Fitness Valleys, Metastability and Changing Environments

Dissertation

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

vorgelegt von

Manuel Esser

aus
Bonn

Bonn, 4. April 2025

Angefertigt mit Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer: Prof. Dr. Anton Bovier

Gutachter: Prof. Dr. Jochen Blath

Tag der Promotion: 2. Juli 2025

Erscheinungsjahr: 2025

Dedicated to my grandmothers

Abstract

Mathematical modelling of biological processes has become an area of high scientific interest. This field has expanded substantially over the past century and developed a wide variety of mathematical approaches and biological applications. The complexity of biological systems requires models taking random effects into account.

This thesis investigates stochastic individual-based models of adaptive dynamics for asexually reproducing populations with mutation, focussing on the interplay between population dynamics, mutation rates, and environmental changes. Evolution is driven by linear birth rates, density-dependent logistic death rates, and mutations occurring along a finite trait graph. The model partially incorporates time-varying parameters, such as periodic changes in the environment or drug concentrations, which impact the evolutionary process. We investigate the behaviour of mutants and their invasion dynamics under small mutation rates and a simultaneously diverging population size, where environmental changes occur on a moderately diverging time scale.

The results of the first part (Appendix A) provide a detailed analysis of transitions between evolutionary stable conditions (ESC) in a constant environment. Here multiple mutations need to be accumulated to cross fitness valleys. The system exhibits metastable behaviour across multiple time scales which are linked to the widths of these fitness valleys. Introducing a meta-graph framework of ESCs, we describe the multi-scale jump chain resulting from concatenated jumps and prove the convergence of the population process to a Markov jump process that visits only ESCs of sufficiently high stability.

We then turn to the study of periodically changing environments. In the second part (Appendix B), we examine the growth of emergent mutants and their invasion of the resident population with a focus on mesoscopic scaling limits and the effective growth rates of mutants. The dynamics are influenced by an averaging effect of invasion fitness across different phases of the environment.

Additionally, we explore the crossing of fitness valleys in a changing environment in the third part (Appendix C), distinguishing two cases: Under the assumption of a strict fitness valley, we can show that the crossing rates are computed as an average taking into account the ability to survive. A particularly interesting scenario is the pit stop phenomenon, where intermediate mutants within a fitness valley experience phases of positive fitness, allowing them to grow to large sizes before going extinct. This accelerates the traversal of the valley and introduces a novel time scale in the evolutionary process.

Acknowledgements

For sure, this thesis would not have been possible without the help of many people. Therefore, I would like to take the opportunity to express my thanks to all of them.

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Anton Bovier. He was the one who supported and guided me from the beginning of my studies in Bonn. He not only introduced me to analysis and probability theory but also showed me the wide spectrum of stochastic models for complex real-world phenomena. Ultimately, he aroused my interest in the mathematical models of evolutionary theory and enabled me to study them intensively during my doctoral studies. His supervision was always characterised by great trust in my autonomous work, while at the same time demanding relevant and ambitious research work. Thank you for your patience, guidance and support, for fostering me in presenting my research at conferences and becoming part of the scientific network and for sharing your enormous intuitions, your immense knowledge and your ideas among solving mathematical problems with me.

My heartfelt thanks go to my co-author, Dr. Anna Kraut, who initiated our collaboration after co-supervising my master's thesis. I am glad that I could profit from her expertise on my PhD topic. Moreover, she always and honestly made me feel like we were working at eye level. I especially enjoyed our mathematical discussions and leisure activities during my stays in Minneapolis. I am very thankful for countless online meetings and all the loops she turned with me.

I would like to thank Prof. Dr. Jochen Blath for taking the job of the second referee and the other members of the committee Prof. Dr. Alexander Effland and Prof. Dr. Michael Hölzel for being part of the jury. Further, I thank Prof. Dr. Angkana Rüland for being my mentor.

Among all members of the probability theory group, I have to highlight Mei-Ling. There are no words describing her dedication to the group and to each of us personally. She is the heart and the soul of the group and always takes care of us, trying to make our life as easy as possible. Thanks to my office mates, Maximilian, Joy and especially Annabell for manoeuvring PhD life together and for many interesting exchanges on more or less scientific questions. Thanks also to Adrien, Dominik, Florian, Francis, Giulia, Katharina, Luis, Min, Saeda and Stefan, for making the time in the group enjoyable.

I want to give my appreciation to my PhD colleagues Uta, Jannik and Paul who shared this journey since our Bachelor studies.

My special thanks go to those who took the time to proofread parts of this work: Anna, Francis, Jannik and Uta.

Thanks go also to the Deutsche Forschungsgemeinschaft (DFG) for the funding under Germany's Excellence Strategy GZ 2047/1, Projekt-ID 390685813 and GZ 2151, Project-ID 390873048, through the Collaborative Research Centre 1060 “The Mathematics of Emergent Effects” Project-ID 211504053 and the Bonn International Graduate School in Mathematics (BIGS). 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.

Eventually, I would like to express my deepest gratitude to my parents and my sister for the love, the encouragement and support I constantly received from them. They have always let me know that they would support me no matter what.

Contents

1. Introduction

This thesis is intended to contribute to an improved mathematical understanding of *evolutionary dynamics*. To this end, we study a *stochastic individual-based model* of an asexually reproducing population, based on the few fundamental mechanisms of Darwinian evolution. Our primary focus is on the analysis of rare events that can only be observed over large time scales. Furthermore, we investigate the implications of environmental changes on the emergence of new traits, the occurrence of rare events, and the whole evolutionary process.

A pivotal rôle in advancing our understanding of the interplay of *ecology* and *evolution* is played by mathematics, particularly through the formalisation and quantification of interactions between individuals. Biology and mathematics share a long history of at least 200 years [47]. Over the past few decades, the integration of biology and mathematics has led to significant advancements in understanding complex biological systems [103]. A key area of interest is local regulation in biological systems, where macroscopic enhancements emerge from the microscopic interactions of individual components with each other and with the environment. This becomes of particular relevance in the study of evolutionary dynamics, where the focus is on how populations adapt to environmental pressures through a series of changes in traits. Both evolutionary mechanisms of how these changes arise and how certain traits are selected to be favoured in the overall population are objects of research. Therefore, mathematical modelling has become an essential tool as it bridges different scales.

The mathematical models used to study these dynamics range from deterministic systems such as differential equations, to stochastic models including diffusions and Markov jump processes. The choice of model depends on the specific biological question at hand, with stochastic individual-based models capturing the detailed behaviour of single individuals within a population. This class of models forms the foundation of our analyses. While individual-based models offer valuable insights, they can be computationally expensive, particularly in large populations. To address this, researchers explore simplified models, such as deterministic equations or less detailed stochastic processes, to approximate the behaviour of large systems more efficiently. By analysing these models, they aim to better understand the long-term evolution of populations.

In order to establish a consistent mathematical theory of evolution, it is essential to link complex microscopic models with those that offer a more macroscopic perspective. In this thesis, we make use of a common systematic approach to introduce scaling parameters and derive macroscopic descriptions as limits of the more intricate models. These limit approximations can hence be applied to improve simulation algorithms for the underlying stochastic models.

We focus on three scaling parameters: The order of the total population size K, the probability of mutations μ_K, and the speed of changes in the environment λ_K. We consider the regime of simultaneously large populations ($K \nearrow \infty$) and rare mutations ($\mu_K \searrow 0$). The frequency of mutations varies depending on the rate at which μ_K decays. This can lead to mutations being either spaced apart, compared to faster ecological dynamics, or overlapping and competing to invade the population. For the environmental changes we choose an intermediate time scale between those of stabilisation of the resident population and exponential growth of mutants ($1 \ll \lambda_K \ll \ln K$). The system can be viewed across multiple time scales, each of which emphasises different aspects of the evolutionary process. As a result, approximating the stochastic individual-based model leads to a number of results with various degrees of resolution as presented in the main part of this thesis.

The interlock of the several effects acting on different time scales is the main concern of Appendix A, with a particular focus on *metastable transitions*. When the evolutionary process reaches a state in which all possible single mutations are harmful, it appears to pause. However, the accumulation of several mutations can produce an individual with higher fitness. Considering a larger time scale, it is possible to escape the metastable state by crossing a *fitness valley*. We look at the evolution on a general finite trait graph and give a precise description of transitions between evolutionary stable states, thereby going beyond the analysis of toy models. The general underlying trait graph is also the basis for the analysis in Appendix B. Here, the focus lies on comprehending the effects of the *changing environment* on the emergence and growth of new mutants. In Appendix C, we finally examine a combination of both, fitness valleys and changing environments in two simplified scenarios.

The remainder of this thesis is structured as follows: In Chapter 2, we give a brief introduction to the history of the biological theory of evolution and present different approaches of its mathematical modelling. Chapter 3 is dedicated to the mathematical model studied in this thesis. In Section 3.1, we introduce the individual-based model followed by the scaling parameters in Section 3.2. Together these sections form the basis for the extensions studied in the Appendices. Section 3.3 explains the notion of fitness used in this thesis, while Section 3.4 is dedicated to a careful overview of the inherent time scales of the model. In Section 3.5, we show how a changing environment can be integrated into the model. We summarise the main results in Chapter 4, accompanied by a discussion of the further outlook in Chapter 5. The Appendices A, B and C contain the two publications and the preprint that together form the heart of the work.

2. Approaches of mathematical modelling

2.1. Foundations in Evolution Theory

The most prominent mastermind of the modern theory of evolution is Charles Darwin. For though some preliminary thoughts were developed by his grandfather Erasmus Darwin [53, 54] and Thomas Malthus [135], C. Darwin presented a completely new perspective in his seminal work *On the origin of species* [51], elaborated in 1839 but published 20 years later. Therein and in the joint work with Alfred Wallace [52] he laid the cornerstone of our modern thinking of evolution and speciation as a process driven by the interplay of three simple mechanisms:

> *Heredity*: reproduction of individuals that pass their traits on to their offspring,
>
> *Variation*: heredity is not perfect, i.e. sometimes the offspring's traits vary from those of its ancestor,
>
> *Natural Selection*: different traits have different rates of survival and reproduction.

While in the retrospective these ideas seem to be natural and intuitive, they stand in opposition to what was argued by other savants. Let us mention exemplarily the theory invented by Lamarck [57], claiming that individuals achieve advantages during their life time which then are passed on to their descendants. The famous representative here is the giraffe, stretching to reach the leaves of taller trees and thus, in Lamarck's view, gaining a small increase of its neck. Over generations, this should result in the long neck characterising giraffes today. Although Lamarck's theory now is outdated, some aspects revive empirically verified in the field epigenetics and horizontal gene transfer, as we discuss below.

Darwin's work is based mostly on observations missing the modern techniques, laboratories and instruments we have at hand. However, his ideas were groundbreaking and foreseeing in view of our modern theory. For example, the traits to be inherited were described by a bundle of the morphology and physiological properties of the individuals. This is called the *phenotype*. Moreover, to Darwin it was not really clear how heredity works. It was the abbot and researcher Gregor Mendel who came up with the concept of heredity on the basis of *genes* [138, 139]. His empirical studies on peas led him to distinguish *alleles* and prove rules on their mixture producing the traits one can observe. In modern language, these rules of *sexual reproduction* of *diploid* organisms are part of what we generally call *recombination*, one of the sources of variation. It took more than 100 years until Crick, Watson, Wilkens and Franklin [85, 165, 168] discovered the material corresponding to Mendel's theory. The whole genetic information, the *genotype*, shaping all features of an individual, is encoded in the large molecule-strings of the DNA. This is also what allows for *heredity* as basic mechanism, since the DNA is usually copied without major changes during reproduction. However, the visible phenotype is not fully determined by the genes, but it is compiled by the complex mechanism of gene expression influenced by the environment surrounding the organism and regulatory mechanisms acting on the gene, which is analysed in Epigenetics.

Besides recombination, there are at least two other sources of variation. A key role is played by *mutations*. These errors arising during DNA replication can produce truly new features and hence enrich the pool of genes, instead of only rearranging what is available (cf. Hardy-Weinberg principle). Thus mutations are crucial for the adaptation of species to environments that permanently undergo changes.

Last but not least, let us mention that some organisms (e.g. bacteria) have the ability to interchange genetic material during their lifetime. In a figurative sense, this phenomenon, called *horizontal gene transfer* (HGT), corresponds to some ideas of Lamarck. For example, if a bacterium has by chance the resistance against a particular antibiotic, it can transfer a copy of the coding gene to a neighbouring bacterium, when triggered by the environment. Hence the latter individual gained a new beneficial feature during its lifetime, which is then inherited by its descendants.

The third component of Darwin's explanation of evolution, the natural selection, leads the random forces of variation in some direction and hence shapes the evolutionary process. It can be divided into two aspects. The *fitness*, meaning the ability to survive and reproduce, is on the one hand determined by the adaptedness to the environment. If an individual has a strong need for unavailable resources it may have a small survival probability, whereas a well-adapted organism with a high fertility is likely to populate most of the available habitat. On the other hand, individuals interact with each other, either through the environment by the competition for resources (nutrients, water, light, space, etc), which turns out to be crucial to explain the limits of growth (cf. Section 2.2). Moreover, the interactions can be immediate in dependent relations such as the asymmetric predator-pray or parasite-host relation or mutually beneficial relations of cooperation or symbiosis.

Although Darwin's theory, that is often summarised by the principle *survival of the fittest*, already covers the key aspects of evolution theory its beauty should not hide the fact that evolution is still a highly complex process challenging us with many open problems on various levels. To bring it to one single question: How do all the chemical processes on subcellular level interact and lead to macroscopically observable speciation in the long run?

Since Darwin, a broad theory of evolution partially answering these questions has been established. In the following, we outline some of the contributions of different mathematical approaches, each of which aims to explain particular aspects of the fascinating phenomenon. For example, the accumulation of variations can be traced back to be the origin of evolutionary jumps and branching. Additionally, it is of no minor interest how the environment, which itself is not constant, influences the genotypic evolution and the phenotypic expression. Though mathematical models are always simplifications or approximations of the real world, the formalisation to abstract problems allows to understand the complex mechanisms on multiple scales and gives insights into the inter-scale dependencies.

Before turning to the mathematical models, let us mention Mukherjee's textbook [144] as a great reference on the history of modern theory of evolution. Moreover, the Encyclopedia Britannica [5] covers different aspects in more detail.

2.2. Population dynamics

One of the oldest directions of biomathematics is *population dynamics*. It aims to analyse the ecological forces generated by the interactions between different species and its origin can be traced back to the end of the 18th century. In [135], Malthus worked out the contrast of population growth following geometric orders and the limited resources of arithmetic order. Notwithstanding the political and socio-ethic conclusions Malthus deduced, his theory pointed out the importance of competitive interaction as a strong force of ecology. Therefore, the first mathematical expression for the dynamics of the population size is the ODE of logistic growth

$$\dot{n}(t) = rn(t) - cn^2(t), \qquad t \geq 0, \tag{2.2.1}$$

where r denotes the expected net growth, i.e. the difference between birth and death rates, and c parametrises the strength of competitive pressure two individuals exert on each other. Let us remark that the function $n : \mathbb{R}_+ \to \mathbb{R}_+$ has to be seen as the (rescaled) mass of the population rather than counting single individuals. In fact, we can identify n to be a scaling limit of an "infinite" population (cf. Section 3.2). In this simple case of a *monomorphic* population (i.e. consisting of individuals all of which have the same trait) and under the assumptions of $r \geq 0$ and $c > 0$, one can easily identify the two fixed points 0 and r/c, the first one unstable and the later one a stable and globally attractive equilibrium.

Turning to the joint ecological dynamics of multiple different species, what we call *polymorphic* populations, directly brings us to the work of Alfred Lotka [130] and Vito Volterra [164]. Considering a finite set V of different traits, they studied the natural generalisation of (2.2.1),

$$\dot{n}_v(t) = n_v(t)\left(r_v - \sum_{w \in V} c_{v,w} n_w(t)\right), \qquad t \geq 0,\ v \in V. \tag{2.2.2}$$

Named *Lotka-Volterra equations* (LVE) after their investigators, these systems display quite different behaviour depending on the parameters; besides the case of the *competitive* LVE ($c_{v,w} \geq 0$), the *predator-prey system* ($V = \{1,2\}, r_1 < 0, r_2 > 0, c_{1,2} > 0, c_{2,1} < 0$ and $c_{1,1} = c_{2,2} = 0$) has gained some popularity due to its periodic solutions.

In the further course of this work, the Lotka-Volterra equations and in particular their (stable) equilibria play a crucial rôle. However, the stability analysis becomes a major challenge for populations consisting of more than two different types. Already for the three-dimensional case, Zeeman lists 33 different scenarios [172] and in dimension $|V| \geq 5$ on can show that the system can exhibit arbitrary complex behaviour in the long term [159].

To get a comprehensive overview of the approaches of population dynamics the readers are referred to the monograph of Hofbauer and Siegmund [102].

2.3. Populations genetics

The main objective of *population genetics* is the understanding of genealogical structures and the changes of gene frequency. Laying the emphasis on heredity, ecological interactions of individuals are usually ignored. An overview of the early theory is given in [67], whereas [77, 74] provide also some modern perspectives.

The mathematical rigorous investigation of the concept of heredity started with the rediscovery of Mendel's work in the beginning of the 20th century by Ronald Fisher [80], Sewall Wright [169] and John B.S. Haldane [97]. An elementary model, that can be seen as the corner stone of many models studied today is the *Wright-Fisher model.* It considers a population of constant size N and time is counted in discrete generations. Each individual of a new generation independently choses a ancestor at random from the previous generation to inherit its trait. To link this model to empirical observations let us take a look on the long-term experiment initiated by Lenski [128]. Its protocol is the following: After bacteria could grow and populate the whole medium over night, in the morning a laboratory worker takes a random fraction of it to be placed on a new medium and each morning the same procedure. If we ask now for the probability of a single bacterium to carry a particular trait of the previous generation, this meets exactly the trait frequency before the laboratory worker put hands on the medium.

Despite the fact, that the Wright-Fisher model seems very simplified, it exhibits a central paradigm. It is a *neutral* model since it does not contain any bias, but the probability of receiving a certain trait is exactly its frequency in the previous generation. One can thus show that the time evolution of the frequencies is a martingale. This is in accordance with the *Hardy-Weinberg principle* [98, 167] stating that in infinite populations with no selection, the allele frequencies remain constant. Then again, due to random fluctuations, one can determine a *genetic drift* driving the finite population towards the states of genetic homogeneity. These are the absorbing states since after losing genetic diversity there is no possibility to regain.

Both of these properties are conserved in the continuous (time and space) analogue, the Wright-Fisher diffusion. Introduced by Kimura [114] it can be derived from the discrete model as a diffusion limit if the time is rescaled as $\lfloor tN \rfloor$, which was shown by Ethier and Norman [76].

Since the Wright-Fisher model laid the cornerstone, many extensions and generalisations have been developed. The continuous time equivalent preserving the finite number of individuals but allowing for exponential distributed inter-generation times is the Moran model [143]. The Cannings model [38, 39] is in some way a joint extension, as it can mimic both. Building up on the classical discrete generation model, it allows for more general offspring distributions. The only assumption is the exchangeability of the offspring distribution to keep the mean trait frequencies constant. Again, as an infinite population limit, but this time from the Moran model, the Flemming-Viot model is deduced [82].

Changing the perspective, the above models and their variations can also be studied backwards in time along the ancestral lines. This analysis of the genealogical tree structure leads to coalescent processes. The first one to mention is the one introduced by Kingmann [115], as it is the natural time-reversal to the Moran model. Here each pair of lineages is assigned an exponential clock, when the first clock rings these two lineages coalesce and are treated from now on as a single lineage. Corresponding to different forward in time models one derives various backward in time processes. They differ in the rate of and the nature of mergers: Namely the Λ-coalescent allows multiple lineages to merge and in the Ξ-coalescent different mergers may happen simultaneously. A particular process to mention is the Bolthausen-Sznitman-coalescent as it is not only some example of the Λ-coalescent, but turns out to be a universal scaling limit of spin glass models of statistical physics. We refer to [12] for an insightful introduction to this field. Let us just mention that this is a vital field of active research bringing out new models integrating further aspects as for example the Brownian spatial coalescent [117] or the weak and strong seedbank coalescent [17].

To understand evolution, it is necessary to overcome the inherent problem of constant allele frequencies, stated in the Hardy-Weinberg principle. As already Darwin proposed, one has to integrate the two missing mechanisms: variation and selection. Mutation allows for an enrichment of the pool of genes. In the models of constant population size, this can be implemented in the following way: At a randomly chosen time, a single individual is affected by a mutation event and henceforth has a new trait, which was not present before. Selection can be included in the form of a bias to not choose parents uniformly at random but to favour some of which we then call to be of higher fitness. Backwards in time, this yields the ancestral selection graph [147] and the ancestral recombination graph [2], if one includes also recombination as a source of variation. Let us point out that this kind of selection does not rely on true interaction of the living individual but relies on a fixed fitness landscape or *adaptive landscape* [170], which comes with a clear ordering of the traits. Therefore we cannot talk about the important phenomena of coexistence and speciation. Instead population is driven by selection towards (local) fitness maxima which one can escape by chance of the genetic drift. We come back to this manifestation of *metastability* when analysing valleys in the fitness landscape in Appendix A.

Lastly to mention is that for a long time, the theories of Darwin and Mendel were argued to exclude each other since Darwin thought about evolution as a continuous process while Mendel's work showed the discrete jumps between generations. Eventually, it was Fisher [80], who laid the basis for understanding that both, Darwin's and Mendel's theories, show two sides of the same coin. The difference was the scale and the otherness of the phenotypic characteristic they studied. As Mendel's focus lies on properties determined just by a few genes, Darwin was interested in features regulated by the interplay of many genes and the environment. The latter is the main concern of the subfield called *quantitative genetics.* It deals with the evolution of phenotypic traits that seem to change gradually. Instead of giving an extensive introduction to this field, we refer to [36, 79], and emphasise that this simultaneousness of different scales is one of the key paradigms to understand evolution.

2.4. Adaptive dynamics

Above we outline that population genetics is very good in dealing with heredity, the genealogical structure and different origins of genetic variation. It sadly lacks the integration of ecological interactions of individuals, which is crucial for a proper description of selective forces. On the opposite, the models of population dynamics are very sensitive to these dependencies between individuals shaping the development of population size. But as they act on an initially fixed set of traits to be distinguished, an inherent problem is the missing variation that would allow an evolution of the pool of genes. Combining the best of both worlds, *adaptive dynamics* aims to take a look at the interesting border line of the *interplay of ecology and evolution.* Initially invented as an outgrow of evolutionary game theory by Hofbauer and Sigmund [101] and parallel contributions of Metz et al. [142] (fitness landscape) as well as Marrow et al. [136] (small mutation steps), which laid the spiritual ground, it has was developed further by Metz et al. [141] (large populations and rare mutations), Dieckmann and Law [61] (canonical equation) and Geritz [88, 89] (evolutionary branching).

A fundamental idea of adaptive dynamics is that the current population can be assumed to be close to an equilibrium, determined by the ecological system, when a new mutant appears. This principle, often referred to as a *separation of time scales*, is quite debatable as it is not clear whether it goes in line with real-world observations. But considering the low effective rate of advantageous mutations and under the reasonable assumption, that the adaptation towards stability of the macroscopic population is faster than the growth of small mutant populations, one can still restrict the dependence of the fixation of new traits on its fitness with respect to a population close to equilibrium. The notion of invasion fitness in adaptive dynamics is oriented on the population dynamics concept of the initial growth rate of small subpopulations. But it also integrates the idea of population genetics to measure selective advantage by the probability of reaching maturity (i.e. producing offspring). Hence fitness is no longer an absolute quantity but becomes dynamic and dependent on the state of the entire population due to interaction (see Section 3.3). Moreover, the invasion fitness and the shape of the corresponding landscape may change totally when the outer environmental conditions change. Let us point out that this notion of invasion fitness landscape has to be distinguished fundamentally from Wright's traditional concept of a fitness landscape [166].

Sometimes the adaptive dynamics approach is specified to study the *meso-evolution* [140] meant in the sense that it connects the micro-evolutionary process of the permanent interplay mutational variation and ecological selection to the macroscopically visible trait substitution. Crucial to guarantee that really one resident trait is replaced by a more advantaged one completely, is to assume that once a mutant is strong enough to form a subpopulation of relevant size, it can fixate in the macroscopic population and long-term coexistence is excluded. This principle is known as *invasion implies fixation* [142]. Mathematically this manifests in the trait-substitution-sequence (see below).

The beauty of adaptive dynamics of course comes by the price of some simplifications. In general, the AD perspective is only interested in observing phenotypes. It usually ignores the genotypic details as well as the mechanism of sexual reproduction. However, as mating is a major source of variation and widely spread among many species, some authors successfully included genetic diploidity, sexual mating-schemes and the Mendelian rules into an AD model [28, 126, 148]. Moreover, it is an ongoing discussion if mutations can be assumed as rare events (see Section 3.2.2). But at least with the focus on advantageous phenotypic mutations, this is broadly accepted.

The *individual-based model of adaptive dynamics* this thesis is based on conceptually relies on the idea of interacting particles. At its core, it is basically a continuous-time Bienaymé-Galton-Watson branching process, but it integrates an additional state-dependent logistic death rate, to take care of the limited resources of the environment and can be seen as an interaction term of the particles. We hence, lose the classical independent additivity which usually characterizes the branching processes. However, this leads to non-trivial equilibria besides extinction or infinite growth one normally deals with in the theory of branching processes. One often refers to these kind of model to be *locally regulated* [18, 40, 43, 55, 72, 83, 94, 104, 127, 132]. Moreover, this self-regulation of the population size is what makes a difference to the Wright-Fisher type models of population genetics as these are often nailed to have a constant population size over time. Together with the possibility to found new subpopulations of different traits due to mutation the ecological regulation via competitive interaction allows for a perceptual adaptation to the environment. This is evolution.

The mathematical foundation of the class of individual-based evolution models can be found in the works of Bolker and Pacala [24, 25], who were originally interested in the spacial dispersion of plants. They utilised the particle-wise description to heuristically derive ODEs Shortly after, Dieckmann and Law [62] enhanced the ODE approach to a stochastic measure-valued process. This representation, actually introduced to deal with the spatial positions, turns out to provide the right framework to consider generally infinite trait spaces as well. Eventually, Fournier and Méléard [83] used Poisson random measures to come up with a path-wise construction of the process and rigorously proved a *law of large numbers* that justifies the former heuristic ODE approaches. Let us emphasise that the convergence of density-dependent Markov processes to the associated ODEs was proven already thirty years before by Kurtz [124].

Since this breakthrough, the individual-based population models are the object of extensive studies by many different authors. Due to its simple and adaptable microscopic description, multiple extensions aim to cover the variety of biological phenomena. To point out just a few of them, Champagnat presented the separation of time scales in [40] by the convergence of the macroscopic evolutionary to a *trait-substitution-sequence* (TSS). Under the collaboration of Méléard this could be extended multi-dimensional *polymorphic-evolution-sequence* (PES) [43]. Both depict the resulting macroscopic evolution as a jump process between evolutionary stable states. Moreover, it were Baar, Bovier and Champagnat, who derived the canonical equation of adaptive dynamics (CEAD) directly as a limit process from the individual-based model. This equation was heuristically introduced by Dieckmann and Law in [61] and describe evolution as continuous movement in the trait space, under the assumption of infinitesimal small mutation steps. Since the driving force of the CEAD is the fitness gradient, it is natural to ask if evolution stops if we run into an *evolutionary singularity*, i.e. if this gradient attains 0. As a rigorous mathematical answer to this question is still part of current research we take it on in Chapter 5 under the keyword of *evolutionary branching.*

Another option of apparent stability is known as *fitness valleys* [29]. Here the evolutionary process has reached a local maximum in a fitness landscape, known as *evolutionary stable condition* (ESC), and one can see this as a combination of an ecological and evolutionary equilibrium. In order to escape a couple of mutation steps have to be accumulated until a trait of higher fitness can invade the resident population. This needs a larger time scale to be observed and it becomes a natural question, which ESC is reached under which condition. We present some answers in Appendix A that is dedicated to discover the *course of evolution* (cf. [140]) and partially ties in with the concept of *adaptive walks* and *adaptive flights.* In this context, it becomes interesting to explore not only the microscopic and the macroscopic state of the population. But particular the *mesosopic* growth in between gives important insights, as well in the stochastic model [48] as in the analysis of the corresponding ODE model [32, 44, 120]. This is also the main point of view of Appendix B.

A very interesting and loosely connected branch that is a topic of recent research aims to take into account the effects of survival of exponentially small subpopulations. Building on early PDE results [63] an initial gate was passed by the rigorous derivation of the *Hamilton-Jacobi equation* from the individual based model with small mutation steps and high mutation probability [44].

For a broad and well structured overview of contributions to various aspects of adaptive dynamics, we refer to the webpage of Kisdi [116]. As important sources for a detailed discussion of the history, the ideas and the aims of adaptive dynamics let us point out the essay of Metz [140] and the thesis of Champagnat [41].

2.5. Further directions

While biological networks are of a highly non-trivial nature also the tools to occupy and investigate evolve continuously. In the last decades, the field of life science expanded quickly aiming to explore and decode the complexity of its object of research. We thus know about the existence of many more interdependences that have an influence on the evolutionary process. In the following, we delineate how different phenomena have been worked out in varying depth. Moreover, we give some hints on additional approaches and model classes.

For instance, the variety of mechanisms to regulate, interchange and store genetic information goes far beyond that of clonal or sexual reproduction with the possibility of mutations. One additional mechanism is the *horizontal gene transfer* (HGT), which allows for the interchange of parts of the genome between individuals during life-time (see [112]). Particularly archaea and bacteria are able to perform HGT, which turns out to be important for the transmission of resistance against antibiotics. covering this an individual-based model has been introduced in [13] and was further developed in [44]. While this part of the superordinate field of *epigenetics* could be reflected well in mathematical models, other aspects like the degree of methylation [96, 162] or further regulatory factors of DNA transcription wait for a rigorous mathematical model.

A quite different phenomenon is known as *dormancy* or the formation of *seed-banks.* In general, one means by this the ability of individuals to switch to a phenotypic state of radically reduced or completely suspended metabolism. Let us point out, that this is not only a single feature but many species have evolved a huge variety of mechanisms that allow for a dormant state. Starting from the formation and spread of seeds by plants, spores of fungus, via the latency phases of infections or cancer cells, up to the adaptive microbial dormancy of bacteria or other unicellular organisms induced by environmental stress. As broad the manifestations of dormancy appear, as broad are also the miscellaneous mathematical models focussing on different aspects. Let us cite [21, 22, 49, 134] to mention just a few of them. Since this is a field of recent research interest, that increases rapidly, we also refer to the map of Blath [16] for a structured overview.

The spatial spread of populations and their forward and backward genealogies are studied intensively in the context of *spatial* Λ*-Flemming-Viot processes* introduced by Etheridge and co-authors [8, 73]. We also mention the class of models relying on the partial differential equation introduced separately by Fisher [81] and Kolmogorov, Petrovsky and Piscunov [118]. These are utilized in particular to model how the spatial habitat is populated. We refer to [122] for a recent applied and to [31] for a more theoretic approache.

Eventually, we like to emphasise that a non-trivial direction is to include the effects of ageing and memory [161]. One approach are the *Hawkes processes* [100], which by construction keep track of (parts of) their history. These are used not only in evolution theory but for example helpful to model the intercellular network of neuron activation [65]. Another interesting strategy is presented in [93], where the authors define a structured seed bank that allows to construct non-exponential (even heavy-tailed) wake-up times but preserves the Markovian evolution.

Of course, this has to be regarded as a non-exhaustive enumeration. As mentioned before life-science and mathematical modelling of biological phenomena is a topic of recent interest and evolves quickly in many directions.

3. Individual-based models of adaptive dynamics

3.1. Basic model

In this thesis we study the evolution of an asexually reproducing population. To this aim, we use a individual-based model of adaptive dynamics, that is introduced subsequently. These models build the foundation for the work presented in Appendix A. In Section 3.5 we explain how we extend the model to the changing environment investigated in Chapters B and C. Note that existence of the jump-processes has been shown by Fournier and Méléard [83] via an explicit pathwise construction of the locally regulated process.

Evolution is observed as macroscopic changes of the population, but its basic mechanisms (heredity, mutation and selection) act on the microscopic level of the individuals. We therefore describe the evolving population by finite but varying number of interacting individuals, each of them characterized by its traits. These can be interpreted as its geno- or phenotype and are taken form a finite, discrete trait space V, which in our case comes as vertex set of a directed graph $G = (V, E)$. The set of edges in E marks possible mutations, meaning that $(v, w) \in E$ if and only if an individual of trait w can be obtained from a v-individual by a single mutation step.

To each trait we associate a number of biological parameters that describe the dynamics of the system:

- $b_v \in \mathbb{R}_+$, 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. This prevents the process from explosion and is also biologically meaningful since individuals of the same trait compete for the same resources.

The original construction of the stochastic individual-based model by Fournier and Méléard [83] represents the population as a measure-valued process composed of Dirac-measures for every individual alive. This brings the advantages to be able to deal with traits from an arbitrary (Polish) space without further complications. However, if the trait space is finite as in this thesis, it is convenient to reduce to a vector-valued process $(N(t))_{t\geq 0}$ taking values in the cádlág space $\mathbb{D}(\mathbb{R}_+, \mathbb{N}_0^V)$, where $N_v(t)$ denotes the number of individuals of trait $v \in V$ that are alive at time $t \geq 0$. The Markov process is characterised by its infinitesimal generator

$$\begin{aligned}\mathcal{L}\phi(N) = &\sum_{v\in V}(\phi(N+e_v)-\phi(N))\left(N_v b_v(1-\mu)+\sum_{w\in V}N_w b_w \mu m_{w,v}\right)\\ &+\sum_{v\in V}(\phi(N-e_v)-\phi(N))N_v\left(d_v+\sum_{w\in V}c_{v,w}N_w\right),\end{aligned} \tag{3.1.1}$$

acting on measurable and bounded functions $\phi : \mathbb{N}_0^V \to \mathbb{R}$. Here, e_v denotes the unit vector at $v \in V$.

Assuming boundedness of the parameters b, d, c and m, which becomes trivial for the finite trait space, one can deduce the existence and uniqueness in law of the process [83]. In fact, Fournier and Méléard give an explicit pathwise construction of the process $(N(t))_{t\geq 0}$ in terms of Poisson random measures. Since we make extensive use of this representation to introduce couplings of the original process to classical birth death processes [4], we recall this description in Sections A.4 and C.4. The process can also be obtained by an explicit algorithmic construction for jump processes (cf. [26, 75]). This can also be seen as a instruction how to simulate the population numerically in terms of a so-called Gillespie algorithm [90].

3.2. Scaling parameters

The stochastic individual-based model presented in the previous section is designed to emulate the basic mechanism of eco-evolution (heredity, mutation and selection). Given this model, we can follow two distinct paths to deduce meaningful information. One approach is to statistically fit the model to some application system by estimating realistic input data (trait graph, parameters, initial condition,...) to eventually run numeric simulations. To not just get a single possibly unlikely realisation one has to evaluate the model multiple times and find the a meaningful statistics to summarize the features of interest. However, simulations of individual-based models can be computationally heavy, which is impractical for applications. In particular, in our situation this seems not to be promising, as we are interested in the long-term evolution of the system and would need to run the algorithm for many iterations.

Instead we aim to analyse the process with mathematical methods, more precisely we prove a number of limit theorems that provide information on the evolution of the population and the key points, where the stochastic fluctuations can have a huge impact on the output. To this end, we introduce three scaling parameters, namely we control the *population size* by the scaling parameter K, the *mutation probability* by the scaling parameter μ and the *mutation step size* by the scaling parameter σ.

In the following, we outline how these are implemented. The limit of large populations ($K \nearrow \infty$) is our central scale of interest and leads to the ecological ODE-system of the Lotka-Volterra equations. For the limit of rare mutations ($\mu \searrow 0$), we distinguish various regimes in relation to the large population limit and the time scales involved. To complete the picture of the main scaling parameters we shortly introduce the limit of small mutation steps ($\sigma \searrow 0$), despite it not being considered in the main part of this thesis. Nevertheless, we comment on to different approaches that lead to the evolutionary ODE, which is the so-called *canonical equation of adaptive dynamics* (CEAD).

3.2.1. Large populations

Since the aim is to study evolution, which can be characterised without further specification as the changes in the composition of a heterogeneous population on the macroscopic scale, it is natural to study populations constituting of many interacting individuals. Evolution should not be effected too much by the specific behaviour of single representatives of a species and these random fluctuations are expected to balance if the total population is large enough. This is exactly what we observe in the limit of large populations.

The respective scaling parameter K is integrated by replacing the former competition kernel by the rescaled version

$$c^K_{v,w} := \frac{c_{v,w}}{K}, \tag{3.2.1}$$

referring to the modified generators by $\mathcal{L}^K$ and the respective processes by N^K, for $K \in \mathbb{N}$. If we take a closer look at the generator given in (3.1.1) it becomes clear why this leads to populations with size scaling linearly with K. Reducing to a monomorphic population, i.e. the trait space $V = \{v\}$ is just a singleton, the process is nothing but a logistic birth death process, which finds its equilibrium at $\bar{n} = (b-d)/c$. Decreasing the competition by the factor K^{-1} implies a linear increase of the expected population size. Motivated by the biological perspective, the parameter K is often called *carrying capacity*, as it can be interpreted as the capacity of the environment to support living organisms. The more resources (space, nutrients, light,...) are available, the less is the competitive pressure between individuals.

In fact, we can make the considerations on the balance of stochastic fluctuations mentioned above rigorous. Fixing a compact time interval $[0, T]$ and assuming convergence of the rescaled initial conditions $N^K(0)/K \to n(0) \in \mathbb{R}^V_+$, Ethier and Kurtz [75, Chap.11, Thm.2.1] provide a law of large numbers). The rescaled processes N^K/K converge almost surely to the solution $(n_v(t),\ v \in V,\ t \geq 0)$ of the Lotka-Volterra equations with mutations:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_v(t) = \left(b_v(1-\mu) - d_v - \sum_{w \in V} c_{v,w} n_w(t) \right) n_v(t) + \mu \sum_{w \in V} b_w m_{w,v} n_w(t), \quad \forall v \in V,\ t \geq 0. \tag{3.2.2}$$

These equations and their solutions are objects of intensive analysis on their own. Because of its deduction from the large population limit, this regime is sometimes referred to as *infinite population*. It is of particular interest to study existence and stability of equilibrium points. Anticipating the rare mutation limit, we can take $\mu = 0$ and ignore the last term of (3.2.2), which accounts for mutations (cf. Section 2.2). Moreover, it is often useful to not consider the full system but restrict the search for equilibria to as subset of traits $\mathbf{v} \subseteq V$. Associated to this, we call $\bar{n} \in \mathbb{R}_+^{\mathbf{v}}$ an equilibrium state of the mutation-free Lotka-Volterra equations if and only if

$$\left(b_v - d_v - \sum_{w \in V} c_{v,w} \bar{n}_w \right) \bar{n}_v = 0, \quad \forall v \in \mathbf{v}. \tag{3.2.3}$$

Note, that for both, the components of the solution to the differential equation and the equilibria, we only consider biologically reasonable values in $\mathbb{R}_+$. If the equilibrium has value 0 for some trait, we sometimes say that this trait does not belong to the equilibrium. On the other hand 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 $|\mathbf{v}| \geq 2$ and there exists a equilibrium $\bar{n}(\mathbf{v}) \in \mathbb{R}_{>0}^{\mathbf{v}}$, we say that the traits of $\mathbf{v}$ can *coexist* and call $\bar{n}(\mathbf{v})$ the *coexistence equilibrium*. The origin $\bar{n} \equiv 0$ is of course always an equilibrium, but in almost all relevant cases it is unstable, because of some traits being able to survive (i.e. $b_v > d_v$). Under the assumption of positive self-competition (i.e. $c_{v,v} > 0$), for all traits, the driving vector field builds a kind of barrier if the population becomes too large. To be more precise, far away from the origin the vector field points towards 0. This ensures boundedness for solutions to 3.2.2 and we can deduce existence, uniqueness, and continuity from Lipschitz continuity of the coefficients.

Although there is a good understanding in dimensions $d = 1$ and $d = 2$ [40], the analysis of the Lotka-Volterra equations becomes highly non-trivial in dimensions $d \geq 3$ (cf. [172]).

Let us close the discussion for the carrying capacity by introducing some terminology for the stochastic model. In the sequel we refer to subpopulations with a size of order K as *macroscopic*, while we call populations with a size of order 1 *microscopic*, and intermediate sizes of order strictly between 1 and K *mesoscopic*. The macroscopic traits that are close to their joint equilibrium size are called *resident*.

3.2.2. Rare mutations

It is an ongoing scientific discussion what is the actual frequency of mutations [**?**, 166]. On the one hand the mutation rates per base pair per cell division is generally estimated to be of order 10^{-9} (cf. [33, 110]). In relation to approximately 3×10^9 base pairs in the human genome for example, this suggests a large probability of finding a genetic variation after cytokinesis. There are various works considering a fixed mutation probability [42, 44]. On the other hand a large part of the genome is non-coding. Thus mutations at irrelevant positions may be much more frequent, but do not change the phenotype. Focussing only on mutations that actually affect the phenotype, suggests to discuss several regimes of rare mutations.

If we simply take the limit $\mu \searrow 0$ in the stochastic system, it converges to the mutation-free stochastic model. This case becomes relevant if only ecological forces are aimed to be studied as in the theory of population dynamics. However, evolution is only possible if mutations occur.

Therefore, a first approach would be to study the deterministic Lotka-Volterra system in the limit $\mu \searrow 0$. As this means a positive mutation probability in an “infinite population”, we interpret the regime still as *frequent mutations*. As a consequence Wang, Bovier and Kraut show in [32, 120] that in finite time every trait in the connected component of the trait space is populated independent of its mutation distance (i.e. the smallest number of mutation steps needed in the trait graph) to the resident population. Under rescaling of time by $\ln(1/\mu)$ they show convergence to a deterministic jump process that can be connected to theory of adaptive walks or flights, respectively. To bound the mutation range, they study in addition a modified differential equation with an artificial lower threshold of population densities that can proliferate, to ensure address the problem of population sizes being discrete. As a consequence this bounds the number of maximally accumulated mutations. Its natural stochastic correspondence we meet again below in the context of fitness valleys.

To study really a small mutation probability it is inevitable to consider a simultaneous limit of $\mu_K \searrow 0$ jointly with $K \nearrow \infty$. It was observed first that under the so-called *Champagnat scaling*,

$$\mathrm{e}^{-CK} \ll \mu_K \ll \frac{1}{K \ln K}, \tag{3.2.4}$$

for some constant $C \in (0, \infty)$, mutation events are rare and well separated in the following sense: The lower bound guarantees the occurrence of mutants before the resident population deviates too much from the equilibrium. After a mutant population of a new trait is founded by a single individual, there is enough time either to fixate and grow to a macroscopic size or to go extinct before the next mutation occurs. The dominating time scale (see Section 3.4) is then $1/K\mu_K$ and it has been shown the convergence of the macroscopic populations to Markov jump processes, the *trait substitution sequence* and the *polymorphic evolution sequence*, respectively [40, 43]. Due to the regime of *very rare mutations*, these cannot be accumulated. Instead from all possibly fit mutants only the nearest neighbours of the residents are reached and get the chance to invade.

The accumulation of mutations before invading the resident becomes possible, when we slightly increase the mutation probability, while still considering a simultaneous limit $\mu_K \searrow 0$. Therefore, we introduce the power-law mutation probability

$$\mu_K = K^{-\frac{1}{\alpha}}, \tag{3.2.5}$$

for some parameter $\alpha \in (0, \infty)$, which is the main regime analysed in this thesis. Note that for $\alpha \in (0, 1)$ this coincides with the Champagnat scaling, but for $\alpha > 1$ mutations are more likely and we thus refer sometimes to this regime as the scaling of *moderately rare mutations*. Depending on the parameter α, we observe multiple mutation steps within time of order 1. Due to consecutive thinning by the mutation probability one can show, that the population size of a mutant trait w reached within $d(\mathbf{v}, w)$ mutation steps is of order $K\mu_K^{d(\mathbf{v},w)}$. Hence, we can make out a neighbourhood $V_\alpha(\mathbf{v})$ around the residents, comprising all traits of mesoscopic size. This allows now to also use non-beneficial mutations to be used as intermediate steps towards a trait of higher fitness. At the same time this phenomenon, which we call the crossing of a fitness valley, is limited to a number of $\lfloor \alpha \rfloor$ accumulated mutations. The individual-based model was studied under this regime first by Bovier, Coquille and Smadi. In [29] they show first that short fitness valleys of length $L < \alpha$ are crossed in a deterministic time of order $\ln K$ (cf. Section 3.4). Secondly larger valleys of length $L > \alpha$, can also be crossed, but on a much larger time scale $1/K\mu_K^L$ and this time is still random. For more general trait graphs he first case has been elaborated in [48], while the large fitness valleys are part of the present thesis.

3.2.3. Small mutation steps

In view of Darwin's concept that evolution happens gradually and incrementally, it is also worthwhile to consider a continuous trait space (e.g. compact subsets of $\mathbb{R}^n$) instead of the discrete trait graph. The third scaling parameter σ that we can hence integrate into the model is hence the mutation step size or the size of the evolutionary advantage of a mutant. The idea of only small changes roots back to diffusion models Fisher already derived in [80]. Mathematically we implement this by defining a scaled mutation kernel such that the mutation distance is proportional to the scaling parameter, i.e. given a probability kernel $M(x, \mathrm{d}y)$, we set

$$\int_V M^\sigma(x, \mathrm{d}y) f(y) \mathrm{d}y = \int_V M(x, \mathrm{d}y) f(\sigma y) \mathrm{d}y, \tag{3.2.6}$$

for all $x \in V$ and all bounded and measurable functions $f : V \to \mathbb{R}$. If a mutation occures at a birth event of an individual of trait $x \in V$, the new mutant carries the trait $x + y$, where y is choosen randomly with distribution $M^\sigma(x, \mathrm{d}y)$.

In the theory of adaptive dynamics, this scaling goes along with the important concept of the canonical equation of adaptive dynamics (CEAD). Heuristically motivated by the biological assumptions of large population and rare mutations with small effects and the assumption that no two different traits can coexist, Dieckmann and Law [61] introduced this ODE. Beside the biological deduction, there are also different rigorous scaling limits of the individual-based model, which lead to the CEAD to encode the limiting dynamics [7, 43]:

$$\dot{x}_t = \int_V h \left[h m(x_t) \bar{n}(x_t) \partial_1 f(x_t, x_t) \right]_+ M(x_t, \mathrm{d}h). \tag{3.2.7}$$

The equation describes the continuous evolution of the trait of a monomorphic population in a fitness landscape driven by the contrary forces of mutation and selection expressed in the fitness gradient. The notion of fitness $f(v, w)$ used in the equation is introduced in Section 3.3. The evolution finally comes to a hold, when approaching an *evolutionary singularity*, i.e. a local extreme value in the fitness landscape. In case of a local minimum one can make some subtle conjectures about evolutionary branching and coexistence.

3.3. The notion of invasion fitness

Talking about fitness valleys it is inevitable to discuss the different notions that have been established in mathematical biology. The aim is in general to quantify the ability of subpopulations to survive and establish in a reasonable manner, respectively. One can distinguish the *individual fitness* from the concept of *invasion fitness* that is assigned to a trait. While the first one is a fixed value describing how well a trait is adapted to its environment without taking into account the other species alive, the latter one takes into account these ecological quantities by integrating the interactions between individuals. These questions have been addressed in the beginning by Metz and co-authors [141, 142] and made precise for the individual-based model by Champagnat and Méléard [40, 43].

Following their approach, one can define, for each trait $v \in V$ the individual fitness by $r_v := b_v - d_v$. This notion is related to classical pure birth-death-processes, if we ignore competition-induced damping effects. Thus r_v reflects the exponential growth rate of such processes as well as of the corresponding deterministic systems.

To make the advantages of the adaptive dynamics approach work, it makes sense to include the ecological population state into a evolutionary meaningful dynamic notion of fitness. This comes somehow natural in the model under consideration by taking care of the competition as interaction terms between individuals. Moreover, let us utilise the idea of adaptive dynamics that the population is close to an equilibrium when a new mutant appears. Then we define for a single individual of trait $w \in V$ within a bulk population of traits $\mathbf{v} \subset V$ close to their equilibrium $\bar{n}(\mathbf{v})$ its invasion fitness by

$$f_{w,\mathbf{v}} := b_w - d_w - \sum_{v \in \mathbf{v}} c_{w,v} \bar{n}_v(\mathbf{v}). \tag{3.3.1}$$

In view of (3.2.2) setting $\mu = 0$, this can again be seen as the exponential growth rate but only in a initial stage. If this value is positive the newly founded subpopulation can and invade the equilibrium population, we therefore refer to $f_{w,\mathbf{v}}$ as invasion fitness. Based on the sign of this invasion fitness of a trait, we can approximate its population either by subcritical or supercritical branching processes.

In the case of equal competition, i.e. $c_{v,w} \equiv c$, for all $v, w \in V$, the two fitness concepts are strongly connected by the relation $f_{w,v} = r_w - r_v$. Hence, r_v can be interpreted as *absolute fitness*, whereas $f_{w,v}$ is now the *relative fitness.* Thus it suffices to give a fitness landscape by the fixed individual fitnesses and a trait w can invade the population of trait v if and only if $r_w > r_v$. For broader discussions see [58], where this is connected to the theory of adaptive walks and flights.

3.4. Time scales

One of the biggest advantages of adaptive dynamics is to study both ecological and evolutionary dynamics of populations in the same model. However this does not necessarily mean to observe those simultaneously, but they mostly occur on different time scales. On the one

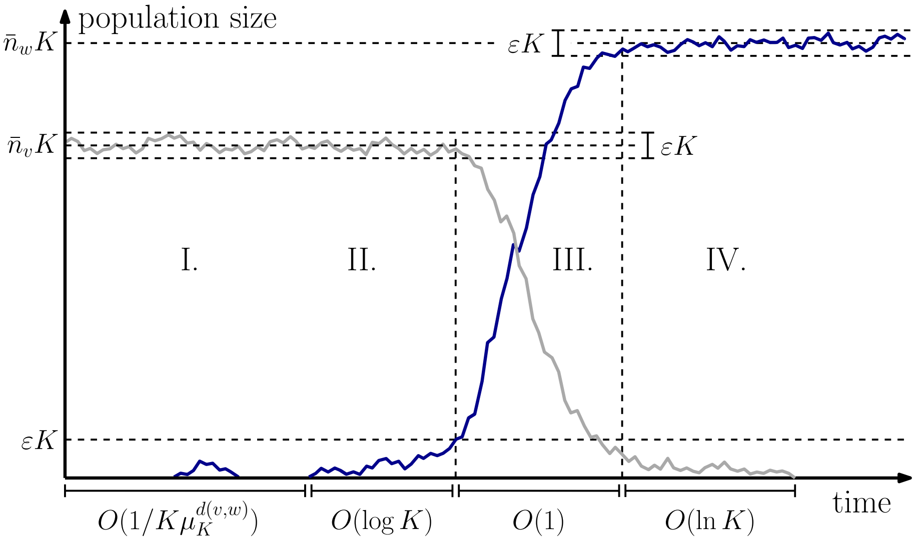


Figure 3.1.: The four phases of the invasion of a mutant trait (blue) in a resident trait (grey) at a distance of $d(v,w)$ mutation steps.

hand the population process covering a certain set of traits of macroscopic population size evolves fast towards a ecological equilibrium without changing the traits being represented. On the other hand the interplay of rare mutations and selection allows for a better adaptation to the (outer) environment in the long run by evolution. This intuitive splitting of time scales also manifests in the individual-based model introduced above and was first shown by Champagnat in [40]. In the sequel we give an overview of the different time scales being involved, as a careful analysis shows these are more than two. We explain them in the order of occurrence in a invasion of a new trait and give some remarks on the mathematical tools utilised. Figure 3.1, which is an adapted version of Figure 1.2 in [121], provides a schematic picture of the population sizes at each phase of the invasion.

3.4.1. Stability of residents

For simplicity let us consider the population process to start near a ecological equilibrium for the traits supporting the macroscopic population. If this was not the case, we could first study the equilibration-phase described below. The key tool to control the dynamics of all other individuals is to show that the macroscopic subpopulations stay close to this equilibrium for a very long time, unless a new trait reaches a macroscopic size. This goes perfectly in line with our biological intuition on adaptive dynamics. This is possible, since we can interpret the subprocess as random perturbation of the underlying ODE-system (2.2.2). Under use of the theory of Freidlin and Wentzell [86] one can establish a *large deviation principle* [66], which yields stability for a exponential long time $\exp(VK)$, for some $V > 0$. In [7] the authors consider a *potential theoretic* approach for the long term stability. We adapt the techniques in the Appendices B and C to show strong results on the speed of convergence.

3.4.2. Arrival of mutants

Since almost stability of the resident population is now warranted, we use an approximation by Poisson point processes to monitor the occurrence of mutants. There are order K-many residents, which each proliferate with linear rate. With probability μ_K a mutation occurs at a birth event and the child carries a different trait. Hence first order mutants (i.e. nearest neighbours to the residents) are produces at a rate of order $K\mu_K$. These individuals themselves produce mutants at a rate of order $K\mu_K^2$ of second order mutants. In general, mutants of trait w are produced at a rate of $K\mu_K^{d(\mathbf{v},w)}$, where $d(\mathbf{v},w)$ is the graph distance of w from the resident set $\mathbf{v}$, i.e. the length of the shortest directed path in G connecting $\mathbf{v}$ to w. Hence, the arrival time for a w-mutant is of order $1/K\mu_K^{d(\mathbf{v},w)}$ and approximately exponential distributed. Depending on the scaling of $\mu_K \searrow 0$ we can distinguish a neighbourhood of $\mathbf{v}$ where $1/K\mu_K^{d(\mathbf{v},w)} \ll 1$, which means that in finite time there are order $K\mu_K^{d(\mathbf{v},w)}$ many mutants present. In our case of the power-law mutation probability $\mu_K = K^{-1/\alpha}$ this is equivalent to $d(\mathbf{v},w) < \alpha$ and we refer to this neighbourhood by $V_\alpha(\mathbf{v})$. Outside of this region, mutants arrive separately on a the diverging time scale $1/K\mu_K^{d(\mathbf{v},w)} \gg 1$.

3.4.3. Mesoscopic growth phase

As long as a mutant does not population become macroscopic, it can be well approximated by classical birth-death-processes. Depending on the sign of the invasion fitness $f_{w,\mathbf{v}}$ these branching processes are either super- or subcritical. In the later case one can use the results of Harris, Atherya and Ney [3, 4, 99] to show extinction in finite time. For supercritical branching processes, the same theory gives us the dichotomy of extinction or fixation with unbounded exponential growth. Moreover, the *probability of fixation* is quantified explicitly by $(f_{w,\mathbf{v}}/b_w)_+$. Because of the exponential growth it takes a time of order $\ln K$ for a finite population to reach a macroscopic size εK. The exact growth behaviour has been studied in [29, 48, 158] a detailed description is provided in the Appendix of [45]. Note that we also use and refine these results for different scales in Appendix B according to the changing environment.

3.4.4. Macroscopic (re-)equilibration

Once a mutant population eventually reaches a macroscopic size εK one follows the early approach of Fournier and Méléard [83] to establish a *law of large numbers* utilising the framework of Ethier and Kurtz [75, 124]. Rescaling the population size by K gives convergence on finite time intervals to the mutation-free Lotka-Volterra equations 2.2.2 with $\mu = 0$. Note that it is crucial, that the initial conditions for the Lotka-Volterra dynamics are of order K, i.e. non-vanishing when rescaled by K, in order to bound the adaptation time uniformly by a time of order 1. Hence, also the stochastic system undergoes a rapid change and gets close to its new (ecological) equilibrium.

Due to the different traits being resident and we have to consider the invasion fitness with respect to the new equilibrium from now on. In particular the former residents may not be fit any longer and thus their population size declines as a subcritical birth-death-process on the $\ln K$-time scale. This is depicted in the last phase of Figure 3.1.

3.5. Changing Environment

The key feature we add to our model and investigate in Chapters B and C is a changing environment. Our biological motivation is rooted in the simple fact, that the outer forces shaping ecology and evolution never stand still, but underlay permanent changes on various scales ranging from dramatic singular catastrophes, like the impact of a meteorite, which implied the extinction of dinosaurs, to (semi-)regular fluctuations like seasons or the day-night-cycle. Out of the different scenarios there have been investigated different approaches to implement these environmental changes into mathematical models. While some works consider more or less random fluctuations in the environment [19, 64, 111, 125, 134], others study a shift or a permanent drift [87, 107, 155] or similar to our approach, periodic changes [37, 145, 160].

Motivated by medical treatment by a regular drug supply, the extension we introduce and analyse in the following is a deterministic, periodically changing environment. This can reflect for example the fluctuations of drug concentration in the blood induced by a regular treatment protocol. For simplicity consider a fixed number ℓ of phases with individual length $T_i > 0$, for each $i = 1, \dots, \ell$. By $T_j^\Sigma := \sum_{i=1}^j T_i$ we refer to the endpoints of these phases. In each phase, we assume the parameters to be constant, so we take birth, death and competition rates $b_v^i, d_v^i, c_{v,w}^i$, for all $i = 1, \dots, \ell$ and all $v, w \in V$. Then we define the time-dependent birth rates as the periodic extension of

$$b_v(t) := \sum_{i=1}^{\ell} \mathbb{1}_{t \in [T_{i-1}^\Sigma, T_i^\Sigma)} b_v^i, \tag{3.5.1}$$

and analogously for the death rates $d_v(t)$ and competition rates $c_{v,w}(t)$.

If we would simply take these parameter functions to define the population process, this corresponds to a quickly changing environment since it takes place in time of order 1. Instead we introduce the scaling parameter λ_K and set

$$b_v^K(t) := b_v(t/\lambda_K), \quad d_v^K(t) := d_v(t/\lambda_K), \quad \text{and} \quad c_{v,w}^K(t) := c_{v,w}(t/\lambda_K) \tag{3.5.2}$$

to control the speed of environmental changes. The aim is to make these happen slow enough such that the resident populations can adapt, but fast enough such that they influence the growth of mutants. In view of the previous section we choose

$$1 \ll \lambda_K \ll \ln K \tag{3.5.3}$$

as an intermediate scale and define for each $K \in \mathbb{N}$ the time dependent generator of the population process

$$\begin{aligned}\left(\mathcal{L}_t^K\phi\right)(N) = &\sum_{v\in V}(\phi(N+e_v)-\phi(N))\left(N_v b_v^K(t)(1-\mu_K)+\sum_{w\in V}N_w b_w^K(t)\mu_K m_{w,v}\right)\\ &+\sum_{v\in V}(\phi(N-e_v)-\phi(N))N_v\left(d_v^K(t)+\sum_{w\in V}\frac{c_{v,w}^K(t)}{K}N_w\right).\end{aligned} \tag{3.5.4}$$

Again it acts on the measurable and bounded functions $\phi:\mathbb{N}_0^V\to\mathbb{R}$ and e_v denotes the unit vector at $v\in V$.

A similar model is introduced in [49] to workout the interplay between changes in the environment and the effect of dormancy on speciation. In contrast to our constraints, they assume very slow changes, i.e. $\lambda_K \gg \ln K$. The consequence of this is that the mutants can grow to a macroscopic size within a single phase of constant parameters and hence invasions add up to a drift. The accelerated changes of our model instead lead to an inter-phase averaging of the key quantities and we can deduce some effective growth rates as a weighted average.

Of course, this comes by the price of some involved technicalities in the growth analysis. For ones, the frequent parameter changes entails a permanent adaptation of the resident population to its phase-dependent equilibrium. We therefore cannot directly use the stability analysis of Freidlin and Wentzell explained in the previous section. But we need good control on the speed of convergence of the probability to get too far from the equilibrium. For the same reason we improve the convergence result of Ethier and Kurtz partially.

Focussing on the mutant growth the changing environment causes the need to extend the approxiamtion theory of branching processes of Champagnat, Méléard and Tran [45] on both time scales. In the time horizon of mutant growth, $\ln K$, we show the averaging effects. In particular the extinction probability turns out to be of non-trivial matter. At the same time we achieve strong estimates on the growth and decline of a population during a single period. In combination with the fitness valley the later one turns out to be crucial in the detailed analysis of a pit stop (cf. Section 4.4).

4. Outline and main results of the thesis

The main body of this thesis is structured into three interconnected chapters, each of which can be read independently, though their contents are related. Appendix A focusses on the analysis of the long-term behaviour of the model of Chapter 3 for a fixed (outer) environment. In Appendix B and C we decode the effective influences of a periodically changing environment. While the first two consider a general trait graph, the results of the last part restrict to a particular setting reflecting the condensed version of a fitness valley. In the following give some remarks on the genesis of the questions addressed in the thesis and summarise the content and main results of the three chapters.

4.1. Historic placement and motivation

The individual-based model defined in Section 3.1 has been the object of intensive analysis over the last two decades. Step by step different authors were able to work out the interplay of ecology and evolution on different time scales and for different mutation regimes.

Building on the preliminary works and fundamental ideas of Metz, Geritz, Bolker, Pacala, Dieckman, Law and coauthors in the 1990s [24, 25, 61, 62, 141, 142] as mentioned before, Fournier and Méléard introduced the model rigorously in [83]. Moreover, they already proved convergence to the deterministic Lotka-Volterra-system and the respective equilibria in the $O(1)$-time scale.

Under the strong assumption of very rare mutations, Champagnat worked out the splitting of time scales between ecology and evolution and described the several phases of invasion we discuss in Section 3.4 in [40]. Therein and later under collaboration with Méléard in [43], it was given a shown convergence of the population process on the time scale $1/K\mu_K$ to the so-called *trait substitution sequence* (TSS) and the *polymorphic evolution sequence* (PES), respectively. Both are Markov jump processes with the state space consisting of the Lotka-Volterra equilibria between different traits. Due to the regime of very rare mutations, the limiting processes are driven by fit mutant traits only in the direct neighbourhood of the present resident traits.

After Smadi gave some insights on the exponential growth of mutant populations in [158] and together with Billard in [14], the joint work of her, Bovier and Coquille [29] worked out the splitting of invasion time scales for the power-law mutation regime for a simplified linear trait space. As explained in Section 3.4.2 traits that are close enough to the residents (by means of mutations to be accumulated) are immediately present with a non-vanishing number of individuals. Therefore, fit subpopulations start growing approximately exponentially and reach a macroscopic size in time of order $\ln K$. Otherwise, if there is no fit mutant in the accessible neighbourhood, the process gets stuck in a so-called *evolutionary stable condition* (ESC). In order to escape one has to wait for a fit mutant to appear outside of the accessible neighbourhood. This requires a time of order $1/K\mu_K^L$, where L is the length of the so-called *fitness valley* to be crossed,i.e. the mutational distance to the fit mutant trait. This goes perfectly in line with the former results for the TSS and PES when setting $L = 1$. Since on a macroscopic level, the process stays stable for a long time, while the actual change takes place comparably fast, this phenomenon can be seen as a characteristic of *metastability*.

While a general description of the evolution on an arbitrary trait graph for the $\ln K$-time scale was achieved by Coquille, Kraut and Smadi in [48], the corresponding result for the metastable transitions between ESCs was still missing. Appendix A aims at closing this gap by investigating fitness valleys in general trait graphs. Considering the model extension of a changing environment introduced in Section 3.5, Appendix B recreates the general picture on the $\ln K$-time scale and Appendix C analyses different cases of crossing a fitness valley on a simplified trait graph.

4.2. A general picture of adaptive dynamics on multiple scales

Appendix A studies the individual-based model for fixed environments, where the trait space is given by an arbitrary finite directed graph $G = (V, E)$. We investigate the full picture of jumps between evolutionary stable conditions (ESC) in two main results. We first construct the *metastability graph* $\mathcal{G}_{\text{ESC}}$ with ESCs as vertices, and describe the jumps on varying time scales depending on the stability of the ESC. Fixing a particular time scale then yields the converges to a Markov jump chain on the so-called *L-scale graph* $\mathcal{G}^L$, which is a collapsed version of the former one. This appendix was published in the Journal of Mathematical Biology as joint work with Anna Kraut [69],

> M. Esser and A. Kraut, *A general multi-scale description of metastable adaptive motion across fitness valleys.* Journal of Mathematical Biology, Volume 89, Number 46, 2024.

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

Metastability is a phenomenon usually known from the field of (statistical) physics and chemistry [30]. It describes the situation of systems to seem stable on short time scales and besides small excursions to return to the same stable state very often. In the long-term run, however, these systems can be triggered to escape from this trap and rapidly evolve towards another stable state. Often one can characterize these metastable states as local minima in the energy landscape and an escape is possible when a critical energy threshold is reached. The seminal work [29] laid the basis for understanding also evolution as a metastable process.

In Appendix A, we follow this perspective and establish the rigorous theory of metastability for evolution on general trait graphs. While the physical notion of stability is closely related to the (absolute) energy landscape, this translates to the shape of the (dynamic) invasion fitness landscape in the biological world. To each evolutionary stable condition $\mathbf{v} \subseteq V$, we associate a stability degree $L(\mathbf{v})$, which is basically the radius of the surrounding fitness valley when traits $\mathbf{v}$ are resident. The critical event to trigger an escape is the birth and fixation of a fit mutant at the boundary of the valley. The fast evolution to the next ESC is then given by the algorithm of [48] on a $\ln K$-time scale. The time scale to observe the metastable transitions $1/K\mu_K^{L(\mathbf{v})}$ depends on the degree of stability and hence varies between the ESCs. We illustrate this peculiarity in two different results.

In Theorem A.2.7 we show that the exit time of an ESC is approximately exponentially distributed and we obtain the exact rates $R(\mathbf{v})$. Moreover, if multiple equidistant mutants can trigger the escape, these are chosen randomly and we compute the respective probabilities. Considering the ESCs as the vertices of a new graph $\mathcal{G}_{\text{ESC}}$, where directed, weighted edges indicate possible transitions, we can describe the trajectory of evolution in the limit of large populations $K \nearrow \infty$ as a multi-scale jump chain on this metastability graph. This can be related to the concept of *adaptive walks* or *adaptive flights* (cf. [123]). Those are stochastic processes that directly study the motion of the macroscopic population on the trait space, focussing on successful invasions and omitting the microscopic dynamics.

In contrast to this general picture spanning multiple time scales, we present another perspective in Theorem A.2.11 Fixing a level of stability L and focussing on a particular timescale $1/K\mu_K^L$ one notices that on the one hand larger valleys $L(\mathbf{v}) > L$ cannot be crossed, while smaller valleys $L(\mathbf{v}) < L$ are crossed immediately. This implies, that edges connecting ESCs of lower stability get concatenated and contracted, but those starting in ESCs of higher stability get removed. The metastability graph hence collapses to a L-scale graph $\mathcal{G}^L$. And we can deduce convergence to a Markov jump process $\mathcal{N}^L$. Since the construction of the meta-graphs $\mathcal{G}_{\text{ESC}}$ and $\mathcal{G}^L$ from the fitness landscape on the original trait graph is a bit involved, we provide a detailed explanation in Section A.3 together with some examples, which exhibit an interesting behaviour.

Particular difficulties we have to overcome in the proofs lie in the fact, that circles and backwards mutations are generally allowed in our trait graph. Heuristically it is clear, that these do not have a valuable effect on the exit times, because the shorter a mutational path is, the fewer mutations need to occur at respective small probability μ_K. We use an inductive approach over the oriented graph distance from the resident population to make the principle of the shortest paths rigorous.

A remarkable insight we gain from the complete analyses is that besides Darwin's principle *survival of the fittest*, there comes additionally the *survival of the first.* By this we mean, that from the possible mutants, which could trigger the exit from ESC, it is the first one, who makes it. This is not necessarily the fittest one.

4.3. The influence of environmental changes on the growth of mutant subpopulations

In Appendix B, we investigate the growth of emergent mutants and their invasion of the resident population in the periodically changing environment introduced in 3.5 under power law mutation probability $\mu_K = K^{-1/\alpha}, \alpha > 0$. Since the growth phase is affected by frequent changes of the parameters, we are able to extract the effective speed of growth as an averaging of the phase-dependent invasion fitnesses. In addition to a mesoscopic scaling limit of the orders of population sizes, we deduce a limit result for the sequence of consecutive macroscopic resident traits that is similar to the so-called trait-substitution-sequence. This appendix is to appear in Stochastic Processes and their Applications as joint work with Anna Kraut [71],

> M. Esser and A. Kraut, *Effective growth rates in a periodically changing environment: From mutation to invasion.* Stochastic Processes and their Applications, Volume 184, Number 104598, 2025.

Appendix B contains the accepted manuscript, with only minor changes to correct some typing errors and adapt the layout to the format of this thesis.

We give a full characterisation of the limiting process under rescaling of time by $\ln K$. Theorem B.2.4 provides the precise description of the orders of the different subpopulation sizes in terms of the exponents $\beta_v^K(t) := \ln(N_v^K(t \ln K) + 1)/\ln K$. When K tends to infinity these converge to deterministic piecewise linear function given by an inductive algorithm. Moreover, in Corollary B.2.5, we depict the “visible” evolution of the population process, i.e. the progression of macroscopic traits that dominate the whole system. The results are parallel to those for the constant environment case in [48]. There are two main differences. First in the algorithmic description of the limiting functions, one has to replace the invasion fitness $f_{w,\mathbf{v}}$, which is the key quantity related to the slopes of the functions, by the so-called *effective growth rate* $f_{w,\mathbf{v}}^{\mathrm{av}} = \sum_{i=1}^{\ell} f_{w,\mathbf{v}}^{i} T_i / T_\ell^{\Sigma}$, which is nothing but the weighted average of the fitness over the different phases of the environment. Second, the characterisation of the macroscopic evolution is limited to determining the supporting traits since the actual value of the equilibrium size fluctuates according to the environmental changes.

The proof relies on the inductive approach of [48], the intermediate time steps introduced there and an analogous approximation by (logistic) birth-and-death processes with and without migration. However, the changing environment faces us with three main technical issues, which necessitate different approaches to be fixed.

First, we require a stability result for the resident population in order to be able to approximate the mutant populations by classical birth-death-processes. However, we cannot simply apply the large deviations approach that is used in the case of constant parameters because due to the changes in the environment, the equilibrium sizes of the resident traits change from phase to phase. Therefore, we must concatenate short times of re-equilibration and longer sections of stability in alternating sequence for all phases. Since the total time horizon under consideration includes an increasing number of phase changes in the limit, it is necessary to get good bounds on the speed of convergence for each approximation as the sum of a diverging number of failure probabilities still needs to vanish in the end. For the re-equilibrations we improved the results of Ethier and Kurtz [75], which were also used in former results. To prove a stability result with sufficiently fast convergence of the probabilities in between parameter changes we decided for a different approach and adapted some results of Baar, Bovier and Champagnat presented in the appendix of [7] using potential theory. Despite these being restricted to monomorphic equilibria they turn out to be very strong and robust.

A second and somehow minor difficulty is to take the swap from the approximation by branching process to the law of large numbers at the right time. When a new emerging mutant becomes macroscopic we have to ensure that the new mutant is in a fit phase to apply the Lotka-Volterra system. We do so by a careful choice of the utilised stopping times, which now include conditions on the fitness, and by establishing results on the short-term growth of the mutant populations.

The third challenge appears in the context of the branching process approximation, where we have to rework the results of the appendix from Champagnat, Méléard, and Tran's paper [45] for the setting of a changing environment. While the adaptation of the results is straightforward, despite needing a lot of notation, the control over small populations, those which newly emerge or are close to extinction, turns out to be particularly different. In the case of a constant environment one utilise the easy extinction probability of simple birth-death-processes several times. Because of the time-dependent parameters, this is no longer a helpful quantity. Instead, we prove estimates on these probabilities by a time-discretisation and an iterative bound for the probability generating function of the corresponding Galton-Watson process, taken from [4].

4.4. The effect of periodic environments on the crossing of fitness valleys

Appendix C extends the considerations of Appendix B by analysing the crossing of fitness valleys in a changing environment. These occur typically on much larger time scales than the growth of mutant populations depending on the length of the valley. By examining the influences of the changing environment carefully on each time scale, we are able to determine the overall crossing rates of fit mutants and their ability to invade the resident population. Building on this, we study an interesting variation, which we call *pit stop*. In this case, an intermediate trait within the valley has temporary phases of positive invasion fitness. We observe a significant acceleration of the traversal of the valley and can precisely characterize the new time scale. This appendix is available as a preprint as joint work with Anna Kraut [70],

> M. Esser and A. Kraut, *Crossing a fitness valley in a changing environment: With and without pit stop.* Preprint, arXiv:2503.19766.

Appendix C contains the preprint, with only minor changes to adapt the layout to the format of this thesis.

In the context of changing environments, there are different notions of fitness. On the one hand for each phase $i = 1, \dots, \ell$ there is the classical phase-dependent invasion fitness $f^i_{w,v} = b^i - d^i - c^i_{w,v}\bar{n}^i_v$ describing the initial growth rate of a new emerging w mutant within a bulk population of trait v near its monomorphic equilibrium under the parameters given for the i-th phase. On the other hand, we determine in Appendix B the effective growth rate $f^{\mathrm{av}}_{w,v} = \sum_{i=1}^{\ell} f^i_{w,v} T_i / T^{\Sigma}_{\ell}$ as the waited average, quantifying the growth of mesoscopic population on the time scale $\ln K$. When defining a fitness valley for changing environments, it turns out to make a huge difference, which of those two quantities we ask to be negative, despite both preventing mutant populations within the valley from growing to macroscopic size. In Appendix C we therefore investigate two different scenarios.

The first and more restrictive assumption is the so-called *strict valley*. For all traits $v \in [\![1, L-1]\!]$ within the valley, we assume the invasion fitness with respect to the resident population at trait 0 to be negative in every phase, i.e. $f^i_{v,0} < 0 \ \forall i = 1, \ldots, \ell$. Note that this immediately implies that also the average fitness $f^{\mathrm{av}}_{v,0}$ is negative. For the trait L at the end of the valley we then assume positive average fitness$f^{\mathrm{av}}_{L,0} > 0$, since this enables the L-trait already to grow to macroscopic size.

Under these fitness valley assumptions we are able to prove a limit result for the crossing time T^K_{inv}, that is similar to [29]. Namely, Theorem C.2.2 shows that the time is approximately exponentially distributed on the time scale $1/K\mu^L_K$ and the *effective crossing rate* R^{eff}_L is an average of the phase-dependent crossing rates R^i_L. In contrast to the effective growth rates of Appendix B one cannot simply take the weighted average to define R^{eff}_L. This is because there are time intervals where mutants of trait L at the end of the valley are born and start growing, due to their positive invasion fitness in this phase ($f^i_{L,0} > 0$). However, if the population was founded too late within the fit phase and in the next phase the trait L is unfit ($f^{i+1}_{L,0} < 0$), the population size at the phase change is not large enough and the L-population goes almost sure extinct. Note, that this situation is possible since we only assumed the average fitness of trait L to be positive.

We deal with this difficulty by a careful study of the initial fixation in a time of order $\sqrt{\lambda_K}$ and the survival over a full period of length $\lambda_K T^{\Sigma}_{\ell}$. Under usage of the probability generating function for time-inhomogeneous branching processes, we characterize a set $A \subseteq [0, \infty)$ of admissible arrival times, that allow for a successful invasion. Using the indicator function of this set as additional filtration finally leads to the proper formula for R^{eff}_L. (cf. C.2.14).

The second and more intricate kind of fitness valley we summarize under the name *valley with pit stop* in Theorem C.2.3. To avoid further complications we restrict the changing environment to only two alternating phases. In contrast to the strict valley, we choose exactly one trait in the more distant part of the valley $w \in [\![\lfloor \alpha \rfloor + 1, L-1]\!]$ which is assumed to have positive invasion fitness in the first phase ($f^1_{w,0} > 0$) but still forced to have negative average fitness ($f^{\mathrm{av}}_{w,0} < 0$). The rest of the valley is of the strict form described in scenario one. To avoid the entanglement of the set A here, the final trait of the valley shall have positive fitness in both phases ($f^1_{L,0}, f^2_{L,0} > 0$).

This little change in the fitness landscape seems harmless but has a significant impact on the speed of traversal of the valley. Formerly, as analysed in [29] a mutant arriving at trait w gave rise to a subcritical excursion, i.e. a subcritical birth-death-process, which is well known to get extinct in finite time almost surely. If a w-mutant is produced in the phase $i = 1$ this is no longer the case because of the positive invasion fitness. Instead, this newly founded population can grow exponentially until the phase changes and it decays until extinction in the second phase. Note that extinction within one period is guaranteed by $f^{\mathrm{av}}_{w,0} < 0$. Since the population grows for a phase of diverging length $1 \ll \lambda_K \ll \ln K$, the peak size is also of diverging order $\exp(c\lambda_K)$, where c depends on the arrival time of the first w-mutant and its fitness. Note that this size is of smaller order, than any power law K^γ.

As there are now many more individuals alive that have the ability to produce mutants of trait $w+1$, this is much more likely than in the purely subcritical case. Therefore, at trait w the former thinning induced by μ_K from the subcritical excursion is replaced by a less restrictive thinning with a probability of order $\exp(c\lambda_K)\mu_K$. Inserted in the complete valley, this yields a significant acceleration of the time scale of traversal. To be precise the impact is the largest if the w-mutant is born at the beginning of the first phase and its descendants produce a mutant of trait $w+1$ at near the phase change. Then the peak size is approximately $\exp(f^1_{w,0}T_1\lambda_K)$. Since the w-mutants arrive almost uniformly the probability for this to happen is of order $1/\lambda_K$. Hence we see an overall speed-up by $\exp(f^1_{w,0}T_1\lambda_K)/\lambda_K$. Moreover, this "typical behaviour" condenses also in the corresponding rate R_L^{pitstop}. Here, averaging of the two phase-dependent crossing rates splits into the part before trait w and thereafter (cf. C.2.17).

In order to make the preceding heuristics rigorous two main challenges have to be addressed. First, we improve the bounds on the stability of the residents. Besides strengthening the estimates on the convergence speed, we also refine the bounding tunnel of allowed fluctuations around the exact equilibrium by replacing the formerly fixed $\varepsilon > 0$ by a vanishing sequence $\varepsilon_K \searrow 0$. The first result is necessary since compared to Appendix B the number of phases to be concatenated increases when enlarging the time scale and hence the time horizon under consideration. The finer tunnel becomes inevitable because the time scale itself depends on the invasion fitness and this quantity and its approximations are strongly connected to the bounds of the fluctuations of the resident population. Unfortunately, this comes at the price of assuming equality of the equilibrium sizes over both phases ($\bar{n}^1_0 = \bar{n}^2_0$). If this was not the case, one had to take into account the adaptation phases after each change of phases, which take too long, when the degree of deviation from the equilibrium at the end of adaptation is too small. However, under this assumption, we can show long-term stability for the finer ε_K-tunnel.

Second, we give a precise description of the growth and decline of the population of the pit stop trait. Under the use of a basic limit result of [4] and refining the techniques established in [45] exploiting Doob's maximum principle, we prove sharp bounds on the population size during one period. This enables us to finally determine the correct time scale and rate of a successful crossing of the valley.

5. Outlook and open questions

It is not a new insight that with every new theory that is proved the number of new emerging questions exceeds the number of given answers. Subsequently, we present a few of the topics.

A somewhat intuitive path forward is to combine the ideas given in the three main chapters in the sense to decode the metastable transitions on general trait graphs for periodically changing environments. It would be of some interest if one can find a similar metastability graph and how the phenomenon of pit stops enters in. This requires a careful analysis of the fitness landscape over the different phases of the environment.

Regarding the environment, it seems promising to consider changes of a more general nature. We are confident that without too much effort, one can extend the results for piece-wise constant parameters to periodic, continuous (or even cádlág) parameter functions, which could bring us closer to realistic scenarios. For the same reason, one could ask for random changes of the phases, either by random times of changes or more generally to model the environment by an ergodic Markov chain. While resident stability can likely be ensured similar to the demonstrated case, the control of the small populations and the decision about their survival or extinction presents as a non-trivial but highly interesting problem.

There are also various scenarios of less stability for the resident population driven by some realistic examples. Motivated by the drug-treatment example, we can think of a setting, where the resident population is also affected by the drug, hence declines and cannot stay macroscopic all the time. In this situation, the possible mutant populations are no longer suppressed by the bulk population and thus have a much larger ability to grow and invade the residents. From a medical point, this is a critical implication of the treatment since the mutants of the pathogen are likely to adapt better to the drug, in the sense of resistance, which means over time they become more aggressive.

The results on changing environments presented in this thesis only consider monomorphic resident populations, due to our different approach to estimating the speed of convergence for the resident stability. Nevertheless, we are of the opinion that the perturbation results of Freidlin and Wentzell [86] can be reworked and adapted to extend the results of Appendix B and C to polymorphic evolution sequences (cf. [43]).

Furthermore, we would like to relax the assumption of the unique stable equilibria to the Lotka-Volterra system involving the resident traits and the new mutant. Instead, the differential equation could approach a stable orbit, such that the resident population is periodically changing. Such limit cycles appear in Lotka-Volterra systems of dimension three and larger [171], for example in predator-prey systems and other constellations of coexistence like collaboration and mutual interaction. Let us mention the modelling of immunotherapy with oscillating cancer-cell and T-cell populations as a relevant application (cf. [121]). The tools established in this thesis seem auspicious to investigate the mutation growth in this context.

Of course, cyclic behaviour can be observed on nearly every time scale and may lead to interesting (average-)results on higher time scales. Besides the changing environment and the Lotka-Volterra limit cycles mentioned above, we heuristically discuss a particular case for a cycle in the metastability graph $\mathcal{G}_{\mathrm{ESC}}$ in Example 5 in Appendix A.

A further branch of mathematical biology, which permanently advances is the theory of dormancy or phenotypic switches in general. This mechanism is studied in full broadness ranging from models of population genetics and coalescent processes [17] over questions of fractional time derivatives [157] in population dynamics up to the spatial dispersal of seeds and seedbanks in ecological project [131]. It ties in with the modelling of cancer and different therapies against [1, 35, 84]. A first stepping stone to study the effects of dormancy in the models of adaptive dynamics is the works of Blath, Tóbiás [22]. The results on dormancy and evolutionary branching, which they developed under the collaboration of Paul and Wilke Berenguer [21] should be extended further to the theory of fitness valleys with larger mutation probabilities. Let us mention in addition, that there are also contributions on the interplay of dormancy and changing environment, e.g. [19, 49]. It seems also worth studying this in more detail regarding the faster environmental time scale of this thesis.

Moreover, let us take a closer view onto an area of adaptive dynamics that seems somewhat disconnected yet. In this thesis, we restrict ourselves to a discrete trait graph. Since mutations are not only rare but mostly cause only small changes, it is also reasonable to consider a continuous trait space. Under an appropriate scaling of small mutation steps (see 3.2.3) one can show convergence to the canonical equation of adaptive dynamics [7, 43, 61]. A non-trivial question is if and how one can achieve comparable results within a changing environment. How does the CEAD depend on the environment and what are the appropriate scales for environmental changes, mutation probability and mutation step size? Already in the constant parameter set-up, the simultaneous scaling requires a variety of new methods. So far there are still subtle heuristic conjectures on the time scale of evolutionary branching. Their proofs let us expect to open a new window towards a better understanding of speciation as a key observable of the evolutionary tree.

We close this outlook on rewarding approaches to adaptive dynamics by mentioning a current and relevant extension of the underlying class of individual-based models. In [154], Popovic and Véber present a particle-based model for chemical *reaction-diffusion processes* and proof a law of large number providing the convergence to an ODE system similar to the Lotka-Volterra equations. Although the biochemical interpretation of this model is meant on a subcellular scale, the individual-based population model outlined in Chapter 3 can also be seen in some sense as a special subclass of this newly established framework, which cannot hide its mathematical origin. It seems that the transfer of results from the adaptive dynamics model is not restricted only to the law of large numbers. Particularly, understanding the long-term behaviour on diverging time scales of these chemical reaction networks is a desirable aim of further investigation.

