LOCAL ADAPTATION IN METAPOPULATIONS

by

Enikő Szép

September 20, 2020

A thesis presented to the
Graduate School
of the
Institute of Science and Technology Austria, Klosterneuburg, Austria
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Institute of Science and Technology

The thesis of Enikő Szép, titled *Local adaptation in metapopulations*, is approved by:

Supervisor: Nicholas H. Barton, IST Austria, Klosterneuburg, Austria

Signature: _____

Committee Member: Beatriz Vicoso, IST Austria, Klosterneuburg, Austria

Signature: _____

Committee Member: Richard A. Nichols, Queen Mary University, London, UK

Signature: _____

Defense Chair: Eva Benková, IST Austria, Klosterneuburg, Austria

Signature: _____

signed page is on file

IST Austria Thesis, ISSN: 2663-337X

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Enikő Szép
September 20, 2020
signed page is on file

Abstract

This thesis concerns itself with the interactions of evolutionary and ecological forces and the consequences on genetic diversity and the ultimate survival of populations. It is important to understand what signals processes leave on the genome and what we can infer from such data, which is usually abundant but noisy. Furthermore, understanding how and when populations adapt or go extinct is important for practical purposes, such as the genetic management of populations, as well as for theoretical questions, since local adaptation can be the first step toward speciation.

In Chapter 2, we introduce the method of maximum entropy to approximate the demographic changes of a population in a simple setting, namely the logistic growth model with immigration. We show that this method is not only a powerful tool in physics but can be gainfully applied in an ecological framework. We investigate how well it approximates the real behavior of the system, and find that is does so, even in unexpected situations. Finally, we illustrate how it can model changing environments.

In Chapter 3, we analyze the co-evolution of allele frequencies and population sizes in an infinite island model. We give conditions under which polygenic adaptation to a rare habitat is possible. The model we use is based on the diffusion approximation, considers eco-evolutionary feedback mechanisms (hard selection), and treats both drift and environmental fluctuations explicitly. We also look at limiting scenarios, for which we derive analytical expressions.

In Chapter 4, we present a coalescent based simulation tool to obtain patterns of diversity in a spatially explicit subdivided population, in which the demographic history of each subpopulation can be specified. We compare the results to existing predictions, and explore the relative importance of time and space under a variety of spatial arrangements and demographic histories, such as expansion and extinction.

In the last chapter, we give a brief outlook to further research.

Acknowledgments

I consider myself fortunate to have been given the opportunity to pursue a PhD at IST Austria. The last five years were the best in my life... and the toughest. I worked on interesting projects, learned a lot, faced unexpected challenges— all of which would not have been possible without the people who supported me and helped me grow.

First and foremost I would like to thank my advisor, Nick Barton, who gave me the chance to join his group, originally as an intern, and later as a PhD student. I have learned a great deal from him and I admire his knowledge and never ending curiosity about science—and all the big questions of life. Despite his busy schedule, he was always available when I needed help. For everything, I am grateful beyond words.

I would also like to thank Beatriz Vicoso and Richard Nichols for serving on my committee and giving me valuable feedback on my work.

Thanks to my collaborators: Barbora, Himani, Kati, and Katka. Working together and seeing how you approach research questions was an inspiration. I would like to say a special thanks to Himani: without you this thesis would have taken forever! You had always time to discuss research and your clear explanations made me appreciate the details instead of drowning in them. Also, I found a great friend in you.

I made many friends at IST, who enriched my life here and I hope we will stay part of each others lives in the future: thanks to Dasha, Anja, Lenka, Gemma, Arinya, Shamsi, Harold, Hande, Javi, Mo, May for being part of this adventure.

Thanks to my amazing research group, past and current members, (and the ,,best office" of course)—for all the interesting conversations and experiences we had together.

Thanks to the lovely "Loser stays" group! Our morning coffees during the quarantine became a nice ritual, possibly bringing us even closer despite the physical distances—almost like playing Werewolf or Foosball...

I would like to thank all the administrative support I got, making my life easier: thanks to Astrid, Christine, Christine, and Elisabeth.

Thanks to my friends who believed in me and supported me, even as borders or the ocean separate us. Thanks to the girls from the Kőrösi, my high school friends, and thanks to my friends from the BUTE.

I would like to thank my teachers from elementary school, and high school, my university professors; they supported my ambitions and furthered my understanding of the world. A special thanks to Ibolya néni, who is not with us anymore, yet her influence accompanied me on the path leading here.

Of course, all this would not have been possible without the constant support of my family, my parents, which is by no means given, and I am grateful. A special mention to the best sister of the world, Teo, who always has an ear when I need to talk. I am looking forward now to meet her little daughter, Nara, the newest addition arriving soon to our family.

And finally, I want to thank Herbert for supporting and loving me.

About the Author

The author of this thesis, Enikő Szép, was born in Budapest, Hungary. She started her undergraduate studies at the Budapest University of Technology and Economics in 2009, where she later received both her Bachelor and Master degrees in Mathematics, with specialization in theoretical mathematics. During the last years of her university years, her main focus shifted to stochastics and its applications. To pursue her interest she applied for a summer internship at the Institute of Science and Technology Austria, where she had the opportunity to work in the research group of Nick Barton. In the following year she got accepted to the graduate program of IST Austria to study evolutionary biology and continue her work with Nick Barton.

