + \lambda_t(y) \leq \ell\} \tag{2.170}$$

and $\Lambda_t := \bigcup_{y\in\mathbb{H}^n} \Lambda_t(y)$.

Since we are assuming constant competition, the population sizes of the different types are approximated by

$$\xi^\mu_{t \ln \frac{1}{\mu}}(y) \approx \mathbb{1}_{y\in\Lambda_t} \mu^{\min_{z\in\Lambda_t}[|y-z|+\lambda_t(z)]}, \tag{2.171}$$

where we drop multiplicative constants and all terms involving η. Figure 2.3 visualises the interplay of $\lambda_t(y)$, $\xi^\mu_{t \ln 1/\mu}(y)$, and the sets $\Lambda_t(y)$ for an easy example.

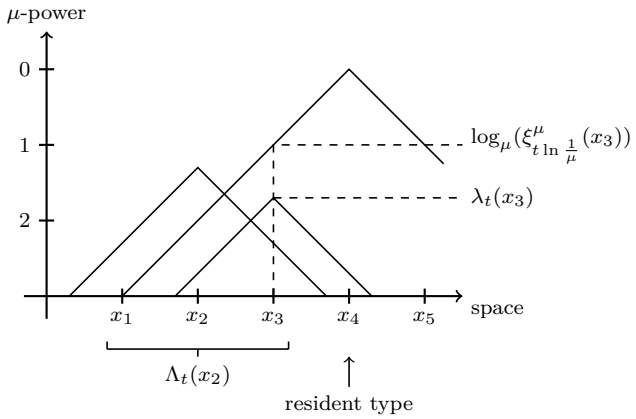


Figure 2.3: Example for the case $\ell = 3$. The mutational influence of x_2 reaches x_1 and x_3. The population size of x_3 is not determined by its own growth rate but by mutants from the resident type x_4.

It is not easy to make general statements about the evolution of this intermediate model. However, we state some first results on the accessibility of types.

Definition 2.29. A type $y \in \mathbb{H}^n$ is called *accessible* if $y \in \Lambda_\infty := \bigcup_{t\geq 0} \Lambda_t$.

Remark 2.30. This is equivalent to $\tau^\mu_y < \infty$.

Since resident types can only produce mutants in a radius of ℓ, in order to be accessible, a type has to be reached on a path with types of increasing fitness and at most distance ℓ. Figure 2.4 gives an example for such a path.

Lemma 2.31. *A necessary condition for a type y to be accessible is the existence of a path $(y_0 = x^0, y_1, ..., y_m = y)$ and indices $i_0 = 0 < i_1 < ... < i_k = m$, such that*

$$\forall\ 1 \leq j \leq k : |i_j - i_{j-1}| \leq \ell, \tag{2.172}$$

$$\forall\ 1 \leq j < k : f_{y_{i_j}, y_{i_{j-1}}} > 0, \tag{2.173}$$

$$\forall\ i_{j-1} < i < i_j : f_{y_{i_{j-1}}, y_i} > 0. \tag{2.174}$$

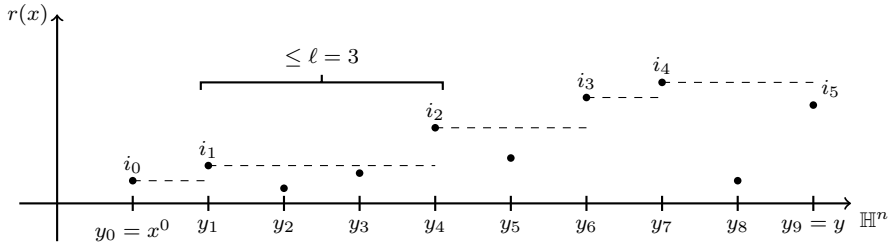


Figure 2.4: A possible path to access y, for $\ell = 3$.

Proof. Assume that $y \neq x^0$. If $y \in \Lambda_0(x^0)$, this implies $|y - x^0| \leq \ell$. Hence we can choose any shortest path from x^0 to y and pick the indices i_j such that the conditions are satisfied.

If y is accessible but $y \notin \Lambda_0(x^0)$, then $\tau_y^\mu > 0$. There is at least one $z \neq y$ such that $y \in \Lambda_{\tau_y^\mu}(z)$. We choose such a z for which the rate $r(z)$ is maximal. Consequently, $\tau_z^\mu < \tau_y^\mu$ and $\xi^\mu_{\tau_y^\mu \ln \frac{1}{\mu}}(z) \approx \mu^{\lambda_{\tau_y^\mu}(z)}$ (else z would just grow due to mutants from a fitter type, which would imply that z was not chosen such that the rate $r(z)$ is maximal). Any direct path from z to y now only goes through types that are unfit in comparison to z. We set $y_{i_k} := z$.

We can now iterate this procedure with z replacing y. In addition, we know that, for the $z' \neq z$ such that $z \in \Lambda_{\tau_z^\mu}(z')$ and $r(z')$ is maximised, $r(z) > r(z')$ (else, we would obtain $\Lambda_t(z) \subset \Lambda_t(z')$, for all $t \geq 0$, and z would not have been chosen maximising $r(z)$). We set $y_{i_{k-1}} := z'$ and continue until we reach x^0. □

Remark 2.32. The condition in Lemma 2.31 is not sufficient. Even if such a path exists, there might be a type z that is reached before y_{i_j} such that $r(z) > r(y_{i_j})$. In this case the population of y_{i_j} is not fit to grow and might never reach the necessary size to induce mutants of type $y_{i_{j+1}}$.

As a corollary, we can consider the non-crossing of fitness valleys. Figure 2.5 gives the example of a non-accessible type, surrounded by a fitness valley.

Corollary 2.33. *If a type y is surrounded by a fitness valley of width at least $\ell+1$, i.e. for all paths $(y_0 = x_0, y_1, ..., y_m = y)$ there exists an $i \leq m-(\ell+1)$ such that $f_{y_i,y_j} > 0, \forall\, i < j < m$, it is non-accessible.*

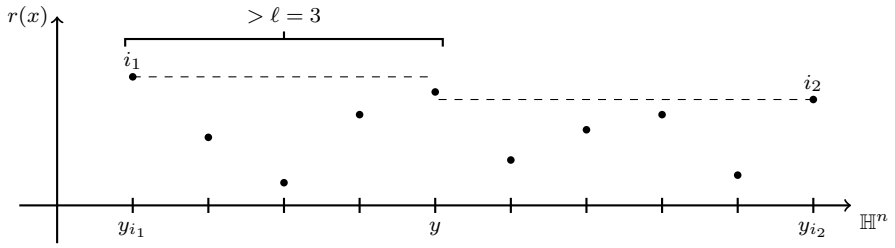


Figure 2.5: Due to the high fitness of y_{i_1} and y_{i_2}, y is not accessible for $\ell = 3$.

Proof. The claim follows directly from Lemma 2.31 since in this case the necessary path cannot exist. □

As a result, at least in the matter of crossing fitness valleys, the intermediate cases interpolate between the extreme cases.

However, as in the case of $\ell = n$, it is still possible to take arbitrarily large steps in the macroscopic process or the limiting jump process, respectively. If there was a series of types with distance smaller than $\ell + 1$ and fast increasing rate r, then each population could be overtaken by its faster growing mutants before it reaches the macroscopic level of μ^0.

Overall, the microscopic types play an important role in defining the limiting process.

3 Stochastic individual-based models with power law mutation rate on a general finite trait space

We consider a stochastic individual-based model for the evolution of a haploid, asexually reproducing population. The space of possible traits is given by the vertices of a (possibly directed) finite graph $\mathcal{G} = (V, E)$. The evolution of the population is driven by births, deaths, competition, and mutations along the edges of $\mathcal{G}$. We are interested in the large population limit under a mutation rate μ_K given by a negative power of the carrying capacity K of the system: $\mu_K = K^{-1/\alpha}, \alpha > 0$. This results in several mutant traits being present at the same time and competing for invading the resident population. We describe the time evolution of the orders of magnitude of each sub-population on the $\log K$ time scale, as K tends to infinity. Using techniques developed in [38], we show that these are piecewise affine continuous functions, whose slopes are given by an algorithm describing the changes in the fitness landscape due to the succession of new resident or emergent types. This work generalises [112] to the stochastic setting, and Theorem 3.2 of [25] to any finite mutation graph. We illustrate our theorem by a series of examples describing surprising phenomena arising from the geometry of the graph and/or the rate of mutations.

3.1 Introduction

Adaptive dynamics is a biological theory that was developed to study the interplay between ecology and evolution. It involves the three mechanisms of heredity, mutations, and natural selection. It was first introduced in the 1990s by Metz, Geritz, Bolker, Pacala, Dieckmann, Law, and coauthors [139, 59, 81, 20, 21, 60], who mostly considered a deterministic setting but also heuristically mentioned first stochastic versions. A paradigm of adaptive dynamics is the separation of the slow evolutionary and the fast ecological time scales, which is a result of reproduction with rare mutations. Invasion, fixation or extinction of a mutant population is determined by its invasion fitness, that describes the exponential growth rate of a single mutant in the current (coexisting) population(s) at equilibrium.

Stochastic individual-based models of adaptive dynamics have been rigorously constructed and first studied in the seminal work of Fournier and Méléard [77], and there is now a growing literature on these models. The population consists of a collection of individuals who reproduce, with or without mutation, or die after random exponential times depending on the current state of the whole population. The population size is controlled by a *carrying capacity* K which represents the amount of available resources. This class of models has first

been studied in the original context of separation between evolutionary and ecological time scales. That is in the joint limit of large populations and rare mutations such that a mutant either dies out or fixates before the next mutation occurs. Mathematically this amounts to considering a probability of mutation satisfying in particular

$$\mu_K \ll 1/K \log K \quad \text{as } K \to \infty. \tag{3.1}$$

We will call this regime 'rare mutation regime' in the sequel. The description of the succession of mutant invasions, on the mutation time scale $1/K\mu_K$, in a monomorphic [31] or polymorphic [37, 15] asexual population gives rise respectively to the so-called *Trait Substitution Sequence* or *Polymorphic Evolution Sequence*. Extensions of the question to sexual populations were then studied, both in the haploid [163, 43] and the diploid [41, 146] cases.

It is natural to consider the effect of a higher mutation rates, where mutation events are no longer separated, if we want to describe several mutant traits being present microscopically at the same time and competing for invading the resident population. The mutation rate given by

$$\mu_K = K^{-\frac{1}{\alpha}}, \quad \text{for } \alpha > 0 \tag{3.2}$$

was considered in different contexts [65, 164, 25, 38] and will be the concern of the present paper. Notice that another mutation scale has been considered in [15, 16] to model the interaction of few mutants in the case without recurrent mutations, namely μ_K of order $1/K \log K$.

Another approach to adaptive dynamics has been introduced by Maynard Smith [132] under the name of *adaptive walks*. This was further developed by Kauffman and Levin [102] and many others, as mentioned below. Here, a given finite graph represents the possible types of individuals (vertices) together with their possibilities of mutation (edges). A fixed, but possibly random, fitness landscape assigns real numbers to the vertices of the graph. The evolution of the population is modelled as a random walk on the graph that moves towards higher fitnesses. This can be interpreted as the adaptation of the population to its environment. In contrast to the adaptive dynamics context, this fitness landscape is not dependent on the current state of the population. Adaptive walks move along edges towards neighbours of increasing fitness, according to some transition law, towards a local or global maximum. In particular, in such models it is not possible for a population to cross a fitness valley. This is partially solved by a variation of this model, called *adaptive flight* [144]. It consists in a walk jumping between local fitness maxima, before eventually reaching a global maximum. The questions of the distribution of maxima [147], the typical length of a walk [150], or the typical accessibility properties of the fitness landscape [113, 161, 13] have been studied under different assumptions on the graph structure, the fitness law, or the transition law of the walk. Moreover, comparisons of these models with actual empirical fitness landscapes have been performed in [167]. As Kraut and Bovier showed [112], adaptive walks and flights arise as the limit of individual-based models of adaptive dynamics, when the large population followed by the rare mutations limit is taken. They also conjecture, and this will be proved in the present article, that similar results hold in the stochastic setting under the mutation rate (3.2), as we detail below.

In this paper, we consider an individual-based Markov process that models the evolution of a haploid, asexually reproducing population. The space of possible traits is given by the vertices of a (possibly directed) finite graph $\mathcal{G} = (V, E)$. The evolution of the population is driven by birth, death, and competition rates, which are fixed and depend on the traits, as well as mutations towards nearest neighbours in the graph $\mathcal{G}$. We start with a macroscopic initial condition (that is to say of order K, see Definition 3.2) and we are interested in the stochastic process given by the large population limit under the mutation rate (3.2). We describe the time evolution of the orders of magnitude of each sub-population on the $\log K$ time scale, as K tends to infinity. We show that the limiting process is deterministic, given by piecewise affine continuous functions, which are determined by an algorithm describing the changes in the fitness landscape due to the succession of new resident or emergent types.

This work constitutes an extension of the paper by Kraut and Bovier [112] to the stochastic setting. They consider the deterministic system resulting from the large population limit of the individual-based model ($K \to \infty$), and let the mutation probability μ tend to zero. By rescaling the time by $\log(1/\mu)$, they prove that the limiting process is a deterministic adaptive walk that jumps between different equilibria of coexisting traits. A corollary of our results gives the same behaviour, on the $\log K$ time scale, for the stochastic process under the scaling (3.2) for α larger than the diameter of the graph $\mathcal{G}$. Kraut and Bovier also study a variation of the model, where they modify the deterministic system such that the subpopulations can only reproduce when their size lies above a certain threshold μ^α. This limits the radius in which a resident population can foster mutants, and mimics the scaling (3.2) that we consider. The resulting limiting processes are adaptive flights (which are not restricted to jumping to nearest neighbours), and thus can cross valleys in the fitness landscape and reach a global fitness maximum. We obtain the same behaviour, on the $\log K$ time scale, for the stochastic process under the scaling (3.2) without any restriction on α.

The results of the present paper can also be seen as a generalisation of Theorem 3.2 in [25] by Bovier, Coquille, and Smadi to any finite trait space. Indeed, they consider the graph with vertices $V = \{0, \dots, L\}$ embedded in $\mathbb{N}$ and choose parameters such that the induced fitness landscape exhibits a valley: mutant individuals with negative fitness have to be created in order for the population to reach a trait with positive fitness. Several speeds of the mutation rate are considered, and in particular, when $\alpha > L$, the exit time of the valley is computed on the $\log K$ time scale. This becomes a corollary of our results, and we can give an algorithmic description of the rescaled process for more general graphs endowed with a fitness valley, as we discuss in several examples in Section 3.3.

Our proof heavily relies on couplings of the original process with logistic birth-and-death processes with non-constant immigration, and the analysis of the latter simpler processes on the $\log K$ time scale. This approach was developed by Champagnat, Méléard, and Tran in [38]. They consider an individual-based model for the evolution of a discrete population performing horizontal gene transfer and mutations on $V = [0, 4] \cap \delta\mathbb{N}, \delta > 0$. Their goal is to analyse the trade-off between natural selection, which drives the population to higher birth-rates, and transfer, which drives the population to lower ones. Under the mutation rate (3.2), they exhibit parameter regimes where different evolutionary outcomes appear, in particular evolutionary suicide and emergence of a cyclic behaviour. As in the present paper, their results characterise the time evolution of the orders of magnitude of each sub-population

on the $\log K$ time scale, which are shown to be piecewise affine continuous functions whose slopes are given by an algorithm describing the succession of phases when a given type is dominant or resident. Their proofs provide us with the main ingredients needed for our results. However, the graph structure they choose simplifies the inductions and we have to generalise their approach to treat the case of more general graphs, in the proof spirit of Kraut and Bovier [112].

Our results are general, and could be applied to have a better understanding of evolutionary trajectories in complex fitness landscapes. There are now more and more empirical studies of fitness landscapes (see [56] for a comprehensive review of data and tools up to 2014 for instance), and the probability and effect of specific mutations in given landscapes are better and better understood. For instance oriented mutation graphs can stem from mutation bias, through codon usage bias or similar molecular phenomena which make some mutations more probable than others [153].

We present a series of specific examples where surprising phenomena arise from the geometry of the graph $\mathcal{G}$ and/or the rate of mutations (3.2). Most of them could not happen under a different scaling of mutation rates.

- In Example 3.10, we describe a scenario where the ancestry of the resident population consists, with high probability, of back mutations towards a previously extinct trait, although the mutations that happen in between are not deleterious. In other words, the final resident individuals, say of trait v, although they can be produced from a wild type directly, come with high probability from a sequence of *non deleterious* mutations which went back to the wild type before mutating to v. This phenomenon can also happen in the regime (3.1), that is for $\alpha \in (0, 1)$, on the mutation time scale ($1/K\mu_K \gg \log K$), where invading mutants fully replace the resident population before a new mutant arises. We show that it can still occur for higher mutation rates of the form (3.2), on a $\log K$ time scale, when parameters are chosen such that temporary extinction of the original trait is likely. Such mutational reversions have been observed (see [57] for instance).

- If evolution and mutation time scales are separated (i.e. in the regime (3.1)), mutations occur one at a time, and the number of successive resident traits from the wild type to the type gathering k successively beneficial mutations is k. This is not the case if mutations are faster, in which case it is possible to observe either more or less successive resident traits. We will show this in Examples 3.11 and 3.12.

- In Example 3.13, we show that adding a new possible mutation path towards a fit trait can increase the time until it appears macroscopically. This is in the spirit of the paradox called *price of anarchy* in game theory or more specifically *Braess paradox* in the study of traffic networks congestion. Motter showed that this paradox may often occur in biological and ecological systems [142]. He studies the removal of part of a metabolic network to ensure its long term persistance, with applications to cancer, antibiotics and metabolic diseases. Another field of application is the food webs management, where selective removal of some species from the network can potentially have a positive outcome of preventing a series of further extinctions [158].

- Another counter-intuitive phenomenon arising from the mutation rate (3.2), presented in Example 3.14, is the possibility to observe, for a cyclic clockwise oriented mutation graph, successive counter-clockwise resident populations. This means that the macroscopic succession of resident traits is not necessarily representative of the mutation graph. In particular, this may call into question the interpretation in terms of mutation graphs of some experiments in experimental evolution (see [124] for instance).
- In Examples 3.15 and 3.16, we show that the mutation rate (3.2) does not restrict the range of the corresponding adaptive flights on the trait space, i.e. the distance that the limiting process can jump, to $\lfloor \alpha \rfloor$.
- We finally study the framework of fitness valley crossings. Combining our results with Theorem 3.3 of [25], we construct Examples 3.17 and 3.19, where effective random walks on the trait space appear on the time scale K^β, for some positive β. Those limiting adaptive flights arise as a result of a "fast" equilibration on the $\log K$ time scale followed by exponential waiting times until fitness valleys get crossed. This makes sense biologically, since there may be traits with positive invasion fitness that can be reached through several consecutive mutation steps [120, 48].

The remainder of this paper is organised as follows. In Section 3.2 we define the model and present our results. In Section 3.3 we illustrate our results by a series of examples describing surprising phenomena arising from the geometry and/or the rate of mutations. Section 3.4 is devoted to the proofs. In the Appendix, we present and extend some technical results.

3.2 Convergence on the $\log K$-time scale

3.2.1 Model

We consider an individual-based Markov process that models the evolution of a haploid, asexually reproducing population. The space of possible traits is given by the vertices of a (possibly directed) finite graph $\mathcal{G} = (V, E)$.

For all traits $v, w \in V$ and every $K \in \mathbb{N}$, we introduce the following parameters:

- $b_v \in \mathbb{R}_+$, the birth rate of an individual of trait v,
- $d_v \in \mathbb{R}_+$, the (natural) death rate of an individual of trait v,
- $c^K_{v,w} = c_{v,w}/K \in \mathbb{R}_+$, the competition imposed by an individual of trait w onto an individual of trait v,
- $\mu_K \in [0, 1]$, the probability of mutation at a birth event,
- $m(v, \cdot) \in \mathcal{M}_p(V)$, the law of the trait of a mutant offspring produced by an individual of trait v.

The process N^K describes the state of the population, where $N^K_v(t)$ denotes the number of individuals of trait $v \in V$ alive at time $t \geq 0$. We assume that edges in E mark the possibility of mutation and hence $m(v, w) > 0$ if and only if $(v, w) \in E$.

Remark 3.1. We could also allow for μ_K to depend on $v \in V$ as long as $\mu_K(v) = \mu_K h(v)$ for some strictly positive function h that is independent of K. However, this would not change the characterisation of the limit, and hence we assume a constant μ_K to simplify the notation.

Moreover, we assume that, for every $v \in V$, $c_{v,v} > 0$. The parameter K is scaling the competitive pressure and, through this self-competition, fixes the equilibrium size of the population to the order of K. K is sometimes called *carrying capacity* and can be interpreted as a scaling parameter for the available sources of food or space.

As a consequence of our parameter definitions, the process N^K is characterised by its infinitesimal generator:

$$\begin{aligned}\mathcal{L}^K\phi(N) = &\sum_{v\in V}(\phi(N+\delta_v)-\phi(N))\left(N_v b_v(1-\mu_K)+\sum_{w\in V}N_w b_w \mu_K m(w,v)\right)\\ &+\sum_{v\in V}(\phi(N-\delta_v)-\phi(N))N_v\left(d_v+\sum_{w\in V}c^K_{v,w}N_w\right),\end{aligned} \tag{3.3}$$

where $\phi : \mathbb{N}^V \to \mathbb{R}$ is measurable and bounded. Such processes have been explicitely constructed in terms of Poisson random measures in [77].

Due to the scaling of the competition c^K, the equilibrium population is of order K. Since the mutation probability μ_K tends to zero as $K \to \infty$, the process N^K/K converges (on finite time intervals) to the mutation-free Lotka-Volterra system (3.5) involving all initial coexisting resident traits. We are interested in the long-term evolution of the population and want to study successive invasions by new mutant populations. Given the fact that a mutant population that is initially of order K^γ, $\gamma < 1$, needs a time of order $\log K$ to grow exponentially to the order of K, we have to rescale the time by $\log K$ to obtain a non trivial limit.

It is convenient to describe the population size of a certain trait $v \in V$ by its K-exponent

$$\beta^K_v(t) := \frac{\log(1+N^K_v(t\log K))}{\log K}, \tag{3.4}$$

which is equivalent to $N^K_v(t\log K) = K^{\beta^K_v(t)} - 1$. Since the population size is restricted to order K by the competition, β^K_v ranges between 0 and 1, as $K \to \infty$ (see Corollary 3.29 for a rigorous statement).

For the sake of readability, we now introduce the terminology we will use in the sequel.

Definition 3.2.

1. A trait $v \in V$ with exponent β^K_v is called *macroscopic* if, for every $\varepsilon > 0$, there exists K_ε such that, for every $K \geq K_\varepsilon$, $\beta^K_v > 1 - \varepsilon$.
2. A trait that is not macroscopic is called *microscopic*.
3. The set of *living traits* is the set $\{v \in V : \beta^K_v > 0\}$.

When K is large enough, the macroscopic traits interact on any finite time interval according to the corresponding mutation-free Lotka-Volterra system (see Chapter 11, Theorem 2.1 in [70] for the proof of this law of large numbers): Let $\mathbf{v} \subset V$, then the mutation-free Lotka-Volterra system associated to $\mathbf{v}$ is

$$\dot{n}_w(t) = \left(b_w - d_w - \sum_{v \in \mathbf{v}} c_{w,v} n_v(t) \right) n_w(t), \quad w \in \mathbf{v},\ t \geq 0. \tag{3.5}$$

For a subset $\mathbf{v} \subset V$ of traits, we denote by $\bar{n}(\mathbf{v}) \in \mathbb{R}_+^V$ the unique equilibrium of the Lotka-Volterra system (3.5), when it exists, and where to simplify notations, we extend it by $\bar{n}_w(\mathbf{v}) = 0$ for $w \notin \mathbf{v}$. In the case where $\mathbf{v} = \{v\}$, we obtain from classical results on Lotka-Volterra models (see [31] for instance)

$$\bar{n}_v(v) = (b_v - d_v)/c_{v,v} \vee 0. \tag{3.6}$$

If $\mathbf{v}$ denotes the set of macroscopic traits, we call the traits $v \in \mathbf{v}$ such that $\bar{n}_v(\mathbf{v}) > 0$ *resident*.

The approximate rate at which a mutant of trait w grows in a population of coexisting resident traits $\mathbf{v}$ is called *invasion fitness* and is denoted by $f_{w,\mathbf{v}}$, where

$$f_{w,\mathbf{v}} := b_w - d_w - \sum_{v \in \mathbf{v}} c_{w,v} \bar{n}_v(\mathbf{v}). \tag{3.7}$$

If $f_{w,\mathbf{v}} > 0$, the trait w is called *fit*. If $f_{w,\mathbf{v}} < 0$, the trait w is called *unfit*. The case $f_{w,\mathbf{v}} = 0$ will be excluded (see Remark 3.4).

Mutants can be produced along (directed) edges of the graph. We denote by $d(v,w)$ the graph distance, i.e. the length of the shortest (directed) path from v to w in $\mathcal{G} = (V,E)$. For a subset $\mathbf{v} \subset V$ we define

$$d(\mathbf{v},w) := \min_{v \in \mathbf{v}} d(v,w) \quad \text{and} \quad d(w,\mathbf{v}) := \min_{v \in \mathbf{v}} d(w,v). \tag{3.8}$$

3.2.2 Results

Let a finite graph $\mathcal{G} = (V,E)$ be given and assume that $\alpha \in \mathbb{R}_{>0} \setminus \mathbb{N}$ and $f_{w,\mathbf{v}} \neq 0$ for any $\mathbf{v} \subset V$ and $w \in V \backslash \mathbf{v}$ (see Remark 3.4). The two following results concern the convergence of the orders of the different subpopulation sizes to a piecewise linear trajectory, whose slopes and times of slope changes can be explicitly expressed in terms of the parameters.

Theorem 3.3. *Let a finite graph $\mathcal{G} = (V,E)$ and $\alpha \in \mathbb{R}_{>0} \setminus \mathbb{N}$ be given and consider the model defined by (3.3). Assume that $f_{w,\boldsymbol{v}} \neq 0$ for any $\boldsymbol{v} \subset V$ and $w \in V \backslash \boldsymbol{v}$. Let $\boldsymbol{v}_0 \subset V$ and assume that, for every $w \in V$,*

$$\beta_w^K(0) \to \left(1 - \frac{d(\boldsymbol{v}_0, w)}{\alpha} \right)_+, \quad (K \to \infty) \quad \textit{in probability.} \tag{3.9}$$

Then, for all $T > 0$, as $K \to \infty$, the sequence $((\beta_w^K(t), w \in V), t \in [0, T \wedge T_0])$ converges in probability in $\mathbb{D}([0, T \wedge T_0], \mathbb{R}_+^V)$ to a deterministic, piecewise affine, continuous function $((\beta_w(t), w \in V), t \in [0, T \wedge T_0])$, which is defined as follows:

(i) *If the mutation-free Lotka-Volterra system* (3.5) *associated to* $\boldsymbol{v}_0$ *has a unique positive globally attractive equilibrium, the initial condition of* β *is set to* $\beta_w(0) := \left(1 - \frac{d(\boldsymbol{v}_0,w)}{\alpha}\right)_+$. *Otherwise, the construction is stopped and* T_0 *is set to* 0.

(ii) *The increasing sequence of invasion times is denoted by* $(s_k)_{k\geq 0}$, *where* $s_0 := 0$ *and, for* $k \geq 1$,

$$s_k := \inf\{t > s_{k-1} : \exists w \in V\backslash \boldsymbol{v}_{k-1} : \beta_w(t) = 1\}. \tag{3.10}$$

Here, $\boldsymbol{v}_k$ *denotes the set of coexisting resident traits of the Lotka-Volterra system that includes* $\boldsymbol{v}_{k-1}$ *and the trait* $w \in V\backslash \boldsymbol{v}_{k-1}$ *that satisfies* $\beta_w(s_k) = 1$.

(iii) *For* $s_{k-1} \leq t \leq s_k$, *for any* $w \in V$, $\beta_w(t)$ *is defined by*

$$\beta_w(t) := \max_{u\in V}\left[\beta_u(s_{k-1}) + (t - t_{u,k} \wedge t) f_{u,\boldsymbol{v}_{k-1}} - \frac{d(u,w)}{\alpha}\right] \vee 0, \tag{3.11}$$

where, for any $w \in V$,

$$t_{w,k} := \begin{cases} \inf\{t \geq s_{k-1} : \exists\ u \in V : d(u,w) = 1, \beta_u(t) = \frac{1}{\alpha}\} & \text{if } \beta_w(s_{k-1}) = 0 \\ s_{k-1} & \text{else} \end{cases} \tag{3.12}$$

is the first time in $[s_{k-1}, s_k]$ *when this trait arises.*

(iv) *The inductive construction is stopped and* T_0 *is set to* s_k *if*

(a) *there is more than one* $w \in V\backslash \boldsymbol{v}_{k-1}$ *such that* $\beta_w(s_k) = 1$;

(b) *the Lotka-Volterra system including* $\boldsymbol{v}_{k-1}$ *and the unique* $w \in V\backslash \boldsymbol{v}_{k-1}$ *such that* $\beta_w(s_k) = 1$ *does not have a unique stable equilibrium;*

(c) *there exists* $w \in V\backslash \boldsymbol{v}_{k-1}$ *such that* $\beta_w(s_k) = 0$ *and* $\beta_w(s_k - \varepsilon) > 0$ *for all* $\varepsilon > 0$ *small enough.*

(d) *there exists* $w \in V\backslash \boldsymbol{v}_{k-1}$ *such that* $s_k = t_{w,k}$.

Remark 3.4. Notice that conditions (a), (c), and (d) of point (iv) are here to exclude very specific and non generic cases where one coordinate reaches 1 while another reaches 1 or reaches 0 from above, or a new trait arises at the exact same time. They are difficult to handle for technical reasons.

Moreover, we exclude the cases where $\alpha \in \mathbb{N}$. They would produce mutant populations, at distance α from the resident traits, that can neither be approximated by sub- nor super-critical branching processes. The same applies to the case $f_{w,\mathbf{v}} = 0$, where the population can both grow and shrink due to fluctuations.

Remark 3.5. The $t_{w,k}$ do not keep track of traits that die out in $[s_{k-1}, s_k]$ and then reappear. However, since the fitnesses do not change between invasions, such a trait would have a negative invasion fitness (else it would not die out). Hence, it would not start growing on its own if it reappears, but only follow along another trait due to mutants. It would therefore not contribute to the maximum over $u \in V$ in (3.11).