A. Publication: A general multi-scale description of metastable adaptive motion across fitness valleys

This appendix was published in the Journal of Mathematical Biology as joint work with Anna Kraut [69],

M. Esser and A. Kraut, *A general multi-scale description of metastable adaptive motion across fitness valleys.* Journal of Mathematical Biology, Volume 89 , Number 46, 2024. `https://doi.org/10.1007/s00285-024-02143-3`

We consider a stochastic individual-based model of adaptive dynamics on a finite trait graph $G = (V, E)$. The evolution is driven by a linear birth rate, a density dependent logistic death rate and the possibility of mutations along the directed edges in E. We study the limit of small mutation rates for a simultaneously diverging population size. Closing the gap between [29] and [48] we give a precise description of transitions between evolutionary stable conditions (ESC), where multiple mutations are needed to cross a valley in the fitness landscape. The system shows a metastable behaviour on several divergent time scales, corresponding to the widths of these fitness valleys. We develop the framework of a meta graph that is constituted of ESCs and possible metastable transitions between them. This allows for a concise description of the multi-scale jump chain arising from concatenating several jumps. Finally, for each of the various time scales, we prove the convergence of the population process to a Markov jump process visiting only ESCs of sufficiently high stability.

A.1. Introduction

The theory of evolution aims to understand the adaptation of biological populations to their environment through mutation and selection. Following the principles originally proposed by Darwin, it associates to each individual a fitness, which characterises their ability to survive and produce a growing population. The path of evolution, tracing the types of individuals that were able to fixate in the population, usually follows a sequence of types of increasing fitness. However, in many cases the mutational path has to pass through a number of deleterious or neutral intermediate types in order to reach a type of higher fitness. This can for example be seen in cancer initiation, where multiple driver mutations need to be accumulated to induce an outgrowing population [137]. Other examples are the formation of complex mechanisms like flagella in bacteria, where only partially functional intermediate stages of flagella yield an evolutionary disadvantage but fully functional apparatuses lead to increased fitness [152]. See also [59] for a review of empirical fitness landscapes arising in nature.

When the population needs to cross types of lower fitness in order to reach a fitter type, many such attempts will be unsuccessful. This is because the intermediate unfit types are destined to go extinct within a short time and might not produce a new mutant type before this happens. As a result, the waiting time to cross a valley in the fitness landscape is much longer than the invasion time of fit mutant types that are directly accessible. Once a fit type is attained, however, it rapidly fixates in the population. These dynamics, which can also be analysed in the framework of metastability, as illustrated below, have already been studied heuristically by Gillespie in the 80s [91]. Since then, fitness valleys have been studied in a variety of mathematical models, ranging from Moran models [119, 92] to multi-type branching processes [149].

The model that we want to focus on in this paper is a stochastic individual-based model of adaptive dynamics, for which Bovier, Coquille and Smadi have studied fitness valleys in the simple case of a linear trait space [29]. This type of model tracks the sizes of different subpopulations and - opposed to many others like the Moran model - does not work under the assumption of a constant overall population size. It is in this aspect closer to branching processes, where the population size varies over time. However, infinite growth is limited due to competitive interactions. Moreover, selective advantages of certain traits are not prescribed by a fixed parameter but arise through these interactions. This is particularly important for the long-term evolution of the population since the fitness landscape depends on the current composition of the dominant population and changes over time.

This study of the interplay of ecology and evolution goes back to ideas from Metz and Geritz (among others) in the early 90s [142]. Shortly after, an individual-based approach has been proposed by Bolker and Pacala [24] and a rigorous construction was first presented by Fournier and Méléard almost 20 years ago [83]. Since then, these models have been the topic of study for scaling limits in a variety of parameter regimes and extensions to the base model (e.g. [40, 43, 7, 158, 28, 120, 45, 48]). We refer to [27] for a comprehensive overview of various scaling limits.

To study the typical long-term behaviour of the population, two scaling parameters are introduced: The carrying capacity K, which scales the order of the population size, and the mutation probability μ_K, which scales the frequency of mutation events. For large populations ($K \to \infty$) and rare mutations ($\mu_K \to 0$), different mechanisms that change the state of the population - like mutations introducing a new type or interactions between individuals that lead to a new equilibrium state of resident traits - act on different time scales. There are three important time scales in this setting: Ecological interactions between well-established subpopulations, like the competition for resources, can change the composition of the overall population within a short time of order 1. This is related to classical Lotka-Volterra dynamics and leads to equilibrium states between the larger traits. Short-range mutations and the initial exponential growth of small mutant populations can be witnessed on a logarithmic time scale of order $\ln K$. Finally, long-range mutations - in particular those that need to traverse a large fitness valley of width L - are quite rare and occur on a time scale of order $1/K\mu_K^L$. The distinction between long and short-range mutations depends on the choice of the mutation probability μ_K, where long ranges L satisfy $K\mu_K^L \ll 1$. To obtain a non-trivial limit as $K \to \infty$, the population size is usually rescaled by K. As a result, only the established resident traits are visible. Since the ecological changes of these traits happen very fast in comparison with the other time scales, the limit of the population process yields a jump process that transitions between different equilibrium states.

The effects of short-range mutations on the $\ln K$-time scale have been studied extensively by Coquille, Kraut, and Smadi in [48]. The authors give a full description of the limiting dynamics for the scenario of a general finite graph as a trait space. As mentioned above, the crossing of fitness valleys through long-range mutations (on the $1/K\mu_K^L$-time scale) has been analysed for a simple linear trait space in [29]. Moreover, the case of very rare mutations, where even neighbouring traits are regarded as long-range mutations, has already been studied by Champagnat and Méléard in [40, 43], who showed convergence to the trait substitution sequence or polymorphic evolution sequence.

The present paper finally closes the gap between the previous works and gives a full description of the jump processes resulting from long-range mutations on general finite trait graphs, thus extending the results of [29] to the more general setting of [48]. This general setting entails that, for a given equilibrium state, there might be several paths to cross the surrounding fitness valley. Concentrating on the decisive, shortest paths we calculate the rate of a transition to the next evolutionary stable condition and give the precise asymptotics in Theorem A.2.7 and Corollary A.2.8. The length of the shortest paths determines the time scale to cross the valley. Based on this, we introduce the notion of a stability degree L to classify the equilibrium states. Combining multiple of these steps gives rise to a jump chain that moves on a so called metastability graph stated in Corollary A.2.10. This graph typically consists of fitness valleys of different width, which can be crossed on different time scales of the form $1/K\mu_K^L$. Depending on the choice of time scale, only some of these transitions are possible (valleys of width strictly larger than L cannot be crossed) or visible (transitions of valleys of width strictly smaller than L are immediate). This leads to different limiting jump processes in Theorem A.2.11.

When long-range mutations are necessary to cross a large fitness valley, the system displays an almost stable behaviour on shorter time scales but can change its state when waiting a long time. This type of phenomenon is also known as *metastability*. It has been studied mathematically mostly in the context of physics and statistical mechanics (e.g. [46]). However, the concept is very versatile and can be applied to many dynamical systems, including models for biological processes. This has for example been mentioned in [29] for models of adaptive dynamics, and in [56] for population dynamics.

In the former case, as well as in this paper, the role of the traditional physical energy (landscape) is taken over by the fitness (landscape). Instead of passing a critical state of high energy, the process has to cross a valley of negative fitness through a sequence of deleterious mutations. Similarly to the fast dynamics after passing a high energy state, the adaptive dynamics system quickly attains a new metastable equilibrium once a fit mutant is reached due to fast exponential growth. The results of [29] and this paper even confirm classical definitions of the mean time for a metastable transition (e.g. [30]), by proving that the waiting times for jumps between equilibrium states are exponentially distributed when considering the correct time scale.

While single jumps across a fitness valley can be regarded as metastable transitions, the limiting jump chain can be related to the concept of *adaptive walks* or *flights*. Those are stochastic processes that directly study the motion of the macroscopic population on the trait space, focussing on successful invasions and omitting the microscopic dynamics (see [123] for an overview). There are two sources of randomness in adaptive walks: A random fitness landscape and a random motion towards neighbours of higher fitness, according to some transition law. Based on these, properties of interest are the distribution and accessibility of fitness maxima [156, 150, 10, 11], as well as the time or path length to reach those maxima [151]. In adaptive flights, transitions are not just possible between neighbouring traits but from one local fitness maximum to another [108, 109, 106, 146]. This relates back to the limiting processes derived in this paper, where the population jumps between equilibrium states that are surrounded by valleys of traits of lower fitness.

A major difference between the models of adaptive walks/flights and adaptive dynamics is that the former assume a fitness landscape that is random but fixed in time, while in the latter case the fitness landscape is dynamic and depends on the current resident traits. As mentioned before, the notion of local fitness maxima can nevertheless be translated. Moreover, if equal competition between all traits is assumed in the adaptive dynamics model, the fitness landscape can again be regarded as fixed. We study this special case in a number of examples. Overall, the results of this paper can be seen as a validation of certain types of adaptive walks or flights, deriving their macroscopic dynamics from a microscopic, individual-based model.

The remainder of this paper is structured as follows: In Chapter A.2, we rigorously define the individual-based model of adaptive dynamics, for which we derive our limit theorems. We introduce key quantities, like the fitness of a trait, and recapitulate the most important results of [48] that lead to a metastable state on the $\ln K$-time scale. Finally, we heuristically derive the limit behaviour on longer time scales and present the formal convergence results, starting with a single metastable transition in Section A.2.3 and treating the full jump process in Section A.2.4. Chapter A.3 is devoted to the discussion of a number of examples that highlight different aspects of the complicated limiting dynamics in an easy set up. The proofs of the main results of this paper can be found in Chapter A.4. A combinatorial result on excursions of subcritical birth death processes and the complete version of the results from [48] are stated in Appendix A.5, for the convenience of the reader.

A.2. Model and Main Results

In this chapter we introduce the individual-based model of adaptive dynamics and develop the main results of this paper. After a rigorous definition of the population process and its driving parameters we give a short overview of the behaviour on the time scales of order 1 and $\ln K$ in Section A.2.2. Moreover, in this section we derive the key quantities that lead us to the definition of the notion of an evolutionary stable condition. Our main results on the transition out of an ESC are stated in Section A.2.3 and we give a heuristic explanation there. Finally, Section A.2.4 is devoted to our results on multi-scale jump chains and the convergence of the population process. For the convenience of the reader, we provide a preview of the different time scales and the main results of this paper at the end of Section A.2.1.

A.2.1. Individual-based model

To study the evolution of a heterogeneous population, we consider a classical stochastic individual-based model of adaptive dynamics. Each individual of our haploid population is characterised by its trait, which can be interpreted as its geno- or phenotype. Note that we assume a one to one correspondence between trait and physical properties. In this paper we consider a finite trait space that is given by a directed graph $G = (V, E)$. Here, the set of vertices V represents the possible traits that individuals can obtain. The set of edges E marks the possibility of mutation between traits.

To each trait we associate a number of parameters that describe the dynamics of the system. For $v, w \in V$ and $K \in \mathbb{N}$, denote by

- $b(v) \in \mathbb{R}_+$, 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.

Here, $\mathcal{M}_p(V)$ denotes the set of probability measures on V. The parameter K scales the competitive interaction between individuals. It is called *carrying capacity* and can be interpreted as the environment's capacity to support life, e.g. through the supply of nutrients or space. The way in which the mutation probability μ_K may depend on K is discussed below.

To ensure a limited population size and to establish the connection between the possibility of mutation and the edges of our trait graph, we make the following assumptions on our parameters.

Assumption 1. For every $v \in V$, $c(v, v) > 0$. Moreover, $m(v, v) = 0$, for all $v \in V$, and $(v, w) \in E$ if and only if $m(v, w) > 0$.

The evolution of the population over time is described by the Markov process N^K with values in $\mathbb{D}(\mathbb{R}_+, \mathbb{N}^V)$. $N_v^K(t)$ denotes the number of individuals of trait $v \in V$ that are alive at time $t \geq 0$. The process is characterised by its infinitesimal generator:

$$\begin{aligned} \mathcal{L}^K \phi(N) = &\sum_{v \in V} (\phi(N + e_v) - \phi(N)) \left(N_v b(v)(1 - \mu_K) + \sum_{w \in V} N_w b(w) \mu_K m(w, v) \right) \\ &+ \sum_{v \in V} (\phi(N - e_v) - \phi(N)) N_v \left(d(v) + \sum_{w \in V} c^K(v, w) N_w \right), \end{aligned} \tag{A.2.1}$$

where $\phi : \mathbb{N}^V \to \mathbb{R}$ is measurable and bounded and e_v denotes the unit vector at $v \in V$. The process can be constructed algorithmically following a Gillespie algorithm [90]. Alternatively the process can be represented via Poisson measures (see [83]), a representation that is used in the proofs of our results. Throughout this paper, we assume that all processes N^K, $K \in \mathbb{N}$, are defined on a common probability space. We give an example of a joint construction in the proof of Lemma A.4.1. However, we emphasize that we do not assume a specific dependence or independence between the different processes in order for our results to hold true.

We want to study the typical behaviour of this process for large populations and moderately rare mutations. We do not have a fixed population size. However, due to our scaling of $c^K(v, w)$, the equilibrium size of the population is always of order K. We therefore consider the limit of the processes $(N^K/K, K \in \mathbb{N})$ as $K \to \infty$ and $\mu_K \to 0$ simultaneously in this paper.

Outlook: In the following sections, we develop the theory to describe the systems behaviour on various time scales. Since the description of each increasing time scale builds on the behaviour on previous shorter time scales, we go through these step by step, introducing the relevant notation as well as previous and new results along the way. To give the reader some orientation, we provide a brief overview of the time scales and preview the main results:

- During times of order 1, the limiting rescaled stochastic process can be approximated by the solution of deterministic differential equations of Lotka-Volterra type. These describe how the larger subpopulations attain an equilibrium state (if existent). Since we consider the regime of $\mu_K \to 0$, mutations cannot be observed on this time scale.
- For moderately rare mutations $\mu_K = K^{-1/\alpha}$, mutations occur on the time scale $1/K\mu_K$ and mutant subpopulations grow from a single individual to a size of order K on the time scale $\ln K \gg 1/K\mu_K$. The limiting dynamics on the $\ln K$-time scale have been described in [48]. We provide the heuristics of this result in Section A.2.2 and give the precise statement in A.5.2. On this time scale, the system evolves until it reaches an equilibrium state, where there are no fit mutant traits of (graph-)distance at most α to the resident traits. This state is what we call an evolutionary stable condition (ESC).
- In Section A.2.3, we discuss how, on a more accelerated time scale $1/K\mu_K^L$ that corresponds to the distance $L > \alpha$ of the closest fit mutant, the process can escape an ESC. Our first result Theorem A.2.7 states that the time to produce a new fit mutant outside of the ESC is of order $1/K\mu_K^L$ and approximately exponentially distributed with a rate that can be calculated precisely. It moreover states the probabilities to produce specific mutant types. Corollary A.2.8 deduces that the time to reach a new ESC has the same distribution as the time of leaving the old ESC and calculates transition probabilities to reach specific new ESCs. These single transitions between ESC states, which can be regarded as metastable transitions, are used to define the (directed) metastability graph $\mathcal{G}_{\text{ESC}}$ in Definition A.2.9, in the beginning of Section A.2.4. It consists of subsets of V that allow for an ESC and the possible transitions between them.
- Since the time scales on which transitions on the metastability graph occur depend on the distances L between fit mutants and current resident traits, the corresponding jump chain (characterised in Corollary A.2.10) cannot be obtained as a limiting process on a single time scale. Instead, if we fix a time scale $1/K\mu_K^L$, only transitions of this

precise distance L are visible in the limit of N^K/K as $K \to \infty$. Shorter jumps occur immediately and longer jumps cannot be observed. To describe these dynamics, we introduce an L-scale graph $\mathcal{G}^L$, consisting of all ESCs that are not left immediately on the time scale $1/K\mu_K^L$ and characterize the limiting jump process on this graph in Theorem A.2.11.

A.2.2. Short-term dynamics and frequent mutations

A law of large numbers result by [75] states that, for $\mu_K \equiv 0$, the rescaled processes N^K/K converge to the solution of a system of Lotka-Volterra equations. The study of these equations is central to determine the short term evolution, i.e. the evolution on a finite time scale, of the process N^K.

Definition A.2.1 (Lotka-Volterra system, equilibrium states, invasion fitness)**.** For a subset $\mathbf{v} \subset V$ we denote by $LVS(\mathbf{v})$ the system of *Lotka-Volterra equations* given by

$$\frac{\mathrm{d}}{\mathrm{d}t} n_v(t) = \left(b(v) - d(v) - \sum_{w \in \mathbf{v}} c(v,w) n_w(t) \right) n_v(t), \quad v \in \mathbf{v}, \ t \geq 0. \tag{A.2.2}$$

By $LVE(\mathbf{v})$, we denote the set of all *equilibrium states* $\bar{n} \in \mathbb{R}^{\mathbf{v}}_{\geq 0}$ such that

$$\left(b(v) - d(v) - \sum_{w \in \mathbf{v}} c(v,w) \bar{n}_w \right) \bar{n}_v = 0, \quad v \in \mathbf{v}, \tag{A.2.3}$$

and by $LVE_+(\mathbf{v}) := LVE(\mathbf{v}) \cap \mathbb{R}^{\mathbf{v}}_{>0}$ the subset of *positive equilibrium states.* If $LVE_+(\mathbf{v})$ consists of a single globally asymptotically stable element, we denote it by $\bar{n}(\mathbf{v})$ and call it *coexistence equilibrium.*

For a trait $w \in V$ and coexistence equilibrium $\bar{n}(\mathbf{v})$, we denote by

$$f(w, \mathbf{v}) = b(w) - d(w) - \sum_{v \in \mathbf{v}} c(w,v) \bar{n}_v(\mathbf{v}) \tag{A.2.4}$$

the *invasion fitness* of w. For a given equilibrium $\bar{n}(\mathbf{v})$, we call a trait w *fit* if $f(w, \mathbf{v}) > 0$ and *unfit* if $f(w, \mathbf{v}) < 0$.

Note that the invasion fitness $f(w, \mathbf{v})$ describes the approximate growth rate of a small population of trait w in a bulk population of coexisting traits $\mathbf{v}$, in the mutation-free system. To simplify notation for later purpose, in the case of monomorphic equilibria, i.e. $\mathbf{v} = \{v\}$, we write

$$\bar{n}(v) := \bar{n}_v(\{v\}) \quad \text{and} \quad f(w,v) := f(w, \{v\}). \tag{A.2.5}$$

Going back to the stochastic process N^K, it is of interest to study the logarithm of the population size as $K \to \infty$. Only subpopulations with a size of order K are visible in the rescaled limit of N^K/K and exponential growth of the absolute population size translates to linear growth of the K-exponent when studying a logarithmic time scale via $e^{t \ln K \cdot f} = K^{t \cdot f}$.

This makes it easier to describe the limiting dynamics. We therefore define $\beta^K = (\beta_v^K)_{v\in V}$, where

$$\beta_v^K(t) := \frac{\ln(1 + N_v^K(t))}{\ln K}, \tag{A.2.6}$$

which is equivalent to $N_v^K(t) = K^{\beta_v^K(t)} - 1$. Note that we add or subtract 1 here respectively to ensure that $\beta_v^K(t) = 0$ if and only if $N_v^K(t) = 0$. As $K \to \infty$, β_v^K ranges between 0 and 1.

Remark 1. In contrast to [45, 48], we do not rescale the time by $\ln K$ in this definition of β^K since we are studying a variety of different time scales.

Based on this definition, we introduce the following subsets of traits.

Definition A.2.2 (macroscopic, microscopic, living and resident traits)**.**

1. A trait $v \in V$ with exponent β_v^K is called *macroscopic* if $\liminf_{K\to\infty} \beta_v^K = 1$.
2. A trait that is not macroscopic is called *microscopic*.
3. The set of *living traits* is the set $V_{\text{living}}^K := \{v \in V : \beta_v^K > 0\}$.
4. A subset of traits $\mathbf{v} \subseteq V$ is called *resident* if all $v \in \mathbf{v}$ are macroscopic and have a population size close to the coexistence equilibrium $\bar{n}(\mathbf{v})$.

Remark 2. Note that these definitions are time dependent when considering an evolving population. The macroscopic traits change according to $\beta^K(t)$ and the varying subset of living traits is denoted by $V_{\text{living}}^K(t)$. Most of the time macroscopic and resident traits coincide. A non-resident macroscopic trait is either unfit and will shrink to an order lower than K within a short time, or it is fit and will therefore induce a change in resident traits according to the short-term Lotka-Volterra dynamics.

To study multi-step mutations we consider paths on the trait graph $G = (V, E)$.

Definition A.2.3 (paths and distances)**.** We denote a *(finite) path* on $G = (V, E)$ by $\gamma = (\gamma_0, ..., \gamma_\ell)$ such that $\gamma_i \in V$, $0 \leq i \leq \ell$, and $(\gamma_i, \gamma_{i+1}) \in E$, $0 \leq i \leq \ell - 1$.
The *length of a path* $\gamma = (\gamma_0, ..., \gamma_\ell)$ is defined as $|\gamma| = \ell$. We write $\gamma : \mathbf{v} \to \mathbf{v}'$ as a short notation for all paths γ that connect $\mathbf{v} \subset V$ to $\mathbf{v}' \subset V$, i.e. that satisfy $\gamma_0 \in \mathbf{v}$ and $\gamma_{|\gamma|} \in \mathbf{v}'$. We introduce the graph distance between two vertices $v, w \in V$ as the length of the shortest connecting path

$$d(v, w) := \min_{\gamma: v\to w} |\gamma|, \tag{A.2.7}$$

where the minimum over an empty set is taken to be ∞. For two subsets $\mathbf{v}, \mathbf{v}' \subset V$ we define

$$d(\mathbf{v}, \mathbf{v}') := \min_{v\in\mathbf{v}, v'\in\mathbf{v}'} d(v, v'). \tag{A.2.8}$$

Remark 3. Note that $d(v, w)$ is not a distance in the classical sense, as it may not be symmetric in the case of a directed graph.

Along these paths γ, mutants can be produced. A macroscopic trait produces subpopulations of a size of order $K\mu_K$ of its neighbouring traits, which then produce subpopulations of a size of order $K\mu_K^2$ of the second order neighbours, and so on. These subpopulations, that are produced along a path γ, can survive as long as $K\mu_K^\ell \gg 1$. This motivates the study of mutation probabilities $\mu_K = K^{-1/\alpha}$, $\alpha > 0$, where mutants can survive within a radius α of the resident traits.

Remark 4. We could also study mutation probabilities $\mu_K = f(K)K^{-1/\alpha}$ such that $|\ln(f(K))| \in o(\ln K)$. This would not change the following results. However, we restrict ourselves to the case of $f(K) \equiv 1$ to simplify notation.

To avoid mutant subpopulations with a size of order $K^0 = 1$ and to ensure that non-resident traits are always either fit or unfit we make the following assumptions.

Assumption 2. (i) The mutation probability satisfies $\mu_K = K^{-1/\alpha}$ for some $\alpha \in \mathbb{R}_+ \setminus \mathbb{N}$.

(ii) For each $\mathbf{v} \subset V$ such that $LVE_+(\mathbf{v}) = \{\bar{n}(\mathbf{v})\}$, it holds $f(w, \mathbf{v}) \neq 0$, for all $w \notin \mathbf{v}$.

Remark 5. Both of these assumptions are purely technical. The first one prevents the case where a fit mutant population of size of order 1 can die out due to stochastic fluctuations such that fixation in the population becomes random. The second one allows us to approximate non-resident subpopulations by branching processes that are either super- or subcritical, but not critical. Note that the second assumption is only required for subsets $\mathbf{v}$ that allow for a unique positive equilibrium state (i.e. such that $LVE_+(\mathbf{v})$ contains exactly one element).

Under these assumptions, the evolution of the population on the time scale $\ln K$ has been studied in [48]. The authors give an algorithmic description of the limiting evolution of $\beta^K(t \ln K)$ as long as there always exists a unique asymptotically stable equilibrium of the Lotka-Volterra system (A.2.2) involving all macroscopic traits. In the following, we give the heuristics of this description. For the precise result we refer to Section A.5.2.

Roughly speaking, for a given set of resident traits $\mathbf{v}$ at their (coexistence) equilibrium $\bar{n}(\mathbf{v})$, every living microscopic trait $w \in V_{\text{living}}$ can grow (or shrink) with rate at least $f(w, \mathbf{v})$. This is due to the fact that the competitive interaction with all microscopic traits can be neglected in comparison with this rate. If there was no mutation (i.e. $\mu_K = 0$), $f(w, \mathbf{v})$ would be the exact growth rate of w. However, due to incoming mutants from neighbouring traits, the population size of w is also at least as big as a μ_K-fraction of the population sizes of its (incoming) neighbours. Since we only consider the order of the population size β_w^K, the largest of these influences dominates the asymptotics and a sum of population sizes (coming from different mutation sources) yields a maximum in the exponent. Overall, we obtain the relation

$$\beta_w^K(t \ln K) \approx \left(\beta_w^K(0) + tf(w, \mathbf{v})\right) \vee \max_{u \in V : d(u,w)=1} \left(\beta_u^K(t \ln K) - \frac{1}{\alpha}\right). \tag{A.2.9}$$

Iterating this argument for traits at increasing distance to w yields that, as long as the resident traits remain unchanged (i.e. traits $\mathbf{v}$ stay close to their equilibrium $\bar{n}(\mathbf{v})$ and no new traits become macroscopic), $\beta^K(t\ln K)$ converges to $\beta(t)$ such that

$$\beta_w(t) = \max_{u\in V} \left[\beta_u(0) + (t-t_u)f(u,\mathbf{v}) - \frac{d(u,w)}{\alpha}\right]_+ . \tag{A.2.10}$$

Here,

$$t_u := \begin{cases} \inf\left\{s \geq 0 : \exists\; u' \in V : \beta_{u'}(s) = \frac{1}{\alpha}, (u',u) \in E\right\} & \text{if } \beta_u(0) = 0, \\ 0 & \text{if } \beta_u(0) > 0. \end{cases} \tag{A.2.11}$$

Once a former microscopic trait w^* becomes macroscopic, the population sizes of $\mathbf{v} \cup w^*$ follow the Lotka-Volterra dynamics of (A.2.2) to reach a new equilibrium associated to the resident traits $\mathbf{v}' \subset \mathbf{v} \cup w^*$ within a time of order 1 (if such a new unique equilibrium does not exist, or in a number of other technical special cases, the algorithm terminates as described in Section A.5.2). During this phase, the orders of population sizes β_w do not change significantly. After the change of resident traits, the population sizes again follow (A.2.10), now with the changed fitnesses $f(u,\mathbf{v}')$.

This algorithmic description yields a series of successive resident traits. The macroscopically visible evolution stops as soon as an equilibrium $\mathbf{v}$ is reached such that $f(w,\mathbf{v}) < 0$ for all $w \in V_{\text{living}}\backslash\mathbf{v}$. All traits $w \in V$ such that $d(\mathbf{v},w) < \alpha$ stay alive due to incoming mutations but all other traits eventually go extinct according to (A.2.10) on the $\ln K$-time scale.

This observation leads us to the following definitions (visualised in Figure A.1).

Definition A.2.4 (mutation spreading neighbourhood)**.** For a subset $\mathbf{v} \subset V$, we denote by $V_\alpha(\mathbf{v}) := \{w \in V : d(\mathbf{v},w) < \alpha\}$ the *mutation spreading neighbourhood* of $\mathbf{v}$. The traits at the boundary of V_α are denoted by $\partial V_\alpha(\mathbf{v}) := \{w \in V : d(\mathbf{v},w) = \lfloor\alpha\rfloor\}$.

Definition A.2.5 ((asymptotic) evolutionary stable condition)**.** (i) A subset $\mathbf{v} \subset V$ and (orders of) population sizes β are called an *evolutionary stable condition (ESC)* if the traits $\mathbf{v}$ can coexist at a unique globally asympotically stable equilibrium $\bar{n}(\mathbf{v})$,

$$f(w,\mathbf{v}) < 0,\ \forall w \in V_\alpha(\mathbf{v})\backslash\mathbf{v}, \tag{A.2.12}$$

and

$$\beta_w = \left(1 - \frac{d(\mathbf{v},w)}{\alpha}\right)_+,\ \forall\ w \in V. \tag{A.2.13}$$

(ii) A subset $\mathbf{v} \subset V$ and population sizes $(\beta^K)_{K\geq 0}$ are called an *asymptotic evolutionary stable condition* if the traits $\mathbf{v}$ can coexist at equilibrium $\bar{n}(\mathbf{v})$, (A.2.12) is satisfied,

$$\left|\beta^K_w - (1 - d(\mathbf{v},w)/\alpha)\right| \in O\left(\frac{1}{\ln K}\right),\ \forall w \in V_\alpha(\mathbf{v}), \tag{A.2.14}$$

and there exists a $K_0 < \infty$ such that $\beta^K_w = 0$, for all $K > K_0$ and $w \in V\backslash V_\alpha(\mathbf{v})$.

Remark 6. (i) Note that (A.2.12) is only a necessary condition for a subset $\mathbf{v} \subset V$ to be able to attain an ESC during the evolution of a population. (A.2.13) are the orders of population sizes that unfit traits stabilise at purely due to (multi-step) mutations from $\mathbf{v}$. (A.2.12) guarantees that these will be reached for $w \in V_\alpha(\mathbf{v})$. To attain an ESC $(\mathbf{v}, \beta)$, in addition all other traits $w \in V_{\text{living}}(\tau_\mathbf{v})$, that are alive at the time $\tau_\mathbf{v}$ when the new equilibrium $\bar{n}(\mathbf{v})$ is reached, have to satisfy $f(w, \mathbf{v}) < 0$. If this is the case, all traits outside of $V_\alpha(\mathbf{v})$ will die out within a time of order $\ln K$ and (A.2.13) will be reached. Otherwise, if there is a $w \in V_{\text{living}}(\tau_\mathbf{v}) \backslash V_\alpha(\mathbf{v})$ such that $f(w, \mathbf{v}) > 0$ (the case $f(w, \mathbf{v}) = 0$ is excluded by Assumption (2)), its subpopulation is able to grow, will not die out, and hence not satisfy (A.2.13). The characterization of ESCs is therefore highly dependent on the state of the whole system.

(ii) Note that the definition of an asymptotic ESC forces the population process to be in an ESC up to a multiplicative error of order one. That is

$$N_w^K = (K^{(1-d(\mathbf{v},w)/\alpha)_+} - 1) \times \mathcal{O}(1). \tag{A.2.15}$$

The reason for introducing this error is that, for finite K, N_w^K might never reach exactly $K^{(1-d(\mathbf{v},w)/\alpha)_+}$. This is for example the case if $\bar{n}_v(\mathbf{v}) < 1$ for some $v \in \mathbf{v}$.

By definition, an evolutionary stable condition is surrounded by unfit traits, at least within an α-radius. This form of a fitness landscape is referred to as a *fitness valley* and has been studied in a special case in [29]. Based on this, we introduce a measure for the stability of a coexistence equilibrium, connected to the width of the surrounding fitness valley.

Definition A.2.6 (Stability degree)**.** For a subset $\mathbf{v} \subseteq V$ we define its *stability degree* $L(\mathbf{v})$ by

$$L(\mathbf{v}) := \begin{cases} \min_{w \in V : f(w,\mathbf{v})>0} d(\mathbf{v}, w) & \text{if } \mathbf{v} \text{ can coexist,} \\ 0 & \text{else.} \end{cases} \tag{A.2.16}$$

Remark 7. A subset $\mathbf{v}$ associated to an ESC satisfies $L(\mathbf{v}) > \alpha$ by definition. The evolution of the population process reaches a final state, independent of the time scale, once the resident traits satisfy $L(\mathbf{v}) = \infty$, i.e. there are no fit traits anymore.

A.2.3. Transitioning out of an ESC and first convergence result

Once an ESC is obtained, there is no further evolution on the $\ln K$-time scale. However, as long as there is a fitter trait that is connected to the resident traits, i.e. that can be reached along a finite path in G, we can witness metastable transitions on an even more accelerated time scale. On this time scale, under certain assumptions on the $\ln K$-dynamics, we observe a direct transition from one ESC to another.

In the following, we consider one of these transitions for an arbitrary initial asymptotic ESC. We split the transition into two phases: In the first phase, a new fit mutant beyond the fitness valley fixates in the population within a time of order $1/K\mu_K^{L(\mathbf{v})}$. In the second phase, a new ESC is obtained, starting with these new initial conditions, which takes a time of order $\ln K$. We assume that $\mathbf{v}$ and $(\beta^K(0))_{K\geq 0}$ are an asymptotic ESC. We could also consider more general initial conditions that lead to an ESC within finitely many steps of the $\ln K$-algorithm in [48], see Remark 11. For the sake of a simpler notation, we stick with the assumption of starting in an (asymptotic) ESC here.

To consider the first phase of the transition, we introduce the set

$$V_{\text{mut}}(\mathbf{v}) := \operatorname*{arg\,min}_{w\in V: f(w,\mathbf{v})>0} d(\mathbf{v},w) = \{w \in V : f(w,\mathbf{v}) > 0, d(\mathbf{v},w) = L(\mathbf{v})\}. \tag{A.2.17}$$

This consists of all fit mutant traits that are closest to $\mathbf{v}$ (visualised in Figure A.1).

Note that $V_{\text{mut}}(\mathbf{v}) \cap V_\alpha(\mathbf{v}) = \emptyset$ by the definition of an ESC. It turns out that the traits $V_{\text{mut}}(\mathbf{v})$ are the only ones that need to be considered for a crossing of the fitness valley since one of them will be the first new trait to fixate in the equilibrium population. If $V_{\text{mut}}(\mathbf{v}) = \emptyset$, i.e. $L(\mathbf{v}) = \infty$, there is no fitter trait connected to the resident traits and the equilibrium associated to $\mathbf{v}$ is the final state of the population.

For $L(\mathbf{v}) < \infty$, we define the stopping time

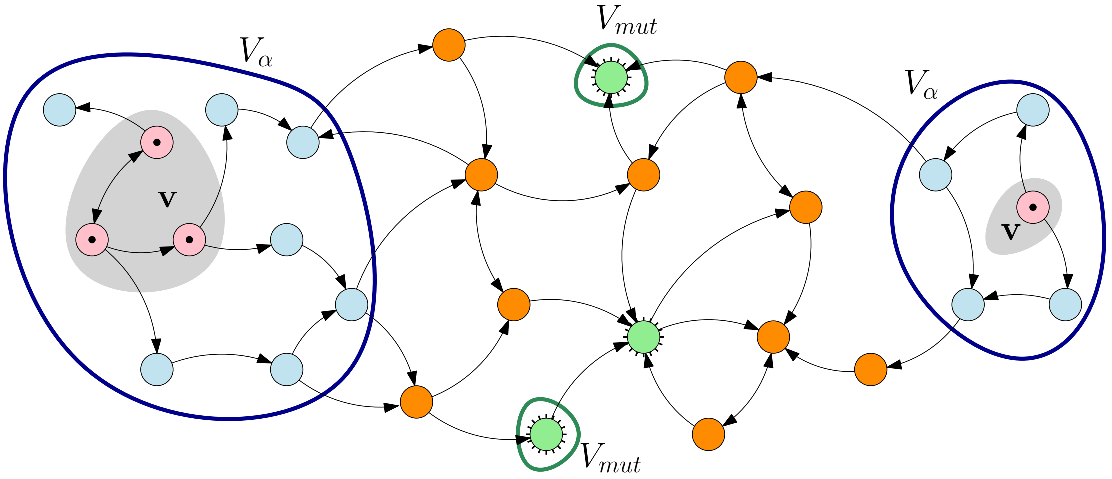
$$T^K_{\text{fix}} := \inf\left\{t \geq 0 : \exists\, w \in V\backslash V_\alpha(\mathbf{v}) : \beta^K_w(t) \geq \frac{1}{\alpha}\right\}, \tag{A.2.18}$$

the first time when a new trait reaches a size of order $K^{1/\alpha}$, can thus produce neighbouring mutants within a time of order 1 and influence the subpopulations of other traits.

Remark 8. Note that the name T^K_{fix} might be a little misleading at first glance. Generally, we speak of the fixation of a trait within a population as the event that the subpopulation corresponding to this trait does not go extinct (due to random fluctuations or negative fitness), as long as the fitness landscape stays unchanged. As this event is determined by the future progression of the population, there is no precise time point to pin it to. In particular, whether a trait fixates or goes extinct is not foreseeable at the time point when the first individual of this trait arises. Therefore, we choose instead the time point when the subpopulation has reached a size that guarantees non-extinction with probability 1, asymptotically as $K \to \infty$. We could choose a much smaller size than $K^{1/\alpha}$ for this, however, this will not influence the event of fixation and only change the stopping time by a time of order $\ln K$, which is negligible compared to the much longer time scale on which mutants arise. We thus pick the first time when mutants can influence the population size of other traits.

Our first result describes the limiting distribution of this stopping time T^K_{fix}.

For a path $\gamma : \mathbf{v} \to V_{\text{mut}}(\mathbf{v})$ such that $|\gamma| = L(\mathbf{v})$, the rate at which a $w = \gamma_{L(\mathbf{v})}$ mutant population arises along this path γ and fixates can be derived as the product of several factors. The rate at which the first trait in γ outside of $V_\alpha(\mathbf{v})$ arises can be calculated in terms of the equilibrium population sizes of the traits in $V_\alpha(\mathbf{v})$ (see Section A.4.1). This rate then has to be multiplied by the probabilities that all of the following unfit traits on the path γ produce mutants of the correct trait before extinction, i.e. during small subcritical



$\alpha \in (2,3)$, $f($○$, \mathbf{v}) < 0$, $f($●$, \mathbf{v}) < 0$, $f($✹$, \mathbf{v}) > 0$

Figure A.1.: Trait graph $G = (V, E)$ with an ESC associated to the non-connected set of spotted resident traits $\mathbf{v}$. Unfit traits inside the mutation spreading neighbourhood $V_\alpha(\mathbf{v})$ are marked light blue while those outside are dark orange. In this case, the stability degree is $L(\mathbf{v}) = 4$. Therefore, only the two nearest of the three fit, spiked green traits make up the set of mutant candidates $V_{\text{mut}}(\mathbf{v})$.

excursions. This yields the rate at which single mutants of trait w arise, which finally has to be multiplied by their probability of fixating in the population, i.e. of non-extinction.

In order to calculate the probability of mutation during a subcritical excursion, we need to introduce some notation. For a subset $\mathbf{v} \subset V$ and a trait $w \in V$ we define

$$\rho(w, \mathbf{v}) := \frac{b(w)}{b(w) + d(w) + \sum_{v \in \mathbf{v}} c(w, v) \bar{n}_v(\mathbf{v})}, \tag{A.2.19}$$

which is connected to the probability of a birth event in the branching process approximating the growth of a mutant w in a bulk population of coexisting traits $\mathbf{v}$. Moreover, we let

$$\lambda(\rho) := \sum_{\ell=1}^{\infty} \frac{(2\ell)!}{(\ell-1)!(\ell+1)!} \rho^\ell (1-\rho)^{\ell+1}, \tag{A.2.20}$$

which is the expected number of birth events before extinction in a subcritical birth death process with birth probability ρ (related to the expected number of positive jumps in a simple random walk on $\mathbb{N}$ before hitting 0, as explained in the proof of this result in Section A.5, Lemma A.5.1). Note that, for $\rho \in [0, 1/2)$, one can explicitely calculate that $\lambda(\rho) = \rho/(1-2\rho) < \infty$. Moreover, the symmetry relation $\lambda(\rho)\rho = \lambda(1-\rho)(1-\rho)$ shows convergence of the series for $\rho \in (1/2, 1]$ as well.

With these definitions, the overall rate of mutants of trait $w \in V_{\text{mut}}(\mathbf{v})$ arising along path γ

and fixating in the population is approximately equal to $R(\mathbf{v},\gamma)K\mu_K^{L(\mathbf{v})}$, where

$$R(\mathbf{v},\gamma) := \bar{n}_{\gamma_0}(\mathbf{v}) \left(\prod_{i=1}^{\lfloor\alpha\rfloor} \frac{b(\gamma_{i-1})m(\gamma_{i-1},\gamma_i)}{|f(\gamma_i,\mathbf{v})|} \right) b(\gamma_{\lfloor\alpha\rfloor})m(\gamma_{\lfloor\alpha\rfloor},\gamma_{\lfloor\alpha\rfloor+1}) \cdot \left(\prod_{j=\lfloor\alpha\rfloor+1}^{L(\mathbf{v})-1} \lambda(\rho(\gamma_j,\mathbf{v}))m(\gamma_j,\gamma_{j+1}) \right) \cdot \frac{f(\gamma_{L(\mathbf{v})},\mathbf{v})}{b(\gamma_{L(\mathbf{v})})}. \tag{A.2.21}$$

Here, the first line is the rate at which the first trait in γ outside of $V_\alpha(\mathbf{v})$ arises, which is related to the equilibrium size of trait $\gamma_{\lfloor\alpha\rfloor}$. The first factor in the second line is the probability of producing consecutive mutants during subcritical excursions and the last factor is the fixation probability of trait $w = \gamma_{L(\mathbf{v})}$. Note that, as $b(\gamma_{L(\mathbf{v})})$ increases, so does $f(\gamma_{L(\mathbf{v})},\mathbf{v})$ (cf. (A.2.4)), and hence this fixation probability is in fact increasing in the birth rate $b(\gamma_{L(\mathbf{v})})$.

The total rate at which a mutant population of trait $w \in V_{\mathrm{mut}}(\mathbf{v})$ arises and fixates collects all shortest paths that end in w and is approximately equal to $R(\mathbf{v},w)\mu_K^{L(\mathbf{v})}$, where

$$R(\mathbf{v},w) := \sum_{\substack{\gamma:\mathbf{v}\to w \\ |\gamma|=L(\mathbf{v})}} R(\mathbf{v},\gamma). \tag{A.2.22}$$

Finally, the total rate at which any mutant population of a trait in $V_{\mathrm{mut}}(\mathbf{v})$ arises and fixates, i.e. the rate at which the population exits the ESC associated to $\mathbf{v}$, is approximately equal to $R(\mathbf{v})\mu_K^{L(\mathbf{v})}$, where

$$R(\mathbf{v}) := \sum_{w\in V_{\mathrm{mut}}(\mathbf{v})} R(\mathbf{v},w). \tag{A.2.23}$$

The probability that this population is of trait $w \in V_{\mathrm{mut}}(\mathbf{v})$ is proportional to the rate $R(\mathbf{v},w)$.

With these heuristics, we can now state the first main result of this paper.

Theorem A.2.7. *Let $G = (V,E)$ be a finite graph. Suppose that Assumption 1 and 2 are satisfied and consider the model defined by* (A.2.1) *with $\mu_K = K^{-1/\alpha}$. Assume that $\mathbf{v} \subset V$ and $(\beta^K(0))_{K\geq 0}$ are an asymptotic ESC. Then there exist constants $\varepsilon_0 > 0$ and $0 < c < \infty$ such that, for all $0 < \varepsilon < \varepsilon_0$, there exist exponential random variables $E_+^K(\varepsilon)$ and $E_-^K(\varepsilon)$ with parameters $R(\mathbf{v})(1+c\varepsilon)$ and $R(\mathbf{v})(1-c\varepsilon)$, such that*

$$\liminf_{K\to\infty} \mathbb{P}(E_-^K(\varepsilon) \leq T_{\mathit{fix}}^K K\mu_K^{L(\mathbf{v})} \leq E_+^K(\varepsilon)) \geq 1 - c\varepsilon. \tag{A.2.24}$$

Moreover, for all $w \in V$, the probability of w being the trait to trigger T_{fix}^K is

$$\lim_{K\to\infty} \mathbb{P}\left(\beta_w^K(T_{\mathit{fix}}^K) = 1/\alpha\right) = \begin{cases} R(\mathbf{v},w)/R(\mathbf{v}) & \text{if } w \in V_{mut}(\mathbf{v}), \\ 0 & \text{else.} \end{cases} \tag{A.2.25}$$

Remark 9. Note that traits in $w \in V_\alpha(\mathbf{v})$ do not attain $\beta_w^K = 1/\alpha$ before T_{fix}^K due to the assumption that $\alpha \notin \mathbb{N}$. Therefore the probability in (A.2.25) is zero for such traits.

Once some $w \in V_{\text{mut}}(\mathbf{v})$ has reached $\beta_w^K \geq 1/\alpha$, the $\ln K$-dynamics evolve as described in [48], initiated with $\beta_w^K = 1/\alpha$ and $\beta_u^K = (1 - d(\mathbf{v}, u)/\alpha)_+$, for $u \in V \backslash w$. These dynamics are deterministic and in case they do not terminate early and if they lead to a new ESC, we denote the associated set of resident traits by $\mathbf{v}_{\text{ESC}}(\mathbf{v}, w)$.

Observe that there is no general formula to express $\mathbf{v}_{\text{ESC}}(\mathbf{v}, w)$ in terms of $\mathbf{v}$ and w and the parameters of the system. An interesting case is illustrated in Example 3.

Under the assumption that all traits $w \in V_{\text{mut}}(\mathbf{v})$ lead to asymptotic ESCs $\mathbf{v}_{\text{ESC}}(\mathbf{v}, w)$, we define the stopping time at which one of these asymptotic ESCs is obtained by

$$\begin{aligned} T_{\text{ESC}}^K := \inf \Bigg\{ t \geq T_{\text{fix}}^K : \exists\, w \in V_{\text{mut}}(\mathbf{v}) : \\ &\forall u \in V_\alpha(\mathbf{v}_{\text{ESC}}(\mathbf{v}, w)) : \left| \beta_u^K(t) - \left(1 - \frac{d(\mathbf{v}_{\text{ESC}}(\mathbf{v}, w), u)}{\alpha}\right) \right| < \varepsilon_K, \\ &\forall u \notin V_\alpha(\mathbf{v}_{\text{ESC}}(\mathbf{v}, w)) : \beta_u^K(t) = 0 \Bigg\}, \end{aligned} \tag{A.2.26}$$

where we pick $\varepsilon_K = C/\ln K$ for a large enough $0 < C < \infty$. Then this definition is precisely in line with the definition of an asymptotic ESC.

Remark 10. The minimal necessary C can be made precise using the prefactors of the population sizes in equilibrium, calculated in Lemma A.4.1. We refrain from doing so here as it is notationally very heavy and does not provide any deeper insight.

Since the time $T_{\text{ESC}}^K - T_{\text{fix}}^K$ is of order $\ln K$, the asymptotics for T_{fix}^K translate to T_{ESC}^K. Moreover, the transition probabilities from one ESC to another can be expressed in terms of the probabilities of traits $w \in V_{\text{mut}}(\mathbf{v})$ fixating in the population. For $\mathbf{w} \subset V$ we define

$$p(\mathbf{v}, \mathbf{w}) := \sum_{\substack{w \in V_{\text{mut}}(\mathbf{v}): \\ \mathbf{v}_{\text{ESC}}(\mathbf{v}, w) = \mathbf{w}}} \frac{R(\mathbf{v}, w)}{R(\mathbf{v})}. \tag{A.2.27}$$

Example 1 treats a case where this probability is indeed the sum over multiple mutant candidates w.

We can now state the result on transitions between ESCs as a direct corollary of Theorem A.2.7.

Corollary A.2.8. *Suppose the same assumptions as in Theorem A.2.7 are satisfied. Moreover, assume that, for every $w \in V_{mut}(\mathbf{v})$, the algorithmic description of the $\ln K$-dynamics in Section A.5.2, initiated with*

$$\beta_u(0) = \begin{cases} \frac{1}{\alpha} & \text{if } u = w, \\ \left(1 - \frac{d(\mathbf{v}, u)}{\alpha}\right)_+ & \text{else} \end{cases} \tag{A.2.28}$$

does not stop early due to one of its termination criteria and reaches an ESC associated to some traits $\mathbf{v}_{ESC}(\mathbf{v}, w)$ *after finitely many steps. Then,* $T^K_{ESC} - T^K_{fix} \in O(\ln K)$ *and therefore, with the same constants* ε_0 *and* c *and with the same random variables* $E^K_+(\varepsilon)$ *and* $E^K_-(\varepsilon)$ *as in Theorem A.2.7,*

$$\liminf_{K\to\infty} \mathbb{P}(E^K_-(\varepsilon) \leq T^K_{ESC} K \mu_K^{L(\mathbf{v})} \leq E^K_+(\varepsilon)) \geq 1 - c\varepsilon. \tag{A.2.29}$$

Moreover, for all $\mathbf{w} \subset V$*,*

$$\lim_{K\to\infty} \mathbb{P}(\{u \in V : \beta^K_u(T^K_{ESC}) > 1 - \varepsilon_K\} = \mathbf{w}) = p(\mathbf{v}, \mathbf{w}). \tag{A.2.30}$$

Remark 11. (i) Note that Theorem A.2.7 and Corollary A.2.8 only consider a specific transition from the ESC associated to some $\mathbf{v}$ to another ESC. The constants ε_0 and c can however be chosen uniformly for all ESCs by reason of the finite trait graph.

(ii) Both results assume that the system starts out in an asymptotic ESC. These are the natural initial conditions, particularly when a first transition between asymptotic ESCs has already occurred. We could however allow for more general initial conditions, as long as they lead to an asymptotic ESC within finitely many steps of the $\ln K$-algorithm.

A.2.4. Multi-scale jump chain and limiting Markov jump processes

Building on the previous description of a single transition step from one ESC to another, we now want to describe the multi-step transitions between ESCs as a jump chain $(v^{(k)})_{k\geq 0}$ on a meta-graph. We first introduce the underlying *metastability graph* $\mathcal{G}_{\mathrm{ESC}}$, consisting of all sufficiently stable macroscopic equilibrium configurations, and then describe the dynamics of the jump chain. Finally, we give a convergence result that derives different Markov jump processes, depending on the chosen time scale.

Definition A.2.9 (Metastability graph)**.** As vertices for the general metastability graph $\mathcal{G}_{\mathrm{ESC}} = (\mathcal{V}_{\mathrm{ESC}}, \mathcal{E}_{\mathrm{ESC}})$ we take all sets of resident traits that correspond to an ESC, i.e. that have stability degree strictly bigger than α, and edges represent possible transitions to other ESCs. More precisely,

$$\mathcal{V}_{\mathrm{ESC}} := \{\mathbf{v} \subseteq V : L(\mathbf{v}) > \alpha\}, \tag{A.2.31}$$

$$\mathcal{E}_{\mathrm{ESC}} := \{(\mathbf{v}, \mathbf{w}) : \exists w \in V_{\mathrm{mut}}(\mathbf{v}) \text{ s.t. } \mathbf{w} = \mathbf{v}_{\mathrm{ESC}}(\mathbf{v}, w)\}. \tag{A.2.32}$$

Recall that $\mathbf{v}_{\mathrm{ESC}}(\mathbf{v}, w)$ stands for the resident traits associated to the new ESC that is attained at the end of the $\ln K$-algorithm being started with resident set $\mathbf{v}$ and invading mutant $w \in V_{\mathrm{mut}}(\mathbf{v})$. We already assigned to each vertex $\mathbf{v} \in \mathcal{V}_{\mathrm{ESC}}$ the exit rate $R(\mathbf{v})$ in (A.2.23) and to each edge $(\mathbf{v}, \mathbf{w}) \in \mathcal{E}_{\mathrm{ESC}}$ the transition probability $p(\mathbf{v}, \mathbf{w})$ in (A.2.27).

Using Corollary A.2.8, we can now work out inductively the multi-scale jump chain $(\mathbf{v}^{(k)})_{k\geq 0}$ on $\mathcal{G}_{\mathrm{ESC}}$. To this end, let $\mathbf{v}^{(0)} \in \mathcal{V}_{\mathrm{ESC}}$ be the resident traits of the initial ESC that the process starts in and set $T_{\mathrm{ESC}}^{(0,K)} := 0$. We describe the k^{th} transition, for $k \geq 1$, conditioned on the knowledge of $\mathbf{v}^{(k-1)}$. We denote the set of closest fit mutant traits by $V_{\mathrm{mut}}^{(k)} = V_{\mathrm{mut}}(\mathbf{v}^{(k-1)})$, the width of the next fitness valley to cross by $L^{(k)} = L(\mathbf{v}^{(k-1)})$, and the exit rate by $R^{(k)} = R(\mathbf{v}^{(k-1)})$. Moreover, we keep track of the time when the first mutant population fixates and when the next ESC is reached by introducing the stopping times

$$T_{\mathrm{fix}}^{(k,K)} := \inf\left\{ t \geq T_{\mathrm{ESC}}^{(k-1,K)} : \exists\, w \in V \backslash V_\alpha(\mathbf{v}^{(k-1)}) : \beta_w^K(t) \geq \frac{1}{\alpha} \right\}, \tag{A.2.33}$$

$$\begin{aligned} T_{\mathrm{ESC}}^{(k,K)} := \inf\Bigg\{ t \geq T_{\mathrm{fix}}^{(k,K)} : \exists\, w \in V_{\mathrm{mut}}^{(k)} :& \\ \forall u \in V_\alpha(\mathbf{v}_{\mathrm{ESC}}(\mathbf{v}^{(k-1)}, w)) :& \left| \beta_u^K(t) - \left(1 - \frac{d(\mathbf{v}_{\mathrm{ESC}}(\mathbf{v}^{(k-1)}, w), u)}{\alpha}\right) \right| < \varepsilon_K, \\ \forall u \notin V_\alpha(\mathbf{v}_{\mathrm{ESC}}(\mathbf{v}^{(k-1)}, w)) :& \, \beta_u^K(t) = 0 \Bigg\}, \end{aligned} \tag{A.2.34}$$

with ε_K as in A.2.26.

With this notation, we can now state the result on the k^{th} transition of the multi-scale jump chain.

Corollary A.2.10. *Assume that we constructed the process up to time $T_{ESC}^{(k-1,K)}$, when the ESC associated to $\mathbf{v}^{(k-1)}$ is obtained, and suppose the same assumptions as in Theorem A.2.7 are satisfied. Moreover, assume that the $\ln K$-dynamics behave as in Corollary A.2.8, for every $w \in V_{mut}^{(k)}$. Then there exist constants $\varepsilon_0 > 0$ and $0 < c < \infty$ such that, for all $0 < \varepsilon < \varepsilon_0$, there are exponential distributed random variables $E_+^{(k,K)}(\varepsilon)$ and $E_-^{(k,K)}(\varepsilon)$ with parameters $R_\pm^{(k)}(\varepsilon) := R^{(k)}(1 \pm c\varepsilon)$ such that*

$$\liminf_{K\to\infty} \mathbb{P}(E_-^{(k,K)}(\varepsilon) \leq (T_{ESC}^{(k,K)} - T_{ESC}^{(k-1,K)}) K \mu_K^{L^{(k)}} \leq E_+^{(k,K)}(\varepsilon) | \mathbf{v}^{(k-1)}) \geq 1 - c\varepsilon. \tag{A.2.35}$$

Moreover, for all $\mathbf{w} \subset V$,

$$\lim_{K\to\infty} \mathbb{P}(\{v \in V : \beta_v^K(T_{ESC}^{(k,K)}) > 1 - \varepsilon_K\} = \mathbf{w} | \mathbf{v}^{(k-1)}) = p(\mathbf{v}^{(k-1)}, \mathbf{w}). \tag{A.2.36}$$

The preceding corollary allows us to construct a limiting random jump chain $(\mathbf{v}^{(k)})_{k\geq 0}$ on the metastability graph $\mathcal{G}_{\mathrm{ESC}}$. To be precise, given the current state $\mathbf{v}^{(k-1)}$, the next ESC $\mathbf{v}^{(k)}$ is taken at random from $\mathcal{V}_{\mathrm{ESC}}$ with probability distribution $p(\mathbf{v}^{(k-1)}, \cdot)$. However, the jumps take place on varying time scales of type $1/K\mu_K^{L^{(k)}}$. The construction is valid until an ESC is obtained such that some mutant $w \in V_{\mathrm{mut}}^{(k)}$ does not induce a unique new ESC, following the deterministic $\ln K$-dynamics. A visualisation of the metastability graph including a particular realisation of the jump chain is given in Figure A.2.

After this general description of the multi-scale jump chain we can now easily elaborate the true Markov jump process on each time scale. To be more precise, for each stability degree

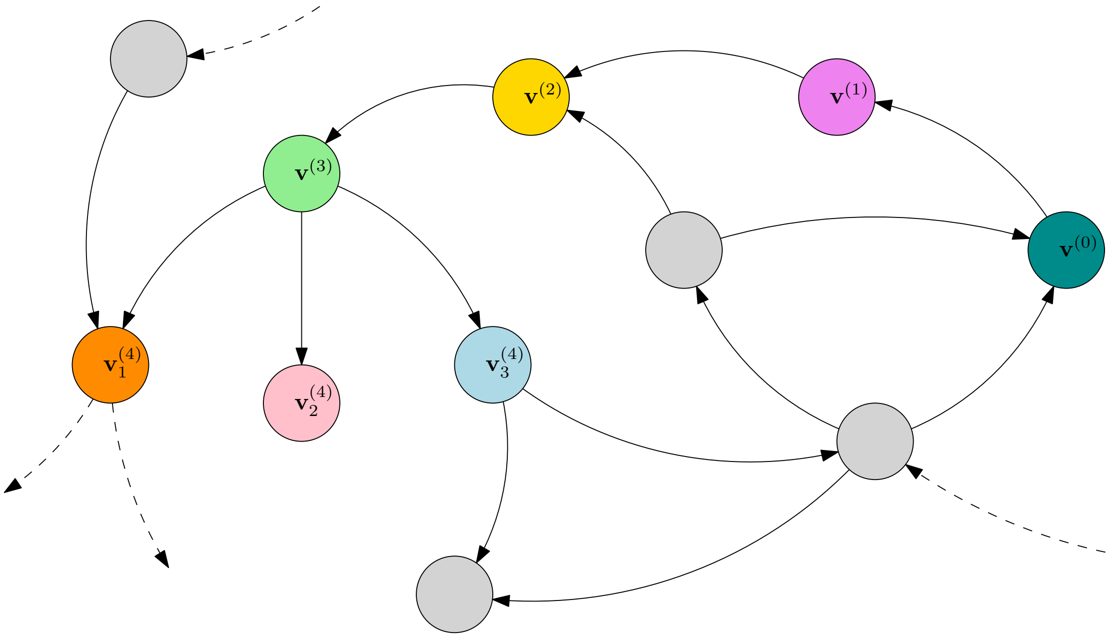


Figure A.2.: Metastability graph $\mathcal{G}_{\text{ESC}}$ including a jump chain $(\mathbf{v}_k)_{k\geq 0}$, where $\mathbf{v}_i^{(4)} = \mathbf{v}_{\text{ESC}}(\mathbf{v}^{(3)}, w_i)$, for $V_{\text{mut}}(\mathbf{v}^{(3)}) = \{w_1, w_2, w_3\}$.

$L > \alpha$, we are looking for the limit process of $N_w^K(t/K\mu_K^L)/K$, for $K \to \infty$. The support of this process jumps between sets of coexisting traits of sufficiently high stability degree, which can only be exited on this time scale. In this context, we define the level sets of equal stability degree as

$$\mathcal{S}^L := \{\mathbf{v} \subseteq V : LVE_+(\mathbf{v}) = \{\bar{n}(\mathbf{v})\}, L(\mathbf{v}) = L\}\,. \tag{A.2.37}$$

Note that, for $L > \alpha$, a stability degree of $L(\mathbf{v}) = L$ ensures that the coexisting traits $\mathbf{v}$ allow for an asymptotic ESC, see Remark 7.

As the state space for the limiting jump process, we introduce the *L-scale graph* $\mathcal{G}^L$, which is a collapsed version of $\mathcal{G}_{\text{ESC}}$. The vertex set consists of all ESCs that are stable enough to be visible on the respective time scale. Therefore, we set

$$\mathcal{V}^L := \bigcup_{L' \geq L} \mathcal{S}^{L'}. \tag{A.2.38}$$

Note that it is possible that the process jumps into an ESC $\mathbf{v} \in \mathcal{S}^{L'}$, for $L' > L$, on the $1/K\mu_K^L$-time scale. However, there is no possibility to escape from those on this time scale, which means that these are absorbing states.

Edges $\mathcal{E}^L$ in $\mathcal{G}^L$ represent possible transitions of the limiting process. To construct these, we study the limiting jump chain from Corollary A.2.10.

In order to use the corollary and in particular the process $(\mathbf{v}^{(k)})_{k\geq 0}$, we have to ensure that, for fixed $L > \alpha$, this process always reaches an ESC of stability degree at least L in finitely many steps.

Assumption 3.

$$\forall \mathbf{v} \in \mathcal{S}^L : \mathbb{P}\left(\exists k \in \mathbb{N}_{>0} : L(\mathbf{v}^{(k)}) \geq L | \mathbf{v}^{(0)} = \mathbf{v}\right) = 1 \tag{A.2.39}$$

Note that, if this assumption is satisfied for some fixed L, this has no implications for the validity for different $L' \neq L$. This is due to the fact, that only the initial conditions $\mathbf{v} \in \mathcal{S}^L$ are considered. One can easily think of counterexamples where $\mathcal{G}_{\text{ESC}}$ is non-connected such that there may be cycles of lower time scale but there is no danger to run into them. For a broader discussion of the assumption we refer to the Examples 4 and 5.

Remark 12. If the process runs into a cycle or stable cluster on a lower time scale, there are still possibilities to escape from these by accelerating and looking at higher time scales. The detailed description of such behaviour is much more involved. This is mainly due to technical reasons: Errors accumulate in the approximation of each transition step. As long as it is ensured that the system reaches a (sufficiently stable) ESC after finitely many steps, these errors can be iteratively bounded to ensure convergence. This however fails if the number of lower time scale transitions between higher time scale jumps is not bounded. Heuristically, if one can observe ergodic behaviour on the L'-scale graph, for some $L' < L$, transitions out of the ergodic cluster will occur along one of the shortest fitness valleys of width L. Transition rates will be weighed according to the stationary distribution on states in $\mathcal{S}^{L'}$ and the transition takes a time of order $1/K\mu_K^L$. Rather than defining vertices of $\mathcal{G}^L$ as single sets of coexisting traits in $\mathcal{S}^L$, one would then choose communication classes of such sets in $\mathcal{S}^{L'}$ (for possibly multiple $L' < L$) that support an ergodic stationary distribution. Rigorously justifying this argument is a topic of current and future research of the authors.

Now asking for possible jumps in $\mathcal{G}^L$ we have to respect again the principle that jumps on lower time scales are absorbed in those happening on the $1/K\mu_K^L$-time scale. This means that the critical event for a transition starting in $\mathbf{v} \in \mathcal{S}^L$ is to escape from $\mathbf{v}$, which needs a time of order $1/K\mu_K^L$. Compared to that, the subsequent transitions in $\mathcal{G}_{\text{ESC}}$ until reaching again a state $\mathbf{w}$ of stability at least $L(\mathbf{w}) \geq L$ take place in very short time. Therefore we say that the (directed) egde $(\mathbf{v}, \mathbf{w})$ is in $\mathcal{E}^L$ if and only if $L(\mathbf{v}) = L$ and there exists a finite path $\Gamma : \mathbf{v} \to \mathbf{w}$ in $\mathcal{G}_{\text{ESC}}$ such that $L(\Gamma_i) < L,\ \forall 1 \leq i < |\Gamma|$.

The probability of possible transitions $(\mathbf{v}, \mathbf{w}) \in \mathcal{E}^L$ is then the sum over all possible paths Γ that give rise to this edge, while the probability of taking a particular path is easily computed as the product of its segments in $\mathcal{G}_{\text{ESC}}$.

$$p^L(\mathbf{v}, \mathbf{w}) := \sum_{\substack{\Gamma:\mathbf{v}\to\mathbf{w} \\ L(\Gamma_i)<L,\ \forall 1\leq i<|\Gamma|}} \prod_{i=1}^{|\Gamma|} p(\Gamma_{i-1}, \Gamma_i) \tag{A.2.40}$$

For an explanatory computation of these probabilities we refer to the Examples 6 and 7.

The transition rate for the jumps on the $1/K\mu^L$ time scale are then given by the over-all rate to escape from $\mathbf{v}$ weighted with the transition probability to end in $\mathbf{w}$.

$$\mathcal{R}^L(\mathbf{v}, \mathbf{w}) := R(\mathbf{v}) p^L(\mathbf{v}, \mathbf{w}) \tag{A.2.41}$$

Now we are prepared to formulate the main result, i.e. the convergence to a Markov jump process on different time scales.

Theorem A.2.11. *Let $L > \alpha$ such that Assumption 3 holds true and take $\mathbf{v}^L(0) \in \mathcal{V}^L$. Suppose the same assumptions as in Theorem A.2.7 are satisfied for $\mathbf{v} = \mathbf{v}^L(0)$ and assume that the $\ln K$-dynamics behave as in Corollary A.2.8, for every $\mathbf{v} \in \bigcup_{L' \leq L} \mathcal{S}^{L'}$ and all $w \in V_{mut}(\mathbf{v})$. Then, for all $T < \infty$, the rescaled process $(N_v^K(t/K\mu_K^L)/K,\ v \in V,\ t \in [0,T])$ converges in the sense of finite dimensional distributions to a jump process of the form*

$$\mathcal{N}_v^L(t) = \mathbb{1}_{v \in \mathbf{v}^L(t)} \bar{n}_v(\mathbf{v}^L(t)), \qquad v \in V,\ t \in [0,T]. \tag{A.2.42}$$

Here $(\mathbf{v}^L(t),\ t \in [0,T])$ is a Markov jump process on the L-scale graph $\mathcal{G}^L = (\mathcal{V}^L, \mathcal{E}^L)$, with transition rates given by (A.2.41).

Remark 13. (i) We like to point out that Assumption 3 does not exclude the cases where we have cycles in $\mathcal{G}^L$, i.e. on the time scale $1/K\mu_K^L$. It only prevents the process from running into a cycle of lower time scale. We even allow for self connecting edges, i.e. edges of the form $(\mathbf{v}, \mathbf{v})$.

(ii) As shown in [40, Prop. 1] it is not possible to get convergence with respect to the Skorohod (J1)-topology since this would imply continuity for the limit of the total mass process, which cannot be true.

A.3. Interesting examples

In this chapter, we present and analyse a variety of examples that aim to highlight different aspects of the complicated dynamics covered in our main results. The first two examples are dedicated to single transition steps from one ESC to another, applying the results of Theorem A.2.7 and Corollary A.2.8. The next three examples focus on the metastability graph $\mathcal{G}_{\text{ESC}}$ that is constructed in Corollary A.2.10 and we study two cases that are concerned with Assumption 3. The final two examples are focussed on applications of Theorem A.2.11, studying the limiting Markov jump processes on different time scales as well as the L-scale-graphs $\mathcal{G}^L$.

In order to give a manageable and clear description of the dynamically changing fitness landscape, we introduce some new notation that helps to simplify the set up of the examples.

Definition A.3.1. We speak of a *regime of equal competition* if and only if $c(v,w) \equiv \text{const} > 0$, for all $v, w \in V$.

This is by no means a necessary assumption to produce the studied phenomena, however, it allows us to characterise the fitness landscape in a much simpler way. In the case of equal competition, the invasion fitness of a trait w with respect to a single resident trait v is fully characterised by

$$f(w,v) = r(w) - r(v), \tag{A.3.1}$$

where we set $r(v) := b(v) - d(v)$ as the *individual fitness* of trait v, i.e. its net growth rate in the absence of competitive interactions. As a consequence, traits w with higher individual fitness than the resident v are able to invade the population. Hence, instead of specifying the invasion fitnesses for all possible resident traits, the fitness landscape is fully described by the individual fitnesses $r(v)$.

To specify the fitness relations between different traits - in particular in the case of non-equal competition - we introduce the following notation.

Definition A.3.2. For $v, w \in V$, we write $v \ll w$ if and only if $f(w,v) > 0$ and $f(v,w) < 0$. Moreover, we write $v_1, ..., v_k \ll w_1, ..., w_l$ whenever $v_i \ll w_j$, for all $1 \leq i \leq k$ and $1 \leq j \leq l$.

This reflects the case where the equilibrium of the Lotka-Volterra system involving v and w is the monomorphic equilibrium $\bar{n}(w)$ of w. In other words w can invade the v population and fixate.

A.3.1. Single transition steps

A first example with multiple mutation paths

Example 1. *Let us consider the directed graph G depicted in Figure A.3. Assume equal competition and the individual fitness r plotted in Figure A.3. Moreover, let $\alpha \in (1,2)$.*

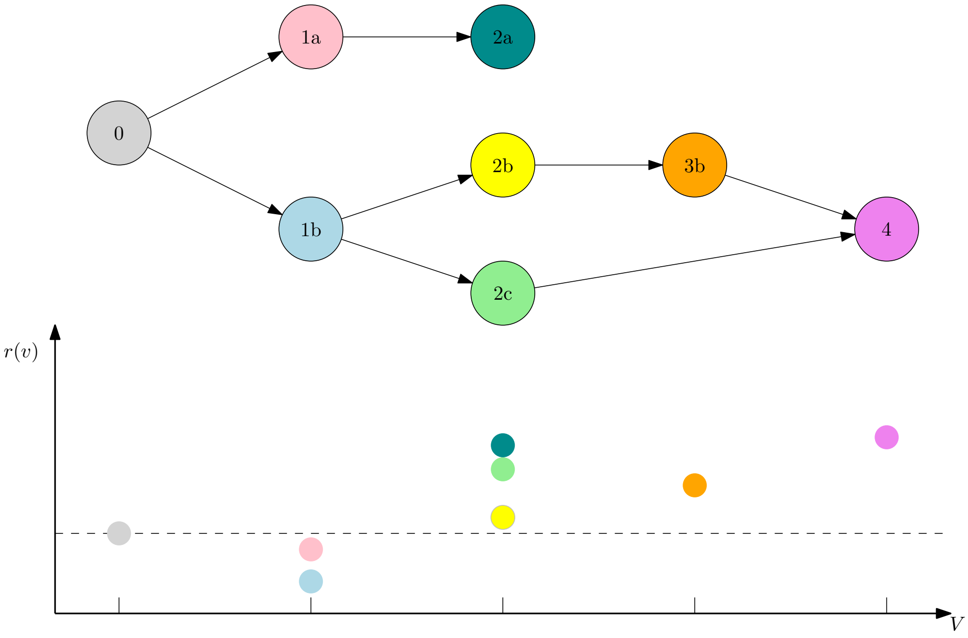


Figure A.3.: Trait graph G and fitness landscape in terms of individual fitness r of Example 1.

In this case, the initial resident trait 0 has stability degree $L(\{0\}) = 2 > \alpha$. This is due to the fact that traits $1a$ and $1b$ are unfit in presence of the resident, while traits $2a$, $2b$ and $2c$ are fit, with connecting paths $\gamma^A = (0, 1a, 2a)$, $\gamma^B = (0, 1b, 2b)$ and $\gamma^C = (0, 1b, 2c)$ of length 2 respectively. Therefore, we have the possible mutant candidates $V_{\text{mut}}(\{0\}) = \{2a, 2b, 2c\}$. An application of Theorem A.2.7 yields that we can observe a new fixating trait at rescaled time $T^K_{\text{fix}} K\mu_K^2$, which is distributed approximately as a exponential random variable with rate $R(\{0\}) = R(\{0\}, 2a) + R(\{0\}, 2b) + R(\{0\}, 2c)$. The probability for say trait $2b$ to be the trait that fixates in the population and triggers the stopping time is $R(\{0\}, 2b)/R(\{0\})$.

Asking for the new ESCs, which are reached after fixation, we have to take into account the subsequent evolution on the $\ln K$ time scale. This allows for jumps towards traits of higher fitness, which are in the mutation spreading neighbourhood, i.e. direct neighbours in this case. Therefore, we end up with

$$\mathbf{v}_{\text{ESC}}(\{0\}, 2a) = \{2a\} \quad \mathbf{v}_{\text{ESC}}(\{0\}, 2b) = \{4\}, \quad \mathbf{v}_{\text{ESC}}(\{0\}, 2c) = \{4\}. \tag{A.3.2}$$

In particular, note that $\{2b\}, \{2c\}$ are not ESCs and thus not part of the metastability graph $\mathcal{G}_{\text{ESC}}$ as plotted in Figure A.4.

This puts us into the setting where the sum in (A.2.27) becomes relevant. In particular, despite the micro-evolutionary branching from $1b$ into $2b$ and $2c$ in the trait graph G , there is no such branching on the macro-evolutionary level in $\mathcal{G}_{\text{ESC}}$. There, we only observe a transition from $\{0\}$ to $\{4\}$. Note also that the different path lengths of $2b \to 4$ and $2c \to 4$ do not matter for the asymptotics of the time T_{ESC} until stabilising in the new ESC. This is because this time is dominated by the waiting time T_{fix} for the first fixation of a fit mutant trait. Since $L(\{0\}) = 2$, this time is of order $1/K\mu_K^2$ and thus absorbs the much faster $\ln K$ evolution.

Note that, since all transitions between ESCs occur on the time scale $1/K\mu_K^2$ here, the metastability graph $\mathcal{G}_{\text{ESC}}$ agrees with the 2-scale graph $\mathcal{G}^2$.

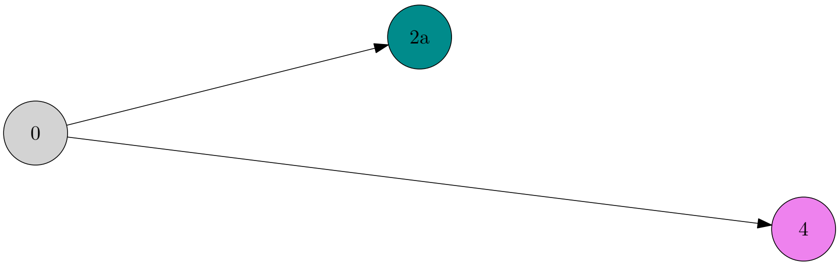


Figure A.4.: Metastability graph $\mathcal{G}_{\text{ESC}}$ and 2-scale graph $\mathcal{G}^2$ of Example 1

An ESC with coexistence

Since in this paper we discuss the occurrence of metastable behaviour in a rather general setting, we like to point out that Definition A.2.5 explicitly allows for ESCs $\mathbf{v}$ that consist of several coexisting traits. This clearly enlarges the mutation spreading neighbourhood $V_\alpha(\mathbf{v})$ and changes the set of mutant candidates V_{mut} in a non-trivial way.

Example 2. *Let us consider the directed graph G depicted in Figure A.5. Let $\alpha \in (1,2)$ and consider a fitness landscape that satisfies*

$$f(0,3), f(3,0) > 0, \tag{A.3.3}$$
$$f(1,\{0,3\}), f(2,\{0,3\}) < 0, \tag{A.3.4}$$
$$f(4,\{0,3\}), f(5,\{0,3\}) > 0, \tag{A.3.5}$$
$$0,1,2,3 \ll 4,5, \tag{A.3.6}$$
$$1,2 \ll 0,3, \tag{A.3.7}$$
$$f(4,5), f(5,4) < 0, \tag{A.3.8}$$

and allows for no polymorphic coexistence equilibria apart from $\{0,3\}$. Moreover, assume that the unique stable equilibrium of the Lotka-Volterra system involving traits $\{0,3,4\}$ is $\bar{n}(4)$ and the same is true for 5 replacing 4.

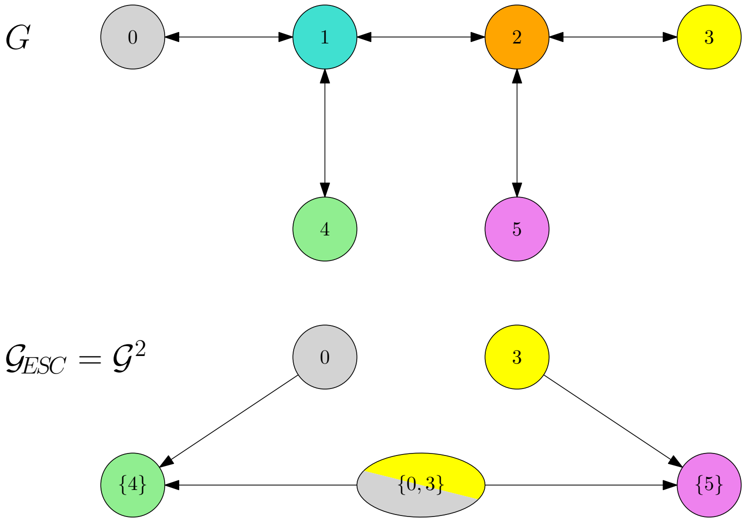


Figure A.5.: Trait graph G and metastability graph $\mathcal{G}_{\mathrm{ESC}}$ (which agrees with the 2-scale graph $\mathcal{G}^2$) of Example 2

Checking for traits that do not have any fitter direct neighbours, and hence do not allow for transitions on the $\ln K$-time scale, the monomorphic ESCs in this case correspond to $\{0\}$, $\{3\}$, $\{4\}$, and $\{5\}$. Classical results on Lotka-Volterra systems yield that under assumption (A.3.3) traits 0 and 3 can coexist, i.e. $\bar{n}(\{0,3\}) \in \mathbb{R}^2_{>0}$. Now the mutation spreading neighbourhood is given by $V_\alpha(\{0,3\}) = \{0,1,2,3\}$. Apart from the resident traits themselves, those traits are by assumption unfit with respect to $\{0,3\}$ and thus $\{0,3\}$ allows for an ESC.

Looking for the stability degree and possible mutant candidates, the assumptions on the fitness landscape imply that

$$L(\{0,3\}) = 2 \quad \text{and} \quad V_{\mathrm{mut}}(\{0,3\}) = \{4,5\}. \tag{A.3.9}$$

By Theorem A.2.7, we can observe a fixating mutant population of one of the traits $w \in V_{\mathrm{mut}}(\{0,3\})$ on the time scale $1/K\mu_K^2$. The corresponding rates are given by

$$R(\{0,3\},4) = \bar{n}_0(\{0,3\})\frac{b(0)m(0,1)}{|f(1,\{0,3\})|}b(1)m(1,4)\frac{f(4,\{0,3\})}{b(4)}, \qquad \text{for } w=4, \tag{A.3.10}$$

$$R(\{0,3\},5) = \bar{n}_3(\{0,3\})\frac{b(3)m(3,2)}{|f(2,\{0,3\})|}b(2)m(2,5)\frac{f(5,\{0,3\})}{b(5)}, \qquad \text{for } w=5. \tag{A.3.11}$$

Note that, although there are also paths connecting 3 → 4 and 0 → 5, only the paths of shortest length $|\gamma| = 2$ do have an impact on the above rates.

To conclude this example, we see that both mutant traits 4 and 5 are fit enough to invade the coexisting resident population. Overall, we obtain the metastability graph $\mathcal{G}_{\mathrm{ESC}}$ pictured in Figure A.5, which in this case agrees with the 2-scale graph $\mathcal{G}^2$. Note that the traits 0 and 3 appear both as monomorphic ESCs, as well as a polymorphic coexistence ESC.

A.3.2. Successive metastable transitions

Self connection in $\mathcal{G}_{\mathrm{ESC}}$

By definition of an ESC, the first fixating mutant has a distance of at least $\lfloor\alpha\rfloor + 1$ from the corresponding resident traits. Despite this fact, the $\ln K$-mechanism triggered by such a mutant may lead to a new ESC that is closer to the old one than $\lfloor\alpha\rfloor + 1$. It can even be the same and thus give rise to a self-connecting edge in $\mathcal{G}_{\mathrm{ESC}}$

Example 3. *Let us consider the directed graph depicted in Figure A.6 and take $\alpha \in (1,2)$. Consider a fitness landscape that satisfies*

$$0 \ll 2 \ll 4 \ll 5 \ll 2, \tag{A.3.12}$$

$$1 \ll 2, \qquad 3 \ll 4, \tag{A.3.13}$$

$$f(1,0), f(3,2), f(3,5) < 0 \tag{A.3.14}$$

and assume that there are no polymorphic coexistence equilibria.

After a first jump from $\mathbf{v}^{(0)} = \{0\}$ to $\mathbf{v}^{(1)} = \{2\}$ on the time scale $1/K\mu_K^2$, the next fixating mutant is of trait 4 and arises on the same time scale. The chosen fitness landscape ensures that it grows and can invade the population of trait 2 within a $\ln K$-time. Since $\alpha \in (1,2)$, we obtain a non-vanishing population of trait 5 on the same time scale, which can grow as soon as trait 4 is the new resident trait. Due to its positive invasion fitness, 5 invades the trait 4 population. Finally, the same argument applies for an invasion by trait 2, where we then get stuck in because $\{2\}$ is an ESC of stability degree $L(\{2\}) = 2 > \alpha$.

Overall, we obtain that

$$\mathbf{v}^{(2)} = \mathbf{v}_{\mathrm{ESC}}(\{2\},4) = \{2\}. \tag{A.3.15}$$

In view of Definition A.2.9, this gives rise to the self-connecting edge $(\{2\},\{2\}) \in \mathcal{G}_{\mathrm{ESC}}$, which is illustrated in Figure A.6.

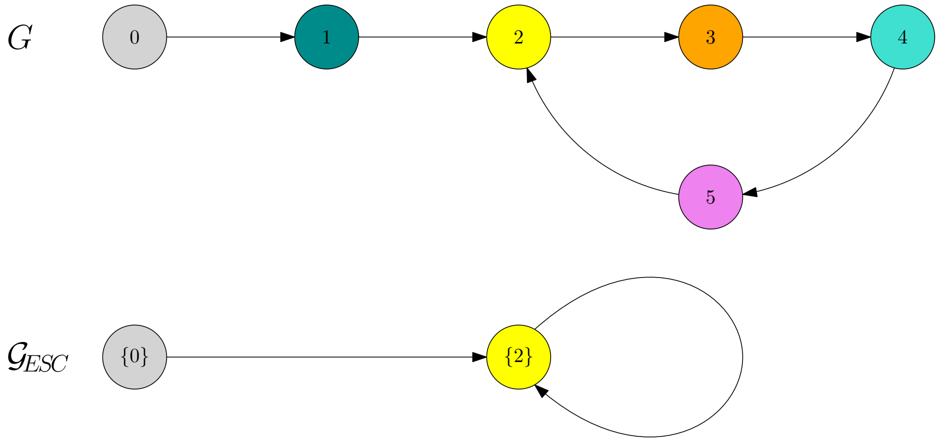


Figure A.6.: Trait graph G and metastability graph $\mathcal{G}_{\mathrm{ESC}}$ of Example 3

On Assumption 3

Since the assumption that prevents the process from getting stuck on a slower time scale is somewhat involved, we give two examples. First, we illustrate in Example 4 that Assumption 3 may hold true even if there is a cycle in the metastability graph $\mathcal{G}_{\mathrm{ESC}}$. Second, we slightly modify the trait graph G and the fitness landscape to get Example 5, where Assumption 3 is not satisfied, and explain why this leads to difficulties.

Example 4. *Let us consider the directed graph depicted in Figure A.7. Let* $\alpha \in (0,1)$ *and consider a fitness landscape that satisfies*

$$0 \ll 2 \ll 3 \ll 4 \ll 6, \tag{A.3.16}$$

$$1 \ll 2, \qquad 5 \ll 6, \tag{A.3.17}$$

$$3 \ll 7 \ll 2, \tag{A.3.18}$$

$$f(1,0), f(5,4) < 0 \tag{A.3.19}$$

and assume that there are no polymorphic coexistence equilibria.

Let us first remark that, because of $\alpha \in (0,1)$, we are in the regime of the trait substitution sequence (cf. [40]). This means that we can neglect the $\ln K$-algorithm. In particular, if $v \ll w$, for some $w \in V_{\mathrm{mut}}(\{v\})$, then $\mathbf{v}_{\mathrm{ESC}}(\{v\}, w) = \{w\}$.