List of Publications

1. Szép*, E., Sachdeva*, H., Barton, N. 2020. Polygenic local adaptation in metapopulations: a stochastic eco-evolutionary model, *submitted to Evolution*

2. Szép*, E., Trubenová*, B., Barton, N., Csilléry, K. 2020. gridCoal: An efficient tool to simulate spatially explicit null models of genetic diversity and divergence, *in preparation*

3. Bodová, K., Szép, E., Barton N. 2020. Dynamic Maximum Entropy provides accurate approximation of structured population dynamics, *in preparation*

Table of Contents

List of Tables

List of Figures

List of Abbreviations

DME Dynamic Maximum Entropy Method

EM Euler-Maruyama method

FPC foliar projective cover

FPE Fokker–Planck Equation

IBD isolation by distance

KFE Kolmogorov Forward Equation

LD linkage disequilibrium

LE linkage equilibrium

ME Maximum Entropy Method

ODE ordinary differential equation

SDE stochastic differential equation

1 Introduction

,,One small step for mankind,
one giant leap for me."

When we look around us in nature, we are immediately struck by the immense diversity we can observe: the various shapes, patterns, behaviors, and skills species have developed in order to function better in a given environment. In general, this process of adaptation happens on evolutionary timescales: taking millions of years to result in what we see today. Evolution is shaped by many forces; however, only selection leads to adaptation. The process of natural selection acts on the genetic variation present in the population and ensures increased reproductive success of individuals carrying beneficial alleles. Selection itself cannot create new variation, but the vast number of possible combinations of already existing genes makes it a powerful force that increases the frequency of the best options. Variation is created by other forces, such as mutation, migration, or in sexually reproducing populations, recombination. However, these forces interact with selection in complex ways: i) mutations are usually deleterious, but occasionally adaptive variation arises, ii) immigrants are likely to be maladapted in a new environment, but simultaneously, they can introduce new variation and alleviate inbreeding depression, iii) recombination breaks down associations between positively selected loci but it can also bring together selected alleles from different lineages. All these neutral and adaptive processes have a simple mechanism, yet they create extraordinary complexity. That

being said, the reverse problem, inference of the evolutionary processes from patterns of diversity, is also of great interest. Indeed, since the development of advanced genotyping methods, many inference frameworks were developed to connect empirical data to theory.

The overarching theme of this thesis is to understand how the interplay of neutral processes (e.g. population dynamics, population subdivision) and adaptive processes (selection) leads to patterns of genetic diversity and how it shapes evolutionary outcomes (e.g. local adaptation, extinction, speciation). Such questions have been central to population genetics for a long time, for example, what evolutionary processes (e.g. balancing selection, mutation) maintain genetic variation in natural populations, how much variation can be maintained (Fisher, 1930; Lande, 1975; Turelli, 1984; Barton & Keightley, 2002), or what fraction is neutral and what fraction is adaptive. Neutral processes affect the whole genome, while adaptive processes affect only targeted regions (loci affecting fitness), however, considering highly polygenic traits where the contributing loci are scattered over a larger part of the genome, this distinction between adaptive and neutral variation blurs.

Of course, the ultimate goal of population genetics is to understand how diversity is shaped by all the different ecological and evolutionary forces (i.e., demography, population structure, selection) together, and what can we infer from patterns of genetic variation about the processes that created them. In the following chapters of this thesis, we will always focus only on a subset of processes to build and understand complex models starting from simple building blocks, such as the logistic growth model, the island model, and directional selection.

1.1 Population dynamics

A key simplification one can make is to consider variation that has no effect on fitness. This means that although demographic processes affect variation, it is not a feedback loop and variation does not influence the demographic changes. Therefore, in the absence of selection, it makes sense to first consider a purely ecological process. Population dynamics describes the number of individuals in a population through time. The study of population dynamics dates back to the end of the eighteenth century, to the work of Malthus (1878). The idea of ,,struggle for existence" led to both the concept of natural selection and the

competition between species for resources - a fundamental topic for early ecologists (Gause et al., 1934). Although evolution and ecology took separate paths and developed into separate disciplines, they are inherently entangled and one can only be fully understood in light of the other. To describe population dynamics, various models with density regulation have been introduced (Henle et al., 2004). While the field of evolution is built from first principles, ecological processes can be unique for a given species in a given environment.

The simplest model of population dynamics is the logistic growth model (Verhulst (1838) in the continuous case, May (1976) in the discrete case), which ensures exponential growth for small populations and density regulation around carrying capacity. Although the study of population dynamics is focused both on single populations with age structure (Lotka, 1924; Leslie, 1966; Lande & Orzack, 1988; Tuljapurkar, 1990; Caswell, 2001; Lande et al., 2017), and on multiple populations with various competitions (Hassell & Comins, 1976), interactions between species (Bairey et al., 2016; Louthan et al., 2015), and community dynamics (Shoemaker et al., 2020), most work considering more than one population is set in a deterministic framework.

I will briefly mention stochastic models of population dynamics, as stochasticity is one of the main underlying feature of all the projects presented in this thesis. There are two types of stochasticity one can distinguish in this framework: demographic and environmental (Turelli, 1977; Lande et al., 2003). *Demographic stochasticity* refers to the random events affecting individual mortality and reproduction and is independent among the individuals, thus its effect averages out in large populations and has a pronounced effect in small populations close to extinction threshold. *Environmental stochasticity*, on the other hand, refers to temporal fluctuation in the environment and affects all individuals equally. Understanding how such random processes influence whether populations survive or go extinct is crucial. Lande (1988) argued that the main interests of conservationists is the maintenance of genetic variation, and avoiding inbreeding depression, however, neglecting demography (population growth and age structure) is a mistake as it may have more immediate effect on the survival of the species than its genetic composition (Frankham et al., 2017).