Proposition 3.6. *Under the same assumptions and with the same notations as in Theorem 3.3, for all $T > 0$, as $K \to \infty$, the sequence $((N_w^K(t \log K)/K, w \in V), t \in [0, T \wedge T_0])$ converges in probability in $\mathbb{D}([0, T \wedge T_0] \backslash \{s_k, k \geq 1\}, \mathbb{R}_+^V)$ to a deterministic jump process $((N_w(t), w \in V), t \in [0, T \wedge T_0])$, which is defined as follows:*

(i) For $t \in [0, T_0]$, $N(t)$ jumps between different Lotka-Volterra equilibria according to

$$N_w(t) := \sum_{k \in \mathbb{N}: s_{k+1} \leq T_0} \mathbf{1}_{s_k \leq t < s_{k+1}} \mathbf{1}_{w \in \boldsymbol{v}_k} \bar{n}_w(\boldsymbol{v}_k). \tag{3.13}$$

(ii) The invasion times s_k and the times $t_{w,k}$ when new mutants arise can be calculated as follows. We define the increasing sequence $(\tau_\ell, \ell \geq 0) = \{s_k, k \geq 0\} \cup \{t_{w,k}, w \in V, k \geq 0\}$ of invasion times or appearance times of new mutants, and $(M_\ell, \ell \geq 0)$ the sets of living traits in the time interval $(\tau_\ell, \tau_{\ell+1}]$. Initially, $\tau_0 = s_0 = 0$ and, according to (3.9), $M_0 = \{w \in V : d(\boldsymbol{v}_0, w) < \alpha\} = \{w \in V : \beta_w(0) > 0\}$. For $s_{k-1} \leq \tau_{\ell-1} < s_k$, τ_ℓ is defined as

$$\tau_\ell := s_k \wedge \min\{t_{w,k} : w \in V, t_{w,k} > \tau_{l-1}\}. \tag{3.14}$$

Given τ_ℓ and $M_{\ell-1}$, we set $M_\ell := (M_{\ell-1} \backslash \{w \in V : \beta_w(\tau_\ell) = 0\}) \cup \{w \in V : \tau_\ell = t_{w,k}\}$. τ_ℓ is then given by

$$\tau_\ell - \tau_{\ell-1} = \min_{\substack{w \in M_{\ell-1}: \\ f_{w, \boldsymbol{v}_{\ell-1}} > 0}} \frac{\left(1 \wedge \frac{d(w, V \backslash M_{\ell-1})}{\alpha}\right) - \beta_w(\tau_{\ell-1})}{f_{w, \boldsymbol{v}_{\ell-1}}}. \tag{3.15}$$

Remark 3.7. We could allow for more general initial conditions of the form

$$\beta_w^K(0) \to \tilde{\beta}_w \in [0, 1], \tag{3.16}$$

with $\tilde{\beta}_w$, $w \in V$, deterministic and $\mathbf{v}_0 := \{w \in V : \tilde{\beta}_w = 1\} \neq \emptyset$. An inductive application of Lemma 3.25, similar to the induction proving (3.75), implies that within a time of order 1, for all $w \in V$, $\beta_w^K \cong \max_{u \in V}[\tilde{\beta}_u - d(u, w)/\alpha]_+$. We therefore set $\beta_w(0) := \max_{u \in V}[\tilde{\beta}_u - d(u, w)/\alpha]_+$ in Theorem 3.3 and $M_0 := \{w \in V : \beta_w(0) > 0\}$ in Proposition 3.6. The rest of the results remains unchanged.

Remark 3.8. The limiting jump process $N(t)$ resembles an adaptive walk or flight, as studied in [150, 144, 161, 147, 13]. For a constant competition kernel $c_{v,w} \equiv c$, we consider the fixed fitness landscape given by $r_v = b_v - d_v$. Since in this case $f_{w,v} = r_w - r_v$, the process jumps along edges towards traits of increasing fitness r.

The above results are in the vein of Theorem 2.1 and Corollary 2.3 in [38]. There are however many differences between the setting considered in [38] and our setting.

Due to the horizontal transfer between individuals, Champagnat and coauthors obtained trajectories where a "dominant" population, i.e. with the size of highest order, could be non resident, i.e. of order negligible with respect to K. They could also witness extinction on a $\log K$ time scale as well as evolutionary suicide. The absence of horizontal transfer in our case prevents such behaviours.

We consider a general finite graph of mutations with possible back mutations, whereas their graph was embedded in $\mathbb{Z}$ and did not allow for back mutations. We also allow for the coexistence of several resident traits in the population at equilibrium. The two main difficulties in the proofs compared to [38] are thus to handle the generality of the graph of mutations, and to extend some approximation results to the multidimensional case.

3.3 Surprising phenomena arising from geometry and mutation rate

In this section, we present some non intuitive behaviours of the population process, which stem from the mutation scale or the generality of the mutational graph that we allow for. They are direct applications of Theorem 3.3 and Proposition 3.6, and provide explicit computations of exponents (3.11) and time intervals (3.15).

Several examples are build on directed graphs. Although this is not a necessary condition to obtain the desired phenomena, it allows a simplified study (especially of the decay phases).

We first introduce some notations for the sake of readability.

Definition 3.9. Let $w, v \in V$ and $\mathbf{v} \subset V$. We write

1. *with high probability* to mean "with a probability converging to 1 as $K \to \infty$",
2. $w > \mathbf{v}$ if and only if $f_{w,\mathbf{v}} > 0$, that is if w can invade in $\mathbf{v}$,
3. $w < \mathbf{v}$ if and only if $f_{w,\mathbf{v}} < 0$, that is if w cannot invade in $\mathbf{v}$,
4. $w \gg v$ (or $v \ll w$) if and only if $f_{w,v} > 0$ and $f_{v,w} < 0$, that is if w can invade in v and fixate,
5. $w \equiv v$ if and only if $f_{w,v} > 0$ and $f_{v,w} > 0$, that is if w and v can coexist,
6. $w \frown v$ if and only if $f_{w,v} < 0$ and $f_{v,w} < 0$, that is if w and v can neither invade in each other.

3.3.1 Back mutations before adaptation

In the following, we build an example where the ancestry of the resident population comes from back mutations from an ancestral trait, even if the mutations happening in between are not deleterious.

Example 3.10. *Let us consider the graph $\mathcal{G}$ depicted in Figure 3.1 where $V = \{0, 1, 2, 3\}$ and $E = \{[0, 1], [1, 2], [2, 0], [0, 3]\}$. Let $\alpha > 2$, an initial condition given by $(\bar{n}(0), 0, 0, 0)$ and*

a fitness landscape given by

$$0 \ll 1 \ll 2, \quad 3 \frown 0, \quad 3 \frown 1 \tag{3.17}$$

$$0 \equiv 2, \quad 3 > \{0, 2\}, \quad 2 < 3 \tag{3.18}$$

$$f_{2,0} < 2f_{1,0} \tag{3.19}$$

$$f_{0,1} \geq f_{3,1}, \quad f_{2,0} \leq f_{1,0} \tag{3.20}$$

$$i_1 := \frac{1 - 1/\alpha}{f_{0,1}} < \frac{-(1 - 4/\alpha)}{f_{2,1}} =: i_2 \tag{3.21}$$

In this case, Proposition 3.6 implies that on the $\log K$ time scale, the rescaled macroscopic population then jumps from traits $0 - 1 - 2$ then to coexistence between 0 and 2, followed by the invasion and fixation of 3 which is produced with high probability, due to Condition (3.21), by individuals of type 0 which have the sequence $0 - 1 - 2$ as ancestry. In other words, the final resident individuals of trait 3, although they can be produced by individuals of trait 0 directly, come from a sequence of mutations which went around the loop $0 - 1 - 2$ of $\mathcal{G}$. Conditions (3.18), summarised in Figure 3.1, imply phase portrait number 8 in the classification of Zeeman [180]. Condition (3.19) ensures that trait 1 becomes resident before 2. Condition (3.20) is not necessary but allows to simplify the setting. The exponents are drawn in Figure 3.1.

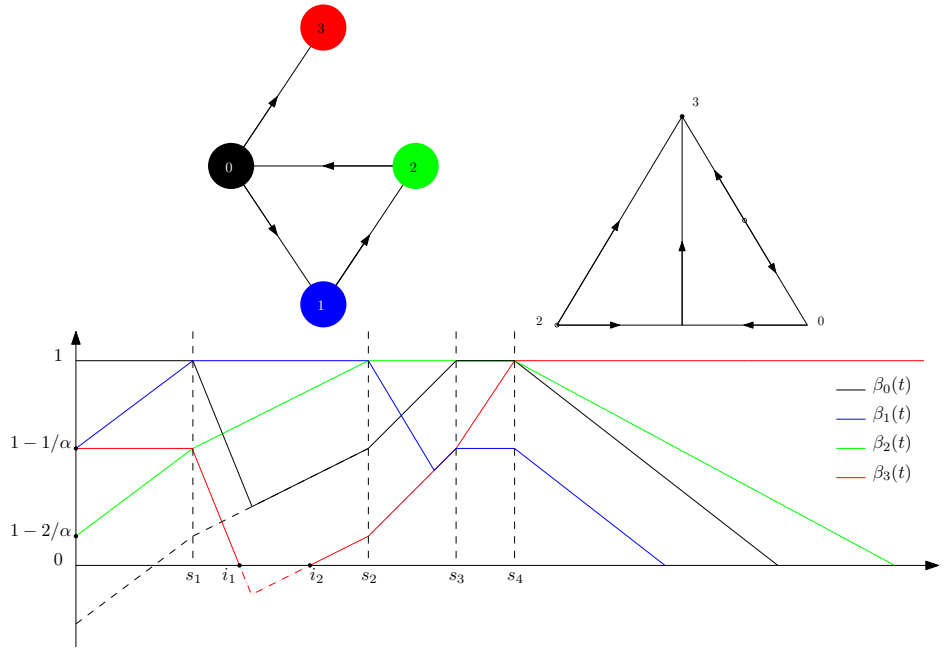


Figure 3.1: Graph $\mathcal{G}$, phase portrait of traits 0-2-3, and exponents $\beta(t)$ of Example 3.10.

Note that this scenario can also happen in the rare mutation regime considered in [31] (for example $\alpha \in (0, 1)$): the average waiting time until a mutant of type 1 appears is then of order $O(1/K\mu_K) = O(K^{-1+1/\alpha}) \gg \log K$. Once it has appeared, it survives with positive probability and the succession of invasions and fixations above takes place on the $\log K$ time scale, separated by mutation events on the $K^{-1+1/\alpha}$ time scale. What is new in our case is

that such a scenario can still take place for higher mutation rates than the ones considered in [31], and on a $\log K$ time scale.

3.3.2 Non-intuitive mutational pathways in the high mutation framework

Longer or shorter path than expected

If evolution and mutation time scales are separated (i.e. in the rare mutation regime), mutations occur one at a time, and the number of successive resident traits from the wild type to the type gathering k successively beneficial mutations is k. This is not the case if mutations are faster, in which case it is possible to observe either more or less resident traits, as the following examples show.

Example 3.11. *Let us consider the directed graph $\mathcal{G}$ depicted in Figure 3.2, where $V = \{00, 01, 10, 11\}$ and $E = \{[00, 01], [00, 10], [01, 11], [10, 11]\}$. Let $\alpha > 2$, an initial condition given by $(\bar{n}(00), 0, 0, 0)$ and a fitness landscape given by*

$$00 \ll 01 \ll 10 \ll 11, \quad \textit{and} \quad 01 \ll 11 \tag{3.22}$$

$$10, 11 \frown 00 \tag{3.23}$$

$$f_{11,00} < f_{01,00} \tag{3.24}$$

$$f_{10,01} > f_{11,01} \tag{3.25}$$

In this case, in the rare mutation regime, the rescaled macroscopic population jumps along $00 - 01 - 11$.

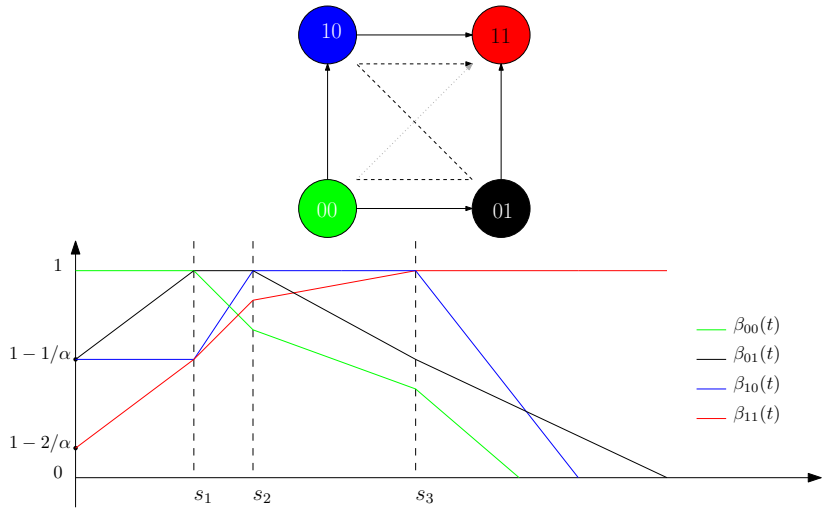


Figure 3.2: Graph $\mathcal{G}$ for Examples 3.11 and 3.12 and exponents $\beta(t)$ for Example 3.11.

In the regime of Theorem 3.3, Proposition 3.6 implies that the rescaled macroscopic population jumps along $00 - 01 - 10 - 11$ on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 3.2. Note that Condition (3.24) ensures that 11 does not invade before

01, it is not necessary but allows to simplify the setting. Condition (3.25) ensures that 11 does not invade before 10.

Example 3.12. *Let us consider the directed graph $\mathcal{G}$ depicted in Figure 3.2, where $V = \{00, 01, 10, 11\}$ and $E = \{[00, 01], [00, 10], [01, 11], [10, 11]\}$. Let $\alpha > 2$, an initial condition given by $(\bar{n}(00), 0, 0, 0)$ and a fitness landscape given by*

$$01 > 00, \quad 11 \gg 00 \tag{3.26}$$

$$10 < 00 \tag{3.27}$$

$$01, 10 < 11 \tag{3.28}$$

$$\frac{2}{f_{11,00}} < \frac{1}{f_{01,00}} \tag{3.29}$$

In this case, in the rare mutation regime, the rescaled macroscopic population still jumps along $00 - 01 - 11$, under the additional assumption that $f_{11,01} > 0$.

In the regime of Theorem 3.3, Proposition 3.6 implies that the rescaled macroscopic population directly jumps from 00 to 11 on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 3.3. Condition (3.29) ensures that 11 fixates before 01. Condition (3.27) is not necessary but allows to simplify the setting. Note that equation (3.15) implies that $s_1 = 2/f_{11,00}$ and $\tilde{s}_1 = 1/f_{01,00}$.

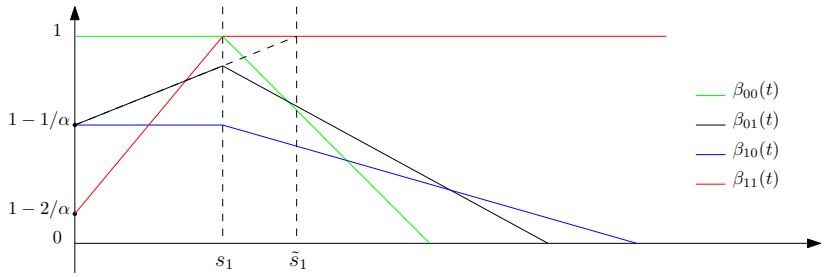


Figure 3.3: Exponents $\beta(t)$ for Example 3.12.

Price of anarchy

We build an example where adding a new possible mutation path to a fit trait increases the time until it appears macroscopically.

Example 3.13. *Let us consider the graph $\mathcal{G}$ depicted in Figure 3.4, where $V = \{1, 2a, 2b, 3\}$ and the edge set is either $E_1 = \{[1, 2a], [2a, 3], [2b, 3], [3, 2b]\}$ or $E_2 = \{[1, 2a], [2a, 3], [2b, 3], [2a, 2b], [3, 2b]\}$. Let $\alpha > 3$, an initial condition given by $(\bar{n}(1), 0, 0, 0)$ and a fitness landscape given by*

$$1 \ll 2a \ll 3, \quad \textit{and} \quad 2a \ll 2b \tag{3.30}$$

$$1 < 2b, \quad \textit{and}, \quad 1, 2b < 3 \tag{3.31}$$

$$f_{2a,1} \geq f_{3,1}, f_{2b,1}, \quad \textit{and} \quad \frac{1}{f_{2b,2a}} < \frac{1}{f_{3,2a}} < \frac{2}{f_{2b,2a}} \tag{3.32}$$

$$0 < f_{3,2b} < f_{3,2a}. \tag{3.33}$$

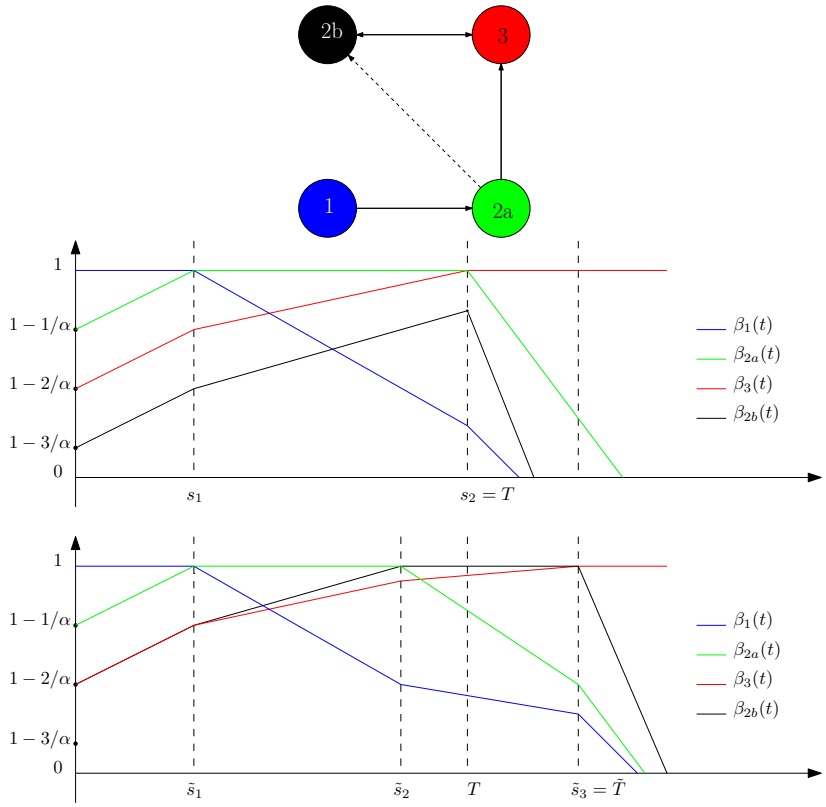


Figure 3.4: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.13, with edge set E_1 (above) and E_2 (below).

In this case, if the edge set is E_1, Proposition 3.6 implies that the rescaled macroscopic population jumps along traits $1-2a-3$ in a time T on the $\log K$ time scale. But if the edge set is E_2, the population jumps along $1-2a-2b-3$ and the time to reach 3 is $\tilde{T} > T$. More precisely, the exponents are drawn in Figure 3.4. Condition (3.32) ensures that $2b$ invades first when the edge set is E_2 but not when it is E_1, in other words β_{2b} reaches 1 before β_3 if started at $1-1/\alpha$ but not at $1-2/\alpha$. And Condition (3.33) enlarges the time of fixation of 3. Note that the first inequality in Condition (3.32) is not necessary but allows to simplify the second one. Moreover, observe that equation (3.15) implies $\tilde{s}_2 - \tilde{s}_1 = 1/f_{2b,2a}$ and $s_2 - s_1 = 1/f_{3,2a}$. Note that in the rare mutation regime we can observe this phenomenon on the mutation time scale, but only with probability strictly smaller than 1, since both $2b$ and 3 are fit with respect to $2a$ and can both invade with positive probability once they are produced.

Counter-cycle

Example 3.14. *Let us consider the graph $\mathcal{G}$ depicted in Figure 3.5, where $V = \{1,2,3\}$ and the edge set is $E = \{[1,2],[2,3],[3,1]\}$. Let $\alpha > 2$, an initial condition given by $(\bar{n}(1),0,0)$ and a fitness landscape given by*

$$1 \gg 2, \quad 2 \gg 3, \quad 3 \gg 1. \tag{3.34}$$

In this case, Proposition 3.6 implies that the rescaled macroscopic population jumps along traits $1-3-2$ (in the clockwise sense) although the mutations are directed counter-clockwise. More precisely, the exponents are drawn in Figure 3.5. Moreover, if Conditions (3.37) below are fulfilled the period is shorter and shorter, and acceleration takes place, as it is depicted in Figure 3.5.

$$f_{2,3} > -f_{1,3} \tag{3.35}$$
$$f_{1,2} > -f_{3,2} \tag{3.36}$$
$$f_{3,1} > -f_{2,1} \tag{3.37}$$

Note that in the rare mutation regime, with the chosen parameters, there would be no evolution since $2 < 1$. Moreover, there are no parameters such that counter-cyclic or accelerating behaviours could arise.

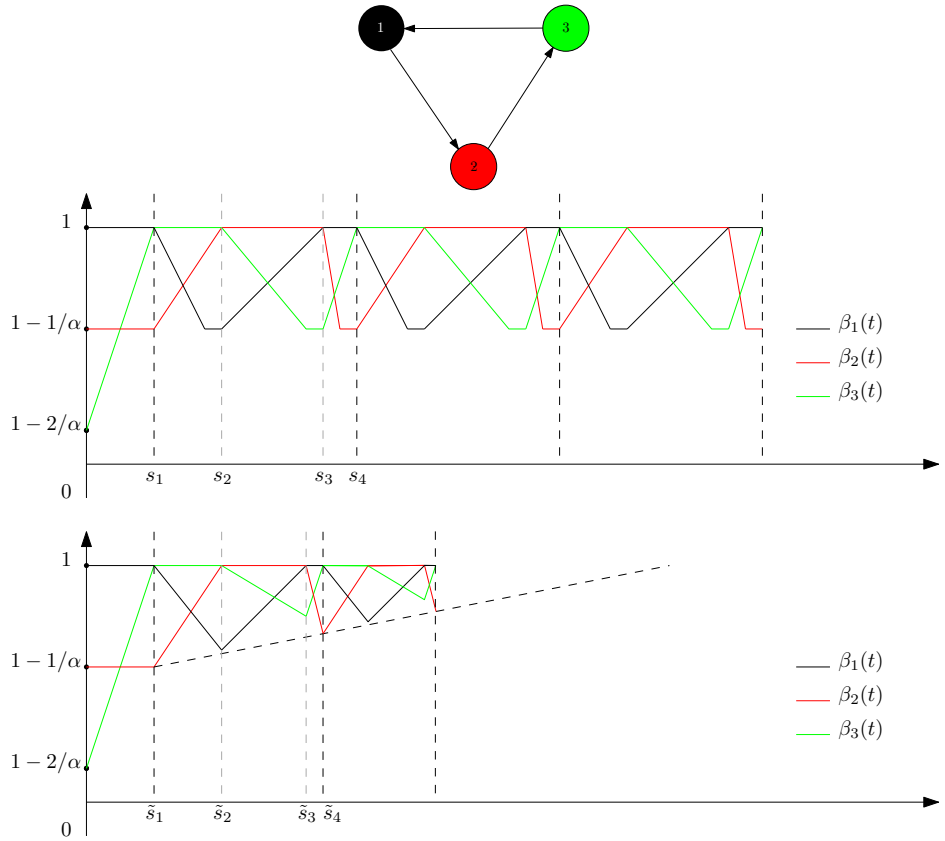


Figure 3.5: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.14, without Assumption 3.37 (above) and with Assumption 3.37 (below).

3.3.3 Arbitrary large jumps on the $\log K$-time scale

A natural question to ask is if the "cut-off" α restricts the range of the jumps, on the $\log K$ time-scale, to traits which are at a distance less than α. The answer is no, as the following

example shows.

Example 3.15. *Let us consider the graph $\mathcal{G}$ depicted in Figure 3.6, where $V = \{0,1,2,3,4\}$ and $E = \{[0,1],[1,2],[2,3],[3,4]\}$. Let $\alpha \in (3,4)$, an initial condition given by $(\bar{n}(0),0,\dots,0)$ and a fitness landscape given by*

$$1,2 < 0, \quad 3,4 > 0, \quad 0,1,2,3 < 4 \tag{3.38}$$

$$\frac{1}{f_{4,0}} + \frac{-1+4/\alpha}{f_{3,0}} < \frac{3/\alpha}{f_{3,0}} \tag{3.39}$$

In this case, the cut-off is in between traits 3 and 4 (meaning that $K\mu_K^i \to 0$ for $i > 3$) and thus the population of trait 4 vanishes at time 0. However, Proposition 3.6 implies that the rescaled macroscopic population jumps from trait 0 to trait 4 in a time

$$s_1 = \frac{-1+4/\alpha}{f_{30}} + \frac{1}{f_{40}} \tag{3.40}$$

on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 3.6. Condition (3.39) ensures that trait 4 fixates before trait 3.

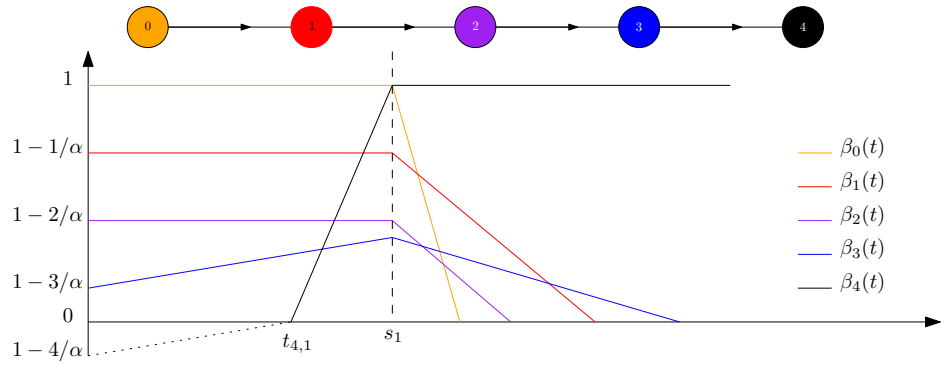


Figure 3.6: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.15.

It is easy to generalise this example to construct jumps to any distance L larger than α, by taking larger and larger fitnesses after the negative fitness region. The condition implying emergence of trait L is then a little more technical to write, since one has to compute the time for the piecewise affine function $\beta_L(t)$ (with multiple slope-breaks) to reach 1 before the other traits. Example 3.15 constitutes the simplest non-trivial example of this phenomenon. Example 3.16 is a further case where a more distant trait fixates, and two intermediate times $t_{4,1}$ and $t_{5,1}$ occur (recall the definition in (3.12)).

Example 3.16. *Let us consider the graph $\mathcal{G}$ depicted in Figure 3.7, where $V = \{0,1,2,3,4,5\}$ and $E = \{[0,1],[1,2],[2,3],[3,4],[4,5]\}$. Let $\alpha \in (3,4)$, an initial condition given by $(\bar{n}(0),0,\dots,0)$ and a fitness landscape given by*

$$1,2 < 0, \quad 3,4,5 > 0, \quad 0,1,2,3,4 < 5 \tag{3.41}$$

$$f_{3,0} < f_{4,0} < f_{5,0} \quad \textit{and} \quad \frac{-1+4/\alpha}{f_{3,0}} + \frac{-4/\alpha+5/\alpha}{f_{4,0}} + \frac{1}{f_{5,0}} < \frac{3/\alpha}{f_{3,0}}. \tag{3.42}$$

In this case, the cut-off is in between traits 3 and 4 (meaning that $K\mu_K^i \to 0$ for $i > 3$) thus population of trait 4 and 5 vanishes at time 0. However, Proposition 3.6 implies that the rescaled macroscopic population jumps from trait 0 to trait 5 in a time

$$s_1 = \frac{-1+4/\alpha}{f_{3,0}} + \frac{-4/\alpha+5/\alpha}{f_{4,0}} + \frac{1}{f_{5,0}} \tag{3.43}$$

on the $\log K$ time scale. More precisely, the exponents are drawn in Figure 3.7. Condition (3.42) ensures that trait 5 fixates before traits 3 and 4. The first inequality is not needed but allows to simplify the second one. The dotted lines in the figures allow to construct the points where some exponents become positive.

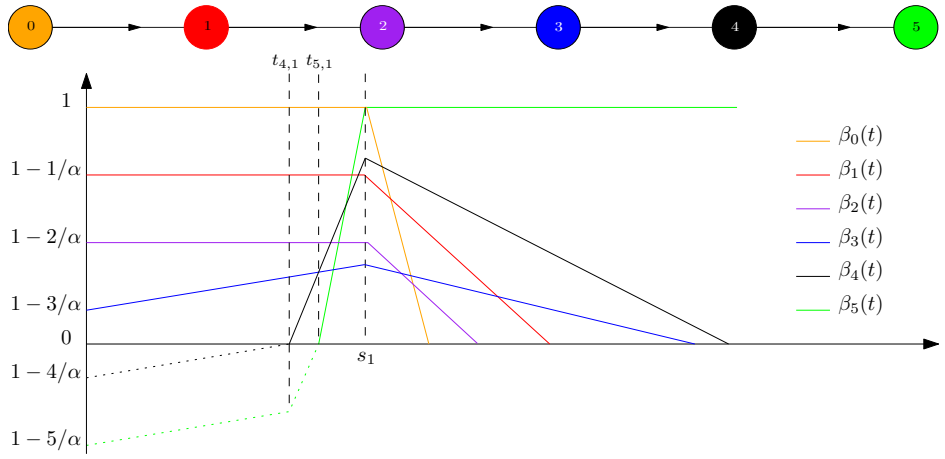


Figure 3.7: Graph $\mathcal{G}$ and exponents $\beta(t)$ for Example 3.16.

3.3.4 Effective random walk across fitness valleys

2 effective sites

Example 3.17. *Let us consider the graph $\mathcal{G}$ depicted in Figure 3.8, where $V = \{0, 1a, 1b, i\}$ and $E = \{[0,i],[i,1a],[1a,1b],[i,1b]\}$. We suppose that whenever there are several outgoing edges from a vertex v, the mutation kernel is uniform among the nearest neighbouring vertices. Let $\alpha \in (0,1)$, an initial condition given by $(\bar{n}(0), 0, \dots, 0)$ and a fitness landscape given by*

$$1a \gg 0 \gg 1b \gg 1a \tag{3.44}$$

$$i < 0, \quad i < 1a, \quad i < 1b. \tag{3.45}$$

In this case, according to [25], the time to cross the fitness valley is of order $O(1/K\mu_K^2) = O(K^{-1+2/\alpha}) \gg \log K$, thus the first mutant of type $1a$ will appear on this time scale, and will invade with positive probability. Then, in a time of order $O(1/K\mu_K)$, type $1b$ fixates, and one has to wait again a time of order $O(K^{-1+2/\alpha})$ until the appearance

of the next mutant of type 0. Thus, on the time scale $O(K^{-1+2/\alpha})$, the population process converges to a jump process between the two states 0 and $1b$ with positive jump rates although the fitness $f_{1b,0}$ is negative. More precisely, following [25] we define

$$\lambda(\rho) := \sum_{\ell=1}^{\infty} \frac{(2\ell)!}{(\ell-1)!(\ell+1)!}\rho^{\ell}(1-\rho)^{\ell+1} \tag{3.46}$$

$$\rho_{w,v} := \frac{b_w}{b_w + d_w + c_{w,v}\bar{n}(v)} \tag{3.47}$$

Theorem 3.18. *As $K \to \infty$, the following convergence holds*

$$(N_0^K, N_{1b}^K)(tK^{-1+2/\alpha}) \Rightarrow \bar{n}(X_t)\delta_{X_t} \tag{3.48}$$

for finite dimensional distributions, where X_t is a continuous time Markov chain on $\{0, 1b\}$ with transition rates

$$r_{0\to 1b} = \bar{n}_0 b_0 \frac{\lambda(\rho_{i,0})}{2}\frac{f_{1a,0}}{b_{1a}} \tag{3.49}$$

$$r_{1b\to 0} = \bar{n}_{1b} b_{1b} \frac{\lambda(\rho_{i,1b})}{2}\frac{f_{0,1b}}{b_0}. \tag{3.50}$$

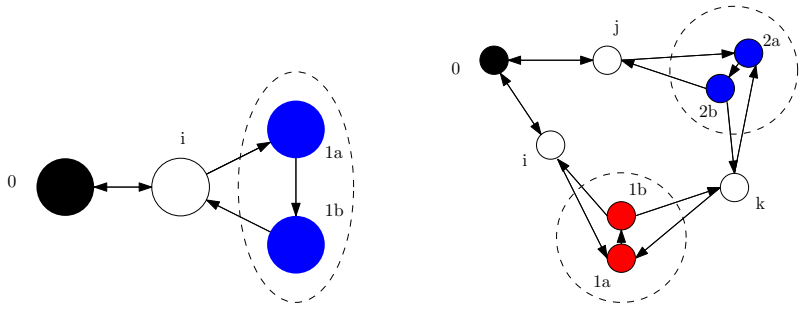


Figure 3.8: Graphs $\mathcal{G}$ of Examples 3.17 and 3.19.