With this knowledge, let us construct the jump chain step by step. The first two jumps are determined easily, noting that

$$\mathbf{v}^{(0)} = \{0\}, \qquad L(\{0\}) = 2, \qquad V_{\mathrm{mut}}(\{0\}) = \{2\}, \tag{A.3.20}$$

$$\mathbf{v}^{(1)} = \{2\}, \qquad L(\{2\}) = 1, \qquad V_{\mathrm{mut}}(\{2\}) = \{3\}, \tag{A.3.21}$$

$$\mathbf{v}^{(2)} = \{3\}, \qquad L(\{3\}) = 1, \qquad V_{\mathrm{mut}}(\{3\}) = \{4,7\}. \tag{A.3.22}$$

For the third jump, there are two possible triggering mutants. If trait 7 fixates first, the process jumps to the ESC $\mathbf{v}^{(3)} = \{7\}$ and then returns to $\mathbf{v}^{(4)} = \{2\}$, all on the time scale

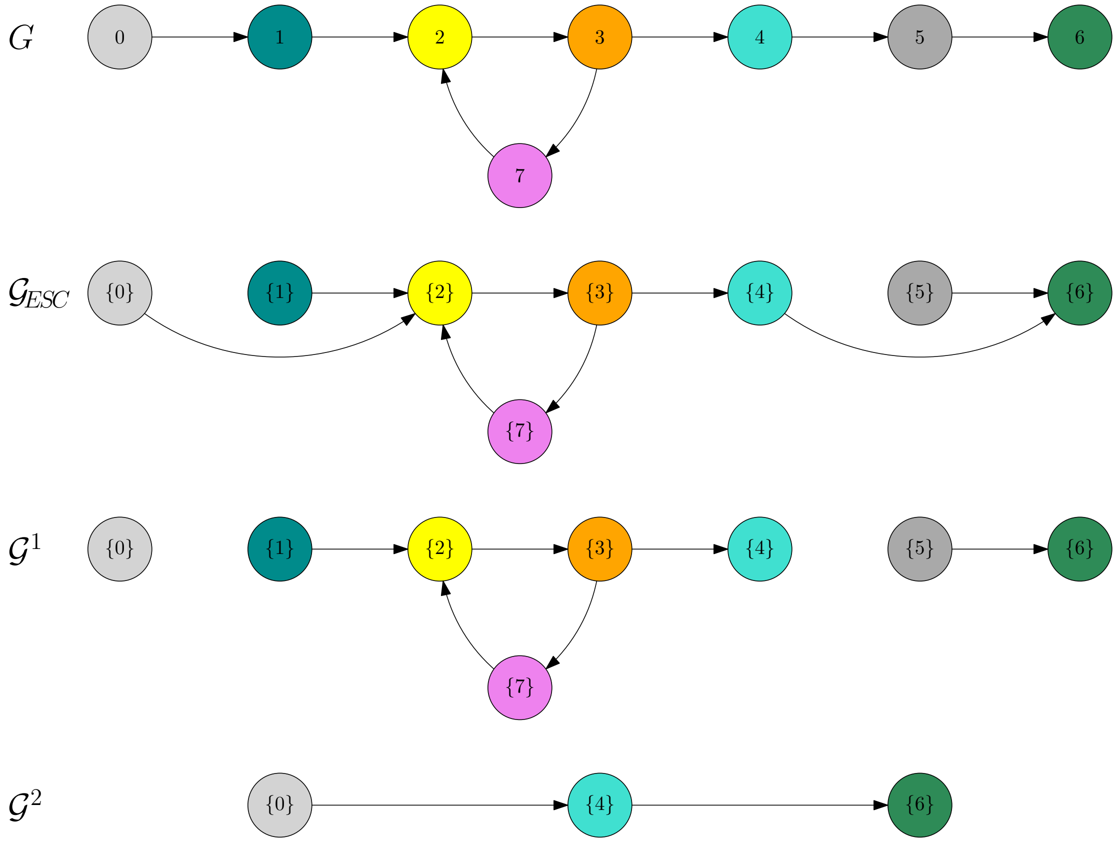


Figure A.7.: Trait graph G, metastability graph $\mathcal{G}_{\mathrm{ESC}}$ and L-scale graphs $\mathcal{G}^1$ and $\mathcal{G}^2$ of Example 4

$1/K\mu_K$. If instead trait 4 fixates earlier, the jump chain continues to $\mathbf{v}^{(3)} = \{4\}$ within a time of order $1/K\mu_K$ and then to $\mathbf{v}^{(4)} = \{6\}$ on the time scale $1/K\mu_K^2$, since $f(5,4) < 0$.

Mentioning that $V_{\mathrm{mut}}(\{1\}) = \{2\}$ and $V_{\mathrm{mut}}(\{5\}) = \{6\}$ gives us the metastability graph drawn in Figure A.7.

To check whether Assumption 3 is satisfied, we decompose the set of ESCs $\mathcal{V}_{\mathrm{ESC}}$ according to the stability degree,

$$\mathcal{S}^1 = \{\{1\},\{2\},\{3\},\{5\},\{7\}\}\,, \qquad \mathcal{S}^2 = \{\{0\},\{4\}\}\,, \qquad \mathcal{S}^\infty = \{\{6\}\}\,. \tag{A.3.23}$$

For all $\mathbf{v} \in \mathcal{S}^1$, one directly sees that an ESC of the same or a higher stability is reached after one jump with probability one. Thus the assumption is true for $L = 1$ and we can construct the graph $\mathcal{G}^1$ as drawn.

In the case of $L = 2$, for $\mathbf{v}^{(0)} = \{4\}$, we obtain that with probability one the process jumps to $\mathbf{v}^{(1)} = \{6\}$, which is of higher stability. Finally, we have to check the most involved case of $\mathbf{v}^{(0)} = \{0\}$. From the metastability graph we identify $\mathbf{v} = \{4\}$ as the only reachable ESC of degree $L \geq 2$. Due to the branching at $\{3\}$, we have to ensure that the process does not get stuck in a cycle of $(\{2\},\{3\},\{7\},\{2\})$ for infinitely many steps. We can see that

$$\mathbb{P}\left(\forall\, k \in \mathbb{N}_{>0} : \mathbf{v}^{(k)} \neq \{4\} | \mathbf{v}^{(0)} = \{0\}\right) = 0 \tag{A.3.24}$$

since the number of cycles that run through before exiting towards $\{4\}$ has a geometric law with success probability $p(\{3\},\{4\}) > 0$. Therefore, Assumption 3 also holds true for $L = 2$. This yields the L-scale graph $\mathcal{G}^2$, depicted in Figure A.7.

Let us now modify the example by inserting an additional trait 8, that can be viewed as an intermediate unfit mutation between 3 and 4. Moreover, for the sake of clarity, we cut off the traits 5 and 6.

Example 5. *Let us consider the directed graph depicted in Figure A.8 and let* $\alpha \in (0,1)$. *Consider a fitness landscape that satisfies*

$$0 \ll 2 \ll 3 \ll 4, \tag{A.3.25}$$

$$3 \ll 7 \ll 2, \tag{A.3.26}$$

$$1 \ll 2, \qquad 8 \ll 4, \tag{A.3.27}$$

$$f(1,0), f(8,3) < 0 \tag{A.3.28}$$

and assume that there are no polymorphic coexistence equilibria.

Since we only changed the trait graph G slightly, also the metastability graph $\mathcal{G}_{\text{ESC}}$ stays almost the same. Apart from the omitted traits 5 and 6, the main difference is that the valley from the ESC $\{3\}$ to the fit mutant 4 is now of width 2. Therefore, trait 4 is no longer one of the nearest fit traits to trait 3 and the set of possible mutants gets reduced to $V_{\text{mut}}(\{3\}) = \{7\}$. In particular, there is no longer an edge $(\{3\},\{4\})$ in the metastability graph.

To check whether Assumption 3 is satisfied, we again separate the stability classes

$$\mathcal{S}^1 = \{\{1\},\{2\},\{3\},\{7\},\{8\}\}, \qquad \mathcal{S}^2 = \{\{0\}\}, \qquad \mathcal{S}^\infty \{\{4\}\}. \tag{A.3.29}$$

For $L = 1$, it is again easy to see from $\mathcal{G}_{\text{ESC}}$ that the assumption holds true. To check this for $L = 2$, we have to consider how the process can get from the initial ESC $\{0\}$ to some ESC of at least the same stability degree. This is not possible since the only candidate would bee $\{4\}$, which is not reachable since the metastability graph is disconnected. As a conclusion, Assumption 3 is not satisfied for $L = 2$ and thus we can neither construct the L-scale graph $\mathcal{G}^2$ nor apply Theorem A.2.11.

Remark 14. Although the population process gets stuck in a cycle between of the ESCs $\{2\},\{3\},\{7\}$ for infinite time, we expect that it might escape through the fitness valley $3 \to 8 \to 4$ eventually, when looking at the time scale $1/K\mu_K^2$. This is due to the fact that, from the microscopic point of view, it is possible to observe mutants of trait 4 in the phases where 3 is the resident trait. Indeed, those mutants appear with a much smaller rate than those of trait 7, but since these phases occur infinitely often, it should only be a question of acceleration to escape from this cycle (c.f. Remark 12).

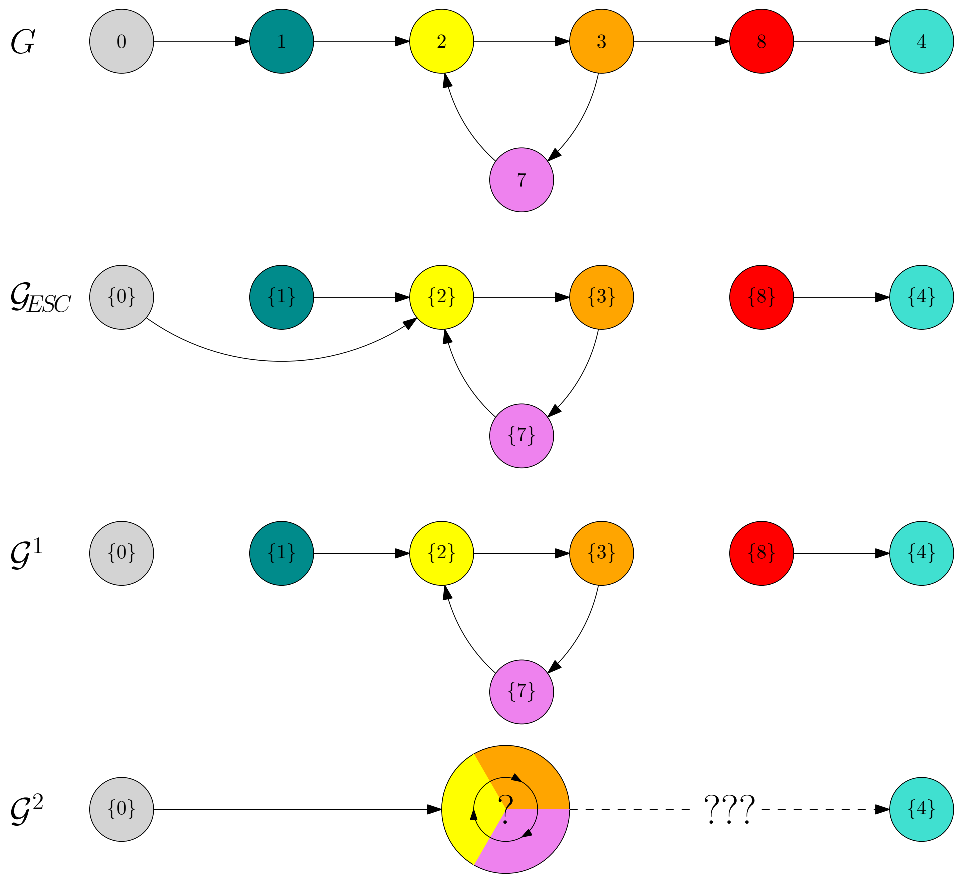


Figure A.8.: Trait graph G, metastability graph $\mathcal{G}_{\text{ESC}}$ and L-scale graphs $\mathcal{G}^1$ and $\mathcal{G}^2$ of Example 5

Collapse on higher time scales

In the two final examples, we demonstrate how paths in the metastability graph that pass through ESCs of different stability degree collapse to a single edge in the L-scale graph when focussing on a particular time scale. To this end we start with an example of a simple linear trait graph with multiple successive fitness valleys of different length. The second example allows for a branching in the metastability graph, which again vanishes in the L-scale graph.

Example 6. *Let us consider the directed graph G depicted in Figure A.9. Assume equal competition and the individual fitness r plotted in Figure A.9. Moreover, let $\alpha \in (1,2)$.*

Due to the linear and directed structure of the trait graph, we can extract the fitness valleys and thus the stability degrees directly from the plotted individual fitness r. The jump chain $(\mathbf{v}^{(k)})_{k\geq 0}$ is the deterministic sequence

$$\mathbf{v}^{(0)} = \{0\}, \qquad \mathbf{v}^{(1)} = \{3\}, \qquad \mathbf{v}^{(2)} = \{5\}, \qquad \mathbf{v}^{(3)} = \{8\}. \tag{A.3.30}$$

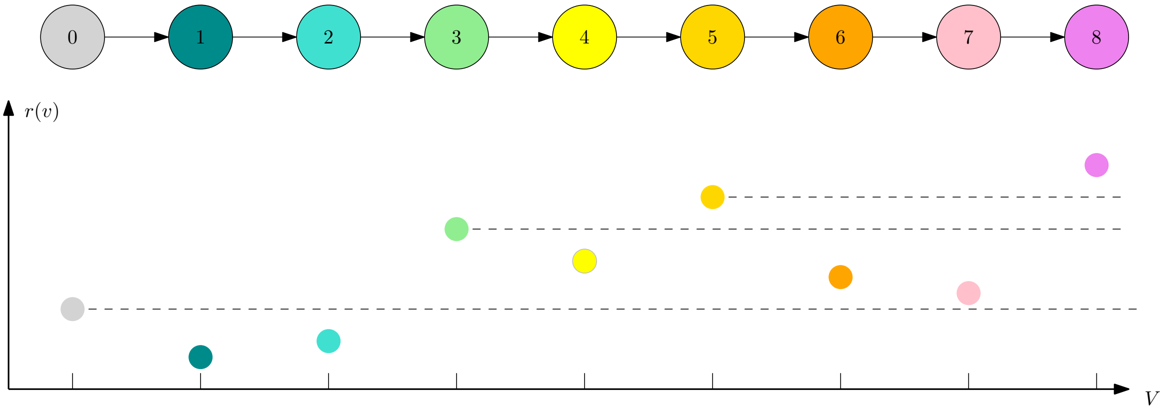


Figure A.9.: Trait graph G and fitness landscape in terms of individual fitness r of Example 6

This is reflected in the metastability graph drawn in Figure A.10. Note that $\{6\}$ is also an ESC of stability degree 2, but it cannot be reached starting from $\{0\}$.

Let us now have a look at the L-scale-graphs, i.e. at how the limiting jump process evolves when fixing a particular time scale. To this end we focus on the sets of ESCs of equal stability degree, namely

$$\mathcal{S}^2 = \{\{3\},\{6\}\}\,, \qquad \mathcal{S}^3 = \{\{0\},\{5\}\}\,, \qquad \mathcal{S}^\infty = \{\{8\}\}\,. \tag{A.3.31}$$

Following our construction in (A.2.38), the L-scale-graph $\mathcal{G}^2$ consists of the vertices $\mathcal{V}^2 = \{\{0\},\{3\},\{5\},\{6\},\{8\}\}$. Since all but $\{3\}$ and $\{6\}$ are of stability degree higher than $L = 2$, the only edges are $\mathcal{E}^2 = \{(\{3\},\{5\}),(\{6\},\{8\})\}$.

The construction of the edges of $\mathcal{G}^3$ is far more interesting. In particular, starting in the initial ESC $\mathbf{v}^{(0)} = \{0\}$, we cannot simply take the edge $(\{0\},\{3\})$ from the metastability graph since $L(\{3\}) < 3$ and thus $\{3\}$ is not stable enough. Instead, we have to consider the whole path $\Gamma = (\{0\},\{3\},\{5\})$ until an ESC of higher stability is reached. This is because the second jump of Γ happens much faster (more precisely on the time scale $1/K\mu_K^2$) and hence becomes absorbed in the slower first jump when rescaling the process with $1/K\mu_K^3$. This gives us one edge in $\mathcal{E}^3$. The second one is given by the jump $(\{5\},\{8\})$. Since $L(\{8\}) = \infty$, no further evolution is possible here.

Overall, these considerations lead to the pictures of $\mathcal{G}^2$ and $\mathcal{G}^3$ in Figure A.10.

Example 7. *Let us consider the directed graph G depicted in Figure A.11. Assume equal competition and the individual fitness r plotted in Figure A.11. Moreover, let $\alpha \in (1,2)$.*

Starting with the resident population in $\mathbf{v}^{(0)} = \{0\}$, we can directly extract from the plotted individual fitness r that only the traits 3 and 5 have positive invasion fitness. Moreover, both can be reached via a path of length $|\gamma| = 3$, namely

$$\gamma^A = (0,1,2,3), \qquad \gamma^B = (0,1,6,5). \tag{A.3.32}$$

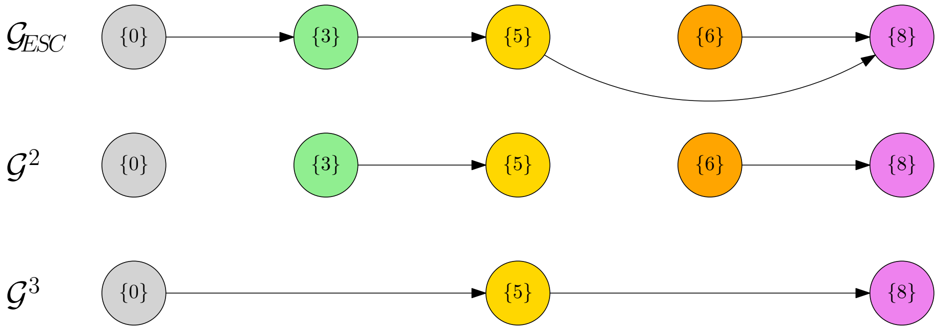


Figure A.10.: Metastability graph $\mathcal{G}_{\text{ESC}}$ and L-scale graphs $\mathcal{G}^2$ and $\mathcal{G}^3$ of Example 6

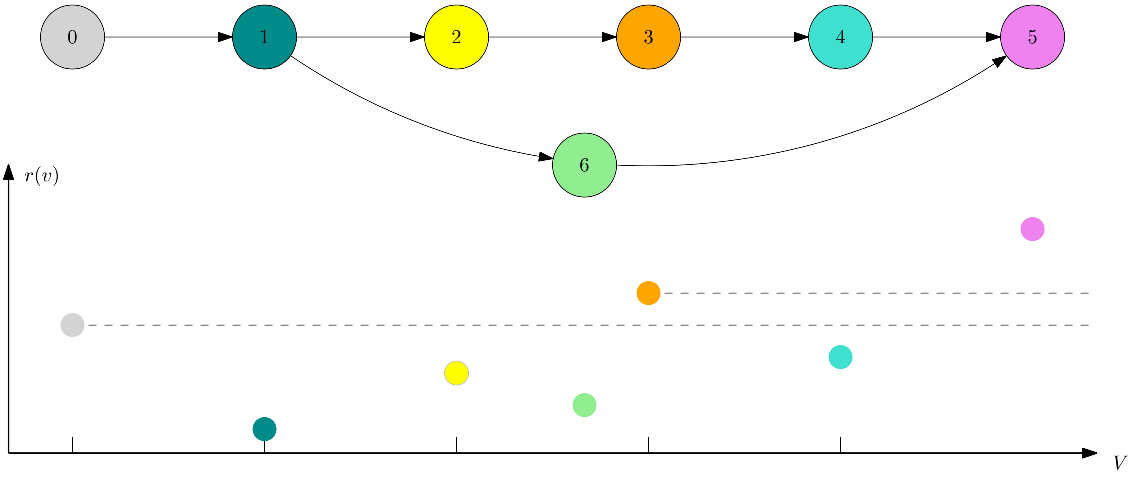


Figure A.11.: Trait graph G and fitness landscape in terms of individual fitness r of Example 7

Hence, we associate to this ESC the stability degree $L(\{0\}) = 3$ and the set of mutant candidates $V_{\text{mut}}(\{0\}) = \{3, 5\}$.

If trait 5 fixates first, there is no further evolution and we end with $\mathbf{v}_{\text{ESC}}(\{0\}, 5) = \{5\}$. In the case where trait 3 fixates, it can grow and becomes macroscopic. Moreover, since $\alpha \in (1, 2)$, the population of trait 4 grows by frequent incoming mutants. However, due to its negative invasion fitness with respect to the resident $\{0\}$ and later against the macroscopic population $\{3\}$, it cannot invade. Hence $\mathbf{v}_{\text{ESC}}(\{0\}, 3) = \{3\}$ is the corresponding ESC and is of stability degree $L(\{3\}) = 2$. From thereon, only trait 5 is a fit reachable mutant, which arises after a waiting time of order $\mathcal{O}(1/K\mu_K^2)$ and replaces 3 as an ESC. Those three jumps form the edges of the drawn metastability graph $\mathcal{G}_{\text{ESC}}$ in Figure A.12.

The L-scale-graph $\mathcal{G}^2$ is constructed easily whereas the really interesting behaviour occurs when asking for the $\mathcal{G}^3$. Since $L(\{0\}) = 3$, the jumps $(\{0\}, \{3\})$ and $(\{0\}, \{5\})$ happen on the visible time scale. The latter one is clearly also an edge in $\mathcal{G}^3$, due to the high stability of the final ESC $L(\{5\}) = \infty$. However, in case of the former, the ESC that the process jumps to is of smaller stability, i.e. $L(\{3\}) = 2$. Therefore, the next jump $(\{3\}, \{5\})$ directly occurs within a time that vanishes under rescaling. The path $\Gamma = (\{0\}, \{3\}, \{5\})$ in $\mathcal{G}_{\text{ESC}}$ thus yields an edge $(\{0\},\{5\})$ for $\mathcal{G}^3$. This edge already exists and we do not allow for double edges in $\mathcal{G}^L$. However, the two edges are merged in the sense of adding up the transition rates and probabilities as in (A.2.40).

Overall, we see that even a branching in the metastability graph can disappear when multiple paths collapse to the same edge on a particular time scale.

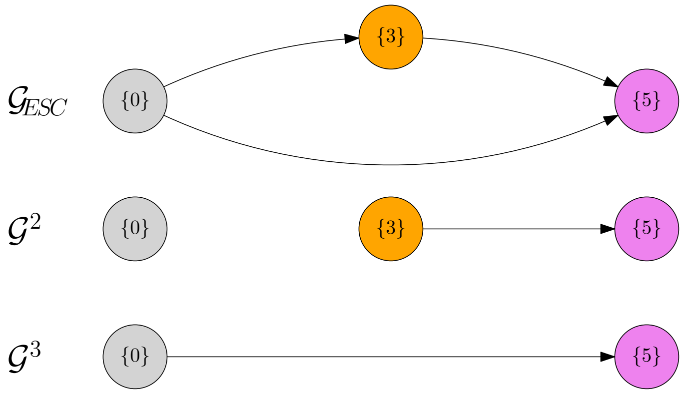


Figure A.12.: Metastability graph $\mathcal{G}_{\text{ESC}}$ and L-scale graphs $\mathcal{G}^2$ and $\mathcal{G}^3$ of Example 7

A.4. Proofs

In this chapter, we discuss the proofs of the results on metastable transitions and limiting jump processes that are presented in Sections A.2.3 and A.2.4. These build on the results in [29] on the crossing of a fitness valley on a linear trait space and in [48] on the faster $\ln K$-dynamics on general finite graphs. The main idea is to extend the techniques from [29] to more complex trait spaces by considering sequential mutations along certain paths. Since mutations are very rare outside of the mutation spreading neighbourhood of the resident traits and unfit traits quickly go extinct, mutations along different paths can essentially be regarded as independent. Consequently, the overall rate of transitioning out of an ESC is obtained by summing over the rates of taking specific paths through the surrounding fitness valley.

The remaining chapter is structured as follows: In Section A.4.1, we determine the precise equilibrium size of the subpopulations with traits inside the mutation spreading neighbourhood. In Section A.4.2, we consider the rates at which mutants of any fitness arise along specific paths and combine these to the overall rate at which single mutants are born. Finally, in Section A.4.3, we combine these rates of producing mutants beyond the fitness valley with the probability of fixation and the faster $\ln K$-dynamics of reaching a new ESC to conclude Theorem A.2.7 and Corollary A.2.8. Section A.4.4 is dedicated to the proof of Corollary A.2.10 and Theorem A.2.11, where we concatenate several jumps across fitness valleys to obtain the multi-scale jump chain and carefully study which transitions are visible on the respective time scales to obtain the dynamics of the limiting Markov jump process.

A.4.1. Estimation of the equilibrium size

In this section we discuss the equilibrium population sizes of the living traits once an ESC is obtained. The results from [48] only characterize the orders of population sizes β_w and the actual size $\bar{n}(\mathbf{v})$ of the resident traits associated to an ESC. In order to calculate the precise transition rates from one ESC to another, we do, however, need a better estimate for the population sizes of the non-resident traits in $V_\alpha(\mathbf{v})$.

We prove that, if the initial conditions of our process satisfy the assumptions of an asymptotic ESC, all living traits in $V_\alpha(\mathbf{v})$ get arbitrarily close to their equilibrium size within a finite time. This equilibrium size preserves the orders of population sizes and is of the form

$$N_v^K(t) = a_v K \mu_K^{d(\mathbf{v},v)} + o\left(K \mu_K^{d(\mathbf{v},v)}\right) \qquad \forall v \in V_\alpha(\mathbf{v}), \tag{A.4.1}$$

for some $a_v \in \mathbb{R}_+$, which can be calculated precisely. The populations of living traits stay close to these equilibrium sizes as long as no new trait arises and reaches a size at which it can influence the population sizes of other traits, i.e. a size of order $K^{1/\alpha}$. To this extend, we recall the definition of the stopping time

$$T_{\text{fix}}^K := \inf\left\{t \geq 0 : \exists w \in V \backslash V_\alpha(\mathbf{v}) : \beta_w^K(t) \geq 1/\alpha\right\}. \tag{A.4.2}$$

Lemma A.4.1 (Equilibrium size inside the α-radius)**.** *Let* $\mathbf{v} \subset V$ *and* $(\beta^K(0))_{K\geq 0}$ *be an asymptotic ESC. Then, for all* $\varepsilon > 0$*, there exist constants* $\tau_\varepsilon < \infty$*,* $U_\varepsilon > 0$ *and Markov processes* $\left(N_v^{(K,\pm)}(t), t \geq 0\right)_{K\geq 0}$ *such that,*

$$\lim_{K\to\infty} \mathbb{P}\left(N_v^{(K,-)}(t) \leq N_v^K(t) \leq N_v^{(K,+)}(t) \ \forall t \in (\tau_\varepsilon, T_{\mathit{fix}}^K \wedge \mathrm{e}^{U_\varepsilon K}), \ v \in V_\alpha(\mathbf{v})\right) = 1 \tag{A.4.3}$$

and

$$\left| \frac{\mathbb{E}\left[N_v^{(K,\pm)}(t)\right]}{K \mu_K^{d(\mathbf{v},v)}} - a_v \right| < \varepsilon \qquad \forall t \geq \tau_\varepsilon, \tag{A.4.4}$$

where

$$a_v := \sum_{\substack{\gamma:\mathbf{v}\to v \\ |\gamma|=d(\mathbf{v},v)}} \bar{n}_{\gamma_0}(\mathbf{v}) \prod_{i=1}^{|\gamma|} \frac{b(\gamma_{i-1}) m(\gamma_{i-1},\gamma_i)}{|f(\gamma_i,\mathbf{v})|}. \tag{A.4.5}$$

Proof. We will prove the claim by induction w.r.t. the distance from the resident traits. For the initialisation let us start with $v \in \mathbf{v}$. That is, we count also a single vertex as a path of length zero together with the convention that an empty product has the value one. In this case $(N_v^K, \ v \in \mathbf{v})$ can be coupled with logistic birth-death processes with immigration, by estimating the incoming and outgoing mutants, which are of order $\mathcal{O}(K\mu_K)$ or smaller. Hence we know already from [48, Lemma A.6(ii)] that the residents stabilize near their Lotka-Volterra-equilibrium within a time of order $\mathcal{O}(1)$. To make this more precise, define, for all $\varepsilon > 0$, the stopping time when the resident populations enter an ε-neighbourhood of their equilibrium size

$$\tau_\varepsilon^K := \inf\left\{t \geq 0 : \forall v \in \mathbf{v} : \left|K^{-1}N_v^K(t) - \bar{n}_v(\mathbf{v})\right| < \varepsilon C\right\}. \tag{A.4.6}$$

Here C is a constant, depending only on the competition rates $c(v,w)$, which compensates the slight shift of the equilibrium due to small fluctuations of non-resident traits. Then there exists a constant time $\tilde{\tau}_\varepsilon < \infty$, such that $\lim_{K\to\infty} \mathbb{P}\left(\tau_\varepsilon^K < \tilde{\tau}_\varepsilon\right) = 1$. After this time $\tilde{\tau}_\varepsilon$, the environment of competitive pressure stays almost constant, unless the fluctuations of the resident populations become too big or the non-residents reach a macroscopic level. These two events are described by the stopping times

$$S_\varepsilon^K := \inf\left\{t \geq \tau_\varepsilon^K : \exists v \in \mathbf{v} : \left|K^{-1}N_v^K(t) - \bar{n}_v(\mathbf{v})\right| > 2\varepsilon C\right\} \tag{A.4.7}$$

and

$$\sigma_\varepsilon^K := \inf\left\{t \geq 0 : \sum_{w\in V\backslash\mathbf{v}} N_w^K(t) \geq \varepsilon K\right\}. \tag{A.4.8}$$

We know from [43, Propostition A.2] that, for some constant $U_\varepsilon > 0$,

$$\lim_{K\to\infty} \mathbb{P}\left(S_\varepsilon^K > \mathrm{e}^{U_\varepsilon K} \wedge \sigma_\varepsilon^K\right) = 1. \tag{A.4.9}$$

For the other traits in the α-radius $v \in V_\alpha\backslash\mathbf{v}$ we prove as the induction step that (A.4.3) is satisfied with

$$a_v = \sum_{\substack{(w,v)\in E\\ d(\mathbf{v},w)=d(\mathbf{v},v)-1}} a_w \frac{b(w)m(w,v)}{|f(v,\mathbf{v})|} \tag{A.4.10}$$

by deriving an upper and a lower bound on the population size through couplings. These bounds then immediately imply the claim.

Following the notation of [83], we represent the population processes in terms of Poisson random measures. For this purpose let $(Q_v^{(b)}, Q_v^{(d)}, Q_{w,v}^{(m)}; v, w, \in V)$ be independent homogeneous Poisson random measures on $\mathbb{R}_+^2$ with intensity $\delta s\delta\theta$. Then we can write

$$\begin{aligned} N_v^K(t) = N_v^K(0) &+ \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta\leq b(v)(1-\mu_K)N_v^K(s^-)} Q_v^{(b)}(\mathrm{d}s, \mathrm{d}\theta) \\ &- \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta\leq [d(v)+\sum_{w\in V} c^K(v,w)N_w^K(s^-)]N_v^K(s^-)} Q_v^{(d)}(\mathrm{d}s, \mathrm{d}\theta) \\ &+ \sum_{(w,v)\in E} \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta\leq \mu_K b(w)m(w,v)N_w^K(s^-)} Q_{w,v}^{(m)}(\mathrm{d}s, \mathrm{d}\theta). \end{aligned} \tag{A.4.11}$$

Note that we use the same Poisson measures to construct the processes for each K here. However, as already pointed out in Section A.2.1, this is not necessary and we do not use any particular correlation between the processes for different K. We can use a specific joint construction here since we are only considering the convergence of probabilities of certain events, rather than of the processes themselves.

Since we already know from [48, Theorem 2.2] that in the equilibrium state the non-resident populations $w \in V_\alpha(\mathbf{v})$ stay of size $\mathcal{O}(K\mu_K^{d(\mathbf{v},w)})$, the main part of the mutations in the last line comes only from traits lying closer to the resident traits. Thus we can adopt the inductive structure of [29, Lemma 7.1] and approximate the population size of v analogously by coupling it, for K large enough, with two processes

$$N_v^{(K,-)}(t) \leq N_v^K(t) \leq N_v^{(K,+)}(t), \quad \forall \tilde{\tau}_\varepsilon \leq t \leq \sigma_\varepsilon^K \wedge T_{\text{fix}}^K \wedge S_\varepsilon^K. \tag{A.4.12}$$

To be precise, we take care of the admissible fluctuations of the residents by defining

$$\bar{n}_v^{(\pm)}(\mathbf{v}) := \bar{n}_v(\mathbf{v}) \pm 2\varepsilon C. \tag{A.4.13}$$

Then, for $v \in V \backslash \mathbf{v}$ and $\mu_K < \varepsilon$, we set

$$\begin{aligned} N_v^{(K,-)}(t) = N_v^K(\tilde{\tau}_\varepsilon) &+ \int_{\tilde{\tau}_\varepsilon}^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq b(v)(1-\varepsilon) N_v^{(K,-)}(s-)} Q_v^{(b)}(\delta s, \delta\theta) \\ &- \int_{\tilde{\tau}_\varepsilon}^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq [d(v) + \sum_{w \in \mathbf{v}} c(v,w) \bar{n}_w^{(+)}(\mathbf{v}) + \varepsilon \max_{\tilde{w} \in V \backslash \mathbf{v}} c(v,\tilde{w})] N_v^{(K,-)}(s-)} Q_v^{(d)}(\delta s, \delta\theta) \\ &+ \sum_{(w,v) \in E} \int_{\tilde{\tau}_\varepsilon}^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq \mu_K b(w) m(w,v) N_w^K(s-)} Q_{w,v}^{(m)}(\delta s, \delta\theta) \end{aligned} \tag{A.4.14}$$

and

$$\begin{aligned} N_v^{(K,+)}(t) =& N_v^K(\tilde{\tau}_\varepsilon) + \int_{\tilde{\tau}_\varepsilon}^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq b(v) N_v^{(K,+)}(s-)} Q_v^{(b)}(\delta s, \delta\theta) \\ &- \int_{\tilde{\tau}_\varepsilon}^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq [d(v) + \sum_{w \in \mathbf{v}} c(v,w) \bar{n}_w^{(-)}(\mathbf{v})] N_v^{(K,+)}(s-)} Q_v^{(d)}(\delta s, \delta\theta) \\ &+ \sum_{(w,v) \in E} \int_{\tilde{\tau}_\varepsilon}^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq \mu_K b(w) m(w,v) N_w^K(s-)} Q_{w,v}^{(m)}(\delta s, \delta\theta), \end{aligned} \tag{A.4.15}$$

where we use the same Poisson measures as in (A.4.11). Note that this coupling satisfies (A.4.12) only on the event $\left\{\tau_\varepsilon^K < \tilde{\tau}_\varepsilon\right\}$. However, as mentioned above, this event's probability converges to 1 and we can hence restrict our considerations to this case to obtain the desired convergence.

On closer inspection, the approximating processes $N_v^{(K,-)}, N_v^{(K,+)}$ are nothing but subcritical branching processes with immigration stemming from incoming mutations.

Similar to the proof of [29, Equation (7.8) et sqq.] we can use the martingale decomposition of $N_v^{(K,+)}$ and $N_v^{(K,-)}$ to derive, for $t > \tilde{\tau}_\varepsilon$, the differential equation

$$\begin{aligned}
&\frac{\delta}{\delta t}\mathbb{E}\left[N_v^{(K,*)}(t)\right] \\
&\quad= \left(b(v)(1-\mathbb{1}_{\{*=-\}}\varepsilon) - d(v) - \sum_{w\in\mathbf{v}} c(v,w)\bar{n}_w^{(\bar{*})}(\mathbf{v}) - \mathbb{1}_{\{*=-\}}\varepsilon \sup_{\tilde{w}\in V\setminus\mathbf{v}} c(v,\tilde{w})\right) \\
&\qquad\times \mathbb{E}\left[N_v^{(K,*)}(t)\right] + \sum_{(w,v)\in E} \mu_K b(w)m(w,v)\mathbb{E}\left[N_w^K(t)\right] \qquad \text{(A.4.16)} \\
&\quad= f^{(*)}(v,\mathbf{v})\mathbb{E}\left[N_v^{(K,*)}(t)\right] + \sum_{(w,v)\in E} \mu_K b(w)m(w,v)\mathbb{E}\left[N_w^K(t)\right], \qquad \text{(A.4.17)}
\end{aligned}$$

where $\bar{*} = \{+,-\}\backslash *$ denotes the inverse sign.

Here, we introduce $f^{(*)}(v,\mathbf{v})$ as a short notation to point out that this is nothing but a perturbation of the invasion fitness. Then we can apply our a priori knowledge on the size of the sub-populations, i.e.

$$\mathbb{E}\left[N_w^K(t)\right] = \mathcal{O}\left(K\mu_K^{\delta(\mathbf{v},w)}\right) \qquad \forall w \in V_\alpha(\mathbf{v}), \tag{A.4.18}$$

to rewrite the ODE system

$$\begin{aligned}
\frac{\delta}{\delta t}\mathbb{E}\left[N_v^{(K,*)}(t)\right] &= f^{(*)}(v,\mathbf{v})\mathbb{E}\left[N_v^{(K,*)}(t)\right] + \sum_{\substack{(w,v)\in E\\ d(\mathbf{v},w)=d(\mathbf{v},v)-1}} \mu_K b(w)m(w,v)\mathbb{E}\left[N_w^K(t)\right] \\
&\qquad + \mathcal{O}\left(K\mu_K^{\delta(\mathbf{v},v)+1}\right) \qquad \text{(A.4.19)} \\
&= f^{(*)}(v,\mathbf{v})\mathbb{E}\left[N_v^{(K,*)}(t)\right] + \sum_{\substack{(w,v)\in E\\ d(\mathbf{v},w)=d(\mathbf{v},v)-1}} b(w)m(w,v)a_w K\mu_K^{\delta(\mathbf{v},v)} \\
&\qquad + o\left(K\mu_K^{\delta(\mathbf{v},v)}\right). \qquad \text{(A.4.20)}
\end{aligned}$$

Here we use the induction hypothesis to estimate the populations with traits lying closer to the residents in the latter equality.

Rescaling with $K\mu_K^{d(v,\mathbf{v})}$ and using (A.4.10), the equation becomes

$$\frac{\delta}{\delta t}\mathbb{E}\left[\frac{N_v^{(K,*)}(t)}{K\mu_K^{\delta(\mathbf{v},v)}}\right] = f^{(*)}(v,\mathbf{v})\mathbb{E}\left[\frac{N_v^{(K,*)}(t)}{K\mu_K^{\delta(\mathbf{v},v)}}\right] + a_v\left|f(v,\mathbf{v})\right| + o(1). \tag{A.4.21}$$

By variation of constants the solution is given by

$$\begin{aligned}
\mathbb{E}\left[\frac{N_v^{(K,*)}(t)}{K\mu_K^{\delta(\mathbf{v},v)}}\right] &= \mathrm{e}^{f^{(*)}(v,\mathbf{v})(t-\tilde{\tau}_\varepsilon)}\left(\mathbb{E}\left[\frac{N_v^K(\tilde{\tau}_\varepsilon)}{K\mu_K^{\delta(\mathbf{v},v)}}\right] - \frac{|f(v,\mathbf{v})|}{|f^{(*)}(v,\mathbf{v})|}a_v + o(1)\right) \\
&\qquad + \frac{|f(v,\mathbf{v})|}{|f^{(*)}(v,\mathbf{v})|}a_v + o(1)
\end{aligned} \tag{A.4.22}$$

Note that the term in brackets can be bounded uniformly in K and ε, for ε small enough. Moreover, the ratio of (perturbed) fitness can be expressed as $(1 \pm \varepsilon\tilde{c}_\varepsilon)$. So (A.4.22) becomes

$$\mathbb{E}\left[\frac{N_v^{(K,*)}(t)}{K\mu_K^{\delta(\mathbf{v},v)}}\right] = \mathrm{e}^{f^{(*)}(v,\mathbf{v})(t-\tilde{\tau}_\varepsilon)}\mathcal{O}(1) + (1 \pm \varepsilon\tilde{c}_\varepsilon)a_v + o(1) \tag{A.4.23}$$

Finally taking into account that the fitness $f^{(*)}(v,\mathbf{v}) < 0$ is negative for $v \in V_\alpha(\mathbf{v})$ the first term vanishes for increasing time. Hence we see that for all $\tilde{\varepsilon} > 0$ there are $\varepsilon > 0$ and $\tau_{\tilde{\varepsilon}} \in (\tilde{\tau}_\varepsilon, \infty)$ and $K_0 \in \mathbb{N}$ such that, for all $t > \tau_{\tilde{\varepsilon}}$ and $K > K_0$

$$\left|\mathbb{E}\left[\frac{N_v^{(K,*)}(t)}{K\mu_K^{\delta(\mathbf{v},v)}}\right] - a_v\right| < \tilde{\varepsilon}. \tag{A.4.24}$$

Finally, we can deduce again from our knowledge on the orders of population sizes that

$$\lim_{K\to\infty} \mathbb{P}\left(\sigma_\varepsilon^K < T_{\text{fix}}^K \wedge \mathrm{e}^{U_\varepsilon K}\right) = 0, \tag{A.4.25}$$

which allows us to drop the stopping time σ_ε^K in the claim. □

A.4.2. Pathwise evolution rates

From the precise description of the population sizes inside the mutation spreading neighbourhood we can now deduce the rate of occurrence of mutants that lay outside.

To observe a new mutant, whose trait is far away from the resident population, a whole sequence of mutation steps is needed. Traits outside the α-neighbourhood $V_\alpha(\mathbf{v})$ cannot avoid extinction only due to incoming mutants. Therefore, if such a trait has negative invasion fitness, mutants only give rise to small excursions approximated by subcritical branching processes. During each of these excursions there is a small probability that a new mutant is produced before extinction.

To overcome the problem of tracking possible back mutations, we not only observe the sizes of the different mutant populations. Instead, we distinguish mutants by the mutational path along which they arose and keep track of the genealogy. We set

$$N_v^K(t) = \sum_{\gamma:\partial V_\alpha \to v} N_{v,\gamma}^K(t) \qquad \forall v \in V \backslash V_\alpha, \tag{A.4.26}$$

where the pathwise mutations can by represented by

$$\begin{aligned} N_{v,\gamma}^K(t) &= \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le b(v)(1-\mu_K)N_{v,\gamma}^K(s^-)} Q_{v,\gamma}^{(b)}(\delta s, \delta\theta) \\ &+ \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le \mu_K b(\tilde{v}) m(\tilde{v},v) N_{\tilde{v},\gamma\backslash v}^K(s^-)} Q_{v,\gamma}^{(m)}(\delta s, \delta\theta) \\ &- \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le [d(v)+\sum_{w\in V} c^K(v,w)N_w^K(s^-)]N_{v,\gamma}^K(s^-)} Q_{v,\gamma}^{(d)}(\delta s, \delta\theta). \end{aligned} \tag{A.4.27}$$

Here $\tilde{v}$ stands for the next-to-last vertex in γ, which is the progenitor of v in γ, and for $\tilde{v} \in \partial V_\alpha$ we set

$$N_{\tilde{v},(\tilde{v})}(t) := N_{\tilde{v}}(t). \tag{A.4.28}$$

As before, $(Q^{(b)}_{v,\gamma}, Q^{(d)}_{v,\gamma}, Q^{(m)}_{v,\gamma}; v \in V,\ \gamma : \partial V_\alpha \to v)$ are independent homogeneous Poisson random measures with constant intensity one.

Remark 15. It suffices to only sum over the paths starting in ∂V_α in the decomposition. By the definition of T^K_{ESC} all populations outside of V_α are extinct at that time. The probability that a mutant of trait $v \in V \backslash V_\alpha$ arises before the finite time τ_ε in Lemma A.4.1, when the populations in V_α reach their equilibrium, goes to zero. After this time we have good bounds on the population sizes of all traits in V_α and it is therefore sufficient to trace back the genealogy of new mutants to the last trait in V_α, i.e. a trait in ∂V_α.

With this representation at hand, we are now able to define the cumulated number of mutant individuals of trait v that arose as mutants of the progenitor $\tilde{v}$, along the path γ

$$M^K_{v,\gamma}(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq \mu_K b(\tilde{v}) m(\tilde{v},v) N^K_{\tilde{v},\gamma\backslash v}(s-)} Q^{(m)}_{v,\gamma}(\delta s, \delta\theta), \tag{A.4.29}$$

as well as the respective occurrence times of these mutants

$$T^{(i,K)}_{v,\gamma} := \inf \left\{ t \geq 0 : M^K_{v,\gamma}(t) \geq i \right\}, \tag{A.4.30}$$

where we set $T^{(0,K)}_{v,\gamma} := 0$.

Our aim is to show that new mutants outside of V_α appear at the end of a mutation path approximately as a Poisson point process with rate scaling with length of the path.

Lemma A.4.2. *Suppose* $\mathbf{v}$ *and* $(\beta^K(0))_{K \geq 0}$ *are an asymptotic ESC and let* T^K_{fix} *be defined as in* (A.4.2)*. Let* $v \in V \backslash V_\alpha$ *and* $\gamma : \partial V_\alpha \to v$ *be such that* $|\gamma| \geq L - \lfloor \alpha \rfloor$ *and* $f(\gamma_i, \mathbf{v}) < 0$*, for all* $i = 0, \dots, |\gamma| - 1$*. Then there exist* $0 < c, C < \infty$ *such that, for each* $\varepsilon > 0$*, there exist two Poisson point processes* $M^{(K,\pm)}_{v,\gamma}$ *with rates* $\tilde{R}^{(\pm)}_{v,\gamma} K \mu_K^{\lfloor \alpha \rfloor + |\gamma|}$ *such that*

$$\liminf_{K \to \infty} \mathbb{P}\left(M^{(K,-)}_{v,\gamma}(t) < M^K_{v,\gamma}(t) < M^{(K,+)}_{v,\gamma}(t),\ \forall t < T^K_{\mathit{fix}} \right) \geq 1 - c\varepsilon, \tag{A.4.31}$$

where the rate parameters are defined as

$$\tilde{R}_{v,\gamma} := a_{\gamma_0} b(\gamma_0) m(\gamma_0, \gamma_1) \prod_{j=1}^{|\gamma|-1} \lambda(\rho(\gamma_j, \mathbf{v})) m(\gamma_j, \gamma_{j+1}), \quad \tilde{R}^{(\pm)}_{v,\gamma} = (1 \pm C\varepsilon) \tilde{R}_{v,\gamma}. \tag{A.4.32}$$

For the definitions of $\lambda(\rho)$ and $\rho(v, \mathbf{v})$ we refer to (A.2.20) and (A.2.19) respectively, while a_{γ_0} is the equilibrium size defined in (A.4.5).

Proof. Note that, throughout the whole proof, we assume that $\tau_\varepsilon < t < T^K_{\text{fix}} \wedge e^{U_\varepsilon K}$, where τ_ε and U_ε are defined in Lemma A.4.1. This can then be extended to all $0 \le t < T^K_{\text{fix}}$ in the limit of $K \to \infty$ since $T^K_{\text{fix}} < e^{U_\varepsilon K}$ with probability converging to 1 and, since $\mu_K \to 0$, there is almost surely no mutation event during the finite time interval $[0, \tau_\varepsilon]$.

Let $v \in V \setminus V_\alpha$ and $\gamma : \partial V_\alpha \to v$ be given as in the Lemma. To better distinguish from the full path γ, we refer to the vertices of the path via $\gamma = (v_0, v_1, \dots, v_{|\gamma|})$. The idea of this proof is to considered the path isolated from the remaining graph and adapt the tools from [29, Ch. 7.3.] to the present situation. We refrain from adding much more notation to our already complicated situation. We try to handle the far more general structure of our trait graph by translating the notation of the central objects between the articles instead.

The first observation is that, for every $t < T^K_{\text{fix}}$, we can bound the mutant counting process of trait v_1 by

$$M^{(K,-)}_{v_1,\gamma}(t) \le M^K_{v_1,\gamma}(t) \le M^{(K,+)}_{v_1,\gamma}(t) \quad \text{a.s.}, \tag{A.4.33}$$

with the bounding processes being defined as

$$M^{(K,\pm)}_{v_1,\gamma}(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le \mu_K b(v_0) m(v_0,v_1) N^{(K,\pm)}_{v_0}(s^-)} Q^{(m)}_{v_0,\gamma}(\delta s, \delta\theta). \tag{A.4.34}$$

Note that the estimate corresponds to equation (7.42) in [29], while the definition is the adapted version of (7.17) therein. In order make use of Lemma A.4.1, we continue temporarily with the simplified processes

$$\bar{M}^{(K,\pm)}_{v_1,\gamma}(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le \mu_K b(v_0) m(v_0,v_1) \mathbb{E}\left[N^{(K,\pm)}_{v_0}(s^-)\right]} Q^{(m)}_{v_0,\gamma}(\delta s, \delta\theta) \tag{A.4.35}$$

and

$$\bar{T}^{(i,K,\pm)}_{v_1,\gamma} := \inf\left\{ t \ge 0 : \bar{M}^{(K,\pm)}_{v_1,\gamma}(t) \ge i \right\}. \tag{A.4.36}$$

In fact, this turns out to be sufficient for our results since a standard application of Doob's martingale inequality shows that, with probability converging to 1, the difference of the processes $M^{(K,\pm)}_{v_1,\gamma}$ and $\bar{M}^{(K,\pm)}_{v_1,\gamma}$ during the relevant time interval stays of sufficiently small order. To be precise there exist sequences of numbers $N_1(K)$ and $N_2(K)$, with

$$N_1(K) \gg (K\mu_K^L)^{-1} \quad \text{and} \quad N_2(K) \ll (\mu_K^{L-1-\lfloor\alpha\rfloor})^{-1} \tag{A.4.37}$$

such that

$$\lim_{K\to\infty} \mathbb{P}\left(\sup_{s \le N_1(K)} \left| M^{(K,\pm)}_{v_1,\gamma}(s) - \bar{M}^{(K,\pm)}_{v_1,\gamma}(s) \right| > N_2(K) \right) = 0. \tag{A.4.38}$$

For details, see [29, p.3583]. At each time $\bar{T}^{(i,K,\pm)}_{v_1,\gamma}$ an individual of trait v_1 is born. In order to track its descendants until potentially a trait $v_{|\gamma|}$ individual is born, in a similar way as done in the previous section, we couple the k-mutant population, for $1 \leq k \leq |\gamma| - 1$, to birth-death-processes with individual birth and death rates

$$b^{(*)}(v_k) = b(v_k)(1 - \mathbb{1}_{\{*=-\}}\varepsilon), \tag{A.4.39}$$

$$d^{(*)}(v_k) = d(v_k) + \sum_{w \in \mathbf{v}} c(v_k, w)\bar{n}_w^{(*)}(\mathbf{v}) + \mathbb{1}_{\{*=-\}}\varepsilon \sup_{\tilde{w} \in V \setminus \mathbf{v}} c(v_k, \tilde{w}). \tag{A.4.40}$$

Note that in contrast to Section A.4.1, these subcritical processes do not gain from any immigration and hence go extinct in finite time. However, there is a small probability during such an excursion of the k-mutant population that an individual of trait $(k+1)$ is born. Analogously to [29, pp. 3581–3582], we can use Lemma A.5.1 (see Appendix A.5.1) to derive

$$\begin{aligned}\mathbb{P}\,&(\text{An excursion of trait } v_k \text{ produces exactly 1 mutant of type } v_{k+1}) \\ &= \mu_K \lambda(\rho(v_k, \mathbf{v}))m(v_k, v_{k+1})(1 + \mathcal{O}(\varepsilon)),\end{aligned} \tag{A.4.41}$$

while on the other hand

$$\mathbb{P}\,(\text{An excursion of trait } v_k \text{ produces at least 2 mutants of } v_{k+1}) = \mathcal{O}(\mu_K^2). \tag{A.4.42}$$

Hence, the probability that the i-th mutant of trait v_1 (i.e. the one triggering $\bar{T}_{v_1,\gamma}^{(i,K,\pm)}$) produces a $v_{|\gamma|}$-mutant is, for large K,

$$\mu_K^{|\gamma|-1} \left(\prod_{k=\lfloor \alpha \rfloor + 1}^{|\gamma|-1} \lambda(\rho(v_k, \mathbf{v}))m(v_k, v_{k+1}) \right) (1 + \mathcal{O}(\varepsilon)). \tag{A.4.43}$$

Since Lemma A.4.1 implies that $\bar{M}_{v_1,\gamma}^{(K,\pm)}$ can be treated as a Poison process with intensity

$$K\mu_K^{d(\mathbf{v},v_0)+1} a_v b(v_0) m(v_0, v_1), \tag{A.4.44}$$

we get appearance of $v_{|\gamma|}$-mutants also as Poison process with thinned intensity

$$K\mu_K^{d(\mathbf{v},v_0)+|\gamma|} a_v b(v_0) m(v_0, v_1) \left(\prod_{k=\lfloor \alpha \rfloor + 1}^{|\gamma|-1} \lambda(\rho(v_k, \mathbf{v}))m(v_k, v_{k+1}) \right) (1 + \mathcal{O}(\varepsilon)) \tag{A.4.45}$$

$$= \tilde{R}_{v,\gamma}^{(\pm)} K \mu_K^{\lfloor \alpha \rfloor + |\gamma|}. \tag{A.4.46}$$

Eventually, the difference between $M_{v_1,\gamma}^{(K,\pm)}$ and $\bar{M}_{v_1,\gamma}^{(K,\pm)}$ is of smaller order than $(\mu_K^{L-1-\lfloor \alpha \rfloor})^{-1}$ and multiplying with the thinning probability (A.4.43), which is of order $\mu_K^{|\gamma|-1-\lfloor \alpha \rfloor}$, this only changes the appearance rate for the $v_{|\gamma|}$-mutants by a vanishing order. $\square$

Remark 16. Note that in general there could be an overlap of two excursions of $N_{v_k,\gamma}^K$, associated to different incoming mutants. Nevertheless in the limit of $K \to \infty$ this does not happen since the time interval between the incoming mutants diverges, while the durations of the excursions stay of order one, i.e. $T_{v_k,\gamma}^{(i+1,K)} - T_{v_k,\gamma}^{(i,K)} \gg 1$.

As a direct corollary we can deduce the law of the appearance times of new mutants with trait $v \in V \backslash V_\alpha$.

Corollary A.4.3. *Suppose* $\mathbf{v}$ *and* $(\beta^K(0))_{K\geq 0}$ *are an asymptotic ESC. Let* $v \in V\backslash V_\alpha$ *be a trait such that all paths* $\gamma : \partial V_\alpha \to v$ *of shortest length* $|\gamma| = d(V_\alpha, v)$ *do only visit traits with negative invasion fitness, excluding the last trait* v*, i.e.* $f(\gamma_i, \mathbf{v}) < 0$ $\forall i = 0, \dots, |\gamma| - 1$*. Denote by* $T_v^{(i,K)}$ *the appearance time of the i-th mutant of trait* v *descended from an nearest neighbour trait. Then there exists a* $0 < c < \infty$ *such that, for each* $\varepsilon > 0$*, there exist sequences of iid. exponential random variables* $E_v^{(i,K,\pm)}$*,* $i \geq 1$ *with rates* $\tilde{R}_v^{(\pm)} = (1 \pm C\varepsilon)\tilde{R}_v$*, where*

$$\tilde{R}_v := \sum_{\substack{\gamma:\partial V_\alpha \to v \\ |\gamma| = d(V_\alpha, v)}} a_{\gamma_0} b(\gamma_0) m(\gamma_0, \gamma_1) \prod_{j=1}^{|\gamma|-1} \lambda(\rho(\gamma_j, \mathbf{v})) m(\gamma_j, \gamma_{j+1}) \tag{A.4.47}$$

Such that

$$\liminf_{K\to\infty} \mathbb{P}\left(E^{(i,K,-)} \leq K\mu_K^{d(\mathbf{v},v)}\left(T_v^{(i,K)} - T_v^{(i-1,K)}\right) \leq E^{(i,K,+)} \middle| T_v^{(i,K)} < T_{\mathit{fix}}^K\right) \geq 1 - c\varepsilon \tag{A.4.48}$$

Proof. Due to Lemma A.4.2, we can describe the arrivals of new v-type mutants approximately as sum of Poisson point processes. Since the Poisson measures $Q^{(\cdot)}_{\cdot,\cdot}$ in our representation (A.4.27) are taken as independent, the resulting mutation counting processes $M^K_{v,\gamma}$ are also independent. Hence their sum can be approximated by a Poisson process with with intensity

$$\sum_{\gamma:\partial V_\alpha \to v} \tilde{R}_{v,\gamma} K \mu_K^{|\gamma| + \lfloor \alpha \rfloor}. \tag{A.4.49}$$

Since each summand scales with the length of the respective path, the first order of the overall rate is given only by the shortest paths (i.e. γ such that $|\gamma| = d(V_\alpha, v) = d(\mathbf{v}, v) - \lfloor \alpha \rfloor$). As a result, the first order becomes (A.4.47) multiplied by $K\mu^{d(\mathbf{v},v)}$. Finally, the waiting times of homogeneous Poisson point processes are exponentially distributed with the same rate. □

A.4.3. Proof of Theorem A.2.7 and Corollary A.2.8

We have now assembled all the tools to finish the proof of Theorem A.2.7 and Corollary A.2.8.

Note that, with the notation from the proof of Lemma A.4.1, all following considerations are only valid up to the stopping time $S^K_\varepsilon \wedge \sigma^K_\varepsilon$, for sufficiently small ε. Since we have seen previously that $T^K_{\text{fix}} \leq S^K_\varepsilon \wedge \sigma^K_\varepsilon$ with probability converging to one, as $K \to \infty$, we do not condition on this anymore in the following. Moreover, constants c and C may vary throughout the proof but are always assumed to satisfy $0 < c, C < \infty$.

Both results assume that the initial conditions $(\beta^K(0))_{K\geq 0}$ compose an asymptotic ESC associated to the coexisting traits $\mathbf{v} \subset V$. In a first step, we study the time until the fixation of the first mutant trait outside of $V_\alpha := V_\alpha(\mathbf{v})$, i.e. T^K_{fix}. Corollary A.4.3 implies that, for all traits $w \in V \backslash V_\alpha$ such that all shortest paths $\gamma : \mathbf{v} \to v$ only pass through unfit traits, new mutants of this trait arise approximately according to a Poisson point process with rate $\tilde{R}_v$. By assumption, $\beta^K_w(0) = 0$, for all $K > K_0$ and $w \in V \backslash V_\alpha$, i.e. all traits outside of V_α are initially extinct. As a result, individuals of such traits w are only present due to the above incoming mutations.

We now argue why it suffices to consider traits $w \in V \backslash V_\alpha$ such that $f(w, \mathbf{v}) > 0$ and $d(\mathbf{v}, w) = L(\mathbf{v})$, i.e. the $w \in V_{\text{mut}} := V_{\text{mut}}(\mathbf{v})$, as candidates to reach $\beta^K_w = 1/\alpha$ first and trigger the stopping time T^K_{fix}.

For all w such that $\lfloor \alpha \rfloor < d(\mathbf{v}, w) < L(\mathbf{v})$, the definition of $L(\mathbf{v})$ yields $f(w, \mathbf{v}) < 0$. Therefore, the descendants of a mutant of such traits can be bounded from above by a subcritical birth-death process with rates that do not depend on K, that dies out within a finite time with probability 1. As a result,

$$\lim_{K \to \infty} \mathbb{P} \left(\sup_{t \in [0, T^K_{\text{fix}} \wedge e^{U_\varepsilon K}]} \beta^K_w(t) \geq \frac{1}{\alpha} \right) = 0. \tag{A.4.50}$$

For w such that $d(\mathbf{v}, w) = L(\mathbf{v})$ and $f(w, \mathbf{v}) < 0$, the same argument can be applied.

Finally, for all w such that $d(\mathbf{v}, w) > L(\mathbf{v})$, for all $T < \infty$, Corollary A.4.3 implies that the arrival time of the first w mutant, $T^{(1)}_w$, satisfies

$$\lim_{K \to \infty} \mathbb{P} \left(T^{(1)}_w \leq \frac{T}{K \mu_K^{L(\mathbf{v})}} \wedge T^K_{\text{fix}} \right) = 0. \tag{A.4.51}$$

Focussing on the $w \in V_{\text{mut}}$, we can use couplings to supercritical birth-death processes (similar to the arguments in the previous sections) to bound the different mutant populations. Using classical results on branching processes (e.g. from [4, Ch. III.4]) we can approximate the probability that the descendants of a single mutant of a particular trait w do not go extinct by $(1 \pm C\varepsilon) f(w, \mathbf{v})/b(w)$. Moreover, conditioned on not going extinct, the time that such a population needs to grow to a size of $K^{1/\alpha}$ can be bounded by $(1 \pm C\varepsilon) \ln K / \alpha f(w, \mathbf{v})$. It is therefore negligible on the time scale $1/K\mu_K^{L(\mathbf{v})}$, on which the w mutants arise.

Overall, we can deduce from Corollary A.4.3 that there is a constant $0 < c < \infty$ and exponential random variables $E^{(K,\pm)}_{w,\text{fix}}$ with parameters $(1 \pm c\varepsilon) \tilde{R}_w f(w, \mathbf{v})/b(w) = (1 \pm c\varepsilon) R(\mathbf{v}, w)$ such that

$$\liminf_{K \to \infty} \mathbb{P} \left(E^{(K,-)}_{w,\text{fix}} \leq K\mu_K^{L(\mathbf{v})} T^K_{\text{fix}} \leq E^{(K,+)}_{w,\text{fix}} \middle| \beta^K_w(T^K_{\text{fix}}) = \frac{1}{\alpha} \right) \geq 1 - c\varepsilon. \tag{A.4.52}$$

Since the mutants arising along different paths are independent (see the proof of Corollary A.4.3), the actual stopping time $K\mu_K^{L(\mathbf{v}))} T^K_{\text{fix}}$ (without conditioning on a trait w) is roughly exponentially distributed with the sum of all rates $R(\mathbf{v}) = \sum_{w \in V_{\text{mut}}} R(\mathbf{v}, w)$. In addition, the probability that a certain trait $w \in V_{\text{mut}}$ triggers the stopping time T^K_{fix} can be approximated by $R(\mathbf{v},w)/R(\mathbf{v})$. More precisely, there are exponential random variables $E^{(K,\pm)}(\varepsilon)$ such that

$$\liminf_{K\to\infty} \mathbb{P}\left(E^{(K,-)}(\varepsilon) \leq K\mu_K^{L(\mathbf{v})} T_{\text{fix}}^K \leq E^{(K,+)}(\varepsilon)\right) \geq 1 - c\varepsilon, \tag{A.4.53}$$

$$\frac{R(\mathbf{v},w)}{R(\mathbf{v})}(1-c\varepsilon) \leq \lim_{K\to\infty} \mathbb{P}\left(\beta_w^K(T_{\text{fix}}^K) = \frac{1}{\alpha}\right) \leq \frac{R(\mathbf{v},w)}{R(\mathbf{v})}(1+c\varepsilon). \tag{A.4.54}$$

Since ε can be picked arbitrarily small, this concludes the proof of Theorem A.2.7.

To deduce Corollary A.2.8, we note that at time T_{fix}^K the population sizes satisfy (A.2.28), for some $w \in V_{\text{mut}}(\mathbf{v})$. Hence the assumption of the corollary and Theorem A.5.2 imply that a new ESC associated to $\mathbf{v}_{\text{ESC}}(\mathbf{v},w)$ is obtained within a time of order $\ln K$. We emphasise that, although Theorem A.5.2 only implies that $\beta_u^K \to 0$ for traits $u \notin V_\alpha(\mathbf{v}_{\text{ESC}}(\mathbf{v},w))$ after this time, these subpopulations can be bounded from above by subcritical branching processes that go extinct within a time of order 1, such that the conditions of T_{ESC}^K are truly satisfied. This yields the first claim of Corollary A.2.8. Since this time is again negligible with respect to the $1/K\mu_K^{L(\mathbf{v})}$-time scale, the second claim follows directly. For the last claim, we realise that a new ESC $\mathbf{w}$ might be reached from multiple $w \in V_{\text{mut}}(\mathbf{v})$, and we therefore add up all corresponding probabilities to obtain $p(\mathbf{v},\mathbf{w})$. This concludes the proof of Corollary A.2.8.

A.4.4. Proof of Corollary A.2.10 and Theorem A.2.11

In order to derive results for the jump chain $(\mathbf{v}^{(k)})_{k\geq 0}$ on $\mathcal{G}_{\text{ESC}}$, we observe that, after a successful transition according to Corollary A.2.8, the final state of the process again satisfies the initial assumptions for another application of the corollary. We simply need to recompute the state-dependent quantities ($L(\mathbf{v}), V_{\text{mut}}(\mathbf{v})$, etc.). As a consequence, the strong Markov property allows us to use Corollary A.2.8 to construct the random sequence $(\mathbf{v}^{(k)})_{k\geq 0}$ as well as derive the asymptotics of the stopping times $T_{\text{ESC}}^{(k,K)}$ by an inductive procedure. This proves Corollary A.2.10.

To extract the limiting process on the time scale $1/K\mu_K^L$ for fixed $L > \alpha$, take an initial configuration of this stability degree, i.e. $\mathbf{v} \in \mathcal{S}^L$. Considering the jump chain $(\mathbf{v}^{(k)})_{k\geq 0}$ with $\mathbf{v}^{(0)} = \mathbf{v}$, Assumption 3 implies that, with probability one, $(\mathbf{v}^{(k)})_{k\geq 0}$ reaches an ESC of stability degree at least L within finitely many steps. We now consider such a finite path $\Gamma : \mathbf{v} \to \mathbf{w}$ in $\mathcal{G}_{\text{ESC}}$, where $L(\mathbf{w}) \geq L$. Without loss of generality we may assume that the intermediate ESCs are of strictly lower stability degree, i.e. $L(\Gamma_i) < L$ $\forall 1 \leq i < |\Gamma|$. Otherwise we could shorten the path. Asking now for the time T_Γ^K that it takes to transition from $\mathbf{v}$ to $\mathbf{w}$ along Γ, we can simply add up the single step transition times $T^{(i,K)} - T^{(i-1,K)}$. By Corollary A.2.10, we know that, on the time scale $1/K\mu_K^{L^{(i)}}$, those are well approximated by exponential random variables $E_\pm^{(i,K)}$. Since $L = L^{(1)} > L^{(i)}$, for $2 \leq i \leq |\Gamma|$, we can deduce that the rescaled transition time $T_\Gamma^K K\mu_K^L$ is dominated by the very first transition and thus well described by exponential random variables.

To compute the respective transition rates, notice that by Corollary A.2.10, on the time scale $1/K\mu_K^L$, the rate to escape from $\mathbf{v} = \Gamma_0$ is given by $R(\mathbf{v}) = R^{(1)}$. Moreover, we have to take into account that we consider the case where the limit process $(\mathbf{v}^{(i)})_{i\geq i}$ takes a particular path, i.e. $\mathbf{v}^{(i)} = \Gamma_i$, for $0 \leq i \leq |\Gamma|$. The probability of this event is simply given by the product of the one-step-probabilities $p(\mathbf{v}^{(i-1)}, \mathbf{v}^{(i)})$. Similarly to previous arguments, there might by different paths $\Gamma : \mathbf{v} \to \mathbf{w}$ and hence we add up their probabilities. This yields the rates $R^L(\mathbf{v}, \mathbf{w})$ in (A.2.41) and therefore the claimed dynamics of the jump process $(\mathbf{v}^L(t))_{t\in[0,T]}$ on the L-scale graph $\mathcal{G}^L$.

To finally deduce the limit of the rescaled population process N^K/K, we note that there is no macroscopic evolution during almost the entire waiting time for a transition on $\mathcal{G}^L$. The set of macroscopic traits $\left\{v \in V : \beta_v^K(t) > 1 - \varepsilon_K\right\}$ only changes after a new mutant fixates, which happens at time $T_{\text{fix}}^{(1,K)}$. The rest of the transition time, which may consist of many chances of the macroscopic traits, vanishes when rescaling with $K\mu_K^L$. Therefore, we obtain the limit process of Theorem A.2.11, which jumps between the Lotka-Volterra-equilibria associated to the state of $(\mathbf{v}^L(t))_{t\in[0,T]}$.

A.5. Technical results

The aim of this chapter is to collect some results on the $\mathcal{O}(1)$- and $\mathcal{O}(\ln K)$-time scale behaviour of the population process. While Section A.5.1 explains the form of $\lambda(\rho)$, Section A.5.2 justifies the notation $\mathbf{v}_{\text{ESC}}(\mathbf{v}, v)$. The statements have been derived in [29] and [48] whereto we refer for detailed proofs.

A.5.1. Excursions of subcritical birth death processes

The first lemma quantifies the mean number of birth events before a subcritical birth death process goes extinct, corresponding to $\lambda(\rho)$. Although we restate an existing result here, we provide a short proof below. This proof is different to the more general scenario that is cited in [29] and gives the reader an intuition behind the expression.

Lemma A.5.1. *([29, Lemma A.3]) Consider a subcritical linear birth death process with individual birth and death rates* $0 < b < d$. *Denote by* Z *the total number of birth events during an excursion of this process initiated with exactly one individual. Then, for* $k \in \mathbb{N}_0$,

$$p^{(b,d)}(k) := \mathbb{P}\left(Z = k\right) = \frac{(2k)!}{k!(k+1)!}\left(\frac{b}{b+d}\right)^k\left(\frac{d}{b+d}\right)^{k+1} \tag{A.5.1}$$

and in particular

$$e^{(b,d)} := \mathbb{E}\left[Z\right] = \sum_{k=1}^{\infty}\frac{(2k)!}{(k-1)!(k+1)!}\left(\frac{b}{b+d}\right)^k\left(\frac{d}{b+d}\right)^{k+1}. \tag{A.5.2}$$

Moreover, we have the following continuity result. There exist two positive constants $c, \varepsilon_0 > 0$, *such that, for all* $0 < \varepsilon < \varepsilon_0$ *and* $0 < b_i < d_i$, *if* $|b_1 - b_2| < \varepsilon$ *and* $|d_1 - d_2| < \varepsilon$, *then*

$$\left|e^{(b_1,d_1)} - e^{(b_2,d_2)}\right| < c\varepsilon. \tag{A.5.3}$$

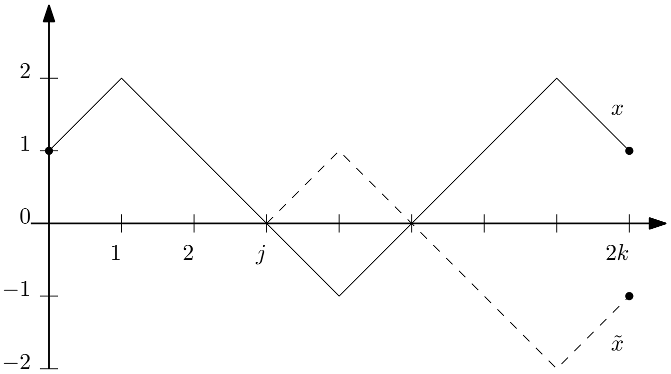


Figure A.13.: Original path x that prematurely goes extinct and its reflection $\tilde{x}$.

Remark 17. Note that (C.5.16) corresponds to (A.2.19) via $e^{(b,d)} = \lambda(\rho)$, where $\rho = b/(b+d)$.

Proof. Although the considered process takes place in continuous time, it suffices to focus on the birth and death events as jump chain in discrete time. This is nothing but a simple random walk on $\mathbb{N}_0$ with probabilities

$$p(x, x+1) = \frac{b}{b+d}, \quad p(x, x-1) = \frac{d}{b+d} \quad \forall x \geq 1 \tag{A.5.4}$$

and absorbing state 0. From this point of view it is only a question of counting the number of paths leading from one individual to extinction consisting of exactly k births and hence $k+1$ death events. As final step there has to happen a death since the population does not vanish before. So the first $2k$ events form a walk from 1 to 1. There are $\binom{2k}{k}$ of such paths but some of them would lead to extinction earlier. To determine their number we apply a reflection principle in the following way. Let $x = (x_0, x_1, \ldots, x_{2k})$ be a path leading from one to one such that there exists a $0 < j < 2k$ with $x_j = 0$. Then we define the partially reflected path $\tilde{x}$ by

$$\tilde{x}_i := \begin{cases} x_i & \text{for } i \leq j, \\ -x_i & \text{for } i > j. \end{cases} \tag{A.5.5}$$

This gives us a unique path from $\tilde{x}_0 = 1$ to $\tilde{x}_{2k} = -1$ (cf. A.13). Moreover, there is a one to one correspondence between prematurely extincting processes and paths leading from 1 to -1. The latter ones consist of only $k-1$ births and hence there are $\binom{2k}{k-1}$ different ones. Finally the total number of legal paths is

$$\#\left\{x = (x_0, x_1, \ldots, x_{2k}) | x_0 = 1, x_{2k} = 1, x_i > 0\right\} = \binom{2k}{k} - \binom{2k}{k-1} = \frac{(2k)!}{k!(k+1)!}. \tag{A.5.6}$$

We now achieve (A.5.1) by multiplying with the probability of k births and $k+1$ death events. The last statement is a simple consequence of the mean value theorem. ☐

A.5.2. Fast evolution until ESC

In this subsection we discuss the first phase of evolution, where an ESC is obtained on the $\ln K$-time scale. The convergence of $N^K(t\ln K)/K$ and $\beta^K(t\ln K)$, as $K\to\infty$, is studied in [48]. In the following we cite the respective results in the notation of this paper.

For a finite graph $\mathcal{G}=(V,E)$ and under Assumptions 1 and 2, the trajectories $(\beta_w(t), w\in V)$ (which turn out to be the limit of $(\beta_w^K(t\ln K), w\in V)$) are defined by an inductive procedure. The construction is valid until a stopping time T_0.

Denote by $\tilde{\mathbf{v}}^{(\ell)}$, $\ell\geq 0$, the sequence of consecutive coexisting resident traits. We emphasize that these are not to be confused with the sequence of resident traits $\mathbf{v}^{(k)}$, $k\geq 0$, that are associated to ESCs. The invasion times, at which the sets of resident traits change due to upcoming mutant traits, are denoted by the increasing sequence $(s_\ell)_{\ell\geq 0}$.

For initial conditions $\tilde{\beta}(0)$, the support of the unique asymptotically stable equilibrium of the Lotka-Volterra system (A.2.2) associated to the traits $\{w\in V:\tilde{\beta}_w(0)=1\}$ (if existent) is denoted by $\tilde{\mathbf{v}}^{(0)}$. The equilibrium $\bar{n}(\tilde{\mathbf{v}}^{(0)})$ is reached within a time of order 1 and we set $s_0:=0$. Moreover, we define $\beta_w(0):=\max_{u\in V}[\tilde{\beta}_u(0)-d(u,w)/\alpha]_+$ as the initial condition of the limiting trajectories. This reflects that, within a time of order 1, living traits produce neighbouring mutant populations with the size of a μ_K-fraction of their own size. This time of order 1 is negligible on the $\ln K$-time scale, which the limit β is defined on.

Assuming that $s_{\ell-1}$, $\tilde{\mathbf{v}}^{(\ell-1)}$ such that $LVE_+(\tilde{\mathbf{v}}^{(\ell-1)})=\bar{n}(\tilde{\mathbf{v}}^{(\ell-1)})$, and $\beta(s_{\ell-1})$ are known, the next phase can be described as follows. The ℓ^{th} invasion time is set to

$$s_\ell:=\inf\{t>s_{\ell-1}:\exists\ w\notin\tilde{\mathbf{v}}^{(\ell-1)}:\beta_w(t)=1\}. \tag{A.5.7}$$

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

$$\beta_w(t):=\max_{u\in V}\left[\beta_u(s_{\ell-1})+(t-t_{u,\ell}\wedge t)f(u,\tilde{\mathbf{v}}^{(\ell-1)})-\frac{d(u,w)}{\alpha}\right]\vee 0, \tag{A.5.8}$$

where, for any $w\in V$,

$$t_{w,\ell}:=\begin{cases}\inf\left\{t\geq s_{\ell-1}:\exists\ u\in V:d(u,w)=1,\beta_u(t)=\frac{1}{\alpha}\right\} & \text{if } \beta_w(s_{\ell-1})=0\\ s_{\ell-1} & \text{else}\end{cases} \tag{A.5.9}$$

is the first time in $[s_{\ell-1},s_\ell]$ when this trait arises. If we define $V_{\text{living}}(t):=\{w\in V:\beta_w(t)>0\}$ equivalently to V_{living}^K (on the $\ln K$-time scale), then this implies $\beta_w(t_{w,\ell})\geq 0$ and $\beta_w(t_{w,\ell}+\delta)>0$, for small $\delta>0$.

The stopping time T_0, that terminates the inductive construction of the limiting trajectories, is set to s_ℓ if

(a) there is more than one $w\in V\backslash\tilde{\mathbf{v}}^{(\ell-1)}$ such that $\beta_w(s_\ell)=1$;

(b) the mutation-free Lotka-Volterra system associated to $\tilde{\mathbf{v}}^{(\ell-1)}$ and the unique $w\in V\backslash\tilde{\mathbf{v}}^{(\ell-1)}$ such that $\beta_w(s_\ell)=1$ does not have a unique globally attractive stable equilibrium (in particular, if such an equilibrium does not exist for $\{w\in V:\tilde{\beta}_w(0)=1\}$, T_0 is set to 0);

(c) there exists $w \in V \backslash \tilde{\mathbf{v}}^{(\ell-1)}$ such that $\beta_w(s_\ell) = 0$ and $\beta_w(s_\ell - \delta) > 0$ for all $\delta > 0$ small enough.

(d) there exists $w \in V \backslash \tilde{\mathbf{v}}^{(\ell-1)}$ such that $s_\ell = t_{w,\ell}$.

These conditions are mostly technical and are discussed in [48].

With this construction, the results can be stated as follows:

Theorem A.5.2. *([48, Theorem 2.7]) Let $\mathcal{G} = (V, E)$ be a finite graph. Suppose that Assumption 1 and 2 hold and consider the model defined by* (A.2.1) *with $\mu_K = K^{-1/\alpha}$. Let $\tilde{\mathbf{v}}_0 \subset V$ and assume that, for every $w \in V$,*

$$\beta_w^K(0) \to \tilde{\beta}_w(0), \quad (K \to \infty) \quad \textit{in probability.} \tag{A.5.10}$$

Then, for all $T > 0$, as $K \to \infty$, the sequence $((\beta_w^K(t \ln K), w \in V), t \in [0, T \wedge T_0])$ converges in probability in $\mathbb{D}([0, T \wedge T_0], \mathbb{R}_+^V)$ to the deterministic, piecewise affine, continuous function $((\beta_w(t), w \in V), t \in [0, T \wedge T_0])$, which is defined in (A.5.8).

Theorem A.5.3. *([48, Proposition 2.8]) Under the same assumptions as in Theorem A.5.2, for all $T > 0$, as $K \to \infty$, the sequence $((N_w^K(t \ln K)/K, w \in V), t \in [0, T \wedge T_0])$ converges in the sense of the finite dimensional distributions to a deterministic jump process $((N_w(t), w \in V), t \in [0, T \wedge T_0])$, which jumps between different Lotka-Volterra equilibria according to*

$$N_w(t) := \sum_{\ell \in \mathbb{N}: s_{\ell+1} \leq T_0} \mathbf{1}_{s_\ell \leq t < s_{\ell+1}} \mathbf{1}_{w \in \tilde{\mathbf{v}}^{(\ell)}} \bar{n}_w(\tilde{\mathbf{v}}^{(\ell)}). \tag{A.5.11}$$

Moreover, the invasion times s_ℓ and the times $t_{w,\ell}$ when new mutants arise are calculated precisely in [48]. This is however not relevant to the discussion in this paper.

We notice that the constructed trajectories $(\beta_w(t), w \in V)$ stay constant precisely once an ESC is obtained. In this case, there is no more visible evolution on the $\ln K$-time scale.

B. Publication: Effective growth rates in a periodically changing environment: From mutation to invasion

This appendix is to appear in Stochastic Processes and their Applications as joint work with Anna Kraut [71],

> M. Esser and A. Kraut, *Effective growth rates in a periodically changing environment: From mutation to invasion.* Stochastic Processes and their Applications, Volume 184 , Number 104598, 2025. `https://doi.org/10.1016/j.spa.2025.104598`

We consider a stochastic individual-based model of adaptive dynamics for an asexually reproducing population with mutation, with linear birth and death rates, as well as a density-dependent competition. To depict repeating changes of the environment, all of these parameters vary over time as piecewise constant and periodic functions, on an intermediate time-scale between those of stabilization of the resident population (fast) and exponential growth of mutants (slow). Studying the growth of emergent mutants and their invasion of the resident population in the limit of small mutation rates for a simultaneously diverging population size, we are able to determine their effective growth rates. We describe this growth as a mesoscopic scaling-limit of the orders of population sizes, where we observe an averaging effect of the invasion fitness. Moreover, we prove a limit result for the sequence of consecutive macroscopic resident traits that is similar to the so-called trait-substitution-sequence.

B.1. Introduction

Mathematical approaches to understanding the long-term evolution of populations have a long history and can even be traced back to ideas of Malthus in 1798 [135]. The study of heterogeneous populations is of particular interest as it allows to analyse the diversity and the interaction of species as they adapt over time. The driving mechanisms, ecology and evolution, which are addressed by models of adaptive dynamics, may strongly depend on the environment a population is living in. Since realistically this environment cannot be assumed to stay constant over time, we study the effects of periodic changes of the model parameters in this paper. From an application point of view, this is for example motivated by the seasonal changes of temperature, humidity, accessibility of nutrition and other resources, which may effect the fertility of individuals and thus directly have an impact on the population's growth [78, 129]. Another interesting example is pulsed drug-based therapy for infectious diseases or cancer. Dependent on the treatment protocol, the concentrations of drugs may vary over time. Assuming a regular supply of drug, this can be described by periodic changes, leading to varying reproduction rates of the pathogen.

While in both of these cases the population's dynamics are directly affected by the environmental changes on a short time-scale, it is reasonable to expect some averaging and thus a macroscopic trend of growth or shrinking on a larger time-scale. This averaging effect is what we study rigorously in the present paper. Starting from a model that describes the individuals' dynamics on a microscopic level, i.e. taking into account interactions between single individuals, we derive results for the effective mesoscopic growth rates of subpopulations of intermediate size, i.e. that consist of a larger number of individuals but are still negligible with respect to the total population size. Moreover, we give an explicit description of the macroscopic evolution of the whole population process, tracing the evolution of the dominating trait, in the large population limit. A crucial aspect in this is to understand under which conditions we can observe the emergence and growth of new types or even the replacement of resident traits by fitter mutants.

We consider a variation of the stochastic individual-based model of adaptive dynamics that has been introduced by Fournier and Méléard [83] and since then was studied for a broad spectrum of scaling limits and model extensions (see e.g. [27] for an overview and [6, 158, 148, 28, 22]). Its aim is to study the interplay of ecology and evolution, i.e. both the short-term effects of competitive interactions of different subpopulations and the long-term effects of occurrence and fixation of new mutant species. Since our interest lies in analysing the effects of time-dependent changes of ecological parameters on the long-term evolution of a population, these models of adaptive dynamics are naturally helpful.

As one of the first results on the individual-based model, Champagnat was able to show that certain assumptions on the scaling of large populations and very rare mutations lead to a separation of the time-scales of ecology and evolution, which is a fundamental principle of adaptive dynamics. Under the aforementioned assumptions, Champagnat derived convergence to the trait-substitution-sequence (TSS) [40] and, together with Méléard, the polymorphic-evolution-sequence (PES) [43]. On an accelerated time-scale, these sequences describe how the macroscopic population essentially jumps between (monomorphic or polymorphic) equilibria of different Lotka-Volterra systems. A broader spectrum of more frequent mutations was investigated by Bovier, Coquille, and Smadi for a simple trait space with a valley in the fitness landscape [29]. This work laid the basis for the more general study of moderately rare mutations in [48], under collaboration of Kraut. The latter provides both the description of a macroscopic limit process, which consists of (deterministic) jumps between equilibrium states, as well as a mesoscopic limit result for the growth and decline of all subpopulations, observable on a logarithmic time-scale.

Despite the variety of different scenarios that have already been analysed, all of these previous works ask for the parameters of the population process to be constant over time. In the present paper, we break with this assumption and allow for periodic parameter changes. As before, we study the limit of a diverging carrying capacity $K \nearrow \infty$, which scales the order of the total population's size, and choose moderate mutation probabilities $\mu_K \searrow 0$. In addition, we introduce a finite number of parameter constellations, which repeat periodically on time intervals (phases) with fixed length of order λ_K, to model a changing environment.

These parameter constellations vary the individual birth, death, and competition rates, which in particular determine the fitness, or growth rate, of the different subpopulations. Consequently, both the sign of the fitness, resulting in growth or shrinkage of the subpopulations, and the fitness relations between different types may change between phases.