Many studies consider extinction dynamics of a single population, for example utilizing branching processes including the effects of demographic stochasticity but neglecting

environmental stochasticity (Wilson & MacArthur, 1967; Karlin & Taylor, 1981; Nåsell, 2001; Ovaskainen & Meerson, 2010), and Markov models for incorporating both demographic and environmental stochasticity with occasional catastrophes (Mangel & Tier, 1993). Stationary distributions of populations and mean times until extinction can also be found under various assumptions (Lande et al., 2003).

Another, different model to deal with multiple subpopulation is the colonization/extinction dynamics. It is based on metapopulation models (Levins, 1969; Hanski et al., 1997) in which each patch in the metapopulation has a probability of going extinct and becoming immediately recolonized. The population persist as a balance of this dynamics. One shortcoming of this model is that it neglects all local dynamics. Such models were explored in variable environments for example by Tuljapurkar (1989, 2013).

The diffusion approximation can also be used to describe population dynamics, and while it has been widely used in population genetics (Fisher, 1922; Kimura, 1955), it remains far less prominent in ecology (e.g., Lande (1993); Mangel & Tier (1993); Turelli (1977)). It works well under certain assumptions, such as continuous space and time, the absence of age structure, and sufficiently small increments.

In Chapter 2, we will use the simplest (but not trivial) model of population dynamics: the logistic growth with immigration (using the diffusion approximation) with the sole purpose of introducing the *principle of maximum entropy* to describe the time evolution of population sizes. The method of maximum entropy originates in thermodynamics (Jaynes, 1957) and it was already used in several different fields, including population genetics (Barton & de Vladar, 2009; Bod'ová et al., 2016), but not yet in ecology. Although the method comes from physics, and the quantities appearing in it are associated to very specific physical meaning, we find that in an ecological context, the same quantities carry relevant biological meaning. With this approximation, one can gain insight to the behavior of a complex system without having to deal with the full problem; that is: this method reduces a high-dimensional problem to an easily tractable one. It approximates the time evolution of certain observable quantities (such as population size, or its variance) through time, which otherwise can only be done if one solves the Fokker-Plank Equation which is usually not feasible. We will explain the details of the method, apply it to this very simple toy model, and show how well it approximates the real process.

Natural populations may experience shifts in the environmental parameters, due to

the cyclic repetition of the seasons, and to random catastrophic events, which may result is scarce resources or increased competition. Understanding how populations react to such changes is an important question (especially now as humanity faces climate change), however, treating it in a mathematically rigorous manner together with other sources of stochasticity is not straightforward.

Population size varying in time resulting in occasional extinctions, in fact, has a profound effect on genetic diversity. Such an effect was discussed by Mayr (1942), called the founder effect: it describes the reduction in genetic variance when a new population gets established by only a few individuals (founders) from a larger population. This effect is similar to what a population experiences after going through a bottleneck. Understanding such events is an important scope of conservation biology: it is crucial in order to identify minimal viable population sizes and plays an important role in agriculture (e.g. pest control).

1.2 Population structure

However, real populations change not only in time, but also in space. Usually they do not form a single well-mixed entity, rather they are distributed across space, where individuals are more likely to mate with individual in close proximity. This sort of division is called *population structure*, which either means that individuals are in a patchy environment, in which the local populations are panmictic and the demes are connected with limited migration, or they are in a continuous range, where the individuals are capable of only a restricted amount of dispersal. This results in a positive correlation between genetic and geographic distances (Wright, 1943), and the partitioning of neutral diversity according to the spatial structure (Charlesworth et al., 2003). When it originally emerged, this idea immediately launched the development of theoretical models for fitting demographic models and inferring gene flow from observed diversity. Although, the role of space and, in particular, spatial autocorrelation in allele frequencies has been recognized early on (Wright, 1943; Malécot, 1948), disproportionately more theoretical and methodological developments focused on understanding the effect of temporal changes in population size and gene flow between spatially non-explicit populations (e.g. island models) (e.g. Hey & Nielsen, 2007; Excoffier et al., 2013).

The most standard measure used for population structure inference is *Wright's fixation index*: F_{ST} (Wright, 1949), as it allows one to estimate Nm under the assumptions of an infinite island model. Later several similar summary statistics appeared that are equivalent to F_{ST} (Nei, 1972; Weir & Cockerham, 1984). In particular, the statistics in terms of diversity between and within populations received increased attention because they can be easily calculated from empirical datasets. The modern population genetics approach is to express F_{ST} in terms of coalescence times (Slatkin 1991, 1993). The *coalescent model* was first proposed to study the Wright-Fisher model in a computationally tractable way (Kingman, 1978). It underlies most modern population genetics (Wakeley, 2009), and it provides the most efficient way to simulate genetic diversity under various evolutionary scenarios (Kelleher et al., 2016).

Another approach to understand the patterns of genetic diversity created in a subdivided population is to look at isolation by distance patterns. Individuals are more genetically related to other individuals that live nearby, as they are more likely to share a recent common ancestor than individuals further apart. The decay in relatedness, in pairwise F_{ST}, for example, gives a good estimate of the neighborhood size (Wright, 1946; Rousset, 1997; Shirk & Cushman, 2014), which is a function of the number of breeding individuals in the local neighborhood, the parent-offspring distance, and the population density. However, Meirmans (2012) urged the development of additional statistical approaches that are based on a spatially explicit null-model instead of the non-spatial ones, as they may introduce biases and misleading conclusions (Battey et al., 2020).