3 effective sites

Example 3.19. *Let us consider the graph $\mathcal{G}$ depicted in Figure 3.8. We suppose that whenever there are several outgoing edges from a vertex v, the mutation kernel is uniform among the nearest neighbouring vertices. Let $\alpha \in (0,1)$, an initial condition given by $(\bar{n}(0), 0, \ldots, 0)$ and a fitness landscape given by*

$$\begin{array}{lll} 1a \gg 0 & 2a \gg 0 & 2a \gg 1b \\ 1b \gg 1a & 2b \gg 2a & 1a \gg 2b \\ 0 \gg 1b & 0 \gg 2b & 1b \gg 2b \\ i < 0 & j < 0 & k < 1a,\ k < 1b \\ i < 1a,\ i < 1b & j < 2a,\ j < 2b & k < 2a,\ k < 2b \end{array} \tag{3.51}$$

Thus, following [25], on the time scale $O(K^{-1+2/\alpha})$, the population process converges to a jump process between the three states $\{0, 1b, 2b\}$ with positive jump rates. More precisely,

Theorem 3.20. *As $K \to \infty$, the following convergence holds*

$$(N_0^K, N_{1b}^K, N_{2b}^K)(tK^{-1+2/\alpha}) \Rightarrow \bar{n}(X_t)\delta_{X_t} \tag{3.52}$$

for finite dimensional distributions, where X_t is a continuous time Markov chain on $\{0, 1b, 2b\}$ with transition rates:

$$\begin{aligned} r_{0\to 1b} &= \tfrac{\bar{n}_0 b_0}{2}\tfrac{\lambda(\rho_{i,0})}{2}\tfrac{f_{1a,0}}{b_{1a}}, & r_{1b\to 0} &= \tfrac{\bar{n}_{1b} b_{1b}}{2}\tfrac{\lambda(\rho_{i,1b})}{2}\tfrac{f_{0,1b}}{b_0} \\ r_{0\to 2b} &= \tfrac{\bar{n}_0 b_0}{2}\tfrac{\lambda(\rho_{j,0})}{2}\tfrac{f_{2a,0}}{b_{2a}}, & r_{2b\to 0} &= \tfrac{\bar{n}_{2b} b_{2b}}{2}\tfrac{\lambda(\rho_{j,2b})}{2}\tfrac{f_{0,2b}}{b_0} \\ r_{1b\to 2b} &= \tfrac{\bar{n}_{1b} b_{1b}}{2}\tfrac{\lambda(\rho_{k,1b})}{2}\tfrac{f_{2a,1b}}{b_{2a}}, & r_{2b\to 1b} &= \tfrac{\bar{n}_{2b} b_{2b}}{2}\tfrac{\lambda(\rho_{k,2b})}{2}\tfrac{f_{1a,2b}}{b_{1a}}. \end{aligned} \tag{3.53}$$

3.4 Proof of Theorem 3.3 and Proposition 3.6

This section is dedicated to the proofs of our main results. As they are technical and involve many stopping times, we begin with a rough outline of the strategy of the proof.

Throughout the proof, we define several stopping times to divide the times between invasions into sub-steps. Heuristically they correspond to the following events:

- σ_k^K, the time when the k^{th} invasion has taken place and a new equilibrium is reached.
- $\theta_{k,m,C}^K$, the first time after σ_{k-1}^K when either the macroscopic traits stray too far from their equilibrium or at least one of the (formerly) microscopic traits becomes macroscopic (recall Definition 3.2)
- s_k^K, the first time after σ_{k-1}^K when a microscopic trait becomes almost macroscopic, i.e. reaches an order of $K^{1-\varepsilon_k}$.
- $t_{w,k}^K$, the first time after σ_{k-1}^K when trait w has a positive population size. ($t_{w,k}^K = \sigma_{k-1}^K$ for all traits that are alive at this time.)

As in Proposition 3.6, $(\tau_\ell^K, \ell \geq 0)$ is the collection of $(s_k^K, k \geq 0)$ and $(t_{w,k}^K, k \geq 0, w \in V)$. Figure 3.9 visualises the different stopping times for the case of one macroscopic and two microscopic traits.

The proof consists of five parts:

1. In the longest and most involved part of the proof, we study the growth dynamics of the different subpopulations in the time interval $[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$, making use of several results from [38], which are restated in the Appendix, and generalised when needed. Similar to [112], we prove lower and upper bounds for $\beta_w^K(t)$ via an induction, successively taking into account incoming mutants originating from traits of increasing distance to w. We prove that $\beta_w^K(t)$ follows the characterisation of $\beta_w(t)$ in Theorem 3.3 up to an error of order ε_k for large K.
2. We construct the sets M_ℓ^K and calculate the value of $\tau_\ell^K - \tau_{\ell-1}^K$, proving part (ii) of Proposition 3.6.

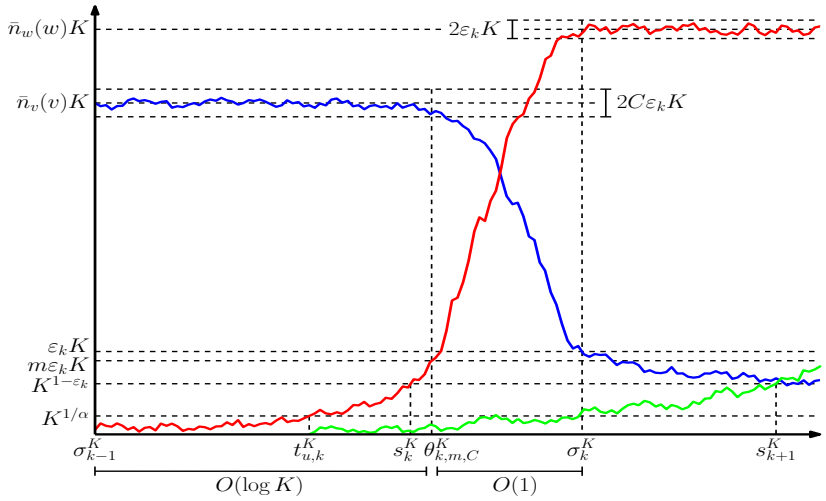


Figure 3.9: Schematic evolution of macroscopic trait v (blue) and microscopic traits $w = l_k^K$ (red) and u (green), where $d(w,u) = 1$, during the k^{th} invasion step.

3. We prove that s_k^K and $\theta_{k,m,C}^K$ are equal up to an error η_k that goes to zero as $\varepsilon_k \to 0$ and conclude that s_k^K converges to s_k when $K \to \infty$.

4. We prove that the stopping time $\theta_{k,m,C}^K$ is triggered by a (formerly) microscopic trait reaching order K, and not by the macroscopic traits deviating from their equilibrium.

5. Knowing that we have non-vanishing population sizes at $\theta_{k,m,C}^K$, we finally consider the Lotka-Volterra phase involving $\mathbf{v}_{k-1}$ and the trait l_k^K that has newly reached order K, proving that the initial conditions for the next step, characterised in the definition of σ_k^K, are satisfied after a time of order 1. This concludes the proof of Theorem 3.3.

Since part (i) of Proposition 3.6 is a direct corollary of Theorem 3.3, this concludes the proofs of both results.

Recall the definitions provided in Theorem 3.3 and Proposition 3.6, and for a given set $\mathbf{v} \subset V$, introduce $\tilde{\mathbf{v}}$ the support of the mutation free Lotka-Volterra equilibrium associated to $\mathbf{v}$, that is to say

$$w \in \tilde{\mathbf{v}} \Leftrightarrow \bar{n}_w(\mathbf{v}) > 0. \tag{3.54}$$

Similarly as in [38], the strategy of the proof consists in performing an induction on successive phases k, during which the population sizes of the set of traits $\tilde{\mathbf{v}}_k$ are close to their equilibrium value and the population sizes of the set of traits $V \setminus \tilde{\mathbf{v}}_k$ are small with respect to K. To be more precise, we will introduce a sequence of stopping times $(\sigma_k^K \log K, k \in \mathbb{N})$ (see definition in (3.140)) satisfying the following conditions, as soon as $s_k < T$:

Assumption 3.21.

1. $\sigma_k^K \to s_k$ in probability when K goes to infinity

2. For any $0 < \varepsilon_k < 1 \wedge \inf_{w\in\tilde{\mathbf{v}}_k} \bar{n}_w(\mathbf{v}_k)$, with high probability

   a) For every $w \in \tilde{\mathbf{v}}_k$,

$$\frac{N_w^K(\sigma_k^K \log K)}{K} \in [\bar{n}_w(\mathbf{v}_k) - \varepsilon_k, \bar{n}_w(\mathbf{v}_k) + \varepsilon_k] \tag{3.55}$$

   b) For every $w \in \mathbf{v}_k \setminus \tilde{\mathbf{v}}_k$,

$$K^{1-\varepsilon_k} \leq N_w^K(\sigma_k^K \log K) \leq \varepsilon_k K. \tag{3.56}$$

   c) There exists $\bar{c}_k < \infty$ such that for every $w \notin \mathbf{v}_k$, either $N_w^K(\sigma_k^K \log K) = 0$ if $\beta_w(s_k) = 0$ or

$$0 < \beta_w(s_k) - \bar{c}_k\varepsilon_k < \frac{\log\left(1 + N_w^K(\sigma_k^K \log K)\right)}{\log K} = \beta_w^K(\sigma_k^K) < \beta_w(s_k) + \bar{c}_k\varepsilon_k < 1. \tag{3.57}$$

To be more precise, for $k \geq 1$, the time interval $[\sigma_{k-1}^K \log K, \sigma_k^K \log K]$ will be divided into two parts:

- a 'stochastic phase' $[\sigma_{k-1}^K \log K, \theta_{k,m,C}^K \log K]$ needed for the trait

$$l_k^K := \mathbf{v}_k^K \setminus \mathbf{v}_{k-1}^K \tag{3.58}$$

  to reach a size of order K,
- a 'deterministic phase' $[\theta_{k,m,C}^K \log K, \sigma_k^K \log K]$ needed for the mutation free Lotka-Volterra system associated to $\tilde{\mathbf{v}}_{k-1}^K \cup l_k^K$ to reach a neighbourhood of its equilibrium.

Initialisation of the induction.

• $\underline{\sigma_0^K}$: By assumption,

$$\beta_w^K(0) \overset{K\to\infty}{\to} \left(1 - \frac{d(\mathbf{v}_0, w)}{\alpha}\right)_+. \tag{3.59}$$

Let us choose a small $\varepsilon_0 > 0$. Then from point (ii) of Lemma 3.28, there exists a deterministic $T(\varepsilon_0) < \infty$ such that

$$\lim_{K\to\infty} \mathbb{P}\left(\|N^K(T(\varepsilon_0))/K - \bar{n}(\mathbf{v}_0)\|_\infty \leq \varepsilon_0\right) = 1. \tag{3.60}$$

Define $\sigma_0^K := T(\varepsilon_0)/\log K$. We can check that σ_0^K is a stopping time converging in probability to $s_0 = 0$ and satisfying Assumption 3.21. Moreover we know that the processes $\beta_w^K, w \in V$, vary on a time scale of order $\log K$ (see [31, 38] for instance). In particular, they do not vary during the time $T(\varepsilon_0)$ in the large K limit. This entails that σ_0^K satisfies

Assumption 3.21.

• $\underline{\sigma_k^K, k \geq 1}$:

Assume that $s_{k-1} < T_0$ and that $\sigma_{k-1}^K \log K$ is a stopping time satisfying Assumption 3.21. We will now construct σ_k^K.

3.4.1 Definitions and first properties

Let us introduce a small $\varepsilon_k > 0$ as well as a stopping time $\theta_{k,m,C}^K \log K$ via

$$\theta_{k,m,C}^K := \inf \left\{ t \geq \sigma_{k-1}^K, \exists w \in \mathbf{v}_{k-1}, \left| \frac{N_w^K(t \log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1}) \right| \geq C\varepsilon_k \right. \\ \left. \text{or} \sum_{w \notin \mathbf{v}_{k-1}} N_w^K(t \log K) \geq m\varepsilon_k K \right\}. \tag{3.61}$$

The conditions satisfied by $m > 0$ and $C > 0$ will be precised later on. m is typically small, see (3.136). The conditions satisfied by C will be specified in Subsection 3.4.4.

We will now finely study the population dynamics on the time interval $[\sigma_{k-1}^K \log K, (\theta_{k,m,C}^K \wedge T) \log K]$. To this aim, we will couple the subpopulations of individuals with a given trait with branching processes with immigration and use results on these processes derived in [38] and recalled (and generalised when needed) in the Appendices. The main difficulty of this step comes from the fact that as we allow for any finite graph of mutations, the immigration rate for a particular subpopulation may vary a lot on the time interval $[\sigma_{k-1}^K \log K, (\theta_{k,m,C}^K \wedge T) \log K]$. This is why we introduced in Proposition 3.6 the sequence of times $(\tau_\ell, \ell \in \mathbb{N})$, which corresponds to the times when mutants of a new type arise or a formerly microscopic trait becomes of order K.

Notice that although we make extensive use of the techniques and results developed in [38], the authors of this paper considered a specific graph embedded in $\mathbb{Z}$, and their proof structure, in particular inductions, relies on their graph structure. The current inductions are more involved and more in the proof spirit of [112].

To begin with, let us recall the rates of the different events for the population N_w^K, with $w \in V$, at time t:

- Reproductions without mutation:

$$\mathfrak{b}_w(t) := b_w(1 - K^{-1/\alpha}) N_w^K(t) \tag{3.62}$$

- Death:

$$\mathfrak{d}_w(t) := \left(d_w + \sum_{x \in V} \frac{c_{w,x}}{K} N_x^K(t) \right) N_w^K(t) \tag{3.63}$$

- Reproductions with mutations towards the trait w:

$$\mathfrak{bm}_w(t) := K^{-1/\alpha} \sum_{x\in V, d(x,w)=1} b_x m(x,w) N_x^K(t). \tag{3.64}$$

Notice that for K large enough, as σ_{k-1}^K satisfies Assumption 3.21 and by definition of $\theta_{k,m,C}^K$, on the time interval $[\sigma_{k-1}^K \log K, (\theta_{k,m,C}^K \wedge T)\log K]$, we have

$$b(w,k,-)N_w^K(t) \le \mathfrak{b}_w(t) \le b(w,k,+)N_w^K(t), \tag{3.65}$$
$$d(w,\tilde{\mathbf{v}}_{k-1},k,+)N_w^K(t) \le \mathfrak{d}_w(t) \le d(w,\tilde{\mathbf{v}}_{k-1},k,-)N_w^K(t) \tag{3.66}$$
$$f(w,\tilde{\mathbf{v}}_{k-1},k,-) \le f_{w,\tilde{\mathbf{v}}_{k-1}} \le f(w,\tilde{\mathbf{v}}_{k-1},k,+), \tag{3.67}$$

where we have introduced the following notations, for any $w \in V$ and $* \in \{-,+\}$,

$$b(w,k,-) := (1-\varepsilon_k)b_w, \quad b(w,k,+) := b_w, \tag{3.68}$$
$$d(w,\tilde{\mathbf{v}}_{k-1},k,-) := d_w + \sum_{x\in\tilde{\mathbf{v}}_{k-1}} c_{w,x}\bar{n}_x(\mathbf{v}_{k-1}) + \left(\sum_{x\in V} c_{w,x}\right)(m+C)\varepsilon_k, \tag{3.69}$$
$$d(w,\tilde{\mathbf{v}}_{k-1},k,+) := d_w + \sum_{x\in\tilde{\mathbf{v}}_{k-1}} c_{w,x}\bar{n}_x(\mathbf{v}_{k-1}) - \left(\sum_{x\in V} c_{w,x}\right)C\varepsilon_k, \tag{3.70}$$
$$f(w,\tilde{\mathbf{v}}_{k-1},k,*) := b(w,k,*) - d(w,\tilde{\mathbf{v}}_{k-1},k,*). \tag{3.71}$$

Hence the rate of reproduction without mutation, as well as the death rate do not vary significantly during the time interval $[\sigma_{k-1}^K \log K, (\theta_{k,m,C}^K \wedge T)\log K]$. The difficulty comes from the rate of mutations towards a given trait, which depends on the population sizes of its neighbours in the graph $\mathcal{G}$, which themselves depend on the population sizes of their neighbours and so on.

Let us introduce the times τ_ℓ^K and the sets M_ℓ^K, which correspond respectively to the times of invasion or appearance of new mutants (and will be the time steps of the algorithm to be described shortly later) and to the sets of living traits in the time interval $(\tau_\ell^K, \tau_{\ell+1}^K]$. To be more precise,

Definition 3.22. Let $s_k^K := \inf\{t \ge \sigma_{k-1}^K : \exists w \in V \setminus \mathbf{v}_{k-1}^K, \beta_w^K(t) > 1-\varepsilon_k\}$, and

$$t_{w,k}^K := \begin{cases} \inf\{t \ge \sigma_{k-1}^K : \exists\, u \in V : d(u,w)=1, \beta_u^K(t) = \frac{1}{\alpha}\} & \text{if } \beta_w^K(\sigma_{k-1}^K) = 0 \\ \sigma_{k-1}^K & \text{else,} \end{cases} \tag{3.72}$$

The sequences $(\tau_\ell^K, \ell \ge 0)$ and $(M_\ell^K, \ell \ge 0)$ are defined as follows: $\tau_0^K = \sigma_0^K$ and, for $\sigma_{k-1}^K \le \tau_{\ell-1}^K < s_k^K$,

$$\tau_\ell^K = s_k^K \wedge \min\{t_{w,k}^K : w \in V, t_{w,k}^K > \tau_{\ell-1}^K\}, \tag{3.73}$$

that is to say the minimum between the time when a previously microscopic population becomes (almost) macroscopic, and the time of appearance of a new mutant. From the definition of the sequence $(\tau_\ell^K, \ell \geq 0)$ we can now define the sequence of sets of living traits $(M_\ell^K, \ell \geq 0)$ via

$$\begin{aligned} M_\ell^K &= \{w \in V : \beta_w^K(\tau_\ell^K) > 0 \text{ or } \tau_\ell^K = t_{w,k}^K\} \\ &= (M_{\ell-1}^K \backslash \{w \in V : \beta_w^K(\tau_\ell^K) = 0\}) \cup \{w \in V : \tau_\ell^K = t_{w,k}^K\}. \end{aligned} \tag{3.74}$$

3.4.2 Dynamics of the process on $[\tau_{\ell-1}^K \log K, \tau_\ell^K \log K]$

We will first prove that there exists a finite and positive constant C_ℓ such that with high probability, for every $w \in M_{\ell-1}^K$ and $t \in [\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$,

$$\begin{aligned} \max_{u \in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+ \\ \leq \beta_w^K(t) \leq \\ \max_{u \in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + C_\ell \varepsilon_k + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, +)\right]_+ \end{aligned} \tag{3.75}$$

Let us thus take t in $[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$. To obtain the lower bound in (3.75), we show by induction that, for any $n \geq 0$ and with high probability,

$$\beta_w^K(t) \geq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n}} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K) f(u, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+ \tag{3.76}$$

Induction lower bound: • $\underline{n = 0:}$ let $w \in M_{\ell-1}^K$. From (3.65) and (3.66), we see that we can couple N_w^K with a process Z^K with law $BP_K\left(b(w,k,-), d(w, \tilde{\mathbf{v}}_{k-1}, k, -), \beta_w^K(\tau_{\ell-1}^K)\right)$ (see the definition of BP_K in Subsection 3.5.1) in such a way that

$$N_w^K(t \log K) \geq Z^K((t - \tau_{\ell-1}^K) \log K). \tag{3.77}$$

Hence, from Corollary 3.27, we obtain that with high probability,

$$\beta_w^K(t) \geq \left[\beta_w^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K) f(w, \tilde{\mathbf{v}}_{k-1}, k, -)\right]_+. \tag{3.78}$$

Remark 3.23. Notice that the application of Lemma 3.24 (which has been derived in [38]) would require $\beta_w^K(\tau_{\ell-1}^K) > 0$ and that this condition may not be satisfied for one of the $w \in M_{\ell-1}^K$ (the trait which becomes macroscopic at time $\tau_{\ell-1}^K \log K$). However, the population of individuals w grows exponentially due to the mutations coming from another trait and there exists a finite c such that, for small $\delta > 0$, $N_w^K((\tau_{\ell-1}^K + \delta) \log K) \geq K^{c\delta}$. We could thus apply Lemma 3.24 at this time, and later on let δ go to 0 to get the result. This is in words the statement of Corollary 3.27.

• $\underline{n \to n+1}$: Let $w, u', u \in M_{\ell-1}^K$ such that $d(u', w) = 1$ and $d(u, u') \leq n$. From now on, we will use the notation BPI_K, which is defined in Subsection 3.5.2. From (3.65), (3.66), and (3.67), by looking only at the immigration coming from u', we see that we can couple N_w^K with a process Z^K with law

$$BPI_K\left(b(w,k,-), d(w,\tilde{\mathbf{v}}_{k-1},k,-), f(u',\tilde{\mathbf{v}}_{k-1},k,-), \beta_{u'}^K(\tau_{\ell-1}^K) - \frac{1}{\alpha}, \beta_w^K(\tau_{\ell-1}^K)\right) \tag{3.79}$$

in such a way that

$$N_w^K(t\log K) \geq Z^K((t-\tau_{\ell-1}^K)\log K). \tag{3.80}$$

By the induction hypothesis, with high probability,

$$\beta_{u'}^K(t) \geq \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')}{\alpha} + (t-\tau_{\ell-1}^K) f(u,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+, \tag{3.81}$$

which implies that we can couple Z^K with a process Y^K with law

$$BPI_K\left(b(w,k,-), d(w,\tilde{\mathbf{v}}_{k-1},k,-), f(u,\tilde{\mathbf{v}}_{k-1},k,-), \beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha}, \beta_w^K(\tau_{\ell-1}^K)\right) \tag{3.82}$$

in such a way that

$$Z^K((t-\tau_{\ell-1}^K)\log K) \geq Y^K((t-\tau_{\ell-1}^K)\log K). \tag{3.83}$$

Hence, from Corollary 3.27, even if we have to work in a time interval $[\tau_{\ell-1}^K + \delta, T]$, for a small positive δ, in the spirit of Remark 3.23, as $w \in M_{\ell-1}^K$ we obtain that with high probability,

$$\begin{aligned}
\beta_w^K(t) &\geq \left[\beta_w^K(\tau_{\ell-1}^K) \vee \left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha}\right) + (t-\tau_{\ell-1}^K) f(w,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+ \\
&\qquad \vee \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (t-\tau_{\ell-1}^K) f(u,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+ \\
&\geq \left[\beta_w^K(\tau_{\ell-1}^K) + (t-\tau_{\ell-1}^K) f(w,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+ \\
&\qquad \vee \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (t-\tau_{\ell-1}^K) f(u,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+ .
\end{aligned} \tag{3.84}$$

As this is true for any u' such that $d(u', w) = 1$ and as the above bound is a decreasing function of $d(u, u')$, by taking the supremum over such u' we obtain

$$\begin{aligned}
\beta_w^K(t) &\geq \left[\beta_w^K(\tau_{\ell-1}^K) + (t-\tau_{\ell-1}^K) f(w,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+ \\
&\qquad \vee \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t-\tau_{\ell-1}^K) f(u,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+ .
\end{aligned} \tag{3.85}$$

Thus, with high probability,

$$\beta_w^K(t) \geq \max_{\substack{u\in M_{\ell-1}^K:\\ d(u,w)\leq n+1}} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t-\tau_{\ell-1}^K) f(u,\tilde{\mathbf{v}}_{k-1},k,-)\right]_+, \tag{3.86}$$

which ends the induction for the lower bound.

Let us now proceed to the induction for the upper bound. We again take t in $[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$ and we will show that for any $n \in \mathbb{N}$ there exists a finite constant $C_{n,\ell}$ such that with high probability,

$$\beta_w^K(t) \leq \max_{\substack{u\in M_{\ell-1}^K:\\ d(u,w)\leq n}} \left[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + C_{n,\ell}\varepsilon_k\right) \vee \left(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k\right) + (t-\tau_{\ell-1}^K) f(u,\tilde{v}_{k-1},k,+)\right]_+ \vee \left(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k\right). \tag{3.87}$$

Notice that, since ε_k can be chosen small enough such that $1-(n+1)/\alpha + (n+2)\varepsilon_k < 0$, for all $n > \lfloor \alpha \rfloor$, this equation is equivalent to the upper bound in (3.75).

Induction upper bound:

Throughout the induction for the upper bound, we will several times make use of the fact that we can approximate the total immigration to one trait, which is the sum of the mutants coming from its neighbours, from above by the number of neighbours times the largest incoming mutation. More precisely, if I_w is the number of incoming neighbours of w,

$$\sum_{\substack{u\in V:\\ d(u,w)=1}} K^{\beta_u^K(t)} \leq I_w \max_{\substack{u\in V:\\ d(u,w)=1}} K^{\beta_u^K(t)} = \max_{\substack{u\in V:\\ d(u,w)=1}} K^{(\log I_w/\log K)+\beta_u^K(t)}. \tag{3.88}$$

Since the trait space is finite, we can assume that $\max_{w\in V} \log I_w/\log K \leq \varepsilon_k$, for K large enough.

• $\underline{n=0}$: We observe that for K large enough $\beta_u^K(t) \leq 1+\varepsilon_k$ for every $u \in M_{\ell-1}^K$ such that $d(u,w)=1$ (see Corollary 3.29).

From (3.65), (3.66), and (3.67), we see that we can couple N_w^K with a process Z^K with law $BPI_K\left(b(w,k,+), d(w,\tilde{\mathbf{v}}_{k-1},k,+), 0, 1-\frac{1}{\alpha}+2\varepsilon_k, \beta_w^K(\tau_{\ell-1}^K)\right)$ in such a way that

$$N_w^K(t\log K) \leq Z^K((t-\tau_{\ell-1}^K)\log K). \tag{3.89}$$

Hence from Corollary 3.27, even if we have to work in a time interval $[\tau_{\ell-1}^K+\delta, T]$, for a small positive δ, in the spirit of Remark 3.23, as $w \in M_{\ell-1}^K$ we obtain that with high probability,

$$\beta_w^K(t) \leq \left[\beta_w^K(\tau_{\ell-1}^K) \vee \left(1-\frac{1}{\alpha}+2\varepsilon_k\right) + (t-\tau_{\ell-1}^K) f(w,\tilde{v}_{k-1},k,+)\right]_+ \vee \left(1-\frac{1}{\alpha}+2\varepsilon_k\right). \tag{3.90}$$

• $\underline{n \to n+1}$: For $w, u' \in M_{\ell-1}^K$ such that $d(u', w) = 1$, by the induction hypothesis we have the existence of a finite constant $C_{n,\ell}$ such that, with high probability,

$$\beta_{u'}^K(t) \le \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,u') \le n}} \Bigg[\Big(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')}{\alpha} + C_{n,\ell}\varepsilon_k \Big) \vee \Big(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k \Big) + (t - \tau_{\ell-1}^K) f(u, \tilde{v}_{k-1}, k, +) \Bigg]_+ \vee \Big(1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k \Big). \quad (3.91)$$

From (3.65), (3.66), and (3.67), by looking at the maximal immigration coming from a neighbouring u' and adding another ε_k in the spirit of (3.88), we thus see that we can couple N_w^K with multiple processes $Z^{K,u,u'}$ and Z^K with respective laws

$$BPI_K\Big(b(w,k,+), d(w, \tilde{\mathbf{v}}_{k-1}, k, +), f(u, \tilde{v}_{k-1}, k, +), \Big(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \Big) \vee \Big(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \Big), \beta_w^K(\tau_{\ell-1}^K) \Big) \quad (3.92)$$

and

$$BPI_K\left(b(w,k,+), d(w, \tilde{\mathbf{v}}_{k-1}, k, +), 0, 1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k, \beta_w^K(\tau_{\ell-1}^K) \right) \quad (3.93)$$

in such a way that

$$N_w^K(t \log K) \le \max_{\substack{u' \in M_{\ell-1}^K: \\ d(u',w)=1}} \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,u') \le n}} Z^{K,u,u'}((t - \tau_{\ell-1}^K) \log K) \vee Z^K((t - \tau_{\ell-1}^K) \log K). \quad (3.94)$$

Hence from Corollary 3.27, even if we have to work in a time interval $[\tau_{\ell-1}^K + \delta, T]$, for a small positive δ, in the spirit of Remark 3.23, as $w \in M_{\ell-1}^K$ we obtain that with high probability,

$$\begin{aligned} \beta_w^K(t) \le \max_{\substack{u' \in M_{\ell-1}^K: \\ d(u',w)=1}} \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,u') \le n}} \Bigg\{ & \Bigg[\beta_w^K(\tau_{\ell-1}^K) \vee \Big(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \Big) \\ & \vee \Big(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \Big) + (t - \tau_{\ell-1}^K) f(w, \tilde{v}_{k-1}, k, +) \Bigg]_+ \\ & \vee \Bigg[\Big(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,u')+1}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \Big) \vee \Big(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \Big) \\ & + (t - \tau_{\ell-1}) f(u, \tilde{v}_{k-1}, k, +) \Bigg]_+ \vee \Big(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \Big) \Bigg\}. \end{aligned} \quad (3.95)$$

Applying that $d(u,w) \leq d(u,u') + 1$, this can be further approximated by

$$\beta_w^K(t) \leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n+1}} \Bigg\{ \Bigg[\beta_w^K(\tau_{\ell-1}^K) \vee \left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}^K) f(w, \tilde{v}_{k-1}, k, +) \Bigg]_+ \vee \Bigg[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}) f(u, \tilde{v}_{k-1}, k, +) \Bigg]_+ \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) \Bigg\}. \tag{3.96}$$

In order to simplify the right hand side of the previous inequality, we will show that for any $\ell \in \mathbb{N}$ there exists a finite and positive constant C_ℓ such that for any $(u,w) \in V^2$, with high probability

$$\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} \leq \beta_w^K(\tau_{\ell-1}^K) + C_\ell \varepsilon_k. \tag{3.97}$$

Combining (3.96) and (3.97) yields that with high probability,

$$\begin{aligned}
\beta_w^K(t) &\leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n+1}} \Bigg\{ \Bigg[\left(\beta_w^K(\tau_{\ell-1}^K) + (C_{n,\ell}+1+C_\ell)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}^K) f(w, \tilde{v}_{k-1}, k, +) \Bigg]_+ \\
&\qquad \vee \Bigg[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell}+1)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}) f(u, \tilde{v}_{k-1}, k, +) \Bigg]_+ \\
&\qquad \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) \Bigg\} \\
&\leq \max_{\substack{u \in M_{\ell-1}^K: \\ d(u,w) \leq n+1}} \Bigg[\left(\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (C_{n,\ell}+1+C_\ell)\varepsilon_k \right) \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right) + (t - \tau_{\ell-1}) f(u, \tilde{v}_{k-1}, k, +) \Bigg]_+ \vee \left(1 - \frac{n+2}{\alpha} + (n+3)\varepsilon_k \right),
\end{aligned} \tag{3.98}$$

which ends the induction for the upper bound.