We choose an intermediate time-scale of $1 \ll \lambda_K \ll \ln K$ for these environmental changes. As a result, the environment stays stable enough for the macroscopic resident population to adapt to it in between parameter changes, but changes occur fast enough to influence the growth of the micro- and mesoscopic subpopulations in between invasions. Under these assumptions, we can observe an averaging effect on the level of mutant growth. Similarly to [48], we prove a mesoscopic convergence result for the orders of population sizes K^β of all subpopulations. In the limit of $K \nearrow \infty$, on the time-scale $\ln K$ of exponential growth, these exponents β converge to deterministic, piecewise affine functions $\bar{\beta}$ that can be described by a recursive algorithm. The slopes of these functions are determined by the effective (time-average) fitness of the subpopulations. Based on this mesoscopic characterisation, we further derive a substitution-sequence on the same time-scale, describing the macroscopic jumps of the population between successive resident traits.

The fact that the environmental parameters now change on an intermediate time-scale at a first glance seems to be only a small variation of the former models. However, a couple of non-trivial difficulties arise in all parts of the established proof strategies: First of all, since time spans of order $\ln K$ consist of asymptotically infinitely many parameter phases that need to be concatenated, the way in which a large deviation principle is usually applied for these types of processes to ensure stability of the resident population in between invasions (see e.g. [40]) is not sufficient. To obtain a quantification of the speed at which the probability of exit from a domain within a λ_K-time span tends to zero, we instead use potential theoretic arguments similar to Baar, Bovier, and Champagnat [7]. Moreover, to take care of the short $O(1)$ times of re-equilibration after a parameter change, we study the speed of convergence in the standard convergence result of Ethier and Kurtz [75].

Second, we need to extend the general growth results for branching processes (with immigration) of Champagnat, Méléard and Tran [45] to periodically changing parameters. This in particular requires a careful consideration of small populations, i.e. in the case of extinction or a newly emerging mutant population. Here we study the distribution function of the extinction time, making use of estimates on the probability generating function of Galton-Watson processes.

Finally, we need to carefully consider the event of a mutant population becoming macroscopic. Here we need to choose the stopping time, after which we start the comparison to the deterministic system, such that the invading mutant is guaranteed to be in a phase with positive invasion fitness and successfully fixate as the new resident trait within a time of order 1.

Changing environmental parameters have been previously introduced to a number of mathematical population models. While we cannot give an extensive review here, we want to mention a few examples. One popular scenario is that of a shifting optimal trait (mostly in deterministic ODE or PDE models), where the fitness of all other traits depends only on the distance to the current optimum [107, 155, 87]. A common observation is the importance of the relation between the time-scales of environmental shifts and trait changes (mutation speed and step size), which determines whether the population can successfully adapt or not. A scenario similar to the one of the present paper is that of periodically changing environments, in both deterministic and stochastic models [145, 160, 37]. Here, previous studies have focussed on the dynamics of a fixed (usually small) number of competing traits without mutation. As we observe in this paper, time-scales again play a crucial role, where all of the above works find that sufficiently fast fluctuations lead to the population evolving according to time-averaged effective parameters. Other questions have been addressed by various authors, for example the dynamics of phenotypic switching and dormancy for non-competitive multi-type systems in more or less randomly fluctuating environment [125, 134, 64, 111, 19]. To the best of our knowledge, the dynamics of stochastic models with periodically changing environment, general fitness landscapes, and newly emerging mutant types are still an open problem.

Similar to some of these approaches it will be interesting to extend the adaptive dynamics model of this paper to more generally changing parameters. Modelling the environmental parameters as continuous functions or as a stochastic process itself, where jump times and transitions are random, can allow for a more realistic depiction of biological scenarios that either only change gradually or less regularly. Our results in this manuscript are meant as a first step to establish techniques of how to study this new class of models and lay the basis for future research that is already in progress.

The remainder of this paper is structured as follows. In Section B.2.1, the individual-based model for a population in a time-dependent environment is introduced rigorously. We point out some key quantities, like equilibrium states and invasion fitness, in Section B.2.2. Finally, we describe the behaviour of the limit process in terms of an inductive algorithm in Section B.2.3 and state our main convergence results. Chapter B.3 provides a discussion of the general heuristics and the necessity of some of our assumptions. Moreover, we give an outlook on possible extensions of this approach. The proofs of the main results of this paper can be found in Chapter B.4. The technical results on birth death processes, which lay the basis for these proofs, are discussed in Appendices B.5, B.6, and B.7.

B.2. Model and Main Results

B.2.1. Individual-based model in a time-dependent environment

We consider a population that is composed of a finite number of asexual reproducing individuals. Each of them is characterized by a genotypic trait, taken from a finite trait space that is given by a (possibly directed) graph $G = (V, E)$. Here, the set of vertices V represents the possible traits that individuals can carry. The set of edges E marks the possibility of mutation between traits. We start out with a microscopic, individual-based model with logistic growth.

To extend the basic model to one with a periodically changing environment, we consider a finite number $\ell \in \mathbb{N}$ of phases. For each phase $i = 1, \cdots, \ell$ and all traits $v, w \in V$, we introduce the following biological parameters:

- $b_v^i \in \mathbb{R}_+$, the *birth rate* of an individual of trait v during phase i,
- $d_v^i \in \mathbb{R}_+$, the *(natural) death rate* of an individual of trait v during phase i,
- $c_{v,w}^i \in \mathbb{R}_+$, the *competition* imposed by an individual of trait w onto an individual of trait v during phase i,
- $K \in \mathbb{N}$, the *carrying capacity* that scales the environment's capacity to support life,
- $\mu_K \in [0,1]$, the *probability of mutation* at a birth event (phase-independent),
- $m_{v,\cdot} \in \mathcal{M}_p(V)$, the *law of the trait of a mutant* offspring produced by an individual of trait v (phase-independent).

To ensure logistic growth and ensure the possibility of mutation only along the edges of our trait graph, we make the following assumptions on our parameters.

Assumption 4. (a) For every $v \in V$ and $i = 1, \cdots \ell$, $c_{v,v}^i > 0$.

(b) $m_{v,v} = 0$, for all $v \in V$, and $m_{v,w} > 0$ if and only if $(v,w) \in E$.

Rescaling the competition by K (cf. (B.2.3) below) leads to a total population size of order K. We adapt the following terminology: As $K \to \infty$, subpopulations of certain traits are referred to as

- *microscopic* if they are of order 1,
- *macroscopic* if they are of order K,
- *mesoscopic* if they are of order strictly between 1 and K.

For a new mutant, reaching a macroscopic population size through exponential growth takes a time of order $\ln K$. For a resident population, it takes a time of order 1 to reach a small neighbourhood of its new equilibrium after an environmental change. In order for environmental changes to happen slow enough such that the resident populations can adapt, but fast enough such that they influence the growth of mutants, we choose

$$1 \ll \lambda_K \ll \ln K \tag{B.2.1}$$

as an intermediate time-scale for the length of the ℓ phases. For each $i = 1, \dots, \ell$, we assume that the i-th phase has length $T_i \lambda_K$, where $T_i > 0$. To refer to the endpoints of the phases, we define $T_j^\Sigma := \sum_{i=1}^j T_i$.

Building on this, we define the time-dependent birth-, death-, and competition rates as the periodic extension of

$$b_v^K(t) = \sum_{i=1}^{\ell} \mathbb{1}_{t \in [T_{i-1}^\Sigma \lambda_K, T_i^\Sigma \lambda_K)} b_v^i, \tag{B.2.2}$$

and analogously for $d_v^K(t)$ and $c_{v,w}^K(t)$. Note that b_v^i and b_v^K are very similar in notation. To make the distinction clear, we always use the upper index i to refer to the constant parameter in phase i and the index K to refer to the time-dependent parameter function for carrying capacity K, and use the same convention in comparable cases.

For any K, the evolution of the population over time is described by a Markov process N^K with values in $\mathbb{D}(\mathbb{R}_+, \mathbb{N}_0^V)$. $N_v^K(t)$ denotes the number of individuals of trait $v \in V$ that are alive at time $t \geq 0$. The process is characterised by its infinitesimal generator

$$\begin{aligned}\left(\mathcal{L}_t^K \phi\right)(N) = &\sum_{v \in V} (\phi(N + e_v) - \phi(N)) \left(N_v b_v^K(t)(1 - \mu_K) + \sum_{w \in V} N_w b_w^K(t) \mu_K m_{w,v} \right) \\ &+ \sum_{v \in V} (\phi(N - e_v) - \phi(N)) N_v \left(d_v^K(t) + \sum_{w \in V} \frac{c_{v,w}^K(t)}{K} N_w \right), \end{aligned} \tag{B.2.3}$$

where $\phi : \mathbb{N}_0^V \to \mathbb{R}$ is measurable and bounded and e_v denotes the unit vector at $v \in V$. The process can be constructed algorithmically following a Gillespie algorithm [90]. Alternatively, the process can be represented via Poisson measures (see [83]), a representation that is used in the proofs of this paper.

B.2.2. Important quantities

In this paper we study the typical behaviour of the processes $(N^K, K \in \mathbb{N})$ for large populations, i.e. as $K \to \infty$. A classical law of large numbers [75] states that the rescaled processes N^K/K converge on finite time intervals to the solution of a system of Lotka-Volterra equations. The study of these equations is central to determine the short term evolution of the processes N^K.

Definition B.2.1 (Lotka-Volterra system)**.** For a phase $i \in \{1, \dots \ell\}$ and a subset $\mathbf{v} \subseteq V$, we refer to the following differential equations as the corresponding *Lotka-Volterra system*:

$$\dot{n}_v(t) = \left(b_v^i - d_v^i - \sum_{w \in \mathbf{v}} c_{v,w}^i n_w(t) \right) n_v(t), \quad v \in \mathbf{v},\ t \geq 0 \tag{B.2.4}$$

In this work, we focus on the case of a sequence of monomorphic resident traits, meaning that, apart from the invasion phases, only one single (fit) subpopulation is of macroscopic size and fluctuates around its equilibrium size. This monomorphism is ensured by a termination criterion in the construction of the limiting process for our main Theorem B.2.4 (criterion (d), see also Remark 19). Taking into account the phase-dependent parameters, we denote these *monomorphic equilibria* by $\bar{n}_v^i := (b_v^i - d_v^i)/c_{v,v}^i$.

Talking about evolution, the most important quantity is fitness. For the individual-based model of adaptive dynamics, the notion of invasion fitness has been shown to be useful. It describes the approximate growth rate of a small population of trait w in a bulk population of trait v in the mutation-free system. To adapt it to the present setting we have to include the phase-dependence.

Definition B.2.2 (Invasion fitness)**.** For each phase $i \in \{1, \cdots, \ell\}$ and for all traits $v, w \in V$ such that the equilibrium size of $\bar{n}_v^i$ is positive, we denote by

$$f_{w,v}^i := b_w^i - d_w^i - c_{w,v}^i \bar{n}_v^i \tag{B.2.5}$$

the *invasion fitness* of trait w with respect to the monomorphic resident v in the i-th phase. Moreover, we define the time-dependent fitness and the average fitness by the periodic extension of

$$f_{w,v}^K(t) := \sum_{i=1}^{\ell} \mathbb{1}_{t \in [T_{i-1}^\Sigma \lambda_K, T_i^\Sigma \lambda_K)} f_{w,v}^i \quad \text{and} \quad f_{w,v}^{\text{av}} := \frac{\sum_{i=1}^{\ell} T_i f_{w,v}^i}{T_\ell^\Sigma}. \tag{B.2.6}$$

Let us now consider multi-step mutations arising along paths within the trait graph $G = (V, E)$. We introduce the graph distance between two vertices $v, w \in V$ as the length of the shortest (directed) connecting path

$$d(v, w) := \min_{\gamma : v \to w} |\gamma|, \tag{B.2.7}$$

where we use the convention that the minimum over an empty set is ∞. Note that $d(v, w)$ is not a distance in the classical sense, as it may not be symmetric in the case of a directed graph.

Since a single birth event causes a mutation with probability μ_K, a macroscopic trait v (size of order K) produces subpopulations of a size of order $K\mu_K$ of its neighbouring traits. These traits themselves produce subpopulations of a size of order $K\mu_K^2$ of second order neighbours of v. In general, v induces mutant populations of trait w of size of order $K\mu_K^{d(v,w)}$. We study mutation probabilities of the form

$$\mu_K = K^{-1/\alpha}, \qquad \alpha \in \mathbb{R}_{>0} \setminus \mathbb{N}. \tag{B.2.8}$$

As a consequence all traits at a distance of at most $\lfloor \alpha \rfloor$ have a size that is non-vanishing for increasing K, which means that they can survive. For technical reasons, we exclude $\alpha \in \mathbb{N}$, see the discussion in Section B.3.

B.2.3. Results

The main result of this paper gives a precise description of the orders of the different subpopulation sizes as K tends to infinity. It is convenient to describe the population size of a certain trait $v \in V$ at time t by its K-exponent, which is given by the following definition.

Definition B.2.3 (Order of the population size)**.** For all $v \in V$ and all $t \geq 0$, we set

$$\beta_v^K(s) := \frac{\ln(N_v^K(s \ln K) + 1)}{\ln K}. \tag{B.2.9}$$

Note that adding one inside the logarithm is only done to ensure that $\beta_v^K(s) = 0$ is equivalent to $N_v^K(s \ln K) = 0$. Before we state the result below, let us describe the limiting functions $(\bar{\beta}_v, v \in V)$. We can define these trajectories up to a stopping time T_0 by the following inductive procedure:

Let $v_0 \in V$ be the initial macroscopic trait. For simplicity, we assume that the initial orders of population sizes converge in probability to $\bar{\beta}_w(0)$ satisfying the constraints

$$\bar{\beta}_w(0) = \max_{u \in V} \left[\bar{\beta}_u(0) - \frac{d(u,w)}{\alpha} \right] \vee 0, \tag{B.2.10}$$

$$\bar{\beta}_w(0) = 1 \;\Leftrightarrow\; w = v_0. \tag{B.2.11}$$

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

$$s_k := \inf\{t > s_{k-1} : \exists\, w \in V \backslash v_{k-1} : \bar{\beta}_w(t) = 1\}. \tag{B.2.12}$$

Moreover, we set v_k to be the trait $w \in V \backslash v_{k-1}$ that satisfies $\bar{\beta}_w(s_k) = 1$, which we assume to be unique in order to proceed (cf. termination criteria below).

For $s_{k-1} \leq t \leq s_k$, for any $w \in V$, $\bar{\beta}_w(t)$ is defined by

$$\bar{\beta}_w(t) := \max_{u \in V} \left[\bar{\beta}_u(s_{k-1}) + (t - t_{u,k})_+ f^{\mathrm{av}}_{u,v_{k-1}} - \frac{d(u,w)}{\alpha} \right] \vee 0, \tag{B.2.13}$$

where, for any $w \in V$,

$$t_{w,k} := \begin{cases} \inf\{t \geq s_{k-1} : \exists\, u \in V : d(u,w) = 1, \bar{\beta}_u(t) = \frac{1}{\alpha}\} & \text{if } \bar{\beta}_w(s_{k-1}) = 0 \\ s_{k-1} & \text{else} \end{cases} \tag{B.2.14}$$

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

Remark 18. The formula in (B.2.13) can heuristically be explained as follows: From time s_{k-1} on, on the $\ln K$-time-scale, every living trait u (i.e. traits such that $t \geq t_{u,k}$) grows/shrinks at least at the rate of its own fitness $f^{\mathrm{av}}_{u,v_{k-1}}$, which would yield $\bar{\beta}_u(t) \approx \bar{\beta}_u(s_{k-1}) + (t - s_{k-1}) f^{\mathrm{av}}_{u,v_k}$. On top of this, every living trait spreads a $\mu_K = K^{-1/\alpha}$ portion of its population size to its neighbouring traits through mutation. These then pass on a μ_K^2 portion to the second order neighbours and so on. Overall, a trait $w \in V$ receives a $\mu_K^{d(u,w)} = K^{-d(u,w)/\alpha}$ portion of incoming mutants from all living traits u, and its actual population size can hence be determined by taking the leading order term, i.e. the maximum of all these exponents $\bar{\beta}_u(s_{k-1}) + (t - s_{k-1}) f^{\mathrm{av}}_{u,v_k} - d(u,w)/\alpha$.

The stopping time T_0, that terminates the inductive construction of the limiting trajectories, is set to s_k if

(a) there is more than one $w \in V \backslash v_{k-1}$ such that $\bar{\beta}_w(s_k) = 1$;

(b) there exists $w \in V \backslash v_{k-1}$ such that $\bar{\beta}_w(s_k) = 0$ and $\bar{\beta}_w(s_k - \varepsilon) > 0$ for all $\varepsilon > 0$ small enough;

(c) there exists $w \in V \backslash v_{k-1}$ such that $s_k = t_{w,k}$;

(d) there exists an $i \in \{1, \ldots, \ell\}$ such that either $f^i_{v_{k-1},v_k} \geq 0$ or $f^i_{v_k,v_{k-1}} = 0$;

(e) there exists an $i \in \{1, \ldots, \ell\}$ such that $b^i_{v_k} - d^i_{v_k} \leq 0$.

Remark 19. Note that conditions (a)–(c) are purely technical (cf. [48]). The first part of condition (d) is a sufficient criterion to ensure the principle of *invasion implies fixation*, i.e. any mutant trait that reaches a macroscopic populations size replaces the former resident trait and there is no coexistence. The criterion is not necessary and there are other possible scenarios where the invading mutant replaces the resident population (see discussion in Chapter B.3). The second part is again technical and ensures that the εK-threshold (needed for the approximations by birth death processes) is reached at a time when invasion can take place in finite time, i.e. the comparision to the deterministic Lotka-Volterra system is possible (cf. the classical result in [75]) The last condition (e) ensures that the new resident possesses a strictly positive monomorphic equilibrium $\bar{n}^i_{v_k}$ in all phases.

Theorem B.2.4 (Convergence of β)**.** *Let a finite graph $G = (V, E)$ and $\alpha \in \mathbb{R}_{>0} \backslash \mathbb{N}$ be given and consider the model defined by* (B.2.3)*. Let $v_0 \in V$ and assume that, for every $w \in V$, $\beta^K_w(0) \to \bar{\beta}_w(0)$ in probability, as $K \to \infty$, where the limits satisfy* (B.2.10) *and* (B.2.11)*. Then, for all fixed $0 \leq T \leq T_0$, the following convergence holds in probability, with respect to the $L^\infty([0, T], \mathbb{R}^V_{\geq 0})$ norm*

$$(\beta^K_w(t), w \in V)_{t \in [0,T]} \overset{K \to \infty}{\longrightarrow} (\bar{\beta}_w(t), w \in V)_{t \in [0,T]}, \tag{B.2.15}$$

where $\bar{\beta}_w$ are the deterministic, piecewise affine, continuous functions defined in (B.2.13)*.*

Remark 20. Note that we only assume (B.2.10) to ensure convergence at $t = 0$. If the $\beta^K_w(0)$ converge to initial conditions $\hat{\beta}_w(0)$ that do not satisfy this constraint, the orders of the population sizes stabilize in a time of order 1 at $\bar{\beta}_w(0) := \max_{u \in V} \left[\hat{\beta}_u(0) - \frac{d(u,w)}{\alpha}\right] \vee 0$. These new orders satisfy (B.2.10). Because the β^K describe the population on a $\ln K$-time-scale, this means that we still get convergence on the half-open interval $(0, T]$.

Building upon this detailed description of growth of all living traits, it is natural to ask for the “visible” evolution of the population process, i.e. the progression of macroscopic traits that dominate the whole system.

Corollary B.2.5 (Sequence of resident traits)**.** *Let*

$$\nu^K_\varepsilon(t) := \sum_{w \in V : N^K_w(t) > \varepsilon K} \delta_w \quad \textit{and} \quad \nu(s) := \sum_{k \in \mathbb{N}_0} \mathbb{1}_{s_k \leq s < s_{k+1}} \delta_{v_k} \tag{B.2.16}$$

denote point measures having support on the macroscopic traits. Then, under the assumptions of Theorem B.2.4, there exists an $\varepsilon_0 > 0$ such that, for all $0 < \varepsilon < \varepsilon_0$ and all $p \in [1, \infty)$, the following convergence holds in probability, with respect to the $L^p([0, T], \mathcal{M}(V))$ norm

$$\left(\nu^K_\varepsilon(s \ln K)\right)_{s \in [0,T]} \overset{K \to \infty}{\longrightarrow} (\nu(s))_{s \in [0,T]}, \tag{B.2.17}$$

where $\mathcal{M}(V)$ denotes the set of finite, non-negative point measures on V equipped with the weak topology.

B.3. Heuristics and discussion

In this chapter, we give a heuristic idea of the proof strategy for the main Theorem B.2.4 and discuss the necessity of some of the assumptions that are made. Moreover, we present some specific examples for fitness landscapes that do not satisfy these assumptions, which can still be treated with similar techniques.

B.3.1. Heuristics of the proof of Theorem B.2.4

As it is usually the case for adaptive dynamics models, the analysis of the limiting dynamics is split into approximations for the resident and the mutant populations.

First, in Section B.4.1, we prove that - as long as the mutant populations stay below a certain small εK-threshold - the resident population also only deviates from its equilibrium state by an amount of order εK. In previous papers, this is often done by applying large deviation results that guarantee for the stochastic process to stay close to an attractive equilibrium for an exponential time in K. In our case, to bound the probability of failure (i.e. deviating too far from the equilibrium), we need to concatenate these results for an order of $\ln K/\lambda_K$ many λ_K-phases that are necessary to observe mutant growth on the $\ln K$-time-scale. By conditioning on not deviating too much during the previous phases, we can write the overall probability of failure as the sum of the probabilities to deviate during specific phases. We hence need the latter probabilities to converge to 0 faster than $\lambda_K/\ln K$. In previous works, this probability of exit from a domain was bounded through a large deviation principle that ensures a vanishing probability of deviating within an exponential time as $K \to \infty$, but does not specify the exact speed of convergence (see e.g. [40]). In the present paper, we instead apply a potential theoretic approach similar to [7] to study the embedded discrete time Markov chain and bound the probability of deviation during a λ_K-phase in $o(\lambda_K/\ln K)$. We combine this with a revised version of the standard convergence result to the deterministic system of [75, Ch. 11.2] to address the short time spans of order 1 at the beginning of each phase, where the resident population attains its new equilibrium. We prove convergence in probability instead of almost surely but can again quantify the convergence speed and bound the probability of failure in $o(\lambda_K/\ln K)$ in return. Overall, concatenating these two results, which are derived in Appendix B.5, for $O(\ln K/\lambda_K)$ many phases yields a vanishing probability for the resident population to stray from its respective equilibria.

With these bounds on the resident population, in Section B.4.2 we can couple the mutant populations to simpler birth death processes (with immigration) to estimate their growth. In [45] we find a collection of general results on the growth of birth death processes (with immigration) which were formerly used to study similar coupling processes. These results however only cover processes with constant parameters. In Appendix B.6, we argue that we can essentially work with the time-average fitness $f^{\mathrm{av}}_{w,v}$ as the mutants growth rate since $\lambda_K \ll \ln K$, i.e. the parameter fluctuations occur on a faster time-scale than the growth of the mutants. This requires a careful rerun of the proofs in [45] to keep track of the error stemming from this averaging approximation.

Finally, based on the results on the coupling processes, we can derive the piecewise affine growth of the orders of population sizes β_w^K of the mutant populations as in (B.2.13). The equations for $\bar{\beta}_w$ combine the growth of a mutant at the rate of its own fitness $f_{w,v}^{\mathrm{av}}$ with the growth due to incoming mutants from other traits u (which themselves grow at least at rate $f_{u,v}^{\mathrm{av}}$).

B.3.2. Discussion of assumptions

In the following, we address the necessity of some of our assumptions and discuss possible extensions to more general cases.

Mutation kernel and probability

In Section B.2.1 we choose the mutation kernel (or law of the trait of a mutant) $(m_{v,\cdot})_{v\in V}$ to be independent of the phases $i = 1, ..., \ell$ and the carrying capacity K. Moreover, in Section B.2.2 we choose the probability of mutation at birth as

$$\mu_K = K^{-1/\alpha}, \qquad \alpha \in \mathbb{R}_{>0} \setminus \mathbb{N}, \tag{B.3.1}$$

which is independent of the phases and traits and depends on K in a very specific way. Both of these assumptions are not necessary and purely made to simplify notation.

The important part is to ensure that, for each $v, w \in V$, $m_{v,w}^{(i,K)} > 0$ either during all or during none of the phases i (i.e. the mutation graph G does not depend on the phase), that $\mu_K^{(i,v)} > 0$ for all phases and traits and that the (additional) dependence on K does not influence the order of the population sizes. Overall, we can allow for dependences of the form

$$m_{v,w}^{(i,K)} \mu_K^{(i,v)} = h(v,w,i,K) K^{-1/\alpha}, \tag{B.3.2}$$

where, for each $(v,w) \in E$,

$$h(v,w,i,K) > 0 \text{ and } \left| \frac{\ln h(v,w,i,K)}{\ln K} \right| = o(1) \tag{B.3.3}$$

and, for each $(v,w) \notin E$, $h(v,w,i,K) \equiv 0$.

Under these assumptions, the mutation kernel and probability only contribute a (varying) multiplicative lower order constant to the mutant population sizes (beyond the $K^{-1/\alpha}$) and do not affect the traits' fitnesses. As a consequence, neither the asymptotic growth of the order of the population size β^K, which determines the next invading trait, nor the outcome of the invasion according to the Lotka-Volterra dynamics are affected. Therefore, the limiting processes $\bar{\beta}$ and ν would remain unchanged.

The technical assumption of $\alpha \notin \mathbb{N}$ allows for the dichotomy that either $K\mu_K^{d(v,w)} \gg 1$, if $d(v,w) < \alpha$, or $K\mu_K^{d(v,w)} \ll 1$, if $d(v,w) > \alpha$. Hence, one can decide whether a w-population gets founded by mutation or not. However, we think that the critical case where w-mutants arrive at a rate of order $K^0 = 1$ and may go extinct due to stochastic fluctuations can be handled as well. On any diverging time-scale ($\ll \lambda_K$) we see an infinite number of mutations and corresponding attempts to survive and fixate in the population, where survival is also decided on the same short time-scale. As a result, one can heuristically expect fixation with probability converging to one. This type of argument is part of ongoing research. Its details are quite involved and diverge too far from the core of this paper. Therefore, we exclude these cases in our present results.

Monomorphic resident population

Through termination criterion (d), we ensure that $f^i_{v_{k-1},v_k} < 0$, for all phases $i = 1, ..., \ell$. This is a sufficient criterion to imply a monomorphic equilibrium of trait v_k as the outcome of the Lotka-Volterra dynamics involving the former resident trait v_{k-1} and the newly macroscopic mutant v_k. While this criterion is not necessary (as shown in the first example below), we do want to guarantee a monomorphic resident population at all times.

The reason for this is that our potential theoretic approach to proving good bounds on the resident's population size in Appendix B.5 (which is made use of in Section B.4.1 to derive the bounding functions $\phi^{(K,\varepsilon,\pm)}$), relies on estimating the influence of variations in the absolute value of the population size. In the case of a monomorphic population, larger variations can be attributed to either a severe over- or undershoot of the equilibrium population size. In the case of polymorphy however, variations could stem from either of the resident subpopulations or even a mixture of those, which makes the same estimates no longer useful.

We expect that these problems are more of a technical nature and an extension to polymorphic resident populations is part of our ongoing research.

B.3.3. Examples

In this section, we present two examples that provide some more insight into the assumptions made to ensure a monomorphic resident population.

Ensured monomorphism despite temporarily fit resident trait

We consider the example of an invasion step (i.e. the last Lotka-Volterra like step of an invasion, where an already macroscopic mutant population takes over a former resident population) for $\ell = 3$ phases, with resident trait v and mutant trait w, where termination criterion (d) is triggered, i.e. $f^i_{v,w} \geq 0$ for one of the phases $i \in \{1, 2, 3\}$. We impose the following fitness landscape:

$$f^{\mathrm{av}}_{w,v} > 0,\ f^{\mathrm{av}}_{v,w} < 0, \tag{B.3.4}$$

$$f^1_{w,v} > 0,\ f^1_{v,w} < 0; \quad f^2_{w,v} < 0,\ f^2_{v,w} < 0; \quad f^3_{w,v} < 0,\ f^3_{v,w} > 0, \tag{B.3.5}$$

$$T_2 f^2_{v,w} + T_3 f^3_{v,w} < 0. \tag{B.3.6}$$

The first part of (B.3.4) ensures that the mutant w population reaches a macroscopic population size of order K in the presence of a resident v population in the first place. Next, the conditions on $f^i_{w,v}$ in (B.3.5) imply that trait w can only invade the resident population in phase 1 and $f^1_{v,w} < 0$ guarantees that a monomorphic equilibrium of trait w is obtained. Moreover, the population of trait v drops to a mesoscopic size (strictly smaller order than K) by the end of phase 1. In phase 2, due to the respective negative fitnesses, trait w stays in its monomorphic equilibrium while the population size of v shrinks further with rate $f^2_{v,w} < 0$. In phase 3, trait v is indeed fit and can grow again (triggering termination criterion (d)). However, (B.3.6) together with the precise approximations in Appendix B.7 ensures that this growth does not make up for the decrease in population size during phase 2 and hence v will not reach the εK-threshold again. Finally, the second part of (B.3.4) implies that the v population shrinks overall and becomes microscopic on the $\ln K$-time-scale.

To summarise, we have shown that termination criterion (d) is not necessary to guarantee a monomorphic resident population. However, formulating a sharp criterion is much more complicated.

Different possible outcomes for two-phase cycles

As a toy example, to motivate our assumptions/termination criteria for the fitness landscape, we consider all possible behaviours during an invasion step for $\ell = 2$ phases with resident trait v and macroscopic mutant w. For the latter to be able to reach a macroscopic size, we need $f^{\text{av}}_{w,v} > 0$, which implies that there is at least one phase during which $f^i_{w,v} > 0$. Hence, excluding cases of fitness 0, there are seven possible scenarios (up to exchangeability of phases):

scenario	1	2	3	4	5	6	7
$f^1_{v,w}$ and $f^2_{v,w}$	-/-	-/-	+/+	+/+	+/-	+/-	+/-
$f^1_{w,v}$ and $f^2_{w,v}$	+/+	+/-	+/+	+/-	+/+	+/-	-/+

The analysis of the different scenarios again makes use of the estimates in Appendix B.7 and we only present the heuristics here.

Scenarios 1 and 2 are covered by our results and lead to a new monomorphic resident population of trait w.

Scenario 3 yields a polymorphic resident population of coexisting traits v and w. This is because in both phases the respective positive invasion fitnesses imply that there is a unique stable equilibrium point of the two-dimensional Lotka-Volterra system with both components being strictly positive (see e.g. [105, Ch. 2.4.3] for a discussion of stable equilibria for two-dimensional Lotka-Volterra systems).

Scenario 4 does not lead to a fixed resident population. During phase 1, traits v and w obtain a polymorphic coexistence equilibrium as in scenario 3. In phase 2 however, a monomorphic v population is the only stable equilibrium state and the w population starts to shrink again. Since $f^{\text{av}}_{w,v} > 0$ is assumed, we have $T_1 f^1_{w,v} + T_2 f^2_{w,v} > 0$, which implies that w recovers from this decline in the next phase 1 and hence the system keeps switching between a coexistence equilibrium of both traits and a monomorphic equilibrium of trait v. (Note that in the beginning of each phase 1, there is a time of order λ_K during which v is still the monomorphic resident trait, before w reaches a critical size to trigger the Lotka-Volterra dynamics again).

Scenario 5 is in some sense the flipped scenario 4. However, we do not have information about the sign of $f^{\mathrm{av}}_{v,w}$. If $f^{\mathrm{av}}_{v,w} > 0$, this is indeed the opposite version and the resident population switches back and forth between a coexistence equilibrium of v and w and a monomorphic equilibrium of trait w only. If $f^{\mathrm{av}}_{v,w} < 0$, then $T_1 f^1_{v,w} + T_2 f^2_{v,w} < 0$. Hence, once the pure w equilibrium is obtained during a phase 1, at the beginning of the next phase 2, v starts to decrease in size. This decline can not be made up by its growth in phase 1 and hence the v population becomes mesoscopic and shrinks on the $\ln K$-time-scale, making w the new monomorphic resident trait. Note that this is an even smaller example for the phenomenon described in Section B.3.3, i.e. an ensured monomorphic new resident population of the mutant trait despite the former resident trait being fit during some phase. However, it is a little more complicated to describe the exact population sizes here (they depend on whether w first becomes macroscopic during phase 1 or 2, where the previous example guarantees invasion during phase 1). Hence we present both examples and treat this one with less detail.

Scenario 6 gets more complicated. In phase 1, the only stable equilibrium is the polymorphic state involving both v and w. In phase 2 however, both monomorphic equilibria of v or w are stable. Hence the dynamics depend on the relation between the coexistence equilibrium state and the regions of attraction for the two monomorphic states. If the former is attracted to the equilibrium of v, the w population shrinks during phase 2 but can recover to re-attain the coexistence state in phase 1 (i.e. the resident population switches between coexistence and trait v alone). If the coexistence state is attracted to the equilibrium of w in phase 2, the outcome again depends on the fitness $f^{\mathrm{av}}_{v,w}$ and whether trait v can make up for its decrease in phase 2 by its growth in phase 1, similarly to scenario 5.

Scenario 7 again has two possible outcomes. The Lotka-Volterra dynamics would lead to monomorphic equilibria of v in phase 1 and of w in phase 2. It depends on the average fitnesses though whether the respective invading traits reach the critical threshold size to trigger these dynamics within these phases. In case of trait w, this is guaranteed by $f^{\mathrm{av}}_{w,v} > 0$. If also $f^{\mathrm{av}}_{v,w} > 0$, both traits grow faster during their respective fit phases than they shrink during their unfit phases. As a consequence, we observe a switching back and forth between the monomorphic equilibria (not necessarily synced up with the phase changes but delayed by a λ_K time, as above). If $f^{\mathrm{av}}_{v,w} < 0$, trait v cannot recover during phase 1 and the system stays in the monomorphic w equilibrium.

Overall, already in this minimal example of two phases, we can observe a variety of different behaviours, ranging from monomorphic equilibrium states to coexistence and switching between those. Which of these is the case not only depends on the invasion fitnesses of the single traits but also on the precise relation between them and the timing of the different phases.

B.4. Proofs

In this chapter, we conduct the proofs of the main results of this paper, i.e. Theorem B.2.4 and Corollary B.2.5. We utilise a number of technical results on birth death processes with self-competition or immigration. To maintain a better readability of the main proofs, these technical results are stated and proved in the appendices.

This chapter is divided into several sections. In Section B.4.1, we discuss the stability of the resident trait during the mutants' growth phase. In Section B.4.2 we prove Theorem B.2.4, i.e. the convergence of the exponents β^K. Finally, in Section B.4.3, we conclude the result on the sequence of resident traits of Corollary B.2.5.

B.4.1. Stability of the resident trait

Since we are working in a regime of periodically changing parameters, we cannot expect the resident population's size to stay close to one fixed value. Instead, the population size is attracted to the respective equilibrium sizes of the different phases. Since the population needs a short time to adapt to the new equilibrium after a change in parameters, we define two functions $\phi_v^{(K,\varepsilon,+)}$ and $\phi_v^{(K,\varepsilon,-)}$ that bound the population size and take into account these short transition phases of length T_ε. We can then prove that, as long as the mutant populations stay small and as $K \to \infty$, the resident's population size stays between these bounding functions for a time of order $\ln K$ with high probability.

We define

$$\begin{aligned}\phi_v^{(K,\varepsilon,+)}(t) &= \begin{cases}\max\{\bar{n}_v^{i-1}, \bar{n}_v^i\} + M\varepsilon & \text{if } t \in (T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_\varepsilon),\\ \bar{n}_v^i + M\varepsilon & \text{if } t \in [T_{i-1}^\Sigma \lambda_K + T_\varepsilon, T_i^\Sigma \lambda_K],\end{cases}\\ \phi_v^{(K,\varepsilon,-)}(t) &= \begin{cases}\min\{\bar{n}_v^{i-1}, \bar{n}_v^i\} - M\varepsilon & \text{if } t \in (T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_\varepsilon),\\ \bar{n}_v^i - M\varepsilon & \text{if } t \in [T_{i-1}^\Sigma \lambda_K + T_\varepsilon, T_i^\Sigma \lambda_K],\end{cases}\end{aligned} \tag{B.4.1}$$

with periodic extension, where $\bar{n}_v^{i-1} := \bar{n}_v^\ell$ for $i = 1$, and $\phi_v^{(K,\varepsilon,\pm)}(0) = \bar{n}_v^\ell \pm M\varepsilon$. Note that these functions also depend on the choices of M and T_ε. To simplify notation, we however do not include those parameters in the functions' names. To mark the time at which the mutant populations become too large and start to significantly perturb the system, we introduce the stopping time

$$S_v^{(K,\varepsilon)} := \inf\left\{t \geq 0 : \sum_{w \neq v} N_w^K(t) \geq \varepsilon K\right\}. \tag{B.4.2}$$

With this notation, the resident's stability result can be stated as follows.

Theorem B.4.1. *There exists a uniform $M < \infty$ and, for all $\varepsilon > 0$ small enough, there exists a deterministic $T_\varepsilon < \infty$ such that, for all traits $v \in V$ such that $b_v^i > d_v^i$, $1 \leq i \leq \ell$, and for all $T < \infty$,*

$$\lim_{K \to \infty} \mathbb{P}\left(\exists\, t \in [0, T \ln K \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [\phi_v^{(K,\varepsilon,-)}(t), \phi_v^{(K,\varepsilon,+)}(t)] \middle| \frac{N_v^K(0)}{K} \in [\phi_v^{(K,\varepsilon,-)}(0) + \varepsilon, \phi_v^{(K,\varepsilon,+)}(0) - \varepsilon] \right) = 0. \tag{B.4.3}$$

Proof. The proof is based on couplings with bounding single-trait birth death processes with self-competition. We proceed in several steps:

1) For a fixed i phase, prove that N_v^K/K gets ε-close to the new equilibrium $\bar{n}_v^i$ within a finite time T_ε and stays bounded until then.
2) For a fixed i phase, prove that N_v^K/K stays ε-close to its equilibrium $\bar{n}_v^i$ after T_ε until the end of the phase.
3) Use the strong Markov property to concatenate multiple phases to obtain a result for $\ln K$ times.

Note that in the following we conduct the proof for a fixed resident trait $v \in V$. Uniform values for M and T_ε can be obtained by taking the maximum over all such traits since we work with a finite trait space. In steps 1 and 2, we prove that the desired bounds fail with a probability in $o(\lambda_K / \ln K)$. This allows us to concatenate $O(\ln K/\lambda_K)$ phases for an overall time horizon of order $\ln K$ in step 3.

Step 1 (attaining the equilibrium): We fix a phase $1 \leq i \leq \ell$ and, without loss of generality, assume that an i phase starts at time $t = 0$ and lasts until $t = T_i \lambda_K$ (we will "reset" time with the help of the Markov property in step 3). We start by showing that there are constants $\underline{C}^i, \overline{C}^i < \infty$ such that, for any $\varepsilon > 0$ and interval $I = [a_1, a_2] \subset (0, \infty)$, there is a deterministic time $T_\varepsilon^{I,i} < \infty$ such that

$$\begin{aligned} \mathbb{P}\Bigg(& \exists\, t \in [0, T_\varepsilon^{I,i} \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [((\bar{n}_v^i - \varepsilon \underline{C}^i) \wedge a_1) - \varepsilon, ((\bar{n}_v^i + \varepsilon \overline{C}^i) \vee a_2) + \varepsilon] \\ & \text{or } T_\varepsilon^{I,i} \leq S_v^{(K,\varepsilon)} \ \& \ \frac{N_v^K(T_\varepsilon^{I,i})}{K} \notin [\bar{n}_v^i - \varepsilon \underline{C}^i - 2\varepsilon, \bar{n}_v^i + \varepsilon \overline{C}^i + 2\varepsilon] \Bigg| \frac{N_v^K(0)}{K} \in [a_1, a_2] \Bigg) \\ & = o\left(\frac{\lambda_K}{\ln K} \right) \quad \text{as } K \to \infty. \end{aligned} \tag{B.4.4}$$

While these bounds seem quite unintuitive, they come up naturally by first comparing the actual process N_v^K to two branching processes and then comparing these to their deterministic equivalent. These approximations are discussed in detail below and are visualised in Figure B.1. The first line of B.4.4 corresponds to a worst case bound of the population up to the time $T_\varepsilon^{I,i}$ when the new equilibrium is (almost) obtained. The second line corresponds to (almost) reaching the new equilibrium at time $T_\varepsilon^{I,i}$ itself.

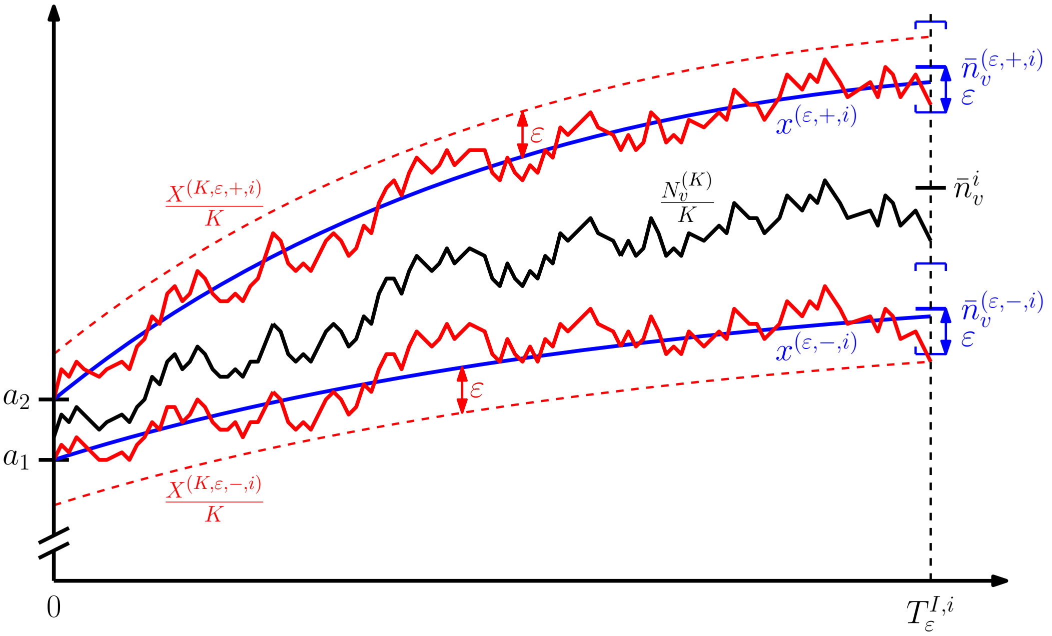


Figure B.1.: Two steps of approximation: Original process N_v^K/K and new equilibrium $\bar{n}_v^i$ in black. Bounding birth death processes with self-competition $X^{(K,\varepsilon,-,i)}/K$ and $X^{(K,\varepsilon,+,i)}/K$ in red. Limiting deterministic solutions $x^{(\varepsilon,-,i)}$ and $x^{(\varepsilon,+,i)}$ with respective perturbed equilibrium sizes $\bar{n}_v^{(\varepsilon,-,i)} = \bar{n}_v^i - \varepsilon \underline{C}^i$ and $\bar{n}_v^{(\varepsilon,+,i)} = \bar{n}_v^i + \varepsilon \overline{C}^i$ in blue.

We want to apply the results from Appendix B.5. To do so, we couple the process N_v^K to two single-trait birth death processes with self-competition $X^{(K,\varepsilon,-,i)}$ and $X^{(K,\varepsilon,+,i)}$ such that

$$X^{(K,\varepsilon,-,i)}(t) \leq N_v^K(t) \leq X^{(K,\varepsilon,+,i)}(t), \quad \forall\, t \in [0, S_v^{(K,\varepsilon)} \wedge \tau_v^{(K,\varepsilon)} \wedge T_i \lambda_K], \tag{B.4.5}$$

where

$$\tau_v^{(K,\varepsilon)} := \inf\left\{t \geq 0 : N_v^K(t) < \varepsilon K\right\} \tag{B.4.6}$$

ensures that the resident does not become too small and incoming mutants can hence be approximated by additional clonal births. We let K be large enough such that $\mu_K < \varepsilon$.

$X^{(K,\varepsilon,-,i)}$ assumes the lowest starting value, maximal competition from other traits, maximal loss due to outgoing mutants and no incoming mutants. It hence has

- initial condition $X^{(K,\varepsilon,-,i)}(0) = \lfloor a_1 K \rfloor$,
- birth rate $b_v^i(1-\varepsilon)$,
- death rate $d_v^i + \varepsilon \max_{w \neq v} c_{vw}^i$ and
- self-competition rate c_{vv}^i/K.

$X^{(K,\varepsilon,+,i)}$ assumes the highest starting value, no competition from other traits, no outgoing mutation and maximal incoming mutation. It hence has

- initial condition $X^{(K,\varepsilon,+,i)}(0) = \lceil a_2 K \rceil$,
- birth rate $b^i_v + \varepsilon \max_{w\neq v} b^i_w$,
- death rate d^i_v and
- self-competition rate c^i_{vv}/K.

These couplings can be explicitly constructed via Poisson process representations, see e.g. [29, Ch. 7.2].

By Theorem B.5.3, on finite time intervals, the rescaled processes $X^{(K,\varepsilon,-,i)}/K$ and $X^{(K,\varepsilon,+,i)}/K$ converge uniformly on finite time intervals to the solutions $(x^{(\varepsilon,-,i)}(t))_{t\geq 0}$ and $(x^{(\varepsilon,+,i)}(t))_{t\geq 0}$ of the ordinary differential equations

$$\begin{aligned} \dot{x}^{(\varepsilon,-,i)}(t) &= x^{(\varepsilon,-,i)}(t)\left(b^i_v(1-\varepsilon) - (d^i_v + \varepsilon \max_{w\neq v} c^i_{vw}) - c^i_{vv} x^{(\varepsilon,-,i)}(t)\right), & x^{(\varepsilon,-,i)}(0) &= a_1, \\ \dot{x}^{(\varepsilon,+,i)}(t) &= x^{(\varepsilon,+,i)}(t)\left(b^i_v + \varepsilon \max_{w\neq v} b^i_w - d^i_v - c^i_{vv} x^{(\varepsilon,+,i)}(t)\right), & x^{(\varepsilon,+,i)}(0) &= a_2. \end{aligned} \tag{B.4.7}$$

These equations have unique attractive equilibrium points

$$\begin{aligned} \bar{n}^{(\varepsilon,-,i)}_v &:= \frac{b^i_v(1-\varepsilon) - (d^i_v + \varepsilon \max_{w\neq v} c^i_{vw})}{c^i_{vv}} = \bar{n}^i_v - \varepsilon \frac{b^i_v + \max_{w\neq v} c^i_{vw}}{c^i_{vv}} =: \bar{n}^i_v - \varepsilon \underline{C}^i, \\ \bar{n}^{(\varepsilon,+,i)}_v &:= \frac{b^i_v + \varepsilon \max_{w\neq v} b^i_w - d^i_v}{c^i_{vv}} = \bar{n}^i_v + \varepsilon \frac{\max_{w\neq v} b^i_w}{c^i_{vv}} =: \bar{n}^i_v + \varepsilon \overline{C}^i, \end{aligned} \tag{B.4.8}$$

which their solutions will attain up to an ε within a finite time $T^{I,i}_\varepsilon < \infty$, i.e.

$$\left| x^{(\varepsilon,-,i)}(T^{I,i}_\varepsilon) - \bar{n}^{(\varepsilon,-,i)} \right| \leq \varepsilon \text{ and } \left| x^{(\varepsilon,+,i)}(T^{I,i}_\varepsilon) - \bar{n}^{(\varepsilon,+,i)} \right| \leq \varepsilon. \tag{B.4.9}$$

Moreover, due to the monotonicity of the solutions, for all $t \in [0, T^{I,i}_\varepsilon]$,

$$x^{(\varepsilon,-,i)}(t) \geq a_1 \wedge \bar{n}^{(\varepsilon,-,i)} \text{ and } x^{(\varepsilon,+,i)}(t) \leq a_2 \vee \bar{n}^{(\varepsilon,+,i)}. \tag{B.4.10}$$

We can hence bound

$$\begin{aligned}
&\mathbb{P}\Bigg(\exists\, t \in [0, T_\varepsilon^{I,i} \wedge S_v^{(K,\varepsilon)} \wedge \tau_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [((\bar{n}_v^i - \varepsilon\underline{C}^i) \wedge a_1) - \varepsilon, ((\bar{n}_v^i + \varepsilon\overline{C}^i) \vee a_2) + \varepsilon] \\
&\qquad \text{or } T_\varepsilon^{I,i} \leq S_v^{(K,\varepsilon)} \wedge \tau_v^{(K,\varepsilon)} \ \&\ \frac{N_v^K(T_\varepsilon^{I,i})}{K} \notin [\bar{n}_v^i - \varepsilon\underline{C}^i - 2\varepsilon, \bar{n}_v^i + \varepsilon\overline{C}^i + 2\varepsilon] \Bigg| \frac{N_v^K(0)}{K} \in [a_1, a_2]\Bigg) \\
&\leq \mathbb{P}\Bigg(\exists\, t \in [0, T_\varepsilon^{I,i}] : \frac{X^{(K,\varepsilon,-,i)}(t)}{K} < ((\bar{n}_v^i - \varepsilon\underline{C}^i) \wedge a_1) - \varepsilon \text{ or } \frac{X^{(K,\varepsilon,+,i)}(t)}{K} > ((\bar{n}_v^i + \varepsilon\overline{C}^i) \vee a_2) + \varepsilon \\
&\qquad \text{or } \frac{X^{(K,\varepsilon,-,i)}(T_\varepsilon^{I,i})}{K} < \bar{n}_v^i - \varepsilon\underline{C}^i - 2\varepsilon \text{ or } \frac{X^{(K,\varepsilon,+,i)}(T_\varepsilon^{I,i})}{K} > \bar{n}_v^i + \varepsilon\overline{C}^i + 2\varepsilon\Bigg) \\
&\leq \mathbb{P}\Bigg(\exists\, t \in [0, T_\varepsilon^{I,i}] : \left|\frac{X^{(K,\varepsilon,-,i)}(t)}{K} - x^{(\varepsilon,-,i)}(t)\right| > \varepsilon \text{ or } \left|\frac{X^{(K,\varepsilon,+,i)}(t)}{K} - x^{(\varepsilon,+,i)}(t)\right| > \varepsilon\Bigg) \\
&\leq \mathbb{P}\left(\sup_{t \leq T_\varepsilon^{I,i}} \left|\frac{X^{(K,\varepsilon,-,i)}(t)}{K} - x^{(\varepsilon,-,i)}(t)\right| > \varepsilon\right) + \mathbb{P}\left(\sup_{t \leq T_\varepsilon^{I,i}} \left|\frac{X^{(K,\varepsilon,+,i)}(t)}{K} - x^{(\varepsilon,+,i)}(t)\right| > \varepsilon\right) \\
&= o\left(\frac{\lambda_K}{\ln K}\right), \qquad\qquad \text{(B.4.11)}
\end{aligned}$$

where we apply Theorem B.5.3 in the last step.

Finally, we note that, if we choose ε small enough such that $\varepsilon < ((\bar{n}_v^i - \varepsilon\underline{C}^i) \wedge a_1) - 2\varepsilon$, then $\tau_v^{(K,\varepsilon)} < T_\varepsilon^{I,i} \wedge S_v^{(K,\varepsilon)}$ implies that $N_v^K(t)/K$ must have left the above intervals prior to this time and hence we can drop the stopping time $\tau_v^{(K,\varepsilon)}$ from the probability on the left hand side.

Step 2 (stability of the equilibrium): We still study a specific i phase from time $t = 0$ up to $t = T_i\lambda_K$ and now consider the time span $[T_\varepsilon^{I,i}, T_i\lambda_K]$. Since this is a time span of divergent length, we can no longer apply Theorem B.5.3 and the convergence to the deterministic system. Instead, we apply Theorem B.5.2 on the stability of equilibrium points for $\ln K$ times to derive

$$\begin{aligned}
\lim_{K\to\infty} \frac{\ln K}{\lambda_K} \cdot \mathbb{P}\Bigg(&\exists\, t \in [T_\varepsilon^{I,i}, T_i\lambda_K \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [\bar{n}_v^i - \varepsilon\underline{C}^i - 16\varepsilon, \bar{n}_v^i + \varepsilon\overline{C}^i + 16\varepsilon]\Bigg| \\
&\qquad\qquad \frac{N_v^K(T_\varepsilon^{I,i})}{K} \in [\bar{n}_v^i - \varepsilon\underline{C}^i - 2\varepsilon, \bar{n}_v^i + \varepsilon\overline{C}^i + 2\varepsilon]\Bigg) = 0 \qquad \text{(B.4.12)}
\end{aligned}$$

We utilise the coupling processes $X^{(K,\varepsilon,-,i)}$ and $X^{(K,\varepsilon,+,i)}$ as in step 1, with the same birth, death and self-competition rates but this time with initial conditions

$$X^{(K,\varepsilon,-,i)}(T_\varepsilon^{I,i}) = \lfloor(\bar{n}_v^i - \varepsilon(\underline{C}^i + 2))K\rfloor \quad \text{and} \quad X^{(K,\varepsilon,+,i)}(T_\varepsilon^{I,i}) = \lceil(\bar{n}_v^i + \varepsilon(\overline{C}^i + 2))K\rceil. \qquad \text{(B.4.13)}$$

Then, if $N_v^K(T_\varepsilon^{I,i})/K \in [\bar{n}_v^i - \varepsilon(\underline{C}^i + 2), \bar{n}_v^i + \varepsilon(\overline{C}^i + 2)]$,

$$X^{(K,\varepsilon,-,i)}(t) \leq N_v^K(t) \leq X^{(K,\varepsilon,+,i)}(t), \quad \forall\, t \in [T_\varepsilon^{I,i}, S_v^{(K,\varepsilon)} \wedge \tau_v^{(K,\varepsilon)} \wedge T_i\lambda_K]. \tag{B.4.14}$$

We apply Theorem B.5.2 with $\varepsilon' = 16\varepsilon$, and hence

$$\left| X_v^{(K,\varepsilon,\pm,i)}(T_\varepsilon^{I,i}) - \bar{n}_v^{(\varepsilon,\pm,i)} K \right| \leq 2\varepsilon K = \frac{\varepsilon' K}{8} < \frac{1}{2} \left\lfloor \frac{\varepsilon' K}{2} \right\rfloor, \tag{B.4.15}$$

to obtain

$$\begin{aligned}
&\mathbb{P}\Bigg(\exists\, t \in [T_\varepsilon^{I,i}, T_i\lambda_K \wedge S_v^{(K,\varepsilon)} \wedge \tau_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [\bar{n}_v^i - \varepsilon\underline{C}^i - 16\varepsilon, \bar{n}_v^i + \varepsilon\overline{C}^i + 16\varepsilon] \Bigg| \\
&\qquad \frac{N_v^K(T_\varepsilon^{I,i})}{K} \in [\bar{n}_v^i - \varepsilon\underline{C}^i - 2\varepsilon, \bar{n}_v^i + \varepsilon\overline{C}^i + 2\varepsilon] \Bigg) \\
\leq& \mathbb{P}\left(\exists\, t \in [T_\varepsilon^{I,i}, T_i\lambda_K] : \frac{X^{(K,\varepsilon,-,i)}(t)}{K} < \bar{n}_v^i - \varepsilon\underline{C}^i - 16\varepsilon \quad \text{or} \quad \frac{X^{(K,\varepsilon,+,i)}(t)}{K} > \bar{n}_v^i + \varepsilon\overline{C}^i + 16\varepsilon \right) \\
\leq& \mathbb{P}\left(\exists\, t \in [T_\varepsilon^{I,i}, T_i\lambda_K] : \left| \frac{X^{(K,\varepsilon,-,i)}(t)}{K} - \bar{n}^{(\varepsilon,-,i)} \right| > \varepsilon' \quad \text{or} \quad \left| \frac{X^{(K,\varepsilon,+,i)}(t)}{K} - \bar{n}_\varepsilon^{(\varepsilon,+,i)} \right| > \varepsilon' \right) \\
\leq& \mathbb{P}\left(\sup_{t \in [T_\varepsilon^{I,i}, T_i\lambda_K]} \left| X^{(K,\varepsilon,-,i)}(t) - \bar{n}^{(\varepsilon,-,i)} K \right| > \varepsilon' K \right) + \mathbb{P}\left(\sup_{t \in [T_\varepsilon^{I,i}, T_i\lambda_K]} \left| X^{(K,\varepsilon,+,i)}(t) - \bar{n}^{(\varepsilon,+,i)} K \right| > \varepsilon' K \right) \\
=& o\left(\frac{\lambda_K}{\ln K} \right).
\end{aligned} \tag{B.4.16}$$

As in step 1, for sufficiently small ε, we can again drop the stopping time $\tau_v^{(K,\varepsilon)}$ from the probability on the left hand side.

Step 3 (concatenating multiple phases): We first piece together steps 1 and 2 to obtain a result for an entire i phase and then concatenate multiple phases to prove the final result of the Theorem.

Applying the Markov property (at $T^{I,i}_\varepsilon$) in the first step, we obtain

$$\begin{aligned}
&\mathbb{P}\Bigg(\exists\ t \in [0, T^{I,i}_\varepsilon \wedge S^{(K,\varepsilon)}_v] : \frac{N^K_v(t)}{K} \notin [((\bar{n}^i_v - \varepsilon\underline{C}^i) \wedge a_1) - \varepsilon, ((\bar{n}^i_v + \varepsilon\overline{C}^i) \vee a_2) + \varepsilon] \\
&\qquad\qquad \text{or } \exists\ t \in [T^{I,i}_\varepsilon \wedge S^{(K,\varepsilon)}_v, T_i\lambda_K \wedge S^{(K,\varepsilon)}_v] : \frac{N^K_v(t)}{K} \notin [\bar{n}^i_v - \varepsilon\underline{C}^i - 16\varepsilon, \bar{n}^i_v + \varepsilon\overline{C}^i + 16\varepsilon] \Bigg| \\
&\qquad\qquad \frac{N^K_v(0)}{K} \in [a_1, a_2]\Bigg) \\
&\leq \mathbb{P}\Bigg(\exists\ t \in [0, T^{I,i}_\varepsilon \wedge S^{(K,\varepsilon)}_v] : \frac{N^K_v(t)}{K} \notin [((\bar{n}^i_v - \varepsilon\underline{C}^i) \wedge a_1) - \varepsilon, ((\bar{n}^i_v + \varepsilon\overline{C}^i) \vee a_2) + \varepsilon] \\
&\qquad\qquad \text{or } T^{I,i}_\varepsilon \leq S^{(K,\varepsilon)}_v\ \&\ \frac{N^K_v(T^{I,i}_\varepsilon)}{K} \notin [\bar{n}^i_v - \varepsilon\underline{C}^i - 2\varepsilon, \bar{n}^i_v + \varepsilon\overline{C}^i + 2\varepsilon] \Bigg| \frac{N^K_v(0)}{K} \in [a_1, a_2]\Bigg) \\
&\quad + \mathbb{P}\Bigg(\exists\ t \in [T^{I,i}_\varepsilon, T_i\lambda_K \wedge S^{(K,\varepsilon)}_v] : \frac{N^K_v(t)}{K} \notin [\bar{n}^i_v - \varepsilon\underline{C}^i - 16\varepsilon, \bar{n}^i_v + \varepsilon\overline{C}^i + 16\varepsilon] \Bigg| \\
&\qquad\qquad \frac{N^K_v(T^{I,i}_\varepsilon)}{K} \in [\bar{n}^i_v - \varepsilon\underline{C}^i - 2\varepsilon, \bar{n}^i_v + \varepsilon\overline{C}^i + 2\varepsilon]\Bigg) \\
&= o\left(\frac{\lambda_K}{\ln K}\right).
\end{aligned} \tag{B.4.17}$$

Here, we impose stronger bounds in the first time period up to $T^{(I,i)}_\varepsilon$ to ensure good initial conditions for the remaining diverging time.

Note that these probabilities are in $o(\lambda_K / \ln K)$ uniformly in $1 \leq i \leq \ell$.

Now we can finally link together multiple phases. For ease of notation, we index the phases by $i \in \mathbb{N}$ instead of $1 \leq i \leq \ell$, where every $(k\ell + i)^{\text{th}}$ phase, $k \in \mathbb{N}$, is of type i and length $T_i\lambda_K$. Similarly, we extend the definitions of T^Σ_i, $\bar{n}^i_v$ and $T^{I,i}_\varepsilon$.

Choosing $T_\varepsilon = \max_{1\leq i\leq\ell} T^{I,i}_\varepsilon$ and $M = \max_{1\leq i\leq\ell}(\underline{C}^i \vee \overline{C}^i) + 17$ in the definition of $\phi^{(K,\varepsilon,\pm)}_v$, and the intervals $I = [a^1_1, a^1_2] = [\bar{n}^\ell_v - \varepsilon(M-1), \bar{n}^\ell_v + \varepsilon(M-1)]$ as well as $I = [a^i_1, a^i_2] = [\bar{n}^{i-1}_v - \varepsilon(\underline{C}^{i-1} + 16), \bar{n}^{i-1}_v + \varepsilon(\overline{C}^{i-1} + 16)]$, $i \geq 2$, in (B.4.17), we deduce the convergence for any $T < \infty$. See Figure B.2 for a visualisation of the concatenation of two phases.

$$
\begin{aligned}
&\mathbb{P}\Bigg(\exists\, t \in [0, T\ln K \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [\phi_v^{(K,\varepsilon,-)}(t), \phi_v^{(K,\varepsilon,+)}(t)] \Bigg| \\
&\quad \frac{N_v^K(0)}{K} \in [\phi_v^{(K,\varepsilon,-)}(0) + \varepsilon, \phi_v^{(K,\varepsilon,+)}(0) - \varepsilon]\Bigg) \\
&= \mathbb{P}\Bigg(\exists\, i \in \mathbb{N} : T_{i-1}^{\Sigma}\lambda_K \leq T\ln K \wedge S_v^{(K,\varepsilon)} \text{ and} \\
&\quad \exists\, t \in (T_{i-1}^{\Sigma}\lambda_K, (T_{i-1}^{\Sigma}\lambda_K + T_{\varepsilon}^{I,i}) \wedge S_v^{(K,\varepsilon)}) : \frac{N_v^K(t)}{K} \notin [(\bar{n}_v^{i-1} \wedge \bar{n}_v^{i}) - M\varepsilon, (\bar{n}_v^{i-1} \vee \bar{n}_v^{i}) + M\varepsilon] \\
&\quad \text{or } \exists\, t \in [(T_{i-1}^{\Sigma}\lambda_K + T_{\varepsilon}^{I,i}) \wedge S_v^{(K,\varepsilon)}, T_i^{\Sigma}\lambda_K \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [\bar{n}_v^{i} - M\varepsilon, \bar{n}_v^{i} + M\varepsilon] \Bigg| \\
&\quad \frac{N_v^K(0)}{K} \in [\phi^{(K,\varepsilon,-)}(0) + \varepsilon, \phi^{(K,\varepsilon,+)}(0) - \varepsilon]\Bigg) \\
&\leq \mathbb{P}\Bigg(\exists\, i \in \mathbb{N} : T_{i-1}^{\Sigma}\lambda_K \leq T\ln K \wedge S_v^{(K,\varepsilon)} \text{ and} \\
&\quad \exists\, t \in [T_{i-1}^{\Sigma}\lambda_K, (T_{i-1}^{\Sigma}\lambda_K + T_{\varepsilon}^{I,i}) \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [((\bar{n}_v^{i} - \varepsilon\underline{C}^{i}) \wedge a_1^i) - \varepsilon, ((\bar{n}_v^{i} + \varepsilon\overline{C}^{i}) \vee a_2^i) + \varepsilon] \\
&\quad \text{or } \exists\, t \in [(T_{i-1}^{\Sigma}\lambda_K + T_{\varepsilon}^{I,i}) \wedge S_v^{(K,\varepsilon)}, T_i^{\Sigma}\lambda_K \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [\bar{n}_v^{i} - \varepsilon\underline{C}^{i} - 16\varepsilon, \bar{n}_v^{i} + \varepsilon\overline{C}^{i} + 16\varepsilon] \Bigg| \\
&\quad \frac{N_v^K(0)}{K} \in [\phi^{(K,\varepsilon,-)}(0) + \varepsilon, \phi^{(K,\varepsilon,+)}(0) - \varepsilon]\Bigg) \\
&\leq \sum_{\substack{i\in\mathbb{N}: \\ T_{i-1}^{\Sigma}\lambda_K < T\ln K}} \mathbb{P}\Bigg(\exists\, t \in [T_{i-1}^{\Sigma}\lambda_K, (T_{i-1}^{\Sigma}\lambda_K + T_{\varepsilon}^{I,i}) \wedge S_v^{(K,\varepsilon)}] : \\
&\quad \frac{N_v^K(t)}{K} \notin [((\bar{n}_v^{i} - \varepsilon\underline{C}^{i}) \wedge a_1^i) - \varepsilon, ((\bar{n}_v^{i} + \varepsilon\overline{C}^{i}) \vee a_2^i) + \varepsilon] \\
&\quad \text{or } \exists t \in [(T_{i-1}^{\Sigma}\lambda_K + T_{\varepsilon}^{I,i}) \wedge S_v^{(K,\varepsilon)}, T_i^{\Sigma}\lambda_K \wedge S_v^{(K,\varepsilon)}] : \frac{N_v^K(t)}{K} \notin [\bar{n}_v^{i} - \varepsilon\underline{C}^{i} - 16\varepsilon, \bar{n}_v^{i} + \varepsilon\overline{C}^{i} + 16\varepsilon] \Bigg| \\
&\quad \frac{N_v^K(T_{i-1}^{\Sigma}\lambda_K)}{K} \in [a_1^i, a_2^i]\Bigg) \\
&= o(1), \qquad\qquad \text{(B.4.18)}
\end{aligned}
$$

where we utilise that we have $O(\ln K/\lambda_K)$ summands that are (uniformly) of order $o(\lambda_K/\ln K)$ to conclude.

Note that, in contrast to the second to last expression, in (B.4.17) the initial time of the phase is set to 0. This however does not change the probability due to the Markov property and the periodic time-homogeneity of the Markov process. Letting K tend to infinity, this yields the proof of Theorem B.4.1. □

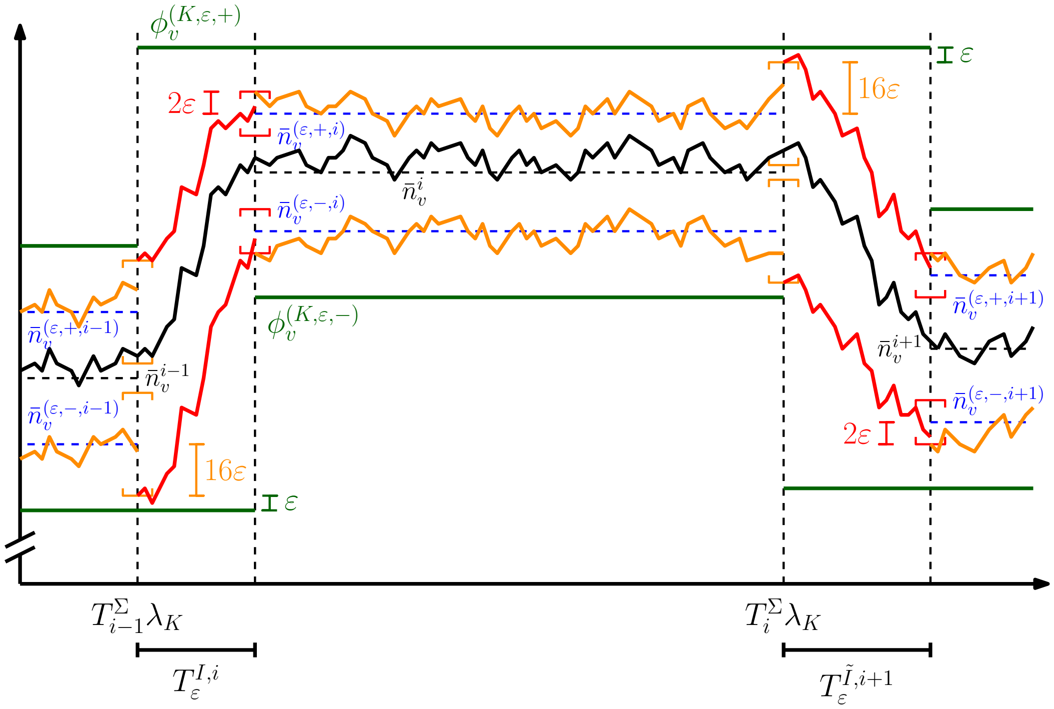


Figure B.2.: Concatenation of phases $i-1$, i and $i+1$. Original process N_v^K/K and corresponding equilibria in black. Bounding birth death processes with self-competition $X^{(K,\varepsilon,-,i)}/K$ and $X^{(K,\varepsilon,+,i)}/K$ in red (fast re-equilibration from step 1) and orange (long stability from step 2). Equilibrium sizes $\bar{n}_v^{(\varepsilon,-,i)} = \bar{n}_v^i - \varepsilon \underline{C}^i$ and $\bar{n}_v^{(\varepsilon,+,i)} = \bar{n}_v^i + \varepsilon \overline{C}^i$ of the corresponding (perturbed) deterministic system in blue. Bounding functions $\phi_v^{(K,\varepsilon,-)}$ and $\phi_v^{(K,\varepsilon,+)}$ in green.

B.4.2. Convergence of the orders of population sizes

The proof of Theorem B.2.4 is based on an induction principle and similar to the proof of the main theorem of [48]. We therefore do not repeat every single detail but point out how to deal with the important difficulties arising from our extended model with time-dependent growth parameters. This is done in five steps:

1) Define the main stopping times and set up the induction.
2) Use the convergence of $\beta_w^K(0)$ for the base case of the induction.
3) Couple the process with non-interacting birth death processes to control the growth of the mutant populations.
4) Ensure that mutants become macroscopic only in a fit phase i.
5) Finish the induction step by comparison to the deterministic Lotka-Volterra system.

Step 1 (preparation): The induction is set up in such a way that each step corresponds to the invasion of a new mutant. We divide these steps into two alternating substeps. During the first one, the resident population is stable in a certain sense and we approximate the growth of the mutant populations on the $\ln K$-time-scale. The second one is started when one of the mutant populations becomes macroscopic and we therefore observe a Lotka-Volterra interaction between the mutant and the former resident population.

In order to make this distinction into substeps rigorous, we introduce, for $k \in \mathbb{N}_0$, the pair of stopping times (visualised in Figure B.3)

$$\begin{aligned} \sigma_k^K &:= \inf\left\{ t \geq \theta_k^K : \frac{N_{v_k}^K(t)}{K} \in [\phi_{v_k}^{(K,\varepsilon_k,-)}(t) + \varepsilon_k, \phi_{v_k}^{(K,\varepsilon_k,+)}(t) - \varepsilon_k] \text{ and } \sum_{w \neq v_k} N_w^K(t) < \varepsilon_k^2 K \right\}, \\ \theta_{k+1}^K &:= \inf\left\{ t \geq \sigma_k^K : \frac{N_{v_k}^K(t)}{K} \notin \left[\phi_{v_k}^{(K,\varepsilon_k,-)}(t), \phi_{v_k}^{(K,\varepsilon_k,+)}(t)\right] \text{ or } \sum_{w \neq v_k} N_w^K(t) \geq \varepsilon_k K \right\}, \end{aligned} \tag{B.4.19}$$

where the $\varepsilon_k > 0$ are chosen at the very end, in reverse order. More precisely, to ensure good estimates until the end of our time horizon $[0, T \ln K]$, one has to keep the accumulating error low from the very beginning and choose each ε_k small enough to provide good initial bounds for the next invasion step.

This means that at time σ_k^K the process has reached the monomorphic Lotka-Volterra-equilibrium of trait $v_k \in V$, and v_k remains the only macroscopic trait until time θ_{k+1}^K. Moreover, its population size lies inside the ε-tunnel described by $\phi^{(K,\varepsilon_k,-)}$ and $\phi^{(K,\varepsilon_k,+)}$ during $[\sigma_k^K, \theta_{k+1}^K]$.

In step 3, we introduce the stopping times s_{k+1}^K, when the first non-resident population becomes *almost macroscopic*, i.e. attains a population size of order $K^{1-\varepsilon_{k+1}}$ for some small $\varepsilon_{k+1} > 0$ (see Figure B.3), as well as the appearance times $t_{w,k+1}^K$ of new mutants. The first one is necessary for technical reasons and gives good bounds for θ_{k+1}^K. The second one is needed to keep track of these new populations. As in [48], let $(\tau_h^K)_{h \geq 0}$ be the collection of both $(s_k^K)_{k \geq 0}$ and $(t_{w,k}^K)_{k \geq 0}$. The main part of the proof then consists of approximating the growth dynamics in the intervals $[\tau_h^K \wedge \theta_{k+1}^K \wedge T, \tau_{h+1}^K \wedge \theta_{k+1}^K \wedge T]$. Subsequently, we estimate the time between θ_{k+1}^K and σ_{k+1}^K, which completes the induction step.

Step 2 (base case): We set $\theta_0^K{=}0$. Then the base case is reminiscent of [48] since, within a finite time horizon $[0, T'_\varepsilon]$, the parameter functions b^K, d^K, c^K stay constant, for K large enough. Hence, we can apply Lemma A.6(ii) of [48] to get, for every $\varepsilon_0 > 0$, a $T'_{\varepsilon_0} < \infty$ such that

$$\lim_{K \to \infty} \mathbb{P}\left(\left\| \frac{N_{v_0}^K(T'_{\varepsilon_0})}{K} - \bar{n}_{v_0}^1 \right\|_\infty < \varepsilon_0 \right) = 1. \tag{B.4.20}$$

Here, our ε_0 corresponds to ε_2 in [48] and a lower bound of order K for the initial population size of trait v_0 gives their $\varepsilon_1 K$. We use that our assumption on the initial condition (B.2.11) guarantees that $\lim_{K \to \infty} \beta_w(0) < 1$, for all $w \neq v_0$. Hence $\lim_{K \to \infty} \beta_w(t) < 1$ for all $w \neq v_0$ and $t \in [0, T'_{\varepsilon_0}]$, and in particular $\sum_{w \neq v_0} N_w^K(t) < \varepsilon_0^2 K$ for such t. Therefore, Assumption A.5 in [48] is satisfied for standard couplings to birth death processes with immigration,

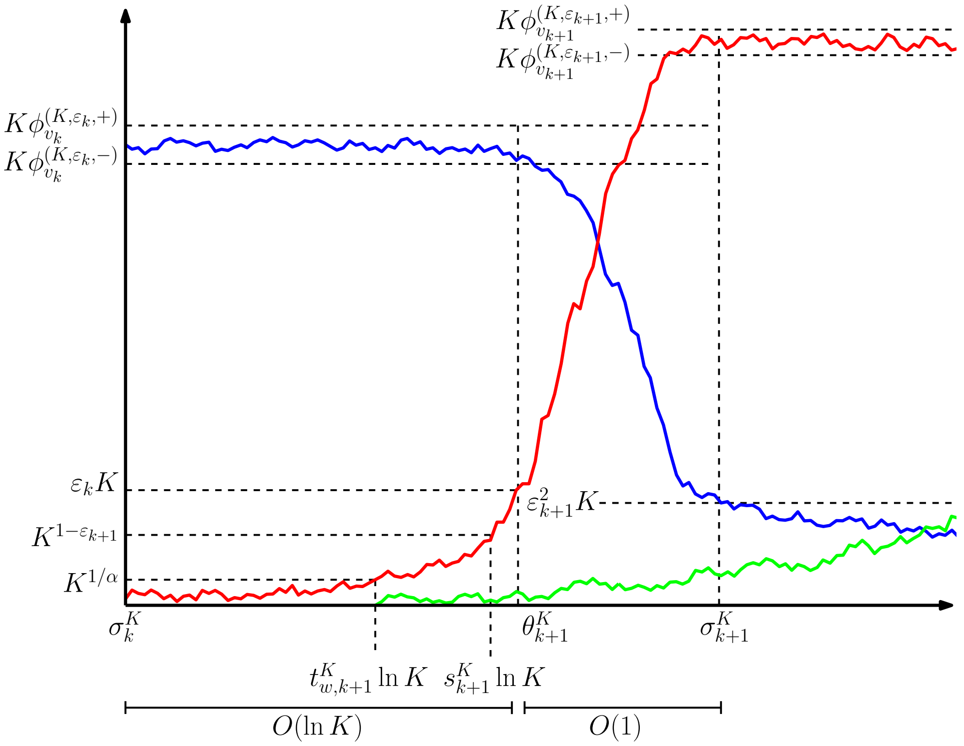


Figure B.3.: Substeps of the $(k+1)^{\text{st}}$ invasion: Resident population $N_{v_k}^K(t)$ (blue), invading mutant $N_{v_{k+1}}^K(t)$ (red) and new emerging subpopulation $N_w^K(t)$ (green), together with the triggered stopping times and the corresponding bounds and thresholds.

where the birth, death and (self)competition rates do not deviate from $b_{v_0}^1$, $d_{v_0}^1$ and c_{v_0,v_0}^1 by more than $\varepsilon_0^2\hat{C}$, and immigration is bounded by $\varepsilon_0^2 K\mu_K\hat{C}$, for some $\hat{C}<\infty$. Overall, we obtain that $\sigma_0^K \leq T'_{\varepsilon_0} < \infty$, for K large enough.

Step 3 (growth of mutants): To show the induction step, let us assume that at time σ_k^K the process has reached the monomorphic Lotka-Volterra-equilibrium of trait $v_k \in V$. Our first goal is to estimate the competitive interaction between the subpopulations in the interval $[\sigma_k^K, \theta_{k+1}^K]$. Recalling the rates of the different events for the population of trait $w \in V$, which has size $N_w^K(t)$ at time t, we have

- Reproduction without mutation:

$$\mathfrak{b}_w^K(t) = b_w^K(t)(1-\mu_K)N_w^K(t), \tag{B.4.21}$$

- Death (natural and by competition):

$$\mathfrak{d}_w^K(t) = \left(d_w^K(t) + \sum_{u\in V}\frac{c_{w,u}^K(t)}{K}N_u^K(t)\right)N_w^K(t), \tag{B.4.22}$$

- Reproduction from mutation:

$$\mathfrak{bm}_w^K(t) = \mu_K \sum_{u \in V} b_u^K(t) m_{u,w} N_u^K(t). \tag{B.4.23}$$

Using the shorthand notation $v := v_k$ and $\hat{c}_w := \max_{u \neq v, 1 \leq i \leq \ell} c_{w,u}^i$, we can introduce the approximating parameter functions, for $w \in V$,

$$\begin{aligned} b_w^{(K,\varepsilon,+)}(t) &:= b_w^K(t), \\ b_w^{(K,\varepsilon,-)}(t) &:= (1-\varepsilon) b_w^K(t), \\ d_w^{(K,\varepsilon,+)}(t) &:= d_w^K(t) + c_{w,v}^K(t) \phi_v^{(K,\varepsilon,-)}(t), \\ d_w^{(K,\varepsilon,-)}(t) &:= d_w^K(t) + c_{w,v}^K(t) \phi_v^{(K,\varepsilon,+)}(t) + \hat{c}_w \varepsilon. \end{aligned} \tag{B.4.24}$$

If K is taken large enough such that $\mu_K < \varepsilon$, we therefore have, for $t \in [\sigma_k^K, \theta_{k+1}^K]$,

$$\begin{aligned} b_w^{(K,\varepsilon,-)}(t) N_w^K(t) \leq \mathfrak{b}_w^K(t) \leq b_w^{(K,\varepsilon,+)}(t) N_w^K(t), \\ d_w^{(K,\varepsilon,+)}(t) N_w^K(t) \leq \mathfrak{d}_w^K(t) \leq d_w^{(K,\varepsilon,-)}(t) N_w^K(t). \end{aligned} \tag{B.4.25}$$

Moreover, defining $f_{w,v}^{(K,\varepsilon,\pm)}(t) := b_w^{(K,\varepsilon,\pm)}(t) - d_w^{(K,\varepsilon,\pm)}(t)$, we have

$$\begin{aligned} f_{w,v}^{(K,\varepsilon,+)}(t) &= \begin{cases} f_{w,v}^i + c_{w,v}^i M\varepsilon + c_{w,v}^i \left(\bar{n}_v^i - \bar{n}_v^{i-1}\right)_+ & : t \in [T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_\varepsilon), \\ f_{w,v}^i + c_{w,v}^i M\varepsilon & : t \in [T_{i-1}^\Sigma \lambda_K + T_\varepsilon, T_i^\Sigma \lambda_K), \end{cases} \\ f_{w,v}^{(K,\varepsilon,-)}(t) &= \begin{cases} f_{w,v}^i - \left(c_{w,v}^i M + b_w^i + \hat{c}_w\right)\varepsilon - c_{w,v}^i \left(\bar{n}_v^{i-1} - \bar{n}_v^i\right)_+ & : t \in [T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_\varepsilon), \\ f_{w,v}^i - \left(c_{w,v}^i M + b_w^i + \hat{c}_w\right)\varepsilon & : t \in [T_{i-1}^\Sigma \lambda_K + T_\varepsilon, T_i^\Sigma \lambda_K). \end{cases} \end{aligned} \tag{B.4.26}$$

We can use these new parameter and fitness functions to define suitable couplings to simpler bounding branching processes to approximate the original processes N_w^K.

In contrast to the estimates of [48], we have to work with periodic functions instead of constants. Another peculiarity is that we only have good estimates in the intervals $[T_{i-1}^\Sigma \lambda_K + T_\varepsilon, T_i^\Sigma \lambda_K)$. On the intervals $[T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_\varepsilon)$, we have to deal with deviations staying macroscopic (i.e. not scaling with ε). Fortunately, these bad estimates are only given for a finite time T_ε (not increasing with λ_K) and for the remaining time, which scales with λ_K, we have the estimates that are arbitrarily accurate, proportional to $\varepsilon > 0$. In Appendix B.6, we work out how to capture both of these characteristics.

In order to make these results applicable, let us first define the stopping time when the first non-resident population becomes almost macroscopic

$$s_{k+1}^K := \inf\left\{ t \geq \sigma_k^K / \ln K : \exists\ w \neq v_k : \beta_w^K(t) > 1 - \varepsilon_{k+1} \right\}, \tag{B.4.27}$$

as well as the time of appearance of a mutant

$$t_{w,k+1}^K := \begin{cases} \inf\left\{ t \geq \sigma_k^K / \ln K : \exists\ u \in V : d(u,w) = 1, \beta_u^K(t) = \frac{1}{\alpha} \right\} & \text{if } \beta_w^K(\sigma_k^K / \ln K) = 0, \\ \sigma_k^K / \ln K & \text{else.} \end{cases} \tag{B.4.28}$$

Building on this, we can define the sequence of important events $(\tau_h)_{h\in\mathbb{N}_0}$ via $\tau_0^K = \sigma_0^K/\ln K$ and, for $\sigma_k^K/\ln K \leq \tau_{h-1}^K < s_{k+1}^K$,

$$\tau_h^K := s_{k+1}^K \wedge \min\left\{t_{w,k+1} : w \in V, t_{w,k+1} > \tau_{h-1}\right\}. \tag{B.4.29}$$

Moreover, we can then define the sequence of sets of living traits $(M_h^K)_{h\in\mathbb{N}_0}$ via

$$\begin{aligned} M_h^K &:= \left\{w \in V : \beta_w^K(\tau_h^K) > 0 \text{ or } \tau_h^K = t_{w,k+1}^K\right\} \\ &= \left(M_{h-1}^K \setminus \left\{w \in V : \beta_w^K(\tau_h^K) = 0\right\}\right) \cup \left\{w \in V : \tau_h^K = t_{w,k+1}^K\right\}. \end{aligned} \tag{B.4.30}$$

After establishing these stopping times and estimates on the rate functions, we are in a similar framework as in [48] but now adapted to the time-dependence of the driving parameters. We can couple the mutant populations N_w^K to time-dependent birth death processes (with immigration) with parameters $b_w^{(K,\varepsilon_k,\pm)}(t)$ and $d_w^{(K,\varepsilon_k,\pm)}(t)$. Together with the results presented in Appendix B.6, one can now follow the arguments of Section 4.2 in [48]. To be precise, we just have to replace their Lemma A.1 by our Theorem B.6.1 and their Corollary A.4 by our Theorem B.6.9 and use similar inductive arguments to show that, for $w \in M_{h-1}^K$ and $t \in [\tau_{h-1}^K \ln K \wedge \theta_{k+1}^K \wedge T\ln K, \tau_h^K \ln K \wedge \theta_{k+1}^K \wedge T \ln K]$, we obtain the bounds

$$\max_{u\in M_{h-1}^K}\left[\beta_u^K(\tau_{h-1}^K) - \frac{d(u,w)}{\alpha} + (t-\tau_{h-1}^K)(f_{u,v_k}^{\text{av}} - C_h\varepsilon_k)\right]_+ \tag{B.4.31}$$

$$\leq \beta_w^K(t) \leq \max_{u\in M_{h-1}^K}\left[\beta_u^K(\tau_{h-1}^K) - \frac{d(u,w)}{\alpha} + C_h\varepsilon_k + (t-\tau_{h-1}^K)(f_{u,v_k}^{\text{av}} + C_h\varepsilon_k)\right]_+. \tag{B.4.32}$$

A heuristic for these bounds has already been given in Remark 18. As a brief reminder, without mutation, every living trait $u \in M_{h-1}^K$ would grow/shrink at the rate of its own fitness f_{u,v_k}^{av} on the $\ln K$-time-scale, yielding $\beta_u^K(t) \approx \beta_u^K(\tau_{h-1}^K) + (t-\tau_{h-1}^K)f_{u,v_k}^{\text{av}}$. Through mutation however, a trait $w \in V$ receives a $\mu_K^{d(u,w)} = K^{-d(u,w)/\alpha}$ portion of incoming mutants from all living traits $u \in M_{h-1}^K$, and its actual population size corresponds the leading order term, i.e. the maximum of all these exponents $\beta_u^K(\tau_{h-1}^K) + (t-\tau_{h-1}^K)f_{u,v_k}^{\text{av}} - d(u,w)/\alpha$.