The relatively young field of *landscape genetics* also aims at understanding how spatial structure shapes genetic diversity (Manel et al., 2003), mostly by correlating measures of diversity to environmental variables. However, much of the landscape genetic literature uses spatial statistical and statistic principles and has no link to population genetic models (e.g. Guillot et al., 2005; Smouse et al., 2008; Forester et al., 2016).

In Chapter 4, we will introduce our simulation tool based on the software *msprime* which accounts for spatial and temporal differences between populations on a two-dimensional grid. We will compare the simulation results of diversity to existing theoretical predictions and assess how adding spatial or temporal changes alters the expectation of models relying on much stricter assumptions. Furthermore, we will use a dataset based on the climate model LPX-Bern (Ruosch et al., 2016), which describes the demographic history of silver

fir populations across Europe since the end of the last ice age. Trees, in general, have large effective population sizes. The coalescent theory, when no demographic changes and explicit spatial structure are assumed, would therefore suggest long coalescence times so that short period of 22000 years could not result in strong genetic differentiation. However, Hewitt (2000) showed overwhelming evidence of genetic changes as a result of the deglaciation and the resulting species range shifts. The main aim of this chapter is to develop a simulation that can provide a spatially explicit null-model for further studies, that more accurately describes the demographic events both in space and time.

1.3 Selection

So far we only considered neutral variation, but to understand how adaptation—even in the case of a single population—occurs, one must consider adaptive variance and selection acting on it. Many traits under selection are polygenic, i.e. affected by many sites of the genome, which makes studies of adaptation complicated as one needs to account for the statistical associations (linkage disequilibrium) building up between the loci. Even in the absence of selection, if the population is subdivided, one needs to account for such associations due to allele frequency differences at multiple loci. That being said, when populations are distributed across larger regions, they experience different local conditions, thus different selective pressures, creating various genotype-environment interactions. In the absence of other forces, divergent selection causes each population to evolve traits that are beneficial under their own local conditions, regardless of the consequences in other environments. This process can be hindered by migration, environmental fluctuations, or the lack of genetic variation, just to mention a few. Adaptation to heterogeneous environments was studied extensively, for example fixation probabilities of new mutations in a metapopulation (Nagylaki, 1980; Whitlock & Gomulkiewicz, 2005), probability of invasion of a new beneficial allele (Holt & Gomulkiewicz, 1997; Gomulkiewicz et al., 1999; Kawecki, 2000), and general conditions under which local adaptation can occur (Blanquart et al., 2012). Local adaptation can be observed in nature (e.g. Antonovics et al., 1971; Leinonen et al., 2011; Mateo et al., 2018) and studying it can also lead to interesting further conclusions, for example, about the rate of adaptive evolution (Gomi & Takeda, 1996), the strength of selection needed to overcome the effects of gene flow (Lenormand,

2002; Edelaar & Bolnick, 2012), the maintenance of genetic variation (Felsenstein, 1976), or speciation (Turelli et al., 2001; Gavrilets, 2003; Barton, 2010).

Most of the local adaptation literature deals with *soft selection*, i.e. when the population size is unaffected by the mean fitness of the population (Kawecki & Ebert, 2004). However, in case of strong environmental gradients or large differences between selective forces in different patches of the environment, populations may experience high genetic load that is comparable to the growth rate. When the loss of individuals due to high genetic load becomes of the same magnitude as the intrinsic growth, one needs to account for feedback mechanisms between the size of the population and its genetic composition. Such a feedback mechanism was described already by Fisher (1930, Ch. 2), which later led to the idea of *hard selection* (Wallace, 1975; Débarre & Gandon, 2011). A simple version of such a feedback loop is when deleterious mutations accumulate in the population: they increase the amount of selective deaths thus decreasing the population size. In a small population selection is less efficient, increasing the effect of drift and gene flow, which results in the population shrinking further (Haldane 1956).

Although evolution was for a long time considered too slow to affect ecological processes, several studies of rapid evolution proved this hypothesis wrong (Thompson, 1998; Grant & Grant, 2006; Kokko & López-Sepulcre, 2007). By now an entire field of eco-evolutionary dynamics (Schoener, 2011; Hendry, 2016) explores the joint effects of population dynamics and evolution. Such interactions play an important role in evolutionary rescue following a sudden environmental shift (Gomulkiewicz & Holt, 1995; Gonzalez et al., 2013), and in the survival of marginal populations (Kawecki, 2008), the colonization of peripheral habitats (Barton & Etheridge, 2018; Sachdeva, 2019), and in the emergence of sharp geographic range margins in continuous environments (Polechová & Barton, 2015; Polechová, 2018).

Although the interplay between genetics and demography can greatly influence the survival or extinction of a population, previous studies omitted key aspects of this dynamics. For example, the conditions, under which local adaptation in a spatially heterogeneous metapopulation is possible were explored by Blanquart et al. (2012), however, only in case of soft selection. In a quantitative genetics framework, Ronce & Kirkpatrick (2001) explicitly consider the coupling between fitness and population size in a metapopulation with multiple ecologically distinct habitats, but neglect all sources of stochasticity. A

different approach, used for instance by Hanski & Mononen (2011), assumes that the patches in a metapopulation go deterministically extinct based on the mean fitness of the population, however, they immediately get recolonized with a fixed rate. That being said, these models do not explicitly include the coupling between the stochastic dynamics of genotype frequencies and population sizes that are likely to influence extinction thresholds in marginal habitats.