Let us now derive inequality (3.97). It is obtained by an induction on ℓ. If $\ell = 1$, by (3.9) and the triangle inequality,

$$\begin{aligned}\lim_{K\to\infty} \beta_u^K(0) - \frac{d(u,w)}{\alpha} &= \left[1 - \frac{d(\mathbf{v}_0,u)}{\alpha}\right]_+ - \frac{d(u,w)}{\alpha} \\ &\leq \left[1 - \frac{d(\mathbf{v}_0,u)}{\alpha} - \frac{d(u,w)}{\alpha}\right]_+ \leq \left[1 - \frac{d(\mathbf{v}_0,w)}{\alpha}\right]_+ = \lim_{K\to\infty} \beta_w^K(0).\end{aligned} \tag{3.99}$$

As the convergence is in probability, it means that for K large enough, there exists a finite $C_{u,w}$ such that with a probability larger than $1 - \varepsilon_k$,

$$\beta_u^K(0) - \frac{d(u,w)}{\alpha} \leq \beta_w^K(0) + C_{u,w}\varepsilon_k. \tag{3.100}$$

As there are only finitely many traits, $\sup_{u,w\in V} C_{u,w} < \infty$. Moreover, as ε_k can be chosen as small as we want and as we want to prove a convergence in probability, we may focus on the event where inequality (3.100) is satisfied. We will do that later on without mentioning it again for the sake of readability.

Now assume that (3.97) is true for $\ell - 1 \in \mathbb{N}$. Let us first prove that it still holds for ℓ.

From the previous step on the time interval $[\tau_{\ell-2}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K]$, we know that if $\tau_{\ell-1}^K \leq T \wedge \theta_{k,m,C}^K$, for any $w \in V$ and K large enough,

$$\max_{u\in M_{\ell-2}^K} \left[\beta_u^K(\tau_{\ell-2}^K) - \frac{d(u,w)}{\alpha} + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(u, \tilde{\mathbf{v}}_{k-2}, k, -)\right]_+ \leq \beta_w^K(\tau_{\ell-1}^K). \tag{3.101}$$

Now let us take $u \in V$. We also deduce from the previous step that for K large enough

$$\beta_u^K(\tau_{\ell-1}^K) \leq \max_{u'\in M_{\ell-2}^K} \left[\beta_{u'}^K(\tau_{\ell-2}^K) - \frac{d(u',u)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(u', \tilde{\mathbf{v}}_{k-2}, k, +)\right]_+. \tag{3.102}$$

In particular there exists $\tilde{u} \in V$ such that $d(\tilde{u}, u) \leq \lfloor \alpha \rfloor$ and for K large enough

$$\beta_u^K(\tau_{\ell-1}^K) \leq \left[\beta_{\tilde{u}}^K(\tau_{\ell-2}^K) - \frac{d(\tilde{u},u)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(\tilde{u}, \tilde{\mathbf{v}}_{k-2}, k, +)\right]_+. \tag{3.103}$$

Thus, for K large enough,

$$\begin{aligned}
&\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} \\
&\quad\le \left[\beta_{\tilde u}^K(\tau_{\ell-2}^K) - \frac{d(\tilde u,u)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(\tilde u, \tilde{\mathbf{v}}_{k-2}, k, +)\right]_+ - \frac{d(u,w)}{\alpha} \\
&\quad\le \left[\beta_{\tilde u}^K(\tau_{\ell-2}^K) - \frac{d(\tilde u,u) + d(u,w)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(\tilde u, \tilde{\mathbf{v}}_{k-2}, k, +)\right]_+ \\
&\quad\le \left[\beta_{\tilde u}^K(\tau_{\ell-2}^K) - \frac{d(\tilde u,w)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(\tilde u, \tilde{\mathbf{v}}_{k-2}, k, +)\right]_+ \\
&\quad\le \left[\beta_{\tilde u}^K(\tau_{\ell-2}^K) - \frac{d(\tilde u,w)}{\alpha} + C_{\ell-1}\varepsilon_k + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(\tilde u, \tilde{\mathbf{v}}_{k-2}, k, -)\right]_+ + C\varepsilon_k \\
&\quad\le \max_{\tilde u \in M_{\ell-2}^K} \left[\beta_{\tilde u}^K(\tau_{\ell-2}^K) - \frac{d(\tilde u,w)}{\alpha} + (\tau_{\ell-1}^K - \tau_{\ell-2}^K) f(\tilde u, \tilde{\mathbf{v}}_{k-2}, k, -)\right]_+ + (C_{\ell-1} + C)\varepsilon_k \\
&\quad\le \beta_w^K(\tau_{\ell-1}^K) + (C_{\ell-1} + C)\varepsilon_k, \qquad (3.104)
\end{aligned}$$

where we used (3.71), (3.76), the bound $\tau_{\ell-1}^K - \tau_{\ell-2}^K \le T$, and

$$C := \max_{\tilde u \in M_{\ell-2}^K} \left(b_{\tilde u} + (m+6)\left(\sum_{x \in V} c_{\tilde u, x}\right)\right) T. \qquad (3.105)$$

This entails (3.97).

To conclude the proof of (3.75), we just need to notice that for $n > \lfloor \alpha \rfloor$, if ε_k is small enough,

$$1 - \frac{n+1}{\alpha} + (n+2)\varepsilon_k < 0. \qquad (3.106)$$

As $\tau_\ell^K - \tau_{\ell-1}^K \le T$, Equation (3.75) tells us that, with an error of order ε_k which is as small as we want, with high probability, the growth of traits $w \in M_{\ell-1}^K$ follows, for $t \in [\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K]$

$$\beta_w^K(t) \cong \max_{u \in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) - \frac{d(u,w)}{\alpha} + (t - \tau_{\ell-1}^K) f_{u, \mathbf{v}_{k-1}}\right]_+. \qquad (3.107)$$

To avoid repetition, we will write $\cong$ in the sequel to indicate approximations with high probability, with an error of order ε_k.

3.4.3 Value of τ_ℓ^K and construction of M_ℓ^K

Let us assume for the moment (it will be proven in Subsection 3.4.4) that the following holds with high probability:

$$[\tau_{\ell-1}^K \wedge T \wedge \theta_{k,m,C}^K, \tau_\ell^K \wedge T \wedge \theta_{k,m,C}^K] = [\tau_{\ell-1}^K, \tau_\ell^K]. \qquad (3.108)$$

Our aim now is to find the duration $\tau_\ell^K - \tau_{\ell-1}^K$ and to construct the set M_ℓ^K knowing the set $M_{\ell-1}^K$.

To reach τ_ℓ^K, two events are possible. Either one living non resident trait reaches a size of order K, or a new mutant appears. Let us consider the first type of event. In fact, we have to be more precise on the time when a new trait has a size which reaches order K, this is why we defined s_k^K as the time when one trait has a size which reaches order $K^{1-\varepsilon_k}$. Notice that we may choose ε_k small enough to be sure that it corresponds to the trait whose exponent reaches 1 at time s_k in the deterministic sequence $(s_j, j \in \mathbb{N})$ defined in Theorem 3.3. (if there exist two such traits, condition $(iv)(a)$ is fulfilled and T_0 is set to s_k). Notice that if $f_{u,\mathbf{v}_{k-1}} < 0$, for any $w \in V$,

$$t \mapsto \beta_u^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K) f_{u,\mathbf{v}_{k-1}} - \frac{d(u,w)}{\alpha} \tag{3.109}$$

is decreasing and thus will not reach $1-\varepsilon_k$ if it is smaller than this value at time $\tau_{\ell-1}^K$. Hence if we denote by u_0 the element of $M_{\ell-1}^K$ such that $\beta_{u_0}^K(\tau_\ell^K) = 1 - \varepsilon_k$, we get

$$1 - \varepsilon_k = \beta_{u_0}^K(\tau_\ell^K) \cong \max_{\substack{u \in M_{\ell-1}^K \\ f_{u,\mathbf{v}_{k-1}} > 0}} \left[\beta_u^K(\tau_{\ell-1}^K) + (t - \tau_{\ell-1}^K) f_{u,\mathbf{v}_{k-1}} - \frac{d(u,u_0)}{\alpha} \right]. \tag{3.110}$$

Now assume by contradiction that there is $u_1 \neq u_0 \in M_{\ell-1}^K$ such that:

$$1 - \varepsilon_k = \beta_{u_0}^K(\tau_\ell^K) \cong \beta_{u_1}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f_{u_1,\mathbf{v}_{k-1}} - \frac{d(u_1,u_0)}{\alpha}. \tag{3.111}$$

This implies

$$\beta_{u_1}^K(\tau_\ell^K) \geq \beta_{u_1}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f_{u_1,\mathbf{v}_{k-1}} > 1, \tag{3.112}$$

as soon as $\varepsilon_k < 1/\alpha$, which yields a contradiction. This implies that if there exists $u_0 \in M_{\ell-1}^K$ such that $\beta_{u_0}^K(\tau_\ell^K) = 1 - \varepsilon_k$, then

$$\beta_{u_0}^K(\tau_\ell^K) \cong \beta_{u_0}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f_{u_0,\mathbf{v}_{k-1}} \tag{3.113}$$

and with high probability, the value of $\tau_\ell^K - \tau_{\ell-1}^K$ satisfies,

$$\tau_\ell^K - \tau_{\ell-1}^K \cong \min_{\substack{w \in M_{\ell-1}^K: \\ f_{w,\mathbf{v}_{k-1}} > 0}} \frac{1 - \beta_w^K(\tau_{\ell-1}^K)}{f_{w,\mathbf{v}_{k-1}}}. \tag{3.114}$$

Let us now consider the second type of event, that is to say that there exist $u_0 \notin M_{\ell-1}^K$ and $u_1 \in M_{\ell-1}^K$ such that $d(u_1,u_0) = 1$ and $\beta_{u_1}^K(\tau_\ell^K) = 1/\alpha$. Notice again that if $f_{u,\mathbf{v}_{k-1}} < 0$, the function defined in (3.109) is decreasing and thus will not reach $1/\alpha$ if it is smaller than this value at time $\tau_{\ell-1}^K$.

By definition we have

$$\frac{1}{\alpha} = \beta_{u_1}^K(\tau_\ell^K) \cong \max_{u \in M_{\ell-1}^K} \left[\beta_u^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f_{u,\mathbf{v}_{k-1}} - \frac{d(u,u_1)}{\alpha} \right]. \tag{3.115}$$

Denote by $u_2 \in M_{\ell-1}^K$ the trait realizing the maximum in the previous equation, that is to say

$$\frac{1}{\alpha} \cong \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f_{u_2,\mathbf{v}_{k-1}} - \frac{d(u_2,u_1)}{\alpha}. \tag{3.116}$$

This equality can be rewritten as

$$\beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f_{u_2,\mathbf{v}_{k-1}} \cong \frac{d(u_2,u_1)+1}{\alpha}. \tag{3.117}$$

Let us now make a *reductio ad absurdum* to prove that $d(u_2,u_1)+1 = d(u_2,u_0)$. Let us thus assume that

$$d(u_2,u_1)+1 > d(u_2,u_0) \Leftrightarrow d(u_2,u_1) \geq d(u_2,u_0), \tag{3.118}$$

and take u_1' such that

$$d(u_2,u_1')+1 = d(u_2,u_0). \tag{3.119}$$

Let us first assume (we will prove it later) that $u_1' \in M_{\ell-1}^K$. In this case, using the proof for the lower bound, we obtain that with high probability

$$\begin{aligned}\beta_{u_1'}^K(\tau_\ell^K) &\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f(u_2,\tilde{\mathbf{v}}_{k-1},k,-) - \frac{d(u_2,u_1')}{\alpha} \\ &\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f(u_2,\tilde{\mathbf{v}}_{k-1},k,-) - \frac{d(u_2,u_1)-1}{\alpha} \cong \frac{2}{\alpha}.\end{aligned} \tag{3.120}$$

As $d(u_1',u_0)=1$, this means that u_0 becomes a living trait before the time τ_ℓ^K, which is in contradiction with the definition of τ_ℓ^K.

Let us now assume that $u_1' \notin M_{l-1}^K$ and consider a sequence of vertices $v_0 = u_2, v_1, ...,$ $v_{d(u_2,u_1')} = u_1'$ such that $d(u_2,v_k) = k$ and $d(v_k,u_1') = d(u_2,u_1') - k$. Let

$$k_0 := \max\{0 \leq k \leq d(u_2,u_1') - 1, v_k \in M_{\ell-1}^K\}. \tag{3.121}$$

Then

$$d(u_2,v_{k_0}) \leq d(u_2,u_1') - 1 \leq d(u_2,u_1) - 2, \tag{3.122}$$

and with high probability

$$\begin{aligned}\beta_{v_{k_0}}^K(\tau_\ell^K) &\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f(u_2,\tilde{\mathbf{v}}_{k-1},k,-) - \frac{d(u_2,v_{k_0})}{\alpha} \\ &\geq \beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f(u_2,\tilde{\mathbf{v}}_{k-1},k,-) - \frac{d(u_2,u_1)-2}{\alpha} \cong \frac{3}{\alpha},\end{aligned} \tag{3.123}$$

and thus v_{k_0+1} becomes a living trait before the time τ_ℓ^K, which again is in contradiction with the definition of τ_ℓ^K. We thus obtain a contradiction and deduce that (3.118) is not satisfied. We conclude that

$$\beta_{u_2}^K(\tau_{\ell-1}^K) + (\tau_\ell^K - \tau_{\ell-1}^K) f_{u_2,\mathbf{v}_{k-1}} \cong \frac{d(u_2,u_1)+1}{\alpha} = \frac{d(u_2,u_0)}{\alpha}. \tag{3.124}$$

Hence, when τ_ℓ^K corresponds to the arrival of a new mutant,

$$\tau_\ell^K - \tau_{\ell-1}^K \cong \min_{\substack{w\in M_{\ell-1}^K \\ f_{w,\mathbf{v}_{\ell-1}}>0}} \frac{\frac{d(w,V\setminus M_{\ell-1}^K)}{\alpha} - \beta_w^K(\tau_{\ell-1}^K)}{f_{w,\mathbf{v}_{k-1}}}. \tag{3.125}$$

Combining (3.114) and (3.125), we finally obtain

$$\tau_\ell^K - \tau_{\ell-1}^K \cong \min_{\substack{w\in M_{\ell-1}^K: \\ f_{w,\mathbf{v}_{k-1}}>0}} \frac{\left(1\wedge\frac{d(w,V\setminus M_{\ell-1}^K)}{\alpha}\right) - \beta_w^K(\tau_{\ell-1}^K)}{f_{w,\mathbf{v}_{k-1}}}. \tag{3.126}$$

To obtain M_ℓ^K from $M_{\ell-1}^K$, we suppress the traits $w \in M_{\ell-1}^K$ such that $\beta_w^K(\tau_\ell^K) = 0$ (if condition $(iv)(c)$ is not satisfied, otherwise T_0 is set to s_k) and if $\tau_\ell \neq s_k$, we add the traits which are at distance 1 from the $w \in V$ satisfying

$$w \in \operatorname*{arg\,min}_{\substack{w\in M_{\ell-1}^K: \\ f_{w,\mathbf{v}_{k-1}}>0}} \frac{\left(1\wedge\frac{d(w,V\setminus M_{\ell-1}^K)}{\alpha}\right) - \beta_w^K(\tau_{\ell-1}^K)}{f_{w,\mathbf{v}_{k-1}}}. \tag{3.127}$$

3.4.4 Value of $\theta_{k,m,C}^K$ and convergence of s_k^K to s_k

Recall the definition of $\theta_{k,m,C}^K$ in (3.61). We thus have constructed, on the time interval $[(\sigma_{k-1}^K\wedge T)\log K, (s_k^K\wedge\theta_{k,m,C}^K\wedge T)\log K]$, the times $(\tau_\ell^K, \ell\in\mathbb{N})$ and the sets $(M_\ell^K, \ell\in\mathbb{N})$ of living traits between times τ_ℓ^K and $\tau_{\ell+1}^K$. We will now study the dynamics of the process on the time interval $[(\sigma_{k-1}^K\wedge T)\log K, (\sigma_k^K\wedge\theta_{k,m,C}^K\wedge T)\log K]$ (σ_k^K to be defined later in order to satisfy Assumption 3.21). Recall that l_k^K is the trait $w\in V$ such that $\beta_w^K(s_k^K) = 1-\varepsilon_k$ and introduce

$$\eta_k := 2\varepsilon_k / \left(f_{l_k^K,\mathbf{v}_{k-1}} - \left(b_{l_k^K} + \left(\sum_{x\in V} c_{l_k^K,x}\right)(C+m)\right)\varepsilon_k\right) \tag{3.128}$$

We will first prove that

$$\lim_{K\to\infty} \mathbb{P}\left(s_k^K \le \theta_{k,m,C}^K \le s_k^K + \eta_k \middle| s_k^K < T\right) = 1. \tag{3.129}$$

The first step consists in showing that

$$\lim_{K\to\infty} \mathbb{P}\left(\theta_{k,m,C}^K < s_k^K \middle| s_k^K < T\right) = 0. \tag{3.130}$$

By definition of s_k^K, we have

$$\sup_{w\in V\setminus\mathbf{v}_{k-1}} \sup_{\sigma_{k-1}^K\le t\le s_k^K} \beta_w^K(t) \le 1-\varepsilon_k. \tag{3.131}$$

Moreover, applying Lemma 3.28 to $\mathbf{v}_{k-1}$ we obtain that

$$\lim_{K\to\infty} \mathbb{P}\left(\forall t \in [\sigma_{k-1}^K, s_k^K], \sup_{w\in\mathbf{v}_{k-1}} \left|\frac{N_w^K(t\log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1})\right| \le C\varepsilon_k \middle| s_k^K < T\right) = 1. \quad (3.132)$$

As a consequence, (3.130) holds true. Notice that the value of C in the definition of $\theta_{k,m,C}^K$ in (3.61) is a consequence of the previous limit. The constant C is the one needed for Lemma 3.28 to hold, and thus depends on the parameters of the process.

Now assume by contradiction that

$$s_k^K + \eta_k \le \theta_{k,m,C}^K < T. \quad (3.133)$$

Then on the time interval $[s_k^K, s_k^K + \eta_k]$, by definition of $\theta_{k,m,C}^K$, the l_k^K population has a growth rate bounded from below by

$$f_{l_k^K,\mathbf{v}_{k-1}} - \left(b_{l_k^K} + \left(\sum_{x\in V} c_{l_k^K,x}\right)(C+m)\right)\varepsilon_k. \quad (3.134)$$

Hence by coupling, with high probability,

$$\beta_{l_k^K}^K(s_k^K + \eta_k) \ge 1 - \varepsilon_k + \left(f_{l_k^K,\mathbf{v}_{k-1}} - \left(b_{l_k^K} + \left(\sum_{x\in V} c_{l_k^K,x}\right)(C+m)\right)\varepsilon_k\right)\eta_k = 1 + \varepsilon_k, \quad (3.135)$$

which leads to a contradiction, as the total population size cannot be of order larger than K in the limit $K \to \infty$, see Corollary 3.29.

This proves (3.129). In particular, this implies that s_k^K converges to s_k in probability when K goes to infinity, as soon as $T > s_k$.

3.4.5 Value of the process at time $\theta_{k,m,C}^K \log K$

We are now interested in the value of the process at time $\theta_{k,m,C}^K \log K$. First notice that according to Proposition A.2 in [37] and (3.129),

$$\begin{aligned}&\lim_{K\to\infty} \mathbb{P}\left(\forall t \in [\sigma_{k-1}^K, (s_k^K + \eta_k) \wedge \theta_{k,m,C}^K], w \in \mathbf{v}_{k-1}, \left|\frac{N_w^K(t\log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1})\right| < C\varepsilon_k\right)\\ &= \lim_{K\to\infty} \mathbb{P}\left(\forall t \in [\sigma_{k-1}^K, \theta_{k,m,C}^K], w \in \mathbf{v}_{k-1}, \left|\frac{N_w^K(t\log K)}{K} - \bar{n}_w(\mathbf{v}_{k-1})\right| < C\varepsilon_k\right) = 1. \quad (3.136)\end{aligned}$$

Notice that m has to be chosen small enough for this limit to hold, and thus depends on the parameters of the Lotka-Volterra deterministic system associated to $\mathbf{v}_{k-1}$. To be more precise, m has to be chosen small enough for the assumption (3.148) in Lemma 3.28 to hold true with ε_k in place of ε. We choose such an m in the definition of $\theta_{k,m,C}^K$ in (3.61).

Hence we obtain that with high probability,

$$\sum_{w\in V\setminus \mathbf{v}_{k-1}} N_w^K(\theta_{k,m,C}^K \log K) \geq m\varepsilon_k K. \tag{3.137}$$

If condition $(iv)(a)$ of Theorem 3.3 is satisfied T_0 is set at s_k and the induction is stopped. Otherwise there exists $\gamma > 0$ such that if ε_k is small enough $\beta_w^K(s_k^K) < 1-\gamma$ for every $w \in V \setminus (\tilde{\mathbf{v}}_{k-1}^K \cup \{l_k^K\})$. Thus again by coupling, as the growth rates of the populations are limited and η_k may be as small as we want, with high probability,

$$\sum_{w\in V\setminus \tilde{\mathbf{v}}_{k-1}^K, w\neq l_k^K} N_w^K(\theta_{k,m,C}^K \log K) \leq K^{1-\gamma/2}. \tag{3.138}$$

From the two last inequalities we deduce that with high probability,

$$N_{l_k^K}^K(\theta_{k,m,C}^K \log K) \geq m\varepsilon_k K/2. \tag{3.139}$$

3.4.6 Construction of σ_k^K and Assumption 3.21

Let us now introduce the stopping time σ_k^K, via:

$$\sigma_k^K := \inf\{t \geq \theta_{k,m,C}^K, \forall w \in \mathbf{v}_k^K, |N_w^K(t\log K)/K - \bar{n}_w(\mathbf{v}_k)| \leq \varepsilon_k\}. \tag{3.140}$$

The last step of the proof consists in showing that σ_k^K indeed satisfies Assumption 3.21. First $\sigma_k^K \log K$ is a stopping time. Second, from (3.136), (3.138), (3.139) and an application of Lemma 3.28 there exists $T(\varepsilon_k) < \infty$ such that

$$\lim_{K\to\infty} \mathbb{P}\left(\left|N_w^K(\theta_{k,m,C}^K \log K + T(\varepsilon_k))/K - \bar{n}_w(\mathbf{v}_k)\right| \leq \varepsilon_k, \forall w \in \mathbf{v}_k\right) = 1. \tag{3.141}$$

Moreover, during a time of order one, the order of population sizes does not vary more than a constant times ε_k (result similar in spirit to Lemma B.9 in [38]). Adding that s_k^K converges to s_k in probability when K goes to infinity, as well as (3.129), we obtain that Assumption 3.21 holds. It ends the proof of Theorem 3.3 and Proposition 3.6.

3.5 Appendix: Couplings with branching processes and logistic processes with immigration

The aim of this section is to collect various couplings of the populations with simpler processes like branching processes and logistic processes with immigration, and to state some properties of these simpler processes. These results have been derived in [38] (note that we need to slightly generalise some of them), and we state them for the sake of readability. For simplicity we keep the notations of [38].

3.5.1 Branching process

In this subsection, we recall Lemma A.1 of [38], which describes the dynamics of a birth-and-death process on a $\log K$ time scale. For $b, d, \beta \geq 0$, let $BP_K(b, d, \beta)$ denote the law of a process $(Z^K(t), t \geq 0)$ with initial state $Z^K(0) = \lfloor K^\beta - 1 \rfloor$, individual birth rate b and individual death rate d.

Lemma 3.24 (Lemma A.1 in [38]). *Let $(Z^K(t), t \geq 0)$ be a $BP_K(b, d, \beta)$ process such that $\beta > 0$. The process $(\log(1 + Z^K(t \log K))/\log K, t > 0)$ converges when K tends to infinity in probability in $L^\infty([0, T])$ for all $T > 0$ to the continuous deterministic function given by*

$$\bar{\beta} : t \mapsto \beta + (b - d)t \vee 0. \tag{3.142}$$

In addition, if $b < d$, for all $t > \beta/(d - b)$,

$$\lim_{K \to \infty} \mathbb{P}\left(Z^K_{t \log K} = 0\right) = 1. \tag{3.143}$$

3.5.2 Branching process with immigration

In this subsection, we recall Lemma B.4 and Theorem B.5 of [38], illustrated in Figure B.1 therein, which describe the dynamics of birth-and-death processes with immigration on a $\log K$ time scale. For $b, d, \beta \geq 0$, $a, c \in \mathbb{R}$, $BPI_K(b, d, a, c, \beta)$ denotes the law of a process $(Z^K(t), t \geq 0)$ with initial state $Z^K(0) = \lfloor K^\beta - 1 \rfloor$, individual birth rate b, individual death rate d, and immigration rate $K^c \mathrm{e}^{as}$ at time $s \geq 0$.

Lemma 3.25 (Lemma B.4 in [38]). *Assume that $\beta < c$. Then for all $\varepsilon > 0$ and all $\bar{a} > |b - d| \vee |a|$,*

$$\lim_{K \to \infty} \mathbb{P}\left(Z^K(\varepsilon \log K) \in \left[K^{c - \bar{a}\varepsilon}, K^{c + \bar{a}\varepsilon}\right]\right) = 1. \tag{3.144}$$

Lemma 3.26 (Theorem B.5 in [38]). *Let $(Z^K(t), t \geq 0)$ be a $BPI_K(b, d, a, c, \beta)$ process with $c \leq \beta$ and assume that $\beta > 0$. The process $(\log(1 + Z^K(t \log K))/\log K, t > 0)$ converges when K tends to infinity in probability in $L^\infty([0, T])$ for all $T > 0$ to the continuous deterministic function $\bar{\beta}$ given by*

$$\bar{\beta} : t \mapsto (\beta + (b - d)t) \vee (c + at) \vee 0. \tag{3.145}$$

In addition, in the case where $c \neq 0$ or $a \neq 0$, for all compact intervals $I \subset \mathbb{R}_+$ which do not intersect the support of $\bar{\beta}$,

$$\lim_{K \to \infty} \mathbb{P}\left(Z^K(t \log K) = 0, \forall t \in I\right) = 1. \tag{3.146}$$

We will mostly use a corollary of those two lemmas, which is valid without the assumption $c \leq \beta$ but on a time interval $[\delta, T]$, for any $\delta > 0$. The idea of the proof has been explained in Remark 3.23.

Corollary 3.27. *Let $(Z^K(t), t \geq 0)$ be a $BPI_K(b, d, a, c, \beta)$ process with $\beta \geq 0$, and either $c > 0$ or both $c = 0$ and $a > 0$. For any $\delta > 0$ and $T > 0$, the process $(\log(1 + Z^K(t \log K))/\log K, t \in [\delta, T])$ converges when K tends to infinity in probability in $L^\infty([\delta, T])$ to the continuous deterministic function $\bar{\beta}$ given by*

$$\bar{\beta} : t \mapsto ((\beta \vee c) + (b - d)t) \vee (c + at) \vee 0. \tag{3.147}$$

3.5.3 Logistic birth-and-death process with immigration

We recall that for a subset $\mathbf{v} \subset V$ of traits that can coexist at a strictly positive equilibrium in the Lotka-Volterra system (3.5), $\bar{n}(\mathbf{v}) \in \mathbb{R}_+^{\mathbf{v}}$ denotes this equilibrium. The next result states that if all traits in $\mathbf{v}$ have an initial population of order K and the immigration of individuals with traits in $\mathbf{v}$ is small enough, the equilibrium $\bar{n}(\mathbf{v})K$ is reached in a time of order 1 and the populations of individuals whose traits belong to $\mathbf{v}$ will keep a size close to its equilibrium during a time of order larger than $\log K$. This result is a generalisation of Lemma C.1 in [38] to the multidimensional case and with (slightly) varying rates.

We consider a subset $\mathbf{v} \subset V$ of traits and denote by $(\mathbf{b}_\mathbf{v}(t), t \geq 0) := ((b_w(t), w \in \mathbf{v}), t \geq 0)$, $(\mathbf{d}_\mathbf{v}(t), t \geq 0) := ((d_w(t), w \in \mathbf{v}), t \geq 0)$, $(\mathbf{c}_\mathbf{v}(t), t \geq 0) := ((c_{w_1,w_2}(t), (w_1, w_2) \in \mathbf{v}^2), t \geq 0)$ its birth, natural death, and death by competition rates that we allow to vary in time, as well as $(\mathbf{g}_\mathbf{v}(t), t \geq 0) := ((g_w(t), w \in \mathbf{v}), t \geq 0)$ a function with values in $\mathbb{R}_+^{\mathbf{v}}$. We denote by $LBDI_K(\mathbf{b}_\mathbf{v}, \mathbf{d}_\mathbf{v}, \mathbf{c}_\mathbf{v}, \mathbf{g}_\mathbf{v})$ the law of a logistic birth-and-death process with immigration $\mathbf{Z}^K := ((Z_w(t)^K, w \in \mathbf{v}), t \geq 0)$ where, at time t, an individual with a trait $w \in \mathbf{v}$ has a birth rate $b_w(t)$, a death rate $d_w(t) + \sum_{x \in \mathbf{v}} c_{w,x}(t) Z_x^K(t)/K$ and an immigration rate $g_w(t)$.

Lemma 3.28. *Let $T > 0$, $\boldsymbol{v} \subset V$ and assume that the mutation-free Lotka-Volterra system (3.5) associated to $\boldsymbol{v}$ and with rates $(\bar{\mathbf{b}}_{\boldsymbol{v}}, \bar{\mathbf{d}}_{\boldsymbol{v}}, \bar{\mathbf{c}}_{\boldsymbol{v}}) \in (\mathbb{R}_{>0})^{\boldsymbol{v}} \times (\mathbb{R}_{>0})^{\boldsymbol{v}} \times (\mathbb{R}_{>0})^{\boldsymbol{v}^2}$ admits a unique positive globally attractive stable equilibrium $\bar{n}(\boldsymbol{v})$. Assume that Z^K follows the law $LBDI_K(\mathbf{b}_{\boldsymbol{v}}, \mathbf{d}_{\boldsymbol{v}}, \mathbf{c}_{\boldsymbol{v}}, \mathbf{g}_{\boldsymbol{v}})$ and that*

$$\sup_{w_1, w_2 \in \boldsymbol{v}} \left\{ \left| \mathbf{b}_{w_1}(t) - \bar{\mathbf{b}}_{w_1} \right|, \left| \mathbf{d}_{w_1}(t) - \bar{\mathbf{d}}_{w_1} \right|, |\mathbf{c}_{w_1,w_2}(t) - \bar{\mathbf{c}}_{w_1,w_2}| \right\} < \varepsilon \tag{3.148}$$

and $g_w(t) \leq K^{1-\eta}$ for all $t \in [0, T \log K]$, $w \in \boldsymbol{v}$ for some $\varepsilon, \eta > 0$.