Note that these estimates on β_w^K are also where the errors accumulate. Namely, knowing the initial value of β_u^K at time $\tau_{h-1}\ln K$ allows for approximations until $\tau_h \ln K$ but at a cost of an additional error term of order ε_k. To eventually ensure the convergence of Theorem B.2.4, which means having good estimates until time $T\ln K$, one has to choose the ε_k in reverse order, such that every approximation step gives good enough bounds to the initial values of the next one.

By analogous arguments to Section 4.3 and 4.4 of [48], we can deduce the formulas for and the convergence of $(\tau_h^K - \tau_{h-1}^K)$ and M_h^K, as well as for s_{k+1}^K and θ_{k+1}^K. Note that we need to introduce the stopping times s_{k+1}^K for the following technical reason: At time θ_{k+1}^K we only know that the overall mutant population (summing over all non-resident traits) has reached the threshold $\varepsilon_k K$. The above bounds on β_w^K only allow us to estimate a single mutant's population size up to a multiplicative factor of $K^{\pm C\varepsilon_k}$, which is not sufficient to imply that v_{k+1} has a non-vanishing population size (when rescaled by K) at this time. This is however necessary to have a good initial condition for the deterministic Lotka-Volterra approximation.

Hence, s^K_{k+1} is chosen to guarantee the existence of one large mutant population, where we choose a threshold slightly smaller than $\beta^K_w = 1$ to ensure that the Lotka-Volterra dynamics are not triggered before this time either.

We can show that, in the limit of $K \nearrow \infty$ and for small ε_{k+1}, the times $\theta^K_{k+1}/\ln K$ and s^K_{k+1} are arbitrarily close. Namely, following again the argument in [48], we can show by contradiction that

$$s^K_{k+1} \ln K < \theta^K_{k+1} < (s^K_{k+1} + \varepsilon_{k+1} C) \ln K. \tag{B.4.33}$$

To be precise, let w^K_{k+1} be the mutant trait that triggers s^K_{k+1} and take

$$\eta^K_{k+1} := 2\varepsilon_{k+1}/(f^{\mathrm{av}}_{w^K_{k+1},\mathbf{v}_k} - C_h \varepsilon_k). \tag{B.4.34}$$

If one assumes that $(s^K_{k+1} + \eta^K_{k+1}) \ln K \leq \theta^K_{k+1}$, then (B.4.31) would still be applicable and directly lead to

$$\beta^K_{w^K_{k+1}}(s^K_{k+1} + \eta^K_{k+1}) \geq 1 + \varepsilon_{k+1}. \tag{B.4.35}$$

This however is a contradiction since $\lim_{K\to\infty} \beta^K_w(s) \leq 1$ for all $w \in V$ and $s \geq 0$, and hence the upper bound in B.4.33 is satisfied for $C = 2/(f^{\mathrm{av}}_{w^K_{k+1},\mathbf{v}_k} - C_h \varepsilon_k)$. The lower bound is satisfied by definition of the stopping times.

Step 4 (time of invasion): Now the last difference to [48] that we have to address is that the trait reaching a macroscopic size at time θ^K_{k+1}, which is with high probability v_{k+1}, might be unfit at that time. In the following, we show that this only happens with vanishing probability. In order to track the sizes of the different subpopulations more carefully, let us introduce the additional stopping times

$$R^K_{k+1} := \inf\left\{t \geq \sigma^K_k : N^K_{v_{k+1}}(t) \geq \varepsilon_k^2 K \text{ and } f^{(K,\varepsilon,-)}_{v_{k+1},v_k}(t) > 0\right\}, \tag{B.4.36}$$

$$\tilde{R}^K_{k+1} := \inf\left\{t \geq \sigma^K_k : N^K_{v_{k+1}}(t) \geq \varepsilon_k^2 K\right\}, \tag{B.4.37}$$

$$\check{R}^K_{k+1} := \inf\left\{t \geq \tilde{R}^K_{k+1} : \int_{\tilde{R}^K_{k+1}}^t f^{(K,\varepsilon_k,-)}_{v_{k+1},v_k}(s)\mathrm{d}s > 0\right\}. \tag{B.4.38}$$

The first time R^K_{k+1} is the time we are ultimately looking for, namely the starting point for the deterministic Lotka-Volterra system involving the resident trait v_k and the (at that time fit) mutant v_{k+1} (see Step 5 below). The second time $\tilde{R}^K_{k+1}$ gives us a first warning before reaching θ^K_{k+1}, with v_{k+1} possibly being unfit. The last time $\check{R}^K_{k+1}$ helps to estimate the first one and can be computed deterministically in relation to the second one.

Our goal in this step is to prove that $R^K_{k+1} < \theta^K_{k+1}$, such that all branching process approximations apply up to this point. While in the next step we deduce from the Lotka-Volterra system $\sigma^K_{k+1} < R^K_{k+1} + O(1)$.

We know that, for $\varepsilon_k > 0$ small and K large enough,

$$\int_{\tilde{R}^K_{k+1}}^{\tilde{R}^K_{k+1}+\lambda_K T^\Sigma} f^{(K,\varepsilon_k,-)}_{v_{k+1},v_k}(s)\mathrm{d}s \geq \lambda_K T^\Sigma_\ell \left(f^{\mathrm{av}}_{v_{k+1},v_k} - \tilde{M}\varepsilon_k\right) - \ell T_{\varepsilon_k} C_{v_{k+1},v_k} > 0, \tag{B.4.39}$$

since $f^{\mathrm{av}}_{v_{k+1},v_k} > 0$. Which implies directly

$$\check{R}^K_{k+1} \leq \tilde{R}^K_{k+1} + \lambda_K T^{\Sigma}. \tag{B.4.40}$$

Moreover, since $f^{(K,\varepsilon,-)}_{v_{k+1},v_k}(s)$ is piecewise constant and the defining inequality of $\check{R}^K_{k+1}$ is strict, there is a small $\delta > 0$ (not scaling with K) such that, with probability 1,

$$f^{(K,\varepsilon,-)}_{v_{k+1},v_k}(\check{R}^K_{k+1} + t) > 0 \qquad \forall t \in (0,\delta). \tag{B.4.41}$$

As argued above, the interval $[\tilde{R}^K_{k+1}, \check{R}^K_{k+1} + \delta]$ is of length $O(\lambda_K)$. Hence from Corollary B.7.2 and an application of the Markov property at time $\tilde{R}^K_{k+1}$, we can deduce that, for δ small enough,

$$\mathbb{P}\left(\sup_{t\in[\tilde{R}^K_{k+1},\check{R}^K_{k+1}+\delta]} \sum_{w\neq v_k} N^K_w(t) < \varepsilon_k K\right) \overset{K\to\infty}{\longrightarrow} 1, \tag{B.4.42}$$

$$\mathbb{P}\left(N^K_{v_{k+1}}(\check{R}^K_{k+1} + \delta) > \varepsilon_k^2 K\right) \overset{K\to\infty}{\longrightarrow} 1. \tag{B.4.43}$$

The statement of (B.4.42) tells us that the mutant population is still bounded from above and thus the assumptions for Theorem B.4.1 are still satisfied up to time $\check{R}^K_{k+1} + \delta$. Hence we know that the resident population only fluctuates inside the ϕ-tube. This implies

$$\check{R}^K_{k+1} + \delta \leq \theta^K_{k+1}. \tag{B.4.44}$$

Finally (B.4.43) together with (B.4.41) leads to $R^K_{k+1} \leq \check{R}^K_{k+1} + \delta$. This eventuelly gives

$$R^K_{k+1} \leq \theta^K_{k+1}, \tag{B.4.45}$$

i.e. we are still allowed to use the couplings with birth death processes to approximate the mutant population up to R^K_{k+1}.

Step 5 (Lotka-Volterra): At time R^K_{k+1}, we are in position to use the convergence result for the fast Lotka-Volterra phase. By definition of this stopping time, we know that the invading trait v_{k+1} is fit with respect to the resident v_k and of a size that does not vanish as $K \to \infty$ when rescaled by K. Moreover, termination criterion (d) of the algorithm in Theorem B.2.4 ensures that the resident trait is unfit with respect to the invading mutant. By standard arguments, the corresponding deterministic system gets close to its equilibrium in finite time and we have convergence of the stochastic process towards the deterministic system on finite time intervals [75]. This implies the existence of a finite and deterministic time $T'_{\varepsilon_{k+1}} < \infty$ such that

$$\sigma^K_{k+1} \leq R^K_{k+1} + T'_{\varepsilon_{k+1}}. \tag{B.4.46}$$

Moreover, the condition $f^i_{v_k,v_{k+1}} < 0$, for all $i = 1 \ldots \ell$, guarantees that (with probability converging to 1 as $K \to \infty$) the former resident population cannot reach the threshold $\varepsilon_{k+1}K$ any more after time σ^K_{k+1}.

Overall, we have proved that, with probability converging to 1 as $K \to \infty$,

$$R^K_{k+1} \leq \theta^K_{k+1} \leq \sigma^K_{k+1} \leq R^K_{k+1} + T'_{\varepsilon_{k+1}}, \tag{B.4.47}$$

which means that on the logarithmic time-scale there is no difference between $R^K_{k,v_{k+1}}, \theta^K_{k+1}$ and σ^K_{k+1} and dividing by $\ln K$ they all converge to s_{k+1} as claimed.

This finishes the proof of Theorem B.2.4.

B.4.3. Sequence of resident traits

We now turn to the proof of Corollary B.2.5. To prove the convergence with respect to $\mathcal{M}(V)$, equipped with the weak topology, we have to study the integrals $\langle \nu, h\rangle = \int h \mathrm{d}\nu$ of all bounded and continuous functions $h : V \mapsto \mathbb{R}$ with respect to the measures $\nu^K_\varepsilon(s \ln K)$. Since V is discrete and finite, all finite functions satisfy these conditions. For later purpose we denote $\bar{h} := \max_{v \in V} |h(v)|$. Under use of (B.4.19), we have

$$\begin{aligned}\sum_{k\in\mathbb{N}_0} \left(-\mathbb{1}_{\theta^K_k \leq s < \sigma^K_k} 2\bar{h} + \mathbb{1}_{\sigma^K_k \leq s < \theta^K_{k+1}} h(v_k)\right) &\leq \left\langle \nu^K_\varepsilon(s \ln K), h\right\rangle \qquad \text{(B.4.48)}\\ &\leq \sum_{k\in\mathbb{N}_0} \left(\mathbb{1}_{\theta^K_k \leq s < \sigma^K_k} 2\bar{h} + \mathbb{1}_{\sigma^K_k \leq s < \theta^K_{k+1}} h(v_k)\right),\\ \langle \nu(s), h\rangle &= \sum_{k\in\mathbb{N}_0} \mathbb{1}_{s_k \leq s < s_{k+1}} h(v_k). \qquad \text{(B.4.49)}\end{aligned}$$

Since we want to show convergence in $L^p([0,T], \mathcal{M}(V))$, for $p \in [1, \infty)$, we have to compute the distance between the two integrals in the $\|\cdot\|_{L^p([0,T])}$-norm, which can be estimated as follows

$$\begin{aligned}&\left\|\left\langle \nu^K_\varepsilon(\cdot \ln K), h\right\rangle - \langle \nu(\cdot), h\rangle\right\|^p_{L^p([0,T])}\\ \leq &\sum_{k\in\mathbb{N}_0 : s_k < T} \left((3\bar{h})^p \left|\frac{\theta^K_k}{\ln K} - \frac{\sigma^K_k}{\ln K}\right| + (2\bar{h})^p \left|\frac{\sigma^K_k}{\ln K} - s_k\right| + (2\bar{h})^p \left|s_{k+1} - \frac{\theta^K_{k+1}}{\ln K}\right|\right)\\ \leq &(5\bar{h})^p \sum_{k\in\mathbb{N}_0 : s_k < T} \left(\left|\frac{\sigma^K_k}{\ln K} - s_k\right| + \left|\frac{\theta^K_{k+1}}{\ln K} - s_{k+1}\right|\right). \end{aligned} \tag{B.4.50}$$

Here the last step consists of an application of triangle inequality at s_k to estimate the first term, followed by a reordering of the sum. Since for fixed $T > 0$ the sum in fact only consists of finitely many summands and moreover $\sigma^K_k / \ln K \to s_k$ and $\theta^K_{k+1}/ \ln K \to s_{k+1}$ in probability, for $K \to \infty$, we deduce, for all $\delta > 0$,

$$\mathbb{P}\left(\left\|\left\langle \nu^K_\varepsilon(\cdot \ln K), h\right\rangle - \langle \nu(\cdot), h\rangle\right\|_{L^p([0,T])} > \delta\right) \overset{K\to\infty}{\longrightarrow} 0, \tag{B.4.51}$$

which is the claimed convergence.

B.5. Birth death processes with self-competition

In this chapter, we prove some general results on birth death processes with self-competition that are used to obtain bounds on the resident's population size in Section B.4.1. In the first section, we quantify the asymptotic probability of such processes to stay close to their equilibrium state for a long time as K tends to infinity. In the second section, we derive asymptotics for the probability of these processes to stay close to the corresponding deterministic system for a finite time.

Both results apply to processes with constant parameters. More precisely, we study stochastic processes $(X_t^K)_{t\geq 0}$ with birth rate b, natural death rate d and self-competition rate c/K, i.e. with infinitesimal generators

$$\left(\mathcal{L}^K f\right)(n) = nb(f(n+1)-f(n)) + n\left(d+\frac{c}{K}n\right)(f(n-1)-f(n)), \tag{B.5.1}$$

for bounded functions $f:\mathbb{N}_0\to\mathbb{R}$.

B.5.1. Attraction to the equilibrium

We study the probability of birth death processes with self-competition to stay close to their equilibrium $(b-d)K/c$ for a long time. In order to be able to concatenate this result for infinitely many phases in Section B.4.1, we need to bound the probability of diverging from the equilibrium by a sequence that tends to zero fast enough as K tends to infinity. We start by proving a general result for time horizons θ_K. The proof uses a potential theoretic approach, similar to the proof of [7, Lem. 6.3].

Theorem B.5.1. *Let X^K be a birth death process with self-competition and parameters $0<d<b$ and $0<c/K$. Define $\bar{n}:=(b-d)/c$. Then there are constants $0<C_1,C_2,C_3<\infty$ such that, for any ε small and any K large enough, any initial value $0\leq |x-\lceil\bar{n}K\rceil|\leq \frac{1}{2}\left\lfloor\frac{\varepsilon K}{2}\right\rfloor$, any $m\geq 0$, and any non-negative sequence $(\theta_K)_{K\in\mathbb{N}}$,*

$$\mathbb{P}_x(\exists\ t\in[0,\theta_K]:|X^K(t)-\lceil\bar{n}K\rceil|>\varepsilon K)\leq mC_1e^{-C_2\varepsilon^2K}+\sum_{l=m}^{\infty}\left(4\left(1-e^{-C_3K\theta_K/l}\right)^{1/2}\right)^l. \tag{B.5.2}$$

Proof. We start by defining a couple of new processes based on X^K. Let

$$V^K(t):=|X^K(t)-\lceil\bar{n}K\rceil| \tag{B.5.3}$$

be the distance of X^K from its equilibrium state $\bar{n}K$ at time t. Note that this is no longer a Markov process. For V^K, let $(Y_n^K)_{n\in\mathbb{N}_0}$ be its discrete jump chain (taking values in $\mathbb{N}$, not Markovian) and $(S_n^K)_{n\in\mathbb{N}}$ its jump times.

The proof is divided into multiple steps:

1) Bound the transition probabilities of Y^K.

2) Define a discrete time Markov chain $(Z_n^K)_{n\in\mathbb{N}_0}$ such that $Z_n^K \geq Y_n^K$, for all $n \in \mathbb{N}$.
3) Derive an upper bound for the probability of Z^K hitting $\lfloor \varepsilon K \rfloor$ before 0.
4) Derive an upper bound for the probability of Z^K returning to 0 at most m times before hitting $\lfloor \varepsilon K \rfloor$.
5) Consider a continuous time version $\tilde{Z}^K$ of Z^K to deduce the final result.

Step 1: The discrete-time process Y^K changes its state due to either a birth or a death event in the original process X^K and hence moves by increments of ± 1 in each step. It is therefore a random walk on $\mathbb{N}_0$ that is reflected in 0. For the boundary case, we obtain

$$\mathbb{P}(Y_{n+1}^K = 1 | Y_n^K = 0) = 1. \tag{B.5.4}$$

For any other $1 \leq i \leq \varepsilon K$, using that $c\lceil \bar{n}K \rceil / K \in [b-d, b-d+c/K]$, we can bound

$$\begin{aligned}
&\mathbb{P}(Y_{n+1}^K = i+1 | Y_n^K = i) \\
=&\mathbb{P}(\text{birth event if } X^K = \lceil \bar{n}K \rceil + i \text{ or death event if } X^K = \lceil \bar{n}K \rceil - i) \\
\leq& \frac{b}{b+d+\frac{c}{K}(\lceil \bar{n}K \rceil + i)} \vee \frac{d + \frac{c}{K}(\lceil \bar{n}K \rceil - i)}{b+d+\frac{c}{K}(\lceil \bar{n}K \rceil - i)} \\
\leq& \frac{b}{2b+\frac{c}{K}i} \vee \frac{b-\frac{c}{K}(i-1)}{2b-\frac{c}{K}i} = \left(\frac{1}{2} - \frac{\frac{c}{2K}i}{2b+\frac{c}{K}i}\right) \vee \left(\frac{1}{2} - \frac{\frac{c}{2K}(i-2)}{2b-\frac{c}{K}i}\right) \\
\leq& \frac{1}{2} - C\frac{i}{K} =: p_+^K(i),
\end{aligned} \tag{B.5.5}$$

for some constant $C > 0$, as long as $\varepsilon \leq \bar{n}$ and hence $ci/K \leq c\varepsilon \leq b-d$.

Step 2: Define a discrete-time process $(Z_n^K)_{n\in\mathbb{N}_0}$ that is coupled to $(Y_n^K)_{n\in\mathbb{N}_0}$ by

- $Z_0^K = Y_0^K$
- Whenever $Z_n^K = Y_n^K = i$ and $Y_{n+1}^K = i+1$, we set $Z_{n+1}^K = i+1$.
- Whenever $Z_n^K = Y_n^K = i$ and $Y_{n+1}^K = i-1$, we set $Z_{n+1}^K = i+1$ with probability $(p_+^K(i) - \mathbb{P}(Y_{n+1}^K = i+1 | Y_n^K = i))/\mathbb{P}(Y_{n+1}^K = i-1 | Y_n^K = i)$ and $Z_{n+1}^K = i-1$ else.
- Whenever $Z_n^K = i > Y_n^K$, we set $Z_{n+1}^K = i+1$ with probability $p_+^K(i)$ and $Z_{n+1}^K = i-1$ else.

Then Z^K is a discrete-time Markov chain such that $Z_n^K \geq Y_n^K$, for all $n \in \mathbb{N}_0$, and

$$p^K(i,j) := \mathbb{P}(Z_{n+1}^K = j | Z_n^K = i) = \begin{cases} 1 & i = 0,\ j = 1, \\ p_+^K(i) & i \geq 1,\ j = i+1, \\ 1 - p_+^K(i) & i \geq 1,\ j = i-1, \\ 0 & \text{else.} \end{cases} \tag{B.5.6}$$

Step 3: For the Markov chain Z^K, we define the stopping times

$$\tau_j^{(Z,K)} := \inf\{n \in \mathbb{N}_0 : Z_n^K = j\}. \tag{B.5.7}$$

By standard potential theoretic arguments (see [30, Ch. 7.1.4]), we obtain, for initial values $0 \leq z \leq \lfloor \varepsilon K \rfloor$,

$$\mathbb{P}_z\left(\tau^{(Z,K)}_{\lfloor \varepsilon K \rfloor} < \tau^{(Z,K)}_0\right) = \frac{\sum_{i=1}^{z} \prod_{j=1}^{i-1} \frac{p(j,j-1)}{p(j,j+1)}}{\sum_{i=1}^{\lfloor \varepsilon K \rfloor} \prod_{j=1}^{i-1} \frac{p(j,j-1)}{p(j,j+1)}} = \frac{\sum_{i=1}^{z} \exp\left(\sum_{j=1}^{i-1} \ln\left(\frac{1+2C\frac{j}{K}}{1-2C\frac{j}{K}}\right)\right)}{\sum_{i=1}^{\lfloor \varepsilon K \rfloor} \exp\left(\sum_{j=1}^{i-1} \ln\left(\frac{1+2C\frac{j}{K}}{1-2C\frac{j}{K}}\right)\right)}. \tag{B.5.8}$$

Using that $\ln(1+\xi) = \xi + O(\xi^2)$, as $\xi \to 0$, and $j \leq \lfloor \varepsilon K \rfloor$, we can approximate, as $\varepsilon \to 0$,

$$\begin{aligned}
\ln\left(\frac{1+2C\frac{j}{K}}{1-2C\frac{j}{K}}\right) &= \ln\left(1 + \frac{4C\frac{j}{K}}{1-2C\frac{j}{K}}\right) = \frac{4C\frac{j}{K}}{1-2C\frac{j}{K}} + O\left(\left(\frac{4C\frac{j}{K}}{1-2C\frac{j}{K}}\right)^2\right) \\
&= 4C\frac{j}{K}\left(1 + \frac{2C\frac{j}{K}}{1-2C\frac{j}{K}}\right) + O\left(\left(\frac{j}{K}\right)^2\right) \\
&= 4C\frac{j}{K} + O\left(\left(\frac{j}{K}\right)^2\right) = 4C\frac{j}{K} + O(\varepsilon^2).
\end{aligned} \tag{B.5.9}$$

Plugging these asymptotics back into the above expression yields

$$\begin{aligned}
\mathbb{P}_z\left(\tau^{(Z,K)}_{\lfloor \varepsilon K \rfloor} < \tau^{(Z,K)}_0\right) &= \frac{\sum_{i=1}^{z} \exp\left(\sum_{j=1}^{i-1} 4C\frac{j}{K} + O(\varepsilon^2)\right)}{\sum_{i=1}^{\lfloor \varepsilon K \rfloor} \exp\left(\sum_{j=1}^{i-1} 4C\frac{j}{K} + O(\varepsilon^2)\right)} \\
&\leq \frac{\sum_{i=1}^{z} \exp\left(4C\frac{i(i-1)}{2K} + O((i-1)\varepsilon^2)\right)}{\sum_{i=\lfloor \frac{\varepsilon K}{2} \rfloor}^{\lfloor \varepsilon K \rfloor} \exp\left(4C\frac{i(i-1)}{2K} + O((i-1)\varepsilon^2)\right)} \\
&\leq \frac{z \exp\left(2C\frac{z^2}{K} + O(z\varepsilon^2)\right)}{\left\lfloor \frac{\varepsilon K}{2} \right\rfloor \exp\left(\frac{2C}{K}\left(\left\lfloor \frac{\varepsilon K}{2} \right\rfloor^2 - \left\lfloor \frac{\varepsilon K}{2} \right\rfloor\right) + O(K\varepsilon^3)\right)} \\
&= \frac{z}{\left\lfloor \frac{\varepsilon K}{2} \right\rfloor} \exp\left(\frac{2C}{K}\left(z^2 - \left\lfloor \frac{\varepsilon K}{2} \right\rfloor^2 + \left\lfloor \frac{\varepsilon K}{2} \right\rfloor\right) + O(z\varepsilon^2) + O(K\varepsilon^3)\right) \\
&\leq \frac{z}{\left\lfloor \frac{\varepsilon K}{2} \right\rfloor} \exp\left(\frac{2C}{K}\left(z^2 - \left\lfloor \frac{\varepsilon K}{2} \right\rfloor^2 + \left\lfloor \frac{\varepsilon K}{2} \right\rfloor + O(K^2\varepsilon^3)\right)\right) \\
&\leq \frac{1}{2} \exp\left(-\frac{2C}{K}\frac{1}{4}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor^2\right) \\
&\leq C_1 e^{-C_2 \varepsilon^2 K},
\end{aligned} \tag{B.5.10}$$

for some uniform constants $C_1, C_2 > 0$, as long as $0 \leq z \leq \frac{1}{2}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor$ and ε small enough such that $\left\lfloor \frac{\varepsilon K}{2} \right\rfloor + O(K^2\varepsilon^3) \leq \frac{1}{2}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor^2$ for large K.

Step 4: Let B^K be the random variable that describes the number of visits to 0 of Z^K before first hitting $\lfloor \varepsilon K \rfloor$ (not counting the first visit/start in case $Z_0^K = 0$). First consider

$1 \le z \le \frac{1}{2}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor$. Then, for ε small and K large enough,

$$\mathbb{P}_z\left(B^K = 0\right) = \mathbb{P}_z\left(\tau^{(Z,K)}_{\lfloor \varepsilon K \rfloor} < \tau^{(Z,K)}_0\right) \le C_1 e^{-C_2 \varepsilon^2 K} \tag{B.5.11}$$

and, for all $l \ge 1$, due to the strong Markov property,

$$\begin{aligned}\mathbb{P}_z\left(B^K = l\right) &= \mathbb{P}_z\left(\tau^{(Z,K)}_0 < \tau^{(Z,K)}_{\lfloor \varepsilon K \rfloor}\right)\left(1 - \mathbb{P}_1\left(\tau^{(Z,K)}_{\lfloor \varepsilon K \rfloor} < \tau^{(Z,K)}_0\right)\right)^{l-1} \mathbb{P}_1\left(\tau^{(Z,K)}_{\lfloor \varepsilon K \rfloor} < \tau^{(Z,K)}_0\right) \\ &\le C_1 e^{-C_2 \varepsilon^2 K}. \end{aligned} \tag{B.5.12}$$

For $z = 0$ and $l \ge 0$,

$$\mathbb{P}_0\left(B^K = l\right) = \mathbb{P}_1\left(B^K = l\right) \le C_1 e^{-C_2 \varepsilon^2 K}.$$

Hence, for $0 \le z \le \frac{1}{2}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor$, ε small and K large enough, and any $m \ge 0$,

$$\mathbb{P}_z\left(B^K \le m\right) \le (m+1) C_1 e^{-C_2 \varepsilon^2 K}.$$

Step 5: Finally, let $\tilde{Z}^K$ be the continuous time process that has $(Z^K_n)_{n \in \mathbb{N}_0}$ as a jump chain and the same jump times $(S^K_n)_{n \in \mathbb{N}}$ as the original process X^K. By the above construction, for all $t \ge 0$, we have

$$\tilde{Z}^K(t) \ge V^K(t) = |X^K(t) - \lceil \bar{n} K \rceil|.$$

We can therefore deduce that, for ε small and K large enough, initial value $x = \lceil \bar{n} K \rceil \pm z$ such that $0 \le z \le \frac{1}{2}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor$, and any $m \ge 0$,

$$\begin{aligned} &\mathbb{P}_x(\exists\, t \in [0, \theta_K] : |X^K(t) - \lceil \bar{n} K \rceil| > \varepsilon K) \\ \le\, &\mathbb{P}_z(\exists\, t \in [0, \theta_K] : \tilde{Z}^K(t) > \varepsilon K) \\ =\, &\sum_{l=0}^{\infty} \mathbb{P}_z(\exists\, t \in [0, \theta_K] : \tilde{Z}^K(t) > \varepsilon K, B^K = l) \\ \le\, &\mathbb{P}_z(B^K \le m-1) + \sum_{l=m}^{\infty} \mathbb{P}_z(\exists\, t \in [0, \theta_K] : \tilde{Z}^K(t) > \varepsilon K, B^K = l) \end{aligned} \tag{B.5.13}$$

Now let $(I^K_j)_{j \ge 1}$ be the times in between visits to 0 of $\tilde{Z}^K$, i.e., for $j \ge 1$,

$$I^K_j := \inf\left\{ t > 0 : \exists\, 0 < s < t : \tilde{Z}^K_{s + \sum_{i=1}^{j-1} I^K_i} \ne 0, \tilde{Z}^K_{t + \sum_{i=1}^{j-1} I^K_i} = 0 \right\}. \tag{B.5.14}$$

Then, since each return takes at least the time of a single jump in the original Markov chain X^K, as long as $\tilde{Z}^K$ does not surpass $2\bar{n}K > \lceil \bar{n} K \rceil + \varepsilon K$, there are independent identically distributed exponential random variables $(E^K_j)_{j \ge 1}$ with parameter $4\bar{n}K(b + d + c\bar{n}) =: \bar{C}K$ such that, for each $a \in \mathbb{R}$,

$$\mathbb{P}(I^K_j < a) \le \mathbb{P}(E^K_j < a) = (1 - e^{-a\bar{C}K})_+. \tag{B.5.15}$$

To bound the probabilities $\mathbb{P}_z(\exists\ t \in [0, \theta_K] : \tilde{Z}^K(t) > \varepsilon K, B^K = l)$, we argue as follows: If there were at least $l/2$ occurrences of $I_j^K \geq 2\theta_K/l$ (out of the l times between visits to 0), the time until $\tilde{Z}^K$ first surpasses εK could be bounded from below by

$$\sum_{j=1}^{l} I_j^K \geq \frac{l}{2}\frac{2\theta_K}{l} = \theta_K. \tag{B.5.16}$$

Hence, by contradiction we can bound

$$\begin{aligned}
&\mathbb{P}_z(\exists\ t \in [0, \theta_K] : \tilde{Z}^K(t) > \varepsilon K, B^K = l) \leq \mathbb{P}_z\left(\sum_{j=1}^{l} \mathbb{1}_{I_j^K < 2\theta_K/l} > \frac{l}{2}, B^K = l\right)\\
&\leq \mathbb{P}_z\Bigg(\underbrace{\sum_{j=1}^{l} \mathbb{1}_{E_j^K < 2\theta_K/l}}_{\sim \mathrm{Bin}\left(l, 1-e^{-\bar{C}K2\theta_K/l}\right)} > \frac{l}{2}\Bigg) = \sum_{j=\lceil l/2 \rceil}^{l} \binom{l}{j} \left(1 - e^{-\bar{C}K2\theta_K/l}\right)^j \left(e^{-\bar{C}K2\theta_K/l}\right)^{l-j}\\
&\leq \frac{l}{2} 2^l \left(1 - e^{-\bar{C}K2\theta_K/l}\right)^{l/2} \leq \left(4\left(1 - e^{-\bar{C}K2\theta_K/l}\right)^{1/2}\right)^l,
\end{aligned} \tag{B.5.17}$$

where we used that $\binom{l}{j} \leq 2^l$ and $l/2 \leq 2^l$.

Combining this with step 4 yields

$$\begin{aligned}
&\mathbb{P}_x(\exists\ t \in [0, \theta_K] : |X^K(t) - \lceil \bar{n}K \rceil| > \varepsilon K)\\
&\leq \mathbb{P}_z(B^K \leq m-1) + \sum_{l=m}^{\infty} \mathbb{P}_z(\exists\ t \in [0, \theta_K] : \tilde{Z}^K(t) > \varepsilon K, B^K = l)\\
&\leq mC_1 e^{-C_2\varepsilon^2 K} + \sum_{l=m}^{\infty} \left(4\left(1 - e^{-\bar{C}K2\theta_K/l}\right)^{1/2}\right)^l.
\end{aligned} \tag{B.5.18}$$

This concludes the proof with $C_3 = 2\bar{C}$. ☐

From this general theorem, we can now derive the result necessary for the proof in Section B.4.1, considering time horizons of size $\theta_K = \ln K$ as an upper bound for phases with length of order λ_K and bounding the probability of diverging from the equilibrium in $o(\lambda_K/\ln K)$ to concatenate $O(\ln K/\lambda_K)$ many phases.

Theorem B.5.2. *Let X^K be a birth death process with self-competition and parameters $0 < d < b$ and $0 < c/K$. Define $\bar{n} := (b-d)/c$ and let $1 \ll \lambda_K \ll \ln K$ as $K \to \infty$. Then, for ε small enough and any sequence of initial values $0 \leq |x^K - \lceil \bar{n}K \rceil| \leq \frac{1}{2}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor$,*

$$\lim_{K\to\infty} \frac{\ln K}{\lambda_K} \cdot \mathbb{P}_{x^K}(\exists\ t \in [0, \ln K] : |X^K(t) - \lceil \bar{n}K \rceil| > \varepsilon K) = 0.$$

Proof. We apply Theorem B.5.1 with $\theta_K = \ln K$ and $m^K = K^2$ to obtain that, for ε small, K large enough and $0 \le |x^K - \lceil \bar{n} K \rceil| \le \frac{1}{2} \left\lfloor \frac{\varepsilon K}{2} \right\rfloor$,

$$\begin{aligned}
&\mathbb{P}_{x^K}(\exists\ t \in [0, \ln K] : |X^K(t) - \lceil \bar{n} K \rceil| > \varepsilon K) \\
&\le K^2 C_1 e^{-C_2 \varepsilon^2 K} + \sum_{l=K^2}^{\infty} \left(4 \left(1 - e^{-C_3 K \ln K / l}\right)^{1/2} \right)^l \\
&\le K^2 C_1 e^{-C_2 \varepsilon^2 K} + \sum_{l=K^2}^{\infty} \Big(\underbrace{4 \left(1 - e^{-C_3 K \ln K / K^2}\right)^{1/2}}_{<1 \text{ for } K \text{ large enough}} \Big)^l \\
&\le K^2 C_1 e^{-C_2 \varepsilon^2 K} + \left(4^2 \left(1 - e^{-C_3 K \ln K / K^2}\right) \right)^{K^2/2} \underbrace{\frac{1}{1 - 4\left(1 - e^{-C_3 K \ln K / K^2}\right)^{1/2}}}_{\le C_4 < \infty \text{ for } K \text{ large enough}} \\
&\le K^2 C_1 e^{-C_2 \varepsilon^2 K} + C_4 \Big(\underbrace{16 C_3 \frac{\ln K}{K}}_{<1 \text{ for } K \text{ large enough}} \Big)^{K^2/2} \\
&\le K^2 C_1 e^{-C_2 \varepsilon^2 K} + C_4 16 C_3 \frac{\ln K}{K}.
\end{aligned} \tag{B.5.19}$$

Now, for fixed ε^2,

$$e^{-C_2 \varepsilon^2 K} \ll \frac{\lambda_K}{K^2 \ln K} \;\Leftrightarrow\; K^2 e^{-C_2 \varepsilon^2 K} \ll \frac{\lambda_K}{\ln K} \tag{B.5.20}$$

and $1 \ll \lambda_K$ implies

$$\frac{(\ln K)^2}{K} \ll \lambda_K \;\Leftrightarrow\; \frac{\ln K}{K} \ll \frac{\lambda_K}{\ln K}. \tag{B.5.21}$$

Hence we obtain that, for fixed $\varepsilon > 0$,

$$\mathbb{P}_{x^K}(\exists\ t \in [0, \ln K] : |X^K(t) - \lceil \bar{n} K \rceil| > \varepsilon K) = O(K^2 e^{-C_2 \varepsilon^2 K}) + O(\ln K / K) = o(\lambda_K / \ln K), \tag{B.5.22}$$

which yields the desired result. □

B.5.2. Convergence to the deterministic system

We now provide a result on the convergence of stochastic birth death processes with competition to the corresponding deterministic system, for finite time horizons. The proof is similar to the one of [75, Ch. 11, Thm. 2.1]. Instead of almost sure convergence, we derive convergence in probability, but are able to quantify the convergence speed in return, which again allows us to concatenate the result for infinitely many phases in Section B.4.1.

Theorem B.5.3. *Let X^K be a birth death process with self-competition and parameters $0 < d < b$ and $0 < c/K$. Assume that $X^K(0)/K \to x_0$ and $1 \ll \lambda_K \ll \ln K$, as $K \to \infty$, and let $(x(t))_{t \geq 0}$ be the solution of the ordinary differential equation*

$$\dot{x}(t) = x(t)\,[b - d - c \cdot x(t)] \tag{B.5.23}$$

with initial value $x(0) = x_0$. Then, for every $0 \leq T < \infty$ and $\varepsilon > 0$,

$$\lim_{K\to\infty} \frac{\ln K}{\lambda_K} \cdot \mathbb{P}\left(\sup_{t \leq T} \left| \frac{X^K(t)}{K} - x(t) \right| > \varepsilon \right) = 0. \tag{B.5.24}$$

Proof. We start be writing X^K in terms of independent standard Poisson processes Y_b and Y_d,

$$X^K(t) \overset{(d)}{=} X^K(0) + Y_b\left(K \int_0^t b\frac{X^K(s)}{K}\,\mathrm{d}s \right) - Y_d\left(K\int_0^t d\frac{X^K(s)}{K} + c\left(\frac{X^K(s)}{K}\right)^2 \mathrm{d}s \right). \tag{B.5.25}$$

Note that we only have equality in distribution here, since we choose Y_b and Y_d uniformly across different values of K and the X^K stand in no specific relation to each other. We will omit this from the notation for the remainder of the proof, as we are only proving convergence in probability and equality in distribution is therefore sufficient.

Setting $\tilde{Y}_b(u) := Y_b(u) - u$ and $\tilde{Y}_d := Y_d(u) - u$ (i.e. centering the Poisson processes at their expectations), this representation yields

$$\begin{aligned} \frac{X^K(t)}{K} = \frac{X^K(0)}{K} &+ \int_0^t \left(b - d - c\frac{X^K(s)}{K} \right) \frac{X^K(s)}{K}\,\mathrm{d}s \\ &+ \frac{1}{K}\tilde{Y}_b\left(K\int_0^t b\frac{X^K(s)}{K}\,\mathrm{d}s \right) - \frac{1}{K}\tilde{Y}_d\left(K\int_0^t d\frac{X^K(s)}{K} + c\left(\frac{X^K(s)}{K}\right)^2 \mathrm{d}s \right). \end{aligned} \tag{B.5.26}$$

Now we introduce the stopping time

$$\tau^K := \inf\left\{ t \geq 0 : \frac{X^K(t)}{K} \neq [0, M] \right\}, \tag{B.5.27}$$

for some large M (e.g. $M \geq 2(\bar{n} \vee x_0)$, where $\bar{n} = (b-d)/c$). Up to time τ^K, the population size of our process is bounded by KM and, using the integral form of (B.5.23), we deduce

$$\begin{aligned}
&\left|\frac{X^K(t\wedge\tau^K)}{K} - x(t\wedge\tau^K)\right| \\
\le\ &\left|\frac{X^K(0)}{K} - x(0)\right| + \int_0^{t\wedge\tau^K} (b+d)\left|\frac{X^K(s)}{K} - x(s)\right| + c\underbrace{\left|\left(\frac{X^K(s)}{K}\right)^2 - x(s)^2\right|}_{\le 2M\left|\frac{X^K(s)}{K}-x(s)\right|}\mathrm{d}s \\
&\qquad + \frac{1}{K}\left|\tilde{Y}_b\left(K\int_0^{t\wedge\tau^K} b\frac{X^K(s)}{K}\mathrm{d}s\right)\right| \\
&\qquad + \frac{1}{K}\left|\tilde{Y}_d\left(K\int_0^{t\wedge\tau^K} d\frac{X^K(s)}{K} + c\left(\frac{X^K(s)}{K}\right)^2\mathrm{d}s\right)\right| \\
\le\ &\left|\frac{X^K(0)}{K} - x(0)\right| + \int_0^t C\left|\frac{X^K(s\wedge\tau^K)}{K} - x(s\wedge\tau^K)\right|\mathrm{d}s \\
&\qquad + \frac{1}{K}\sup_{u\in[0,t\wedge\tau^K]}\left|\tilde{Y}_b\left(KbMu\right)\right| + \frac{1}{K}\sup_{u\in[0,t\wedge\tau^K]}\left|\tilde{Y}_d\left(K(dMu + cM^2u)\right)\right|.
\end{aligned} \tag{B.5.28}$$

Here, $C := b + d + 2Mc$ and in the last line we used that, even though the centred Poisson processes can take positive and negative values, we can bound their absolute value by considering the supremum over all possible rates.

Next, Gronwall's inequality implies that

$$\begin{aligned}
&\left|\frac{X^K(t\wedge\tau^K)}{K} - x(t\wedge\tau^K)\right| \\
\le\ &\left[\left|\frac{X^K(0)}{K} - x(0)\right| + \frac{1}{K}\left(\sup_{u\in[0,t]}\left|\tilde{Y}_b\left(KbMu\right)\right| + \sup_{u\in[0,t]}\left|\tilde{Y}_d\left(K(dMu + cM^2u)\right)\right|\right)\right]e^{Ct}.
\end{aligned} \tag{B.5.29}$$

Now fix a $T \ge 0$. With probability 1, for some large enough K_0 and all $K \ge K_0$, $|X^K(0) - x(0)| \le e^{-CT}\varepsilon/2$.

Hence, for $\varepsilon < M/2$ and $K \geq K_0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\leq T}\left|\frac{X^K(t)}{K} - x(t)\right| \leq \varepsilon\right) \\
=&\,\mathbb{P}\left(\sup_{t\leq T}\left|\frac{X^K(t)}{K} - x(t)\right| \leq \varepsilon, \tau^K > T\right) \\
=&\,\mathbb{P}\left(\sup_{t\leq T}\left|\frac{X^K(t\wedge\tau^K)}{K} - x(t\wedge\tau^K)\right| \leq \varepsilon\right) \\
\geq&\,\mathbb{P}\left(\left[\left|\frac{X^K(0)}{K} - x(0)\right| + \frac{1}{K}\left(\sup_{u\in[0,T]}\left|\tilde{Y}_b\left(KbMu\right)\right| + \sup_{u\in[0,T]}\left|\tilde{Y}_d\left(K(dMu + cM^2u)\right)\right|\right)\right]e^{CT} \leq \varepsilon\right) \\
\geq&\,\mathbb{P}\left(\frac{1}{K}\left(\sup_{u\in[0,T]}\left|\tilde{Y}_b\left(KbMu\right)\right| + \sup_{u\in[0,T]}\left|\tilde{Y}_d\left(K(dMu + cM^2u)\right)\right|\right)e^{CT} \leq \frac{\varepsilon}{2}\right) \\
\geq&\,1 - \mathbb{P}\left(\sup_{u\in[0,T]}\left|\tilde{Y}_b\left(KbMu\right)\right| + \sup_{u\in[0,T]}\left|\tilde{Y}_d\left(K(dMu + cM^2u)\right)\right| > \frac{Ke^{-CT}\varepsilon}{2}\right) \\
\geq&\,1 - \mathbb{P}\left(\sup_{u\in[0,T]}\left|\tilde{Y}_b\left(KbMu\right)\right| > \frac{Ke^{-CT}\varepsilon}{4}\right) - \mathbb{P}\left(\sup_{u\in[0,T]}\left|\tilde{Y}_d\left(K(dMu + cM^2u)\right)\right| > \frac{Ke^{-CT}\varepsilon}{4}\right).
\end{aligned} \tag{B.5.30}$$

To finish the proof, it is left to show that the two probabilities in the last line are of order $o(\lambda_K/\ln K)$. We run through the calculation for $\tilde{Y}_b$, the other summand works equivalently.

Since $\tilde{Y}_b$ is a martingale, $|\tilde{Y}_b|$ is a submartingale. We set $\tilde{T} = bMT$ and $\tilde{\varepsilon} = e^{-CT}\varepsilon/4$. Then, using Doob's maximum inequality [30, Thm. 3.87], we obtain

$$\begin{aligned}
&\mathbb{P}\left(\sup_{u\in[0,T]}\left|\tilde{Y}_b\left(KbMu\right)\right| > \frac{Ke^{-CT}\varepsilon}{4}\right) = \mathbb{P}\left(\sup_{u\in[0,\tilde{T}]}\left|\tilde{Y}_b(Ku)\right|^2 > (K\tilde{\varepsilon})^2\right) \\
\leq&\,\frac{1}{(K\tilde{\varepsilon})^2}\mathbb{E}\left(\left|\tilde{Y}_b(K\tilde{T})\right|^2\right) = \frac{1}{(K\tilde{\varepsilon})^2}\mathbb{E}\left(\langle\tilde{Y}_b(K\tilde{T})\rangle\right) \\
=&\,\frac{K\tilde{T}}{(K\tilde{\varepsilon})^2} = \frac{\tilde{T}}{K\tilde{\varepsilon}^2} = o\left(\frac{\lambda_K}{\ln K}\right).
\end{aligned} \tag{B.5.31}$$

This concludes the proof. □

B.6. Branching processes at varying rates

In this chapter, we collect some technical results for birth death processes with time-dependent rates. These are used to approximate the micro- and mesoscopic populations in the proof of the main result of this paper in Section B.4.2. In Section B.6.1 we focus on pure birth death processes and then add on the effects of immigration in Section B.6.2. We build on the results of [45] and work out the averaging effects of growth in a periodically changing environment.

The particular form of time-dependent parameters in this chapter depicts two different effects. Firstly, the system parameters jump periodically between finitely many values on the divergent time-scale λ_K. Secondly, after each parameter change, the macroscopic subpopulation restabilizes at the corresponding equilibrium, which takes a finite time (independent of K). During this short re-equilibration time, we only have weaker estimates on the effective parameters for the growth of the micro- and mesoscopic subpopulations. However, part of the following results is that the general behaviour of the branching processes is not effected significantly by these short phases.

B.6.1. Pure birth death processes

Let us first consider pure birth death processes with time-dependent rates. As before, take $1 \ll \lambda_K \ll \ln K$ and $\ell \in \mathbb{N}$. Let $b_i, d_i, b_{i,*}, d_{i,*}, > 0$, $T_i > 0$ and $T_{*,i} \geq 0$, for $i = 1, \ldots, \ell$. Writing $T_i^\Sigma := \sum_{j=1}^{i} T_j$, we define the rate functions for birth and death to be the periodic extensions of

$$\begin{aligned} b^K(t) &= \begin{cases} b_{*,i} & : t \in [T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_{*,i}), \\ b_i & : t \in [T_{i-1}^\Sigma \lambda_K + T_{*,i}, T_i^\Sigma \lambda_K), \end{cases} \\ d^K(t) &= \begin{cases} d_{*,i} & : t \in [T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_{*,i}), \\ d_i & : t \in [T_{i-1}^\Sigma \lambda_K + T_{*,i}, T_i^\Sigma \lambda_K). \end{cases} \end{aligned} \tag{B.6.1}$$

Moreover, for $i = 1, \ldots, \ell$, we write $r_i := b_i - d_i$ and $r^K(t) := b^K(t) - d^K(t)$ to refer to the net growth rate. Finally, we define the average growth rate by $r_{\mathrm{av}} := (\sum_{i=1}^{\ell} r_i T_i)/T_\ell^\Sigma$.

We analyse the processes $\left(Z_t^K\right)_{t\geq 0}$, which are Markov processes with $Z_0^K = \left\lfloor K^\beta - 1 \right\rfloor$, for some $\beta \geq 0$, and with generators

$$\left(\mathcal{L}_t^K f\right)(n) = b^K(t) n \left(f(n+1) - f(n)\right) + d^K(t) n \left(f(n-1) - f(n)\right), \tag{B.6.2}$$

acting on all bounded functions $f : \mathbb{N}_0 \to \mathbb{R}$. We refer to the law of these processes by $Z^K \sim \mathrm{BD}\left(b^K, d^K, \beta\right)$.

Our aim is to show that, under logarithmic rescaling of time and size, such population processes grow (or shrink) according to their average net growth rate. Note that the process becomes trivial if $\beta = 0$. We therefore exclude this case in the entire section without further announcement.

Theorem B.6.1. *Let Z^K follow the law of* $\mathrm{BD}\left(b^K, d^K, \beta\right)$*, where $\beta > 0$. Then, for all fixed $T \in (0, \infty)$, the following convergence holds in probability, with respect to the $L^\infty([0,T])$ norm,*

$$\left(\frac{\ln\left(Z_{s \ln K}^K + 1\right)}{\ln K}\right)_{s\in[0,T]} \overset{K\to\infty}{\longrightarrow} \left((\beta + r_{\mathrm{av}} s) \vee 0\right)_{s\in[0,T]}. \tag{B.6.3}$$

The rest of this section is dedicated to the proof of this theorem and we split up the claim into several lemmas.

Remark 21. To avoid complicated notation, we only conduct the proofs for the case of $\ell = 2$ and $T_{*,1} = 0$. The general case is proven analogously and there is no deeper insight or additional difficulty to it. This choice allows us to use the shorthand notation $b_* := b_{*,2}$, $d_* := d_{*,2}$, and $T_* := T_{*,2}$, which leads to the rate functions taking the form

$$b^K(t) = \begin{cases} b_1 & t \in [0, \lambda_K T_1), \\ b_* & t \in [\lambda_K T_1, \lambda_K T_1 + T_*), \\ b_2 & t \in [\lambda_K T_1 + T_*, \lambda_K (T_1 + T_2)), \end{cases} \qquad d^K(t) = \begin{cases} d_1 & t \in [0, \lambda_K T_1), \\ d_* & t \in [\lambda_K T_1, \lambda_K T_1 + T_*), \\ d_2 & t \in [\lambda_K T_1 + T_*, \lambda_K (T_1 + T_2)), \end{cases} \tag{B.6.4}$$

with periodic extension.

We start by stating an explicit representation of the processes in terms of Poisson measures and derive the corresponding Doob's martingale decomposition.

Poisson representation: Let $Q^{(b)}(\mathrm{d}s, \mathrm{d}\theta)$ and $Q^{(d)}(\mathrm{d}s, \mathrm{d}\theta)$ be independent homogenous Poisson random measures on $(\mathbb{R}^2_{\geq 0}, \mathrm{d}s, \mathrm{d}\theta)$ and denote by $\tilde{Q}^{(*)} = Q^{(*)} - \mathrm{d}s\mathrm{d}\theta$, for $* \in \{b, d\}$, their normalized versions. Then we can represent Z^K as

$$Z^K_t = Z^K_0 + \int_0^t \int_{\mathbb{R}_{\geq 0}} \mathbb{1}_{\theta \leq b^K(s^-) Z^K_{s^-}} Q^{(b)}(\mathrm{d}s, \mathrm{d}\theta) - \int_0^t \int_{\mathbb{R}_{\geq 0}} \mathbb{1}_{\theta \leq d^K(s^-) Z^K_{s^-}} Q^{(d)}(\mathrm{d}s, \mathrm{d}\theta). \tag{B.6.5}$$

In particular we have the martingale decomposition $Z^K_t = Z^K_0 + M^K_t + A^K_t$, where

$$M^K_t = \int_0^t \int_{\mathbb{R}_{\geq 0}} \mathbb{1}_{\theta \leq b^K(s^-) Z^K_{s^-}} \tilde{Q}^{(b)}(\mathrm{d}s, \mathrm{d}\theta) - \int_0^t \int_{\mathbb{R}_{\geq 0}} \mathbb{1}_{\theta \leq d^K(s^-) Z^K_{s^-}} \tilde{Q}^{(d)}(\mathrm{d}s, \mathrm{d}\theta). \tag{B.6.6}$$

and

$$A^K_t = \int_0^t (b^K(s) - d^K(s)) Z^K_s \mathrm{d}s. \tag{B.6.7}$$

In terms of Itô's calculus this implies $\mathrm{d}Z^K_t = \mathrm{d}M^K_t + r^K(t) Z^K_t \mathrm{d}t$. Therefore, we directly obtain the bracket process

$$\langle M^K \rangle_t = \int_0^t (b^K(s) + d^K(s)) Z^K_s \mathrm{d}s. \tag{B.6.8}$$

Towards proving Theorem B.6.1, we first determine the expected value of the process and check that it satisfies the desired convergence.

Lemma B.6.2. *Let Z^K follow the law of* $\mathrm{BD}\left(b^K, d^K, \beta\right)$. *Then*

$$\mathbb{E}\left[Z^K_t\right] = \left\lfloor K^\beta - 1 \right\rfloor \mathrm{e}^{g^K(t)}, \qquad \textit{where } g^K(t) = \int_0^t r^K(s) \mathrm{d}s. \tag{B.6.9}$$

Proof. Using the martingale decomposition and (B.6.7) we obtain the integral equation

$$\mathbb{E}\left[Z_t^K\right] = \mathbb{E}\left[Z_0^K\right] + \mathbb{E}\left[A_t^K\right] = \left\lfloor K^\beta - 1\right\rfloor + \int_0^t r^K(s)\mathbb{E}\left[Z_s^K\right] \mathrm{d}s. \tag{B.6.10}$$

Due to existence and uniqueness of the solution to this integral equation, this directly gives the claim. □

Lemma B.6.3. *For all fixed $T < \infty$ we have the uniform convergence*

$$\sup_{s \le T} \left| \frac{\ln\left(K^\beta \mathrm{e}^{g^K(s \ln K)}\right)}{\ln K} - (\beta + r_{\mathrm{av}} s)\right| \overset{K\to\infty}{\longrightarrow} 0. \tag{B.6.11}$$

Proof. We slice the whole time span $[0, T\ln K]$ into equal pieces of order λ_K. Since r^K has a period of length $\lambda_K(T_1 + T_2)$ and is piecewise constant, we obtain, for all $n \in \mathbb{N}$,

$$\begin{aligned} g^K(n\lambda_K(T_1+T_2)) &= n[\lambda_K T_1 r_1 + T_* r_* + (\lambda_K T_2 - T_*) r_2] \\ &= n[\lambda_K(T_1 r_1 + T_2 r_2) + T_*(r_* - r_2)]. \end{aligned} \tag{B.6.12}$$

Moreover, within such a period the growth of g^K is linearly bounded, i.e. for all $0 \le s \le t$ such that $t - s \le \lambda_K(T_1 + T_2)$, one has

$$\left|g^K(t) - g^K(s)\right| \le (t-s)\max\{r_1, r_*, r_2\} \le C\lambda_K, \tag{B.6.13}$$

for some uniform finite constant C. Hence, we can estimate (in the case of $r_{\mathrm{av}} > 0$),

$$\begin{aligned} g^K(s\ln K) &\ge g^K\left(\left\lfloor \frac{s\ln K}{\lambda_K(T_1+T_2)}\right\rfloor \lambda_K(T_1+T_2)\right) - C\lambda_K \\ &= \left\lfloor \frac{s\ln K}{\lambda_K(T_1+T_2)}\right\rfloor [\lambda_K(T_1 r_1 + T_2 r_2) + T_*(r_* - r_2)] - C\lambda_K \\ &\ge r_{\mathrm{av}} s \ln K + s\frac{\ln K}{\lambda_K}\frac{T_*(r_* - r_2)}{T_1 + T_2} - 2C\lambda_K, \end{aligned} \tag{B.6.14}$$

where we use in the last inequality that the term in the brackets is positive for K large enough. Similarly, we obtain

$$g^K(s\ln K) \le r_{\mathrm{av}} s\ln K + s\frac{\ln K}{\lambda_K}\frac{T_*(r_* - r_2)}{T_1+T_2} + 2C\lambda_K. \tag{B.6.15}$$

Both estimates can be achieved in the same way for $r_{\mathrm{av}} < 0$. With this at hand, the claim can be shown directly. □

Our next aim is to study the deviation of the original process Z^K from its expected value.

Lemma B.6.4. *For all $T < \infty$ and all $\delta > 0$, there exists a $K_0 \in \mathbb{N}$ such that, for all $K \ge K_0$ and all $\eta \in (0, \beta)$, it holds that*

$$\mathbb{P}\left(\sup_{t \le T\ln K}\left|\mathrm{e}^{-g^K(t)} Z_t^K - K^\beta\right| \ge K^\eta\right) \le (\bar{b} + \bar{d})K^{\beta+\delta-2\eta}\frac{1}{r_{\mathrm{av}}}\left(1 - K^{-r_{\mathrm{av}}T}\right), \tag{B.6.16}$$

where $\bar{b} := \max_{t\in[0,T_\ell^\Sigma\lambda_K]} b^K(t)$ and $\bar{d} := \max_{t\in[0,T_\ell^\Sigma\lambda_K]} d^K(t)$.

Proof. The main idea is to use Doob's maximum inequality for a rescaled martingale. We introduce the process

$$\hat{M}_t^K := \int_0^t \mathrm{e}^{-g^K(s)} \mathrm{d}M_s^K, \tag{B.6.17}$$

which is a martingale since M^K is a martingale. Following the techniques of proof step 1 of [45, Lem. A.1] using Itô's isometry, Itô's formula and Doob's maximum inequality, we get

$$\begin{aligned}\mathbb{P}\left(\sup_{t\leq T\ln K}\left|\mathrm{e}^{-g^K(t)}Z_t^K - K^\beta\right| \geq K^\eta\right) &\leq \left(\bar{b}+\bar{d}\right)K^{\beta-2\eta}\int_0^{T\ln K}\mathrm{e}^{-g^K(t)}\mathrm{d}t \\ &\leq \left(\bar{b}+\bar{d}\right)K^{\beta-2\eta}\int_0^{T\ln K}\mathrm{e}^{-r_{\mathrm{av}}t}K^\delta\mathrm{d}t \\ &= \left(\bar{b}+\bar{d}\right)K^{\beta+\delta-2\eta}\frac{1}{r_{\mathrm{av}}}\left(1-K^{-r_{\mathrm{av}}T}\right). \end{aligned} \tag{B.6.18}$$

Here, for the last inequality, K has to be choosen large enough, such that

$$\frac{g^K(t)}{\ln K} \geq r_{\mathrm{av}}\frac{t}{\ln K} - \delta, \qquad \forall t \in [0, T\ln K], \tag{B.6.19}$$

which is possible by Lemma B.6.3. □

With the above results, we are now able to prove the desired convergence for populations that tend to grow (i.e. for $r_{\mathrm{av}} > 0$).

Lemma B.6.5. *Let* Z^K *follow the law of* $\mathrm{BD}\left(b^K, d^K, \beta\right)$ *and assume that* $r_{\mathrm{av}} > 0$*, then the convergence of Theorem B.6.1 holds true.*

Proof. Fix $T < \infty$ and choose $\delta = \beta/4$ and $\eta = 3\beta/4$. Moreover, define the set

$$\Omega_1^K := \left\{\sup_{t\leq T\ln K}\left|\mathrm{e}^{-g^K(t)}Z_t^K - K^\beta\right| \leq K^\eta\right\}. \tag{B.6.20}$$

Then $\lim_{K\to\infty}\mathbb{P}\left(\Omega_1^K\right) = 1$ by Lemma B.6.4 since $\beta+\delta-2\eta < 0$ and an analogous computation to proof step 2 of [45, Lem. A.1] together with Lemma B.6.3 and the triangle inequality yield the claim. □

In order to study birth death processes with tendency to shrink (i.e. with $r_{\mathrm{av}} < 0$), we have to take care of the extinction event. Let us point out that in our situation of the changing environment, this is a little more involved and we cannot use the results for time-homogenoeus branching processes. To this end, we first determine the probability generating function of general birth death processes with piecewise constant rates, which we then use to establish bounds on the distribution function of the extinction time.

Lemma B.6.6 (generating function)**.** *For $\ell \in \mathbb{N}$, let $b_i, d_i, T_i > 0$, $1 \leq i \leq \ell$, and write $T := \sum_{i=1}^{\ell} T_i$ and $r_i := b_i - d_i$. We consider the birth death process $(X_t)_{t\in[0,T]}$ with initial value $X_0 = 1$ that is driven by the birth and death rates b_i, d_i on $\left[\sum_{j=1}^{i-1} T_j, \sum_{j=1}^{i} T_j\right)$. Then the (probability) generating function g of X_T is given by*

$$\begin{aligned} g(s) &:= \mathbb{E}\left[s^{X_T}\right] && \text{(B.6.21)} \\ &= 1 - \frac{\mathrm{e}^{r_1T_1+\cdots+r_\ell T_\ell}}{\frac{b_1}{r_1}\left(\mathrm{e}^{r_1T_1}-1\right)\mathrm{e}^{r_2T_2+\cdots+r_\ell T_\ell} + \frac{b_2}{r_2}\left(\mathrm{e}^{r_2T_2}-1\right)\mathrm{e}^{r_3T_3+\cdots+r_\ell T_\ell} + \cdots + \frac{b_\ell}{r_\ell}\left(\mathrm{e}^{r_\ell T_\ell}-1\right) - \frac{1}{s-1}}. \end{aligned}$$

Proof. For homogeneous birth death processes $(Y_t)_{t\geq 0}$ with constant rates b and d and initial value $Y_0 = 1$, the probability generating function at time $t > 0$ is given by

$$\begin{aligned} F(s,t) := \mathbb{E}\left[s^{Y_t} | Y_0 = 1\right] &= \frac{d(s-1) - \mathrm{e}^{-rt}(bs-d)}{b(s-1) - \mathrm{e}^{-rt}(bs-d)} \\ &= 1 - \frac{\mathrm{e}^{rt}}{\frac{b}{r}\left(\mathrm{e}^{rt}-1\right) - \frac{1}{s-1}}. && \text{(B.6.22)} \end{aligned}$$

Due to independence, for initial values $Y_0 = k \in \mathbb{N}$ we obtain $\mathbb{E}\left[s^{Y_t} | Y_0 = k\right] = F(s,t)^k$.

Let us first consider the case $\ell = 2$. Studying X at time $T_1 + T_2$, we can interpret it as a birth death process on $[T_1, T_1 + T_2]$ with rates b_2 and d_2, initialized with X_{T_1} individuals. Letting F_1 and F_2 be the generating functions corresponding to the parameters of the two phases (again assuming initial values of 1), this leads to

$$\begin{aligned} \mathbb{E}\left[s^{X_{T_1+T_2}}\right] = \mathbb{E}\left[\mathbb{E}\left[s^{X_{T_1+T_2}} | X_{T_1}\right]\right] &= \mathbb{E}\left[F_2(s,T_2)^{X_{T_1}}\right] = F_1(F_2(s,T_2),T_1) \\ &= 1 - \frac{\mathrm{e}^{r_1T_1}}{\frac{b_1}{r_1}\left(\mathrm{e}^{r_1T_1}-1\right) + \frac{(b_2/r_2)(\mathrm{e}^{r_2T_2}-1)-1/(s-1)}{\mathrm{e}^{r_2T_2}}} \\ &= 1 - \frac{\mathrm{e}^{r_1T_1+r_2T_2}}{\frac{b_1}{r_1}\left(\mathrm{e}^{r_1T_1}-1\right)\mathrm{e}^{r_2T_2} + \frac{b_2}{r_2}\left(\mathrm{e}^{r_2T_2}-1\right) - \frac{1}{s-1}} && \text{(B.6.23)} \end{aligned}$$

The claim for larger $\ell \in \mathbb{N}$ follows by induction. □

With this preparation, we can now prove the following helpful bounds for the extinction time.

Lemma B.6.7 (extinction time)**.** *Let Z^K be the birth death process with varying rates defined above. Denote by $T_{\mathrm{ext}}^{Z^K} := \inf\left\{t \geq 0 : Z_t^K = 0\right\}$ the extinction time of Z^K. Then, for all $\delta, \delta_1, \delta_2 > 0$ and all $M > (T_1|r_1| + T_2|r_2|)/(T_1+T_2)$ there exists a $K_0 < \infty$ such that, for all $K \geq K_0$ and all $t \geq 0$,*

$$\mathbb{P}\left(T_{\mathrm{ext}}^{Z^K} > t\ln K | Z_0^K = 1\right) \leq \exp\left[\left(\left(r_{\mathrm{av}} + \delta_1\right)t + \delta_2\right)\ln K\right], \qquad \text{(B.6.24)}$$

$$\mathbb{P}\left(T_{\mathrm{ext}}^{Z^K} > t\ln K | Z_0^K = 1\right) \geq \exp\left[-\left(Mt + \delta\right)\ln K\right]. \qquad \text{(B.6.25)}$$

Proof. Let us first consider a time discretisation $\left(Y_n^K\right)_{n\in\mathbb{N}_0}$ of Z^K, namely $Y_n^K := Z_{n\lambda_K(T_1+T_2)}^K$. Then, by periodicity of the rate functions, Y^K is a Galton-Watson process. An application of Lemma B.6.6 (with $\ell = 3$) yields that the one-step offspring distribution is determined by the generating function

$$\begin{aligned} g(s) &= \mathbb{E}\left[s^{Z^K_{\lambda_K(T_1+T_2)}} | Z_0^K = 1\right] \\ &= 1 - \frac{\mathrm{e}^{\lambda_K T_1 r_1 + T_* r_* + (\lambda_K T_2 - T_*) r_2}}{\frac{b_1}{r_1}\left(\mathrm{e}^{\lambda_K r_1 T_1} - 1\right)\mathrm{e}^{r_* T_* + (\lambda_K T_2 - T_*) r_2} + \frac{b_*}{r_*}\left(\mathrm{e}^{r_* T_*} - 1\right)\mathrm{e}^{(\lambda_K T_2 - T_*) r_2} + \frac{b_2}{r_2}\left(\mathrm{e}^{(\lambda_K T_2 - T_*) r_2} - 1\right) - \frac{1}{s-1}}. \end{aligned} \tag{B.6.26}$$

Therefore, the probability for Z^K to go extinct in the first step is

$$\begin{aligned} p_0^K &= g(0) \\ &= 1 - \frac{\mathrm{e}^{\lambda_K T_1 r_1 + T_* r_* + (\lambda_K T_2 - T_*) r_2}}{\frac{b_1}{r_1}\left(\mathrm{e}^{\lambda_K r_1 T_1} - 1\right)\mathrm{e}^{r_* T_* + (\lambda_K T_2 - T_*) r_2} + \frac{b_*}{r_*}\left(\mathrm{e}^{r_* T_*} - 1\right)\mathrm{e}^{(\lambda_K T_2 - T_*) r_2} + \frac{b_2}{r_2}\left(\mathrm{e}^{(\lambda_K T_2 - T_*) r_2} - 1\right) + 1}. \end{aligned} \tag{B.6.27}$$

Moreover, for the mean offspring we obtain

$$m^K = g'(1) = \mathrm{e}^{\lambda_K T_1 r_1 + T_* r_* + (\lambda_K T_2 - T_*) r_2} = \mathrm{e}^{\lambda_K (r_1 T_1 + r_2 T_2) + T_*(r_* - r_2)}. \tag{B.6.28}$$

Denote now by $g_n(s) = \mathbb{E}\left[s^{Y_n^K}\right]$ the generating function of the n-th generation of Y^K. Then it is well known [4] that $g_n = g_{n-1} \circ g = g \circ g_{n-1}$. Since $g : [0,1] \to [0,1]$ is convex and strictly increasing, we can deduce that (see Figure B.4)

$$(1 - p_0^K)(1 - s) \leq 1 - g(s) \leq m^K (1 - s) \qquad \forall s \in [0,1]. \tag{B.6.29}$$

This can be iterated (cf. [4, eq. I.11.7]) to obtain

$$(1 - p_0^K)^n (1 - s) \leq 1 - g_n(s) \leq (m^K)^n (1 - s) \qquad \forall s \in [0,1], \forall n \in \mathbb{N}. \tag{B.6.30}$$

Together with $g_n(0) = \mathbb{P}\left(Y_n^K = 0\right) = \mathbb{P}\left(T_{\text{ext}}^{Y^K} \leq n\right)$, where $T_{\text{ext}}^{Y^K} := \inf\{n \geq 0 : Y_n^K = 0\}$ is the extinction time of Y^K, this leads to

$$(1 - p_0^K)^n \leq \mathbb{P}\left(T_{\text{ext}}^{Y^K} > n\right) \leq (m^K)^n. \tag{B.6.31}$$

Let us now check the upper bound. For $r_{\text{av}} \geq 0$, the claim is trivially satisfied since the right hand side is larger than 1. In the case of $r_{\text{av}} < 0$, which, for K large enough, implies

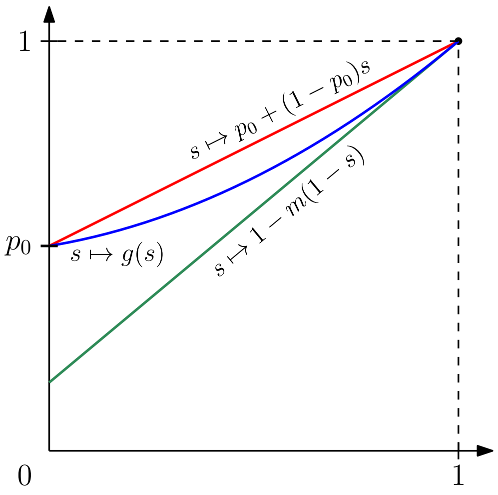


Figure B.4.: Generating function $g(s)$ and the corresponding affine upper and lower bounds from (B.6.29).

$m^K < 1$, we can estimate

$$\begin{aligned}
\mathbb{P}\left(T_{\text{ext}}^{Z^K} > t\ln K \middle| Z_0^K = 1\right) &\leq \mathbb{P}\left(T_{\text{ext}}^{Z^K} > \left\lfloor \frac{t\ln K}{\lambda_K(T_1+T_2)} \right\rfloor \lambda_K(T_1+T_2) \middle| Z_0^K = 1\right)\\
&= \mathbb{P}\left(T_{\text{ext}}^{Y^K} > \left\lfloor \frac{t\ln K}{\lambda_K(T_1+T_2)} \right\rfloor \middle| Y_0^K = 1\right)\\
&\leq \left(\mathrm{e}^{\lambda_K(r_1T_1+r_2T_2)+T_*(r_*-r_2)}\right)^{\left\lfloor \frac{t\ln K}{\lambda_K(T_1+T_2)} \right\rfloor}\\
&\leq \left(\mathrm{e}^{\lambda_K(r_1T_1+r_2T_2)+T_*(r_*-r_2)}\right)^{\left(\frac{t\ln K}{\lambda_K(T_1+T_2)}-1\right)}\\
&\leq \exp\left[r_{\text{av}}t\ln K + \frac{T_*(r_*-r_2)}{\lambda_K(T_1+T_2)}t\ln K + C\lambda_K\right]\\
&\leq \exp\left[r_{\text{av}}t\ln K + \delta_1 t\ln K + \delta_2\ln K\right]\\
&\leq \exp\left[\left(\left(r_{\text{av}}+\delta_1\right)t+\delta_2\right)\ln K\right], \qquad \text{(B.6.32)}
\end{aligned}$$

again for K large enough.

To obtain the lower bound, for $r_{\text{av}} \in \mathbb{R}$, we we first calculate that

$$\begin{aligned}
\left(1-p_0^K\right)^{-1} &= \frac{b_1}{r_1}(1-\mathrm{e}^{-\lambda_K T_1 r_1}) + \frac{b_*}{r_*}(1-\mathrm{e}^{-T_* r_*})\mathrm{e}^{-\lambda_K T_1 r_1}\\
&\quad + \frac{b_2}{r_2}(1-\mathrm{e}^{-(\lambda_K T_2 - T_*)r_2})\mathrm{e}^{-\lambda_K T_1 r_1 - T_* r_*} + \mathrm{e}^{-\lambda_K T_1 r_1 - T_* r_* - (\lambda_K T_2 - T_*)r_2}\\
&\leq \left(\frac{b_1}{|r_1|} + \frac{b_*}{|r_*|} + \frac{b_2}{|r_2|} + 1\right)\exp\left[\lambda_K T_1 |r_1| + T_* |r_*| + (\lambda_K T_2 - T*)|r_2|\right]\\
&= C\exp\left[\lambda_K(T_1|r_1| + T_2|r_2|) + T_*(|r_*| - |r_2|)\right]\\
&\leq \exp[\lambda_K \tilde{M}], \qquad \text{(B.6.33)}
\end{aligned}$$

for all $\tilde{M} > T_1 |r_1| + T_2 |r_2|$ and K large enough. This gives $1 - p_0^K \geq \exp[-\lambda_K \tilde{M}]$. Using again the connection between Z^K and Y^K, we can finally estimate similarly to (B.6.32)

$$\begin{aligned}\mathbb{P}\left(T_{\text{ext}}^{Z^K} > t \ln K | Z_0^K = 1\right) &\geq \exp\left[-t \frac{\tilde{M}}{T_1 + T_2} \ln K - \lambda_K \tilde{M}\right] \\ &\geq \exp\left[-(tM + \delta) \ln K\right], \end{aligned} \tag{B.6.34}$$

for all $M > (T_1 |r_1| + T_2 |r_2|)/(T_1 + T_2)$ and K large enough. ☐

Now we have collected all the tools to derive the convergence for shrinking populations and thus conclude the proof of Theorem B.6.1.

Lemma B.6.8. *Let Z^K follow the law of* $\text{BD}\left(b^K, d^K, \beta\right)$ *and assume that $r_{\text{av}} < 0$, then the convergence of Theorem B.6.1 holds true.*

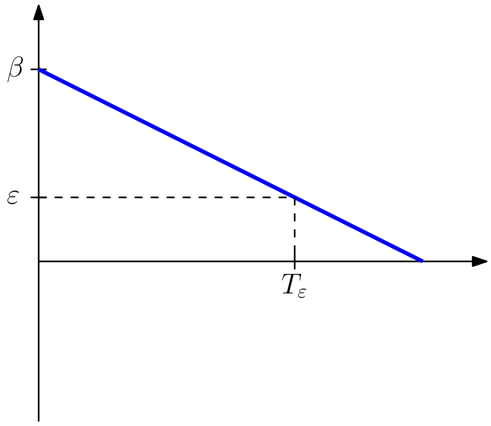


Figure B.5.: Graph of the limiting exponent $\beta + r_{\text{av}} s$ and the time T_ε that separates steps 1 and 2 of the proof of Lemma B.6.8.

Proof. Fix $\varepsilon \in (0, \beta)$ and set $\delta = \varepsilon/4, \eta = \beta - \varepsilon/3$ and $T_\varepsilon = (\beta - \varepsilon)/|r_{\text{av}}|$ (see Figure B.5).

Step 1 We first show convergence on $[0, T_\varepsilon]$. The computations are very similar to the proof of Lemma B.6.5 and proof step 3(i) of [45, Lem. A.1], respectively. Here one uses the event

$$\Omega_2^K := \left\{ \sup_{t \leq T_\varepsilon \ln K} \left| \mathrm{e}^{-g^K(t)} Z_t^K - K^\beta \right| \leq K^\eta \right\}, \tag{B.6.35}$$

which by Lemma B.6.4 has probability converging to 1.

Step 2 Next up, we show that extinction occurs before $T_\varepsilon + 3\varepsilon/|r_{\text{av}}|$. From the previous step, we know that, with high probability, the population at time $t = T_\varepsilon \ln K$ can be bounded from above by $2K^\varepsilon$, for large K. Moreover, the first part of Lemma B.6.7 provides an estimate on the extinction time of a subcritical birth death process initialized with one individual. Thus, we obtain the upper bound

$$\begin{aligned}\mathbb{P}\left(T_{\text{ext}}^{Z^K} > s\ln K \middle| Z_0^K = 2K^\varepsilon\right) &= 1 - \left(1 - \mathbb{P}\left(T_{\text{ext}}^{Z^K} > s\ln K \middle| Z_0^K = 1\right)\right)^{2K^\varepsilon} \\ &\leq 1 - \left(1 - \exp\left[\left(\left(r_{\text{av}} + \delta_1\right)s + \delta_2\right)\ln K\right]\right)^{2K^\varepsilon} \\ &= 1 - \left(1 - K^{(r_{\text{av}}+\delta_1)s+\delta_2}\right)^{2K^\varepsilon} \\ &= 1 - \left(1 - K^{-2\varepsilon}\right)^{2K^\varepsilon} \\ &\sim 2K^{-\varepsilon} \overset{K\to\infty}{\longrightarrow} 0. \end{aligned} \tag{B.6.36}$$

Here, the inequality is a consequence of (B.6.24), while for the last equality we set

$$s = \frac{2\varepsilon + \delta_2}{|r_{\text{av}} + \delta_1|} < \frac{3\varepsilon}{|r_{\text{av}}|}, \tag{B.6.37}$$

for proper choice of $\delta_1, \delta_2 > 0$. The final asymptotic behaviour can be deduced using the limit representation of the exponential function and a first-order approximation.
This now allows to deduce

$$\mathbb{P}\left(T_{\text{ext}}^{Z^K} > \frac{3\varepsilon}{|r_{\text{av}}|}\ln K \middle| Z_0^K = 2K^\varepsilon\right) \overset{K\to\infty}{\longrightarrow} 0. \tag{B.6.38}$$

Step 3 To finally conclude the convergence on the whole interval $[0, T]$, we have to ensure that the process stays bounded in the time interval $[T_\varepsilon, T_\varepsilon + 3\varepsilon/|r_{\text{av}}|]$. To this end, it is sufficient to consider the very rough upper bound obtained from the coupling with a Yule process (a pure birth process) with constant birth rate $\bar{b} = \max\{b_1, b_*, b_2\}$. Since those processes are non-decreasing, we just need to control the endpoint. The size of a family at time $(T_\varepsilon + 3\varepsilon/|r_{\text{av}}|)\ln K$, stemming from a single (i^{th}) individual at time $T_\varepsilon \ln K$, is given by a geometric random variable G_i^K with expectation $K^{3\bar{b}\varepsilon/|r_{\text{av}}|}$. Since families evolve independently of each other, we obtain by Chebyshev's inequality, for $d := \bar{b} + |r_{\text{av}}|$,

$$\mathbb{P}\left(\sup_{s\in[T_\varepsilon, T_\varepsilon+3\varepsilon/|r_{\text{av}}|]} Z_{s\ln K}^K \geq K^{3d\varepsilon/|r_{\text{av}}|}\right) \leq \mathbb{P}\left(\sum_{i=1}^{2K^\varepsilon} G_i^K \geq K^{3d\varepsilon/|r_{\text{av}}|}\right) \overset{K\to\infty}{\longrightarrow} 0. \tag{B.6.39}$$

Overall, this means that, with probability converging to 1, we have

$$\sup_{s\in[T_\varepsilon, T_\varepsilon+3\varepsilon/|r_{\text{av}}|]} \frac{\ln(1 + Z_{s\ln K}^K)}{\ln K} \leq \frac{3d}{|r_{\text{av}}|}\varepsilon. \tag{B.6.40}$$

□

B.6.2. Branching processes with immigration

We now turn to the study of birth death processes with immigration. In addition to the birth and death rates defined in the beginning of Section B.6.1, we introduce the following parameters connected to the effects of immigration. Let $c \in \mathbb{R}$ describe the initial order of incoming migration, i.e. initially immigrants arrive at overall rate K^c. When applied in Section B.4.2, this is representing the initial size of the neighbouring population multiplied by the mutation rate. Let $a_i, a_{*,i} \in \mathbb{R}$, for $i \in \{1, \dots, \ell\}$, be the immigrants net growth rates in the respective (sub-)phases. Thus, we define the time-dependent net growth rate of the immigrants as the periodic extension of

$$\tilde{a}^K(s) := \begin{cases} a_{*,i} & : s \in [T^{\Sigma}_{i-1}\lambda_K, T^{\Sigma}_{i-1}\lambda_K + T_{*,i}), \\ a_i & : s \in [T^{\Sigma}_{i-1}\lambda_K + T_{*,i}, T^{\Sigma}_{i}\lambda_K). \end{cases} \tag{B.6.41}$$

Moreover, (corresponding to $g_K(t)$ and r_{av}) we define the time integral

$$a^K(t) := \int_0^t \tilde{a}^K(s)\mathrm{d}s. \tag{B.6.42}$$

and the average growth rate of the immigrants $a_{\text{av}} := \left(\sum_{i=1}^{\ell} a_i T_i\right)/T^{\Sigma}_{\ell}$. Hence, the overall rate of immigration at time t is given by $K^c \mathrm{e}^{a^K(t)}$. We can define the Markov processes $(Z^K_t)_{t\geq 0}$ generated by

$$\left(\mathcal{L}^K_t f\right)(n) = \left(b^K(t)n + K^c\mathrm{e}^{a^K(t)}\right)(f(n+1) - f(n)) + d^K(t)n\,(f(n-1) - f(n)) \tag{B.6.43}$$

and with $Z^K_0 = \left\lfloor K^{\beta} - 1\right\rfloor$. We refer to the law of such processes by $Z^K \sim \text{BDI}\left(b^K, d^K, \beta, a^K, c\right)$.

As in the previous section, we derive a convergence result for the logarithmically rescaled process.

Theorem B.6.9. *Let Z^K follow the law of* BDI $\left(b^K, d^K, \beta, a^K, c\right)$. *Then, for all fixed $T \in (0, \infty)$, the following convergence holds in probability, with respect to the $L^{\infty}([0,T])$ norm,*

$$\left(\frac{\ln\left(1 + Z^K_{s \ln K}\right)}{\ln K}\right)_{s\in[0,T]} \overset{K\to\infty}{\longrightarrow} \left(\bar{\beta}_s\right)_{s\in[0,T]}, \tag{B.6.44}$$

where

$$\bar{\beta}_s = \begin{cases} (c + (r_{\text{av}} \vee a_{\text{av}})s) \vee 0 & : c > \beta, \\ (\beta + r_{\text{av}}s) \vee (c + a_{\text{av}}s) \vee (c + r_{\text{av}}s) \vee 0 & : \beta > 0, c \leq \beta, \\ (r_{\text{av}} \vee a_{\text{av}})(s - |c|/a_{\text{av}}) \vee 0 & : \beta = 0, c \leq 0, a_{\text{av}} > 0, \\ 0 & : \beta = 0, c < 0, a_{\text{av}} \leq 0, \\ 0 & : \beta = c = 0, a_{\text{av}} \leq 0, r_{\text{av}} \leq 0. \end{cases} \tag{B.6.45}$$

Note that the one case that is not cover by this result is that of $\beta = c = 0$, $a_{\text{av}} \leq 0$, $r_{\text{av}} > 0$.

The remainder of this section is dedicated to the proof of this theorem. We first study a number of specific cases in a series of lemmas and then outline how these can be combined to prove the final general result.

Since a^K is of the same form as g^K, we directly obtain.