In Chapter 3—which forms the main body of this thesis—we specify under what conditions local adaptation can be maintained in a rare environment and when maladapted sink populations emerge. Understanding evolution under such circumstances is crucial if one is to understand the long term survival of marginal populations, or the first steps leading toward speciation. We will use the diffusion approximation in an eco-evolutionary framework to describe the joint evolution of allele frequencies and population size in a heterogeneous environment. We will incorporate both demographic stochasticity and drift. The full solution requires numerical techniques, however in many interesting limits explicit analytical solutions exist.

The main aim of this thesis is to connect evolutionary processes to the patterns they result in: to gain a better understanding for how evolution and ecology shape the patterns of diversity or the evolutionary outcomes, such as extinctions or adaptation, and what we can infer from the vast amount of data available today. In two chapters, we develop accurate mathematical approximations for given problems to obtain analytical results and identify scaling parameters. These are both important aspects as they provide a deeper insight to complex problems. Of course, not all the problems are solved, and in Chapter 5 we will briefly discuss possible directions for future research.

2 Population dynamics in a changing environment

This work is a joint work with Katka Bodova and Nick Barton. Parts of this chapter will be part of a methodology paper planed for submission in the early fall.

2.1 Introduction

Finding exact solutions of complicated dynamical systems is, in general, nearly impossible. Even finding solutions numerically can be an equally challenging problem. The high dimensionality, the interplay of various forces with different effects and magnitudes, the nonlinear dependencies, and the role of stochasticity all contribute to the difficulties. Here we introduce a powerful method to deal with such situations. Instead of solving the complete model with all its complexity, we solve a low-dimensional set of ordinary differential equations in a way that the dynamics of interest will remain accurate.

Our goal in this chapter is to capture the dynamics of mean population size and other such macroscopic quantities in a changing environment. Such estimates are very well needed to aid conservation efforts, especially in our current time, facing climate change, habitat destructions, fragmentations, and species extinctions.

The method of maximum entropy (ME) has a long history by now. It was set in motion by two papers in thermodynamics by Jaynes (1957). He introduced the principle of maximum entropy using information theory to construct a probability distribution on the basis of partial knowledge, that leads to a statistical inference method. This is the estimate that expresses maximum ignorance, i.e. it makes the least claim to be informed

beyond the information that we have.

Originally, information theory and statistical physics had not much in common, even though the same value of $-\sum p_i \log(p_i)$appeared. This is in itself, of course, not enough to establish any further connections but Jaynes found a way to introduce thermodynamic entropy and the information entropy as the exact same concepts. This became possible only after Shannon (1948) showed a deeper meaning of information entropy, independent of thermodynamics, namely that it expresses uncertainty in a very natural way. The idea that there is a distribution that maximizes entropy subject to constraints justifies using this distribution for a method of inference. Freeing the theory of maximum entropy from its dependence on actual physical hypotheses allowed statistical mechanics to be now seen in a more general light.

The method of maximum entropy (ME) is a method of inference and has a form of a variational problem, where the entropy of the microscopic distribution is maximized, while enforcing macroscopic constraints (e.g. on the average energy, train mean, or average population size). This method was used to study bird flocking (Bialek et al., 2012), neuronal firing (Schneidman et al., 2006), or protein variability (Mora et al., 2010). This method assumes stationarity, which is not satisfied in many systems. Jaynes (1980) introduced another way of using the same principle: he applied the ME to trajectories instead of data points and used constraints on the dynamical features. This method is called the Maximum Caliber method (MC) and was reviewed by Dixit et al. (2018), and Ghosh et al. (2020) gave examples of biological system where MC was successfully used.

The previous methods, ME and MC, are methods of inference: they can only be used when data is available. However, one may use a similar idea when instead of data, the dynamics of the system (the Fokker–Planck Equation) is known. Solving this equation is difficult in general, even numerically. The method presented here, so called Dynamic Maximum Entropy (DME), will allow us to, instead of solving a complicated stochastic differential equation, to solve a set of ordinary differential equations, which as a much simpler task. The method combines the classic ME approach with the FPE assuming quasi-stationarity. This DME was introduced for polygenic systems to describe evolutionary processes already in 1997 (Prügel-Bennett & Shapiro, 1997), but only gained attention later (Barton & de Vladar, 2009; de Vladar & Barton, 2011; Bod'ová et al., 2016). It was shown that although this approximation should work well when the forces

are changing slowly, even with abrupt changes, the method still gives accurate estimates of the macroscopic dynamics.

In many cases, the macroscopic dynamics of a system can be measured, whereas the underlining microscopic behavior remains hidden. DME offers a general framework in which we can understand how these observable values (e.g. functions of the moments) evolve without exact knowledge about the distributions of the microscopic variables themselves.

We use an information entropy measure (Boltzmann, 1872; Shannon, 1948), that ensures an exact solution at equilibrium, and follow the dynamics of the observables that are governed by the ecological forces. At each time point, DME generates a distribution of microscopic variables that would be stationary, given the estimated parameters (in our case the growth rate, density regulation, and migration rate). These parameters we refer to as the effective forces, that maximize entropy subject to the information we have. This naturally introduces a coupling between the measurable quantities and the ecological forces shaping the dynamics, i.e the system can be described either by the observables or by the effective forces, as in physics.