(i) There exists $C, \varepsilon_0 > 0$ such that if $\varepsilon \leq \varepsilon_0$ and $\|\mathbf{Z}^K(0)/K - \bar{n}(\boldsymbol{v})\|_\infty \leq \varepsilon$, then

$$\lim_{K \to \infty} \mathbb{P}\left(\forall t \in [0, T \log K], \|\mathbf{Z}^K(t)/K - \bar{n}(\boldsymbol{v})\|_\infty \leq C\varepsilon \right) = 1. \tag{3.149}$$

(ii) For all $\varepsilon_1, \varepsilon_2 > 0$, there exists $\varepsilon_0 > 0$ and $T(\varepsilon_1, \varepsilon_2) < \infty$ such that for all $\varepsilon < \varepsilon_0$ and initial condition $\mathbf{Z}^K(0)$ such that $Z_v^K(0)/K \geq \varepsilon_1$, $\forall v \in \boldsymbol{v}$, we have that

$$\lim_{K \to \infty} \mathbb{P}\left(\|\mathbf{Z}^K(T(\varepsilon_1, \varepsilon_2))/K - \bar{n}(\boldsymbol{v})\|_\infty \leq \varepsilon_2 \right) = 1. \tag{3.150}$$

Proof. The case where the functions $\mathbf{b}_\mathbf{v}$, $\mathbf{d}_\mathbf{v}$, $\mathbf{c}_\mathbf{v}$ are constant is a direct generalisation of Lemma C.1 in [38], whose proof follows arguments similar to the ones given in [31, 37] or in the Proposition 4.2 in [34] to handle the addition of (negligible) immigration. We do not

provide it. Let us explain how we deal with varying rates for point (i). Let us choose $w_0 \in \mathbf{v}$, and introduce for $w_1, w_2 \in \mathbf{v}$:

$$\tilde{b}_{w_1} = \begin{cases} \bar{b}_{w_1} + \varepsilon & \text{if } w_1 \neq w_0 \\ \bar{b}_{w_1} - \varepsilon & \text{if } w_1 = w_0 \end{cases} \tag{3.151}$$

$$\tilde{d}_{w_1} = \begin{cases} \bar{d}_{w_1} - \varepsilon & \text{if } w_1 \neq w_0 \\ \bar{d}_{w_1} + \varepsilon & \text{if } w_1 = w_0 \end{cases} \tag{3.152}$$

$$\tilde{c}_{w_1,w_2} = \begin{cases} \bar{c}_{w_1,w_2} - \varepsilon & \text{if } w_1 \neq w_0 \\ \bar{c}_{w_1,w_2} + \varepsilon & \text{if } w_1 = w_0 \end{cases} \tag{3.153}$$

Then we can couple a process Z^K with the law $LBDI_K(\mathbf{b}_\mathbf{v}, \mathbf{d}_\mathbf{v}, \mathbf{c}_\mathbf{v}, \mathbf{g}_\mathbf{v})$ with a process $\tilde{Z}^K$ with the law $LBDI_K(\tilde{\mathbf{b}}_\mathbf{v}, \tilde{\mathbf{d}}_\mathbf{v}, \tilde{\mathbf{c}}_\mathbf{v}, \mathbf{g}_\mathbf{v})$ such that for every $t \geq 0$, $\tilde{Z}^K_{w_0}(t) \leq Z^K_{w_0}(t)$ and $\tilde{Z}^K_w(t) \geq Z^K_w(t)$ for every $w \in \mathbf{v} \setminus w_0$. Moreover, as the equilibrium of a Lotka-Volterra system is continuous with respect to its coefficient, there is a positive $\tilde{C}$ such that for ε small enough, and if we denote by $\bar{n}^{(w_0)}(\mathbf{v})$ the equilibrium of the Lotka-Volterra system with the coefficients $\tilde{\mathbf{b}}_\mathbf{v}, \tilde{\mathbf{d}}_\mathbf{v}, \tilde{\mathbf{c}}_\mathbf{v}$ we have just introduced, $\|\bar{n}^{(w_0)}(\mathbf{v}) - \bar{n}(\mathbf{v})\| \leq \tilde{C}\varepsilon$. Hence applying the point (i) for the process $\tilde{Z}^K$, we obtain upper bounds for coordinates $w \neq w_0$ and a lower bound for the coordinate w_0, for the process Z^K. Doing the same and the reverse bounds for the other elements of $\mathbf{v}$ gives the result for some $C > \tilde{C}$ that takes into account the fluctuations around the varied equilibria. □

We end this section with a result stating that the time needed for the total population size of a logistic birth-and-death process (with or without mutations) to reach an order smaller or equal to K is of order one for K large enough.

Corollary 3.29. *Let us consider a subset $\boldsymbol{v} \subset V$ of traits, $(\mathbf{b}_\boldsymbol{v}, \mathbf{d}_\boldsymbol{v}, \mathbf{c}_\boldsymbol{v})$ in $(\mathbb{R}_{>0})^{\boldsymbol{v}} \times (\mathbb{R}_{>0})^{\boldsymbol{v}} \times (\mathbb{R}_{>0})^{\boldsymbol{v}^2}$ and let Z^K follow the law $LBDI_K(\mathbf{b}_\boldsymbol{v}, \mathbf{d}_\boldsymbol{v}, \mathbf{c}_\boldsymbol{v}, 0)$, and $\mathcal{Z}^K$ denote the total population size of the process Z^K. For every $\varepsilon > 0$ there exists $T(\varepsilon) < \infty$ such that for $t > T(\varepsilon)$*

$$\lim_{K\to\infty} \mathbb{P}\left(\frac{\log(1 + \mathcal{Z}^K(t))}{\log K} < 1 + \varepsilon\right) = 1. \tag{3.154}$$

Remark 3.30. Notice that this result only treats mutation-free logistic birth-and-death processes. However, mutation within $\mathbf{v}$ does not affect the total population size and hence the result can be transferred to such cases. Considering $\mathbf{v} = V$, Corollary 3.29 therefore implies the same asymptotic bound for the total population size of the process that we consider in Theorem 3.3 and Proposition 3.6, and hence also for each subpopulation.

Proof. The process $\mathcal{Z}^K$ increases by 1 at a rate

$$\sum_{w\in\mathbf{v}} b_w Z^K_w \leq (\sup_{w\in\mathbf{v}} b_w)\mathcal{Z}^K =: \mathcal{B}\mathcal{Z}^K \tag{3.155}$$

and decreases by 1 at a rate

$$\sum_{w\in\mathbf{v}}\left(d_w+\sum_{u\in\mathbf{v}}\frac{c_{w,u}}{K}Z_u^K\right)Z_w^K \geq \frac{1}{K}(\inf_{u\in\mathbf{v}}c_{u,u})\sum_{w\in\mathbf{v}}\left(Z_w^K\right)^2$$
$$\geq \frac{1}{K}(\inf_{u\in\mathbf{v}}c_{u,u})\frac{1}{Card(\mathbf{v})}\left(\mathcal{Z}^K\right)^2 =: \frac{\mathcal{C}}{K}\left(\mathcal{Z}^K\right)^2. \qquad (3.156)$$

Hence the process $\mathcal{Z}^K$ can be coupled with a logistic birth-and-death process $\mathcal{N}^K$ with individual birth rate $\mathcal{B}$ and individual death rate $\mathcal{C}\mathcal{N}^K/K$ in such a way that for every $t\geq 0$, if $\mathcal{Z}^K(0)=\mathcal{N}^K(0)$

$$\mathcal{Z}^K(t)\leq\mathcal{N}^K(t). \qquad (3.157)$$

But from Chapter 11, Theorem 2.1 in [70], we know that on any finite time interval, the rescaled process $\mathcal{N}^K/K$ converges in probability to the solution to the logistic equation $\dot{\varkappa}=\varkappa(\mathcal{B}-\mathcal{C}\varkappa)$, $\varkappa(0)=\varkappa_0$ if $\mathcal{N}^K(0)/K$ converges in probability to $\varkappa_0$. The one dimensional logistic equation has an explicit solution, and in particular, we know that its equilibrium is $\mathcal{B}/\mathcal{C}$, that it comes down from infinity, and that it takes a time

$$\frac{1}{\mathcal{B}}\log\left(\frac{\bar{\varkappa}}{\bar{\varkappa}-\mathcal{B}/\mathcal{C}}\right) \qquad (3.158)$$

to reach $\bar{\varkappa}>\mathcal{B}/\mathcal{C}$ from an infinite initial condition. As a consequence, $\mathcal{N}^K$ takes a time of order one to become smaller than $2\bar{\varkappa}K$, and as $\mathcal{B}/\mathcal{C}$ is a globally hyperbolic equilibrium for the function $\varkappa$, classical large deviation results (see [63] for instance) entail that $\mathcal{N}^K/K$ will stay an exponential (in K) time in any compact interval of $R_{>0}$ including $\mathcal{B}/\mathcal{C}$. This concludes the proof. □

4 A stochastic model for melanoma T-cell therapy

In this chapter we present an individual-based Markov process that models immunotherapy of melanoma. It is an extension of the stochastic model for population dynamics that was discussed in Chapters 2 and 3 and allows to analyse evolutionary dynamics within the tumour as well as the interaction with cells and molecules of the immune system that arise during therapy. In particular, we study Adoptive Cell Transfer (ACT) therapy, where cytotoxic T-cells are injected to infiltrate the tumour tissue and specifically target the melanoma cells, which are characterised by a distinct surface marker (antigen). The tumour cells show various mechanisms to avoid detection by the immune cells and thus escape from therapy.

The current model is based on the model of Baar et al. in [11]. In a previous project, they have studied the escape from ACT therapy through phenotypic plasticity, where the melanoma cells can reversibly down-regulate the presentation of their specific antigen in an inflammatory environment. In the new experimental and mathematical models, we compare this adaptive phenotype switching to a permanent genetic antigen loss. The fitness of these genetic variants is highly variable and dependent on their environment and thus creates an interesting evolutionary scenario. We provide a framework to better understand how subclone heterogeneity in tumours and evolutionary dynamics influence immune selection by T-cell therapy through stochastic events. This work emphasises the need to take these dynamics into account when interpreting variant allele frequencies in therapy resistant tumour specimens.

In the present thesis we focus more on the mathematical modelling and demonstrate how theoretical approaches and simulations can benefit experimental research through the characterisation of important mechanisms, identification of likely causes for observed phenomena, and validation of clinical relevance. In Section 4.1, we give a brief summary of the biological background and the experimental results that are the basis of our model. In Section4.2, we present the extended stochastic model and explain the hybrid algorithm and parameter choices that we apply to run simulations. The results of the simulations are discussed extensively in Section 4.3. Finally, in Section 4.4, we consider the clinical relevance and the implications for future treatment strategies of our studies.

Larger parts of this chapter were previously made available as a preprint [88]. For the purpose of this thesis, the content is adapted to a mostly mathematical audience. Besides rearranging some sections, the paragraphs on the experimental methods and results are rewritten in a condensed way. For a more detailed description we refer to [88]. The paragraph on the deterministic approximation and the hybrid algorithm was extended substantially, now including a rigorous justification of the applicability of the limit result from [70], as well

as a more detailed description of the algorithm along with the corresponding pseudo code. In the discussion of the simulation results, more details on the necessity of including T-cell inhibition and a shielding effect into the model are included. Moreover, the critical threshold for the tumour size at treatment onset and stochastic effects in simulations for spontaneous mutations are analysed in more detail.

4.1 Medical background and experimental results

In this section we want to give a short overview of the biological context of our research, as well as the experimental setup and results. A more general background on cancer, therapy approaches, and the immune system is given in Section 1.4.

Cytotoxic $CD8^+$ T-cells play an important role in tumour immune surveillance. They recognise peptides (epitopes) derived from tumour cell-encoded gene products (antigens) which are presented on major histocompatibility complex (MHC) class I molecules. Epitope-specific activation of $CD8^+$ T-cells leads to tumour cell killing through the release of cytotoxic granules and cytokines, amongst others, which can be exploited therapeutically by different strategies. One approach is genetic engineering of autologous $CD8^+$ T-cells by introducing a tumour antigen-specific T-cell receptor (TCR) [61, 118, 178]. After *ex vivo* expansion these TCR-transgenic (TCRtg) $CD8^+$ T-cells can be re-infused into the same patient, a procedure known as adoptive T-cell transfer (ACT). Clinical trials have been conducted in patients with various types of cancer showing that the emergence of resistant tumour cell variants restrains durable responses [40, 134, 168].

Malignant melanoma, an aggressive type of skin cancer, is a paradigm disease for the development of novel immunotherapies including ACT. Using mouse models, it was previously found that melanomas can escape from ACT targeting the melanocyte differentiation antigen Pmel (also known as gp100) by dedifferentiation [117], an epigenetic mechanism also known as phenotype switching or phenotypic plasticity [93, 94]. Briefly, infiltrating activated Pmel-specific $CD8^+$ T-cells (Pmel-1 T-cells) instigated pro-inflammatory cytokine release which induced dedifferentiation and down-regulation of the Pmel antigen in melanoma cells, which impaired immune recognition and killing [117, 156]. Thus, phenotypic plasticity of melanoma cells emerges as a relevant mechanism of resistance to immunotherapy [2, 97].

A better understanding how the various cell populations interact over time could help to improve current ACT regimens, but longitudinal analyses impose challenges with regard to tissue sampling, both in patients and animal models. We therefore reason that mathematical modeling as in [67, 116] can complement the experimental approaches. Mathematical models allowed us to easily manipulate the system and potentially make novel predictions. Baar et al. have proposed an individual-based stochastic model of ACT [11]. Using adjusted parameters the model faithfully recapitulated tumour growth kinetics and melanoma cell state transitions as reported in the previous experimental study [117].

Here, we advanced the experimental and mathematical models of ACT in order to compare the evolution of gene edited Pmel antigen loss variants (Pmel knockout; Pmel^{KO}) with adaptive dedifferentiation, because genetically hardwired loss of target antigen expression or presentation is another key mechanism of resistance to immunotherapy [155, 166, 179]. Pmel^{KO} melanoma cells were positively selected by ACT, but unexpectedly they exhibited a growth defect in the absence of therapy imposed by the bulk wild type (WT) melanoma cell population. This established an evolutionary scenario of competing tumour cell populations, where Pmel^{KO} and WT melanoma cells switch their fitness in response to ACT.

4.1.1 Experimental setup

Experiments were generally conducted both with HCmel12 and B16F1 melanoma cells. The latter were used to study the effects of METi on the immune system since they do not respond to an increased level of METi themselves. However, since the aim of the mathematical modelling approach was mostly to understand evolutionary dynamics within the tumour, we focus on the HCmel12 melanoma cells, which were used for the generation of Pmel^{KO} cells. The setup for all the experiments is described extensively in [88].

Therapy protocol

The adoptive T-cell transfer immunotherapy (ACT) protocol that is applied here is similar to clinical protocols [40, 134] and is an improved version of the ACT therapy described in [110, 117]. It adds injections of small molecule inhibitors (METi, e.g. capmatinib) of the c-MET receptor tyrosine kinase and was shown to enhance ACT efficacy in [87]. We only give the short version of the protocol here (visualised in Figure 4.1) with the information that is relevant to the mathematical model.

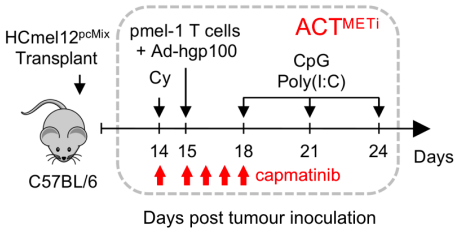


Figure 4.1: Exemplary ACT^{METi} therapy protocol.

On day 1, cohorts of syngeneic C57BL/6J mice were injected with a total of $2 \cdot 10^5$ HCmel12 melanoma cells (WT, Pmel^{KO} or mixtures thereof at different ratios) into the skin of the flank. When the transplanted tumours reached a size of 3-5 mm diameter (usually around day 14-18), the advanced ACT^{METi} therapy was initiated. For five consecutive days, the mice were injected with METi into the body cavity every 12 hours. On the second day of METi injections, $2 \cdot 10^6$ naïve Pmel-specific T-cells were injected into the blood and activated *in vivo* by an adenoviral vector expressing Ad-hgp100. For animal welfare reasons mice were sacrificed when tumours exceeded 10 mm in diameter or when signs of illness were observed.

Generation of genetic knockout variants

The polyclonal HCmel12 Pmel^{KO} variants were generated using the CRISPR-Cas9 genome engineering technology, targeting the Pmel gene upstream of the region encoding for the gp100 epitope separately with three different guide RNA sequences in parental HCmel12 melanoma cells. With this technique, the DNA is cut in a specific position and several nucleotides are deleted or inserted at random, producing Pmel^{indel} cells. Depending on whether the number of nucleotides differs from that of the wild type by a multiple of three or not, the reading frame for the production of amino acids stays intact or is shifted, resulting in so-called in-frame or frame-shift indels. Only frame-shift indels (or in-frame indels that inserted/deleted a rather long sequence of nucleotides, which are quite rare) truly change the structure of the resulting epitope and produce a functional knockout. We hence focus on the frame-shift variants and call those Pmel^{KO} cells. In-frame indels are considered as equivalent to wild type cells, which is supported by the observation that they are preserved in untreated tumours and almost eradicated under ACT^{METi} therapy (see [88] for details). To generate monoclonal HCmel12 Pmel^{KO} single cell clones, individual cells were selected and expanded *in vitro*. Frequencies of frame-shift and in-frame Pmel^{indel} cells and functional knockouts, in the case of single cell clones, were determined by amplicon next generation sequencing (NGS).

The described procedure results in multiple subpopulations of Pmel^{KO} cells, both due to different positions of cutting the DNA and different insertion and deletion of nucleotides. Experiments were conducted both with polyclonal (genetically slightly different, corresponding to the same guide RNA) and monoclonal (established from a single cell and hence genetically identical) HCmel12 Pmel^{KO} cultures. They were either injected into the mice alone or mixed with different Pmel^{KO} and/or parental HCmel12 (WT) cultures at different ratios.

4.1.2 Experimental results

In this subsection we summarise the most important experimental results. The types of data that were collected and led to the conclusions are: The initial composition of the melanoma cell population at tumour inoculation, the tumour size measured as diameter in mm twice weekly, and (in most cases) the tumour composition at the time of harvesting. The composition of the melanoma cell population, i.e. the frequency of WT cells and different types of Pmel^{KO} cells, was determined by amplicon NGS.

Effects of MET inhibitor (METi) treatment

Regular injections of METi alter the evolution of HCmel12 melanomas in several ways. On the one hand, they directly influence the melanoma cells by limiting their population growth (through inhibition of c-MET signalling) and partially counteracting the TNF-α induced dedifferentiation. The growth deficit was witnessed comparing *in vivo* growth curves of untreated HCmel12 tumours with and without METi injections (Figure 4.2A), as well as *in vitro* cell growth of WT and METi-resistant HCmel12 melanoma cells under METi treatment (Figure 4.2B). The Pmel level of HCmel12 cells, i.e. their (de)differentiated phenotype, was

studied *in vitro*, where cell cultures were treated with TNF-α, METi, or a combination of both, as well as control vehicles (Figure 4.2C). The results suggest that METi has little effect on the differentiation of melanoma cells in the absence of TNF-α. However, it mostly reverses the cytokine induced dedifferentiation in the presence of TNF-α.

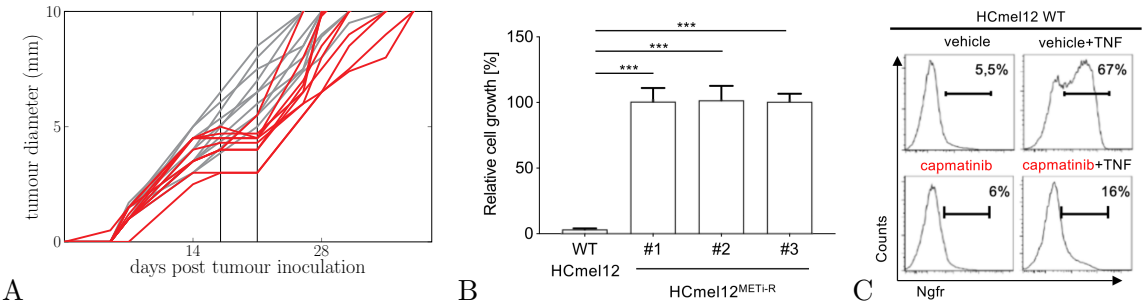


Figure 4.2: (A) Tumour growth curves of HCmel12 melanomas, untreated (gray) or treated with METi between day 17 and 21 (red) [87]. (B) Quantification of cell growth of WT and METi-resistant HCmel12 cells *in vitro* over 6 days, percentages compare growth under METi treatment to untreated controls (n=5). (C) Flow cytometric analyses of Ngfr (melanoma dedifferentiation maker) surface expression in HCmel12 WT melanoma cells exposed to TNF-α or left untreated in combination with vehicle (DMSO) or METi (capmatinib) treatment.

On the other hand, experiments with HCmel12 METi-resistant variants (HCmel12$^{\text{METi-R}}$) and B16F1 melanoma cells (which are METi-resistant by nature) have shown that, upon ACT immunotherapy, METi injections also influence the melanoma-T-cell-dynamics beyond their direct influence on the melanoma cells, namely through an increased Pmel-1 T-cell activity. A higher level of METi both accelerates T-cell expansion (Figure 4.3B) and increases T-cell efficacy (Figure 4.3A) by blocking the recruitment of T-cell-suppressive neutrophils early during treatment [87].

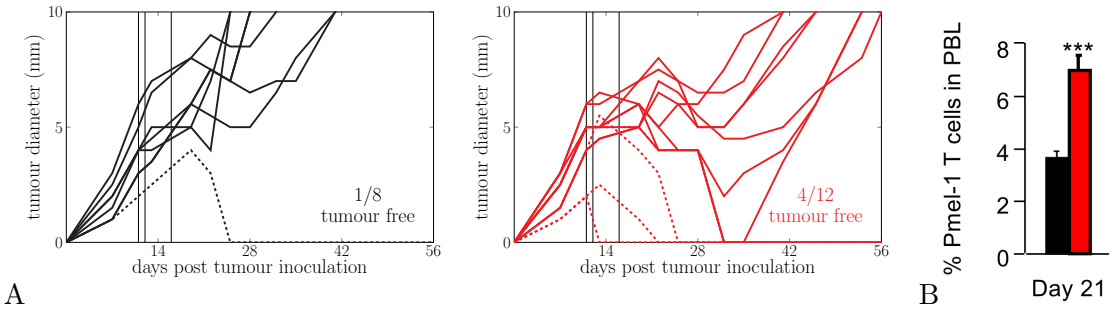


Figure 4.3: (A) Individual tumour growth curves of experiments comparing the therapeutic efficacy of ACT (black) and ACT$^{\text{METi}}$ (red) in mice bearing HCmel12$^{\text{METi-R}}$ melanomas. Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections. Dashed lines indicate tumours undergoing eradication. (B) Flow cytometry-based quantification of the percentage of Pmel-1 T-cells in peripheral blood leukocytes from mice treated as in A (1 week after onset of ACT regimens). (n=8 for ACT; n=12 for ACT$^{\text{METi}}$).

Enrichment of Pmel^{KO} variants in untreated and ACT^{METi}-recurrent melanomas

Experiments with mixed melanomas suggest an interesting evolutionary scenario with varying context-dependent fitness of Pmel^{KO}cells. Polyclonal HCmel12 Pmel^{indel} cells were mixed with HCmel12 WT cells (HCmel12^{pcMix}) to obtain a Pmel^{indel} allele frequency of 20% (roughly 17% of Pmel^{KO}, see Figure 4.4). This mixture was injected into mice and animals bearing established HCmel12^{pcMix} melanomas were treated with ACT^{METi}, as described above, or left untreated for control purposes. When tumours reached a size of 10 mm in diameter, the mice were sacrificed and genomic DNA was isolated to determine the Pmel^{KO} allele frequencies. The results are visualised in Figure 4.5.

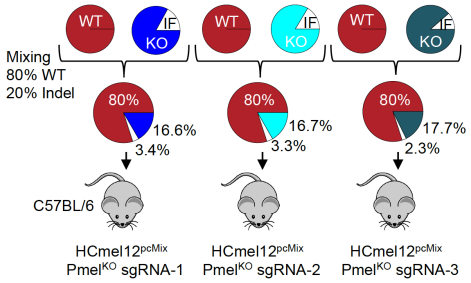


Figure 4.4: Composition of HCmel12^{pcMix} melanomas at tumour inoculation for three different guide RNAs.

In untreated HCmel12^{pcMix} tumours, the average frequency of Pmel^{KO} alleles declined to a level of 7.8%, which suggests a reduced fitness in the context of a bulk HCmel12 WT population when the immunological selection pressure by Pmel-1 T-cells is absent. When compared to pure HCmel12 WT tumours, pure HCmel12 Pmel^{KO} tumours exhibited very similar growth rates (see Figure 4.6). This result argues for a scenario where HCmel12 WT cells impose a large competitive pressure on HCmel12 Pmel^{KO} cells.

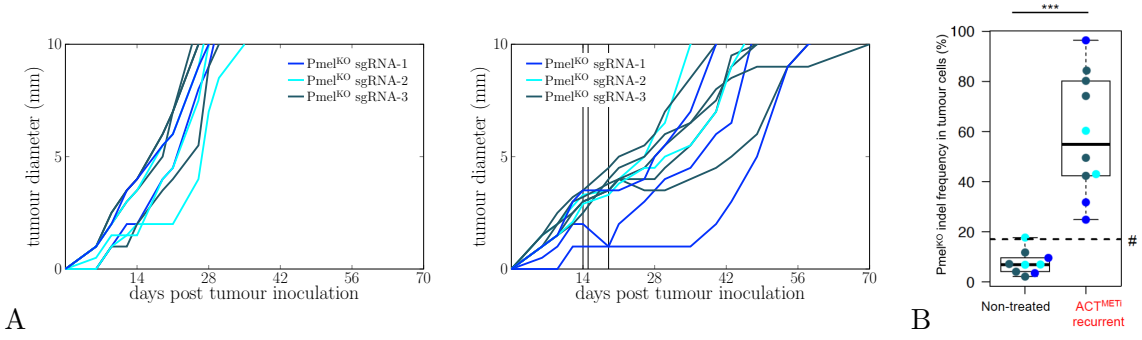


Figure 4.5: (A) Individual tumour growth curves of HCmel12^{pcMix} melanomas left untreated or treated with ACT^{METi}. Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections. (B) Pmel^{KO} allele percentages in untreated or recurrent ACT^{METi}-treated HCmel12^{pcMix} melanomas, $^{\#}$average Pmel^{KO} frequency at tumour inoculation (17%).

When treated with ACT^{METi} therapy, HCmel12^{pcMix} melanomas could no longer be eradicated (as they were in the pure WT experiments in [87]). Hence, despite the fact that

melanoma cells can down-regulate the Pmel expression by dedifferentiation, a complete genetic abrogation of antigen expression provides a survival benefit for the tumour. The experiments show a strong enrichment of Pmel^{KO} cells in recurrent $\text{HCmel12}^{\text{pcMix}}$ melanomas with an average of 58.5%, which suggests a gain in fitness in comparison to HCmel12 WT cells in response to ACT^{METi}. However, the enrichment was surprisingly highly variable, ranging from 20% to almost 100%.

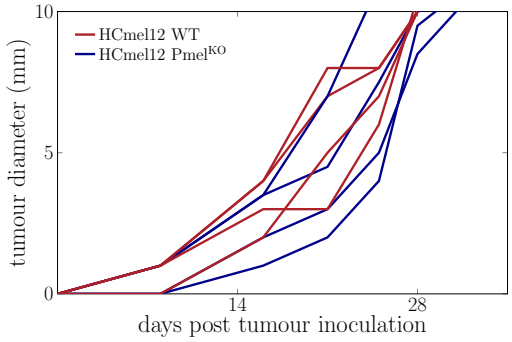


Figure 4.6: Tumour growth curves of HCmel12 WT (red) versus mixed HCmel12 Pmel^{KO} (blue, sgRNA-1/2/3 at 1:1:1) melanomas in untreated mice.

In conclusion, this finding supported the notion that genetic Pmel loss reduced HCmel12 fitness in untreated melanomas, but strongly increased their fitness under ACT^{METi}.

4.2 Extension of the mathematical model

The experimental approach with $\text{HCmel12}^{\text{pcMix}}$ melanomas, described in the previous section, establishes an evolutionary scenario of competing tumour cell populations undergoing a reciprocal fitness switch upon ACT^{METi} therapy. This prompts us to use mathematical modelling in order to better understand cell population dynamics and potentially explain the highly variable enrichment of Pmel^{KO} cells found in recurrent melanomas.

4.2.1 Stochastic model of ACT^{METi} with Pmel^{KO} variants

To study the evolution of melanomas under the improved ACT^{METi} therapy, we use an individual-based continuous-time Markov process that is an extension of the model by Baar et al. [11]. This model itself is based on an individual-based model of adaptive dynamics, that we introduce in Section 1.2 and study in Chapters 2 and 3.

In the following, we adjust the model in [11] according to the context of $\text{HCmel12}^{\text{pcMix}}$ melanomas treated with ACT^{METi}. The main changes are the addition of negative feedback in the immune system, i.e. cytokine-mediated inhibition of T-cells, the possibility of genetic mutations, and the shielding effect against T-cell infiltration of the tumour. We consider six different types of interacting cell populations or molecules, collectively termed as individuals:

Differentiated (Diff) and dedifferentiated (Dedi) WT melanoma cells, Pmel^{KO} melanoma cells (KO), CD8^+ Pmel-1 T-cells (CD8), cytokines (Cyto), and dead melanomas cells (Dead). The latter are included, because they contribute to the measured tumour sizes, although they do not influence the evolution of the other cells in the mathematical model. Cytokines comprise a variety of different molecules, in particular T-cell effector cytokines such as TNF-α and IFN-γ, that evoke a pro-inflammatory microenvironment promoting melanoma cell dedifferentiation and up-regulation of negative immune checkpoint molecules [39, 117, 154, 156]. The state of the mathematical process

$$N(t) = (N_{Diff}(t), N_{Dedi}(t), N_{KO}(t), N_{Dead}(t), N_{CD8}(t), N_{Cyto}(t)) \tag{4.1}$$

describes the numbers of these different types of individuals (cells, cytokines) that are present in the tumour tissue as a function of time t starting with tumour cell inoculation at $t = 0$.