Lemma B.6.10. *For all fixed $T < \infty$ we have the uniform convergence*

$$\sup_{s\le T}\left|\frac{\ln\left(K^c \mathrm{e}^{a^K(s\ln K)}\right)}{\ln K}-(c+a_{\mathrm{av}}s)\right|\xrightarrow{K\to\infty}0. \tag{B.6.46}$$

Proof. See the proof of Lemma B.6.3. ☐

As before, we can construct the processes Z^K in terms of the Poisson random measures $Q^{(b)}(\mathrm{d}s,\mathrm{d}\theta)$ and $Q^{(d)}(\mathrm{d}s,\mathrm{d}\theta)$ and derive the martingale decomposition $Z^K_t = Z^K_0 + M^K_t + A^K_t$, where

$$Z^K_t = Z^K_0 + \int_0^t\int_{\mathbb{R}_{\ge 0}} \mathbb{1}_{\theta\le b^K(s-)Z^K_{s-}+K^c\mathrm{e}^{a^K(s)}}\,Q^{(b)}(\mathrm{d}s,\mathrm{d}\theta) - \int_0^t\int_{\mathbb{R}_{\ge 0}} \mathbb{1}_{\theta\le d^K(s-)Z^K_{s-}}\,Q^{(d)}(\mathrm{d}s,\mathrm{d}\theta), \tag{B.6.47}$$

$$M^K_t = \int_0^t\int_{\mathbb{R}_{\ge 0}} \mathbb{1}_{\theta\le b^K(s-)Z^K_{s-}+K^c\mathrm{e}^{a^K(s)}}\,\tilde{Q}^{(b)}(\mathrm{d}s,\mathrm{d}\theta) - \int_0^t\int_{\mathbb{R}_{\ge 0}} \mathbb{1}_{\theta\le d^K(s-)Z^K_{s-}}\,\tilde{Q}^{(d)}(\mathrm{d}s,\mathrm{d}\theta), \tag{B.6.48}$$

$$A^K_t = \int_0^t r^K(s)Z^K_s + K^c\mathrm{e}^{a^K(s)}\mathrm{d}s, \tag{B.6.49}$$

$$\langle M^K\rangle_t = \int_0^t \left(b^K(s)+d^K(s)\right)Z^K_s + K^c\mathrm{e}^{a^K(s)}\mathrm{d}s, \tag{B.6.50}$$

$$\mathrm{d}Z^K_t = \mathrm{d}M^K_t + \left(r^K(t)Z^K_t + K^c\mathrm{e}^{a^K(t)}\right)\mathrm{d}t. \tag{B.6.51}$$

Note that, as in Section B.6.1, $\tilde{Q}^{(*)} := Q^{(*)} - \mathrm{d}s\mathrm{d}\theta$ and equalities only hold in distribution.

As in the case without immigration, we first take a look at the expected value. Moreover, we derive a bound on the variance of the process.

Lemma B.6.11. *Let Z^K follow the law of* $\mathrm{BDI}\left(b^K, d^K, \beta, a^K, c\right)$ *and assume that* $r_{\mathrm{av}} \neq a_{\mathrm{av}}$. *Then, for fixed $T < \infty$ and all $t \in [0, T\ln K]$,*

$$z^K_t := \mathbb{E}\left[Z^K_t\right] \approx \left(\mathrm{e}^{r_{\mathrm{av}}t}\left(K^\beta - 1\right) + \frac{K^c}{a_{\mathrm{av}} - r_{\mathrm{av}}}\left[\mathrm{e}^{a_{\mathrm{av}}t} - \mathrm{e}^{r_{\mathrm{av}}t}\right]\right)K^{\pm 3\delta}. \tag{B.6.52}$$

Moreover, under the additional assumption of $r_{\mathrm{av}} \neq 0$ *and* $2r_{\mathrm{av}} \neq a_{\mathrm{av}}$, *one obtains*

$$\mathrm{Var}\left(Z^K_t\right) \le \left[(\bar{b}+\bar{d})\left(K^\beta - 1 + \frac{K^c}{r_{\mathrm{av}} - a_{\mathrm{av}}}\right)\frac{\mathrm{e}^{2r_{\mathrm{av}}t} - \mathrm{e}^{r_{\mathrm{av}}t}}{r_{\mathrm{av}}} + K^c\left(1 - \frac{\bar{b}+\bar{d}}{r_{\mathrm{av}} - a_{\mathrm{av}}}\right)\frac{\mathrm{e}^{a_{\mathrm{av}}t} - \mathrm{e}^{2r_{\mathrm{av}}t}}{a_{\mathrm{av}} - 2r_{\mathrm{av}}}\right]K^{5\delta} \tag{B.6.53}$$

Here and in the remainder of the appendix, these kind of approximation results are meant in the following way: For any $\delta > 0$, there exists $K_0 \in \mathbb{N}$ such that, for all $K \ge K_0$, plugging in $K^{+3\delta}$ yields an upper bound and $K^{-3\delta}$ a lower bound for the left hand side.

Proof. As in Lemma B.6.2, we use the martingale decomposition to derive the integral equation

$$z_t^K = \left\lfloor K^\beta - 1 \right\rfloor + \int_0^t r^K(s) z_s^K + K^c \mathrm{e}^{a^K(s)} \mathrm{d}s. \tag{B.6.54}$$

By variation of constants, this leads to

$$z_t^K = \mathrm{e}^{g^K(t)} \left(\left\lfloor K^\beta - 1 \right\rfloor + K^c \int_0^t \mathrm{e}^{-g^K(s)} \mathrm{e}^{a^K(s)} \mathrm{d}s \right). \tag{B.6.55}$$

Using the convergence of g^K and a^K (Lemmas B.6.3 and B.6.10), we can estimate, for any $t \in [0, T \ln K]$, fixed $\delta > 0$, and K large enough,

$$z_t^K \approx \mathrm{e}^{r_{\mathrm{av}} t} \left(K^\beta - 1 + K^c \int_0^t \mathrm{e}^{(a_{\mathrm{av}} - r_{\mathrm{av}})s} \mathrm{d}s \right) K^{\pm 3\delta}. \tag{B.6.56}$$

This directly gives the first claim.

For the estimate on the variance, we see that

$$\mathrm{d}\left(Z_t^K\right)^2 = 2 Z_t^K \mathrm{d} Z_t^K + \mathrm{d}\langle Z^K \rangle_t = 2 Z_t^K \mathrm{d} M_t^K + 2 Z_t^K \left(r^K(t) Z_t^K + K^c \mathrm{e}^{a^K(t)} \right) \mathrm{d}t + \mathrm{d}\langle Z^K \rangle_t. \tag{B.6.57}$$

Define $u_t^K := \mathbb{E}\left[\left(Z_t^K\right)^2\right]$, then $u_0^K = \left\lfloor K^\beta - 1 \right\rfloor^2$ and

$$\begin{aligned} \dot{u}_t^K &= 0 + 2 r^K(t) u_t^K + 2 K^c \mathrm{e}^{a^K(t)} z_t^K + \left(\left(b^K(t) + d^K(t) \right) z_t^K + K^c \mathrm{e}^{a^K(t)} \right) \\ &= 2 r^K(t) u_t^K + \left(2 K^c \mathrm{e}^{a^K(t)} + b^K(t) + d^K(t) \right) z_t^K + K^c \mathrm{e}^{a^K(t)}. \end{aligned} \tag{B.6.58}$$

Using variation of constants, we deduce

$$u_t^K = \mathrm{e}^{2 g^K(t)} \left(\left\lfloor K^\beta - 1 \right\rfloor^2 + \int_0^t \mathrm{e}^{-2 g^K(s)} \left[\left(2 K^c \mathrm{e}^{a^K(s)} + b^K(s) + d^K(s) \right) z_s^K + K^c \mathrm{e}^{a^K(s)} \right] \mathrm{d}s \right). \tag{B.6.59}$$

We now focus on the integral term, where we plug in B.6.55 and treat each summand separately. For the first summand, involving $2 K^c \mathrm{e}^{a^K(s)}$, we obtain

$$\begin{aligned} & 2 K^c \int_0^t \mathrm{e}^{-g^K(s)} \mathrm{e}^{a^K(s)} \left(\left\lfloor K^\beta - 1 \right\rfloor + K^c \int_0^s \mathrm{e}^{-g^K(w)} \mathrm{e}^{a^K(w)} \mathrm{d}w \right) \mathrm{d}s \\ &= 2 \left\lfloor K^\beta - 1 \right\rfloor K^c \int_0^t \mathrm{e}^{-g^K(s)} \mathrm{e}^{a^K(s)} \mathrm{d}s + 2 K^{2c} \int_0^t \mathrm{e}^{-g^K(s)} \mathrm{e}^{a^K(s)} \int_0^s \mathrm{e}^{-g^K(w)} \mathrm{e}^{a^K(w)} \mathrm{d}w \mathrm{d}s \\ &= 2 \left\lfloor K^\beta - 1 \right\rfloor K^c \int_0^t \mathrm{e}^{-g^K(s)} \mathrm{e}^{a^K(s)} \mathrm{d}s + \left(K^c \int_0^t \mathrm{e}^{-g^K(s)} \mathrm{e}^{a^K(s)} \mathrm{d}s \right)^2. \end{aligned} \tag{B.6.60}$$

Together with the term $\left\lfloor K^{\beta}-1\right\rfloor^2$ and the prefactor $\mathrm{e}^{2g^K(t)}$ from B.6.59, this is equals to the square of z_t^K. Thus, the other two summands of B.6.59 give us the desired variance. Overall, using the convergence of g^K and a^K as above and plugging in B.6.55, this yields

$$\begin{aligned}
&\operatorname{Var}\left(Z_t^K\right)=u_t^K-(z_t^K)^2\\
=&\mathrm{e}^{2g^K(t)}\left(\int_0^t \mathrm{e}^{-2g^K(s)}\left(b^K(s)+d^K(s)\right)z_s^K\mathrm{d}s+K^c\int_0^t \mathrm{e}^{-2g^K(s)}\mathrm{e}^{a^K(s)}\mathrm{d}s\right)\\
\leq&\mathrm{e}^{2g^K(t)}\left(\left(\bar{b}+\bar{d}\right)\int_0^t \mathrm{e}^{-g^K(s)}\left(K^{\beta}-1+K^c\int_0^s \mathrm{e}^{-g^K(w)}\mathrm{e}^{a^K(w)}\mathrm{d}w\right)\mathrm{d}s+K^c\int_0^t \mathrm{e}^{-2g^K(s)}\mathrm{e}^{a^K(s)}\mathrm{d}s\right)\\
\leq&K^{5\delta}\mathrm{e}^{2r_{\mathrm{av}}t}\left(\left(\bar{b}+\bar{d}\right)\int_0^t \mathrm{e}^{-r_{\mathrm{av}}s}\left(K^{\beta}-1+K^c\int_0^s \mathrm{e}^{-r_{\mathrm{av}}w}\mathrm{e}^{a_{\mathrm{av}}w}\mathrm{d}w\right)\mathrm{d}s+K^c\int_0^t \mathrm{e}^{-2r_{\mathrm{av}}s}\mathrm{e}^{a_{\mathrm{av}}s}\mathrm{d}s\right)\\
=&K^{5\delta}\left(\left(\bar{b}+\bar{d}\right)\left(K^{\beta}-1+\frac{K^c}{r_{\mathrm{av}}-a_{\mathrm{av}}}\right)\frac{\mathrm{e}^{2r_{\mathrm{av}}t}-\mathrm{e}^{r_{\mathrm{av}}t}}{r_{\mathrm{av}}}+K^c\left(1-\frac{\bar{b}+\bar{d}}{r_{\mathrm{av}}-a_{\mathrm{av}}}\right)\frac{\mathrm{e}^{a_{\mathrm{av}}t}-\mathrm{e}^{2r_{\mathrm{av}}t}}{a_{\mathrm{av}}-2r_{\mathrm{av}}}\right).
\end{aligned}\tag{B.6.61}$$

□

Similar to the model without immigration (c.f. Lemma B.6.4), the starting point for proving the different parts of Theorem B.6.9 is an estimate on the bracket of the rescaled martingale, which will be used in combination with Doob's maximum inequality.

Lemma B.6.12. *Let $Z^K\sim\mathrm{BDI}\left(b^K,d^K,\beta,a^K,c\right)$. Then the process*

$$\tilde{M}_t^K:=\mathrm{e}^{-g^K(t)}\left(Z_t^K-z_t^K\right),\tag{B.6.62}$$

with z_t^K being defined in (B.6.52)*, is a martingale. Assuming that $r_{\mathrm{av}}\notin\{a_{\mathrm{av}},2a_{\mathrm{av}}\}$, for all $T<\infty$ and all $\delta>0$, there is a $K_0\in\mathbb{N}$ such that, for all $K\geq K_0$, it holds that*

$$\begin{aligned}
&\mathbb{E}\left[\langle\tilde{M}^K\rangle_{T\ln K}\right]K^{-5\delta}\leq K^c\frac{K^{(a_{\mathrm{av}}-2r_{\mathrm{av}})T}-1}{a_{\mathrm{av}}-2r_{\mathrm{av}}}\\
&+\left(\bar{b}+\bar{d}\right)\left(\left(K^{\beta}-1\right)\frac{1-K^{-r_{\mathrm{av}}T}}{r_{\mathrm{av}}}+\frac{K^c}{a_{\mathrm{av}}-r_{\mathrm{av}}}\left[\frac{K^{(a_{\mathrm{av}}-2r_{\mathrm{av}})T}-1}{a_{\mathrm{av}}-2r_{\mathrm{av}}}-\frac{1-K^{-r_{\mathrm{av}}T}}{r_{\mathrm{av}}}\right]\right).
\end{aligned}\tag{B.6.63}$$

Proof. Using Itô's formula and the martingale decomposition (B.6.51), we get

$$\mathrm{d}\tilde{M}_t^K=\mathrm{e}^{-g^K(t)}\mathrm{d}M_t^K,\tag{B.6.64}$$

which yields that $\tilde{M}^K$ is a martingale. Moreover, an application of Itô's isometry gives

$$\mathrm{d}\langle\tilde{M}^K\rangle_t=\mathrm{e}^{-2g^K(t)}\mathrm{d}\langle M^K\rangle_t=\mathrm{e}^{-2g^K(t)}\left(\left(b^K(t)+d^K(t)\right)Z_t^K+K^c\mathrm{e}^{a^K(t)}\right)\mathrm{d}t.\tag{B.6.65}$$

Thus,

$$\mathbb{E}\left[\langle\tilde{M}^K\rangle_{T\ln K}\right]=\int_0^{T\ln K}\mathrm{e}^{-2g^K(t)}\left(\left(b^K(t)+d^K(t)\right)z_t^K+K^c\mathrm{e}^{a^K(t)}\right)\mathrm{d}t,\tag{B.6.66}$$

and finally, using Lemma B.6.11 and the asymptotics of g^K, we obtain, for K large enough,

$$\int_0^{T\ln K} \mathrm{e}^{-2g^K(t)} z_t^K \mathrm{d}t \leq K^{5\delta} \int_0^{T\ln K} \mathrm{e}^{-r_{\mathrm{av}}t} \left(K^c \frac{\mathrm{e}^{(a_{\mathrm{av}}-r_{\mathrm{av}})t}-1}{a_{\mathrm{av}}-r_{\mathrm{av}}} + (K^\beta - 1) \right) \mathrm{d}t, \tag{B.6.67}$$

which yields the claimed estimate. □

Now we can start checking the convergence of Theorem B.6.9 in the cases without extinction or newly emerging populations.

Lemma B.6.13. *The claim of Theorem B.6.9 holds true for $c \leq \beta, \beta > 0$ and all $T < \infty$ such that*

$$\inf_{t\in[0,T]} (\beta + r_{\mathrm{av}}t) \vee (c + a_{\mathrm{av}}t) > 0. \tag{B.6.68}$$

Proof. We split the proof into several steps. First, in step 1, we apply Doob's maximum inequality for the rescaled martingales $\tilde{M}^K$ to prove the convergence in most of the possible cases. Next, in step 2, we make use of another maximum inequality to check the convergence for some other cases. Finally, in step 3, we go through a case distinction of all the possible scenarios of the lemma and explain the strategy of glueing together parts 1 and 2 with the help of the Markov property to cover the remaining cases.

Step 1: Let us consider the case where we can find an η such that

$$\frac{1}{2}\left[\beta \vee (\beta - r_{\mathrm{av}}T) \vee (c + (a_{\mathrm{av}} - 2r_{\mathrm{av}})T)\right] < \eta < \beta. \tag{B.6.69}$$

Then, applying Doob's maximum inequality to $\tilde{M}^K$ and adapting the computations of case 1(b) of the proof of [45, Lem. B.1] one obtains the desired convergence.

Step 2: Let us now consider the specific case of $\beta = c$ and $a_{\mathrm{av}} > r_{\mathrm{av}}$. Then we can apply the maximum inequality of [60, Ch. VI.1.2. p. 66] to the supermartingales $\left(\mathrm{e}^{-(a_{\mathrm{av}}t - r_{\mathrm{av}}t)}\tilde{M}_t^K\right)_{t\geq 0}$ to obtain, for K large enough,

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\leq T\ln K} \mathrm{e}^{-a_{\mathrm{av}}t}\left|Z_t^K - z_t^K\right| > K^\eta\right) &= \mathbb{P}\left(\forall t \leq T\ln K : \mathrm{e}^{-(a_{\mathrm{av}}t - r_{\mathrm{av}}t)}\left|\tilde{M}_t^K\right| > \mathrm{e}^{r_{\mathrm{av}}t - g^K(t)}K^\eta\right) \\
&\leq \mathbb{P}\left(\sup_{t\leq T\ln K} \mathrm{e}^{-(a_{\mathrm{av}}t - r_{\mathrm{av}}t)}\left|\tilde{M}_t^K\right| > K^{\eta-\delta}\right) \\
&\leq 3K^{-\eta+\delta} \sup_{t\leq T\ln K} \mathrm{e}^{-(a_{\mathrm{av}}t - r_{\mathrm{av}}t)}\mathbb{E}\left[\langle \tilde{M}^K\rangle_t\right]^{\frac{1}{2}} \\
&\leq CK^{-\eta+\delta} \sup_{s\leq T} K^{-(a_{\mathrm{av}} - r_{\mathrm{av}})s + \frac{1}{2}(\beta \vee (\beta - r_{\mathrm{av}}s) \vee (c + (a_{\mathrm{av}} - 2r_{\mathrm{av}})s))}.
\end{aligned} \tag{B.6.70}$$

To conclude, we again proceed as in proof step 2 of [45, Lem. B.1].

Step 3: Now we can check whether all cases are covered. Under the assumptions of the lemma, i.e. $\beta > 0, c \leq \beta, T \in (0, \infty)$ and $\inf_{s\in[0,T]}(\beta + r_{\text{av}}s) \vee (c + a_{\text{av}}s) > 0$, we see that the constraint (B.6.69) of step 1, which is equivalent to

$$r_{\text{av}} > -\frac{\beta}{T} \quad \text{and} \quad c + (a_{\text{av}} - 2r_{\text{av}})T < 2\beta, \tag{B.6.71}$$

holds true, if

(i) $r_{\text{av}} \geq 0, a_{\text{av}} \leq 2r_{\text{av}}$,

(ii) $r_{\text{av}} \geq 0, a_{\text{av}} > 2r_{\text{av}}$ and $T < T^* := (2\beta - c)/(a_{\text{av}} - 2r_{\text{av}})$,

(iii) $r_{\text{av}} < 0, a_{\text{av}} \leq r_{\text{av}}$,

(iv) $r_{\text{av}} < 0, a_{\text{av}} > r_{\text{av}}, c + a_{\text{av}}\beta/\left|r_{\text{av}}\right| \leq 0$ (in this case $T < \beta/|r_{\text{av}}|$),

(v) $r_{\text{av}} < 0, a_{\text{av}} > r_{\text{av}}, c + a_{\text{av}}\beta/\left|r_{\text{av}}\right| > 0$ and $T < T^*$ (in this case $T^* \leq \beta/|r_{\text{av}}|$).

Hence the only remaining cases are (see Figure B.6)

- $r_{\text{av}} \geq 0, a_{\text{av}} > 2r_{\text{av}}$ and $T > T^* := (2\beta - c)/(a_{\text{av}} - 2r_{\text{av}})$,
- $r_{\text{av}} < 0, a_{\text{av}} > r_{\text{av}}, c + a_{\text{av}}\beta/\left|r_{\text{av}}\right| > 0$ and $T > T^*$.

The strategy for these is the following: Take $t^* < T_1 < T^*$, where $t^* := (\beta - c)/(a_{\text{av}} - r_{\text{av}}) \geq 0$ is the first time when $(s \mapsto \beta + r_{\text{av}}s)$ crosses $(s \mapsto c + a_{\text{av}}s)$. Now apply step 1 on the interval $[0, T_1]$ to get the desired convergence up to T_1. In particular, $1 + Z^K_{T_1 \ln K} \approx K^{c+a_{\text{av}}T_1 \pm \varepsilon}$ for K large enough. On $[T_1, T]$ we apply step 2 to the approximating processes $Z^{(K,+)}, Z^{(K,-)}$, with parameters $b_i^\pm := b_i, d_i^\pm := d_i, a_i^\pm := a_i$, for $i \in \{*, 1, 2\}$, and $\beta^\pm = c^\pm := c + a_{\text{av}}T_1 \pm \varepsilon$. Together with the Markov property these approximations give the claim. ☐

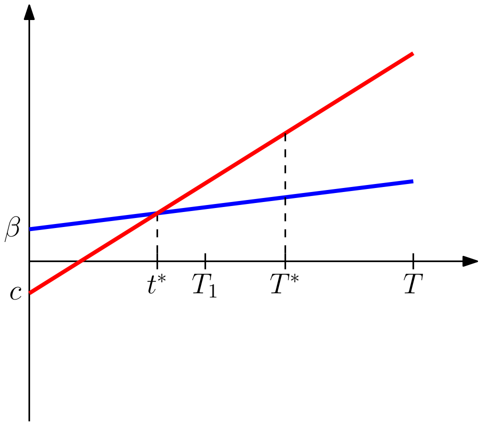


Figure B.6.: Illustration of the two remaining cases of step 3 of the proof of Lemma B.6.13.

Let us now collect some technical results on the (non-)emergence, extinction and instantaneous immigration of populations, as well as the continuity of the exponent, to complete the proof of Theorem B.6.9.

Lemma B.6.14 (Non-emergence of any new population)**.** *Let* $Z^K \sim \mathrm{BDI}\left(b^K, d^K, \beta, a^K, c\right)$ *such that* $\beta = 0$ *and* $c < 0$*. Then, for all* $T > 0$ *such that* $c + a_{\mathrm{av}}T < 0$ *it holds*

$$\lim_{K\to\infty} \mathbb{P}\left(\forall s \in [0, T] : Z^K_{s \ln K} = 0\right) = 1. \tag{B.6.72}$$

Proof. This can be shown by following the proof of [45, Lem. B.7]. ☐

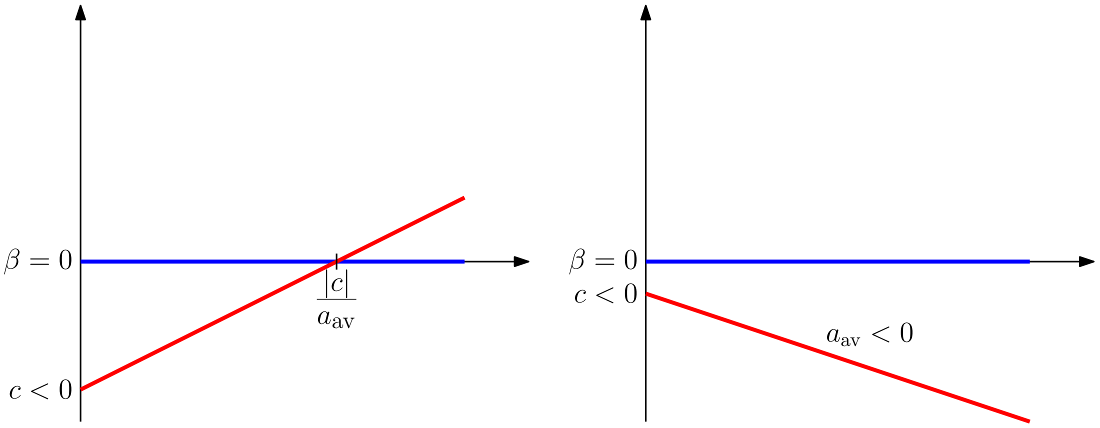


Figure B.7.: Illustrations of the two possible situations for non-emergence of any new population, as in Lemma B.6.14.

Lemma B.6.15 (Emergence of a new population)**.** *Let* $Z^K \sim \mathrm{BDI}\left(b^K, d^K, \beta, a^K, c\right)$ *such that* $\beta = 0, c = -\varepsilon < 0$ *and* $a_{\mathrm{av}} > 0$*. Then, for all* $\eta > (1 \vee 2r_{\mathrm{av}}/a_{\mathrm{av}})\varepsilon$*, it holds that*

$$\lim_{K\to\infty} \mathbb{P}\left(K^{\varepsilon/2} - 1 \leq Z^K_{\frac{2\varepsilon}{a_{\mathrm{av}}} \ln K} \leq K^\eta - 1\right) = 1. \tag{B.6.73}$$

Proof. Let us first consider the lower bound. The number of immigrant families that arrive within the time interval $[0, (2\varepsilon/a_{\mathrm{av}}) \ln K]$ and that survived up to time $(2\varepsilon/a_{\mathrm{av}}) \ln K$, is, by thinning, a Poisson random variable with parameter

$$\vartheta = \int_0^{\frac{2\varepsilon}{a_{\mathrm{av}}} \ln K} K^{-\varepsilon} \mathrm{e}^{a^K(t)} \mathbb{P}\left(T^{\tilde{Z}^K}_{\mathrm{ext}} > \frac{2\varepsilon}{a_{\mathrm{av}}} \ln K - t \middle| \tilde{Z}^K_0 = 1\right) \mathrm{d}t, \tag{B.6.74}$$

where $\tilde{Z}^K$ is the corresponding birth death process without immigration. Hence we can apply the second part of Lemma B.6.7 and bound this from below, for all $\delta > 0$ and K large enough,

$$\begin{aligned}
\vartheta &= \int_0^{\frac{2\varepsilon}{a_{\mathrm{av}}}} K^{-\varepsilon} \mathrm{e}^{a^K(s\ln K)} \mathbb{P}\left(T_{\mathrm{ext}}^{\tilde{Z}^K} > \left(\frac{2\varepsilon}{a_{\mathrm{av}}} - s\right) \ln K \middle| \tilde{Z}_0^K = 1\right) \ln K \mathrm{d}s \\
&\geq \int_0^{\frac{2\varepsilon}{a_{\mathrm{av}}}} K^{-\varepsilon} K^{-\delta} \mathrm{e}^{a_{\mathrm{av}} s \ln K} \exp\left[-\left(M\left(\frac{2\varepsilon}{a_{\mathrm{av}}} - s\right) + \delta\right) \ln K\right] \ln K \mathrm{d}s \\
&= K^{-\varepsilon - 2\delta - M\frac{2\varepsilon}{a_{\mathrm{av}}}} \int_0^{\frac{2\varepsilon}{a_{\mathrm{av}}}} \mathrm{e}^{(a_{\mathrm{av}}+M)s\ln K} \ln K \mathrm{d}s \\
&= K^{-\varepsilon - 2\delta - M\frac{2\varepsilon}{a_{\mathrm{av}}}} \frac{K^{\frac{2\varepsilon}{a_{\mathrm{av}}}(a_{\mathrm{av}}+M)} - 1}{a_{\mathrm{av}} + M} \\
&\geq \frac{1}{2} \frac{1}{a_{\mathrm{av}} + M} K^{-\varepsilon - 2\delta - M\frac{2\varepsilon}{a_{\mathrm{av}}} + \frac{2\varepsilon}{a_{\mathrm{av}}}(a_{\mathrm{av}}+M)} \\
&= CK^{\varepsilon - 2\delta}. \qquad \text{(B.6.75)}
\end{aligned}$$

Therefore, taking $\delta > 0$ small enough, we have

$$\lim_{K\to\infty} \mathbb{P}\left(Z^K_{\frac{2\varepsilon}{a_{\mathrm{av}}}\ln K} \geq K^{\frac{\varepsilon}{2}}\right) = 1 \qquad \text{(B.6.76)}$$

and can continue as in the proof of [45, Lem. B.8]. ☐

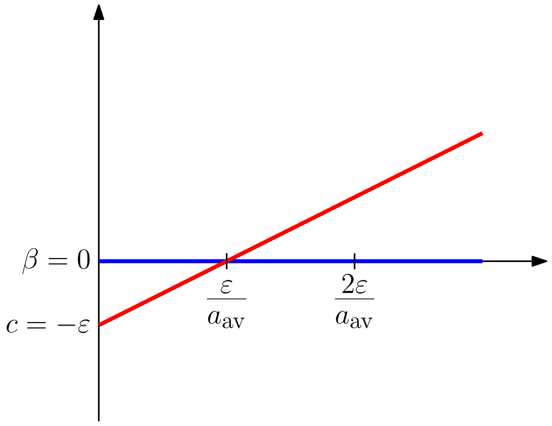


Figure B.8.: Illustration of the situation for emergence of a new population, as in Lemma B.6.15.

Lemma B.6.16 (Continuity of the exponent). *Let* $Z^K \sim \mathrm{BDI}\left(b^K, d^K, \beta, a^K, c\right)$ *such that* $c \leq \beta$. *Then there exists a constant* $\bar{c} = \bar{c}(\bar{b}, \bar{d}, a_{\mathrm{av}})$ *such that, for all* $\varepsilon > 0$, *it holds that*

$$\lim_{K\to\infty} \mathbb{P}\left(\forall s \in [0,\varepsilon] : K^{\beta - \bar{c}\varepsilon} - 1 \leq Z^K_{s\ln K} \leq K^{\beta + \bar{c}\varepsilon} - 1\right) = 1. \qquad \text{(B.6.77)}$$

Proof. Using the estimate of Lemma B.6.10 one can adapt the proof of [45, Lem. B.9]. ☐

Lemma B.6.17 (Extinction). *Let* $Z^K \sim \mathrm{BDI}\left(b^K, d^K, \beta, a^K, c\right)$ *such that* $r_{\mathrm{av}} < 0$.

(a) If in addition $c < 0$ *and* $c + a_{\mathrm{av}}\beta / |r_{\mathrm{av}}| < 0$*, then, for all* $0 < \eta < T$ *with the property* $c + a_{\mathrm{av}}(\beta/|r_{\mathrm{av}}| + T) < 0$*, it holds that*

$$\lim_{K\to\infty} \mathbb{P}\left(\forall s \in \left[\frac{\beta}{|r_{\mathrm{av}}|} + \eta, \frac{\beta}{|r_{\mathrm{av}}|} + T\right] : Z^K_{s\ln K} = 0\right) = 1. \tag{B.6.78}$$

(b) If in addition $a_{\mathrm{av}} < 0$ *and* $c + a_{\mathrm{av}}\beta/|r_{\mathrm{av}}| > 0$*, then, for all* $0 < \eta < T$*, we have*

$$\lim_{K\to\infty} \mathbb{P}\left(\forall s \in \left[\frac{c}{|a_{\mathrm{av}}|} + \eta, \frac{c}{|a_{\mathrm{av}}|} + T\right] : Z^K_{s\ln K} = 0\right) = 1. \tag{B.6.79}$$

Proof. The proof follows the lines of the one for [45, Lem. B.10]. We emphazise, that the results of Lemma B.6.7 on the extinction time are crucial for the processes studied in the present paper. □

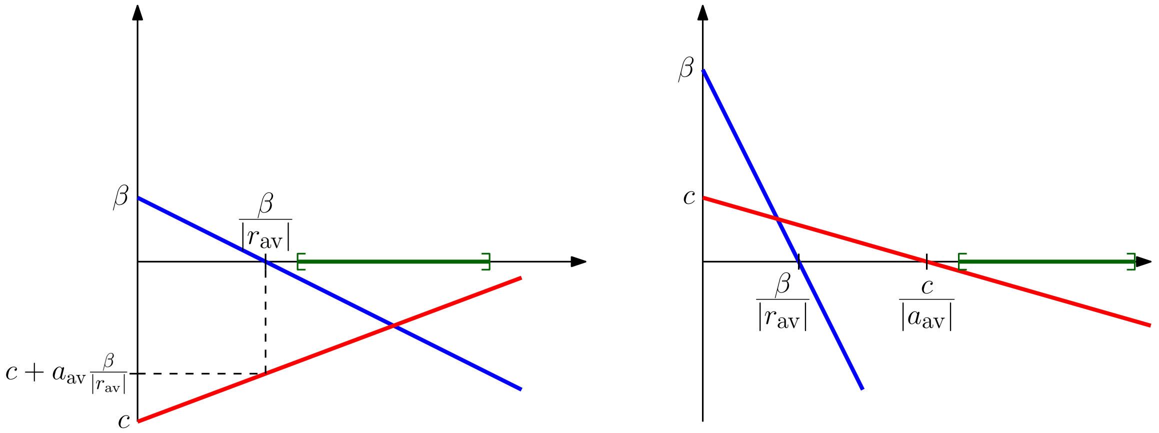


Figure B.9.: Illustration of cases (a) and (b) for extinction in Lemma B.6.17.

Lemma B.6.18 (Instantaneous immigration)**.** *Let* $Z^K \sim \mathrm{BDI}\left(b^K, d^K, \beta, a^K, c\right)$ *such that* $0 \le \beta < c$*. Then, for all* $\varepsilon > 0$ *and all* $\bar{a} > |r_{\mathrm{av}}| \vee |a_{\mathrm{av}}|$ *it holds that*

$$\lim_{K\to\infty} \mathbb{P}\left(Z^K_{\varepsilon\ln K} \in [K^{c-\bar{a}\varepsilon}, K^{c+\bar{a}\varepsilon}]\right) = 1. \tag{B.6.80}$$

Proof. With the help of Lemma B.6.2, this is verified as for [45, Lem. B.4]. □

Having all tools at hand, we can now prove the main convergence result step by step.

Proof of Theorem B.6.9. The previous lemmas cover all the different scenarios of Theorem B.6.9 either directly or in combination, under application of the Markov property and the continuity result in Lemma B.6.16. For convenience, we provide a summary of how to treat each case:
First, we check all the different cases with $c \le \beta$.

- If $\beta = 0, c < 0$ and $a_{\mathrm{av}} \leq 0$, we apply Lemma B.6.14.
- If $\beta = 0, c < 0$ and $a_{\mathrm{av}} > 0$, we apply Lemma B.6.14 up to time $(|c| - \varepsilon)/a_{\mathrm{av}}$. On $[(|c| - \varepsilon)/a_{\mathrm{av}}, (|c| + \varepsilon)/a_{\mathrm{av}}]$, we use Lemma B.6.15, and finally, Lemma B.6.13 on $[(|c| + \varepsilon)/a_{\mathrm{av}}, T]$.
- If $\beta > 0$, the case $r_{\mathrm{av}} \geq 0$ is already checked in Lemma B.6.13.
- If $\beta > 0, r_{\mathrm{av}} < 0$ and $a_{\mathrm{av}} \geq 0$ such that $c + a_{\mathrm{av}}\beta / |r_{\mathrm{av}}| > 0$, we can also apply Lemma B.6.13.
- If $\beta > 0, r_{\mathrm{av}} < 0$ and $a_{\mathrm{av}} > 0$ such that $c + a_{\mathrm{av}}\beta / |r_{\mathrm{av}}| < 0$, we use Lemma B.6.13 on $[0, \beta / |r_{\mathrm{av}}| - \varepsilon]$. After that time, an application of Lemma B.6.17 leads to extinction until $(|c| - \varepsilon)/a_{\mathrm{av}}$. On $[(|c| - \varepsilon)/a_{\mathrm{av}}, (|c| + \varepsilon)/a_{\mathrm{av}}]$, we use Lemma B.6.15, and finally, Lemma B.6.13 on $[(|c| + \varepsilon)/a_{\mathrm{av}}, T]$.
- If $\beta > 0$ and either $(r_{\mathrm{av}} < 0, a_{\mathrm{av}} < 0)$ or $(r_{\mathrm{av}} < 0, a_{\mathrm{av}} = 0, c < 0)$, we use Lemma B.6.13 on $[0, \beta / |r_{\mathrm{av}}| \vee c/|a_{\mathrm{av}}| - \varepsilon]$ or $[0, \beta / |r_{\mathrm{av}}| - \varepsilon]$, respectively. After that time, an application of Lemma B.6.17 leads to extinction.
- Cases $(\beta = c = 0, a_{\mathrm{av}} > 0)$, $(\beta = c = 0, a_{\mathrm{av}} \leq 0, r_{\mathrm{av}} \leq 0)$, $(\beta > 0, r_{\mathrm{av}} < 0, c = a_{\mathrm{av}} = 0)$, as well as $(\beta > 0, r_{\mathrm{av}} < 0, a_{\mathrm{av}} > 0, c < 0, \beta / |r_{\mathrm{av}}| = |c| / a_{\mathrm{av}})$ can be treated using comparisons.

Finally, for $c > \beta$, we apply Lemma B.6.18 on $[0, \varepsilon]$ and then continue with the previous results. ☐

B.7. Phase of invasion

In contrast to the results of Appendix B.6, in this chapter we focus on birth and death processes with many individuals on shorter but still divergent time horizons. These restrictions allow for relatively strong bounds on the growth of these processes. The results are used in the proof of the main theorem of this paper to control the mutant population when it gets almost macroscopic, to ensure that this happens in the right (fit) phase.

Lemma B.7.1. *Let Z^K be birth death processes with constant parameters $b, d \geq 0$ and $r = b - d$. Moreover, for $q \in (0, \infty)$ and $\alpha \in \mathbb{R}$, assume initial values $Z^K_0 = q\mathrm{e}^{\alpha\lambda_K} K$. Then, for all $T \geq 0$ and all $\gamma \in (0, 1)$,*

(a) $\mathbb{P}\left(Z^K_{T\lambda_K} < p\mathrm{e}^{rT\lambda_K} Z^K_0\right) = o(K^{-\gamma}) \overset{K\to\infty}{\longrightarrow} 0, \qquad$ *for* $p \in (0, 1)$,

(b) $\mathbb{P}\left(Z^K_{T\lambda_K} > p\mathrm{e}^{rT\lambda_K} Z^K_0\right) = o(K^{-\gamma}) \overset{K\to\infty}{\longrightarrow} 0, \qquad$ *for* $p \in (1, \infty)$.

Proof. For this proof we make use of the rescaled martingale $\hat{M}_t^K = \mathrm{e}^{-rt} Z_t^K - Z_0^K$. Then $\mathrm{d}\hat{M}_t^K = \mathrm{e}^{-rt}\mathrm{d}M_t^K$ and thus $\mathrm{d}\langle \hat{M}^K \rangle_t = \mathrm{e}^{-2rt}\mathrm{d}\langle M^K \rangle_t$ (cf. (B.6.8) and corresponding discussion). Therefore, in the case of $r \neq 0$, we can compute

$$\begin{aligned}\mathbb{E}\left[\langle \hat{M}^K \rangle_t\right] &= \int_0^t \mathrm{e}^{-2rs}\,(b+d)\,\mathbb{E}\left[Z_s^K\right]\mathrm{d}s \\ &= (b+d)\,Z_0^K \int_0^t \mathrm{e}^{-rs}\mathrm{d}s \\ &= (b+d)\,Z_0^K \frac{1-\mathrm{e}^{-rt}}{r}. \end{aligned} \tag{B.7.1}$$

An application of Doob's maximum inequality [30, Thm. 3.87] yields that, for $0 < p < 1$,

$$\begin{aligned}\mathbb{P}\left(Z_{T\lambda_K}^K < p\mathrm{e}^{rT\lambda_K} Z_0^K\right) &= \mathbb{P}\left(\mathrm{e}^{-rT\lambda_K} Z_{T\lambda_K}^K - Z_0^K < -(1-p)Z_0^K\right) \\ &\leq \mathbb{P}\left(\sup_{t\leq T\lambda_K} \left|\mathrm{e}^{-rt}Z_t^K - Z_0^K\right| > (1-p)Z_0^K\right) \\ &= \mathbb{P}\left(\sup_{t\leq T\lambda_K} \left|\hat{M}_t^K\right| > (1-p)Z_0^K\right) \\ &\leq (1-p)^{-2}\left(Z_0^K\right)^{-2}\mathbb{E}\left[\langle \hat{M}^K \rangle_{T\lambda_K}\right] \\ &= C\mathrm{e}^{-\alpha\lambda_K}K^{-1}\frac{\left|\mathrm{e}^{-rT\lambda_K}-1\right|}{|r|} \\ &\leq \tilde{C}\mathrm{e}^{-\alpha\lambda_K}K^{-1}\mathrm{e}^{|r|T\lambda_K} \\ &= \tilde{C}\exp\left(-\alpha\lambda_K - \ln K + |r|\,T\lambda_K\right) \\ &\leq \tilde{C}\exp\left(-\gamma\ln K\right) \xrightarrow{K\to\infty} 0. \end{aligned} \tag{B.7.2}$$

The last inequality is true for every $\gamma \in (0,1)$ and K large enough since, in the limit of large K, one has $(|r|\,T-\alpha)\lambda_K/\ln K < 1-\gamma$ in the limit of large K.

If we now consider $r = 0$, we see that in this case

$$\mathbb{E}\left[\langle \hat{M}^K \rangle_t\right] = (b+d)\,Z_0^K t. \tag{B.7.3}$$

Plugging this into the above estimate of the probability leads to an even better bound.

For $p > 1$, we adapt the first calculations as

$$\begin{aligned}\mathbb{P}\left(Z_{T\lambda_K}^K > p\mathrm{e}^{rT\lambda_K} Z_0^K\right) &= \mathbb{P}\left(\mathrm{e}^{-rT\lambda_K} Z_{T\lambda_K}^K - Z_0^K > -(1-p)Z_0^K\right) \\ &\leq \mathbb{P}\left(\sup_{t\leq T\lambda_K} \left|\mathrm{e}^{-rt}Z_t^K - Z_0^K\right| > (p-1)Z_0^K\right), \end{aligned} \tag{B.7.4}$$

from where we can continue as before. □

Remark 22. Note that we do not only prove the convergence to zero here but also determine the speed of convergence, which is $o(K^{-\gamma})$, for all $\gamma \in (0,1)$.

Corollary B.7.2. *Let Z^K be birth death process with time-dependent rates b^K, d^K as introduced in Appendix B.6.1 and recall that $g^K(t) = \int_0^t r^K(s)\mathrm{d}s$, where $r^K(s) = b^K(s) - d^K(s)$ is the net growth rate. For initial values $Z_0^K = \varepsilon^2 K$, for all $0 < p_1 < 1 < p_2$, all $T \geq 0$, and all $\gamma \in (0,1)$, we obtain*

$$\mathbb{P}\left(p_1 \mathrm{e}^{g^K(t)} Z_0^K < Z_t^K < p_2 \mathrm{e}^{g^K(t)} Z_0^K \quad \forall t \in [0, T\lambda_K]\right) = 1 - o(K^{-\gamma}) \overset{K\to\infty}{\longrightarrow} 1. \tag{B.7.5}$$

Proof. The corollary can be deduced easily by iterative application of Lemma B.7.1, combined with the Markov property at the times T_i^{Σ}. □

C. Preprint: Crossing a fitness valley in a changing environment: With and without pit stop

This appendix is available as a preprint as joint work with Anna Kraut [70],

> M. Esser and A. Kraut, *Crossing a fitness valley in a changing environment: With and without pit stop.* Preprint, arXiv:2503.19766.

We consider a stochastic individual-based model of adaptive dynamics for an asexually reproducing population with mutation. Biologically motivated by the influence of seasons or the variation of drug concentration during medical treatment, the model parameters vary over time as piecewise constant and periodic functions. We study the typical evolutionary behavior of the population by looking at limits of large populations and rare mutations. An analysis of the crossing of valleys in the fitness landscape in a changing environment leads to various interesting phenomena on different time scales, which depend on the length of the valley. By carefully examining the influence of the changing environment on each time scale, we are able to determine the crossing rates of fit mutants and their ability to invade the resident population. In addition, we investigate the special scenario of pit stops, where single intermediate mutants within the valley have phases of positive fitness and can thus grow to a diverging size before going extinct again. This significantly accelerates the traversal of the valley and leads to an interesting new time scale.

C.1. Introduction

Adaptation to the environment is one of the key factors of biological evolution. Condensed in the principle of *survival of the fittest*, it is known since Charles Darwin [51], that among several individuals of a species, the ones that are better adapted to their natural environment transmit their characteristics to a larger number of descendants than the ones that are less adapted. In the long run, this leads to the persistence of the adapted individual traits and the disappearance of disadvantageous traits. This general principle seems to be nicely short and satisfying. However, the observation of nature gives suggests that the underling mechanisms are somewhat more involved. There are two specific aspects that we like to point out in the following.

First, let us turn to the micro evolutionary perspective by looking at a cell's DNA. Most of the time, the DNA is replicated exactly during cell division, however, sometimes this process is effected by errors, called mutations. Changing a single base-pair can likely cause a defect in the encoded gene, which makes us believe that most mutations are disadvantageous. In some cases, the accumulation of multiple mutations can lead to an advantage by changing the function of a particular gene. Since effective mutations (altering the coding region of the DNA) are rare, these mutations have to be collected one by one. This means that, in order to reach a state of higher fitness, there is a temporary decrease in fitness in between. This phenomenon is called a *fitness valley* and is for example observed in the initiation of cancer [137], the formation of the flagella apparatus of bacteria [152], and other fields [50, 59, 133].

A second observation is that the environment that populations adapt to underlies ongoing changes. Even if we restrict to purely abiotic factors such as temperature, humidity, or accessibility and concentration of nutrients, fluctuations are ubiquitous and have a big impact on the process of selection. In addition to random or chaotic fluctuations of the environment, there are cases of regular and recurrent changes. One can think of seasonal changes or the variation of drug concentration during medical treatment as simple examples.

The present article aims to study the interplay of these two aspects, extending the basic picture of selection. Stochastic individual-based models of adaptive dynamics, as introduced by Fournier and Méléard [83], have turned out to be a useful model type that allows to depict many different mechanisms. A first key result about the basic model was the separation of ecological and evolutionary time scales, studied by Champagnat [40]. In the last decades, this model has been developed and extended in multiple directions, e.g. studying diploidy [28, 126, 148], dormancy [22, 20], the canonical equation of adaptive dynamics [7, 42, 153], or Hamilton-Jacobi equations [44]. At its core, these model rely on the simple biological principles of asexual clonal birth, natural death, additional competition-induced deaths, depending on the population density, and the possibility of mutation at birth.

From the various scaling parameters that have been studied for this class of models, we focus on large populations of order $K \to \infty$ and small mutation probabilities $\mu_K \to 0$ that vanishes as power law, i.e. $\mu_K = K^{-1/\alpha}$, for some $\alpha > 0$. This regime has been studied in various works e.g. [29, 48, 34, 153, 69, 71]. Under these assumptions, it has been shown that the dominating types within the population move fast towards an equilibrium, in a time of order 1, while newly appearing mutants with a positive (invasion) fitness need a time of order $\ln K$ to reach a macroscopic size.

To depict repeating changes of the environment, we let all of the model parameters vary over time as piecewise constant, periodic functions and introduce a new parameter λ_K to control the speed of environmental changes. Branching processes in changing environments have previously been studied in the discrete-time, single-type setting [15, 23, 113, 163], answering questions about population size growth, genealogies and tree structures. A deterministic differential equation model for a multi-type non-competitive population spreading across a sink of negative fitness was considered in [9]. Other works have focused on either fast changes on time scales $O(1)$ (cf. [78, 87] for deterministic models of interacting populations), which hinder the resident population's ability to stabilize close to an equilibrium, or very slow changes on time scales larger than the $\ln K$-times of mutant growth and invasion (cf. [49] for a multi-type Moran-like model). Our work, on the other hand, allows for an intermediate speed of environmental changes, choosing $1 \ll \lambda_K \ll \ln K$. As previously worked out in [71], this means that the effective growth rates on the $\ln K$-time scale of mutant populations are given as weighted averages over all phases. On this time scale, the population's traits evolve until it gets stuck in a local fitness maximum. In this present work, we study how the population can leave such a local maximum, traversing a valley in the fitness landscape on a more accelerated time scale.

Based on the notion of phase-dependent and average fitness, we distinguish two scenarios. We first consider a *strict fitness valley*, which means that all intermediate traits between the current resident trait and the advantageous mutant are unfit in every phase, resulting in a scenario as in the (constant environment) considerations of [29]. In this case, successfully invading mutants can be observed on the time scale of $1/K\mu_K^L$, where L describes the width of the valley. Since the environment changes on a much shorter time scale, the rate of crossing the valley is given by the weighted average of the crossing rates computed for constant environments in [29]. The main difficulty arising in this case is the fact that the probability for the mutant population to fixate and finally grow to a macroscopic size is not only determined by its average fitness or its fitness in the phase of arrival alone. Instead, it strongly depends on the arrival time within the phase since one has to ensure that the new mutant grows enough during fit phases to not go extinct during potential unfit phases. In our result, we make this precise by defining a set $A \subseteq [0, \infty)$ of possible arrival times of successful mutants, and incorporating it when computing the effective crossing rate.

To relax the assumptions of the strict valley, the second scenario allows for a single *pit stop* within the fitness valley. This means that there is a single trait w in the valley that has a positive fitness during one phase, while maintaining a negative average fitness. In contrast to the approximating subcritical birth death processes in [29], this trait can grow for a short but diverging time of order λ_K. Therefore, we see a speed up in the crossing rates for the fitness valley and the respective time scale. Since the growth behavior of w, and hence also the acceleration of the time scale, strongly depends on the equilibrium size of the resident population, we need to derive more accurate estimates on the resident's stability. Another challenge in this second scenario is to distinguish typical crossings from other possibilities. A crossing is more likely when the population of trait w can grow the most. This is exactly the case when a mutant of trait w is born at the very beginning of its fit phase and hence produces the next order mutants at the highest possible rate when it is at its peak population size, at the end of the fit phase or the beginning of the next one, respectively.

The remainder of this article is structured as follows. In Section C.2.1, we introduce the individual-based model for a population in a time-dependent environment and point out some key quantities, such as equilibrium states and invasion fitness. Section C.2.2 and Section C.2.3 provide our two main convergence results for strict fitness valleys and valleys with a pit stop, respectively. We discuss the proof heuristics, the necessity of some assumptions, and possible generalizations of our results in Chapter C.3. Chapter C.4 is dedicated to the proofs of the main results, and in the Appendix C.5 we collected some technical results on birth death processes.

C.2. Model and Results

C.2.1. Model introduction: Individual-based adaptive dynamics in changing environment

We consider a population that is composed of a finite number of asexually reproducing individuals. Denote by $V = [\![0, L]\!] := \{0, 1, \dots, L\}$ the space of possible traits, characterising the individuals. To model a periodically changing environment, we consider a finite number $\ell \in \mathbb{N}$ of phases. For each phase $i = 1, \dots, \ell$ and all traits $v, w \in V$, we introduce the following biological parameters:

- $b_v^i \in \mathbb{R}_+$, the *birth rate* of an individual of trait v during phase i,
- $d_v^i \in \mathbb{R}_+$, the *(natural) death rate* of an individual of trait v during phase i,
- $c_{v,w}^i \in \mathbb{R}_+$, the *competition* imposed by an individual of trait w onto an individual of trait v during phase i,
- $K \in \mathbb{N}$, the *carrying capacity* that scales the environment's capacity to support life,
- $\mu_K \in [0, 1]$, the *probability of mutation* at a birth event (phase-independent),
- $m_{v,\cdot} \in \mathcal{M}_p(V)$, the *law of the trait of a mutant* offspring produced by an individual of trait v (phase-independent).

For simplicity, we focus on the situation of nearest neighbour forward mutation without backwards mutation. That is $m_{v,\cdot} = \delta_{v+1,\cdot}$, for $v \in [\![0, L-1]\!]$, and $m_{L,\cdot} = \delta_{L,\cdot}$, where δ denotes the Dirac measure. Moreover, to ensure logistic growth of the total population and thus in particular non-explosion we assume that $c_{v,v}^i > 0$, for all $v \in V$ and all $i = 1, \dots, \ell$.

To describe the time-dependent environment, we take, for each $i = 1, \dots, \ell$, $T_i > 0$ as the length of the i-th phase and refer to the endpoints of these phases by $T_j^\Sigma := \sum_{i=1}^j T_i$. Now we can define the time-dependent birth rates as the periodic extension of

$$b_v(t) := \sum_{i=1}^{\ell} \mathbb{1}_{t \in [T_{i-1}^\Sigma, T_i^\Sigma)} b_v^i, \tag{C.2.1}$$

and analogously for the death rates $d_v(t)$ and competition rates $c_{v,w}(t)$.

In the following, we consider three scaling parameters. As already mentioned, K denotes the carrying capacity of the environment and will correspond to the typical population size, see below. The probability of mutation at birth is denoted by μ_K and is chosen as a power law $\mu_K = K^{-1/\alpha}$, for some $\alpha \in \mathbb{R}_+ \backslash \mathbb{N}_0$, here. Lastly, we let λ_K describe the time scale on which parameter changes occur. In order for environmental changes to happen slow enough such that the resident populations can adapt, but fast enough such that they influence the growth of mutants, we choose

$$1 \ll \lambda_K \ll \ln K \tag{C.2.2}$$

as an intermediate scale and set

$$b_v^K(t) := b_v(t/\lambda_K), \quad d_v^K(t) := d_v(t/\lambda_K), \quad \text{and} \quad c_{v,w}^K(t) := c_{v,w}(t/\lambda_K). \tag{C.2.3}$$

This means that the parameters of the i-th phase now apply for a time of rescaled length $T_i\lambda_K$. Note that b_v^i and b_v^K are very similar in notation. To make the distinction clear, we always use the upper index i to refer to the constant parameter in phase i and the index K to refer to the time-dependent parameter function for carrying capacity K, and use the same convention for the other parameters.

For any K, the evolution of the population over time is described by a Markov process N^K with values in $\mathbb{D}(\mathbb{R}_+, \mathbb{N}_0^V)$. $N_v^K(t)$ denotes the number of individuals of trait $v \in V$ that are alive at time $t \geq 0$. The process is characterised by its infinitesimal generator

$$\begin{aligned}\left(\mathcal{L}_t^K \phi\right)(N) = &\sum_{v\in V}(\phi(N+e_v)-\phi(N))\left(N_v b_v^K(t)(1-\mu_K) + \sum_{w\in V} N_w b_w^K(t)\mu_K m_{w,v}\right)\\ &+ \sum_{v\in V}(\phi(N-e_v)-\phi(N))N_v\left(d_v^K(t) + \sum_{w\in V}\frac{c_{v,w}^K(t)}{K}N_w\right),\end{aligned} \tag{C.2.4}$$

where $\phi : \mathbb{N}_0^V \to \mathbb{R}$ is measurable and bounded and e_v denotes the unit vector at $v \in V$.

Dividing the competition kernel by K in the quadratic term of the stated generator leads to a total population size of order K. In the following, we will refer to subpopulations with a size of order K as *macroscopic*, while we call populations with a size of order 1 *microscopic*, and intermediate sizes of order strictly between 1 and K *mesoscopic*. We are interested in studying the typical behaviour of the processes $(N^K, K \in \mathbb{N})$ for large populations (i.e. as $K \to \infty$). A classical law of large numbers result states that the rescaled processes N^K/K converge on finite time intervals to the solution of a system of Lotka-Volterra equations.

$$\dot{n}_v(t) = \left(b_v^i - d_v^i - \sum_{w\in V} c_{v,w}^i n_w(t)\right) n_v(t), \quad v \in V,\ t \geq 0. \tag{C.2.5}$$

We are interested in the process started with a monomorphic resident population of trait 0, studying the transition towards a new monomorphic subpopulation of trait L. This means that, apart from the invasion phase, only one single (fit) subpopulation is of macroscopic size and fluctuates around its equilibrium size. Taking into account the phase-dependent parameters, we denote these *monomorphic equilibria* by

$$\bar{n}_v^i := \frac{b_v^i - d_v^i}{c_{v,v}^i}, \qquad v \in V,\ i = 1, \dots, \ell, \tag{C.2.6}$$

and introduce the corresponding time-dependent versions

$$\bar{n}_v(t) := \sum_{i=1}^{\ell} \mathbb{1}_{t\in[T_{i-1}^\Sigma, T_i^\Sigma)} \bar{n}_v^i \qquad \text{and} \qquad \bar{n}_v^K(t) = \bar{n}_v(t/\lambda_K). \tag{C.2.7}$$

Starting with such a monomorphic equilibrium, a natural question is to ask for the approximate growth rate of a smaller population of different trait w in the presence of the bulk population of trait v. This leads to the concept of invasion fitness.

Definition C.2.1 (Invasion fitness)**.** For each phase $i = 1, \cdots, \ell$ and for all traits $v, w \in V$ such that the equilibrium size of $\bar{n}_v^i$ is positive, we denote by

$$f_{w,v}^i := b_w^i - d_w^i - c_{w,v}^i \bar{n}_v^i \tag{C.2.8}$$

the *invasion fitness* of trait w with respect to the monomorphic resident v in the i-th phase. Moreover, we define the time-dependent fitness and its rescaled version by the periodic extension of

$$f_{w,v}(t) := \sum_{i=1}^{\ell} \mathbb{1}_{t \in [T_{i-1}^{\Sigma}, T_i^{\Sigma})} f_{w,v}^i \quad \text{and} \quad f_{w,v}^K(t) := f_{w,v}(t/\lambda_K). \tag{C.2.9}$$

C.2.2. Main Result 1: Strict fitness valley

Our aim is to study the crossing of a fitness valley of length L. By this we mean to start initially with a monomorphic wild-type population of trait 0, near its equilibrium $\bar{n}_0^1 K$, and wait until mutants have transitioned through a number of unfit intermediate traits to eventually produce a mutant of trait L that forms a subpopulation of macroscopic order K and replaces the wild-typ as the resident trait. To depict this situation, we fix the initial condition as follows.

Assumption 5 (Initial condition)**.**

(i) $N_0^K(0) = \lfloor \bar{n}_0^1 K \rfloor$,

(ii) $N_v^K(0) = 0$, for all $v \in [\![1, L]\!]$.

Moreover, we introduce the following stopping time that marks the time when the L-trait has taken over the population.

$$T_{\text{inv}}^{(K,\varepsilon)} = \inf \left\{ t \geq 0 : \left| \frac{N_L^K(t)}{K} - \bar{n}_L^K(t) \right| < \varepsilon \text{ and } \frac{1}{K} \sum_{j=0}^{L-1} N_j^K(t) < \varepsilon \right\}. \tag{C.2.10}$$

To ensure that an L-mutant subpopulation is able to fixate and invade in a phase when it is fit with respect to the resident 0-trait, we make the following assumptions.

Assumption 6 (Guaranteed invasion)**.**

(i) $f_{0,L}^i < 0$, whenever $f_{L,0}^i > 0$,

(ii) $f_{L,0}^i \neq 0$, for all $i = 1, \ldots, \ell$.

Note that while the first part of the assumption prevents coexistence, the second part is only technical and avoids the situation of critical branching process approximations.

To precisely define the notion of a fitness valley, let us note that, as shown in [71], the growth of a mutant subpopulation is effectively driven by its *average fitness*

$$f_{v,0}^{\mathrm{av}} := \frac{\sum_{i=1}^{\ell} f_{v,0}^{i} T_i}{T_\ell^\Sigma}. \tag{C.2.11}$$

One might now simply require this quantity to be negative for all intermediate traits in $[\![1, L-1]\!]$ to define a fitness valley. However, a negative average fitness only prevents long-term growth on the $\ln K$-time scale, as studied in [71]. On the λ_K-time scale of environmental changes, there might still be phases i of positive invasion fitness $f_{v,0}^{i} > 0$, for some trait $v \in [\![1, L-1]\!]$, which would allow for temporary growth to a mesoscopic size of this mutant subpopulation. Such a short-term growth significantly complicates the study of a fitness valley transition. We therefore distinguish two scenarios: Our first result is restricted to the case of a *strict fitness valley* in the sense that the traits within the valley are unfit in every phase (cf. Assumption 7). In the second result we then present an extension by allowing exactly one trait to have positive fitness in one phase (cf. Assumption 8) and call this conditions a *pit stop*.

Assumption 7 (Strict fitness valley)**.**

(i) $\bar{n}_0^i > 0$, for all $i = 1, \dots, \ell$,

(ii) $f_{w,0}^{i} < 0$, for all $w \in [\![1, L-1]\!]$ and all $i = 1, \dots, \ell$,

(iii) $f_{L,0}^{\mathrm{av}} > 0$.

As outlined in the heuristics in Section C.3.1, the crossing of the fitness valley is very rare but itself a fast event. Therefore, we can treat it phase by phase and define the phase-dependent crossing rates, for $i = 1 \dots \ell$,

$$R_L^i := \bar{n}_0^i \left(\prod_{v=1}^{\lfloor \alpha \rfloor} \frac{b_{v-1}^i}{\left| f_{v,0}^i \right|} \right) b_{\lfloor \alpha \rfloor}^i \left(\prod_{w=\lfloor \alpha \rfloor+1}^{L-1} \lambda(\rho_w^i) \right) \frac{\left(f_{L,0}^i \right)_+}{b_L^i}, \tag{C.2.12}$$

where

$$\rho_w^i = \frac{b_w^i}{b_w^i + d_w^i + c_{w,0}^i \bar{n}_0^i} \quad \text{and} \quad \lambda(\rho_w^i) = \frac{\rho_w^i}{1 - 2\rho_w^i} = \frac{b_w^i}{\left| f_{w,0}^i \right|}. \tag{C.2.13}$$

The *effective crossing rate* is then given by

$$R_L^{\mathrm{eff}} = \frac{1}{T_\ell^\Sigma} \int_0^{T_\ell^\Sigma} \left(\sum_{i=1}^{\ell} R_L^i \mathbb{1}_{t \in [T_{i-1}^\Sigma, T_i^\Sigma)} \right) \mathbb{1}_{t \in A} \mathrm{d}t, \tag{C.2.14}$$

where A denotes the set of possible arrival times of successful mutants and is given by

$$A := \left\{ t \geq 0 : \int_t^{t+s} f_{L,0}(u) \mathrm{d}u > 0 \ \forall s \in (0, T_\ell^\Sigma] \right\}. \tag{C.2.15}$$

Again, we refer to Section C.3.1 for a heuristic explanation of these rates and the corresponding time scale for crossing the fitness valley.

Using the above notation, we can describe the crossing times of a strict fitness valley as follows.

Theorem C.2.2. *Suppose that Assumptions 5, 6, and 7 is satisfied. Then there exist* $\varepsilon_0 > 0$ *and* $c \in (0, \infty)$ *such that, for all* $0 < \varepsilon < \varepsilon_0$*, there are exponential random variables* $E^{(K,\pm)}(\varepsilon)$ *with parameters* $(1 \pm c\varepsilon) R_L^{\text{eff}}$ *such that*

$$\liminf_{K \to \infty} \mathbb{P}\left(E^{(K,-)}(\varepsilon) \leq T_{\text{inv}}^{(K,\varepsilon)} K \mu_K^L \leq E^{(K,+)}(\varepsilon) \right) \geq 1 - c\varepsilon. \tag{C.2.16}$$

Remark 23. Originally, it was introduced $\lambda(\rho) = \sum_{k=1}^{\infty} \frac{(2k)!}{(k-1)!(k+1)!} \rho^k (1-\rho)^{k+1}$ in [29], which incorporates the combinatorial origin related to the number of birth events in a subcritical branching process excursion. We decide for the simpler representation here, as it points out the similarity to the other factors. Using complex integration, one can show that both definitions are equivalent.

C.2.3. Main Result 2: Valley with pit stop

After the analysis of the crossing of a strict fitness valley in the previous section, it is natural to ask how we can extend this result to more general fitness landscapes. In order to stay in the setting of a fitness valley, we still ask for the traits within the valley to be unfit in the sense of average fitness, i.e. $f_{v,0}^{\text{av}} < 0$, for all $v \in [\![1, L-1]\!]$. In contast to the previous setting, this does allow for a positive invasion fitness of intermediate traits in the valley for some phases. Since this little change leads to a totally different development of the crossing, we keep the situation manageable by restricting to an environment changing only between two different phases and allowing only one stand-out trait w in the valley to be fit in one of the phases. Moreover, we assume that the equilibrium size of the wild-type trait 0 is the same in both phases. In Section C.3, we discuss some conjectures of how these assumptions might be relaxed in future work.

Assumption 8 (Fitness valley with pit stop)**.**

(0) $\ell = 2$,

(i) $\bar{n}_0^1 = \bar{n}_0^2 > 0$,

(ii) $f_{w,0}^1 > 0$, $f_{w,0}^{\text{av}} < 0$, for a unique $w \in [\![\lfloor \alpha \rfloor + 1, L-1]\!]$, and
$f_{v,0}^i < 0$, for all $v \in [\![1, L-1]\!] \setminus \{w\}$ and $i = 1, 2$,

(iii) $f_{L,0}^i > 0$, for $i = 1, 2$.

The short but significant growth phases of trait w in phase 1, before going extinct again in phase 2, give rise to a partially changed crossing rate,

$$\begin{aligned} R_L^{\text{pitstop}} = \bar{n}_0^1 \left(\prod_{v=1}^{\lfloor\alpha\rfloor} \frac{b^1_{v-1}}{\left|f^1_{v,0}\right|} \right) b^1_{\lfloor\alpha\rfloor} \left(\prod_{z=\lfloor\alpha\rfloor+1}^{w-1} \lambda(\rho^1_z) \right) \frac{1}{f^1_{w,0}} \\ \times \left[\frac{b^1_w}{f^1_{w,0}} \left(\prod_{z=w+1}^{L-1} \lambda(\rho^1_z) \right) \frac{f^1_{L,0}}{b^1_L} + \frac{b^2_w}{|f^2_{w,0}|} \left(\prod_{z=w+1}^{L-1} \lambda(\rho^2_z) \right) \frac{f^2_{L,0}}{b^2_L} \right] \frac{1}{T_2^{\Sigma}}. \end{aligned} \tag{C.2.17}$$

Moreover, the refreshments at this pit stop causes a speed up of the crossing that is depicted in an additional term in the corresponding time scale. Overall, we obtain the following result.

Theorem C.2.3. *Suppose that Assumptions 5, 6 and 8 is satisfied. Then there exist* $\varepsilon_0 > 0$ *and* $c \in (0, \infty)$ *such that, for all* $0 < \varepsilon < \varepsilon_0$, *there are exponential random variables* $E^{(K,\pm)}(\varepsilon)$ *with parameter* $(1 \pm c\varepsilon) R_L^{pitstop}$ *such that*

$$\liminf_{K\to\infty} \mathbb{P}\left(E^{(K,-)}(\varepsilon) \leq T_{\text{inv}}^{(K,\varepsilon)} K \mu_K^L e^{\lambda_K T_1 f^1_{w,0}} / \lambda_K \leq E^{(K,+)}(\varepsilon) \right) \geq 1 - c\varepsilon. \tag{C.2.18}$$

A heuristic explanation of the rate and the time scale can be found in Section C.3.1.

C.3. Heuristics and Discussion

The proofs in the field of adaptive dynamics are often quite technical. Therefore, we use this chapter to first provide some heuristics behind the main results of this paper and work out the details in the next chapter. Moreover, we have kept our results in their simplest form to avoid even more technicalities but want to discuss possible extensions or generalizations here.

C.3.1. Explanation of the main results

Theorem C.2.2

We begin by explaining the rational behind the phase-dependent crossing rate in (C.2.12).

Under the assumption that $\alpha < L$, all mutant traits within an α-distance of the initial resident trait 0 (and beyond, up to trait $L - 1$) are initially unfit. As a consequence, their population size is fed by incoming mutants from left neighbors but otherwise declines. During a given phase i, and as long as all mutant traits are small enough such that they essentially do not contribute to competitive interactions, we can hence iteratively estimate their sizes as follows:

The resident trait 0, which does not get any incoming mutants, is approximately at its equilibrium size $N_0^K = \bar{n}_0^i K$. Trait 1 has incoming mutants at rate $N_0^K \cdot b_0^i \mu_K$ and otherwise decays at rate $N_1^K \cdot f_{1,0}^i$, which yields an equilibrium size of $N_1^K = N_0^K b_0^i \mu_K / |f_{1,0}^i| = K\mu_K \bar{n}_0^i b_0^i / |f_{1,0}^i|$. Trait 2 then has incoming mutants at rate $N_1^K \cdot b_1^i \mu_K$ and decays at rate $N_2^K \cdot f_{2,0}^i$, yielding an equilibrium of $N_2^K = K\mu_K^2 \bar{n}_0^i (b_0^i/|f_{1,0}^i|)(b_1^i/|f_{2,0}^i|)$ and so on. Iterating, we obtain an equilibrium of trait $\lfloor \alpha \rfloor$ of

$$N_{\lfloor \alpha \rfloor}^K = K\mu_K^{\lfloor \alpha \rfloor} \bar{n}_0^i \prod_{v=1}^{\lfloor \alpha \rfloor} \frac{b_{v-1}^i}{\left| f_{v,0}^i \right|}. \tag{C.3.1}$$

Note that, since $K\mu_K^{\lfloor \alpha \rfloor} \gg 1$, as $K \to \infty$, all of these traits have a diverging population size and hence an argument via a deterministic approximation can be applied.

As above, trait $\lfloor \alpha \rfloor$ produces mutants of type $\lfloor \alpha \rfloor + 1$ at rate $N_{\lfloor \alpha \rfloor}^K \cdot b_{\lfloor \alpha \rfloor}^i \mu_K$. This rate however is now of order $K\mu_K^{\lfloor \alpha \rfloor + 1} \ll 1$. As a consequence, mutation events are separated and occur on a longer time scale of order $1/K\mu_K^{\lfloor \alpha \rfloor + 1} \gg 1$. Assuming that trait $\lfloor \alpha \rfloor + 1 < L$ is unfit, its population can be approximated by a subcritical birth death process and the descendants of a single arriving mutant go extinct within a finite time of order 1. The only chance for an $\lfloor \alpha \rfloor + 2$ mutant to occur is therefore the unlikely case that the $\lfloor \alpha \rfloor + 1$ population produces a mutant in this order 1 time before its extinction. The probability of this event can be estimated by $\lambda(\rho_{\lfloor \alpha \rfloor + 1}^i)\mu_K$, where $\lambda(\rho_{\lfloor \alpha \rfloor + 1}^i)$ is the expected number of birth events in an excursion of a subcritical birth death process with birth probability of $\rho_{\lfloor \alpha \rfloor + 1}^i$.

In order for an L-mutant to occur across the fitness valley, every mutant trait in between $\lfloor \alpha \rfloor$ and L must produce the next mutant before going extinct in finite time, which has a combined probability of

$$\mu_K^{L - \lfloor \alpha \rfloor - 1} \prod_{w = \lfloor \alpha \rfloor + 1}^{L-1} \lambda(\rho_w^i). \tag{C.3.2}$$

Note that, since extinction occurs within a time of order 1 and phases change on a time scale of order $\lambda_K \gg 1$, this crossing of the fitness valley will take place within a single i-phase and hence all parameters are chosen accordingly.

Finally, if an L-mutant occurs in an i-phase, according to classical branching process theory, it has a chance of $(f_{L,0}^i)_+ / b_L^i$ to initially survive and not go extinct within a finite time due to random fluctuations (or being unfit, which is covered by taking only the positive part of $f_{L,0}^i$ here). Overall, the rate at which successful L-mutants - those that foster an initially growing population - occur in phase i can be found as the product of the rate at which $\lfloor \alpha \rfloor + 1$ mutants occur, times the probability of crossing the valley and producing an L-mutant, times the survival probability of that L-mutant, i.e.

$$K\mu_K^L R_L^i = K\mu_K^L \bar{n}_0^i \left(\prod_{v=1}^{\lfloor \alpha \rfloor} \frac{b_{v-1}^i}{\left| f_{v,0}^i \right|} \right) b_{\lfloor \alpha \rfloor}^i \left(\prod_{w = \lfloor \alpha \rfloor + 1}^{L-1} \lambda(\rho_w^i) \right) \frac{\left(f_{L,0}^i \right)_+}{b_L^i}. \tag{C.3.3}$$

To conclude the effective rate at which an L-mutant occurs and not only initially survives but invades the population - i.e. reaches a size of order K and out-competes the current resident trait - we need to consider the growth dynamics of an L-population over the course of many phases. During an i-phase, the L-population grows approximately at exponential rate $f_{L,0}^i$. Hence, starting with a size of order 1 at time $T\lambda_K$, after a time $S\lambda_K$ the population would have grown to a size of order

$$e^{\int_{T\lambda_K}^{(T+S)\lambda_K} f_{L,0}(t/\lambda_K)\mathrm{d}t} = e^{\lambda_K \int_T^{(T+S)} f_{L,0}(u)\mathrm{d}u}. \tag{C.3.4}$$

To guarantee survival, this order of the population size needs to stay larger than 1 (and in fact almost sure extinction can be shown in the case where it drops below 1), i.e. one needs

$$\int_T^{(T+S)} f_{L,0}(u)\mathrm{d}u > 0. \tag{C.3.5}$$

Since by assumption $f_{L,0}^{\mathrm{av}} > 0$, this can only fail within the first cycle of phases and we therefore introduce the set of possible arrival times of successful L-mutants of

$$A = \left\{ t \geq 0 : \int_t^{t+s} f_{L,0}(u)\mathrm{d}u > 0 \ \forall\ s \in (0, T_\ell^\Sigma] \right\}. \tag{C.3.6}$$

Finally, the effective crossing rate, i.e. the rate at which L-mutants occur, initially survive, and grow to a population size of order K, can be calculated by averaging the phase-dependent rates over a full cycle of phases and taking the above set A into account, which yields

$$K\mu_K^L R_L^{\mathrm{eff}} = \frac{1}{T_\ell^\Sigma} \int_0^{T_\ell^\Sigma} \left(\sum_{i=1}^{\ell} K\mu_K^L R_L^i \mathbb{1}_{t \in [T_{i-1}^\Sigma, T_i^\Sigma)} \right) \mathbb{1}_{t \in A} \mathrm{d}t. \tag{C.3.7}$$

Since this is an exponential rate of order $K\mu_K^L$, the crossing event itself occurs on a time scale of order $1/K\mu_K^L$. The exponential growth of the L-mutant from a population size of order 1 to a size of order K occurs within a $\ln K$-time and the Lotka-Volterra dynamics of the L-mutant taking over the resident population plays out in a time of order 1 once both populations are of the same order. Both of these events are negligible on the $1/K\mu_K^L$ time scale, which leads to Theorem C.2.2.

Theorem C.2.3

We now turn to the case of a fitness valley with a pit stop trait $\lfloor\alpha\rfloor < w < L$ and the heuristics for (C.2.17). For technical reasons, we restrict this result to the case of only two parameter phases, where w is fit during phase 1 and unfit during phase 2. Possible extensions are discussed below.

During phase i, new mutants of trait w occur at approximate rate

$$K\mu_K^w \bar{n}_0^i \left(\prod_{v=1}^{\lfloor\alpha\rfloor} \frac{b_{v-1}^i}{\left| f_{v,0}^i \right|} \right) b_{\lfloor\alpha\rfloor}^i \left(\prod_{z=\lfloor\alpha\rfloor+1}^{w-1} \lambda(\rho_z^i) \right) \tag{C.3.8}$$

and mutants of trait $w+1$ foster a successfully invading L population with probability

$$\mu_K^{L-(w+1)} \left(\prod_{z=w+1}^{L-1} \lambda(\rho_z^i) \right) \frac{\left(f_{L,0}^i\right)_+}{b_L^i}, \tag{C.3.9}$$

as in the previous case of Theorem C.2.2 (here the set A is dropped since L is assumed to be fit in both phases). However, the probability of a trait w mutant fostering a $w+1$ mutant is only $\lambda(\rho_w^i)\mu_K$ in phase $i=2$.

If the w mutant occurs in phase 1, it is temporarily fit and grows exponentially at rate $f_{w,0}^1$ until the next phase change. To get the average/effective rate of crossing the fitness valley, one needs to average the crossing rate over all possible arrival times of the w mutant. Since trait w produces $w+1$ mutants at rate $N_w^K b_w^i \mu_K$, the dominating rate - and hence typical case - occurs when the w population reaches its highest possible population size before becoming subcritical and going extinct again. This is the case when w mutants arise right at the beginning of a phase 1 and hence grow to an approximate size of $e^{\lambda_K T_1 f_{w,0}^1}$, yielding the maximal mutation rate of $e^{\lambda_K T_1 f_{w,0}^1} b_w^i \mu_K$ at its highest peak at the transition from phase 1 to phase 2. Here both values of $i=1,2$ are relevant since the $w+1$ mutant typically arises either right before or after the change from phase 1 to phase 2. The probability of a w mutant to occur right at the beginning of phase 1 is of order $1/\lambda_K$ since arrival times are roughly uniform within a phase. Up to some remaining constants that stem from the averaging integration and that we do not want to discuss in detail here, these heuristics combine to the overall crossing rate of $R_L^{\text{pitstop}} K \mu_K^L e^{\lambda_K T_1 f_{w,0}^1} / \lambda_K$ in (C.2.17), which yields Theorem C.2.3.

On some simplifying assumptions

To simplify the already complicated proofs, we have made some assumptions on the initial condition of population sizes (Assumption 5) and the possible directions of mutations ($m_{v,z} = \delta_{v+1,z}$). These are not necessary assumptions and we want to briefly explain why relaxing them would not change the overall results.

First, we assume that the population starts out with a monomorphic population of trait 0, close to its equilibrium size $K\bar{n}_0^1$. This assumption could be relaxed to a trait 0 population of order K and traits $v \in [\![1, \lfloor \alpha \rfloor]\!]$ of any order smaller or equal to $K\mu_K^v$. In this case, trait 0 gets close to its equilibrium within a time of order 1, following the deterministic single-trait Lotka-Volterra dynamics. Within an additional time of order 1, traits $v \in [\![1, \lfloor \alpha \rfloor]\!]$ also reach their respective (lower-order) equilibria due to incoming mutants from traits $v-1$, see Lemma C.4.2. This order 1 time is negligible on the time scale of our result and the probability of a fitness valley crossing to occur during this time converges to zero.

In addition, we could also allow for positive initial population sizes for traits $v \in [\![\lfloor \alpha \rfloor + 1, L-1]\!]$, as long as they are of order 1, as $K \to \infty$. Any one of these finitely many individual has a probability of producing a successful L-mutant that converges to 0, as $K \to \infty$, and hence the probability of all of their offspring going extinct (in a time of order 1) without crossing the fitness valley converges to 1. The important heuristic here is that the each of the finitely many initial individuals only has a one time shot to cross the valley (through its offspring) that is unlikely to succeed. A successful crossing only occurs through infinitely many of such unlikely attempts, occurring on the diverging time scale of our results.

Finally, note that we do need to assure that $N_L^K(0) = 0$ in order to guarantee that a successful L population stems from a crossing of the fitness valley and does not just start growing immediately.

On another note, we assume that mutation can only occur to neighboring higher traits, i.e. from v to $v+1$. We could allow for backwards mutation, i.e. from v to $v-1$, without changing the outcome of the main results. This is true because the crossing rate of order $K\mu_K^L$ in Theorem C.2.2 and the respective adjusted rate in Theorem C.2.3 stem from tracking mutations along the shortest possible path from 0 to L. Taking a "detour" via forwards and backwards mutation would only add additional factors of μ_K due to additional mutation steps and hence produce lower order crossing rates. When determining the overall crossing rates in this case, one can write them as the sum of rates of L-mutants arising along different paths from 0 to L, where the dominant summands will be exactly the respective rates of our theorems here. We refer to [29, 69] for the precise arguments in the case of constant parameters.

C.3.2. Possible generalizations of the pitstop result

There are a number of ways in which we conjecture Theorem C.2.3 could be extended and that we briefly want to discuss in the following.

Non-constant resident trait

In contrast to Theorem C.2.2, for Theorem C.2.3 we require that $\bar{n}_0^1 = \bar{n}_0^2$ in Assumption 8. We conjecture that the same result is still true for $\bar{n}_0^1 \neq \bar{n}_0^2$, as long as both equilibria are strictly positive. However, this cannot be argued with our current proof techniques for the following reason: In order to ensure the correct order of the crossing rate of $K\mu_K^L e^{\lambda_K T_1 f_{w,0}^1}/\lambda_K$, one needs to approximate the w population by birth death processes with a fitness that only deviates from $f_{w,0}^1$ by an error that vanishes as $K \to \infty$. To do so, we pick a threshold of $\varepsilon_K K$, where $\varepsilon_K \to 0$, to bound both the size of the mutant populations and the deviation between the resident 0 trait and its equilibrium, since these two quantities are the source of errors in the actual fitness of w. Now for Theorem C.2.2, the proof relies on bounding the resident population size N_0^K in two ways. Once the population is close to its equilibrium, potential theoretic arguments are applied to ensure it staying close. Initially after a parameter change however, the approximating deterministic system is used to ensure that the population gets close to its new equilibrium in a negligible time of order 1. If one requires this "closeness" to be of an order $\varepsilon_K K$ for the pitstop result, it would take a diverging time in the deterministic system to be achieved. The classical results for deterministic approximations are however only valid on a finite time scale and, moreover, this adaptation time would now no longer be a negligible order 1 time, during which mutations do not occur with probability 1 (a fact that is necessary to justify using only the equilibrium population sizes in the transition rate).

More than two distinct phases

When there are more than two parameter phases ($\ell > 2$), even if the target mutant trait L remains fit throughout all of them, the description of the crossing rate becomes more intricate and the proof would require a lot more notation. Heuristically, for every specific example, one must determine the corresponding maximal possible population size of the pitstop trait w, which replaces the factor $e^{\lambda_K T_1 f^1_{w,0}}$ in the transition rate in Theorem C.2.3. In accordance with (C.3.4), and setting $g_w(t,s) = \int_t^{t+s} f_{w,0}(u)\mathrm{d}u$, this population size can be written as

$$\max_{t\in[0,T_\ell^\Sigma]} \max_{\substack{s\in[0,T_\ell^\Sigma]: \\ g_w(t,s')>0\ \forall s'\in(0,s]}} e^{\lambda_K g_w(t,s)}. \tag{C.3.10}$$

The maximizers t^* and and s^* correspond to the optimal time $t^*\lambda_K$ of occurrence of a w mutant with a successive growth period of length $s^*\lambda_K$, at the end of which the w population reaches its peak size before shrinking again and eventually going extinct (see Figure C.1 for an exemplary plot). Note that it is possible that the w population temporarily has a negative fitness during this period, as long as it grows to a larger size afterwards and never shrinks to a size of order 1 in between. Moreover, t^* and $t^* + s^*$ will always coincide with beginning and endpoints of fit phases for trait w, respectively, in order to maximize the time of growth.

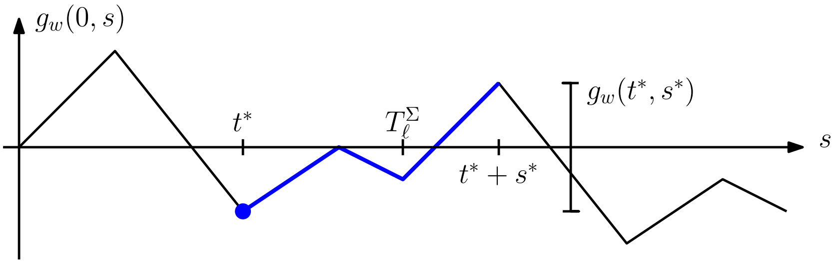


Figure C.1.: Exemplary plot of $g_w(0,s)$ for $\ell = 4$ parameter phases. The blue dot marks the optimal/typical occurrence time t^* of a w mutant to initiate a population reaching its highest possible size. The blue line marks this growth phase, at the end of which (at time $t^* + s^*$), L-mutants are produced with the highest possible rate.

Temporarily unfit trait L

Another possible generalization of the pitstop result is to drop the assumption that trait L is always fit. In this case, one needs to determine the maximal possible population size of the transitional trait w within the times of set A, as defined for Theorem C.2.2 (see heuristics above).

For two parameter phases, there are two distinct scenarios, see Figure C.2 below (where g_L is defined analogously to g_w, using $f_{L,0}(u)$). If the fit phases of trait w and L are asynchronous, the typical transition time from w to L is at the end of the fit phase of trait w, or rather right at the beginning of the fit phase of trait L. This is when the w population has is maximal population size, while the L trait is also fit and has a positive fixation probability. If the fit phases are synchronized, the typical transition time would be at the point within the set A, when w is not at its global maximum but the largest population size that also allows the L-trait to survive the first cycle of phases (and hence long-term) after the transition.

In the general case of more than two phases, the typical transition time from w to L will still either be the time of a phase change and/or at the boundary of the set A. To our knowledge there is no nice and concise general formula to describe this time point and the corresponding population size of w but for any specific case it can be determined similar considerations to the above two-phase examples.