First, we introduce briefly the well known logistic equation with immigration, and the general DME framework. Then we apply this to population size evolution, show its accuracy, and point out interesting features of the method.

2.2 The method

The principle of maximum entropy states that the probability distribution that best represents the current state of the system compared to a baseline distribution, only assuming knowledge about certain observed values coming from this distribution and nothing more, is the one that maximizes relative entropy. The method offers freedom in choosing this baseline distribution, but in the absence of any prior information this can be chosen as uniform. Maximum entropy can be interpreted as maximum ignorance, i. e. this distribution is the one that makes the least claim of being informed beyond the stated observables: this is the maximum likelihood estimate (MLE) of the distribution, given the observables. Mathematically, this becomes an optimization problem in which we need to find the dis-

tribution that maximizes information entropy subject to constraints on this information. This can be done using the method of Lagrange multipliers.

The definition of entropy was introduced by Shannon (1948), originally for discrete spaces, as part of his theory of communication. The entropy of a random variable measures an average level of information, surprise, or uncertainty that is inherent to that random variable. However, our problem lies in a continuous space, therefore following the method introduced by Jaynes (1957), we use a slightly different entropy measure, the Kullback-Leibler divergence, or also known as relative entropy.

$$S_H(\Psi) = -\int_{[0,\infty]^L} \Psi(n) \log\left(\frac{\Psi(n)}{\Phi(n)}\right) \mathrm{d}n \tag{2.1}$$

This measure assumes a reference distribution Φ, that we in our case have a freedom to choose as convenient. Jaynes called this distribution an invariant measure. The relative entropy can be interpreted as the amount of information that is lost when the distribution Ψ is used to approximate Φ: how much information is added to the reference distribution due to specific choice of Ψ.

Now we can state the maximum entropy principle formally. We need to maximize the following functional in Ψ:

$$\max_{\Psi} S_H(\Psi) = \max_{\Psi} \left(-\int_0^\infty \Psi(n) \log\left(\frac{\Psi(n)}{\Phi(n)}\right) \mathrm{d}n \right),$$

with $m+1$ constraints:

$$\int_0^\infty \Psi(n) A_k(n) \mathrm{d}n = \langle A_k(n) \rangle_{obs}, \quad k \in \{1, \ldots, m\}, \quad \int_0^\infty \Psi(n) \mathrm{d}n = 1. \tag{2.2}$$

This constrained problem can be translated into an unconstrained maximization of the Lagrange function, with multipliers λ and α:

$$\mathcal{L}(\Psi, \lambda, \alpha) = S_H(\Psi) - \lambda \left(\int_0^\infty \Psi(n) \mathrm{d}n - 1 \right) - \sum_{k=1}^{n} \left(\int_0^\infty \Psi(n) A_k(n) \mathrm{d}n - \langle A_k \rangle_{obs} \right). \tag{2.3}$$

Let us take the variational derivative of $\mathcal{L}$:

$$\begin{aligned} \frac{\delta \mathcal{L}(\Psi, \lambda, \alpha)}{\delta \Psi} &= \frac{\mathrm{d}}{\mathrm{d}\epsilon} \mathcal{L}(\Psi + \epsilon \xi(n)) \Big|_{\epsilon=0} = \\ &= \int_0^\infty \xi(n) \left[-1 - \log\left(\frac{\Psi(n)}{\Phi(n)}\right) - \lambda - \sum_{k=1}^{m} \alpha_k A(n) \right] \mathrm{d}n. \end{aligned}$$

To have this integral equal zero, the term in the bracket must equal zero. Solving for $\Psi(n)$:

$$\Psi(n) = \frac{1}{\Phi(n)} \exp\{-1 - \lambda - \sum_{k=1}^{m} \alpha_k A_k\}, \tag{2.4}$$

where the normalization constant is $\exp\{-1-\lambda\}$ and the Lagrange multipliers are chosen such that the observables are accurately matched. We will show, that the distribution $\Psi(n)$ is the same as the stationary solution of the FPE written for the original diffusion problem, if the stationary distribution can be written in a form of a potential function, i.e. the system is in detailed balance. The ecological forces are the Lagrange multipliers (growth rate, density regulation, and migration) and the observables A are associated with the underlying processes.

2.3 Logistic growth in a continent-island model

In natural populations several factors impose bounds on exponential population growth, such as predation, or scarce resources—all these effects become more severe with higher population densities. The logistic growth model describes this process: the total growth of the population decreases linearly with increasing population size. When the population size is low, in the absence of competition for resources, the population grows with its intrinsic growth rate, r. When the population reaches the point where each individual replaces itself in each generation, we say the population is at carrying capacity, K. This is the maximal value of population size, at which the population can sustain itself. Above this value competition for resources increases, and the population shrinks in size. Migration can be introduced, as the additional source of population growth. A constant influx of individuals can prevent the extinction of the population, which is unavoidable when each individual has, in expectation, only one offspring.