The dynamics of the Markov process are determined by a number of events that change the state of the population and occur at exponential rates. Those rates depend on fixed parameters as well as the current state of the population. The evolution of the Markov process is described by its infinitesimal generator that is of the form

$$L\phi(N) = \sum_{e \in \mathcal{E}} (\phi(N + v_e) - \phi(N)) R_e(N), \tag{4.2}$$

where $\mathcal{E}$ is the set of possible events, v_e is the change in the population associated to an event, $R_e(N)$ is the rate at which the event occurs, and ϕ is a measurable bounded function. It can be constructed rigorously similar to the process in [77]. In the model we consider the following events and rates ($\delta_i = i^{\text{th}}$ unit vector), visualised in Figure 4.7:

Differentiated WT melanoma cells

- reproduce clonal: $v_e = \delta_{Diff}$, $R_e(N) = (1-m) b_{Di} N_{Diff}$
- switch to dedifferentiated state (naturally): $v_e = -\delta_{Diff} + \delta_{Dedi}$, $R_e(N) = s_{Di,De} N_{Diff}$
- die (naturally and due to competitive pressure): $v_e = -\delta_{Diff} + \delta_{Dead}$,
  $R_e(N) = (d_{Di} + c_{Di,Di} N_{Diff} + c_{Di,De} N_{Dedi} + c_{Di,KO} N_{KO}) N_{Diff}$

Dedifferentiated WT melanoma cells

- reproduce clonal: $v_e = \delta_{Dedi}$, $R_e(N) = (1-m) b_{De} N_{Dedi}$
- switch to differentiated state (naturally): $v_e = -\delta_{Dedi} + \delta_{Diff}$, $R_e(N) = s_{De,Di} N_{Dedi}$
- die (naturally and due to competitive pressure): $v_e = -\delta_{Dedi} + \delta_{Dead}$,
  $R_e(N) = (d_{De} + c_{De,Di} N_{Diff} + c_{De,De} N_{Dedi} + c_{De,KO} N_{KO}) N_{Dedi}$

Pmel^{KO} melanoma cells

- reproduce clonal: $v_e = \delta_{KO}$, $R_e(N) = b_{KO} N_{KO}$
- arise as (spontaneous) mutants from (de)differentiated melanoma cells: $v_e = \delta_{KO}$,
  $R_e(N) = m(b_{Di} N_{Diff} + b_{De} N_{Dedi})$
- die (naturally and due to competitive pressure): $v_e = -\delta_{KO} + \delta_{Dead}$,
  $R_e(N) = (d_{KO} + c_{KO,Di} N_{Diff} + c_{KO,De} N_{Dedi} + c_{KO,KO} N_{KO}) N_{KO}$

Dead melanoma cells

- get disintegrated: $v_e = -\delta_{Dead}$, $R_e(N) = d_{Dead}N_{Dead}$

CD8$^+$ Pmel-1 T-cells (abbreviated as T-cells)

- reproduce/get activated/get recruited and simultaneously secrete ℓ cytokines: $v_e = \delta_{CD8}+\ell\delta_{Cyto}$, $R_e(N) = b_{CD}N_{Diff}N_{CD8}[N_{Diff}/(N_{Diff}+N_{Dedi}+N_{KO})]^{\alpha}(1-hN_{Cyto})_+$
- kill differentiated melanoma cells: $v_e = -\delta_{Diff} + \delta_{Dead}$, $R_e(N) = k_{Di}N_{Diff}N_{CD8}[N_{Diff}/(N_{Diff} + N_{Dedi} + N_{KO})]^{\alpha}(1 - hN_{Cyto})_+$
- die/get inactive: $v_e = -\delta_{CD8}$, $R_e(N) = d_{CD}N_{CD8}$

Cytokines

- induce an additional dedifferentiation of melanoma cells: $v_e = -\delta_{Diff} + \delta_{Dedi}$, $R_e(N) = s^{Cy}_{Di,De}N_{Cyto}N_{Diff}$
- get disintegrated: $v_e = -\delta_{Cyto}$, $R_e(N) = d_{Cy}N_{Cyto}$

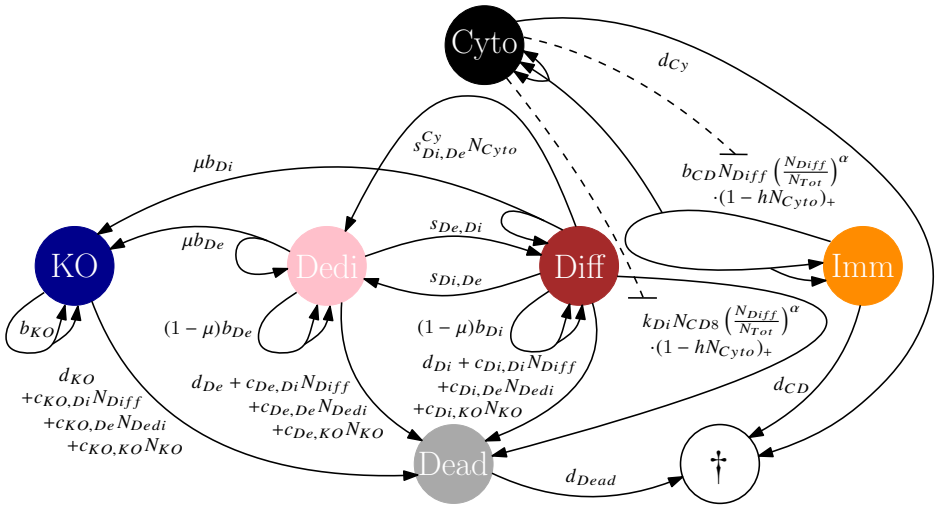


Figure 4.7: Interaction diagram displaying cells/molecules (filled circles) and mechanisms that are incorporated in the mathematical model. Arrows represent possible changes to the state of the system, i.e. events $e \in \mathcal{E}$, like cell division, cytokine secretion, or cell death. Formulae describe the rates at which these population changes occur (per individual from which the arrow emanates).

Most rates correspond to a standard birth-and-death process with competition and switching/mutation. The unusual rates are the ones of T-cell reproduction and melanoma cell killing. They do not only depend on the number of differentiated melanoma cells and T-cells but include two additional factors. $[N_{Diff}/(N_{Diff} + N_{Dedi} + N_{KO})]^{\alpha}$ represents the effect of differentiated cells being shielded from the T-cells by other melanoma cells, not susceptible to the T-cells. In an otherwise non-spatial model that assumes the individuals to be well-mixed, it partially takes into account the spatial structure of the tumour and the fact that dedifferentiated WT and Pmel$^{\mathrm{KO}}$ melanoma cells can physically prevent an effective infiltration of the differentiated WT melanoma cells by the T-cells. The proportion

of differentiated cells $[N_{Diff}/(N_{Diff} + N_{Dedi} + N_{KO})]$ is weighed by a parameter α that determines the influence of this shielding effect. The factor $(1 - hN_{Cyto})_+$ corresponds to cytokine-induced T-cell inhibition. As the number of inflammatory cytokines increases, the melanoma cells up-regulate PD-1 ligands (PD-L1) that inhibit the immune reaction. We comment on the necessity of adding the shielding effect and T-cell inhibition to the model in the next section, where we discuss the simulation results.

The ACT^{METi} therapy protocol is modelled through an addition of T-cells on the respective day of adoptive T-cell transfer and a temporary change of parameters during the five days of METi injections. These injections of METi are not modelled by an additional particle, which is reasonable since regular injections ensure a relatively stable level of METi, which is then degraded quickly after the injections stop. During the five days, melanoma cell reproduction, cytokine-induced dedifferentiation of WT melanoma cells, and T-cell inhibition are down-regulated, while reproduction and killing activity of T-cells are up-regulated, based on the experimental data in [87].

4.2.2 Law of large numbers/deterministic approximation and hybrid algorithm

In simulations, we use a Gillespie-type algorithm that generates a realisation of the stochastic process by simulating single events [83]. Since this method is computationally intensive in large populations with frequent events, we combine stochastic simulations of rare events and deterministic simulations (employing Runge-Kutta methods [27, 157, 115]) of frequent events to reduce the running time while keeping random effects such as subpopulations dying out. Similar approaches have been discussed in [126, 159].

Deterministic approximation

A deterministic approximation of frequent events is reasonable due to the law of large numbers result from Ethier and Kurtz [70]. For an extensive discussion of limit approximations of individual-based Markov processes see the previous two chapters of this thesis. In this particular context, the result can be stated as follows:

Theorem 4.1 ([70], Ch. 11, Thm. 2.1)**.** *Let N^K, $K \in \mathbb{N}$, be Markov processes on $\Omega \subset \mathbb{R}^d$ with generators*

$$\tilde{\mathcal{L}}^K \phi(N) = \sum_{e \in \mathcal{E}} (\phi(N + v_e) - \phi(N)) K \tilde{R}_e \left(\frac{N}{K} \right). \tag{4.3}$$

Suppose that, for each compact $\Omega_0 \subset \Omega$,

$$\sum_{e \in \mathcal{E}} |v_e| \sup_{n \in \Omega_0} \tilde{R}_e(n) < \infty \tag{4.4}$$

and there exists $C_{\Omega_0} < \infty$ such that

$$\sup_{n_1, n_2 \in \Omega_0} \left| \sum_{e \in \mathcal{E}} v_e (\tilde{R}_e(n_1) - \tilde{R}_e(n_2)) \right| \leq C_{\Omega_0} |n_1 - n_2|. \tag{4.5}$$

Assume that $N^K(0)/K \to n_0 \in \Omega$ almost surely, as $K \to \infty$. Then, for every $T \geq 0$, $(N^K(t)/K)_{t\in[0,T]}$ almost surely converges uniformly to $(n(t))_{t\in[0,T]}$, which is the unique solution to the differential equation

$$\dot{n}(t) = \sum_{e\in\mathcal{E}} v_e \tilde{R}_e(n(t)) \tag{4.6}$$

with initial condition $n(0) = n_0$.

We want to apply this result to approximate the stochastic process that is described above with the deterministic solution to a differential equation. If, for some large but fixed $K \in \mathbb{N}$, we can rewrite the generator of $N(t)$ with rates $\tilde{R}_e$ as

$$\mathcal{L}\phi(N) = \sum_{e\in\mathcal{E}} (\phi(N + v_e) - \phi(N))R_e(N) = \sum_{e\in\mathcal{E}} (\phi(N + v_e) - \phi(N))K\tilde{R}_e\left(\frac{N}{K}\right) \tag{4.7}$$

and check the conditions of the theorem for these rates $\tilde{R}_e$, then $N(t)/K \approx n(t)$ with

$$\dot{n}(t) = \sum_{e\in\mathcal{E}} v_e \tilde{R}_e(n(t)). \tag{4.8}$$

To construct the new rates $\tilde{R}_e$, we have to rescale some of the parameters in the non-linear original rates R_e with the fixed K. More precisely, we replace all competition parameters $c_{*,*}$ with $Kc_{*,*}$, the parameter for cytokine-induced dedifferentiation $s^{Cy}_{Di,De}$ with $Ks^{Cy}_{Di,De}$, the parameter for T-cell inhibition h with Kh, the parameter for T-cell proliferation b_{CD} with Kb_{CD}, and the parameter for the killing of differentiated WT cells k_{Di} with Kk_{Di}. For example, if e is the event of cytokine-induced dedifferentiation,

$$R_e(N) = s^{Cy}_{Di,De} N_{Cyto} N_{Diff} = K(Ks^{Cy}_{Di,De})\frac{N_{Cyto}}{K}\frac{N_{Diff}}{K}, \tag{4.9}$$

and hence $\tilde{R}_e(n) = Ks^{Cy}_{Di,De} n_{Cyto} n_{Diff}$. Apart from these changes, the new rates are the same as the original ones.

We check the conditions for these $\tilde{R}_e$ on

$$\Omega := \left\{ n \in \mathbb{R}_+^{\{Diff,Dedi,KO,Dead,CD8,Cyto\}} : n_{Diff} + n_{Dedi} + n_{KO} > \varepsilon \right\}, \tag{4.10}$$

for some small $\varepsilon > 0$.

Condition (4.4) is automatically satisfied since $\mathcal{E}$ is finite, Ω_0 is compact, and the only non-polynomial factor $n_{Diff}/(n_{Diff} + n_{Dedi} + n_{KO})$ is bounded by 1 for all $n \in \Omega$.

To check condition (4.5), we note that, since $\mathcal{E}$ is finite, we can look at each event e separately. We can also consider different factors individually, as long as they are bounded on compact subsets $\Omega_0 \subset \Omega$, since

$$\begin{aligned}
&\left| v_e(\tilde{R}^1_e(n_1)\tilde{R}^2_e(n_1) - \tilde{R}^1_e(n_2)\tilde{R}^2_e(n_2)) \right| \\
&\leq |v_e| \left(\left|\tilde{R}^1_e(n_1)\right| \cdot \left|\tilde{R}^2_e(n_1) - \tilde{R}^2_e(n_2)\right| + \left|\tilde{R}^2_e(n_2))\right| \cdot \left|\tilde{R}^1_e(n_1) - \tilde{R}^1_e(n_2)\right| \right).
\end{aligned} \tag{4.11}$$

Since $|(1 - hn_{1,Cyto})_+ - (1 - hn_{2,Cyto})_+| \leq h|n_1 - n_2|$ and $1/(n_{Diff} + n_{Dedi} + n_{KO})$ is both bounded and Lipschitz-continuous away from $n_{Diff} + n_{Dedi} + n_{KO} = 0$, condition (4.5) is satisfied for our choice of Ω.

Hybrid algorithm

We now want to apply this approximation result to speed up simulations for our stochastic model. In collaboration with the group of Prof. Martin Rumpf at the Institute for Numerical Simulations, in particular his student Kai Echelmeyer, we have developed a hybrid algorithm that combines a classical Gillespie algorithm with Runge-Kutta methods [66].

While a Gillespie algorithm produces an exact realisation of the stochastic process, it is computationally very heavy since every event is simulated separately. This particularly takes effect in large populations, where there are many reproduction and death events in a short period of time. In this case, the algorithm must perform many steps in order to simulate a time span of interest.

In contrast to this, the Runge-Kutta methods for the deterministic approximation of the stochastic process run much faster. However, they neglect any stochasticity, which is particularly relevant in small populations. In those scenarios, random fluctuations could cause extinction of a subpopulation, while the deterministic solution will never quite hit zero. We want to keep these stochastic effects since for example the T-cells could die out and through this induce a relapse of the tumour.

In our algorithm we do not distinguish between small and large subpopulations, but rare and frequent events. We apply the deterministic approximation to the frequent events, i.e. events with rates $\tilde{R}_e$ above a certain threshold, which is not directly in line with the choice of Ω above. However, since the rates depend on the sizes of the different subpopulations, there is a direct correspondence and we can choose ε such that all population states that produce (partially) super-critical rates lie in Ω.

The stochastic-deterministic hybrid algorithm is implemented in C++. In the algorithm, the rates $\tilde{R}$ depend on the time as well as the state of the population. This is due to the fact that we allow for a change of parameters, e.g. during METi-injections. The threshold values $lower(K)$ and $upper(K)$, that determine whether an event is treated stochastically or deterministically, are chosen of order $K^{-1/2}$. This is because fluctuations of the rescaled stochastic process are of order $K^{-1/2}$ (compare for example the diffusion approximation in [171]) and we want to make sure that we allow for extinction of a subpopulation due to stochastic fluctuations. The parameters α_{ij}, β_i, and γ_i determine the Runge-Kutta method and are in our case chosen as follows (summarised in the corresponding Butcher-tableau):

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$$

In the following we present the pseudo code for our stochastic-deterministic hybrid algorithm.

More details can be found in [66].

Algorithm 1: Hybrid Algorithm

Input: number of types d, capacity K,
event set $\mathcal{E}$ with event vectors v_e and rates $\tilde{R}_e(t,n)$,
stochastic events $\mathcal{E}_s$, deterministic events $\mathcal{E}_d$,
thresholds for update $lower(K) < upper(K)$,
maximal number of iterations k_{max}, end time T, step size τ
initial condition n_0
Output: Evolution of types $\{t_k, n(t_k)\}_k$
$k = 0$, $t_0 = 0$, $n(0) = n_0$
while $k \leq k_{max}$ *and* $t_k \leq T$ *and* $\sum_{i=1}^{d} n_i(t_k) > 0$ **do**
  for $e \in \mathcal{E}_s$ **do**
    if $\tilde{R}_e(t_k, n(t_k)) > upper(K)$ **then**
      $\mathcal{E}_s = \mathcal{E}_s \backslash e$, $\mathcal{E}_d = \mathcal{E}_d \cup e$
  for $e \in \mathcal{E}_d$ **do**
    if $\tilde{R}_e(t_k, n(t_k)) < lower(K)$ **then**
      $\mathcal{E}_d = \mathcal{E}_d \backslash e$, $\mathcal{E}_s = \mathcal{E}_s \cup e$
  $t_{k+1} := t_k + \tau$, $t \leftarrow 0$, $w \leftarrow 0$
  while $t < \tau$ **do**
    $\tilde{R}_{tot} \leftarrow \sum_{e \in \mathcal{E}_s} \tilde{R}_e(t_k + t, n(t_k) + w))$
    sample $t_{tot} \sim \mathrm{Exp}(K\tilde{R}_{tot})$
    if $t + t_{tot} \leq \tau$ **then**
      sample $e^* \in \mathcal{E}_s$ proportional to $\{\tilde{R}_e(t_k + t, n(t_k) + w)\}_{e \in \mathcal{E}_s}$
      $w \leftarrow w + \frac{v_{e^*}}{K}$
    $t \leftarrow t + t_{tot}$
  for $1 \leq i \leq 4$ **do**
    $\kappa_i \leftarrow 0$
    for $e \in \mathcal{E}_d$ **do**
      $\kappa_i \leftarrow \kappa_i + \tilde{R}_e\left(t_k + \gamma_i\tau, n(t_k) + \tau \sum_{j=1}^{i-1} \alpha_{ij}\kappa_j\right)$
  $n(t_{k+1}) := \frac{\left[K\left(n(t_k) + \tau \sum_{i=1}^{4} \beta_i \kappa_i\right)\right]}{K}$
  $n(t_{k+1}) \leftarrow n(t_{k+1}) + w$
  $k \leftarrow k + 1$

4.2.3 Parameter choices

We have two sets of parameters corresponding to two sets of experiments. First, the experiments with pure wild type tumours from Glodde et al. [87], which are displayed in Figure 4.11A, and second, the experiments including Pmel$^{\mathrm{KO}}$ cells that are new to [88]. Between the sets of experiments, the melanoma cells showed slightly different behaviour (e.g. different speed of growth), however this can be achieved through variation of the parameters, leaving the systemic level (different events, structure of the rates) unchanged.

In the following, we describe how the parameter values were either derived from the experiments to fit the model to the data or chosen to simulate scenarios beyond the experiments. The parameter choices for the two experimental setups and with or without the influence of METi injections are summarised in Table 4.1.

Initial cell numbers and scaling parameter K

We assume that the tumour takes the form of a 3-dimensional ball to relate the number of cells in our model to the tumour diameter measured in the experiments. A 5 mm tumour contains roughly $7 \cdot 10^7$ cells and since the measured tumour sizes vary from 1 mm $\approx 5.6 \cdot 10^5$ cells to 10 mm $\approx 5.6 \cdot 10^8$ cells, we choose $K = 10^7$ as the typical size for the deterministic approximation. The number of initially injected melanoma cells is $2 \cdot 10^5$. Since likely not all of these cells survive and contribute to form the growing tumour, the initial condition $N_{Diff}(0) + N_{Dedi}(0) + N_{KO}(0)$ is varied between 10^4 and $2 \cdot 10^5$.

Reproduction, death, and competition rates

The growth parameters of the different melanoma cells can be approximated from *in vivo* experiments where tumour cells are injected into mice and then left to grow without treatment. We set the death rates to 0.1 and assume the same rates for differentiated and dedifferentiated wild type tumour cells, since we cannot distinguish them in the experiments. Through a logistic fitting, we determine the birth and self-competition rates. The cross-competition between Pmel^{KO} and WT melanoma cells is only relevant to the new experiments. As mentioned in the previous section and seen in Figure 4.6, the growth rates of pure WT and Pmel^{KO} tumours are very similar and hence the competitive pressure that the WT melanoma cells impose on the Pmel^{KO} melanoma cells is chosen relatively large to fit the percentages in Figure 4.5B (results in Figure 4.16B).

Switch between differentiated and dedifferentiated melanoma cells

The rates for natural and cytokine-induced switching between differentiated and dedifferentiated WT melanoma cells are chosen to fit experiments from Figure 4.2C, where WT melanoma cells are stimulated *in vitro* with TNF-α and/or METi for 72 hours, during which an equilibrium is attained. Without treatment, in a vehicle control experiment, a ratio of roughly 0.95:0.05 of differentiated to dedifferentiated melanoma cells is observed, which determines the ratio between the natural switch rates. The actual size of the rates is then chosen to make sure that an equilibrium is attained within 72 hours. This ratio shifts to approximately 0.65:0.35 under the influence of TNF-α, which indicates an additional dedifferentiation, induced by inflammatory cytokines. Under addition of METi alone, the ratio is about the same as in the control case, while a combination with TNF-α results in a ratio of roughly 0.85:0.15. This suggests that METi has no influence on the natural switch rates, while partially cancelling the cytokine-induced switch.

T-cell activity and cytokines

The therapy parameters (reproduction, death, killing efficiency, and inhibition of T-cells, secretion and degradation of cytokines) are chosen to fit the experiments in Figure 4.11A and experiments with pure WT tumours corresponding to the protocol in Figure 4.1. During the course of therapy, $2 \cdot 10^6$ Pmel-1 T-cells are injected into the mice, which do not all infiltrate the tumour tissue. Simulations show that a variation of the number of T-cells has little to no effect on the evolution of the tumour since the amount of differentiated WT cells that can be killed is limited by the T-cell inhibition. Therefore, we fix the number of T-cells that are added on the respective day to 10^5 for all simulation runs. The variation between different mice is obtained by running simulations with different initial conditions (initial number of melanoma cells). The effect of METi on the proliferation, killing efficiency, and inhibition of T-cells is determined comparing experiments with therapy protocols including and excluding METi injections.

Dead melanoma cells

The parameter for the degradation of dead melanoma cells is chosen to fit the descent of the tumour size during the 5 days of METi injections. During this phase, the most dead melanoma cells are produced due to effective killing by T-cells, which contributes to the measured tumour diameter.

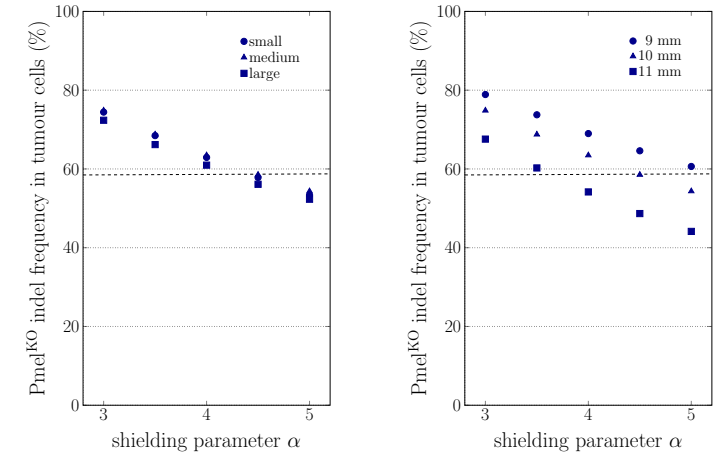


Figure 4.8: (A) Percentage of Pmel^{KO} cells in simulations, taken at tumour size of 10 mm, varying parameter α and initial size of the tumour. Initial portion of Pmel^{KO} cells always at 17.1%. (B) Percentage of Pmel^{KO} cells in simulations, taken at varying tumour sizes and parameters α. Initial tumour of medium size with 17.1% Pmel^{KO} cells.

Shielding effect

The parameter α, which scales the effect of differentiated cells being shielded from T-cell infiltration by other melanoma cells, is chosen to fit the percentages in Figure 4.5B (results in Figure 4.16B). The parameter has little influence in the simulations with pure WT tumours, which is why we determine it after fitting the other therapy parameters. An increase in α corresponds to a larger shielding effect. Since the experimental data in Figure 4.5B has a broad spectrum, we investigate several sources of variation to determine α. In Figure 4.8A, the composition of the tumour is analysed at a diameter of 10 mm, while the initial conditions are varied (keeping the initial percentage of Pmel^{KO} cells at 17,1%). Even though we compare initial sizes that lie below and above the critical threshold for therapy success, the percentages are very similar. Therefore, we fix the initial number of cells to a medium amount of 10^5 cells in Figure 4.8B and vary the time point at which the percentage is taken between the times when 9 and 11 mm diameter are reached. Since, in the experiments, the tumour is sequenced at the time when it has a size of at least 10 mm diameter, we set α to 4 to obtain an average of around 60% Pmel^{KO} cells between 10 mm and 11mm.

Table 4.1 summarises the parameter sets corresponding to the pure wild type experiments from [87] and the experiments with Pmel^{KO} variants from [88]. Notice that these parameters correspond to the rescaled rates $\tilde{R}_e$ and the deterministic differential equation. Therefore, some of them are rescaled by K (see the previous subsection for details).

	[87], no METi	[87], METi	[88], no METi	[88], METi
K	10e7	10e7	10e7	10e7
$b_{Di} = b_{De}$	0.5	0.113	0.45	0.113
b_{KO}	-	-	0.46	0.33
$d_{Di} = d_{De} = d_{KO}$	0.1	0.1	0.1	0.1
$Kc_{Di,Di} = Kc_{Di,De} = Kc_{De,Di} = Kc_{De,De}$	0.0053	0.0002	0.0047	0.0002
$Kc_{KO,KO}$	-	-	0.0048	0.003
$Kc_{Di,KO} = c_{De,KO}$	-	-	0.00002	0.00002
$Kc_{KO,Di} = c_{KO,De}$	-	-	0.0117	0.0117
d_{Dead}	0.5	0.5	0.5	0.5
$s_{Di,De}$	0.05	0.05	0.05	0.05
$s_{De,Di}$	0.95	0.95	0.95	0.95
$Ks^{Cy}_{Di,De}$	2	0.02	2	0.02
Kb_{CD}	60	70	70	100
d_{CD}	0.4	0.4	0.4	0.4
Kk_{Di}	200	250	100	120
Kh	1.2	0.6	1.5	1
ℓ	3	3	3	3
d_{Cy}	1	1	1	1
α	4	4	4	4

Table 4.1: Parameter choices for the different simulation scenarios.

Natural mutation and variation of parameters

In our model, we introduce the event of a natural mutation to investigate the spontaneous occurrence of Pmel^{KO} mutants due to mutation from the WT population (in contrast to artificially mixed $\text{HCmel12}^{\text{pcMix}}$ tumours in experiments). We choose the mutation probability such that the occurrence of a mutant is likely but fixation is not ensured. With $m = 10^{-7}$, the probability of at least one mutant occurring before the therapy starts at day 14 is approximately 80% but the probability of more than three mutants occurring is only 10%. This can be calculated using the fact that the growth of the WT melanoma cell population is roughly exponential in the beginning and the occurrence of mutants is hence a Poisson point process with a known distribution. Note that for most simulations, e.g. the ones of $\text{HCmel12}^{\text{pcmix}}$ tumours, we set $m = 0$. This is however not significant since single naturally occurring mutants are of no consequence in a larger Pmel^{KO} cell population.

In Figures 4.17 and 4.21, the birth rate b_{KO} and natural death rate d_{KO} of the Pmel^{KO} mutants are varied between [0.3,0.55] and [0.01,0.26] respectively to obtain individual fitness values $r_{KO} = b_{KO} - d_{KO}$ in [0.2,0.45].

4.3 Simulation results and comparison to experiments

In this section we want to discuss the results of our simulations. We are able to characterise important mechanisms that need to be included in the model in order to reproduce the experimental data, like T-cell inhibition and the shielding of differentiated WT melanoma cells from T-cells. Moreover, we identify the most likely causes for the variable enrichment of Pmel^{KO} melanoma cells in ACT^{METi}-recurrent $\text{HCmel12}^{\text{pcMix}}$ tumours, namely the time point of sequencing the tumour and subclone fitness variability within the Pmel^{KO} cell population. Finally, in simulations we can predict the evolution of the tumour in the case of a spontaneous occurrence of single Pmel^{KO} cells through mutation, and thus validate the clinical relevance of the experimental results.

4.3.1 T-cell inhibition and threshold for therapy success

The first step in deriving the current model (as described in the previous section) was to adjust the model of Baar et al. [11] to the new experiments with pure WT melanomas in [87]. On the one hand, the HCmel12 melanoma cells in those experiments display much more aggressive growth dynamics than the HCmel3 melanoma cells in the experiments of [117], which are the foundation for the simulations of Baar et al. On the other hand, injections of METi were added to the therapy protocol.

Our first approach was to leave the systemic level of the model intact, i.e. keep the structure of the rates the same and only adjust the parameters to fit the experimental data. However, simulations of the tumour evolution under ACT^{METi} therapy could not reproduce the experimental data (compare Figure 4.11A). As Figure 4.9 shows, depending on the parameters for T-cell efficiency (reproduction rate and killing of differentiated WT cells), either all tumours went into a relapse or all tumours could be contained apart from exceptionally

small tumours. Both of these scenarios do not fit the experimental findings, where tumours both respond well to treatment and escape the therapy, with a tendency of better treatment success in small tumours.

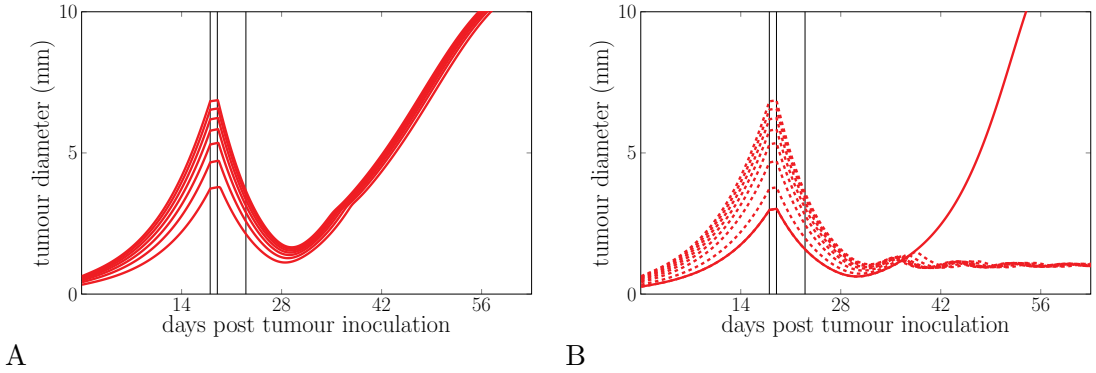


Figure 4.9: Growth curves of tumours undergoing ACT^{METi} therapy according to the old model, generated by simulations for different initial tumour sizes, shown as tumour diameter [mm]. T-cell efficiency either low (A) or high (B). Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections. Dashed lines indicate tumours undergoing eradication.

An explanation for the type of behaviour that we see in the simulations is the fact that, in the old model, T-cells and differentiated WT melanoma cells interact in a predator-prey like manner, where the T-cell proliferation and melanoma cell killing rates are proportional to the number of differentiated WT melanoma cells and T-cells, i.e. of the form $b_{CD}N_{Diff}N_{CD8}$ and $k_{Diff}N_{Diff}N_{CD8}$. As a result, T-cell activity increases with the size of the targeted cell population and large tumours trigger a more effective treatment. Independent of the initial conditions, the system converges to the same stable state balancing T-cells and melanoma cells. Depending on the parameters for T-cell efficiency, this stable state is either at a high level of melanoma cells, causing a tumour relapse, or a low level, resulting in containment of the tumour. This can be compared to the discussion of the simpler pure predator-prey system (1.7) in Section 1.2. Figure 1.1B shows the same fluctuating convergence towards an equilibrium state that depends on the rate at which predators kill their prey, which is the equivalent of T-cell efficiency. The relapse of very small tumours under high T-cell efficiency is caused by the extinction of T-cells due to random fluctuations, as shown in Figure 4.10.

To counteract this effect, and since T-cells cannot physically proliferate infinitely fast, we include aspects of negative feedback within the immune system into our new model. Inflammatory cytokines like INF-γ promote the up-regulation of negative immune checkpoint molecules (PD-L1) on melanoma cells, which inhibit the immune reaction. To model this, we introduce the factor $(1 - hN_{Cyto})_+$ into the rates for T-cell proliferation and melanoma cell killing, which decreases T-cell activity as the number of cytokines increases until eventually, at a cytokine level of $N_{Cyto} = 1/h$, the T-cells are shut down. Since the cytokines themselves are secreted at T-cell proliferation in our model, this constitutes a negative feedback loop within the immune system that limits the efficacy of treatment.

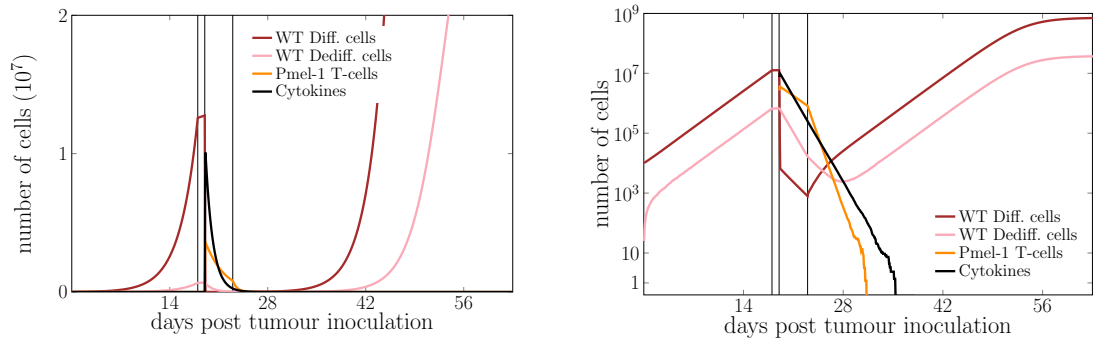
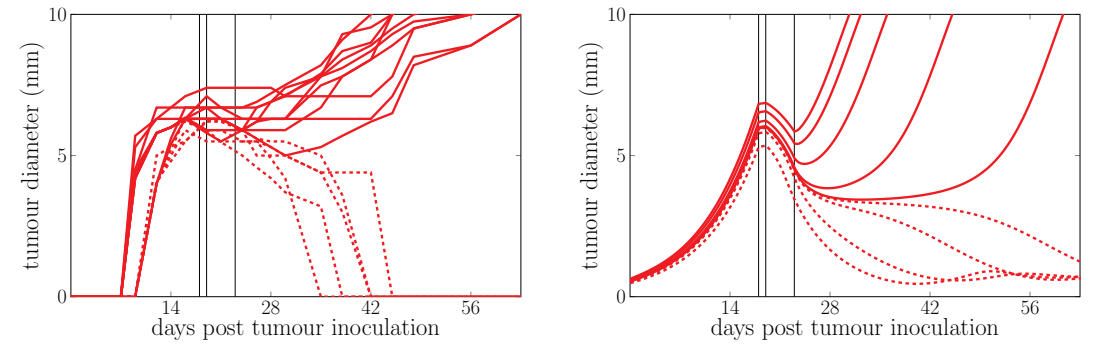


Figure 4.10: Simulation (and log-plot) of the evolution of different cell/molecule types for for tumours undergoing ACT^{METi}therapy in a very small tumour for high T-cell efficiency. Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections.