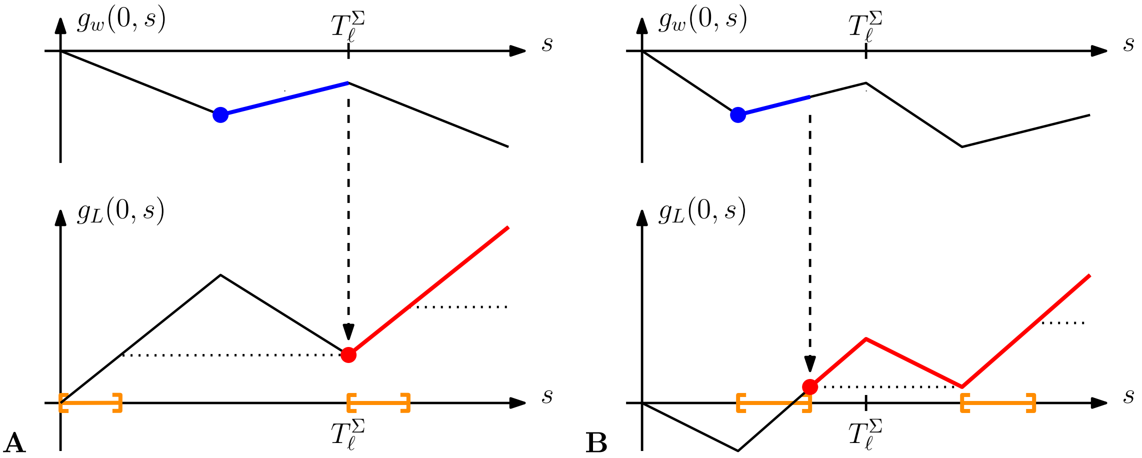


Figure C.2.: Exemplary plot of $g_w(0,s)$ and $g_L(0,s)$ for $\ell = 2$ parameter phases, temporarily unfit trait L and A) asynchronous or B) synchronous fit phases. Blue dots and trajectories mark the optimal/typical occurrence time w mutants and their successive growth phase. Dashed arrows mark the typical transition time to trait L and red dots and trajectories mark the occurrence and growth of L-mutants. The set A of possible arrival times of successful L-mutants is marked in orange.

Multiple pitstops

One could also study a scenario where more than one, e.g. two, intermediate pitstop traits $\lfloor \alpha \rfloor < w_1 < w_2 < L$ exist. In the case of two parameter phases, with L always being fit, there are again two scenarios. If the fit phases of w_1 and w_2 are asynchronous, the considerations are similar to above. The typical time to transition from w_1 to w_2 is the end of the fit phase of w_1, while the typical transition time from w_2 to L is at the end of the fit phase of w_2. Assuming that w_1 is fit in the first phase, this would then lead to a transition time scale of $K\mu_K^L e^{\lambda_K T_1 f^1_{w_1,0}} e^{\lambda_K T_2 f^2_{w_2,0}} / \lambda_K$.

The case of synchronized fit phases becomes more involved. The transition from w_2 to L will still occur at the end of the fit phase. For the transition from w_1 to L however, a late transition in the fit phase would result in a large w_1 population but would not give w_2 much time to grow, while the situation is reversed for an early transition. Which of these is more beneficial (in terms of maximizing the corresponding factor of the transition rate) depends on the relation between $f^1_{w_1,0}$ and $f^1_{w_2,0}$. Essentially, the time span $T_1\lambda_K$ for growth gets split between the two traits $w_{1/2}$ and the highest transition rate is obtained when the trait with the higher fitness grows for almost the full duration of $T_1\lambda_K$. Hence, the corresponding time scale for crossings of the fitness valley ends up as $K\mu_K^L e^{\lambda_K T_1 \max\{f^1_{w_1,0}, f^1_{w_2,0}\}}/\lambda_K$.

Pitstop trait $w < \alpha$

Lastly, one could also consider a pitstop trait $w \in [\![1, \lfloor\alpha\rfloor]\!]$. This will be an interesting topic of future research but will require quite different considerations to the present paper since it is no longer a matter of small excursions of populations before going extinct again. Instead, a temporarily growing w population would also trigger a temporary growth of the neighboring $w+1$ population through mutation, and so on. We hypothesise that, at least under similarly restrictive assumptions of a single fit phase for a single intermediate trait, the transition rate and time scale will look very similar to the one in Theorem C.2.3. This is because again only the peak possible population size of trait w needs to be considered for the dominating rate.

C.3.3. Beyond the valley

The results of this paper only consider the transition of a fitness valley up to the point when the (single) fit trait L beyond the valley takes over the resident population. Similar to the results in [29], one could also consider the following decay and eventual extinction of the remaining traits on the $\ln K$-time scale. To ensure this extinction however, one would need to make the additional assumption of $f^{\text{av}}_{v,L} < 0$, for all $v \in [\![0, L-1]\!]$.

If this assumption is not satisfied, or if there were more traits beyond the valley ($L+1, L+2$, etc), one could apply the results of [71] to study the following dynamics of consecutively invading mutant traits on the $\ln K$-time scale in a changing environment. Note that, in case there are multiple traits L_1, L_2 that have a positive average fitness with respect to trait 0, the shortest fitness valley, i.e. the trait closest to 0, will determine the time scale of the first transition.

These kinds of considerations, as well as the option of a more complicated trait space (e.g. a finite graph instead of a simple line of traits) lead to considerations as in [69], where metastable transitions through fitness valleys of varying width are studied as transitions within a meta graph of evolutionary stable conditions. These results apply to the case of constant model parameters but could be generalised to changing environments as future work.

C.4. Proofs

C.4.1. Proof of Theorem C.2.2

The proof of Theorem C.2.2 is split into several steps:

- First, we ensure that the resident population stays close to its phase-dependent equilibrium size, except for very short adaptation times at the beginning of each phase, yielding bounding functions $\phi_0^{(K,\varepsilon,\pm)}(t)$.
- Next, we show that the subpopulations of traits close to 0 ($v \in [\![1, \lfloor\alpha\rfloor]\!]$) follow a periodic equilibrium $a_v^{(K,\pm)}(t)$, scaled with their respective mesoscopic orders of population sizes $K\mu_K^v$.
- These approximations allows us to precisely determine the rate $\tilde{R}^{(K,\pm)}(t)$ at which single L-mutants arise.
- We then analyze how and under which conditions a single L-mutant can fixate and grow to a macroscopic size $\varepsilon^2 K$.
- Finally, we show that a macroscopic L-mutant quickly outcompetes and replaces the resident trait 0.
- Combining these steps allows for the computation of the overall time scale $1/K\mu_K^L$ and effective rate R_L^{eff} of crossing the fitness valley.

Resident stability

To bound the population size of the resident trait $v = 0$, we define the threshold-functions

$$\phi_0^{(K,\varepsilon,+)}(t) = \begin{cases} \max\{\bar{n}_0^{i-1}, \bar{n}_0^i\} + M\varepsilon & \text{if } t \in (T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_\varepsilon), \\ \bar{n}_0^i + M\varepsilon & \text{if } t \in [T_{i-1}^\Sigma \lambda_K + T_\varepsilon, T_i^\Sigma \lambda_K], \end{cases} \tag{C.4.1}$$

$$\phi_0^{(K,\varepsilon,-)}(t) = \begin{cases} \min\{\bar{n}_0^{i-1}, \bar{n}_0^i\} - M\varepsilon & \text{if } t \in (T_{i-1}^\Sigma \lambda_K, T_{i-1}^\Sigma \lambda_K + T_\varepsilon), \\ \bar{n}_0^i - M\varepsilon & \text{if } t \in [T_{i-1}^\Sigma \lambda_K + T_\varepsilon, T_i^\Sigma \lambda_K], \end{cases} \tag{C.4.2}$$

with periodic extension, where, for $i = 1$, $\bar{n}_v^{i-1} := \bar{n}_v^\ell$, and $\phi_0^{(K,\varepsilon,\pm)}(0) = \bar{n}_0^1 \pm M\varepsilon$. Note that these functions also depend on the choices of M and T_ε. To simplify notation, we however do not include those parameters in the functions' names. We denote the first time that these bounds on the resident 0-population fail by

$$T_\phi^{(K,\varepsilon)} = \inf\left\{ t \geq 0 : \frac{N_0^K(t)}{K} \notin [\phi_0^{(K,\varepsilon,-)}(t), \phi_0^{(K,\varepsilon,+)}(t)] \right\}. \tag{C.4.3}$$

To mark the time at which the mutant populations become too large and start to significantly perturb the system, we moreover introduce the stopping time

$$S^{(K,\varepsilon)} := \inf\left\{ t \geq 0 : \sum_{w \neq 0} N_w^K(t) \geq \varepsilon K \right\}. \tag{C.4.4}$$

With this notation, the resident's stability result can be stated as follows.

Lemma C.4.1. *There exists a uniform $M < \infty$ and, for all $\varepsilon > 0$ small enough, there exists a deterministic $T_\varepsilon < \infty$ such that, for all $T < \infty$,*

$$\lim_{K\to\infty} \mathbb{P}\left(T_\phi^{(K,\varepsilon)} \leq \frac{T}{K\mu_K^L} \wedge S^{(K,\varepsilon)} \middle| \frac{N_0^K(0)}{K} \in [\phi_0^{(K,\varepsilon,-)}(0)+\varepsilon, \phi_0^{(K,\varepsilon,+)}(0)-\varepsilon]\right) = 0. \quad \text{(C.4.5)}$$

Proof. We can proceed exactly as in the proof of [71, Theorem 4.1] and make use of the improved estimates of Corollaries C.5.2 and C.5.5 (replacing Theorems A.2 and A.3 in [71]) to concatenate the increased number of phases due to the longer time horizon. □

Equilibrium of mesoscopic traits

Despite the negative fitness of the traits $v \in [\![1, L-1]\!]$ inside the valley, we can observe non-vanishing subpopulation of the traits $v \in [\![1, \alpha]\!]$ that are close to the resident trait 0. This is due to the frequent influx of new mutants. Because of the changing environment, these populations vary in size over time. By the following lemma, we can determine not only their order of population size, which only depends on the mutational distance from the resident, but also their exact equilibrium size that is reached (up to a small error) within each phase.

Lemma C.4.2 (Equilibrium size of mesoscopic traits)**.** *Fix $\varepsilon > 0$, let the initial condition be given by Assumption 5 and let the fitness landscape satisfy either Assumption 7 or Assumption 8. Then, for all $v \in [\![0, \lfloor\alpha\rfloor]\!]$, there exist constants $c_v, C_v^\pm, \tau_v^\varepsilon \in [0,\infty)$ and Markov processes $\left(N_v^{(K,\pm)}(t), t \geq 0\right)_{K\geq 1}$ such that, for all $T < \infty$,*

$$\lim_{K\to\infty} \mathbb{P}\left(\forall t \in (0, (T/K\mu_K^L) \wedge S^{(K,\varepsilon)}),\ \forall v \in [\![0, \lfloor\alpha\rfloor]\!] :\ N_v^{(K,-)}(t) \leq N_v^K(t) \leq N_v^{(K,+)}(t)\right) = 1 \quad \text{(C.4.6)}$$

and

$$a_v^{(K,-)}(t)K\mu_K^v \leq \mathbb{E}\left[N_v^{(K,-)}(t)\right] \leq \mathbb{E}\left[N_v^{(K,+)}(t)\right] \leq a_v^{(K,+)}(t)K\mu_K^v, \quad \text{(C.4.7)}$$

where the bounding functions are the periodic extensions of

$$a_v^{(K,\pm)}(t) = \begin{cases} C_v^\pm & : t \in [\lambda_K T_{i-1}^\Sigma, \lambda_K T_{i-1}^\Sigma + \sum_{w=0}^v \tau_w^\varepsilon), \\ a_v^{(i,\pm)} & : t \in [\lambda_K T_{i-1}^\Sigma + \sum_{w=0}^v \tau_w^\varepsilon, \lambda_K T_i^\Sigma), \end{cases} \quad \text{(C.4.8)}$$

$$a_v^{(i,\pm)} = (1 \pm c_v\varepsilon)\bar{n}_0^i \prod_{w=1}^v \frac{b_{w-1}^i}{\left|f_{w,0}^i\right|}. \quad \text{(C.4.9)}$$

Remark 24. Note that for initial conditions $N_v^K(0) = 0$, $v \in [\![1, L]\!]$, we need to choose $C_v^- = 0$ in (C.4.8). After the first phase however, C_v^- can be chosen as strictly positive. Since the number of traits is finite, the constants $C_v^\pm, c_v$ can be chosen uniformly for all traits by simply taking the minimum and the maximum, respectively. The same is true for the times $\sum_{w=0}^v \tau_w^\varepsilon$, $v \in [\![0, \lfloor\alpha\rfloor]\!]$. We therefore no longer indicate this trait dependence when applying Lemma C.4.2 in the following considerations.

Proof. This proof follows the strategy of [29] and [69], which goes as follows: We first define the event, on which we have good estimates on the resident population size for a sufficiently large time horizon. Then we represent the process by an explicit construction involving Poisson measures and use this to introduce the estimating processes $N_v^{(K,\pm)}$ by couplings. Finally, we deduce an ODE for the expectation of the coupled processes that can be solved approximately to derive the desired bounds.

Let us first define the event on which we have good control on the resident population

$$\Omega^K := \left\{(T/K\mu_K^L) \wedge S^{(K,\varepsilon)} < T_\phi^{(K,\varepsilon)}\right\}. \tag{C.4.10}$$

Since Lemma C.4.1 states that $\lim_{K\to\infty} \mathbb{P}\left(\Omega^K\right) = 1$, we can restrict our considerations to this event for the remainder of this proof. Moreover, this already provides the desired bounds for $v = 0$ with $N_0^{(K,\pm)} = N_0^K$, $\tau_0^\varepsilon = T_\varepsilon$ from Lemma C.4.1, and appropriate choices of c_0, $C_0^\pm$.

To define the coupled processes, we follow the notation of [83] and give an explicit construction of the population process in terms of Poisson random measures. Let $(Q_v^{(b)}, Q_v^{(d)}, Q_{w,v}^{(m)} : v, w \in V)$ be independent homogeneous Poisson random measures on $\mathbb{R}^2$ with intensity $\mathrm{d}s\mathrm{d}\theta$. Then we can write

$$\begin{aligned} N_v^K(t) = N_v^K(0) &+ \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le b_v^K(s)(1-\mu_K)N_v^K(s-)} Q_v^{(b)}(\mathrm{d}s, \mathrm{d}\theta) \\ &- \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le [d_v^K(s) + \sum_{w\in V} c_{v,w}^K(s) N_w^K(s-)/K] N_v^K(s-)} Q_v^{(d)}(\mathrm{d}s, \mathrm{d}\theta) \\ &+ \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le \mu_K b_{v-1}^K(s) N_{v-1}^K(s-)} Q_{v-1,v}^{(m)}(\mathrm{d}s, \mathrm{d}\theta). \end{aligned} \tag{C.4.11}$$

Using the shorthand notation $\check{c}_v := \max_{w\in V\setminus 0,\ i=1,\dots,\ell} c_{v,w}^i$ and the same Poisson measures as before, we inductively, for $v \in [\![1, \lfloor\alpha\rfloor]\!]$, introduce the coupled processes

$$\begin{aligned} N_v^{(K,-)}(t) = N_v^K(0) &+ \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le b_v^K(s)(1-\varepsilon)N_v^{(K,-)}(s-)} Q_v^{(b)}(\mathrm{d}s, \mathrm{d}\theta) \\ &- \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \le [d_v^K(s) + c_{v,0}^K(s)\phi_0^{(K,\varepsilon,+)}(s) + \varepsilon\check{c}_v] N_v^{(K,-)}(s-)} Q_v^{(d)}(\mathrm{d}s, \mathrm{d}\theta) \\ &+ \int_0^t \int_{\mathbb{R}_-} \mathbb{1}_{\theta \le \mu_K b_{v-1}^K(s) N_{v-1}^{(K,-)}(s-)} Q_{v-1,v}^{(m)}(\mathrm{d}s, \mathrm{d}\theta) \end{aligned} \tag{C.4.12}$$

and

$$\begin{aligned} N_v^{(K,+)}(t) = N_v^K(0) &+ \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq b_v^K(s) N_v^{(K,+)}(s-)} Q_v^{(b)}(\mathrm{d}s, \mathrm{d}\theta) \\ &- \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq [d_v^K(s) + c_{v,0}^K(s)\phi_0^{(K,\varepsilon,-)}(s)] N_v^{(K,+)}(s-)} Q_v^{(d)}(\mathrm{d}s, \mathrm{d}\theta) \\ &+ \int_0^t \int_{\mathbb{R}_-} \mathbb{1}_{\theta \leq \mu_K b_{v-1}^K(s) N_{v-1}^{(K,+)}(s-)} Q_{v-1,v}^{(m)}(\mathrm{d}s, \mathrm{d}\theta). \end{aligned} \tag{C.4.13}$$

Restricting to the event Ω^K and times up to $S^{(K,\varepsilon)}$, these coupled processes then satisfy

$$N_v^{(K,-)}(t) \leq N_v^K(t) \leq N_v^{(K,+)}(t), \qquad \forall v \in [\![1, \lfloor \alpha \rfloor]\!], \tag{C.4.14}$$

for K large enough such that $\mu_K < \varepsilon$.

On closer inspection, the approximating processes $N_v^{(K,-)}, N_v^{(K,+)}$ are nothing but subcritical birth death processes with immigration stemming form incoming mutations.

Similar to the proof of [29, Equation (7.8) et seq.], we can use the martingale decomposition of $N_v^{(K,+)}$ and $N_v^{(K,-)}$ to derive the differential equation

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[N_v^{(K,*)}(t)\right] &= \left(b_v^K(t)(1 - \mathbb{1}_{\{*=-\}}\varepsilon) - d_v^K(t) - c_{v,0}^K(t)\phi_0^{(K,\varepsilon,\bar{*})}(t) - \mathbb{1}_{\{*=-\}}\varepsilon \check{c}_v\right) \times \mathbb{E}\left[N_v^{(K,*)}(t)\right] \\ &\quad + \mu_K b_{v-1}^K(t) \mathbb{E}\left[N_{v-1}^{(K,*)}(t)\right] \\ &= f_{v,0}^{(K,*)}(t) \mathbb{E}\left[N_v^{(K,*)}(t)\right] + \mu_K b_{v-1}^K(t) \mathbb{E}\left[N_{v-1}^{(K,*)}(t)\right], \end{aligned} \tag{C.4.15}$$

where $\bar{*} = \{+,-\} \backslash *$ denotes the inverse sign. Moreover, we introduce $f_{v,0}^{(K,*)}(t)$ as a shorthand notation for the first factor to indicate that this is nothing but a perturbation of the invasion fitness $f_{v,0}^K(t)$.

The solution to this ODE is generally given in a closed form by the variation of constants formula. However, it makes more sense here to study the solution phase by phase and use the estimates we already have. To this end, assume that we had shown (C.4.7) already for the sub-population of trait $v-1$, for all times $t \in [0, \infty)$, and for the trait under observation $v < \alpha$ up to time $\lambda_K T_{i-1}^{\Sigma}$, for some $1 \leq i \leq \ell$, which is the beginning of the i-th phase. We now show that it also holds true for trait $v < \alpha$ during the interval $[\lambda_K T_{i-1}^{\Sigma}, \lambda_K T_i^{\Sigma})$. Since we only have rough bounds on the ancestor $v-1$ and the resident 0 populations at the beginning of the phase, up to time $\lambda_K T_{i-1}^{\Sigma} + \sum_{w=0}^{v-1} \tau_w^{\varepsilon}$, the ODE for the upper bound can be estimated by

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[N_v^{(K,+)}(t)\right] \leq \left(b_v^i - d_v^i - c_{v,0}^i \left((\bar{n}_0^{i-1} \wedge \bar{n}_0^i) - M\varepsilon\right)\right) \mathbb{E}\left[N_v^{(K,+)}(t)\right] + b_{v-1}^i C_{v-1}^+ K \mu_K^v, \tag{C.4.16}$$

with initial condition

$$\mathbb{E}\left[N_v^{(K,+)}(\lambda_K T_{i-1}^{\Sigma})\right] \leq a_v^{(i-1,+)} K \mu_K^v. \tag{C.4.17}$$

This implies at most exponential growth for a finite time and thus we can bound the expectation of $N_v^{(K,+)}$ at the beginning of the phase by

$$\mathbb{E}\left[N_v^{(K,+)}(t)\right] \leq C_v^+ K\mu_K^v, \qquad t \leq \lambda_K T_{i-1}^\Sigma + \sum_{w=0}^{v-1} \tau_w^\varepsilon. \tag{C.4.18}$$

From this time on until the end of the i-th phase, we have good bounds on both the resident and the ancestor. Hence the ODE for the upper bound reads as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left[N_v^{(K,+)}(t)\right] \leq \left(f_{v,0}^i + \varepsilon C\right)\mathbb{E}\left[N_v^{(K,+)}(t)\right] + b_{v-1}^i a_{v-1}^{(i,+)} K\mu_K^v. \tag{C.4.19}$$

Together with the estimate on the initial condition, this gives

$$\begin{aligned}\mathbb{E}\left[N_v^{(K,+)}(t)\right] \leq \mathrm{e}^{\left(f_{v,0}^i+\varepsilon C\right)\left(t-\left(\lambda_K T_{i-1}^\Sigma+\sum_{w=0}^{v-1}\tau_w^\varepsilon\right)\right)} &\left(C_v^+ - \frac{b_{v-1}^i}{\left|f_{v,0}^i+\varepsilon C\right|}a_{v-1}^{(i,+)}\right)K\mu_K^v \\ &+ \frac{b_{v-1}^i}{\left|f_{v,0}^i+\varepsilon C\right|}a_{v-1}^{(i,+)}K\mu_K^v.\end{aligned} \tag{C.4.20}$$

Note that the term in brackets can be bounded uniformly, for ε small enough, and is independent of K. Together with the fact that the perturbed fitness $f_{v,0}^i + \varepsilon C < 0$ is still negative, for ε small enough, the first summand can be made smaller than $\varepsilon K\mu_K^v$ by waiting an additional finite time $\tau_v^\varepsilon < \infty$. Finally one just has to take c_v slightly larger than c_{v-1} to bound this small term and the perturbation of the fitness to achieve the claim

$$\mathbb{E}\left[N_v^{(K,+)}(t)\right] \leq a_v^{(i,+)} K\mu_K^v, \qquad t \geq \lambda_K T_{i-1}^\Sigma + \sum_{w=0}^{v} \tau_w^\varepsilon. \tag{C.4.21}$$

Note that during the additional time of length τ_v^ε, we can still use the rough bound instead, potentially taking C_v^+ a bit larger. This procedure can now be continued periodically for times $t \geq \lambda_K T_\ell^\Sigma$. Moreover, the estimates for the $N_v^{(K,-)}(t)$ follow exactly the same steps, using the lower bounds for all relevant parameters. □

Crossing the fitness valley

To see a successful invasion of the mutant trait L, several attempts of crossing the fitness valley might be necessary. We track this carefully by introducing the processes

$$M_v^K(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq \mu_K b_{v-1}^K(s) N_{v-1}^K(s^-)} Q_{v-1,v}^{(m)}(\mathrm{d}s, \mathrm{d}\theta), \tag{C.4.22}$$

which are the cumulative numbers of mutant individuals of trait v that arose as mutants of the progenitor trait $v-1$, as well as the respective occurrence times of these mutants,

$$T_{v,j}^K := \inf\left\{t \geq 0 : M_v^K(t) \geq j\right\}. \tag{C.4.23}$$

Lemma C.4.3. *Fix $\varepsilon > 0$, let the initial condition be given by Assumption 5 and let the fitness landscape satisfy Assumption 7. Then there exist constants $0 < c, C < \infty$ (independent of ε) such that, for each $K \in \mathbb{N}$, there exist two Poisson counting processes $M^{(K,\pm)}$ with intensity functions $t \mapsto \tilde{R}^{(K,\pm)}(t)K\mu_K^L$ such that, for all $T < \infty$,*

$$\liminf_{K\to\infty} \mathbb{P}\left(\forall t \in [0, (T/K\mu_K^L) \wedge S^{(K,\varepsilon)}) : \ M^{(K,-)}(t) < M_L^K(t) < M^{(K,+)}(t)\right) \geq 1 - c\varepsilon, \tag{C.4.24}$$

where the rescaled intensity functions are given by

$$\tilde{R}^{(K,\pm)}(t) = a^{(K,\pm)}_{\lfloor\alpha\rfloor}(t) b^K_{\lfloor\alpha\rfloor}(t) \prod_{w=\lfloor\alpha\rfloor+1}^{L-1} \frac{b_w^K(t)}{|f_{w,0}^K(t)|}(1 \pm C\varepsilon). \tag{C.4.25}$$

Proof. We apply the same arguments as previously used in the case of a constant environment (cf. [29, Ch. 7.3]). In order to reduce to this situation, we have to first ensure that, with high probability, the mutants of type $\lfloor\alpha\rfloor + 1$ appear after the living populations of types $0, ..., \lfloor\alpha\rfloor$ have adapted to the new environment in a particular phase. The second step is then to show that, in the case of a successful cascade of accumulating mutations, the mutant of trait L is born before the environment changes again.

On the event Ω_K, defined in the proof of Lemma C.4.2, we bound the mutant counting process of trait $\lfloor\alpha\rfloor + 1$ by

$$M^{(K,-)}_{\lfloor\alpha\rfloor+1}(t) \leq M^K_{\lfloor\alpha\rfloor+1}(t) \leq M^{(K,+)}_{\lfloor\alpha\rfloor+1}(t), \qquad \forall t \leq (T/K\mu_K^L) \wedge S^{(K,\varepsilon)}, \tag{C.4.26}$$

where the bounding processes are given by

$$M^{(K,\pm)}_{\lfloor\alpha\rfloor+1}(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq \mu_K b^K_{\lfloor\alpha\rfloor}(s) N^{(K,\pm)}_{\lfloor\alpha\rfloor}(s-)} Q^{(m)}_{\lfloor\alpha\rfloor,\lfloor\alpha\rfloor+1}(\mathrm{d}s, \mathrm{d}\theta). \tag{C.4.27}$$

Note that, in contrast to $M^K_{\lfloor\alpha\rfloor+1}$, this definition is based on the bounding processes $N^{(K,\pm)}_{\lfloor\alpha\rfloor}$ from Lemma C.4.2.

As explained in detail in [29, 69], for the following considerations it is sufficient to continue with a simplified version of these processes, based on the expectation of $N^{(K,\pm)}_{\lfloor\alpha\rfloor}$,

$$\bar{M}^{(K,\pm)}_{\lfloor\alpha\rfloor+1}(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\theta \leq \mu_K b^K_{\lfloor\alpha\rfloor}(s) \mathbb{E}\left[N^{(K,\pm)}_{\lfloor\alpha\rfloor}(s-)\right]} Q^{(m)}_{\lfloor\alpha\rfloor,\lfloor\alpha\rfloor+1}(\mathrm{d}s, \mathrm{d}\theta) \tag{C.4.28}$$

and

$$\bar{T}^{(K,\pm)}_{\lfloor\alpha\rfloor+1,j} := \inf\left\{t \geq 0 : \bar{M}^{(K,\pm)}_{\lfloor\alpha\rfloor+1}(t) \geq j\right\}, \tag{C.4.29}$$

since they do not differ too much from the original processes, particularly on the considered time scales. For details, see [29, p. 3583]. Lemma C.4.2 guarantees, that $\bar{M}^{(K,\pm)}_{\lfloor\alpha\rfloor+1}$ are Poisson counting processes with intensity functions bounded by $b^K_{\lfloor\alpha\rfloor}(t) a^{(K,\pm)}_{\lfloor\alpha\rfloor}(t) K\mu_K^{\lfloor\alpha\rfloor+1}$. Moreover, we know that these functions are constant for phases with length of order $O(\lambda_K)$, while the short adaptation intervals after an environmental change are only of order $O(1)$. Therefore, the number of possible mutants appearing within these adaptation intervals is negligible compared to the ones falling into the long constant phases, as $K \to \infty$.

Now at each time $\bar{T}^{(K,\pm)}_{\lfloor\alpha\rfloor+1,j}$ an individual of trait $\lfloor\alpha\rfloor+1$ is born, e.g. during an i-th phase, and its descendant population can be approximated by classical sub-critical birth death processes with constant rates

$$b^{(i,+)}_{\lfloor\alpha\rfloor+1} = b^i_{\lfloor\alpha\rfloor+1}, \qquad d^{(i,+)}_{\lfloor\alpha\rfloor+1} = d^i_{\lfloor\alpha\rfloor+1} + c^i_{\lfloor\alpha\rfloor+1,0}(\bar{n}^i_0 - M\varepsilon), \tag{C.4.30}$$

$$b^{(i,-)}_{\lfloor\alpha\rfloor+1} = b^i_{\lfloor\alpha\rfloor+1}(1-\varepsilon), \qquad d^{(i,-)}_{\lfloor\alpha\rfloor+1} = d^i_{\lfloor\alpha\rfloor+1} + c^i_{\lfloor\alpha\rfloor+1,0}(\bar{n}^i_0 + M\varepsilon) + \varepsilon\check{c}_{\lfloor\alpha\rfloor+1}. \tag{C.4.31}$$

Here we utilise that such processes go extinct within a time of order $O(1)$ almost surely, i.e. before the next phase change, and hence the parameters can be assumed to be constant. This approximation allows us to continue exactly as in [29] and apply Lemma C.5.6, which shows that a single mutant of trait $\lfloor\alpha\rfloor+2$ is produced before the family of trait $\lfloor\alpha\rfloor+1$ goes extinct with probability

$$\mu_K \frac{b^{(i,\pm)}_{\lfloor\alpha\rfloor+1}}{d^{(i,\pm)}_{\lfloor\alpha\rfloor+1} - b^{(i,\pm)}_{\lfloor\alpha\rfloor+1}}, \tag{C.4.32}$$

while the probability of two or more such mutants is of smaller order $O(\mu_K^2)$. Since the total excursion of the trait $\lfloor\alpha\rfloor+1$-population only lasts a time of order $O(1)$, we conclude that an $\lfloor\alpha\rfloor+2$-mutant, if it arises, does so shortly after $\bar{T}^{(K,\pm)}_{\lfloor\alpha\rfloor+1,j}$ and we can assume the same constant phase-i-environment also for its descendants. Iterating this thinning mechanism for the whole cascade of mutations from trait $\lfloor\alpha\rfloor+1$ to trait L then yields that a mutant of trait $\lfloor\alpha\rfloor+1$ leads to a mutant of trait L with probability

$$\mu_K^{L-\lfloor\alpha\rfloor-1} \prod_{v=\lfloor\alpha\rfloor+1}^{L-1} \frac{b^i_v}{d^i_v + c^i_{v,0}\bar{n}^i_0 - b^i_v}(1+O(\varepsilon)) = \mu_K^{L-\lfloor\alpha\rfloor-1} \prod_{v=\lfloor\alpha\rfloor+1}^{L-1} \frac{b^i_v}{-f^i_{v,0}}(1+O(\varepsilon)) \tag{C.4.33}$$

and this chain of mutations occurs within a finite time, not scaling with K.

Thus the mutant counting process M^K_L can be approximated by the corresponding thinnings of the processes $\bar{M}^{(K,\pm)}_{\lfloor\alpha\rfloor+1}$. We denote these thinnings by $M^{(K,\pm)}$ to deduce the claim of the lemma. The small correction term $c\varepsilon$ in (C.4.24) stems from the approximation of the birth and death rates used to compute the thinning-probability under use of (C.5.17) (see [29]). □

Fixation and growth to a macroscopic size

From the previous lemma, we know that mutants of type L are born at a rate of order $K\mu_K^L$ with a specific phase-dependent prefactor. However, we cannot expect the L-individual appearing first to necessarily be the ancestor of a successfully invading new subpopulation. Instead, the subpopulation founded by a single L-mutant appearing might go extinct in finite time. This can happen for multiple reasons: Firstly, we do not assume that the invasion fitness $f^i_{L,0}$ of trait L is positive in all phases. Secondly, even in phases of positive invasion fitness, we have to account for the risk of extinction due to stochastic fluctuations. Lastly, in the case of a changing environment, even if the L population initially survives with a positive invasion fitness, it might still go extinct in a subsequent phase if the fitness becomes too negative.

In the following, to simplify notation, we only study the fate of the first L-mutant's subpopulation and its probability to go extinct or reach a macroscopic size. It turns out that one of these outcomes is obtained in a time of order $O(\ln K)$. Since new L-mutants arise on the longer time scale of order $O(1/K\mu^L_K)$, all later L-mutant subpopulations following previous extinction events can be regarded as independent and the same probabilities of different outcomes carry over (with probability tending to 1 as $K \to \infty$).

To state the lemma on the first mutant's fate, we require a number of stopping times. Recall that $S^{(K,\varepsilon)}$ is the first time when the total population of mutants of traits $[\![1, L]\!]$ reaches the size of εK and that $T^{(K,\varepsilon)}_\phi$ is the first time that the bounds on the resident 0-population fail. In addition, we introduce the first time that the L-mutant population goes extict after the j-th mutation,

$$T^K_{\mathrm{ext},j} = \inf\left\{t \geq T^K_{L,j} : N^K_L(t) = 0\right\}, \tag{C.4.34}$$

and the first time that the L-mutant population reaches a certain size M,

$$T^K_M = \inf\left\{t \geq 0 : N^K_L(t) \geq M\right\}. \tag{C.4.35}$$

Finally, to characterize the mutation times for which an L-invasion is possible, we introduce the function

$$g(t) = \int_0^t f_{L,0}(u)\mathrm{d}u, \quad t \in [0, \infty) \tag{C.4.36}$$

and sets

$$\tilde{A} = \{t \geq 0 : \ \exists\ s \in (0, T^\Sigma_\ell] : g(t+s) < g(t)\}, \tag{C.4.37}$$

$$A = \{t \geq 0 : \ \forall\ s \in (0, T^\Sigma_\ell] : g(t+s) > g(t)\}. \tag{C.4.38}$$

These definitions allow us to distinguish the following cases in our lemma, where it will be part of the claim to argue that these are exhaustive for large K:

$$\Omega^{K,\tilde{A}} = \left\{T^K_{L,1}/\lambda_K \in \tilde{A}\right\} \cap \left\{\left(T^K_{L,1} + \frac{2}{f^{\mathrm{av}}_{L,0}} \ln K\right) \wedge T^K_{\varepsilon^2 K} \leq T^K_{L,2} \wedge S^{(K,\varepsilon)} \wedge T^{(K,\varepsilon)}_\phi\right\} \tag{C.4.39}$$

$$\begin{aligned}\Omega^{K,A,i} = &\left\{T^K_{L,1}/\lambda_K \in A, (T^K_{L,1}/\lambda_K \bmod T^\Sigma_\ell) \in [T^\Sigma_{i-1}, T^\Sigma_i)\right\}\\ &\cap \left\{\left(T^K_{L,1} + \frac{2}{f^{\mathrm{av}}_{L,0}} \ln K\right) \wedge T^K_{\varepsilon^2 K} \leq T^K_{L,2} \wedge S^{(K,\varepsilon)} \wedge T^{(K,\varepsilon)}_\phi\right\}, \qquad 1 \leq i \leq \ell.\end{aligned} \tag{C.4.40}$$

Lemma C.4.4. *Fix $\varepsilon > 0$ small enough, let the initial condition be given by Assumption 5 and let the fitness landscape satisfy Assumption 7. Then there exist constants $C, \hat{C} < \infty$ (independent of ε) such that we obtain the following:*

(i) The sets $\Omega^{K,\tilde{A}}$ and $\Omega^{K,A,i}$, $1 \leq i \leq \ell$, are pairwise disjoint and

$$\liminf_{K\to\infty} \mathbb{P}\left(\Omega^{K,\tilde{A}} \cup \bigcup_{i=1}^{\ell} \Omega^{K,A,i}\right) \geq 1 - 2c\varepsilon, \tag{C.4.41}$$

where $c < \infty$ is the constant from Lemma C.4.3.

(ii) The probability of extinction for a mutation event at a (rescaled) time in $\tilde{A}$ satisfies

$$\lim_{K\to\infty} \mathbb{P}\left(T^K_{\mathrm{ext},1} < T^K_{L,1} + \lambda_K T^{\Sigma}_{\ell} \mid \Omega^{K,\tilde{A}}\right) = 1. \tag{C.4.42}$$

(iii) The probability of extinction for a mutation event at a (rescaled) time in A satisfies

$$\limsup_{K\to\infty} \left| \mathbb{P}\left(T^K_{\mathrm{ext},1} < (T^K_{L,1} + \lambda_K T^{\Sigma}_{\ell}) \wedge T^K_{\varepsilon^2 K} \mid \Omega^{K,A,i}\right) - \left(1 - \frac{f^i_{L,0}}{b^i_L}\right)\right| \leq C\varepsilon, \quad 1 \leq i \leq \ell. \tag{C.4.43}$$

(iv) The probability of growth to a macroscopic size for a mutation event at a (rescaled) time in A satisfies

$$\limsup_{K\to\infty} \left| \mathbb{P}\left(T^K_{\varepsilon^2 K} < T^K_{L,1} + \frac{1+\hat{C}\varepsilon}{f^{\mathrm{av}}_{L,0}} \ln K \mid \Omega^{K,A,i}\right) - \frac{f^i_{L,0}}{b^i_L}\right| \leq C\varepsilon, \quad 1 \leq i \leq \ell. \tag{C.4.44}$$

Essentially, what this lemma entails is the following: If an L-mutant arises at a (rescaled) time in $\tilde{A}$, its offspring is guaranteed to go extinct within one cycle of parameter phases. If it occurs at a (rescaled) time in A, during an i-phase, its offspring can still go extinct, at a probability of roughly $1 - f^i_{L,0}/b^i_L$. It again does so within one cycle of parameter phases and in the meantime never reaches a population size of $\varepsilon^2 K$. If the offspring population survives, which it does at the counter probability of roughly $f^i_{L,0}/b^i_L$, it grows to a macroscopic size of $\varepsilon^2 K$ within a time that is not much larger than $\ln K / f^{\mathrm{av}}_{L,0}$. Moreover, these are all the possible cases.

Proof. The proof can be broken down into six steps:

1. Proof of claim (i)
2. Introduction of coupled birth death processes $N^{(K,\pm)}_L$ with time-dependent parameters to bound N^K_L
3. Proof of claim (ii)
4. Introduction of coupled birth death processes $Z^{i,\pm}$ with constant parameters to bound N^K_L during an i-th phase

5. Lower bound for extinction probability in claim (iii)
6. Lower bound for fixation probability in claim (iv) and conclusion of claims (iii)&(iv)

Step 1: By Lemma C.4.2 and the fact that all mutant traits $[\![\lfloor\alpha\rfloor+1, L-1]\!]$ are unfit, at the time when the total mutant population surpasses an εK threshold, $S^{(K,\varepsilon)}$, the L-mutant population is required to be of order K and all other mutant populations are of lower order. In particular, this cannot occur before time $T^K_{L,1}$ or $T^K_{\varepsilon^2 K}$.

Moreover, from the result of Lemma C.4.3, it is not hard to see that there exists a $\hat{T}_\varepsilon < \infty$ such that

$$\liminf_{K\to\infty} \mathbb{P}\left(T^K_{L,1} \leq \hat{T}_\varepsilon / K\mu^L_K\right) \geq 1 - c\varepsilon. \tag{C.4.45}$$

Taking $\hat{T}_\varepsilon < \check{T}_\varepsilon < \infty$ slightly larger,

$$\liminf_{K\to\infty} \mathbb{P}\left(T^K_{L,1} + \frac{2}{f^{\mathrm{av}}_{L,0}} \ln K \leq \check{T}_\varepsilon / K\mu^L_K\right) \geq 1 - c\varepsilon \tag{C.4.46}$$

holds true as well. Again by Lemma C.4.3, we obtain that

$$\liminf_{K\to\infty} \mathbb{P}\left(T^K_{L,1} + \frac{2}{f^{\mathrm{av}}_{L,0}} \ln K \leq T^K_{L,2}\right) \geq 1 - c\varepsilon. \tag{C.4.47}$$

Finally, from Lemma C.4.1 we deduce that

$$\lim_{K\to\infty} \mathbb{P}\left(S^{(K,\varepsilon)} \wedge \check{T}_\varepsilon / K\mu^L_K \leq T^{(K,\varepsilon)}_\phi\right) = 1. \tag{C.4.48}$$

Combining all of the above facts eventually yields

$$\liminf_{K\to\infty} \mathbb{P}\left(\left(T^K_{L,1} + \frac{2}{f^{\mathrm{av}}_{L,0}} \ln K)\right) \wedge T^K_{\varepsilon^2 K} \leq T^K_{L,2} \wedge S^{(K,\varepsilon)} \wedge T^{(K,\varepsilon)}_\phi\right) \geq 1 - 2c\varepsilon. \tag{C.4.49}$$

Claim (i) now immediately follows since the rate of newly arriving L-mutants is bounded uniformly (of order $K\mu^L_K$) and, by Assumption 6(ii), $(A \cup \tilde{A})^C$ is a Lebesgue nullset.

Step 2: To prove claims (ii)-(iv), we introduce coupled pure birth death processes to bound the population N^K_L. For $\varepsilon > 0$ and some small $\delta > 0$ that will be fixed later, define

$$b^+_L(t) = b_L(t), \tag{C.4.50}$$

$$b^-_L(t) = b_L(t)(1-\varepsilon), \tag{C.4.51}$$

$$D^+_L(t) = d_L(t) + c_{L,0}(t)\left[-M\varepsilon + \sum_{i=1}^{L}\left(\mathbb{1}_{t\in[T^\Sigma_{i-1}, T^\Sigma_{i-1}+\delta)} \min\{\bar{n}^{i-1}_0, \bar{n}^i_0\} + \mathbb{1}_{t\in[T^\Sigma_{i-1}+\delta, T^\Sigma_i)} \bar{n}^i_0\right)\right], \tag{C.4.52}$$

$$D^-_L(t) = d_L(t) + \varepsilon\check{c}_L + c_{L,0}(t)\left[M\varepsilon + \sum_{i=1}^{L}\left(\mathbb{1}_{t\in[T^\Sigma_{i-1}, T^\Sigma_{i-1}+\delta)} \max\{\bar{n}^{i-1}_0, \bar{n}^i_0\} + \mathbb{1}_{t\in[T^\Sigma_{i-1}+\delta, T^\Sigma_i)} \bar{n}^i_0\right)\right], \tag{C.4.53}$$

with periodic extensions, $\bar{n}_v^0 := \bar{n}_v^\ell$, $\check{c}_L := \max_{1\leq w\leq L, 1\leq i\leq \ell} c^i_{L,w}$, and M the (ε-independent) constant from Lemma C.4.1. Then, for $\varepsilon > 0$ small enough, K large enough such that $T_\varepsilon < \delta\lambda_K$ and $\mu_K < \varepsilon$, and all times $t \leq S^{(K,\varepsilon)} \wedge T_\phi^{(K,\varepsilon)}$,

$$b_L^-(t/\lambda_K) \leq b_L^K(t) \leq b_L^+(t/\lambda_K), \tag{C.4.54}$$

$$D_L^-(t/\lambda_K) \geq d_L^K(t) + \sum_{w=0}^{L} \frac{c^K_{L,w}(t)}{K} N_w^K(t) \geq D_L^+(t/\lambda_K) \geq \underline{D}, \tag{C.4.55}$$

for some $\underline{D} > 0$. We can hence, for all K large enough, define a collection of pure birth death processes $(N_L^{(K,\pm)}(t))_{t\geq T^K_{L,1}}$ with time-inhomogeneous birth rates $b_L^\pm(t/\lambda_K)$ and death rates $D_L^\pm(t/\lambda_K)$, coupled to $(N_L^K(t))_{t\geq T^K_{L,1}}$ such that

$$N_L^{(K,-)}(T^K_{L,1}) = N_L^K(T^K_{L,1}) = N_L^{(K,+)}(T^K_{L,1}) = 1, \tag{C.4.56}$$

$$N_L^{(K,-)}(t) \leq N_L^K(t) \leq N_L^{(K,+)}(t), \text{ for } T^K_{L,1} \leq t \leq T^K_{L,2} \wedge S^{(K,\varepsilon)} \wedge T_\phi^{(K,\varepsilon)}. \tag{C.4.57}$$

This coupling can for example be constructed using a Poisson measure representation, as in the proof of Lemma C.4.2.

Moreover,

$$\begin{aligned} &|(b_L^\pm(t/\lambda_K) - D_L^\pm(t/\lambda_K)) - f^K_{L,0}(t)| \\ &\qquad \leq \begin{cases} C_1 & \text{for } t/\lambda_K \in [jT^\Sigma_\ell + T^\Sigma_{i-1}, jT^\Sigma_\ell + T^\Sigma_{i-1} + \delta),\ j \in \mathbb{N},\ i = 1, ..., \ell, \\ C_2\varepsilon & \text{for } t/\lambda_K \in [jT^\Sigma_\ell + T^\Sigma_{i-1} + \delta, jT^\Sigma_\ell + T^\Sigma_i),\ j \in \mathbb{N},\ i = 1, ..., \ell, \end{cases} \end{aligned} \tag{C.4.58}$$

for some constants $C_1, C_2 < \infty$.

Step 3: With these couplings in place, we start by considering the case of claim (ii), i.e. $T^K_{L,1}/\lambda_K \in \tilde{A}$. We set

$$g^{(K,\pm)}(s) := \int_0^s b_L^\pm(u/\lambda_K) - D_L^\pm(u/\lambda_K)\mathrm{d}u. \tag{C.4.59}$$

Then, for all $s \in [T^K_{L,1}, T^K_{L,1} + T^\Sigma_\ell\lambda_K]$,

$$\begin{aligned} g^{(K,+)}(s) - g^{(K,+)}(T^K_{L,1}) &\leq \int_{T^K_{L,1}}^s f_{L,0}(u/\lambda_K)\mathrm{d}u + \int_{T^K_{L,1}}^s \left| b_L^+(u/\lambda_K) - D_L^+(u/\lambda_K) - f^K_{L,0}(u) \right| \mathrm{d}u \\ &\leq \int_{T^K_{L,1}/\lambda_K}^{s/\lambda_K} \lambda_K f_{L,0}(u)\mathrm{d}u + C_1\delta\ell\lambda_K + C_2\varepsilon T^\Sigma_\ell \lambda_K \\ &\leq \lambda_K \left[g\left(s/\lambda_K\right) - g\left(T^K_{L,1}/\lambda_K\right) + C_3(\delta + \varepsilon) \right], \end{aligned} \tag{C.4.60}$$

for some $C_3 < \infty$. Since $T^K_{L,1}/\lambda_K \in \tilde{A}$, there exists a $u_0 \in (T^K_{L,1}, T^K_{L,1} + T^\Sigma_\ell \lambda_K]$ such that

$$g(u_0/\lambda_K) - g(T^K_{L,1}/\lambda_K) < 0. \tag{C.4.61}$$

Note that the choice of u_0 depends on the random stopping time $T^K_{L,1}$ and is hence also random. Since g is a continuous function, for $\delta, \varepsilon > 0$ small enough, there exist $0 < u_1 < u_2 \leq T^\Sigma_\ell$ and $c > 0$ (each also dependent on $T^K_{L,1}$) such that, for all $s \in [T^K_{L,1} + u_1\lambda_K, T^K_{L,1} + u_2\lambda_K]$,

$$g^{(K,+)}(s) - g^{(K,+)}(T^K_{L,1}) \leq -\lambda_K c. \tag{C.4.62}$$

Using an identity for the generating function of time-inhomogeneous birth death process from [95, Ch. 6.12] and the fact that $\mathrm{d}/\mathrm{d}s \left[g^{(K,\pm)}(s) - g^{(K,\pm)}(T^K_{L,1})\right] = b^\pm_L(s/\lambda_K) - D^\pm_L(s/\lambda_K)$, we conclude that, on the event $\Omega^{K\tilde{A}}$,

$$\begin{aligned}
&\mathbb{P}\left(T^K_{\text{ext},1} < T^K_{L,1} + \lambda_K T^\Sigma_\ell \mid T^K_{L,1}\right) \geq \mathbb{P}\left(N^{(K,+)}_L(T^K_{L,1} + \lambda_K T^\Sigma_\ell) = 0 \mid T^K_{L,1}\right) \\
&= 1 - \left(e^{-\left(g^{(K,+)}(T^K_{L,1}+\lambda_K T^\Sigma_\ell) - g^{(K,+)}(T^K_{L,1})\right)} + \int_{T^K_{L,1}}^{T^K_{L,1}+\lambda_K T^\Sigma_\ell} b^+_L(s/\lambda_K) e^{-\left(g^{(K,+)}(s) - g^{(K,+)}(T^K_{L,1})\right)} \mathrm{d}s\right)^{-1} \\
&= 1 - \left(1 + \int_{T^K_{L,1}}^{T^K_{L,1}+\lambda_K T^\Sigma_\ell} D^+_L(s/\lambda_K) e^{-\left(g^{(K,+)}(s) - g^{(K,+)}(T^K_{L,1})\right)} \mathrm{d}s\right)^{-1} \\
&\geq 1 - \left(\int_{T^K_{L,1}}^{T^K_{L,1}+\lambda_K T^\Sigma_\ell} \underline{D} e^{-\left(g^{(K,+)}(s) - g^{(K,+)}(T^K_{L,1})\right)} \mathrm{d}s\right)^{-1} \geq 1 - \left(\int_{T^K_{L,1}+u_1\lambda_K}^{T^K_{L,1}+u_2\lambda_K} \underline{D} e^{\lambda_K c} \mathrm{d}s\right)^{-1} \\
&= 1 - \frac{1}{\lambda_K(u_2 - u_1)\underline{D} e^{\lambda_K c}},
\end{aligned} \tag{C.4.63}$$

which converges to 1 as $K \to \infty$. This convergence holds true for every $T^K_{L,1}/\lambda_K \in \tilde{A}$ and hence the conditioning on $T^K_{L,1}$ on the left hand side can be dropped.

Step 4: Next, we turn to the case of claims (iii) and (iv), i.e. $T^K_{L,1}/\lambda_K \in A$, with the L-mutant appearing during an i-th phase, such that $b^K_L(T^K_{L,1}) = b^i_L$ etc. Note that the definition of the set A automatically implies that $f^i_{L,0} > 0$, hence we can make use of couplings to supercritical birth death processes and existing results for the latter.

For the following argument, we restrict to the event of

$$\Omega^{K,A,i}_\delta = \Omega^{K,A,i} \cap \left\{(T^K_{L,1}/\lambda_K \bmod T^\Sigma_\ell) \in [T^\Sigma_{i-1} + \delta, T^\Sigma_i - \delta)\right\}, \tag{C.4.64}$$

i.e. exclude the cases where $T^K_{L,1}$ falls into the short $\delta\lambda_K$-interval at the beginning of a new phase or close to its end. Since, by Lemma C.4.3, L-mutants arrive at a uniformly bounded rate of order $K\mu^L_K$, it follows that

$$\lim_{\delta\to 0} \lim_{K\to\infty} \mathbb{P}\left(\Omega^{K,A,i}_\delta | \Omega^{K,A,i}\right) = 1. \tag{C.4.65}$$

Hence it is sufficient to derive the claim on $\Omega^{K,A,i}_\delta$ and pick $\delta > 0$ arbitrarily small in the end.

For large enough K, $\sqrt{\lambda_K} < \delta\lambda_K$ and hence, on $\Omega^{K,A,i}_\delta$, time $T^K_{L,1} + \sqrt{\lambda_K}$ is smaller that the time point of the next phase change. As a result, the i-phase parameters are applicable for the entire time horizon of $[T^K_{L,1}, T^K_{L,1} + \sqrt{\lambda_K})$. Moreover, during this time, for $\varepsilon > 0$ small enough,

$$b^{\pm}_L(t/\lambda_K) - D^{\pm}_L(t/\lambda_K) \geq f^i_{L,0} - C_2\varepsilon > 0. \tag{C.4.66}$$

Considering the coupled processes defined above, this implies that $(N^{(K,\pm)}_L(T^K_{L,1}+s))_{s\in[0,\sqrt{\lambda_K})}$ have the same distribution as supercritical birth death processes $(Z^{i,\pm}(s))_{s\in[0,\sqrt{\lambda_K})}$ with initial condition $Z^{i,\pm}(0) = 1$, birth rate b^i_L or $b^i_L(1-\varepsilon)$, and death rate $d^i_L + c^i_{L,0}(\bar{n}^i_0 - M\varepsilon)$ or $d^i_L + \varepsilon\check{c}_L + c^i_{L,0}(\bar{n}^i_0 + M\varepsilon)$, respectively. Importantly, the same processes $Z^{i,\pm}$ can be chosen for all (large enough) K here.

Step 5: One can now bound the probability of extinction from below by

$$\begin{aligned}
\lim_{K\to\infty} \mathbb{P}\left(T^K_{\text{ext},1} < T^K_{L,1} + \lambda_K T^\Sigma_\ell | \Omega^{K,A,i}_\delta\right) &\geq \lim_{K\to\infty} \mathbb{P}\left(T^K_{\text{ext},1} < T^K_{L,1} + \sqrt{\lambda_K} | \Omega^{K,A,i}_\delta\right) \\
&\geq \lim_{K\to\infty} \mathbb{P}\left(N^{(K,+)}_L(T^K_{L,1} + \sqrt{\lambda_K}) = 0 | \Omega^{K,A,i}_\delta\right) \\
&= \lim_{K\to\infty} \mathbb{P}\left(Z^{i,+}(\sqrt{\lambda_K}) = 0\right) \\
&= \lim_{s\to\infty} \mathbb{P}\left(Z^{i,+}(s) = 0\right) \\
&= \mathbb{P}\left(\lim_{s\to\infty} Z^{i,+}(s) = 0\right).
\end{aligned} \tag{C.4.67}$$

By a standard branching process results (e.g. Theorem 1 in Chapter III.4 of [4]), this extinction probability is equal to

$$\frac{d^i_L + c^i_{L,0}(\bar{n}^i_0 - M\varepsilon)}{b^i_L} = 1 - \frac{f^i_{L,0}}{b^i_L} - \frac{c^i_{L,0}M\varepsilon}{b^i_L} \geq 1 - \frac{f^i_{L,0}}{b^i_L} - C\varepsilon, \tag{C.4.68}$$

for some $C < \infty$ independent of i, $\varepsilon > 0$, and $\delta > 0$.

To ensure that this extinction occurs before reaching a threshold of $\varepsilon^2 K$, we can bound $(N^K_L(T^K_{L,1} + s))_{s\in[0,\lambda_K T^\Sigma_\ell)}$ from above by a coupled pure birth process $(\overline{Z}(s))_{s\in[0,\lambda_K T^\Sigma_\ell)}$ with birth rate $\bar{b} = \max_{1\leq j\leq \ell} b^j_L$ and $\overline{Z}(0) = 1$ and deduce

$$\begin{aligned}
\lim_{K\to\infty} \mathbb{P}\left(T^K_{L,1} + \lambda_K T^\Sigma_\ell < T^K_{\varepsilon^2 K} \mid \Omega^{K,A,i}_\delta\right) &\geq 1 - \lim_{K\to\infty} \mathbb{P}\left(\overline{Z}(\lambda_K T^\Sigma_\ell) \geq \varepsilon^2 K\right) \\
&= 1 - \lim_{K\to\infty} \mathbb{P}\left(\overline{Z}(\lambda_K T^\Sigma_\ell) e^{-\bar{b}\lambda_K T^\Sigma_\ell} \geq \varepsilon^2 K e^{-\bar{b}\lambda_K T^\Sigma_\ell}\right).
\end{aligned} \tag{C.4.69}$$

On one hand, by Theorems 1 and 2 in Chapter III.7 of [4], $\lim_{K\to\infty} \overline{Z}^i(\lambda_K T^\Sigma_\ell) e^{-\bar{b}\lambda_K T^\Sigma_\ell}$ exists almost surely and has expectation 1. On the other hand, since $\lambda_K \ll \ln K$, $\lim_{K\to\infty} \varepsilon^2 K e^{-\bar{b}\lambda_K T^\Sigma_\ell} = \infty$, for any $\varepsilon > 0$. Hence the limit on the right hand side above is equal to 0 and consequentially, together with (C.4.67) we can conclude that

$$\lim_{K\to\infty} \mathbb{P}\left(T^K_{\text{ext},1} < (T^K_{L,1} + \lambda_K T^\Sigma_\ell) \wedge T^K_{\varepsilon^2 K} | \Omega^{K,A,i}_\delta\right) \geq 1 - \frac{f^i_{L,0}}{b^i_L} - C\varepsilon. \tag{C.4.70}$$

Step 6: Deriving the corresponding lower bound on the fixation probability of the L-mutant is a little more involved and can be broken down into three substeps: First, we consider the probability of initial survival, similar to Step 5, to prove (C.4.72). Second, once a small but diverging population size is obtained, we can show that a size of order $K^{\varepsilon\gamma}$ is reached within a small $\ln K$-time, proving (C.4.80). Finally, once this positive K-power is reached, the time to grow to a macroscopic size of order K can be approximated using results from [71], concluding (C.4.84).

To derive the lower bound, we consider the coupled birth death processes $N_L^{(K,-)}$ and $Z^{i,-}$. On the event of $\Omega_\delta^{K,A,i}$, setting $f_{L,0}^{i,-} := b_L^i(1-\varepsilon) - (d_L^i + \varepsilon\check{c}_L + c_{L,0}^i(\bar{n}_0^i + M\varepsilon)) > 0$,

$$\begin{aligned}\lim_{K\to\infty} \mathbb{P}\left(N_L^{(K,-)}(T_{L,1}^K + \sqrt{\lambda_K}) \geq e^{f_{L,0}^{i,-}\sqrt{\lambda_K}/2} \mid \Omega_\delta^{K,A,i}\right) &= \lim_{K\to\infty} \mathbb{P}\left(Z^{i,-}(\sqrt{\lambda_K}) \geq e^{f_{L,0}^{i,-}\sqrt{\lambda_K}/2}\right)\\ &= \lim_{s\to\infty} \mathbb{P}\left(Z^{i,-}(s) \geq e^{f_{L,0}^{i,-}s/2}\right)\\ &= \lim_{s\to\infty} \mathbb{P}\left(Z^{i,-}(s)e^{-f_{L,0}^{i,-}s} \geq e^{-f_{L,0}^{i,-}s/2}\right).\end{aligned} \tag{C.4.71}$$

Now again, by Theorems 1 and 2 in Chapter III.7 of [4], $\lim_{s\to\infty} Z^{i,-}(s)e^{-f_{L,0}^{i,-}s} = W$ exists almost surely (and hence in distribution), is non-negative, has expectation 1, and has a density on $\{W > 0\}$. Moreover, $\mathbb{P}(W > 0) = f_{L,0}^{i,-}/(b_L^i(1-\varepsilon))$. Consequentially, we can find $c_\varepsilon > 0$ such that

$$\begin{aligned}\lim_{K\to\infty} \mathbb{P}\left(N_L^{(K,-)}(T_{L,1}^K + \sqrt{\lambda_K}) \geq e^{f_{L,0}^{i,-}\sqrt{\lambda_K}/2} \mid \Omega_\delta^{K,A,i}\right) &\geq \lim_{s\to\infty} \mathbb{P}\left(Z^{i,-}(s)e^{-f_{L,0}^{i,-}s} \geq e^{-f_{L,0}^{i,-}s/2}\right)\\ &\geq \lim_{s\to\infty} \mathbb{P}\left(Z^{i,-}(s)e^{-f_{L,0}^{i,-}s} \geq c_\varepsilon\right)\\ &= \mathbb{P}\left(\lim_{s\to\infty} Z^{i,-}(s)e^{-f_{L,0}^{i,-}s} \geq c_\varepsilon\right)\\ &\geq \mathbb{P}(W > 0) - \varepsilon = \frac{f_{L,0}^{i,-}}{b_L^i(1-\varepsilon)} - \varepsilon\\ &\geq \frac{f_{L,0}^i}{b_L^i} - C\varepsilon,\end{aligned} \tag{C.4.72}$$

where $C < \infty$ can be chosen independently of i, $\varepsilon > 0$, and $\delta > 0$ (possibly larger than in Step 5).

Recalling (C.4.58), for all $s \in [T_{L,1}^K + \sqrt{\lambda_K}, T_{L,1}^K + \sqrt{\lambda_K} + \lambda_K T_\ell^\Sigma]$ we obtain

$$\begin{aligned}&g^{(K,-)}(s) - g^{(K,-)}(T_{L,1}^K + \sqrt{\lambda_K})\\ &\geq \int_{T_{L,1}^K+\sqrt{\lambda_K}}^{s} f_{L,0}(u/\lambda_K)\mathrm{d}u - \int_{T_{L,1}^K+\sqrt{\lambda_K}}^{s} \left|b_L^-(u/\lambda_K) - D_L^-(u/\lambda_K) - f_{L,0}^K(u)\right| \mathrm{d}u\\ &\geq \int_{(T_{L,1}^K+\sqrt{\lambda_K})/\lambda_K}^{s/\lambda_K} \lambda_K f_{L,0}(u)\mathrm{d}u - C_1\delta\ell\lambda_K - C_2\varepsilon T_\ell^\Sigma\lambda_K\\ &\geq \lambda_K\left(g(s/\lambda_K) - g\left((T_{L,1}^K + \sqrt{\lambda_K})/\lambda_K\right) - C_3(\delta+\varepsilon)\right),\end{aligned} \tag{C.4.73}$$

for some $C_3 < \infty$. Since $T^K_{L,1}/\lambda_K \in A$, it follows that

$$g(s/\lambda_K) - g(T^K_{L,1}/\lambda_K) > 0, \quad \forall\ s \in (T^K_{L,1}, T^K_{L,1} + \lambda_K T^\Sigma_\ell]. \tag{C.4.74}$$

Since $\sqrt{\lambda_K}/\lambda_K \to 0$ and g is continuous, this implies that, for $\delta, \varepsilon > 0$ small enough and K large enough,

$$g^{(K,-)}(s) - g^{(K,-)}(T^K_{L,1} + \sqrt{\lambda_K}) > 0, \quad \forall\ s \in (T^K_{L,1} + \sqrt{\lambda_K}, T^K_{L,1} + \sqrt{\lambda_K} + \lambda_K T^\Sigma_\ell]. \tag{C.4.75}$$

Hence we can apply Lemma C.5.8 to deduce that, since $1 \ll e^{f^{i,-}_{L,0}\sqrt{\lambda_K}/2} \ll K^\varepsilon$, for any $p \in (0,1)$,

$$\begin{aligned}\lim_{K\to\infty} \mathbb{P}\Big(N^{(K,-)}_L(T^K_{L,1} + \sqrt{\lambda_K} + \varepsilon \ln K) \geq p e^{g^{(K,-)}(T^K_{L,1}+\sqrt{\lambda_K}+\varepsilon\ln K) - g^{(K,-)}(T^K_{L,1}+\sqrt{\lambda_K})} e^{f^{i,-}_{L,0}\sqrt{\lambda_K}/2} \Big| \\ N^{(K,-)}_L(T^K_{L,1} + \sqrt{\lambda_K}) \geq e^{f^{i,-}_{L,0}\sqrt{\lambda_K}/2}\Big) = 1.\end{aligned} \tag{C.4.76}$$

Similar to above, for some $C_4 < \infty$,

$$\begin{aligned}&g^{(K,-)}(T^K_{L,1} + \sqrt{\lambda_K} + \varepsilon \ln K) - g^{(K,-)}(T^K_{L,1} + \sqrt{\lambda_K}) \\ &\geq \int_{(T^K_{L,1}+\sqrt{\lambda_K})/\lambda_K}^{(T^K_{L,1}+\sqrt{\lambda_K}+\varepsilon\ln K)/\lambda_K} \lambda_K f_{L,0}(u)\mathrm{d}u - \int_{T^K_{L,1}+\sqrt{\lambda_K}}^{T^K_{L,1}+\sqrt{\lambda_K}+\varepsilon\ln K} \Big| b^-_L(u/\lambda_K) - D^-_L(u/\lambda_K) - f^K_{L,0}(u)\Big| \mathrm{d}u \\ &\geq \varepsilon \ln K f^{\mathrm{av}}_{L,0} - T^\Sigma_\ell \lambda_K \max_{1\leq i\leq\ell} |f^i_{L,0} - f^{\mathrm{av}}_{L,0}| - C_1 \ell \frac{\varepsilon \ln K}{T^\Sigma_\ell \lambda_K} \delta \lambda_K - C_2 \varepsilon^2 \ln K \\ &\geq \varepsilon \ln K \left(f^{\mathrm{av}}_{L,0} - C_4(\delta + \varepsilon)\right) \geq \varepsilon \ln K \frac{f^{\mathrm{av}}_{L,0}}{2},\end{aligned} \tag{C.4.77}$$

as long as $\delta, \varepsilon > 0$ small enough and K large enough. This yields

$$\lim_{K\to\infty} \mathbb{P}(N^{(K,-)}_L(T^K_{L,1} + \sqrt{\lambda_K} + \varepsilon \ln K) \geq K^{\varepsilon f^{\mathrm{av}}_{L,0}/3} \mid N^{(K,-)}_L(T^K_{L,1} + \sqrt{\lambda_K}) \geq e^{f^{i,-}_{L,0}\sqrt{\lambda_K}/2}) = 1. \tag{C.4.78}$$

Summarizing so far, setting $\gamma = f^{\mathrm{av}}_{L,0}/3$ and

$$T^{(K,-)}_M = \inf\{t \geq 0 :\ N^{(K,-)}_L(t) \geq M\}, \tag{C.4.79}$$

the last limit and (C.4.72) yield

$$\begin{aligned}&\lim_{K\to\infty} \mathbb{P}\left(T^{(K,-)}_{K^{\varepsilon\gamma}} \leq T^K_{L,1} + \sqrt{\lambda_K} + \varepsilon \ln K \mid \Omega^{K,A,i}_\delta\right) \\ &\geq \lim_{K\to\infty} \mathbb{P}\left(T^{(K,-)}_{K^{\varepsilon\gamma}} \leq T^K_{L,1} + \sqrt{\lambda_K} + \varepsilon \ln K \mid N^{(K,-)}_L(T^K_{L,1} + \sqrt{\lambda_K}) \geq e^{f^{i,-}_{L,0}\sqrt{\lambda_K}/2}\right) \\ &\qquad \times \mathbb{P}\left(N^{(K,-)}_L(T^K_{L,1} + \sqrt{\lambda_K}) \geq e^{f^{i,-}_{L,0}\sqrt{\lambda_K}/2} \mid \Omega^{K,A,i}_\delta\right) \\ &\geq \frac{f^i_{L,0}}{b^i_L} - C\varepsilon.\end{aligned} \tag{C.4.80}$$

Finally, now that a population size of some positive power of K is reached, we can apply Theorem B.1 of [71]. Setting

$$r_{\mathrm{av}} = \frac{1}{T_\ell^\Sigma} \int_0^{T_\ell^\Sigma} b_L^-(t) - D_L^-(t) \mathrm{d}t > f_{L,0}^{\mathrm{av}} - c_r(\delta + \varepsilon), \tag{C.4.81}$$

for some $c_r < \infty$ independent of ε and δ, we obtain that, for any $\eta > 0$ and $S > 0$,

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{P}\Big(\forall s \in [0,S] :\ N_L^{(K,-)}(T_{K^{\varepsilon\gamma}}^{(K,-)} + s\ln K) > K^{\varepsilon\gamma + [f_{L,0}^{\mathrm{av}} - c_r(\delta+\varepsilon)]s - \eta}\Big) \\
\geq &\lim_{K\to\infty} \mathbb{P}\Big(\forall s \in [0,S] :\ N_L^{(K,-)}(T_{K^{\varepsilon\gamma}}^{(K,-)} + s\ln K) + 1 > K^{\varepsilon\gamma + r_{\mathrm{av}}s - \eta}\Big) \\
= &\lim_{K\to\infty} \mathbb{P}\Big(\forall s \in [0,S] :\ \ln\Big(N_L^{(K,-)}(T_{K^{\varepsilon\gamma}}^{(K,-)} + s\ln K) + 1\Big) > (\varepsilon\gamma + r_{\mathrm{av}}s - \eta)\ln K\Big) \\
\geq &\lim_{K\to\infty} \mathbb{P}\left(\forall s \in [0,S] :\ \left|\frac{\ln\Big(N_L^{(K,-)}(T_{K^{\varepsilon\gamma}}^{(K,-)} + s\ln K) + 1\Big)}{\ln K} - (\varepsilon\gamma + r_{\mathrm{av}}s)\right| < \eta\right) = 1.
\end{aligned} \tag{C.4.82}$$

Hence,

$$\begin{aligned}
1 &= \lim_{K\to\infty} \mathbb{P}\left(T_{\varepsilon^2 K}^{(K,-)} < T_{K^{\varepsilon\gamma}}^{(K,-)} + \frac{1 - \varepsilon\gamma + \eta + \log_K(\varepsilon^2)}{f_{L,0}^{\mathrm{av}} - c_r(\delta+\varepsilon)} \ln K\right) \\
&\leq \lim_{K\to\infty} \mathbb{P}\left(T_{\varepsilon^2 K}^{(K,-)} < T_{K^{\varepsilon\gamma}}^{(K,-)} + \frac{1 + \eta}{f_{L,0}^{\mathrm{av}} - c_r(\delta+\varepsilon)} \ln K\right) \\
&\leq \lim_{K\to\infty} \mathbb{P}\left(T_{\varepsilon^2 K}^{(K,-)} < T_{K^{\varepsilon\gamma}}^{(K,-)} + \frac{1 + \hat{c}\varepsilon}{f_{L,0}^{\mathrm{av}}} \ln K\right)
\end{aligned} \tag{C.4.83}$$

for some $\hat{c} < \infty$ and as long as $0 < \delta, \eta < \varepsilon$ sufficiently small. Combining this with (C.4.80) yields

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{P}\left(T_{\varepsilon^2 K}^{K} < T_{L,1}^{K} + \frac{1 + \hat{C}\varepsilon}{f_{L,0}^{\mathrm{av}}} \ln K \mid \Omega_\delta^{K,A,i}\right) \\
\geq &\lim_{K\to\infty} \mathbb{P}\left(T_{\varepsilon^2 K}^{(K,-)} < T_{L,1}^{K} + \frac{1 + \hat{C}\varepsilon}{f_{L,0}^{\mathrm{av}}} \ln K \mid \Omega_\delta^{K,A,i}\right) \geq \frac{f_{L,0}^i}{b_L^i} - C\varepsilon,
\end{aligned} \tag{C.4.84}$$

for some $\hat{c} < \hat{C} < \infty$.

Since all the above bounds in the limiting probabilities hold true for any choice of $\delta > 0$ small enough and C and $\hat{C}$ can be chosen independent of $\delta > 0$ and $\varepsilon > 0$, we can pick δ arbitrarily small in the end and combine (C.4.65), (C.4.70), and (C.4.84) to deduce claims (iii) and (iv), for a possibly slightly larger choice of C.

□

Lotka-Volterra step and conclusion

To finally conclude Theorem C.2.2, we have to show that if a mutant population of trait L has successfully grown up to the macroscopic size $\varepsilon^2 K$, it invades and finally outcompetes the resident trait 0 population very fast and with high probability.

Lemma C.4.5. *For $\varepsilon > 0$ small enough and under the Assumptions 5, 6 and 7, there exists $C < \infty$ such that*

$$\lim_{K\to\infty} \mathbb{P}\left(T^K_{\varepsilon^2 K} \le T^{(K,\varepsilon)}_{\mathrm{inv}} \le T^K_{\varepsilon^2 K} + C\lambda_K \middle| T^K_{\varepsilon^2 K} < S^{(K,\varepsilon)} \wedge T^{(K,\varepsilon)}_{\phi}\right) = 1. \tag{C.4.85}$$

Proof. We would like to make use of the macroscopic population size of the L-mutant to approximate the total population process under use of a law of large numbers for dynamical systems, valid on time intervals of finite length (not scaling with K). However, this can only be helpful if the mutant trait L is currently fit with respect to the resident and thus has a positive growth rate, ensuring that the invasion takes place in such a finite time span. Unfortunately, the assumptions only guarantee the average fitness $f^{\mathrm{av}}_{L,0} > 0$ to be positive. We work around this by introducing the alternative stopping time

$$T^{(K,\mathrm{fit})}_{\varepsilon^2 K} := \inf\left\{t \ge 0 : N^K_L(t) \ge \varepsilon^2 K \text{ and } f^{(K,-)}_{L,0}(t) > 0\right\}, \tag{C.4.86}$$

which indicates the starting point of the approximation with the corresponding deterministic Lotka-Volterra system. For the definition of $f^{(K,-)}_{L,0}(t)$, we refer to (C.4.15) in the proof of Lemma C.4.2, where we use analog bounding processes.

Following the lines of Step 4 in the proof of Theorem 2.4 in [71], one can show that, conditioned on fixation,

$$T^K_{\varepsilon^2 K} \le T^{(K,\mathrm{fit})}_{\varepsilon^2 K} < T^K_{\varepsilon^2 K} + O(\lambda_K). \tag{C.4.87}$$

The main idea is to utilize that $g^{(K,-)}$ is a continuous function and, because of the assumption $f^{\mathrm{av}}_{L,0} > 0$, it holds that $g^{(K,-)}(T^K_{\varepsilon^2 K} + \lambda_K T^{\Sigma}_{\ell}) - g^{(K,-)}(T^K_{\varepsilon^2 K}) > 0$, for $\varepsilon > 0$ small enough. Looking at the first time after $T^K_{\varepsilon^2 K}$ when this difference is positive, one can show that this must fall into a phase of positive fitness, i.e. $f^i_{L,0} > 0$. Moreover, by Lemma C.5.8, the population size of the mutants must exceed $\varepsilon^2 K$, possibly shortly afterwards but still during the same phase. Therefore, $T^{(K,\mathrm{fit})}_{\varepsilon^2 K}$ is hit within a time of order $O(\lambda_K)$ after $T^K_{\varepsilon^2 K}$. Notably, between $T^K_{\varepsilon^2 K}$ and $T^{(K,\mathrm{fit})}_{\varepsilon^2 K}$ the total mutant population does also not exceed a size of εK and hence the approximating birth death processes can still be used for this argument.

At time $T^{(K,\mathrm{fit})}_{\varepsilon^2 K}$, it is now guaranteed that on the one hand, exactly the resident trait 0 and the mutants of trait L have a macroscopic population size, and on the other hand, the invading trait L is fit with respect to the resident trait while the resident is unfit with respect to trait L. This puts us into the position to apply the standard arguments of [75] to approximate the system by the corresponding deterministic two-type Lotka-Volterra system. This yields the existence of a finite and deterministic time $T(\varepsilon) < \infty$ such that

$$T^{(K,\varepsilon)}_{\mathrm{inv}} \le T^{(K,\mathrm{fit})}_{\varepsilon^2 K} + T(\varepsilon), \tag{C.4.88}$$

with probability converging to 1 as $K \to \infty$. For more details on this type of argument we refer to e.g. [40, Prop. 2(b)]. $\square$

We are now well prepared to put everything together and finally prove Theorem C.2.2. From Lemma C.4.3 we know that, until time $S^{(K,\varepsilon)}$, single L-mutants appear approximately as a Poisson process with intensity function $K\mu_K^L \tilde{R}^{(K,\pm)}(t)$. From Lemma C.4.4, remembering that the claim transfers from $T_{L,1}^K$ to general $T_{L,j}^K$ due to a separation of the time scales of mutant appearance and invasion/extinction, we know however that not all of these L-mutants lead to a macroscopic mutant population. Instead, there is a thinning probability (dependent on the appearance time), that can be estimated by

$$\mathbb{1}_{t/\lambda_K \in A} \frac{f_{L,0}^K(t)}{b_L^K(t)}(1 \pm C\varepsilon). \tag{C.4.89}$$

Therefore, successful L-mutants are born approximately according to a Poisson process with new intensity function given by the product of the former one and the thinning probability.

Now, asking for birth of the first successful mutant, we see directly that this happens on a time scale of order $1/K\mu_K^L$. Moreover, we know that the new intensity function is periodic with period length $T_\ell^\Sigma \lambda_K$, which is much shorter than the expected waiting time. Thus, effectively the intensity function can be replaced by its average over one period, i.e. for every $T < \infty$, we have

$$\begin{aligned}
&\int_0^{T/K\mu_K^L} K\mu_K^L \tilde{R}^{(K,\pm)}(t) \mathbb{1}_{t/\lambda_K \in A} \frac{f_{L,0}^K(t)}{b_L^K(t)}(1 \pm C\varepsilon) \mathrm{d}t \\
&= (1 \pm C\varepsilon) \frac{T}{T_\ell^\Sigma \lambda_K} \int_0^{T_\ell^\Sigma \lambda_K} a_{\lfloor\alpha\rfloor}^{(K,\pm)}(t) b_{\lfloor\alpha\rfloor}^K(t) \prod_{w=\lfloor\alpha\rfloor+1}^{L-1} \frac{b_w^K(t)}{|f_{w,0}^K(t)|} \frac{f_{L,0}^K(t)}{b_L^K(t)} \mathbb{1}_{t/\lambda_K \in A} \mathrm{d}t + O(\lambda_K K\mu_K^L) \\
&= (1 \pm C\varepsilon) \frac{T}{T_\ell^\Sigma} \int_0^{T_\ell^\Sigma} a_{\lfloor\alpha\rfloor}^{(K,\pm)}(t\lambda_K) b_{\lfloor\alpha\rfloor}(t) \prod_{w=\lfloor\alpha\rfloor+1}^{L-1} \frac{b_w(t)}{|f_{w,0}(t)|} \frac{f_{L,0}(t)}{b_L(t)} \mathbb{1}_{t \in A} \mathrm{d}t + O(\lambda_K K\mu_K^L) \\
&= (1 \pm C\varepsilon) \frac{T}{T_\ell^\Sigma} \int_0^{T_\ell^\Sigma} \left(\sum_{i=1}^{\ell} R_L^i \mathbb{1}_{t \in [T_{i-1}^\Sigma, T_i^\Sigma)} \right) \mathbb{1}_{t \in A} \mathrm{d}t + O(1/\lambda_K) + O(\lambda_K K\mu_K^L) \\
&= T(1 \pm C\varepsilon) R_L^{\mathrm{eff}} + o(1).
\end{aligned} \tag{C.4.90}$$

Here we utilize in the first equality the periodicity of all integrands and have to pay the error of counting at most one integral too much. In the second equality we make a change of variables to reduce from the K-dependent the functions $b_w^K(t), f_{w,0}^K$, to the unscaled versions $b_w(t), f_{w,0}$. The additional error of order $1/\lambda_K$ stems from the short $O(1)$ phases in the definition of $a_{\lfloor\alpha\rfloor}^{K,\pm}$. Finally we realize in the last step, that all errors vanish as $K \to \infty$ and we remember the definition of R_L^{eff} in (C.2.14).

Lemma C.4.5 now states that, if the L-population reaches a macroscopic size, it directly invades into the resident population and stabilizes near its equilibrium, with probability converging to one. All in all, this means that the appearance of a single mutant of trait L that grows and eventually invades and replaces the former resident population can be approximated by two exponentially distributed random variables with constant rate $(1 \pm C\varepsilon)R_L^{\mathrm{eff}}$, on the time scale $1/K\mu_K^L$. Compared to this, the total growth time between birth of the founding successful mutant and the final invasion time is of lower order, namely $O(\ln K) + O(1)$. Hence, it can be neglected and we can approximate the rescaled invasion time $T_{\mathrm{inv}}^{(K,\varepsilon)} K\mu_K^L$ by exponential random variables with rate $(1 \pm c\varepsilon)R_L^{\mathrm{eff}}$, by possibly enlarging the constant slightly. This observation implies the claim of Theorem C.2.2.

C.4.2. Proof of Theorem C.2.3

For the proof of Theorem C.2.3, we can re-use some parts of the previous section, with small extensions and refinements. To obtain better bounds on the resident population and hence the approximate invasion fitnesses, we replace the previous ε by an $\varepsilon_K \to 0$, satisfying $K^{-1/\max\{\alpha,2\}} \ll \varepsilon_K \ll \lambda_K{}^{-1}$. In particular, this yields

$$\mu_K \ll \varepsilon_K, \qquad \varepsilon_K^2 K \gg 1, \qquad \varepsilon_K \lambda_K \ll 1. \tag{C.4.91}$$

Assuming that the resident equilibrium sizes $\bar{n}_0^1 = \bar{n}_0^2$ coincide for both phases (Assumption 8(i)), one can show that Lemma C.4.1 still holds true when using such an ε_K and considering the slightly extended time horizon $T/K\mu_K^{L+1}$. Notably, the proof even slightly simplifies since there is no adaptation step at the beginning of each phase and one only need to apply a version of Corollary C.5.2. The introduction of a decaying ε_K is necessary to achieve better approximations for the invasion fitness and a precise time scale at the end.

Moreover, for the traits $v \in [\![1, \lfloor\alpha\rfloor]\!]$ close to the resident, Lemma C.4.2 is still valid when extending the time horizon to $T/K\mu_K^{L+1}$.

The crucial part of proving Theorem C.2.3 lies in analyzing the probability of a successful crossing of the valley. We proceed by first estimating the population size of the pit stop trait w population in Lemma C.4.6, dependent on the time the first mutant arises. Second, in Lemma C.4.7, we use this and arguments adapted from [29] to compute the probability that a single w-mutant induces the fixation of an L-mutant population.

Due to the Assumption 6, we can conclude exactly as in Lemma C.4.4 and C.4.5 that after fixation the L-population grows to a macroscopic size and finally replaces the resident population quickly.

Growth of the pit stop population

Recall that $T_{w,j}^K = \inf\left\{t \geq 0 : M_w^K(t) = j\right\}$ is the time when the j-th mutant the trait w is born as offspring of an individual of trait $w-1$. Since w is the only trait within the valley that has some phases of positive invasion fitness ($f_{w,0}^1 > 0$), the descendant population might start growing significantly. However, due to the negative average fitness ($f_{w,0}^{\mathrm{av}} < 0$) it is clear that it will die out again within one period. An explicit quantification is given in the following Lemma. As before, in Lemma C.4.4, we focus on the case of the first arriving w mutant. Due to separation of time scales, the results are transferable to all following mutants that occur before the invasion of the L trait.

Similar to before, we can couple the population process of trait w to branching processes such that

$$N_w^{(K,-)}(t) = N_w^K(t) = N_w^{(K,+)}(t) = 0, \quad \forall\, 0 \leq t < T_{w,1}^K, \tag{C.4.92}$$

$$N_w^{(K,-)}(t) \leq N_w^K(t) \leq N_w^{(K,+)}(t), \qquad \forall\, T_{w,1}^K \leq t \leq T_{w,2}^K \wedge S^{(K,\varepsilon_K)} \wedge T_\phi^{(K,\varepsilon_K)}, \tag{C.4.93}$$

where $N_w^{(K,-)}(T_{w,1}^K) = N_w^{(K,+)}(T_{w,1}^K) = N_w^K(T_{w,1}^K) = 1$ and the coupled processes follow the law of time-dependent birth death processes with rate functions

$$B_w^{(K,*)}(t) = b_w^K(t)(1-\mu_K), \tag{C.4.94}$$

$$D_w^{(K,*)}(t) = d_w^K(t) + c_{w,0}^K(t)\phi_0^{(K,\varepsilon_K,\bar{*})}(t) + \mathbb{1}_{*=-}\varepsilon_K \check{c}_w, \tag{C.4.95}$$

for $* \in \{+,-\}$ and $\bar{*}$ denoting the inverse sign. This coupling can be made explicit through a construction via Poisson measures, as outlined in the proof of Lemma C.4.2. We refer to this section for further details.

To formulate the growth results precisely, let us introduce the time-dependent invasion fitness for the coupled processes, as well as their integrals, which appear as the exponent of the expected population size.

$$f_{w,0}^{(K,\pm)}(t) := B_w^{(K,\pm)}(t) - D_w^{(K,\pm)}(t) = f_{w,0}^K(t) + O(\varepsilon_K) \tag{C.4.96}$$

$$h^{(K,\pm)}(t) := \int_{T_{w,1}^K}^t f_{w,0}^{(K,\pm)}(s)\mathrm{d}s \tag{C.4.97}$$

Note that $h^{(K,\pm)}$ depends on the random time $T_{w,1}^K$.