2.3.1 DME applied to the logistic growth model

We apply the DME method for the logistic growth model with demographic stochasticity. The following SDE describes the population size changes:

$$\mathrm{d}n = [n(r - \lambda n) + m]\,\mathrm{d}t + \sqrt{\gamma n}\mathrm{d}\xi, \tag{2.5}$$

where r is the intrinsic growth rate, λ is the density regulation, m is the number of incoming immigrants, and γ describes the variance in population size. For the sake of simplicity, we fix $\gamma = 1$ for the rest of the chapter, this corresponds to the assumption that individuals have Poisson(1) offspring, therefore the total variance of the system is n.

The corresponding Forward Kolmogorov Equation, or Fokker–Planck Equation, describing the time evolution of the probability distribution $\Psi(n,t)$ is

$$\partial_t \Psi = -\partial_n \left[(n(r - \lambda n) + m)\Psi \right] + \frac{1}{2}\partial^2_{nn}(n\Psi). \tag{2.6}$$

Stationary distribution. The stationary solution of Equation 2.6 can be found in the form of a potential function. Note that it indeed has the same form as the distribution that we observed earlier to maximize entropy (Equation 2.4):

$$\Psi(n) = \frac{1}{Z}\frac{1}{n}\exp\left\{ 2(rn - \frac{\lambda n^2}{2} + m\log(n)) \right\} = \frac{1}{Z}\Phi(n)\mathrm{e}^{2\alpha A(n)},$$

where $\Phi(n) = \frac{1}{n}$ is the baseline distribution (the stationary solution without any forces acting on the system), $\mathbf{A} = (n, -\frac{n^2}{2}, \log(n))$ is a set of observables, and $\alpha = (r, \lambda, m)$ is a set of the ecological forces governing the system. The potential function $\boldsymbol{\alpha} \cdot \mathbf{A}$ consists of the effects of growth, density regulation, and migration. Here we point out, that it is a crucial assumption that the forces act through only the observable quantities, and all changes occur exclusively due to changes in the forces α. The expectations of the observables may have biologically meaningful interpretations, and can, in principle, be measured. In our case, $\langle n \rangle$ corresponds to the average population size, $\langle n^2 \rangle$ to the second moment of population size, and the third term, $\langle \log(n) \rangle$ is the logarithm of population size. The normalizing constant Z, which is the function of the effective forces α, plays an important role, as a generating function for quantities of interest:

$$\frac{\partial \log(Z)}{\partial(2\alpha_j)} = \langle A_j(n) \rangle, \qquad \frac{\partial \langle A_j(n) \rangle}{\partial(2\alpha_i)} = \mathrm{Cov}(A_i(n), A_j(n) = C_{i,j}. \tag{2.7}$$

The covariance matrix, denoted by C, plays an important role in the dynamical approximation.

Dynamics. Given a set of forces α, the system evolves to a stationary state that is distributed according to $\frac{\Phi(n)}{Z}\exp\{2\alpha \cdot A(n)\}$. This is the distribution that maximizes entropy with constraints on the observables, where $2\boldsymbol{\alpha}$ serves as the Lagrange-multiplier.

We are interested in how the system changes when a set of initial forces α_0 suddenly changes to a new set of values α_1. The observables will evolve towards the new stationary state, which creates a path between α_0 and α_1 in the space of effective forces. We will later see that although the transition in the observable quantities is rather slow and monotonic, the changes in the effective forces can be abrupt and non-monotonic. We can derive ordinary differential equations for the changes in the mean of the observables, however, the system will not be closed. Following (Barton & de Vladar, 2009), we derive the mean and mean-square changes in an infinitesimal time δt:

$$\langle \delta n \rangle = n \frac{\partial(\alpha \cdot A)}{\partial n}, \qquad \langle \delta n^2 \rangle = n.$$

Under the diffusion approximation, the rate of change in the mean of the observables $\langle A_i(n) \rangle$ is the following:

$$\frac{\partial \langle A_i(n) \rangle}{\partial t} = \frac{\partial A_i}{\partial n} \langle \delta n \rangle + \frac{1}{2} \frac{\partial^2 A_i}{\partial n^2} \langle \delta n^2 \rangle = \sum_j B_{i,j} \alpha_j + \frac{1}{2} V_i. \tag{2.8}$$

This equation is exact, when B and V are evaluated at the real distribution of population sizes. Applying Equation 2.8 to $\mathbf{A} = (n, -\frac{n^2}{2}, \log(n))$ and $\alpha = (r, \lambda, m)$, we obtain the following set of ODEs:

$$\begin{aligned}
\langle n \rangle' &= r\langle n \rangle - \lambda \langle n^2 \rangle + m, \\
\langle n^2 \rangle' &= (2m+1)\langle n \rangle + 2r\langle n^2 \rangle - 2\lambda \langle n^3 \rangle, \\
\langle \log(n) \rangle' &= r - \lambda \langle n \rangle + \left(m - \frac{1}{2} \right) \left\langle \frac{1}{n} \right\rangle,
\end{aligned}$$

that can be written using the matrix notation:

$$\langle \mathbf{A} \rangle' = \begin{pmatrix} \frac{1}{2}\langle n \rangle & \langle n^2 \rangle & \frac{1}{2} \\ \langle n^2 \rangle & 2\langle n^3 \rangle & \langle n \rangle \\ \frac{1}{2} & \langle n \rangle & \frac{1}{2}\left\langle \frac{1}{n} \right\rangle \end{pmatrix} \boldsymbol{\alpha} + \begin{pmatrix} 0 \\ \langle n \rangle \\ -\frac{1}{2}\left\langle \frac{1}{n} \right\rangle \end{pmatrix} = \mathbf{B}\boldsymbol{\alpha} + \mathbf{V}. \tag{2.9}$$