Simulations using this adjusted model recapitulated the tumour growth kinetics of HCmel12 WT melanomas treated with ACT^{METi} from [87] as shown in Figure 4.11.

When comparing the experimantal and simulated data, note that tumours with a diameter below 1 mm will likely not be detected in the experiments. Tumours that are marked as undergoing eradication in Figure 4.11 are more likely to exists at a small population size, contained by the T-cell population and undergoing predator-prey like oscillations. This is predicted by the mathematical model and has also been witnessed in experiments, see [152], where this state was described as a so-called *immune equilibrium*. While our simulations



A B

Figure 4.11: (A) Tumour growth curves from experimental data published in Glodde et al. [87]. (B) Tumour growth curves generated by simulations for different initial tumour sizes. Shown as tumour diameter [mm], vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections. Dashed lines indicate tumours undergoing eradication.

can recapitulate the size of the tumour at treatment onset and the approximate duration of tumour remission and eradication, it is of note that they also predicted a critical threshold for treatment success. Below a critical tumour size at treatment onset (roughly 6 mm), ACT^{METi} achieves long-term melanoma control or eradication. Above this threshold, the T-cells cannot contain the tumour and a remission occurs. This is in line with the experimental

data.

In the corresponding deterministic system, this threshold marks a critical point. Slight perturbations at the time of treatment onset amplify and lead to very different paths of tumour evolution. In our simulations, we can witness both courses of treatment for the same initial conditions, see Figure 4.12. This is a case of stochastic behaviour in the interior of the state space, where no subpopulation is small and in danger of dying out due to fluctuations. However, simulations also show that the fate of the melanoma cell population is decided early on, due to fluctuations in the beginning of tumour growth that determine whether the tumour size at treatment onset is sub- or super-critical.

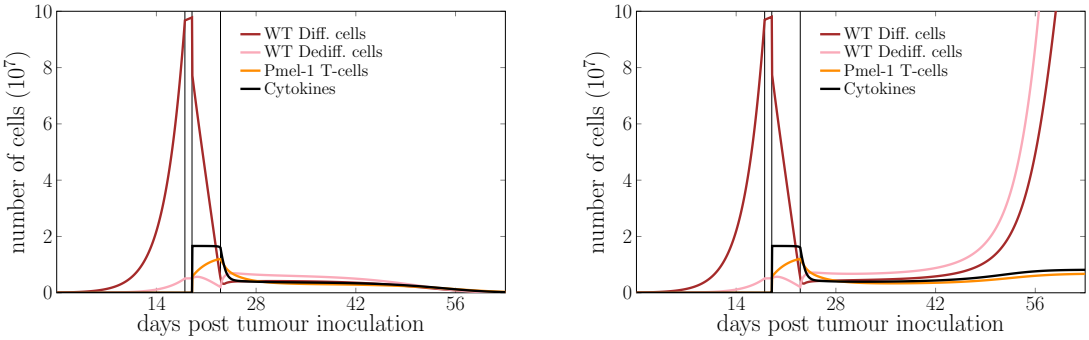


Figure 4.12: Simulations of the evolution of different cell/molecule types for tumours undergoing ACT^{METi}therapy. Both trajectories are generated with the same initial conditions that produce tumours near the critical threshold for size at treatment onset. Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections.

Overall, our mathematical modelling approach for pure WT melanomas was able to identify T-cell inhibition as a key mechanism in the immune response and predict a critical threshold for therapy success.

4.3.2 Competitive pressure and shielding effect

Next, we simulate the dynamics of pre-existing Pmel^{KO} melanoma cells within a bulk WT melanoma cell population, because we seek to explain the different enrichment of Pmel^{KO} cells found in untreated or ACT^{METi}-recurrent $\text{HCmel12}^{\text{pcMix}}$ melanomas from the new experiments (see Figure 4.5B). In the mathematical model, we distinguish two notions of fitness, similar to the concepts in our theoretical work in the previous chapters. The fixed *individual fitness* r of a certain cell type is composed of its cell division rate b and natural death rate d and describes its growth rate in a competition-free environment. The *context-dependent fitness* $f(N)$ describes the growth rate of a cell type taking into account the competition c between individuals in a population of state N. For a single mutant individual within a bulk population at equilibrium, the context-dependent fitness is equal to its so-called *invasion fitness* (first introduced in [139] under a different name). For example, the fixed fitness of Pmel^{KO} cells is $r_{KO} = b_{KO} - d_{KO}$ while their context-dependent fitness in a population of

state N is

$$f_{KO}(N) = b_{KO} - d_{KO} - c_{KO,Di}N_{Diff} - c_{KO,De}N_{Dedi} - c_{KO,KO}N_{KO}. \tag{4.12}$$

Since the overall growth rate is very similar for untreated pure WT or Pmel$^{\mathrm{KO}}$ melanomas (Figure 4.6), their individual fitness and self-competition parameters are also chosen as similar. The reduced (context-dependent) fitness of Pmel$^{\mathrm{KO}}$ cells in HCmel12$^{\mathrm{pcMix}}$ tumours is modelled via a high competitive pressure $c_{KO,Di}$ and $c_{KO,De}$ that WT melanoma cells impose on Pmel$^{\mathrm{KO}}$ melanoma cells. Figure 4.13 shows how the growth of the unfit Pmel$^{\mathrm{KO}}$ melanoma cell population slows down until a critical number of WT melanoma cells is reached and its overall growth rate (context-dependent fitness) becomes negative. The number of Pmel$^{\mathrm{KO}}$ melanoma cells then decreases and would eventually reach zero. In the experiments, however, mice needed to be sacrificed before this would happen.

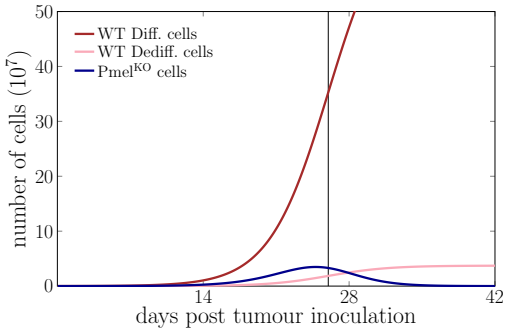


Figure 4.13: Simulation of the evolution of different cell/molecule types for the untreated case (number of cells in 10^7). The vertical line marks the time that a 10 mm tumour diameter is reached.

If Pmel-1 T-cell killing of WT$^{\mathrm{Diff}}$ melanoma cells was not influenced by the presence of Pmel$^{\mathrm{KO}}$ melanoma cells, our simulations predict that the enrichment of Pmel$^{\mathrm{KO}}$ cells in ACT$^{\mathrm{METi}}$-recurrent melanomas would be a lot higher than what was measured experimentally, making up nearly the entire melanoma cell population (see Figure 4.14).

However, if the portion of Pmel$^{\mathrm{KO}}$ melanoma cells is high, one may assume that differentiated WT melanoma cells will be 'shielded' from Pmel-1 T-cell recognition and killing due to physical obstruction, an immunosuppressive tumour microenvironment, and other mechanisms. Our model is per se non-spatial and assumes a well-mixed population. We therefore represent this *shielding effect* by introducing the percentage of differentiated WT melanoma cells (out of all tumour cells) as a factor into the killing rate of Pmel-1 T-cells and also their proliferation rate since Pmel-1 T-cell activation depends on Pmel antigen presentation. The influence of this factor is determined by its exponent α, which we termed *shielding parameter*. The killing rate thus takes the form

$$R_e(N) = k_{Di}N_{Diff}N_{CD8}\left(\frac{N_{Diff}}{N_{Diff} + N_{Dedi} + N_{KO}}\right)^{\alpha}(1 - hN_{Cyto})_+. \tag{4.13}$$

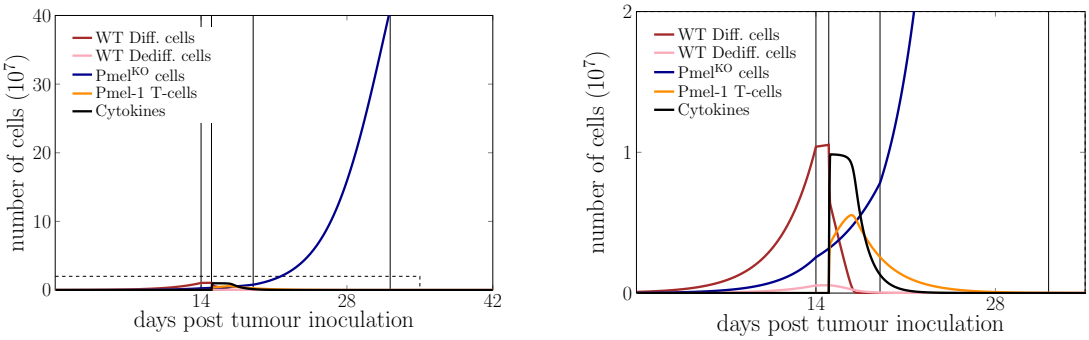


Figure 4.14: Simulation (and zoom-in) of the evolution of different cell/molecule types under ACT^{METi}, not taking into account a shielding effect (number of cells in 10^7). Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, end of METi injections, and the time that a 10 mm tumour diameter is reached.

The T-cell proliferation rate has the same structure. A higher value of α causes a larger shielding effect through dedifferentiated WT and Pmel^{KO} melanoma cells. Details on the determination of the parameter α are given in the previous section.

Figure 4.15 shows how, under ACT^{METi}, the Pmel^{KO} cell population grows rapidly, as soon as the number of WT cells drops low enough to no longer impose a large competitive pressure (right zoom-in panel). As a result, the few remaining differentiated WT cells are protected by the abundant Pmel^{KO} cells, which allows for the recovery of the WT population. As soon as WT cells surpass the critical level for competition, the number of Pmel^{KO} cells decreases again.

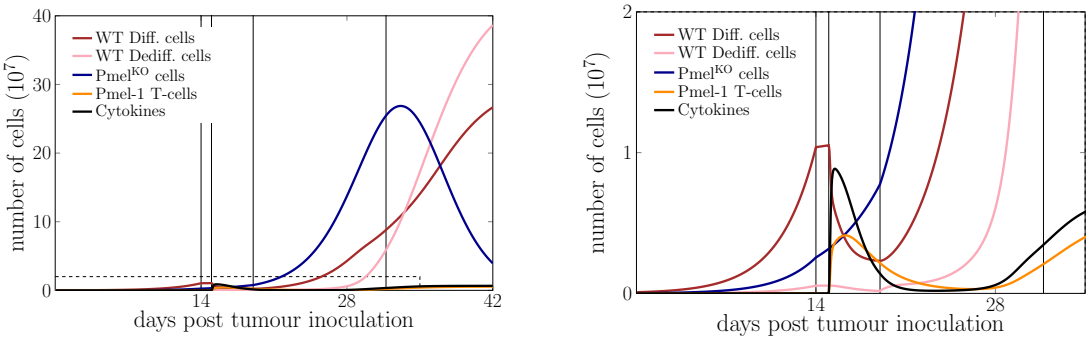


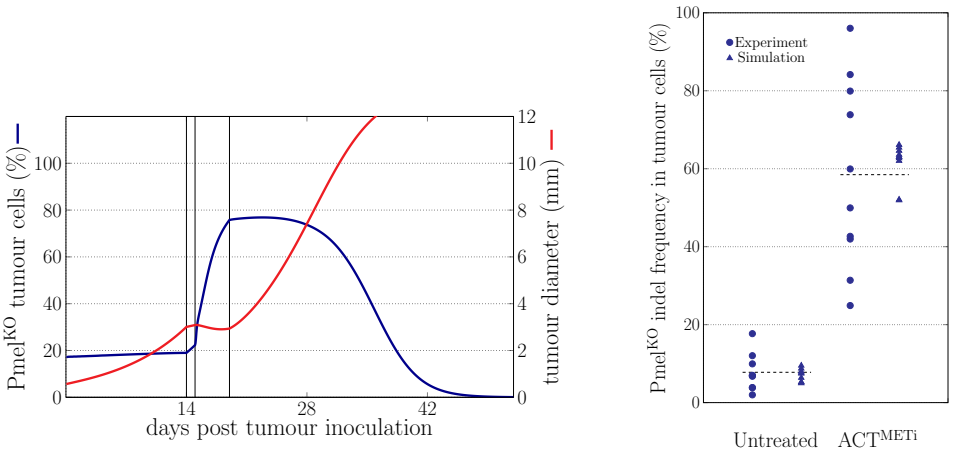
Figure 4.15: Simulation (and zoom-in) of the evolution of different cell/molecule types under ACT^{METi} (number of cells in 10^7). Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, end of METi injections, and the time that a 10 mm tumour diameter is reached.

To sum up the results of this subsection, we model the reduced fitness of Pmel^{KO} cells in the untreated case through a high competitive pressure that is imposed by the WT cell population. We identify the shielding of differentiated WT melanoma cells from Pmel-1 T-cells through other melanoma cells as an important mechanism to prevent a full invasion

by Pmel^{KO} cells under ACT^{METi} therapy, which is not witnessed in experiments.

4.3.3 Variable enrichment of KO through sequencing time and subclone fitness variability

In this subsection we try to explain the highly variable immune selection of Pmel^{KO} cells in ACT^{METi}-recurrent HCmel12^{pcMix} melanomas. Figure 4.8A has shown that the tumour size at therapy onset surprisingly has very little influence on the Pmel^{KO} cell percentage, even when lying above and below the critical threshold for tumour eradication. However, Figure 4.16A shows that the enrichment of Pmel^{KO} cells is highly dependent on the time point of tumour tissue harvesting and sequencing analysis. Temporarily, those cells make up a large portion of the tumour with up to 80%, before the wild type cells recover and the percentage of Pmel^{KO} cells drops down to almost 0%. Due to these high fluctuations, part of the variation in the measured percentage of Pmel^{KO} cells can be explained.



A B

Figure 4.16: (A) Simulation of the evolution of tumour size, shown as diameter [mm], and percentage of Pmel^{KO} cells. Initial tumour of medium size and $\alpha = 4$. Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections. (B) Comparison of experimental data from Figure 4.5B and simulation results for the frequency of Pmel^{KO} alleles in ACT^{METi}-recurrent HCmel12^{pcMix} melanomas. Percentage determined at a random time ±0.5 mm of the measured diameter in experiments.

Figure 4.16B compares the percentage of Pmel^{KO} cells in harvested tumours between experiments and simulations, with and without ACT^{METi}. The parameters of the mathematical model were chosen such that they match the mean percentage of 7.8% Pmel^{KO} cells in the untreated case and 58.5% under ACT^{METi} (from Figure 4.5B). In the simulations, the time point of sequencing (harvesting) was varied by picking a random point close to the documented diameter from experiments (±0.5 mm). This is tenable since the diameter of the tumour is only measured twice a week and will therefore never be exactly at 10 mm. Nevertheless, especially in the case of ACT^{METi}, we still witness a much higher variation in the experimental data compared to the simulations.

Tumour cell heterogeneity is another possible cause of variation. Experimentally, it is plausible since, in the polyclonal approach, each of the generated HCmel12 Pmel$^{\mathrm{KO}}$ subclones may display a slightly different fitness due to pre-existing genetic or epigenetic heterogeneity or even CRISPR-Cas9 off-target effects, amongst others (for more details see the first section of this chapter). Therefore, we run simulations for varying (individual) fitness r_{KO} of the Pmel$^{\mathrm{KO}}$ cells by changing the birth rate b_{KO} or the natural death rate d_{KO}. Figure 4.17 displays the simulated percentage of Pmel$^{\mathrm{KO}}$ cells in untreated and ACT$^{\mathrm{METi}}$-recurrent melanomas, determined at tumour sizes between 9 and 11 mm, plotted against their individual fitness r_{KO}.

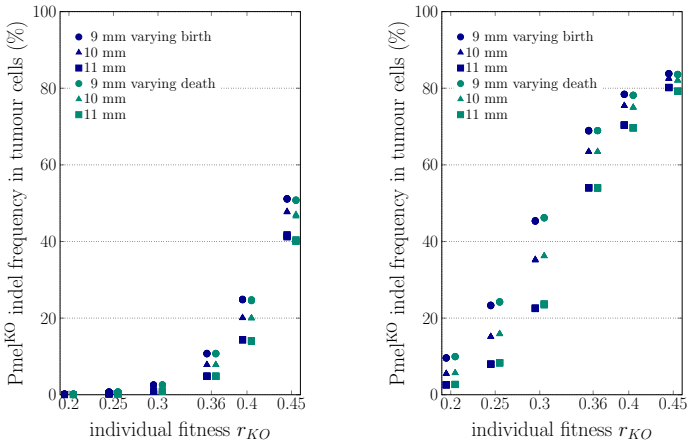


Figure 4.17: Predictions for enrichment of Pmel$^{\mathrm{KO}}$ cells, under varying subclone fitness r_{KO} and different tumour sizes at time point of harvesting, for untreated (A) and ACT$^{\mathrm{METi}}$-recurrent (B) HCmel12$^{\mathrm{pcMix}}$ melanomas.

The results show no major difference between the two approaches of varying b_{KO} and d_{KO}. They do, however, account for most of the variation seen in the experimental data as, especially in the case of ACT$^{\mathrm{METi}}$ treatment, the percentage of Pmel$^{\mathrm{KO}}$ cells at the time point of sequencing (harvesting) largely increases with increasing fitness r_{KO}.

To experimentally confirm this prediction of subclone fitness variability, the Pmel$^{\mathrm{indel}}$ distributions from ACT$^{\mathrm{METi}}$-recurrent HCmel12$^{\mathrm{pcMix}}$ melanomas were analysed. They showed a predominance of only few frame-shift indels, in contrast to the widespread input distribution at tumour cell inoculation, which strongly supports a scenario where selection of a few Pmel$^{\mathrm{KO}}$ subclones with superior fitness has occurred (for more details see [88]).

To further substantiate this finding, experiments with HCmel12 Pmel$^{\mathrm{KO}}$ single cell subclones were conducted, asking whether they exhibited differences in fitness or not. For this purpose, three Pmel$^{\mathrm{KO}}$ subclones were mixed with WT cells at different proportions (HCmel12$^{\mathrm{sccMix}}$), injected into syngeneic mice, and treated with ACT$^{\mathrm{METi}}$ (following the protocol in Figure 4.1) or left untreated. The frequencies of the three Pmel$^{\mathrm{KO}}$ subclones were determined by NGS and dual-color imaging and showed that, in ACT$^{\mathrm{METi}}$-recurrent melanomas, clone #1 was predominantly enriched in all cases, whereas enrichment of the other subclones occurred

only when the input frequency was increased (see Figure 4.18 and [88] for details). These results confirm the prediction from our simulations that subclones substantially differ in fitness.

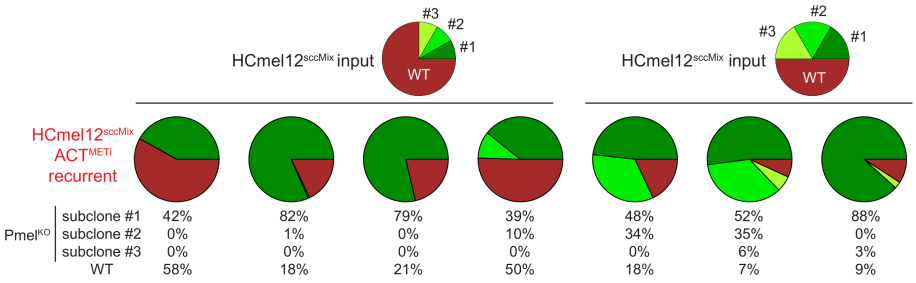


Figure 4.18: Percentages of individual Pmel^{KO} single cell clones (subclones) in ACT^{METi}-recurrent $\text{HCmel12}^{\text{sccMix}}$ tumours determined by amplicon NGS.

Overall, in contrast to the tumour size at therapy onset, the time point of sequencing (or rather the tumour size at this time) and varying subclone fitness are likely explanations for the highly variable enrichment of Pmel^{KO} cells found in ACT^{METi}-recurrent melanomas.

4.3.4 Validation of clinical relevance through study of spontaneous mutations

So far, we only addressed scenarios were a substantial amount of pre-existing Pmel^{KO} cells was mixed with WT melanoma cells before tumour inoculation. However, in a clinical scenario, a mutation would occur spontaneously and start out with a single cell. In order to determine whether a single Pmel^{KO} cell can fixate although it is unfit compared to the bulk WT cell population, we introduced the possibility of spontaneous mutations from WT cells to Pmel^{KO} cells into our stochastic model.

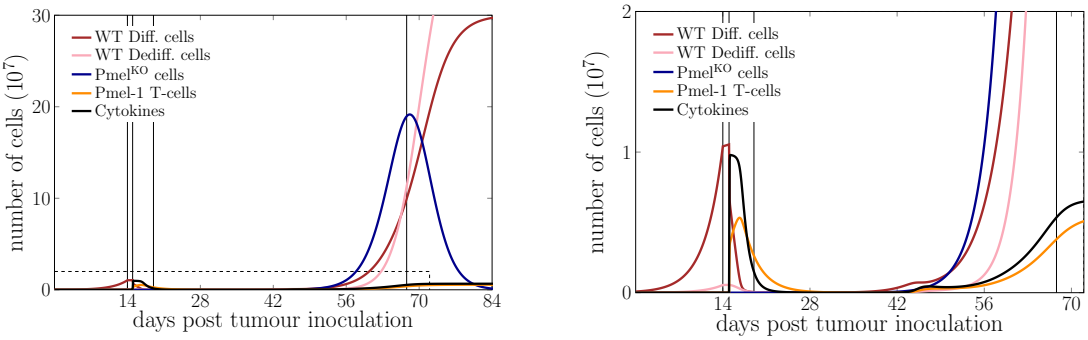


Figure 4.19: Simulation (and zoom-in) of the evolution of different cell/molecule types under ACT^{METi}, shown as number of cells in 10^7. Initial pure WT tumour size below critical threshold for therapy success. Natural mutation to Pmel^{KO} cells at rate of $m = 10^{-7}$. Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, end of METi injections, and the time that a 10 mm tumour diameter is reached.

With 1.5 mutations per day on average in a tumour of 3 mm in diameter (typical melanoma size at treatment onset in our experiments), we have chosen a relatively low frequency of mutational events to obtain Figure 4.19. For this choice of parameters, in more than half of the simulation runs the Pmel^{KO} cells fixate and cause a relapse within the first 100 days after tumour inoculation. Even when further decreasing the probability of mutation, this still happens, but in less cases and at a later time points.

Whenever the Pmel^{KO} cells fixate, i.e. surpass a detectable number of cells, the same effects as in the experimental setup can be witnessed. For a smaller pure WT tumour, below the critical threshold for tumour eradication, the Pmel^{KO} cell population grows and thus protects the WT cells from dying out. The latter can then recover and eventually expand within the Pmel^{KO} tumour. Compared to the situations with a pre-existing portion of Pmel^{KO} cells, this happens much later, because spontaneously occurring Pmel^{KO} cells start to grow from a much lower number. However, the relapse phase itself takes a very similar course (Figure 4.19, right zoom-in panel).

Figure 4.20 shows the results of a number of different simulation runs, where the Pmel^{KO} cells fixate at different times to cause a relapse. The occurrence of the mutation that caused the relapse is marked with a cross. This variability is due to the stochasticity of our model where mutations occur randomly at different time points and Pmel^{KO} mutants may die out before they fixate and hence the first mutation may not always be successful.

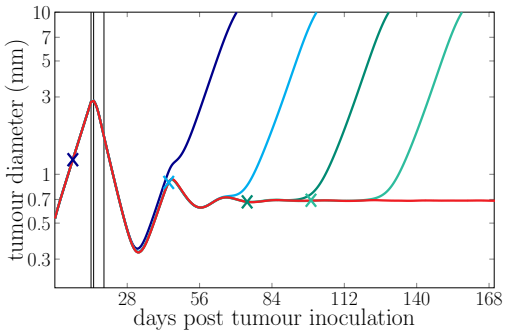


Figure 4.20: Simulations of tumour growth curves, shown as diameter [mm] on a log-scale. Red curve shows no successful mutation while first successful mutations of blue and green curves are marked with crosses. Vertical lines mark beginning of METi injections, injection of Pmel-1 T-cells, and end of METi injections.

As above, in Figure 4.17, we again varied the individual fitness r_{KO} of the Pmel^{KO} cells. For the percentage of Pmel^{KO} cells in ACT^{METi}-recurrent melanoma at 9 to 11 mm diameter we obtain a similar picture (Figure 4.21A). We see slightly more variability between simulation runs, particularly for high fitness, and on average lower percentages for the intermediate fitness values. However, the overall range between the highest and lowest values for r_{KO} remains the same and there is no major difference between the variation of b_{KO} and d_{KO}.

As mentioned before, not every simulation run shows a successful mutation where Pmel^{KO} cells fixate in the population and and cause a relapse. We have analysed the time points of occurring relapses (as the time when the tumour reaches a diameter of 10 mm) and

summarised the results in Figure 4.21B. With increasing fitness r_{KO}, the Pmel$^{\mathrm{KO}}$ population can grow faster and thus causes an earlier relapse. As already pointed out in Figure 4.20, mutations (and thus relapses) can arise at different time points. However, Figure 4.21B shows that the bulk of the relapses occurs at an early time point and can therefore be traced back to a mutation event during the first growth phase of the tumour, before the treatment was initiated.

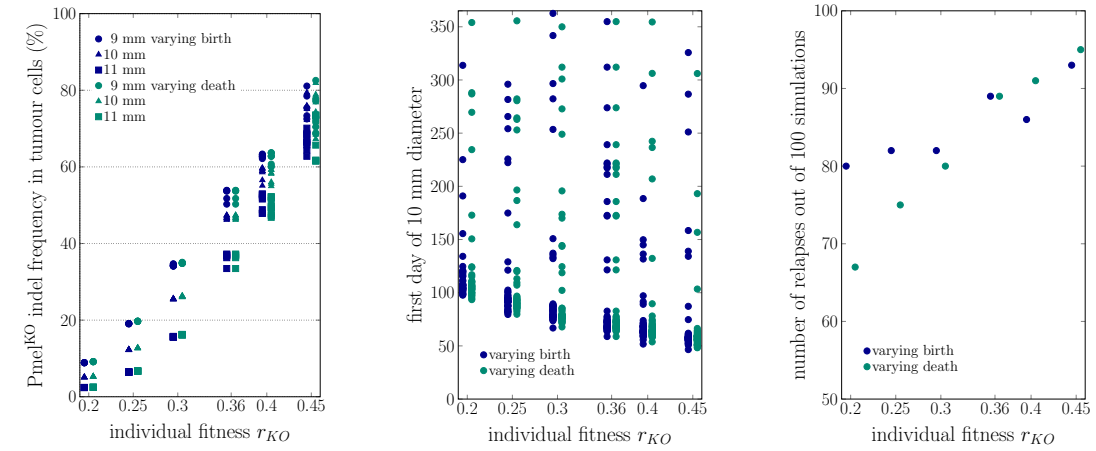


Figure 4.21: Analysis of ACT$^{\mathrm{METi}}$ melanomas for sub-critical initial wild type tumour and natural mutation at rate $m = 10^{-7}$ under varying subclone fitness $r_{KO} = b_{KO} - d_{KO}$. (A) Prediction of enrichment of Pmel$^{\mathrm{KO}}$ cells for different tumour sizes at time point of harvesting. (B) Time points when 10 mm diameter are reached (within one year after tumour inoculation). (C) Number of successful relapses (within one year after tumour inoculation) out of 100 simulation runs.

In addition to the time point of a relapse and the percentage of Pmel$^{\mathrm{KO}}$ cells in ACT$^{\mathrm{METi}}$-recurrent melanoma, we also studied the likelihood of those relapses. Figure 4.21C shows the number of simulation runs (out of 100) that exhibit a relapse within one year after tumour inoculation. Besides the expectable higher number of relapses for higher fitness r_{KO}, we also observe different behaviour between the variation of b_{KO} and d_{KO}. In the cases where the fitness decrease (compared to $r_{KO} = 0.36$) is obtained by an increased death rate d_{KO} or the fitness increase is due to a higher birth rate b_{KO}, we detect fewer relapses than in the other cases. This is due to the fact that we have higher rates b_{KO} and d_{KO} (while keeping r_{KO} the same), which causes more birth and death events in the same time interval and thus higher fluctuations in the Pmel$^{\mathrm{KO}}$ population. This makes it more likely for the mutant to die out before fixation, i.e. before it reaches a population size at which it cannot go extinct due to random fluctuations, and thus relapses become less likely.

Mathematically this is in line with theoretical results calculating the probability of fixation, within a population at state N, to be proportional to $[f_{KO}(N)]_+/b_{KO}$ (see [31]) and thus to decrease with increasing rate b_{KO} (for constant r_{KO} and hence $f_{KO}(N)$). Once a certain population size is reached, the fluctuations have less influence and the Pmel$^{\mathrm{KO}}$ cells grow according to their average dynamics, hence their enrichment at relapse is less sensitive to the different approaches.

In conclusion, simulations with spontaneously occurring antigen loss mutations have shown the clinical relevance of the conducted experiments since similar phenomena can be witnessed during treatment. In this scenario, random events play an important role as they decide if and when a relapse of the tumour may occur. Possible implications of these stochastic fluctuations for treatment strategies are discussed in the next section.

4.4 Discussion

In this section we discuss the clinical relevance and implications of our results.

The experiments that are considered in this chapter have all been conducted in mouse models. Nevertheless, we argue the relevance of our results. The epitope derived from murine Pmel protein is a low affinity epitope and rather poorly recognised by Pmel-1 T-cells, in contrast to the corresponding high affinity epitope derived from human Pmel [68, 151]. Thus, the complex dynamics of antigen down-regulation and ablation under ACT therapy and the facilitation of tumour immune escape are of even greater importance in the context of human high affinity epitopes.

Our simulation results emphasise the importance of an early tumour discovery and therapy initiation. Due to negative feedback within the immune system, there is a threshold for tumour size at treatment onset, above which successful eradication through $\mathrm{ACT}^{\mathrm{METi}}$ therapy or at least control of the melanoma cells in an immune equilibrium is impossible. However, in those scenarios, it could be promising to first remove a portion of the tumour by surgery and afterwards treat the remaining cells with immunotherapy.

Moreover, our results underscore the importance of tumour heterogeneity on tumour immune surveillance of melanomas, as also demonstrated by recent studies analyzing melanoma patient samples or using UVB-induced mouse melanomas as a models system [133, 176]. We believe that our findings have important implications for the analysis of patient samples, because the genomic comparison of pre- and post-treatment tumour specimens, untreated and recurrent melanomas in our experimental setting, is a standard approach to identify genetic changes in tumour cells that cause resistance to immunotherapy [179]. A highly variable enrichment of resistant tumour cell variants limits their detection likelihood, because the chance of being identified as a recurrent event decreases. We therefore postulate that many resistance mechanisms to immunotherapy remain to be discovered, in particular those genetic events that are associated with a reduced tumour cell fitness prior to but an increased fitness upon treatment. Hence, we envision that implementing mathematical models for evolution into genomic analysis pipelines could help to identify such resistance mechanisms more reliably.