In what follows, we are only interested in the case of the first w-mutant being born in a fit phase, i.e. there exists an $n \in \mathbb{N}_0$ such that $\lambda_K n T_2^\Sigma \leq T_{w,1}^K < \lambda_K(nT_2^\Sigma + T_1)$. In this situation, we know that the function $h^{(K,\pm)}$ grows linearly with slope $f_{w,0}^1 + O(\varepsilon_K)$ until the change of phases. At that time, an approximate value of $\left(\lambda_K(nT_2^\Sigma + T_1) - T_{w,1}^K\right) f_{w,0}^1$ is reached. Afterwards, $h^{(K,\pm)}$ decays with approximate slope $f_{w,0}^2$ and crosses the x-axis before the end of the second phase. Let us denote this time by

$$T_{h=0}^K := \inf\left\{t > T_{w,1}^K : h^{(K,-)}(t) = 0\right\}, \tag{C.4.98}$$

which is the predicted time when the subpopulation of trait w becomes extinct again.

Lemma C.4.6. *Fix $\varepsilon > 0$ small enough, let the initial condition be given by Assumption 5 and let the fitness landscape satisfy Assumption 8. Assume that $T_{w,1}^K$ falls into a fit phase, i.e. there exists an $n \in N$ such that $\lambda_K n T_2^\Sigma \leq T_{w,1}^K < \lambda_K(nT_2^\Sigma + T_1) - \sqrt{\lambda_K}$. Then we have the following limit results:*

(a) (Fixation probability)

$$\lim_{K\to\infty} \mathbb{P}\left(N_w^{(K,-)}\left(T_{w,1}^K + \sqrt{\lambda_K}\right) \geq \mathrm{e}^{f_{w,0}^1\sqrt{\lambda_K}/2}\right) \geq \frac{f_{w,0}^1}{b_w^1}, \tag{C.4.99}$$

$$\lim_{K\to\infty} \mathbb{P}\left(N_w^{(K,+)}\left(T_{w,1}^K + \sqrt{\lambda_K}\right) = 0\right) \geq 1 - \frac{f_{w,0}^1}{b_w^1}. \tag{C.4.100}$$

(b) *(Initial boundedness) For every diverging sequence* $A_K \to \infty$,

$$\lim_{K\to\infty} \mathbb{P}\left(\forall t \in \left[T^K_{w,1}, T^K_{w,1}+\sqrt{\lambda_K}\right] : N^{(K,+)}_w(t) \le \mathrm{e}^{f^{(1,+)}_{w,0}(t-T^K_{w,1})} A_K\right) = 1. \tag{C.4.101}$$

(c) *(Short-term growth) There exist families of independent random variables* $\left(W^{(K,\pm)}_n\right)_{n\in\mathbb{N}}$ *with distribution*

$$W^{(K,\pm)}_n \overset{d}{=} \mathrm{Ber}\left(\frac{f^{(1,\pm)}_{w,0}}{b^1_w}\right) \otimes \mathrm{Exp}\left(\frac{f^{(1,\pm)}_{w,0}}{b^1_w}\right) \tag{C.4.102}$$

such that, for $0 < p_1 < 1 < p_2 < \infty$ *and* $I^K := \left[T^K_{w,1}+\sqrt{\lambda_K}, T^K_{h=0} - \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|\right]$,

$$\lim_{K\to\infty} \mathbb{P}\left(\forall t \in I^K : N^{(K,-)}_w(t) \ge p_1 \mathrm{e}^{h^{(K,-)}(t)} W^{(K,-)}_n\right) = 1, \tag{C.4.103}$$

$$\lim_{K\to\infty} \mathbb{P}\left(\forall t \in I^K : N^{(K,+)}_w(t) \le p_2 \mathrm{e}^{h^{(K,+)}(t)} W^{(K,+)}_n\right) = 1. \tag{C.4.104}$$

(d) *(Extinction)*

$$\lim_{K\to\infty} \mathbb{P}\left(N^{(K,+)}_w\left(T^K_{h=0} + \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|\right) = 0\right) = 1. \tag{C.4.105}$$

(e) *(Final boundedness) There exists a constant* $C < \infty$*, such that, for* $J^K := [T^K_{h=0} - \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|, T^K_{h=0} + \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|]$,

$$\lim_{K\to\infty} \mathbb{P}\left(\sup_{t\in J^K} N^{(K,+)}_w(t) \le \mathrm{e}^{C\sqrt{\lambda_K}}\right) = 1. \tag{C.4.106}$$

Proof. (a) We imitate the strategy of Step 6 in the proof of Lemma C.4.4 and improve the estimates slightly. To this end, fix some $\varepsilon > 0$ and let

$$B^{(\varepsilon,-)}_w = b^1_w(1-\varepsilon), \qquad D^{(\varepsilon,-)}_w = d^1_w + c^1_{w,0}\bar{n}_0 + \varepsilon\left(M c^1_{w,0} + \check{c}_w\right), \tag{C.4.107}$$

be the time-independent rates of a birth death process $Z^{(\varepsilon,-)}$. Moreover, set $f^{(\varepsilon,-)}_w = B^{(\varepsilon,-)}_w - D^{(\varepsilon,-)}_w$. Then $f^{(\varepsilon,-)}_w > f^1_{w,0}/2$ for $\varepsilon > 0$ small enough.

Since $\varepsilon_K \to 0$ and $\mu_K \to 0$, this process $(Z^{(\varepsilon,-)}(s))_{s\ge 0}$ is stochastically dominated by the processes $(N_w^{(K,-)}(T_{w,1}^K+s))_{s\geq 0}$, for K large enough. Therefore we can estimate

$$\begin{aligned}\lim_{K\to\infty}\mathbb{P}\left(N_w^{(K,-)}\left(T_{w,1}^K+\sqrt{\lambda_K}\right)\geq \mathrm{e}^{f_{w,0}^1\sqrt{\lambda_K}/2}\right) &\geq \lim_{s\to\infty}\mathbb{P}\left(Z^{(\varepsilon,-)}(s)\mathrm{e}^{-f_w^{(\varepsilon,-)}s}\geq \mathrm{e}^{\left(f_{w,0}^1/2-f_w^{(\varepsilon,-)}\right)s}\right)\\ &\geq \limsup_{\delta\downarrow 0}\lim_{s\to\infty}\mathbb{P}\left(Z^{(\varepsilon,-)}(s)\mathrm{e}^{-f_w^{(\varepsilon,-)}s}\geq\delta\right)\\ &=\limsup_{\delta\downarrow 0}\mathbb{P}\left(\lim_{s\to\infty}Z^{(\varepsilon,-)}(s)\mathrm{e}^{-f_w^{(\varepsilon,-)}s}\geq\delta\right)\\ &=\limsup_{\delta\downarrow 0}\mathbb{P}\left(W^{(\varepsilon,-)}\geq\delta\right)=\mathbb{P}\left(W^{(\varepsilon,-)}>0\right)\\ &=\frac{f_w^{(\varepsilon,-)}}{B_w^{(\varepsilon,-)}}=\frac{f_{w,0}^1-C\varepsilon}{b_w^1(1-\varepsilon)}. \qquad \text{(C.4.108)}\end{aligned}$$

Here, $W^{(\varepsilon,-)}=\lim_{s\to\infty}Z^{(\varepsilon,-)}(s)\mathrm{e}^{-f_w^{(\varepsilon,-)}s}$, as in Theorems 1 and 2 in Chapter III.7 of [4]. This gives a lower bound for every $\varepsilon>0$, and the limit on the left hand side is independent of ε. Hence, we can take the limes superior as $\varepsilon\downarrow 0$ of the inequality to obtain the sharp bound that is claimed in the lemma.

The opposite estimate for $N^{(K,+)}$ can be shown in the same manner (similar to Step 5 in the proof of Lemma C.4.4), under use of the birth death process corresponding to the rates

$$B_w^{(\varepsilon,+)}=b_w^1, \qquad D_w^{(\varepsilon,+)}=d_w^1+c_{w,0}^1\bar{n}_0-\varepsilon Mc_{w,0}^1. \qquad \text{(C.4.109)}$$

(b) The proof is similar to Lemma C.1 in [71] and relies on an application of Doob's maximum inequality to the the rescaled martingales $\hat{M}^{(K,+)}(t)=\mathrm{e}^{-h^{(K,+)}(t)}N_w^{(K,+)}(t)$. By assumption, the considered time interval is entirely part of the fit 1-phase. For the counter event of the desired probability, we hence obtain

$$\begin{aligned}&\mathbb{P}\left(\exists\ t\in\left[T_{w,1}^K,T_{w,1}^K+\sqrt{\lambda_K}\right]: N_w^{(K,+)}(t)>\mathrm{e}^{f_{w,0}^{(1,+)}(t-T_{w,1}^K)}A_K\right)\\ &=\mathbb{P}\left(\sup_{t\in\left[T_{w,1}^K,T_{w,1}^K+\sqrt{\lambda_K}\right]}\left|\hat{M}^{(K,+)}(t)\right|>A_K\right)\\ &\leq A_K^{-2}\mathbb{E}\left[\left\langle\hat{M}^{(K,+)}\right\rangle_{T_{w,1}^K+\sqrt{\lambda_K}}\right]\\ &=A_K^{-2}C\int_{T_{w,1}^K}^{T_{w,1}^K+\sqrt{\lambda_K}}\mathrm{e}^{-f_{w,0}^{(1,+)}(t-T_{w,1}^K)}\mathrm{d}t\\ &=A_K^{-2}\frac{C}{f_{w,0}^{(1,+)}}\left(1-\mathrm{e}^{-f_{w,0}^{(1,+)}\sqrt{\lambda_K}}\right)\xrightarrow{K\to\infty}0, \qquad \text{(C.4.110)}\end{aligned}$$

which proves the claim.

(c) Again, this proof is similar to Lemma C.1 in [71], this time applying Doob's maximum inequality to both rescaled martingales $\hat{M}^{(K,\pm)}(t)=\mathrm{e}^{-h^{(K,\pm)}(t)}N_w^{(K,\pm)}(t)$. As a preparation we remind ourselves of the results of [4, Ch. III.7], already mentioned in the proof of part (a), from which we deduce that, at the divergent time $T^K_{w,1} + \sqrt{\lambda_K}$, $\hat{M}^{(K,\pm)}$ is close to a random variable $W_n^{(K,\pm)}$. It has been shown in [68] that this random variable has exactly the distribution stated in this lemma.

Let us focus on the first claim and consider the counter event. It suffices to condition on the non-extinction-event from part (a) since under extinction the claim is trivial. Instead of comparing directly to $W^{(K,-)}$ we insert the exact rescaled population size of $\hat{M}^{(K,-)}$ at the initial time of the interval. We use the short notation $I^K = \left[T^K_{w,1} + \sqrt{\lambda_K}, T^K_{h=0} - \sqrt{\lambda_K} f^1_{w,0} / \left| f^2_{w,0} \right|\right]$, as introduced in the lemma. Then we have

$$
\begin{aligned}
&\mathbb{P}\left(\exists\, t \in I^K : N_w^{(K,-)}(t) < p_1 \mathrm{e}^{h^{(K,-)}(t)} \mathrm{e}^{-h^{(K,-)}(T^K_{w,1}+\sqrt{\lambda_K})} N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)\right.\\
&\qquad\left.\Big| N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)\right)\\
\leq{}& \mathbb{P}\left(\sup_{t \in I^K} \left| \mathrm{e}^{-h^{(K,-(t)}} N_w^{(K,-)}(t) - \mathrm{e}^{-h^{(K,-)}(T^K_{w,1}+\sqrt{\lambda_K})} N_w^{(K,-)}(T^K_{w,1} + \sqrt{\lambda_K}) \right|\right.\\
&\qquad\left. > (1-p_1)\mathrm{e}^{-h^{(K,-)}\left(T^K_{w,1}+\sqrt{\lambda_K}\right)} N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right) \Big| N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)\right)\\
={}& \mathbb{P}\left(\sup_{t \in I^K} \left| \hat{M}^{(K,-)}(t) - \hat{M}^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right) \right|\right.\\
&\qquad\left. > (1-p_1)\mathrm{e}^{-h^{(K,-)}\left(T^K_{w,1}+\sqrt{\lambda_K}\right)} N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right) \Big| N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)\right)\\
\leq{}& \frac{\mathbb{E}\left[\left\langle \hat{M}^{(K,-)} \right\rangle_{T^K_{h=0}-\sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|} - \left\langle \hat{M}^{(K,-)} \right\rangle_{T^K_{w,1}+\sqrt{\lambda_K}} \Big| N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)\right]}{\left[(1-p_1)\mathrm{e}^{-h^{(K,-)}\left(T^K_{w,1}+\sqrt{\lambda_K}\right)} N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)\right]^2}\\
={}& C \frac{(1-p_1)\mathrm{e}^{-h^{(K,-)}\left(T^K_{w,1}+\sqrt{\lambda_K}\right)} N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)}{\left[(1-p_1)\mathrm{e}^{-h^{(K,-)}\left(T^K_{w,1}+\sqrt{\lambda_K}\right)} N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right)\right]^2} \int_{T^K_{w,1}+\sqrt{\lambda_K}}^{T^K_{h=0}-\sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|} \mathrm{e}^{-h^{(K,-)}(t)} \mathrm{d}t\\
\leq{}& C \mathrm{e}^{-f^1_{w,0}\sqrt{\lambda_K}/2} \mathrm{e}^{f^{(1,-)}_{w,0}\sqrt{\lambda_K}} \left(\frac{1}{f^{(1,-)}_{w,0}} + \frac{1}{\left|f^{(2,-)}_{w,0}\right|}\right) \left(\mathrm{e}^{-f^{(1,-)}_{w,0}\sqrt{\lambda_K}} - \mathrm{e}^{-f^{(1,-)}_{w,0}\left[\lambda_K\left(nT^\Sigma_2+T_1\right)-T^K_{w,1}\right]}\right)\\
={}& C \mathrm{e}^{-f^1_{w,0}\sqrt{\lambda_K}/2} \left(1 - \mathrm{e}^{-f^{(1,-)}_{w,0}\left[\lambda_K\left(nT^\Sigma_2+T_1\right)-\left(T^K_{w,1}+\sqrt{\lambda_K}\right)\right]}\right)\\
\leq{}& C \mathrm{e}^{-f^1_{w,0}\sqrt{\lambda_K}/2} \overset{K\to\infty}{\longrightarrow} 0,
\end{aligned}
\tag{C.4.111}
$$

where the value of the constant $C < \infty$ changes between lines. Here we apply Doob's maximum-inequality, make use of the bracket computations in [71] and recall that $\varepsilon_K \lambda_K \ll 1$. Finally, we utilize that, on the non-extinction-event, the population size at the beginning of the interval can be bounded from below by $N_w^{(K,-)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right) > \mathrm{e}^{f^1_{w,0}\sqrt{\lambda_K}/2}$, for K large enough.

The second claim is proven analogously.

(d) Using part (c) and noticing that

$$\lim_{K\to 0} h^{(K,+)}\left(T^K_{h=0} - \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|\right) - h^{(K,+)}\left(T^K_{w,1} + \sqrt{\lambda_K}\right) = 0, \tag{C.4.112}$$

since $\varepsilon_K \lambda_K \ll 1$, we can bound the number of individuals being alive shortly before we expect extinction from above by

$$\lim_{K\to\infty} \mathbb{P}\left(N_w^{(K,+)}\left(T^K_{h=0} - \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|\right) \le \mathrm{e}^{f^{(1,+)}_{w,0}\sqrt{\lambda_K}} A_K\right) = 1, \tag{C.4.113}$$

for every diverging sequence $A_K \to \infty$.

We then remind ourselves that within the time interval

$$\left[T^K_{h=0} - \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|, T^K_{h=0} + \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|\right] \tag{C.4.114}$$

the process $N_w^{(K,+)}$ is just a subcritical birth death process with parameters $B_w^{(2,+)} < D_w^{(2,+)}$. If we denote by $Z^{(K,+)}$ a process with the same birth and death rates but initialized with a single individual, i.e. $Z^{(K,+)}(0) = 1$, it is well known for the probability of extinction up to time t (cf. [95]) that

$$\mathbb{P}\left(Z^{(K,+)}(t) = 0 \middle| Z^{(K,+)}(0) = 1\right) = 1 - \frac{\left|f^{(2,+)}_{w,0}\right| \mathrm{e}^{f^{(2,+)}_{w,0} t}}{D_w^{(2,+)} - B_w^{(2,+)} \mathrm{e}^{f^{(2,+)}_{w,0} t}}. \tag{C.4.115}$$

Since the families of all individuals alive at the beginning of the interval evolve independently of each other, we can estimate the probability of extinction by

$$\begin{aligned}
&\mathbb{P}\left(N_w^{(K,+)}\left(T^K_{h=0} + \sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|\right) = 0\right) \\
&\ge \left[\mathbb{P}\left(Z^{(K,+)}\left(2\sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|\right) = 0\right)\right]^{\mathrm{e}^{f^{(1,+)}_{w,0}\sqrt{\lambda_K}} A_K} \\
&= \left[1 - \frac{\left|f^{(2,+)}_{w,0}\right| \mathrm{e}^{f^{(2,+)}_{w,0} 2\sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|}}{D_w^{(2,+)} - B_w^{(2,+)} \mathrm{e}^{f^{(2,+)}_{w,0} 2\sqrt{\lambda_K} f^1_{w,0}/\left|f^2_{w,0}\right|}}\right]^{\mathrm{e}^{f^{(1,+)}_{w,0}\sqrt{\lambda_K}} A_K} \\
&\approx \left[1 - \frac{\left|f^{(2,+)}_{w,0}\right| \mathrm{e}^{-2 f^1_{w,0}\sqrt{\lambda_K}}}{D_w^{(2,+)} - B_w^{(2,+)} \mathrm{e}^{-2 f^1_{w,0}\sqrt{\lambda_K}}}\right]^{\mathrm{e}^{f^1_{w,0}\sqrt{\lambda_K}} A_K}
\end{aligned} \tag{C.4.116}$$

Here we used in the last line, that $f^{(1,+)}_{w,0} = f^1_{w,0}(1 + C\varepsilon_K)$ and $f^{(2,+)}_{w,0} = f^2_{w,0}(1 - C\varepsilon_K)$ as well as $\varepsilon_K \sqrt{\lambda_K} \ll 1$. As the only condition on A_K is to be a diverging sequence, we choose $A_K := \mathrm{e}^{\frac{1}{2} f^1_{w,0}\sqrt{\lambda_K}}$. Then, for K large enough, the above probability can be bounded by

$$\left[1 - \frac{\left|f^{(2,+)}_{w,0}\right| A_K^{-4}}{D_w^{(2,+)} - B_w^{(2,+)} A_K^{-4}}\right]^{A_K^3} \ge \left[1 - \frac{\left|f^{(2,+)}_{w,0}\right|}{D_w^{(2,+)} A_K^4/2}\right]^{A_K^3} \xrightarrow{K\to\infty} \mathrm{e}^0 = 1. \tag{C.4.117}$$

(e) The strategy is the same as in the proof of Lemma A.1 in [45], Step 3(iii). As already seen in the proof of part (d), the number of individuals alive at time $T^K_{h=0} - \sqrt{\lambda_K} f^1_{w,0} / \left| f^2_{w,0} \right|$ is bounded from above by $A_K \mathrm{e}^{f^{(1,+)}_{w,0} \sqrt{\lambda_K}}$, which is still diverging. By the nature of branching processes, we can consider the evolving family of each individual at this time independently. Now we disregard possible death events, which leads to a collection of independent Yule-processes Y_i with birth rate b^2_w since the considered time interval lies entirely within the second parameter phase. Since the Y_i are monotonously increasing, it is sufficient to look at their endpoints. We use the fact that these have the same distribution as iid. geometric random variables $G_i \sim \mathrm{Geo}(p)$ with $p = \mathrm{e}^{-2\sqrt{\lambda_K} b^2_w f^1_{w,0} / \left| f^2_{w,0} \right|}$. An application of the law of large numbers finally yields

$$\lim_{K \to \infty} \mathbb{P} \left(\frac{\sum_{i=0}^{A_K \mathrm{e}^{f^{(1,+)}_{w,0} \sqrt{\lambda_K}}} \left(G_i - \mathrm{e}^{2\sqrt{\lambda_K} b^2_w f^1_{w,0} / \left| f^2_{w,0} \right|} \right)}{A_K \mathrm{e}^{f^{(1,+)}_{w,0} \sqrt{\lambda_K}}} \leq \mathrm{e}^{2\sqrt{\lambda_K} b^2_w f^1_{w,0} / \left| f^2_{w,0} \right|} \right) = 1. \tag{C.4.118}$$

Choosing A_K appropriately and rearranging this estimate allows us finally to conclude the claim. □

Crossing the fitness valley and fixation

Lemma C.4.7. *Let the initial condition be given by Assumption 5 and let the fitness landscape satisfy Assumption 8. Then there exists a* $C < \infty$ *such that, for every* $\varepsilon > 0$ *small enough, for all* $0 < p_1 < 1 < p_2 < \infty$*, and* K *large enough,*

$$\mathcal{P}^K(T^K_{w,1}) = \mathbb{P} \left(T^K_{w,1} < T^{(K,\varepsilon)}_{\mathrm{inv}} < T^K_{w,1} + \frac{1 + C\varepsilon}{f^{\mathrm{av}}_{L,0}} \ln K \;\middle|\; T^K_{w,1} \right). \tag{C.4.119}$$

satisfies

$$\mathcal{P}^K(T^K_{w,1}) \geq \sum_{n=0}^{\infty} \Bigg[\mathbb{1}_{\lambda_K n T^\Sigma_2 \leq T^K_{w,1} < \lambda_K (n T^\Sigma_2 + T_1)} p_1 W^{(K,-)}_n \mu_K^{L-w} \tag{C.4.120}$$
$$\times \left(\frac{b^1_w}{f^{(1,-)}_{w,0}} \Lambda^1 \frac{f^1_{L,0}}{b^1_L} + \frac{b^2_w}{\left| f^{(2,-)}_{w,0} \right|} \Lambda^2 \frac{f^2_{L,0}}{b^2_L} \right) \left(\mathrm{e}^{f^{(1,-)}_{w,0} \left[\lambda_K (n T^\Sigma_2 + T_1) - T^K_{w,1} \right]} - 1 \right) \Bigg],$$

$$\mathcal{P}^K(T^K_{w,1}) \leq \sum_{n=0}^{\infty} \Bigg[\mathbb{1}_{\lambda_K n T^\Sigma_2 \leq T^K_{w,1} < \lambda_K (n T^\Sigma_2 + T_1)} p_2 W^{(K,+)}_n \mu_K^{L-w} \tag{C.4.121}$$
$$\times \left(\frac{b^1_w}{f^{(1,+)}_{w,0}} \Lambda^1 \frac{f^1_{L,0}}{b^1_L} + \frac{b^2_w}{\left| f^{(2,+)}_{w,0} \right|} \Lambda^2 \frac{f^2_{L,0}}{b^2_L} \right) \left(\mathrm{e}^{f^{(1,+)}_{w,0} \left[\lambda_K (n T^\Sigma_2 + T_1) - T^K_{w,1} \right]} - 1 \right) \Bigg],$$

where $W^{(K,\pm)}_n$ *are the same iid. random variables as in Lemma C.4.6.*

Proof. In order to simplify the following proof, we do not document the exact form of the approximation error in every step. All estimates stemming from the branching process approximation enter the proposed probability as a multiplicative error of the form $(1 \pm C\varepsilon_K)$ and can hence, for large K, be treated by slightly changing the choices of p_1 and p_2.

We know from [29], for the case of fixed environments $i = 1, 2$, that once a mutant of trait $w+1$ is born the probability of accumulating further mutants during the subcritical excursions and finally producing a mutant of trait L is approximately given by

$$\mu_K^{L-w-1} \prod_{v=w+1}^{L-1} \lambda(\rho_v^i) = \mu_K^{L-w-1} \frac{b_{w+1}^i \cdots b_{L-1}^i}{\left|f_{w+1}^i\right| \cdots \left|f_{L-1}^i\right|} =: \mu_K^{L-w-1} \Lambda^i. \tag{C.4.122}$$

Moreover, we know that, if this happens, the transition of the valley takes only a short time of order $O(1)$ and thus is finished within one phase. This inspires us to introduce a time-dependent periodic version of this probability in the same way as for previous quantities:

$$\Lambda^K(t) := \begin{cases} \Lambda^1 & : t \in [0, \lambda_K T_1), \\ \Lambda^2 & : t \in [\lambda_K T_1, \lambda_K T_2^\Sigma). \end{cases} \tag{C.4.123}$$

The next question we address is the probability that a w-mutant born in the first phase (which is the fit one) leads to a successfully growing population of trait L. To this end we use the results of the preceding lemma to estimate the size N_w^K of the founded w-subpopulation until its extinction, which is with high probability before the end of the period. During this time, it produces $w+1$-mutants at rate $N_w^K(t) b_w^K(t) \mu_K$. These mutants then get thinned by the probability $\Lambda^K(t)$ and moreover we have to multiply by the probability that an L-mutant fixates and invades successfully, which is the well known fixation probability $(f_{L,0}^K(t))_+ / b_L^K(t)$. Overall, we obtain that, in the case of $\lambda_K n T_2^\Sigma \leq T_{w,1}^K < \lambda_K(nT_2^\Sigma + T_1) - \sqrt{\lambda_K}$, the probability to observe a fixating L-subpopulation is approximately

$$\mathcal{P}_n^K(T_{w,1}^K) := \int_{\lambda_K n T_2^\Sigma}^{\lambda_K (n+1) T_2^\Sigma} N_w^K(t) b_w^K(t) \mu_K \mu_K^{L-w-1} \Lambda^K(t) \frac{f_{L,0}^K(t)}{b_L^K(t)} \mathrm{d}t. \tag{C.4.124}$$

Our first observation is that the population size N_w^K vanishes before $T_{w,1}^K$ and shortly after $T_{h=0}^K$ by definition and part (d) of Lemma C.4.6, respectively. Moreover, it is not hard to see from part (b) and (e) of the same lemma that the contribution of the intervals

$$\left[T_{w,1}^K, T_{w,1}^K + \sqrt{\lambda_K}\right] \quad \text{and} \quad \left[T_{h=0}^K - \sqrt{\lambda_K} \frac{f_{w,0}^1}{\left|f_{w,0}^2\right|}, T_{h=0}^K + \sqrt{\lambda_K} \frac{f_{w,0}^1}{\left|f_{w,0}^2\right|}\right] \tag{C.4.125}$$

is negligible compared to the rest of the integral.

On the remaining interval, we can use the bounds of part (c) of the lemma already mentioned to estimate $N_w^K(t) \leq \mathrm{e}^{h^{(K,+)}(t)} p_2 W_n^{(K,+)}$, with high probability. Inserting this bound into the integral yields, with probability converging to 1, as $K \to \infty$,

$$\begin{aligned}\mathcal{P}_n^K(T_{w,1}^K) &\leq \int_{\lambda_K n T_2^\Sigma}^{\lambda_K (n+1) T_2^\Sigma} p_2 W_n^{(K,+)} \mathrm{e}^{h^{(K,+)}(t)} b_w^K(t) \mu_K^{L-w} \Lambda^K(t) \frac{f_{L,0}^K(t)}{b_L^K(t)} \mathrm{d}t \qquad \text{(C.4.126)} \\ &= p_2 W_n^{(K,+)} \mu_K^{L-w} \left(\int_{T_{w,1}^K}^{\lambda_K (n T_2^\Sigma + T_1)} \mathrm{e}^{h^{(K,+)}(t)} b_w^1 \Lambda^1 \frac{f_{L,0}^1}{b_L^1} \mathrm{d}t + \int_{\lambda_K (n T_2^\Sigma + T_1)}^{T_{h=0}^K} \mathrm{e}^{h^{(K,+)}(t)} b_w^2 \Lambda^2 \frac{f_{L,0}^2}{b_L^2} \mathrm{d}t \right).\end{aligned}$$

We notice that the only non-constant term in both integrals is $h^{(K,+)}(t)$, which is piecewise linear. To be precise, in the first integral it growths linearly with slope $f_{w,0}^{(1,+)} > 0$ starting at 0 and decays in the second integral with slope $f_{w,0}^{(2,+)} < 0$ until getting close to 0 again. Therefore, evaluating the integrals gives, with probability converging to 1,

$$\mathcal{P}_n^K(T_{w,1}^K) \leq p_2 W_n^{(K,+)} \mu_K^{L-w} \left(\frac{b_w^1}{f_{w,0}^{(1,+)}} \Lambda^1 \frac{f_{L,0}^1}{b_L^1} + \frac{b_w^2}{\left| f_{w,0}^{(2,+)} \right|} \Lambda^2 \frac{f_{L,0}^2}{b_L^2} \right) \left(\mathrm{e}^{f_{w,0}^{(1,+)} \left[\lambda_K (n T_2^\Sigma + T_1) - T_{w,1}^K \right]} - 1 \right). \tag{C.4.127}$$

By the same strategy we achieve a lower bound.

$$\mathcal{P}_n^K(T_{w,1}^K) \geq p_1 W_n^{(K,-)} \mu_K^{L-w} \left(\frac{b_w^1}{f_{w,0}^{(1,-)}} \Lambda^1 \frac{f_{L,0}^1}{b_L^1} + \frac{b_w^2}{\left| f_{w,0}^{(2,-)} \right|} \Lambda^2 \frac{f_{L,0}^2}{b_L^2} \right) \left(\mathrm{e}^{f_{w,0}^{(1,-)} \left[\lambda_K (n T_2^\Sigma + T_1) - T_{w,1}^K \right]} - 1 \right). \tag{C.4.128}$$

We notice that the dependency on $T_{w,1}^K$, only enters the bounds for $\mathcal{P}_n^K(T_{w,1}^K)$ in the difference $\lambda_K(nT_2^\Sigma + T_1) - T_{w,1}^K$. Consequently, only the point of time within the parameter phase, and not the cycle n, is important. Hence we can approximate $\mathcal{P}^K(t)$ by

$$\sum_{n=0}^{\infty} \mathcal{P}_n^K(t) \mathbb{1}_{\lambda_K n T_2^\Sigma \leq t < \lambda_K (n T_2^\Sigma + T_1)} \tag{C.4.129}$$

and conclude the claim. □

We can now argue to conclude the final result of Theorem C.2.3. The function $\mathcal{P}^K(t)$ in Lemma C.4.7 can be seen as a thinning probability of the arrival rate of w-mutants. Moreover, let us notice that mutants arriving in the second phase of a period are always unfit and thus get thinned by a probability that is of strictly lower order than the $\mathcal{P}^K$, namely μ_K^{L-w}. We can hence neglect those cases. By Lemma C.4.3, new w-mutants are known to occur approximately as a Poisson process with rate function

$$K \mu_K^w a_{\lfloor \alpha \rfloor}^{(K,\pm)}(t) b_{\lfloor \alpha \rfloor}^K(t) \prod_{v=\lfloor \alpha \rfloor + 1}^{w-1} \frac{b_v^K(t)}{\left| f_{v,0}^K(t) \right|} (1 \pm C\varepsilon). \tag{C.4.130}$$

Hence the birth times of successfully invading L-mutants follow approximately a Poisson process with intensity function $\mathcal{R}^K$, which we can estimate by the product of the above terms,

$$\begin{aligned}\mathcal{R}^{(K,\pm)}(t) = K\mu_K^L p_*(1 \pm C\varepsilon) a_{\lfloor\alpha\rfloor}^{(K,\pm)}(t) b_{\lfloor\alpha\rfloor}^K(t) &\left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1} \frac{b_v^K(t)}{\left|f_{v,0}^K(t)\right|} \right) \\ &\times \left(\frac{b_w^1}{f_{w,0}^{(1,\pm)}} \Lambda^1 \frac{f_{L,0}^1}{b_L^1} + \frac{b_w^2}{\left|f_{w,0}^{(2,\pm)}\right|} \Lambda^2 \frac{f_{L,0}^2}{b_L^2} \right) \\ &\times \sum_{n=0}^{\infty} \mathbb{1}_{\lambda_K n T_2^\Sigma \leq t < \lambda_K (n T_2^\Sigma + T_1)} W_n^{(K,\pm)} \left(\mathrm{e}^{f_{w,0}^{(1,\pm)} \left[\lambda_K (n T_2^\Sigma + T_1) - t\right]} - 1 \right).\end{aligned} \tag{C.4.131}$$

Here we set $p_* = p_1$ for the lower bound and $p_* = p_2$ for the upper bound. Due to our previous observations, this is almost a periodic function. It differs between the periods only by the iid. random variables $W_n^{(K,\pm)}$. Moreover, we see that the leading order term is of order $K\mu_K^L \mathrm{e}^{f_{w,0}^1 T_1 \lambda_K} \ll 1$, when integratet over one period of length $\lambda_K T_2^\Sigma$. We therefore expect the first successful L-mutant to be born on the time scale $\lambda_K \mathrm{e}^{-f_{w,0}^1 T_1 \lambda_K} / K\mu_K^L$. As argued in the final step of proof of Theorem C.2.2, the periodic variations of the intensity function average out since these are on the much shorter time scale λ_K. Effectively, for every $T < \infty$, we compute the Poisson intensity of successfully fixating L-mutants by

$$\begin{aligned}&\int_0^{T\lambda_K \mathrm{e}^{-f_{w,0}^1 T_1 \lambda_K} / K\mu_K^L} \mathcal{R}^{(K,\pm)}(t) \mathrm{d}t \\ =&p_*(1 \pm C\varepsilon) K\mu_K^L \left(\frac{b_w^1}{f_{w,0}^{(1,\pm)}} \Lambda^1 \frac{f_{L,0}^1}{b_L^1} + \frac{b_w^2}{\left|f_{w,0}^{(2,\pm)}\right|} \Lambda^2 \frac{f_{L,0}^2}{b_L^2} \right) \\ &\times \int_0^{T\lambda_K \mathrm{e}^{-f_{w,0}^1 T_1 \lambda_K} / K\mu_K^L} \sum_{n=0}^{\infty} \mathbb{1}_{\lambda_K n T_2^\Sigma \leq t < \lambda_K (n T_2^\Sigma + T_1)} a_{\lfloor\alpha\rfloor}^{(K,\pm)}(t) b_{\lfloor\alpha\rfloor}^K(t) \\ &\qquad \left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1} \frac{b_v^K(t)}{\left|f_{v,0}^K(t)\right|} \right) W_n^{(K\pm)} \left(\mathrm{e}^{f_{w,0}^{(1,\pm)} \left[\lambda_K (n T_2^\Sigma + T_1) - t\right]} - 1 \right) \mathrm{d}t.\end{aligned} \tag{C.4.132}$$

If we now focus on the integral term, this can be rewritten and bounded from below by

$$\begin{aligned}
&\sum_{n=0}^{\left\lfloor \frac{T}{T_2^\Sigma} \frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu_K^L}\right\rfloor} W_n^{(K,-)} \int_{\lambda_K n T_2^\Sigma}^{\lambda_K(nT_2^\Sigma+T_1)} a_{\lfloor\alpha\rfloor}^{(K,-)}(t) b_{\lfloor\alpha\rfloor}^K(t) \left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1} \frac{b_v^K(t)}{\left|f_{v,0}^K(t)\right|}\right)\left(\mathrm{e}^{f_{w,0}^{(1,-)}\left[\lambda_K(nT_2^\Sigma+T_1)-t\right]}-1\right)\mathrm{d}t \\
&= \sum_{n=0}^{\left\lfloor \frac{T}{T_2^\Sigma} \frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu_K^L}\right\rfloor} W_n^{(K,-)}\lambda_K \int_0^{T_1} a_{\lfloor\alpha\rfloor}^{(K,-)}(t\lambda_K) b_{\lfloor\alpha\rfloor}^K(t\lambda_K) \left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1} \frac{b_v^K(t\lambda_K)}{\left|f_{v,0}^K(t\lambda_K)\right|}\right)\left(\mathrm{e}^{f_{w,0}^{(1,-)}[T_1\lambda_K-t\lambda_K]}-1\right)\mathrm{d}t \\
&= \sum_{n=0}^{\left\lfloor \frac{T}{T_2^\Sigma} \frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu_K^L}\right\rfloor} W_n^{(K,-)}\lambda_K \left[\int_0^{T_1} a_{\lfloor\alpha\rfloor}^{(1,-)} b_{\lfloor\alpha\rfloor}^1 \left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1} \frac{b_v^1}{\left|f_{v,0}^1\right|}\right)\left(\mathrm{e}^{f_{w,0}^{(1,-)}\lambda_K(T_1-t)}-1\right)\mathrm{d}t \right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\left. + O\left(\mathrm{e}^{f_{w,0}^{(1,\pm)}\sum_{w=0}^{\lfloor\alpha\rfloor}\tau_w^\varepsilon}/\lambda_K\right)\right] \\
&= \sum_{n=0}^{\left\lfloor \frac{T}{T_2^\Sigma} \frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu_K^L}\right\rfloor} W_n^{(K,-)}\lambda_K \left[a_{\lfloor\alpha\rfloor}^{(1,-)} b_{\lfloor\alpha\rfloor}^1 \left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1} \frac{b_v^1}{\left|f_{v,0}^1\right|}\right)\left(\frac{1}{f_{w,0}^{(1,-)}\lambda_K}\left[\mathrm{e}^{f_{w,0}^{(1,-)}\lambda_K T_1}-1\right]-T_1\right)\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\left. + O\left(\mathrm{e}^{f_{w,0}^{(1,-)}\sum_{w=0}^{\lfloor\alpha\rfloor}\tau_w^\varepsilon}/\lambda_K\right)\right] \\
&= a_{\lfloor\alpha\rfloor}^{(1,-)} b_{\lfloor\alpha\rfloor}^1 \left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1} \frac{b_v^1}{\left|f_{v,0}^1\right|}\right) \frac{1}{f_{w,0}^{(1,-)}} \mathrm{e}^{f_{w,0}^{(1,-)}\lambda_K T_1} \sum_{n=0}^{\left\lfloor \frac{T}{T_2^\Sigma} \frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu_K^L}\right\rfloor} W_n^{(K\pm)} + o\left(1/K\mu_K^L\right).
\end{aligned} \tag{C.4.133}$$

Here, in the first equality, we used the periodicity of the integrands and a change of variables. In the second step, we reduced the K-dependent functions $a_{\lfloor\alpha\rfloor}^{(K,\pm)}, b_v^K, f_{v,0}^K$ to their unscaled versions, which are constant. Note that this comes at the expense of adding an error of order $O(\mathrm{e}^{f_{w,0}^{(1,\pm)}\sum_{w=0}^{\lfloor\alpha\rfloor}\tau_w^\varepsilon}/\lambda_K)$, stemming from the short $O(1)$ phases of adaptation in the definition of $a_{\lfloor\alpha\rfloor}^{(K,\pm)}$. Finally, we just rearrange the constant prefactor at the front of the sum and estimate the lower order terms.

A corresponding upper bound is obtained by considering the sum running up to $\left\lceil \frac{T}{T_2^\Sigma} \frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu_K^L}\right\rceil$ and using the parameters $a_{\lfloor\alpha\rfloor}^{(1,+)}$ etc., corresponding to the upper bounding branching process.

Putting things together, we obtain

$$\begin{aligned}
&\int_0^{T\lambda_K \mathrm{e}^{-f^1_{w,0}T_1\lambda_K}/K\mu^L_K} \mathcal{R}^{(K,\pm)}(t)\mathrm{d}t \\
&\geq p_1(1-C\varepsilon)\left(\frac{b^1_w}{f^{(1,-)}_{w,0}}\Lambda^1\frac{f^1_{L,0}}{b^1_L}+\frac{b^2_w}{\left|f^{(2,-)}_{w,0}\right|}\Lambda^2\frac{f^2_{L,0}}{b^2_L}\right)a^{(1,-)}_{\lfloor\alpha\rfloor}b^1_{\lfloor\alpha\rfloor}\left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1}\frac{b^1_v}{\left|f^1_{v,0}\right|}\right)\frac{1}{f^{(1,-)}_{w,0}} \\
&\qquad\times K\mu^L_K\mathrm{e}^{f^{(1,-)}_{w,0}\lambda_K T_1}\sum_{n=0}^{\left\lfloor\frac{T}{T^\Sigma_2}\frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu^L_K}\right\rfloor}W^{(K,-)}_n+o(1) \\
&=p_1(1-C\varepsilon)a^{(1,-)}_{\lfloor\alpha\rfloor}b^1_{\lfloor\alpha\rfloor}\left(\prod_{v=\lfloor\alpha\rfloor+1}^{w-1}\frac{b^1_v}{\left|f^1_{v,0}\right|}\right)\frac{1}{f^{(1,-)}_{w,0}}\left(\frac{b^1_w}{f^{(1,-)}_{w,0}}\Lambda^1\frac{f^1_{L,0}}{b^1_L}+\frac{b^2_w}{\left|f^{(2,-)}_{w,0}\right|}\Lambda^2\frac{f^2_{L,0}}{b^2_L}\right)\frac{T}{T^\Sigma_2} \\
&\qquad\times\left(\frac{T}{T^\Sigma_2}\frac{\mathrm{e}^{-f^1_{w,0}\lambda_K T_1}}{K\mu^L_K}\right)^{-1}\sum_{n=0}^{\left\lfloor\frac{T}{T^\Sigma_2}\frac{\mathrm{e}^{-f^1_{w,0}T_1\lambda_K}}{K\mu^L_K}\right\rfloor}W^{(K,-)}_n+o(1) \\
&\overset{K\to\infty}{\longrightarrow}(1-\tilde{C}\varepsilon)TR^{\mathrm{pitstop}}_L.
\end{aligned}\tag{C.4.134}$$

Here, besides using the fact that $f^{(i,\pm)}_{w,0}\to f^i_{w,0}$, for $K\to\infty$, we applied the law of large numbers to the sum of iid. random variables $W^{(K,\pm)}_n$, which have expected value 1. Implementing the corresponding upper bounds results in a limit of $(1+\tilde{C}\varepsilon)TR^{\mathrm{pitstop}}_L$ accordingly.

Choosing ε arbitrarily small and remembering that growth of the L-mutant population and invasion of the resident population occur on a shorter time scale, as analogously to the proof of Theorem C.2.2, yields the claim of Theorem C.2.3.

C.5. Appendix

In this chapter, we collect and prove a number of technical results about branching processes that are related to the resident trait's stability, excursions of subcritical processes, and the short-term growth dynamics of mutants in a changing environment.

C.5.1. Resident stability

The following results build on and extend the results of [71]. They apply to what we refer to as *birth death processes with self-competition*, i.e. birth death processes X with individual birth rate b and a density-dependent individual death rate $d+cX$. In the results, the competitive term cX is rescaled by the carrying capacity K, as it is for the processes introduced in Section C.2.1. We start by citing a bound for the probability of deviating from the equilibrium population size on an arbitrary time scale θ_K, based on a potential theoretic argument.

Theorem C.5.1 ([71, Lemma A.1]). *Let X^K be a birth death process with self-competition and parameters $0 < d < b$ and $0 < c/K$. Define $\bar{n} := (b-d)/c$. Then there are constants $0 < C_1, C_2, C_3 < \infty$ such that, for any ε small and any K large enough, any initial value $0 \leq |x - \lceil \bar{n}K \rceil| \leq \frac{1}{2}\left\lfloor \frac{\varepsilon K}{2} \right\rfloor$, any $m \geq 0$, and any non-negative sequence $(\theta_K)_{K\in\mathbb{N}}$,*

$$\mathbb{P}_x\left(\exists\, t \in [0,\theta_K] : |X^K(t) - \lceil \bar{n}K \rceil| > \varepsilon K\right) \leq mC_1 e^{-C_2\varepsilon^2 K} + \sum_{l=m}^{\infty}\left(4\left(1 - e^{-C_3 K\theta_K/l}\right)^{1/2}\right)^l. \tag{C.5.1}$$

We can now apply this result to general time scales of the form K^p, $p > 0$, which in particular covers the time scale of interest $1/K\mu_K$, on which the crossing of the fitness valley occurs.

Corollary C.5.2. *Let X^K be the processes from Theorem C.5.1. Then, for all $p, q > 0$,*

$$\mathbb{P}_x\left(\exists\, t \in [0,K^p] : |X^K(t) - \lceil \bar{n}K \rceil| > \varepsilon K\right) = O(1/K^q). \tag{C.5.2}$$

In particular, for all $L > \alpha$,

$$\lim_{K\to\infty} \frac{1}{\lambda_K K \mu_K^L} P_x\left(\exists\, t \in \left[0, 1/K\mu_K^L\right] : |X^K(t) - \lceil \bar{n}K \rceil| > \varepsilon K\right) = 0. \tag{C.5.3}$$

Proof. We use the estimate of Theorem C.5.1 with $\theta_K = K^p$. Choosing $m = m_K = K^{p+q+1}$ we see that the first term $m_K C_1 e^{-C_2\varepsilon^2 K}$ is still exponentially decaying in K. Moreover, note that, for $l \geq m_K$,

$$\left(4\left(1 - e^{-C_3 K\theta_K/l}\right)^{1/2}\right)^l \leq \left(4\left(1 - e^{-C_3 K^{p+1}/m_K}\right)^{1/2}\right)^l \leq \left(4\left(1 - e^{-C_3/K^q}\right)^{1/2}\right)^l. \tag{C.5.4}$$

This allows us to estimate the sum by a geometric series

$$\begin{aligned}
&\sum_{l=m_K}^{\infty}\left(4\left(1 - e^{-C_3 K\theta_K/l}\right)^{1/2}\right)^l \leq \sum_{l=m_K}^{\infty}\left(4\left(1 - e^{-C_3/K^q}\right)^{1/2}\right)^l \\
\leq &\frac{\left(4\left(1 - e^{-C_3/K^q}\right)^{1/2}\right)^{m_K}}{1 - 4\left(1 - e^{-C_3/K^q}\right)^{1/2}} \leq C_4\left(16\left(1 - e^{-C_3/K^q}\right)\right)^{m_K/2} \\
\leq &C_4\left(16 C_3 K^{-q}\right)^{m_K/2} \leq 16 C_3 C_4 K^{-q}.
\end{aligned} \tag{C.5.5}$$

Here we used that, for K large enough, $1 - 4\left(1 - e^{-C_3/K^q}\right)^{1/2} \geq C_4^{-1}$ to get rid of the fraction. Moreover, we made use of the standard estimate $1 - \mathrm{e}^{-x} > x$. This yields the first claim. To conclude the second claim, we simply take $p = q = (L/\alpha) - 1$ and use that $\lambda_K \gg 1$. □

To estimate the process during the short adaptation phase after a parameter change, we derive a comparison result to the corresponding deterministic differential equation. We begin by providing two technical lemmas on properties of the Poisson distribution and Poisson processes, respectively.

Lemma C.5.3. *Let Y be a Poisson distributed random variable with parameter $\lambda > 0$ and denote its central moments by*

$$\mu_p := \mathbb{E}\left[(Y-\lambda)^p\right], \quad p \in \mathbb{N}_0. \tag{C.5.6}$$

Then we have, for $n \in \mathbb{N}_0$, the following leading order result in λ,

$$\mu_{2n} = a_n \lambda^n + O\left(\lambda^{n-1}\right), \qquad \mu_{2n+1} = b_n \lambda^n + O\left(\lambda^{n-1}\right), \tag{C.5.7}$$

where the prefactors are given by

$$a_n = \prod_{k=0}^{n-1}(2k+1) = (2n-1)!!, \qquad b_n = \sum_{k=0}^{n-1} \frac{k+1}{2k+1}\left(\prod_{i=0}^{k-1}(2i+1)\right). \tag{C.5.8}$$

Proof. It is easy to verify that all moments exist. By differentiating with respect to $\lambda > 0$, we obtain, for $p \geq 1$, the recursion

$$\mu_{p+1} = \lambda\left(\frac{\mathrm{d}\mu_p}{\mathrm{d}\lambda} + p\mu_{p-1}\right). \tag{C.5.9}$$

From this, we get the induction step

$$\mu_{2n+2} = (2n+1)a_n\lambda^{n+1} + O\left(\lambda^n\right), \qquad \mu_{2n+3} = \left[(n+1)a_n + (2n+1)b_n\right]\lambda^{n+1} + O\left(\lambda^n\right), \tag{C.5.10}$$

which, together with the base cases $\mu_0 = 1$ and $\mu_1 = 0$, directly implies the claim. □

Lemma C.5.4. *Let $Y \sim \mathrm{PPP}\left([0,\infty), \mathrm{d}u\right)$ be a homogeneous Poisson point process on $[0,\infty)$ and denote by $\tilde{Y}$ its compensated version, i.e. $\tilde{Y}(u) = Y(u) - u$. Then, for all $n \in \mathbb{N}$, all $1 \leq T < \infty$ and all $\xi \in (0,\infty)$,*

$$\mathbb{P}\left(\sup_{u\in[0,T]}\left|\tilde{Y}(u)\right| > \xi\right) \leq C_n \xi^{-2n} T^n, \tag{C.5.11}$$

where $C_n \in (0,\infty)$ only depends on n.

Proof. Since $\tilde{Y}$ is a martingale, $\left|\tilde{Y}\right|^{2n}$ is a submartingale. Therefore, we can apply Doob's maximum inequality

$$\mathbb{P}\left(\sup_{u\in[0,T]}\left|\tilde{Y}(u)\right| > \xi\right) = \mathbb{P}\left(\sup_{u\in[0,T]}\left|\tilde{Y}(u)\right|^{2n} > \xi^{2n}\right) \leq \xi^{-2n}\mathbb{E}\left[\left|\tilde{Y}(T)\right|^{2n}\right] \leq C_n \xi^{-2n} T^n. \tag{C.5.12}$$

Here we used in the last step that $\tilde{Y}(T)$ is a centered Poisson random variable with parameter $\lambda = T$ and we know from Lemma C.5.3 that its $(2n)$-th moment is a polynomial of degree n in T. □

This bound now allows us to extend a previous result from [71] on the convergence of the stochastic process to the solution of the corresponding differential equation, which is itself a quantification of the classical convergence result in [75].

Theorem C.5.5. *Let* X^K *be a birth death process with self-competition and parameters* $0 < d < b$ *and* $0 < c/K$. *Assume that* $X^K(0)/K \to x_0$ *as* $K \to \infty$ *and let* $(x(t))_{t \geq 0}$ *be the solution to the ordinary differential equation*

$$\dot{x}(t) = x(t)\left[b - d - c \cdot x(t)\right] \tag{C.5.13}$$

with initial value $x(0) = x_0$. *Then, for all* $n \in \mathbb{N}$, *there exists* $\tilde{C}_n \in (0, \infty)$ *such that, for every* $0 \leq T < \infty$ *and* $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{t \in [0,T]} \left| \frac{X^K(t)}{K} - x(t) \right| > \varepsilon \right) \leq \tilde{C}_n T^n \varepsilon^{-2n} K^{-n}. \tag{C.5.14}$$

Proof. The proof follows along the lines of Theorem A.3 in [71], with the only difference of using the higher moment estimates of Lemma C.5.4 in the final step. ☐

C.5.2. Subcritical excursions

The following result describes the distribution of the number birth events in a subcritical birth death process before extinction. While the result is already derived in [29], we want to mention a simplification of the expected value.

Lemma C.5.6 (extension of [29, Lemma A.3])**.** *Consider a subcritical birth death process with individual birth and death rates* $0 < b < d$. *Denote by* B *the total number of birth events during an excursion of this process initiated with exactly one individual. Then, for* $k \in \mathbb{N}_0$,

$$\mathbb{P}\left(B = k\right) = \frac{(2k)!}{k!(k+1)!} \left(\frac{b}{b+d}\right)^k \left(\frac{d}{b+d}\right)^{k+1} \tag{C.5.15}$$

and in particular

$$e^{(b,d)} := \mathbb{E}\left[B\right] = \frac{b}{d-b}. \tag{C.5.16}$$

Moreover, we have the following continuity result: There exist two positive constants $c, \varepsilon_0 > 0$, *such that, for all* $0 < \varepsilon < \varepsilon_0$ *and* $0 < b_i < d_i$, *if* $|b_1 - b_2| < \varepsilon$ *and* $|d_1 - d_2| < \varepsilon$, *then*

$$\left| e^{(b_1,d_1)} - e^{(b_2,d_2)} \right| < c\varepsilon. \tag{C.5.17}$$

Proof. The claim of (C.5.15) and the continuity result can be obtained by studying a discrete-time simple random walk on $\mathbb{Z}$ with probability of jumping up equal to $\rho = b/(b+d)$, which is the probability that the next event in the birth death process is a birth. Details can be found for example in [69, Lemma 17]. This also implies that

$$\mathbb{E}\left[B\right] = \sum_{k=1}^{\infty} \frac{(2k)!}{(k-1)!(k+1)!} \rho^k (1-\rho)^{k+1}. \tag{C.5.18}$$

This expression can be shown to be equal to $\rho/(1-2\rho)$, for $\rho < 1/2$, e.g. by rewriting the binomial coefficients using the residue theorem. Plugging back in the value of ρ then yields

$$\mathbb{E}\left[B\right] = \frac{b/(b+d)}{1-2b/(b+d)} = \frac{b}{d-b}. \tag{C.5.19}$$

□

C.5.3. Short-term growth

Finally, we present a result on the short-term growth dynamics for birth death processes with time-dependent rates on the $\ln K$-time scale. As introduced in C.2.1, the rates of the birth death processes vary on the time scale $1 \ll \lambda_K \ll \ln K$ with $\ell \in \mathbb{N}$ different parameter phases, where ℓ is possibly different from the one in the main results. Denoting by $T_i > 0$ the single and by $T_i^\Sigma := \sum_{j=1}^{i} T_j$ the accumulated lengths of parameter phases, and by b^i and d^i the corresponding birth and death rates, the time-dependent rate function are given by the periodic extensions of

$$b(t) := \sum_{i=1}^{\ell} \mathbb{1}_{t \in [T_{i-1}^\Sigma, T_i^\Sigma)} b^i, \qquad d(t) := \sum_{i=1}^{\ell} \mathbb{1}_{t \in [T_{i-1}^\Sigma, T_i^\Sigma)} d^i. \tag{C.5.20}$$

We set $r^i := b^i - d^i$ and $r(t) := b(t) - d(t)$ to refer to the net growth rate and $r^{\text{av}} := (\sum_{i=1}^{\ell} r^i T_i)/T_\ell^\Sigma$ to refer to the average net growth rate. Moreover, on the time scale λ_K we consider $b^K(t) := b(t/\lambda_K)$, $d^K(t) := d(t/\lambda_K)$, and $r^K(t) := r(t/\lambda_K)$.

To prove the desired result, we first derive an equivalent formulation of the set of possible arrival times of successful mutants.

Lemma C.5.7. *For a piecewise constant, right-continuous, periodic function r such as the one above, let*

$$g(t) := \int_0^t r(u) \mathrm{d}u. \tag{C.5.21}$$

The the following definitions of the set $A \subset [0,\infty)$ of possible arrival times of successful mutants are equivalent:

$$A_1 = \left\{t \geq 0 : \forall s \in (0, T_\ell^\Sigma]\ g(t+s) > g(t)\right\}, \tag{C.5.22}$$

$$A_2 = \left\{t \geq 0 : \forall s \in (0, \infty)\ g(t+s) > g(t)\right\}, \tag{C.5.23}$$

$$A_3 = \left\{t \geq 0 : \exists\ \gamma > 0\ \forall s \in (0, \infty)\ g(t+s) > g(t) + \gamma s\ \right\}. \tag{C.5.24}$$

Proof. The inclusions $A_3 \subseteq A_2 \subseteq A_1$ are somewhat trivial and we hence focus on $A_1 \subseteq A_3$. To this end, take $t \in A_1$ and note that it suffices to show that

$$\inf_{s>0} \frac{g(t+s)-g(t)}{s} > 0. \tag{C.5.25}$$

Since g is continuous and the defining inequality of A_1 is strict, it still holds true for $\tilde{t} = t+\varepsilon$, with $\varepsilon > 0$ sufficiently small. Hence, for $s \in [\varepsilon, T_\ell^\Sigma]$, we get

$$g(t+s) - g(t) = g(t+s) - g(t+\varepsilon) + g(t+\varepsilon) - g(t) > g(t+\varepsilon) - g(t). \tag{C.5.26}$$

Moreover, for such s, we can make the rough estimate

$$\frac{g(t+s)-g(t)}{s} > \frac{g(t+\varepsilon)-g(t)}{T_\ell^\Sigma} =: \tilde{\gamma} > 0. \tag{C.5.27}$$

Now, since r is piecewise constant and right-continuous, we can take $\varepsilon > 0$ sufficiently small such that r is constant on $[t, t+\varepsilon]$, which implies that $g(t+s) = g(t) + r^* s$, for $s \in [0, \varepsilon]$, where $r^* \in \{r^i : i = 1, \dots, \ell\}$. The defining inequality of A_1 immediately implies that $r^* > 0$. Lastly, we note that every $s \geq T_\ell^\Sigma$ can be split uniquely into $s = nT_\ell^\Sigma + \tilde{s}$, with $n \in \mathbb{N}$ and $0 \leq \tilde{s} < T_\ell^\Sigma$. Thus

$$g(t+s) - g(t) > g(t+s) - g(t+\tilde{s}) = r^{\text{av}} n T_\ell^\Sigma \tag{C.5.28}$$

and hence

$$\frac{g(t+s)-g(t)}{s} > \frac{r^{\text{av}} n T_\ell^\Sigma}{(n+1)T_\ell^\Sigma} \geq \frac{r^{\text{av}}}{2} > 0. \tag{C.5.29}$$

The positivity of r^{av} is a direct consequence of $A_1 \neq \emptyset$. We can thus take $\gamma = \min\{\tilde{\gamma}, r^*, r^{\text{av}}/2\}$ to show that $t \in A_3$ and hence $A_1 \subseteq A_3$, which concludes the proof. $\square$

With this characterization at hand, we can now prove the following lemma, which is an extension of Lemma C.1 in [71].

Lemma C.5.8. *Let Z^K be birth death process with time-dependent rates b^K, d^K and let $g^K(t) = \int_0^t r^K(s)\mathrm{d}s$, where r^K is the net growth rate. Assume that $r^{\text{av}} > 0$ and the initial time lies in the set of possible arrival times of successful mutants defined in Lemma C.5.7 corresponding to the growth function $f = r$, i.e. $0 \in A$. Then, for all $\varepsilon > 0$, $0 < p_1 < 1 < p_2$, and all initial values satisfying $1 \ll Z^K(0) \ll K^\varepsilon$, we obtain*

$$\mathbb{P}\left(p_1 \mathrm{e}^{g^K(t)} Z^K(0) < Z^K(t) < p_2 \mathrm{e}^{g^K(t)} Z^K(0) \quad \forall t \in [0, \varepsilon \ln K]\right) = 1 - O((Z^K(0))^{-1}) \xrightarrow{K\to\infty} 1. \tag{C.5.30}$$

Proof. Checking the counter probabilities, we observe that

$$\begin{aligned}
&\mathbb{P}\left(\exists\, t \leq \varepsilon \ln K :\ Z^K(t) \leq p_1 \mathrm{e}^{g^K(t)} Z^K(0)\right)\\
&= \mathbb{P}\left(\exists\, t \leq \varepsilon \ln K :\ Z^K(0) - \mathrm{e}^{-g^K(t)} Z^K(t) \geq (1-p_1) Z^K(0)\right)\\
&\leq \mathbb{P}\left(\sup_{t \leq \varepsilon \ln K} \left|\mathrm{e}^{-g^K(t)} Z^K(t) - Z^K(0)\right| \geq q Z^K(0)\right), \qquad \text{(C.5.31)}\\
&\mathbb{P}\left(\exists\, t \leq \varepsilon \ln K :\ Z^K(t) \geq p_2 \mathrm{e}^{g^K(t)} Z^K(0)\right)\\
&= \mathbb{P}\left(\exists\, t \leq \varepsilon \ln K :\ \mathrm{e}^{-g^K(t)} Z^K(t) - Z^K(0) \geq (p_2-1) Z^K(0)\right)\\
&\leq \mathbb{P}\left(\sup_{t \leq \varepsilon \ln K} \left|\mathrm{e}^{-g^K(t)} Z^K(t) - Z^K(0)\right| \geq q Z^K(0)\right), \qquad \text{(C.5.32)}
\end{aligned}$$

for some $q > 0$. For both bounds we apply Doob's maximum inequality to the rescaled martingale $\hat{M}^K(t) = \mathrm{e}^{-g^K(t)} Z^K(t) - Z^K(0)$ to obtain (see [71, Lemma C.1] for details)

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t \leq \varepsilon \ln K} \left|\mathrm{e}^{-g^K(t)} Z^K(t) - Z^K(0)\right| \geq q Z^K(0)\right) = \mathbb{P}\left(\sup_{t \leq \varepsilon \ln K} \left|\hat{M}^K(t)\right| \geq q Z^K(0)\right)\\
&\leq \frac{C}{Z^K(0)} \int_0^{\varepsilon \ln K} \mathrm{e}^{-g^K(s)} \mathrm{d}s \leq \frac{C}{Z^K(0)} \int_0^{\varepsilon \ln K} \mathrm{e}^{-\gamma s} \mathrm{d}s = \frac{C}{Z^K(0)} \frac{1 - K^{-\gamma\varepsilon}}{\gamma} \leq \frac{\tilde{C}}{Z^K(0)}. \qquad \text{(C.5.33)}
\end{aligned}$$

Here we used that, by Lemma C.5.7, $g^K(s) = \lambda_K g(s/\lambda_K) \geq \gamma s$, for some $\gamma > 0$ and all $s \geq 0$, since $0 \in A$. $\square$

Bibliography

[1] A stochastic population model for the impact of cancer cell dormancy on therapy success. *J. Theor. Biol.*

[2] F. Alberti. Asymptotic sampling distributions made easy: loose linkage in the ancestral recombination graph. *Electron. J. Probab.*, 29:1–30, 2024.

[3] K. B. Athreya. Some results on multitype continuous time Markov branching processes. *Ann. Math. Stat.*, 39:347–357, 1968.

[4] K. B. Athreya and P. E. Ney. *Branching processes.* Die Grundlehren der mathematischen Wissenschaften, Vol. 196. Springer-Verlag, New York-Heidelberg, 1972.

[5] F. J. Ayala. Evolution. *Encyclopædia Britannica*, 2019. Available at `https://www.britannica.com/science/evolution-scientific-theory`, accessed: 03/01/2025.

[6] M. Baar and A. Bovier. The polymorphic evolution sequence for populations with phenotypic plasticity. *Electron. J. Probab.*, 23(72):1–27, 2018.

[7] 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.

[8] N. H. Barton, A. M. Etheridge, and A. Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab.*, 15:no. 7, 162–216, 2010.

[9] M. Benaim, C. Lobry, T. Sari, and E. Strickler. When can a population spreading across sink habitats persist? *Journal of Mathematical Biology*, 88(19), 2024.

[10] J. Berestycki, E. Brunet, and Z. Shi. The number of accessible paths in the hypercube. *Bernoulli*, 22(2):653–680, 2016.

[11] J. Berestycki, E. Brunet, and Z. Shi. Accessibility percolation with backsteps. *ALEA Lat. Am. J. Probab. Math. Stat.*, 14(1):45–62, 2017.

[12] N. Berestycki. Recent progress in coalescent theory. *Preprint*, arXiv:0909.3985, 2009.

[13] S. Billiard, P. Collet, R. Ferriére, S. Méléard, and V. C. Tran. The effect of competition and horizontal trait inheritance on invasion, fixation, and polymorphism. *J. Theor. Biol.*

[14] 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.

[15] A. Blancas and S. Palau. Coalescent point process of branching trees in a varying environment. *Electronic Communications in Probability*, 29(none):1 – 15, 2024.

[16] J. Blath. Dimensions of dormancy. 2024. Available at `https://www.uni-frankfurt.de/155323294/dormancy_map_jb_03_07_2024.pdf`accessed: 03/23/2025.

[17] J. Blath, A. G. Casanova, N. Kurt, and M. Wilke-Berenguer. A new coalescent for seed-bank models. *Ann. Appl. Probab.*, 26(2):857–891, 2016.

[18] J. Blath, A. Etheridge, and M. Meredith. Coexistence in locally regulated competing populations and survival of branching annihilating random walk. *Ann. Appl. Probab.*, 17(5-6):1474 – 1507, 2007.

[19] J. Blath, F. Hermann, and M. Slowik. A branching process model for dormancy and seed banks in randomly fluctuating environments. *J. Math. Biol.*, 83(2):17, 2021.

[20] J. Blath, T. Paul, A. Tóbiás, and M. W. Berenguer. The impact of dormancy on evolutionary branching, 2022.

[21] J. Blath, T. Paul, A. Tóbiás, and M. Wilke Berenguer. The impact of dormancy on evolutionary branching. *Theor. Popul. Biol.*, 156:66–76, 2024.

[22] J. Blath and A. Tóbiás. Invasion and fixation of microbial dormancy traits under competitive pressure. *Stoch. Process. Appl.*, 130(12):7363–7395, 2020.

[23] F. Boenkost, F. Foutel-Rodier, and E. Schertzer. The genealogy of nearly critical branching processes in varying environment, 2024.

[24] 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.

[25] 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.

[26] A. Bovier. Stochastic individual based models: From adaptive dynamics to modelling of cancer therapies. Lecture Notes, 2019.

[27] A. Bovier. Stochastic models for adaptive dynamics: Scaling limits and diversity. In *Probabilistic Structures in Evolution,* E. Baake and A. Wakolbinger, Eds., volume 17 of *EMS Series of Congress Reports*, pages 127–149. EMS Press, Berlin, 2021.

[28] 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.

[29] 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.

[30] A. Bovier and F. den Hollander. *Metastability, A Potential-Theoretic Approach.* Die Grundlehren der Mathematischen Wissenschaften, Vol. 351. Springer Cham Heidelberg New York Dortrecht London, 2015.

[31] A. Bovier and L. Hartung. The speed of invasion in an advancing population. *J. Math. Biol.*, 87(4):Paper No. 56, 2023.

[32] A. Bovier and S.-D. Wang. Trait substitution trees on two time scales analysis. *Markov Process. Relat. Fields*, 19(4):607–642, 2013.

[33] I. Bozic and M. A. Nowak. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. *PNAS*, 111(45):15964–15968, 2014.

[34] V. Brouard. Genetic composition of supercritical branching populations under power law mutation rates, 2024.

[35] A. Bukkuri, K. J. Pienta, R. H. Austin, E. U. Hammarlund, S. R. Amend, and J. S. Brown. A mathematical investigation of polyaneuploid cancer cell memory and cross-resistance in state-structured cancer populations. *Scientific reports*, 13(1):15027–11, 2023.

[36] M. Bulmer. *The mathematical theory of quantitative genetics*. Clarendon Press, Oxford, 1980.

[37] T. Burkart and E. Frey. Periodic temporal environmental variations induce coexistence in resource competition models. *arXiv:2202.11635*, 2022.

[38] 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.

[39] 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.

[40] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Process. Appl.*, 116(8):1127–1160, 2006.

[41] N. Champagnat. *Approches stochastiques et déterministes en biologie: dynamique adaptative, modélisation pour l'écologie, génétique des populations et dynamique moléculaire; caractere bien posé d'équations différentielles ordinaires et stochastiques*. Habilitation, Université de Lorraine, 2014.

[42] N. Champagnat and V. Hass. Convergence of population processes with small and frequent mutations to the canonical equation of adaptive dynamics. *Ann. Appl. Probab.*, 35(1):1–63, 2025.

[43] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probab. Theory Relat. Fields*, 151(1-2):45–94, 2011.

[44] N. Champagnat, S. Méléard, S. Mirrahimi, and V. C. Tran. Filling the gap between individual-based evolutionary models and hamilton-jacobi equations. *Journal de l'École polytechnique. Mathématiques*, 10:1247–1275, 2023.

[45] N. Champagnat, S. Méléard, and V. C. Tran. Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer. *Ann. Appl. Probab.*, 31(4):1820–1867, 2021.

[46] E. N. M. Cirillo and F. R. Nardi. Relaxation height in energy landscapes: An application to multiple metastable states. *J.Stat.Phys.*, 150:1080–1114, 2013.

[47] J. E. Cohen. Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. *PLoS biology*, 2(12):e439–e439, 2004.

[48] L. Coquille, A. Kraut, and C. Smadi. Stochastic individual-based models with power law mutation rate on a general finite trait space. *Electron. J. Probab.*, 26:1–37, 2021.

[49] F. Cordero, A. G. Casanova, and J. Schweinsberg. Two waves of adaptation: speciation induced by dormancy in a model with changing environment. *Preprint*, arXiv:2410.10890, 2024.

[50] 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.

[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] D. A. Dawson and A. Greven. Multiscale analysis: Fisher–wright diffusions with rare mutations and selection, logistic branching system. In J.-D. Deuschel, B. Gentz, W. König, M. von Renesse, M. Scheutzow, and U. Schmock, editors, *Probability in Complex Physical Systems*, pages 373–408, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[56] D. A. Dawson and A. Greven. *Spatial Fleming-Viot models with selection and mutation.* Lecture Notes in Mathematics, Vol. 2092. Springer International Publishing, 2014.

[57] J.-B. P. A. d. M. de Lamarck. *Philosophie zoologique.* Dentu, Muséum d'Histoire Naturelle, 1809.

[58] J. A. G. de Visser and J. Krug. Empirical fitness landscapes and the predictability of evolution. *Nature reviews. Genetics*, 15(7):480–490, 2014.

[59] J. A. G. De Visser and J. Krug. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.*, 15(7):480–490, 2014.

[60] C. Dellacherie and P. A. Meyer. *Probabilities and Potential, B: Theory of Martingales.* Elsevier Science & Technology, Amsterdam, 1982.

[61] 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.

[62] 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.

[63] O. Diekmann, P. E. Jabin, S. Mischler, and B. Perthame. The dynamics of adaptation: An illuminating example and a hamilton-jacobi approach. *Theo. Pop. Biol.*, 67(4):257–271, jun 2005.

[64] C. Dombry, C. Mazza, and V. Bansaye. Phenotypic diversity and population growth in a fluctuating environment. *Adv. Appl. Probab.*, 43(2):375–398, 2011.

[65] S. Donnet, V. Rivoirard, and J. Rousseau. Nonparametric baysian estimation for multivariate Hawkes processes. *Ann. Stat.*, 48(5):2698–2727, 2020.

[66] 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.

[67] R. Durrett. *Probability models for DNA sequence evolution.* Probability and its applications. Springer, 2nd ed. edition, 2013.

[68] R. Durrett and S. Moseley. Evolution of resistance and progression to disease during clonal expansion of cancer. *Theoretical Population Biology*, 77(1):42–48, 2010.

[69] M. Esser and A. Kraut. A general multi-scale description of metastable adaptive motion across fitness valleys. *J. Math. Biol.*, 89(4):Article No. 46, 2024.

[70] M. Esser and A. Kraut. Crossing a fitness valley in a changing environment: With and without pit stop. *Preprint*, arXiv:2503.19766, 2025.

[71] M. Esser and A. Kraut. Effective growth rates in a periodically changing environment: From mutation to invasion. *Stoch. Proc. Appl.*, 184:Article No. 104598, 2025.

[72] A. M. Etheridge. Survival and extinction in a locally regulated population. *The Annals of Applied Probability*, 14(1):188–214, 2004.

[73] A. M. Etheridge. Drift, draft and structure: some mathematical models of evolution. In *Stochastic models in biological sciences*, volume 80 of *Banach Center Publ.*, pages 121–144. Polish Acad. Sci. Inst. Math., Warsaw, 2008.

[74] A. M. Etheridge. Some mathematical models from population genetics, volume 2012 of lecture notes in mathematics, 2011.

[75] S. N. Ethier and T. G. Kurtz. *Markov processes.* Wiley Ser. in Probab. and Math. Stat. John Wiley & Sons, Inc., New York, 1986.

[76] 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.

[77] W. J. Ewens. *Mathematical Population Genetics: I. Theoretical Introduction.* Interdisciplinary Applied Mathematics, Vol. 27. Springer, New York, 2004.

[78] D. Ewing, C. Cobbold, B. Purse, M. Nunn, and S. White. Modelling the effect of temperature on the seasonal population dynamics of temperate mosquitoes. *Jour. Theo. Biol.*, 400:65–79, 2016.

[79] D. S. Falconer and T. F. C. Mackay. *Introduction To Quantitative Genetics.* Longman Group UK Limited, 1980.

[80] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *T. Roy. Soc. Edin.*, 52(2):399–433, 1918.

[81] R. A. Fisher. The wave of advance of advantageous genes. *Ann. Eugenic.*, 7(4):355–369, 1937.

[82] W. H. Fleming and M. Viot. Some measure-valued Markov processes in population genetics theory. *Indiana U. Math. J.*, 28(5):817–843, 1979.

[83] 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.

[84] F. Francescangeli, M. L. De Angelis, R. Rossi, A. Cuccu, A. Giuliani, R. De Maria, and A. Zeuner. Dormancy, stemness, and therapy resistance: interconnected players in cancer evolution. *Cancer and metastasis reviews*, 42(1):197–215, 2023.

[85] R. E. Franklin and R. G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.

[86] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems.* Die Grundlehren der Mathematischen Wissenschaften, Vol.260. Springer, Heidelberg, third edition, 2012.

[87] J. Garnier, O. Cotto, E. Bouin, T. Bourgeron, T. Lepoutre, O. Ronce, and V. Calvez. Adaptation of a quantitative trait to a changing environment: New analytical insights on the asexual and infinitesimal sexual models. *Theor. Popul. Biol.*, 152:1–22, 2023.

[88] S. A. Geritz, É. Kisdi, G. Meszéna, and J. o. A. Metz. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.*

[89] S. A. Geritz, J. A. Metz, É. Kisdi, and G. Meszéna. Dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.*, 78(10):2024, 1997.

[90] 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.

[91] J. H. Gillespie. Molecular evolution over the mutational landscape. *Evolution*, 38(5):1116–1129, 1984.

[92] C. S. Gokhale, Y. Iwasa, M. A. Nowak, and A. Traulsen. The pace of evolution across fitness valleys. *Journal of Theoretical Biology*, 259(3):613–620, 2009.

[93] A. Greven, F. den Hollander, and M. Oomen. Spatial populations with seed-bank: well-posedness, duality and equilibrium. *Electron. J. Probab.*, 27:1–88, 2022.

[94] A. Greven, A. Sturm, A. Winter, and I. Zähle. Multi-type spatial branching models for local self-regulation i: Construction and an exponential duality. *Preprint*, arXiv:1509.04023, 2015.

[95] G. Grimmett and D. Stirzaker. *Probability and random processes.* Oxford University Press, 2020.

[96] J. O. Haerter, C. Lövkvist, I. B. Dodd, and K. Sneppen. Collaboration between cpg sites is needed for stable somatic inheritance of dna methylation states. *Nucleic Acids Res.*, 42(4):2235–2244, 11 2013.

[97] 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.

[98] G. H. Hardy. Mendelian proportions in a mixed population. *Science*, 28:49–50, 1908.

[99] T. E. Harris. *The theory of branching processes.* Die Grundlehren der mathematischen Wissenschaften, Vol. 119. Springer, Berlin, 1963.

[100] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[101] J. Hofbauer and K. Sigmund. Adaptive dynamics and evolutionary stability. *Appl. Math. Lett.*, 3(4):75–79, 1990.

[102] J. Hofbauer and K. Sigmund. *Evolutionary games and population dynamics.* Camb. Univ. Press, 1998.

[103] P. Hunter. Biology is the new physics. *EMBO reports*, 11(5):350–352, 2010.

[104] M. Hutzenthaler and A. Wakolbinger. Ergodic behavior of locally regulated branching populations. *Ann. Appl. Probab.*, 17(2):474–501, 2007.

[105] J. Istas. *Mathematical Modeling for the Life Sciences.* Universitext. Springer Berlin, Heidelberg, 2005.

[106] 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.

[107] K. Jain and A. Devi. Polygenic adaptation in changing environments. *Europhys. Lett.*, 123(4):48002, 2018.

[108] K. Jain and J. Krug. Evolutionary trajectories in rugged fitness landscapes. *J Stat Mech: Theory and Exp.*, 2005(04):P04008, apr 2005.

[109] K. Jain and J. Krug. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. *Genetics*, 175:1275–88, 03 2007.

[110] S. Jones, W.-d. Chen, G. Parmigiani, F. Diehl, N. Beerenwinkel, T. Antal, A. Traulsen, M. A. Nowak, C. Siegel, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, J. Willis, and S. D. Markowitz. Comparative lesion sequencing provides insights into tumor evolution. *PNAS*, 105(11):4283–4288, 2008.

[111] J. Jost and Y. Wang. Optimization and phenotype allocation. *Bull. Math. Biol.*, 76:184–200, 2014.

[112] P. J. Keeling and J. D. Palmer. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Drug Discov.*

[113] G. Kersting and V. Vatutin. *Discrete Time Branching Processes in Random Environment.* Wiley, Newark, 1 edition, 2017.

[114] M. Kimura. Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci. USA*, 41(3):144–150, 1955.

[115] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, (Special Vol. 19A):27–43, 1982.

[116] E. Kisdi. Adaptive dynamics papers. `https://www.mv.helsinki.fi/home/kisdi/addyn.htm`. Accessed: 03/19/2025.

[117] P. Koepernik. The brownian spatial coalescent, 2024.

[118] 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.

[119] N. L. Komarova. Loss- and gain-of-function mutations in cancer: Mass-action, spatial and hierarchical models. *J. Stat. Phys.*, 128:413–446, 2007.

[120] A. Kraut and A. Bovier. From adaptive dynamics to adaptive walks. *J. Math. Biol.*, 79(5):1699–1747, 2019.

[121] A. K. Kraut. *Stochastic and deterministic models for the evolution of heterogeneous populations: Multiscale approximation and applications to melanoma T-cell therapy.* PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2020.

[122] F. P. Kreten. *Self-organized branching morphogenesis: study of a related FKPP-system and a case study of prostate cancer.* PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2023.

[123] J. Krug. Accessibility percolation in random fitness landscapes. In *Probabilistic Structures in Evolution,* E. Baake and A. Wakolbinger, Eds., volume 17 of *EMS Series of Congress Reports*, pages 1–22. EMS Press, Berlin, 2021.

[124] T. G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *J. Appl. Probab.*, 7(1):49–58, 1970.

[125] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078, 2005.

[126] L. A. La Rocca, J. Frank, H. B. Bentzen, J. T. Pantel, K. Gerischer, A. Bovier, and P. M. Krawitz. Understanding recessive disease risk in multi-ethnic populations with different degrees of consanguinity. *Am. J. Med. Genet. Part A*, 194(3):e63452, 2024.

[127] A. Lambert. The branching process with logistic growth. *Ann. Appl. Probab.*, 15(2):1506–1535, 2005.

[128] R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami. The evolutionary origin of complex features. *Nature*, 423(6936):139–144, 2003.

[129] E. Litchman and C. A. Klausmeier. Competition of phytoplankton under fluctuating light. *The American Naturalist*, 157(2):170–187, 2001.

[130] A. J. Lotka. Quantitative studies in epidemiology. *Nature*, 88:497–498, 1912.

[131] A. Louvet, C. Mantoux, and N. Machon. Assessing the extinction risk of the spontaneous flora in urban tree bases. *PLOS Comput. Biol.*, 20(6):1–20, 06 2024.

[132] P. Maillard, G. Raoul, and J. Tourniaire. Spreading speed of locally regulated population models in macroscopically heterogeneous environments. *Preprint*, arXiv:2105.06985, 2021.

[133] S. Maisnier-Patin, O. G. Berg, L. Liljas, and D. I. Andersson. Compensatory adaptation to the deleterious effect of antibiotic resistance in salmonella typhimurium. *Molecular microbiology*, 46(2):355–366, 2002.

[134] T. Malik and H. L. Smith. Does dormancy increase fitness of bacterial populations in time-varying environments? *Bull. Math. Biol.*, 70(4):1140–1162, 2008.

[135] 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.

[136] 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.

[137] I. Martincorena, K. M. Raine, M. Gerstung, K. J. Dawson, K. Haase, P. Van Loo, H. Davies, M. R. Stratton, and P. J. Campbell. Universal patterns of selection in cancer and somatic tissues. *Cell*, 171(5):1029–1041, 2017.

[138] G. Mendel. Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 4:3–47, 1865.

[139] G. Mendel. Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde. *Verhandlungen des naturforschenden Vereines in Brünn*, 8:26–31, 1869.

[140] J. A. J. Metz. Adaptive dynamics. In *Encyclopedia of Theoretical Ecology*, pages 7–17. University of California Press, 2012.

[141] 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.

[142] 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.

[143] P. A. P. Moran. Random processes in genetics. *Math. Proc. Cambridge*, 54(1):60–71, 1958.

[144] S. Mukherjee. *The Gene: An Intimate History.* Bodley, London, 2016.

[145] R. Murray and G. Young. Neutral competition in a deterministically changing environment: revisiting continuum approaches. *J. Theor. Biol.*, 486:110104, 2020.

[146] J. Neidhart and J. Krug. Adaptive walks and extreme value theory. *Phys. Rev. Lett.*, 107:178102, Oct 2011.

[147] C. Neuhauser and S. M. Krone. Ancestral processes with selection. *Theor. Popul.. Biol.*, 51(3):210–237, 1997.

[148] R. Neukirch and A. Bovier. Survival of a recessive allele in a Mendelian diploid model. *J. Math. Biol.*, 75(1):145–198, 2017.

[149] M. Nicholson and T. Antal. Competing evolutionary paths in growing populations with applications to multidrug resistance. *PLoS Comput Biol*, 15(4), 2019.

[150] 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.

[151] H. A. Orr. A minimum on the mean number of steps taken in adaptive walks. *J. Theor. Biol.*, 220(2):241–247, 2003.

[152] M. N. Pallen, M. From the origin of species to the origin of bacterial flagella. *Nat. Rev. Microbiol.*, 4:784–790, 2006.

[153] T. Paul. The canonical equation of adaptive dynamics in individual-based models with power law mutation rates. *Preprint*, arXiv:2309.02148, 2024.

[154] L. Popovic and A. Véber. A spatial measure-valued model for chemical reaction networks in heterogeneous systems. *Ann. Appl. Probab.*, 33(5):3706–3754, 2023.

[155] L. Roques, F. Patout, O. Bonnefon, and G. Martin. Adaptation in general temporally changing environments. *SIAM J. App. Math.*, 80(6):2420–2447, 2020.

[156] B. Schmiegelt and J. Krug. Evolutionary accessibility of modular fitness landscapes. *J. Stat. Phys.*, 154(1-2):334–355, 2014.

[157] J. G. Silva, A. C. Ribeiro, R. F. Camargo, P. F. Mancera, and F. L. Santos. Stability analysis and numerical simulations via fractional calculus for tumor dormancy models. *Commun. Nonlinear Sci. Nummer. Simulat.*, 72:528–543, 2019.

[158] C. Smadi. The effect of recurrent mutations on genetic diversity in a large population of varying size. *Acta Appl. Math.*, 149:11–51, 2017.

[159] S. Smale. On the differential equations of species in competition. *J. Math. Biol.*, 3(1):5–7, 1976.

[160] A. Taitelbaum, R. West, M. Assaf, and M. Mobilia. Population dynamics in a changing environment: random versus periodic switching. *Phys. Rev. Lett.*, 125(4):048105, 2020.

[161] 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.

[162] K. Utsey and J. P. Keener. A mathematical model for inheritance of dna methylation patterns in somatic cells. *Bull. Math. Biol.*, 82(7):84–84, 2020.

[163] V. A. Vatutin and C. Smadi. Critical branching processes in a random environment with immigration: The size of the only surviving family. *Proceedings of the Steklov Institute of Mathematics,*, 316(4):336–355, 2022.

[164] V. Volterra. Variations and fluctuations of the number of individuals in animal species living together. *ICES J. Mar. Sci.*, 3(1):3–51, 1928.

[165] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[166] D. Waxman and S. Gavrilets. 20 questions on adaptive dynamics. *J. Evol. Biol.*, 18(5):1139–1154, 2005.

[167] W. Weinberg. Über den Nachweis der Vererbung beim Menschen. *Jahresh. Ver. Vaterl. Natkd. Württ.*, 64:369–382, 1908.

[168] 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.

[169] S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97–157, 1931.

[170] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 6th Int. Congr. Genet.*, 1:356–366, 1932.

[171] D. Xiao, W. Li, and M. Han. Dynamics in a ratio-dependent predator–prey model with predator harvesting. *J. Math. Anal. Appl.*, 324(1):14–29, 2006.

[172] M. L. Zeeman. Hopf bifurcations in competitive three-dimensional Lotka-Volterra systems. *Dynam. Stab. Syst.*, 8(3):189–217, 1993.