Finally, our simulations of spontaneous antigen loss mutations further emphasised the importance of stochastic events, in particular in small tumour cell populations. We have shown that stochastic fluctuations can cause extinction of spontaneously occurring antigen loss variants even though their context-dependent fitness might be positive (as it is in smaller WT tumours). This observation is applicable to any small cell population, as for example tumours in remission. It is obvious to target tumours with treatments that impair the overall

growth of the cell population. We argue that it is most promising to do so with approaches that actively induce cell death (rather than impair cell division) since this increases fluctuations. In other words, treatments that enforce tumour cell death such as Bcl-2 family antagonists [3] are predicted to efficiently eliminate residual, possibly resistant, tumour cells and thus prevent melanoma recurrence. This scenario reminds of a recent study, where tissue-resident memory CD8 T-cells were shown to achieve long-term immune surveillance of residual melanoma cells in the skin of mice [152]. Experimental models like this seem to be suitable to confirm the prediction regarding cell death induction and may eventually contribute to improved therapeutic strategies that prevent tumour recurrences after successful immunotherapy.

Bibliography

[1] P. M. Altrock, L. L. Liu, and F. Michor. The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer*, 15(12):730–745, 2015.

[2] I. Arozarena and C. Wellbrock. Phenotype plasticity as enabler of melanoma progression and therapy resistance. *Nat. Rev. Cancer*, 19(7):377–391, 2019.

[3] A. Ashkenazi, W. J. Fairbrother, J. D. Leverson, and A. J. Souers. From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors. *Nat. Rev. Drug Discov.*, 16(4):273, 2017.

[4] K. B. Athreya. Some results on multitype continuous time Markov branching processes. *Ann. Math. Stat.*, 39:347–357, 1968.

[5] K. B. Athreya and P. E. Ney. *Branching processes*. Die Grundlehren der mathematischen Wissenschaften, Vol. 196. Springer-Verlag, New York-Heidelberg, 1972.

[6] F. J. Ayala. Evolution. *Encyclopædia Britannica*, 2019. Available at `https://www.britannica.com/science/evolution-scientific-theory`, accessed: 06.04.2020.

[7] E. Baake and M. Baake. Ancestral lines under recombination. *arXiv:2002.08658v1; to appear in: Probabilistic Structures in Evolution, E. Baake and A. Wakolbinger (eds.), EMS Publishing House, Zurich*, 2020.

[8] E. Baake and A. Wakolbinger. Lines of descent under selection. *J. Stat. Phys.*, 172(1):156–174, 2018.

[9] M. Baar and A. Bovier. The polymorphic evolution sequence for populations with phenotypic plasticity. *Electron. J. Probab.*, 23(72):1–27, 2018.

[10] M. Baar, A. Bovier, and N. Champagnat. From stochastic, individual-based models to the canonical equation of adaptive dynamics in one step. *Ann. Appl. Probab.*, 27(2):1093–1170, 2017.

[11] M. Baar, L. Coquille, H. Mayer, M. Hölzel, M. Rogava, T. Tüting, and A. Bovier. A stochastic model for immunotherapy of cancer. *Sci. Rep.*, 6:24169, 2016.

[12] N. H. Barton and J. Polechová. The limitations of adaptive dynamics as a model of evolution. *J. Evol. Biol.*, 18(5):1186–1190, 2005.

[13] J. Berestycki, E. Brunet, and Z. Shi. The number of accessible paths in the hypercube. *Bernoulli*, 22(2):653–680, 2016.

[14] J. Berestycki, E. Brunet, and Z. Shi. Accessibility percolation with backsteps. *ALEA Lat. Am. J. Probab. Math. Stat.*, 14(1):45–62, 2017.

[15] S. Billiard and C. Smadi. The interplay of two mutations in a population of varying size: A stochastic eco-evolutionary model for clonal interference. *Stoch. Process. Appl.*, 127(3):701–748, 2017.

[16] S. Billiard and C. Smadi. Stochastic dynamics of three competing clones: Conditions and times for invasion, coexistence, and fixation. *Am. Nat.*, 0(0):000–000, 2020.

[17] C. U. Blank, W. N. Haining, W. Held, P. G. Hogan, A. Kallies, E. Lugli, R. C. Lynn, M. Philip, A. Rao, N. P. Restifo, et al. Defining 'T cell exhaustion'. *Nat. Rev. Immunol.*, 19(11):665–674, 2019.

[18] J. Blath and M. Ortgiese. The symbiotic branching model: duality and interfaces. *arXiv:2002.02825v1; to appear in: Probabilistic Structures in Evolution, E. Baake and A. Wakolbinger (eds.), EMS Publishing House, Zurich*, 2020.

[19] J. Blath and A. Tóbiás. Invasion and fixation of microbial dormancy traits under competitive pressure. *arXiv:1910.13156v2*, 2019.

[20] 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.

[21] 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.

[22] A. Bovier. Stochastic individual based models: From adaptive dynamics to modelling of cancer therapies. Lecture Notes, 2019.

[23] A. Bovier. Stochastic models for adaptive dynamics: Scaling limits and diversity. *arXiv:1909.02456v1; to appear in: Probabilistic Structures in Evolution, E. Baake and A. Wakolbinger (eds.), EMS Publishing House, Zurich*, 2019.

[24] A. Bovier, L. Coquille, and R. Neukirch. The recovery of a recessive allele in a Mendelian diploid model. *J. Math. Biol.*, 77(4):971–1033, 2018.

[25] A. Bovier, L. Coquille, and C. Smadi. Crossing a fitness valley as a metastable transition in a stochastic population model. *Ann. Appl. Probab.*, 29(6):3541–3589, 2019.

[26] A. Bovier and S.-D. Wang. Trait substitution trees on two time scales analysis. *Markov Process. Relat. Fields*, 19(4):607–642, 2013.

[27] J. C. Butcher. Coefficients for the study of Runge-Kutta integration processes. *J. Aust. Math. Soc. A*, 3:185–201, 1963.

[28] C. Cannings. The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models. *Adv. Appl. Probab.*, 6:260–290, 1974.

[29] C. Cannings. The latent roots of certain Markov chains arising in genetics: A new approach. II. Further haploid models. *Adv. Appl. Probab.*, 7:264–282, 1975.

[30] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.*, 124(4):044109, 2006.

[31] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Process. Appl.*, 116(8):1127–1160, 2006.

[32] N. Champagnat, R. Ferrière, and G. Ben Arous. The canonical equation of adaptive dynamics: A mathematical view. *Selection*, 2(1-2):73–83, 2001.

[33] N. Champagnat, R. Ferrière, and S. Méléard. From individual stochastic processes to macroscopic models in adaptive evolution. *Stoch. Models*, 24(suppl. 1):2–44, 2008.

[34] 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.

[35] N. Champagnat, P.-E. Jabin, and G. Raoul. Convergence to equilibrium in competitive Lotka-Volterra and chemostat systems. *C. R. Math. Acad. Sci. Paris*, 348(23-24):1267–1272, 2010.

[36] N. Champagnat and S. Méléard. Invasion and adaptive evolution for individual-based spatially structured populations. *J. Math. Biol.*, 55(2):147–188, 2007.

[37] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probab. Theory Relat. Fields*, 151(1-2):45–94, 2011.

[38] N. Champagnat, S. Méléard, and V. C. Tran. Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer. *arXiv:1901.02385v2*, 2019.

[39] S. Chen, G. A. Crabill, T. S. Pritchard, T. L. McMiller, P. Wei, D. M. Pardoll, F. Pan, and S. L. Topalian. Mechanisms regulating PD-L1 expression on tumor and immune cells. *J. Immunotherapy Cancer*, 7(1):1–12, 2019.

[40] T. Chodon, B. Comin-Anduix, B. Chmielowski, R. C. Koya, Z. Wu, M. Auerbach, C. Ng, E. Avramis, E. Seja, A. Villanueva, et al. Adoptive transfer of MART-1 T-cell receptor transgenic lymphocytes and dendritic cell vaccination in patients with metastatic melanoma. *Clin. Cancer Res.*, 20(9):2457–2465, 2014.

[41] 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.

[42] L. Coquille, A. Kraut, and C. Smadi. Stochastic individual-based models with power law mutation rate on a general finite trait space. *arXiv:2003.03452v1*, 2020.

[43] C. Coron, M. Costa, H. Leman, and C. Smadi. A stochastic model for speciation by mating preferences. *J. Math. Biol.*, 76(6):1421–1463, 2018.

[44] A. L. Correia and M. J. Bissell. The tumor microenvironment is a dominant force in multidrug resistance. *Drug Resist. Updat.*, 15(1-2):39–49, 2012.

[45] 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.

[46] J. Couzin-Frankel. Cancer immunotherapy. *Science*, 342(6165):1432–1433, 2013.

[47] J. Coville and F. Fabre. Convergence to the equilibrium in a Lotka-Volterra ODE competition system with mutations. *arXiv:1301.6237v2*, 2013.

[48] M. C. Cowperthwaite, J. J. Bull, and L. A. Meyers. From bad to good: Fitness reversals and the ascent of deleterious mutations. *PLoS Comput. Biol.*, 2(10):e141, 2006.

[49] I. Cvijović, B. H. Good, E. R. Jerison, and M. M. Desai. Fate of a mutation in a fluctuating environment. *P. Natl. Acad. Sci. USA*, 112(36):E5021–E5028, 2015.

[50] J. M. J. da Costa, H. R. B. Orlande, and W. B. da Silva. Model selection and parameter estimation in tumor growth models using approximate Bayesian computation-ABC. *Comput. Appl. Math.*, 37(3):2795–2815, 2018.

[51] 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.

[52] 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.

[53] E. Darwin. *Zoonomia, Or, The Laws of Organic Life.* J. Johnson in St Paul's Church-Yard, London, 1794.

[54] E. Darwin. *The Temple of Nature; or, the Origin of Society, a Poem, with Philosophical Notes.* John W. Butler, and Bonsal & Niles, Baltimore, 1804.

[55] J.-B. P. A. d. M. de Lamarck. *Philosophie zoologique.* Dentu, Muséum d'Histoire Naturelle, 1809.

[56] J. A. G. De Visser and J. Krug. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.*, 15(7):480–490, 2014.

[57] M. A. DePristo, D. L. Hartl, and D. M. Weinreich. Mutational reversions during adaptive protein evolution. *Mol. Biol. Evol.*, 24(8):1608–1610, 2007.

[58] M. Diabaté, L. Coquille, and A. Samson. Parameter estimation and treatment optimization in a stochastic model for immunotherapy of cancer. *arXiv:806.01915v3*, 2018.

[59] 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.

[60] U. Dieckmann and R. Law. Moment approximations of individual-based models. In *The geometry of ecological interactions: simplifying spatial complexity*, pages 252–270. Camb. Univ. Press, 2000.

[61] E. D'Ippolito, K. Schober, M. Nauerth, and D. H. Busch. T cell engineering for adoptive T cell therapy: safety and receptor avidity. *Cancer Immunol. Immun.*, 68(10):1701–1712, 2019.

[62] A. Duncan, R. Erban, and K. Zygalakis. Hybrid framework for the simulation of stochastic chemical kinetics. *J. Comp. Phys.*, 326:398–419, 2016.

[63] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley Ser. in Probab. and Math. Stat. John Wiley & Sons, Inc., New York, 1997.

[64] R. Durrett, J. Foo, and K. Leder. Spatial moran models, II: cancer initiation in spatially structured tissue. *J. Math. Biol.*, 72(5):1369–1400, 2016.

[65] R. Durrett and J. Mayberry. Traveling waves of selective sweeps. *Ann. Appl. Probab.*, 21(2):699–744, 2011.

[66] K. Echelmeyer. *Kombinierte deterministisch-stochastische Simulationen eines Populationsmodelles mit Anwendung in der Krebsimmuntherapie.* Bachelor thesis, Bonn University, 2017.

[67] R. Eftimie, J. L. Bramson, and D. J. D. Earn. Interactions between the immune system and cancer: a brief review of non-spatial mathematical models. *Bull. Math. Biol.*, 73(1):2–32, 2011.

[68] B. Engels, V. H. Engelhard, J. Sidney, A. Sette, D. C. Binder, R. B. Liu, D. M. Kranz, S. C. Meredith, D. A. Rowley, and H. Schreiber. Relapse or eradication of cancer is predicted by peptide-major histocompatibility complex affinity. *Cancer cell*, 23(4):516–526, 2013.

[69] A. Etheridge. Some mathematical models from population genetics, volume 2012 of lecture notes in mathematics, 2011.

[70] S. N. Ethier and T. G. Kurtz. *Markov processes.* Wiley Ser. in Probab. and Math. Stat. John Wiley & Sons, Inc., New York, 1986.

[71] S. N. Ethier and M. F. Norman. Error estimate for the diffusion approximation of the Wright–Fisher model. *Proc. Natl. Acad. Sci. USA*, 74(11):5096–5098, 1977.

[72] W. J. Ewens. *Mathematical Population Genetics: I. Theoretical Introduction.* Interdisciplinary Applied Mathematics, Vol. 27. Springer, New York, 2004.

[73] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *T. Roy. Soc. Edin.*, 52(2):399–433, 1918.

[74] R. A. Fisher. The wave of advance of advantageous genes. *Ann. Eugenic.*, 7(4):355–369, 1937.

[75] W. H. Fleming and M. Viot. Some measure-valued Markov processes in population genetics theory. *Indiana U. Math. J.*, 28(5):817–843, 1979.

[76] J. Foo, K. Leder, and J. Schweinsberg. Mutation timing in a spatial model of evolution. *arXiv:2001.01175v1*, 2020.

[77] 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.

[78] R. E. Franklin and R. G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.

[79] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems.* Grundlehren der Mathematischen Wissenschaften, Vol.260. Springer, Heidelberg, third edition, 2012.

[80] F. Fröhlich, T. Kessler, D. Weindl, A. Shadrin, L. Schmiester, H. Hache, A. Muradyan, M. Schütte, J.-H. Lim, M. Heinig, F. J. Theis, H. Lehrach, C. Wierling, B. Lange, and J. Hasenauer. Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model. *Cell Syst.*, 7(6):567–579, 2018.

[81] S. A. Geritz, J. A. Metz, É. Kisdi, and G. Meszéna. Dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.*, 78(10):2024, 1997.

[82] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, 104(9):1876–1889, 2000.

[83] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4):403–434, 1976.

[84] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115(4):1716–1733, 2001.

[85] J. H. Gillespie. Molecular evolution over the mutational landscape. *Evolution*, 38(5):1116–1129, 1984.

[86] 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.

[87] N. Glodde, T. Bald, D. van den Boorn-Konijnenberg, K. Nakamura, J. S. O'Donnell, S. Szczepanski, M. Brandes, S. Eickhoff, I. Das, N. Shridhar, et al. Reactive neutrophil responses dependent on the receptor tyrosine kinase c-MET limit cancer immunotherapy. *Immunity*, 47(4):789–802, 2017.

[88] N. Glodde, A. Kraut, D. van den Boorn-Konijnenberg, S. Vadder, F. Kreten, J. L. Schmid-Burgk, P. Aymans, K. Echelmeyer, M. Rumpf, J. Landsberg, T. Bald, T. Tüting, A. Bovier, and M. Hölzel. Experimental and stochastic models of melanoma T-cell therapy define impact of subclone fitness on selection of antigen loss variants. *bioRxiv:10.1101/860023*, 2019.

[89] E. B. Gunnarsson, S. De, K. Leder, and J. Foo. Understanding the role of phenotypic switching in cancer drug resistance. *arXiv:2001.07690v1, to appear in J. Theor. Biol.*, 2020.

[90] J. B. S. Haldane. A mathematical theory of natural and artificial selection (series of ten papers). *T. Camb. Philos. Soc., P. Camb. Philos. Soc., and Genetics*, 1924–1934.

[91] D. Hanahan and R. A. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674, 2011.

[92] G. H. Hardy. Mendelian proportions in a mixed population. *Science*, 28:49–50, 1908.

[93] K. S. Hoek, O. M. Eichhoff, N. C. Schlegel, U. Döbbeling, N. Kobert, L. Schaerer, S. Hemmi, and R. Dummer. In vivo switching of human melanoma cells between proliferative and invasive states. *Cancer Res.*, 68(3):650–656, 2008.

[94] K. S. Hoek and C. R. Goding. Cancer stem cells versus phenotype-switching in melanoma. *Pigm. Cell Melanoma R.*, 23(6):746–759, 2010.

[95] J. Hofbauer and K. Sigmund. Adaptive dynamics and evolutionary stability. *Appl. Math. Lett.*, 3(4):75–79, 1990.

[96] J. Hofbauer and K. Sigmund. *Evolutionary games and population dynamics*. Camb. Univ. Press, 1998.

[97] 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.

[98] K. Jain. Evolutionary dynamics of the most populated genotype on rugged fitness landscapes. *Phys. rev. E, Stat., nonlinear, and soft matter phys.*, 76 3 Pt 1:031922, 2007.

[99] K. Jain and J. Krug. Evolutionary trajectories in rugged fitness landscapes. *J Stat Mech: Theory and Exp.*, 2005(04):P04008, apr 2005.

[100] K. Jain and J. Krug. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics*, 175:1275–88, 03 2007.

[101] J. L. Jameson, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and J. Loscalzo. Harrison's principles of internal medicine, 20th edition. 2018.

[102] S. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.*, 128(1):11–45, 1987.

[103] S. A. Kauffman. The origins of order: Self-organization and selection in evolution. In *Spin glasses and biology*, pages 61–100. World Scientific, 1992.

[104] P. J. Keeling and J. D. Palmer. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.*, 9(8):605–618, 2008.

[105] G. Kersting and A. Wakolbinger. Probabilistic aspects of Λ-coalescents in equilibrium and in evolution. *arXiv:2002.05250v1; to appear in: Probabilistic Structures in Evolution, E. Baake and A. Wakolbinger (eds.), EMS Publishing House, Zurich*, 2020.

[106] G. J. Kimmel, F. L. Locke, and P. M. Altrock. Response to CAR T cell therapy can be explained by ecological cell dynamics and stochastic extinction events, version 3. *bioRxiv:10.1101/717074*, 2020.

[107] M. Kimura. Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci. USA*, 41(3):144–150, 1955.

[108] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, (Special Vol. 19A):27–43, 1982.

[109] E. Kisdi. Adaptive dynamics papers. `https://www.mv.helsinki.fi/home/kisdi/addyn.htm`. Accessed: 06.04.2020.

[110] J. Kohlmeyer, M. Cron, J. Landsberg, T. Bald, M. Renn, S. Mikus, S. Bondong, D. Wikasari, E. Gaffal, G. Hartmann, et al. Complete regression of advanced primary and metastatic mouse melanomas following combination chemoimmunotherapy. *Cancer Res.*, 69(15):6265–6274, 2009.

[111] A. N. Kolmogorov, I. G. Petrovsky, and N. S. Piskunov. Investigation of the equation of diffusion combined with increasing of the substance and its application to a biology problem. *Mosc. Univ. Math. Bull.*, 1(6):1–25, 1937.

[112] A. Kraut and A. Bovier. From adaptive dynamics to adaptive walks. *J. Math. Biol.*, 79(5):1699–1747, 2019.

[113] J. Krug. Accessibility percolation in random fitness landscapes. *arXiv:1903.11913v1; to appear in: Probabilistic Structures in Evolution, E. Baake and A. Wakolbinger (eds.), EMS Publishing House, Zurich*, 2019.

[114] J. Krug and C. Karl. Punctuated evolution for the quasispecies model. *Physica A: Stat. Mech. Appl.*, 318(1):137 – 143, 2003.

[115] W. Kutta. Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. *Z. Math. Phys.*, 46:435–453, 1901.

[116] V. A. Kuznetsoz, L. Makalkin, M. Talor, and A. S. Perelson. Nonlinear dynamics of immunogenic tumors: parameter estimation and global bifurcation analysis. *Bull. Math. Biol.*, 56:295–321, 1994.

[117] J. Landsberg, J. Kohlmeyer, M. Renn, T. Bald, M. Rogava, M. Cron, M. Fatho, V. Lennerz, T. Wölfel, M. Hölzel, et al. Melanomas resist T-cell therapy through inflammation-induced reversible dedifferentiation. *Nature*, 490(7420):412–416, 2012.

[118] M. Leisegang, B. Engels, K. Schreiber, P. Y. Yew, K. Kiyotani, C. Idel, A. Arina, J. Duraiswamy, R. R. Weichselbaum, W. Uckert, et al. Eradication of large solid tumors by gene therapy with a T-cell receptor targeting a single cancer-specific point mutation. *Clin. Cancer Res.*, 22(11):2734–2743, 2016.

[119] H. Leman. Convergence of an infinite dimensional stochastic process to a spatially structured trait substitution sequence. *Stoch. Partial Differ. Equ. Anal. Comput.*, 4(4):791–826, 2016.

[120] R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami. The evolutionary origin of complex features. *Nature*, 423(6936):139–144, 2003.

[121] U. Lenz, S. Kluth, E. Baake, and A. Wakolbinger. Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution. *Theor. Popul. Biol.*, 103:27–37, 2015.

[122] M. I. Lind and F. Spagopoulou. Evolutionary consequences of epigenetic inheritance. *Heredity*, 121:205–209, 2018.

[123] A. J. Lotka. Quantitative studies in epidemiology. *Nature*, 88:497–498, 1912.

[124] R. Maddamsetti, R. E. Lenski, and J. E. Barrick. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with escherichia coli. *Genetics*, 200(2):619–631, 2015.

[125] T. Malthus. *An essay on the principle of population, as it affects the future improvement of society, with remarks on the speculations of Mr. Goodwin, M. Condorcet and other writers.* J. Johnson in St Paul's Church-Yard, London, 1798.

[126] L. Marchetti, C. Priami, and V. H. Thanh. HRSSA–Efficient hybrid stochastic simulation for spatially homogeneous biochemical reaction networks. *J. Comput. Phys.*, 317:301–317, 2016.

[127] 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.

[128] J. S. Marshall, R. Warrington, W. Watson, and H. L. Kim. An introduction to immunology and immunopathology. *Allergy Asthma Cl. Im.*, 14(Suppl. 2):49, 2018.

[129] A. Marusyk, V. Almendro, and K. Polyak. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, 12(5):323–334, 2012.

[130] H. Mayer and A. Bovier. Stochastic modelling of T-cell activation. *J. Math. Biol.*, 70(1-2):99–132, 2015.

[131] J. Maynard Smith. *The scientist speculates: An anthology of partly-baked ideas.* Basic Books, 1962.

[132] J. Maynard Smith. Natural selection and the concept of a protein space. *Nature*, 225:563–564, 1970.

[133] N. McGranahan, A. J. Furness, R. Rosenthal, S. Ramskov, R. Lyngaa, S. K. Saini, M. Jamal-Hanjani, G. A. Wilson, N. J. Birkbak, C. T. Hiley, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 351(6280):1463–1469, 2016.

[134] A. Mehta, Y. J. Kim, L. Robert, J. Tsoi, B. Comin-Anduix, B. Berent-Maoz, A. J. Cochran, J. S. Economou, P. C. Tumeh, C. Puig-Saus, et al. Immunotherapy resistance by inflammation-induced dedifferentiation. *Cancer Discov.*, 8(8):935–943, 2018.

[135] I. Mellman, G. Coukos, and G. Dranoff. Cancer immunotherapy comes of age. *Nature*, 480(7378):480–489, 2011.

[136] G. Mendel. Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 4:3–47, 1865.

[137] G. Mendel. Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde. *Verhandlungen des naturforschenden Vereines in Brünn*, 8:26–31, 1869.

[138] J. A. J. Metz. Adaptive dynamics. In *Encyclopedia of Theoretical Ecology*, pages 7–17. 2012.

[139] 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)*, volume 45 of *Konink. Nederl. Akad. Wetensch. Verh. Afd. Natuurk. Eerste Reeks*, pages 183–231. North-Holland, Amsterdam, 1996.

[140] J. A. J. Metz, R. M. Nisbet, and S. A. H. Geritz. How should we define 'fitness' for general ecological scenarios? *Trends Ecol. Evol.*, 7(6):198–202, 1992.

[141] P. A. P. Moran. Random processes in genetics. *Math. Proc. Cambridge*, 54(1):60–71, 1958.

[142] A. E. Motter. Improved network performance via antagonism: from synthetic rescues to multi-drug combinations. *BioEssays*, 32(3):236–245, 2010.

[143] K. M. Murphy and C. Weaver. *Janeway's Immunobiology, 9th edition.* Taylor & Francis, 2017.

[144] J. Neidhart and J. Krug. Adaptive walks and extreme value theory. *Phys. Rev. Lett.*, 107:178102, Oct 2011.

[145] C. Neuhauser and S. M. Krone. Ancestral processes with selection. *Theor. Popul.. Biol.*, 51(3):210–237, 1997.

[146] R. Neukirch and A. Bovier. Survival of a recessive allele in a Mendelian diploid model. *J. Math. Biol.*, 75(1):145–198, 2017.

[147] S. Nowak and J. Krug. Analysis of adaptive walks on NK fitness landscapes with different interaction schemes. *J. Stat. Mech. Theory Exp.*, (6):P06014, 27, 2015.

[148] P. C. Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28, 1976.

[149] H. Ochman, J. G. Lawrence, and E. A. Groisman. Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405(6784):299–304, 2000.

[150] H. A. Orr. A minimum on the mean number of steps taken in adaptive walks. *J. Theor. Biol.*, 220(2):241–247, 2003.

[151] W. W. Overwijk, A. Tsung, K. R. Irvine, M. R. Parkhurst, T. J. Goletz, K. Tsung, M. W. Carroll, C. Liu, B. Moss, S. A. Rosenberg, et al. gp100/pmel 17 is a murine tumor rejection antigen: Induction of 'self'-reactive, tumoricidal T cells using high-affinity, altered peptide ligand. *J. Exp. Med.*, 188(2):277–286, 1998.

[152] S. L. Park, A. Buzzai, J. Rautela, J. L. Hor, K. Hochheiser, M. Effern, N. McBain, T. Wagner, J. Edwards, R. McConville, et al. Tissue-resident memory CD8+ T cells promote melanoma–immune equilibrium in skin. *Nature*, 565(7739):366–371, 2019.

[153] J. B. Plotkin and G. Kudla. Synonymous but not the same: The causes and consequences of codon bias. *Nat. Rev. Genet.*, 12(1):32–42, 2011.

[154] J. Reinhardt, J. Landsberg, J. L. Schmid-Burgk, B. B. Ramis, T. Bald, N. Glodde, D. Lopez-Ramos, A. Young, S. F. Ngiow, D. Nettersheim, et al. MAPK signaling and inflammation link melanoma phenotype switching to induction of CD73 during immunotherapy. *Cancer Res.*, 77(17):4697–4709, 2017.

[155] N. P. Restifo, F. M. Marincola, Y. Kawakami, J. Taubenberger, J. R. Yannelli, and S. A. Rosenberg. Loss of functional beta2-microglobulin in metastatic melanomas from five patients receiving immunotherapy. *J. Natl. Cancer I.*, 88(2):100–108, 1996.

[156] S. Riesenberg, A. Groetchen, R. Siddaway, T. Bald, J. Reinhardt, D. Smorra, J. Kohlmeyer, M. Renn, B. Phung, P. Aymans, et al. MITF and c-Jun antagonism interconnects melanoma dedifferentiation with pro-inflammatory cytokine responsiveness and myeloid cell recruitment. *Nat. Commun.*, 6:8755, 2015.

[157] C. Runge. Über die numerische Auflösung von Differentialgleichungen. *Math. Ann.*, 46(2):167–178, 1895.

[158] S. Sahasrabudhe and A. E. Motter. Rescuing ecosystems from extinction cascades through compensatory perturbations. *Nat. Commun.*, 2(1):1–8, 2011.

[159] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J. Chem. Phys.*, 122(5):054103, 2005.

[160] K. R. Sanft and H. G. Othmer. Constant-complexity stochastic simulation algorithm with optimal binning. *J. Chem. Phys.*, 143:074108, 2015.

[161] B. Schmiegelt and J. Krug. Evolutionary accessibility of modular fitness landscapes. *J. Stat. Phys.*, 154(1-2):334–355, 2014.

[162] B. A. Sewastjanow. *Verzweigungsprozesse*. R. Oldenbourg Verlag, Munich-Vienna, 1975.

[163] C. Smadi. An eco-evolutionary approach of adaptation and recombination in a large population of varying size. *Stoch. Process. Appl.*, 125(5):2054–2095, 2015.

[164] C. Smadi. The effect of recurrent mutations on genetic diversity in a large population of varying size. *Acta Appl. Math.*, 149:11–51, 2017.

[165] S. Smale. On the differential equations of species in competition. *J. Math. Biol.*, 3(1):5–7, 1976.

[166] A. Sucker, F. Zhao, N. Pieper, C. Heeke, R. Maltaner, N. Stadtler, B. Real, N. Bielefeld, S. Howe, B. Weide, et al. Acquired IFNγ resistance impairs anti-tumor immunity and gives rise to T-cell-resistant melanoma lesions. *Nat. Commun.*, 8(1):1–15, 2017.

[167] I. G. Szendro, M. F. Schenk, J. Franke, J. Krug, and J. A. G. M. de Visser. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech. Theory Exp.*, (1):P01005, 22, 2013.

[168] E. Tran, P. F. Robbins, Y.-C. Lu, T. D. Prickett, J. J. Gartner, L. Jia, A. Pasetto, Z. Zheng, S. Ray, E. M. Groh, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *New Engl. J. Med.*, 375(23):2255–2262, 2016.

[169] V. C. Tran. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM Probab. Stat.*, 12:345–386, 2008.

[170] V. Volterra. Variations and fluctuations of the number of individuals in animal species living together. *ICES J. Mar. Sci.*, 3(1):3–51, 1928.

[171] S.-D. Wang. *Multi-scale analysis of adaptive population dynamics*. PhD thesis, Bonn University, 2011.

[172] H. W. Watson and F. Galton. On the probability of the extinction of families. *J. R. Anthropol. Inst. G.*, 4:138–144, 1875.

[173] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[174] W. Weinberg. Über den Nachweis der Vererbung beim Menschen. *Jahresh. Ver. Vaterl. Natkd. Württ.*, 64:369–382, 1908.

[175] M. H. F. Wilkins, A. R. Stokes, and H. R. Wilson. Molecular structure of nucleic acids: molecular structure of deoxypentose nucleic acids. *Nature*, 171:738–740, 1953.

[176] Y. Wolf, O. Bartok, S. Patkar, G. B. Eli, S. Cohen, K. Litchfield, R. Levy, A. Jiménez-Sánchez, S. Trabish, J. S. Lee, et al. UVB-induced tumor heterogeneity diminishes immune response in melanoma. *Cell*, 179(1):219–235, 2019.

[177] S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97–157, 1931.

[178] J. C. Yang and S. A. Rosenberg. Adoptive T-cell therapy for cancer. In *Adv. Immunol.*, volume 130, pages 279–294. Elsevier, 2016.

[179] J. M. Zaretsky, A. Garcia-Diaz, D. S. Shin, H. Escuin-Ordinas, W. Hugo, S. Hu-Lieskovan, D. Y. Torrejon, G. Abril-Rodriguez, S. Sandoval, L. Barthly, et al. Mutations associated with acquired resistance to PD-1 blockade in melanoma. *New Engl. J. Med.*, 375(9):819–829, 2016.

[180] M. L. Zeeman. Hopf bifurcations in competitive three-dimensional Lotka-Volterra systems. *Dynam. Stab. Syst.*, 8(3):189–217, 1993.