Quasi-stationarity. We introduce here the main assumption of the DME method, namely, the quasi-stationarity assumption. At each time point, we approximate the elements of B and V from Equation 2.8 by the values, that they would have if they were evaluated at the corresponding stationary distribution that generates the actual observables. This means, that for every time point, there are effective forces $\boldsymbol{\alpha}^*$, such that $\mathbf{B}_{\boldsymbol{\alpha}^*}\boldsymbol{\alpha}^* + \mathbf{V}^* = 0$ ($\mathbf{B}_{\boldsymbol{\alpha}^*}$ is where the moments in matrix B are evaluated at the stationary

distribution with parameters $\boldsymbol{\alpha}^*$). Substituting $\mathbf{V} = -\mathbf{B}_{\boldsymbol{\alpha}^*}\boldsymbol{\alpha}^*$ into (2.9) and using that $\mathbf{B} \approx \mathbf{B}_{\boldsymbol{\alpha}^*}$, we obtain

$$\frac{\partial\langle A_i(n)\rangle}{\partial t} \approx \sum_j B^*_{i,j}(\alpha_j - \alpha^*_j). \tag{2.10}$$

The moments of n appearing in matrix $\mathbf{B}_{\boldsymbol{\alpha}^*}$ can be expressed analytically under the condition that the migration rate is not too low ($m > \frac{1}{2}$). Let us call the kth moment of the stationary distribution $G(k)$, this can be expressed analytically in terms of hypergeometric functions:

$$G(k) = \langle n^k\rangle = \frac{1}{2}\lambda^{\frac{1}{2}(-1-k-2m)}\left(\sqrt{\lambda}\Gamma\left(\frac{k}{2}+m\right){}_1F_1\left(\frac{k}{2}+m,\frac{1}{2},\frac{r^2}{\lambda}\right) + \right.$$
$$\left. + 2r\Gamma\left(\frac{k+1}{2}+m\right){}_1F_1\left(\frac{k+1}{2}+m;\frac{3}{2};\frac{r^2}{\lambda}\right)\right),$$

if $\mathrm{Re}\,(k + 2m/\gamma) > 0$. Using the function G, all the moments of interest can be expressed such as:

$$G(0) = Z, \qquad \frac{G(1)}{G(0)} = \mathbb{E}(n), \qquad \frac{G(2)}{G(0)} = \mathbb{E}(n^2). \tag{2.11}$$

Furthermore, we can express $\mathbb{E}(\log(n))$ analytically by taking the j^{th} derivative of $G(k)$ with respect to m:

$$H(k,j) = \mathbb{E}(n^k \log(n)^j) = \frac{1}{2^j}\frac{\partial^{(j)}G_k}{\partial m^j}. \tag{2.12}$$

Changing variables in Equation 2.10 yields the dynamics of $\boldsymbol{\alpha}^*$:

$$\frac{\mathrm{d}\boldsymbol{\alpha}^*}{\mathrm{d}t} = \left[\frac{d\langle\mathbf{A}\rangle_{\boldsymbol{\alpha}^*}}{d\boldsymbol{\alpha}^*}\right]^{-1}\frac{\mathrm{d}\langle\mathbf{A}\rangle_{\boldsymbol{\alpha}^*}}{\mathrm{d}t}. \tag{2.13}$$

The matrix $\mathbf{C}_{\boldsymbol{\alpha}^*} = \frac{d\langle\mathbf{A}\rangle_{\boldsymbol{\alpha}^*}}{d\boldsymbol{\alpha}^*}$ is the covariance matrix of the observables, and can be expressed in terms of the previously defined functions G and H evaluated at the stationary distribution with parameters $\boldsymbol{\alpha}^*$:

$$\mathbf{C}_{\boldsymbol{\alpha}^*} = \begin{pmatrix} \frac{G(2)}{G(0)} - \frac{G(1)^2}{G(0)^2} & \frac{1}{2}\left(\frac{G(1)G(2)}{G(0)^2} - \frac{G(3)}{G(0)}\right) & \frac{H(1,1)}{H(0,0)} - \frac{G(1)H(0,1)}{G(0)H(0,0)} \\ \frac{1}{2}\left(\frac{G(1)G(2)}{G(0)^2} - \frac{G(3)}{G(0)}\right) & \frac{1}{4}\left(\frac{G(4)}{G(0)} - \frac{G(2)^2}{G(0)^2}\right) & \frac{1}{2}\left(\frac{G(2)H(0,1)}{G(0)H(0,0)} - \frac{H(2,1)}{H(0,0)}\right) \\ \frac{H(1,1)}{H(0,0)} - \frac{G(1)H(0,1)}{G(0)H(0,0)} & \frac{1}{2}\left(\frac{G(2)H(0,1)}{G(0)H(0,0)} - \frac{H(2,1)}{H(0,0)}\right) & \frac{H(0,2)}{H(0,0)} - \frac{H(0,1)^2}{H(0,0)^2}. \end{pmatrix} \tag{2.14}$$

Now we change variables in Equation 2.10 and using the formula 2.7, we obtain the dynamics of $\boldsymbol{\alpha}^*$ as a set of ODEs:

$$\frac{\mathrm{d}\boldsymbol{\alpha}^*}{\mathrm{d}t} = \left[\frac{\mathrm{d}\langle\mathbf{A}\rangle_{\boldsymbol{\alpha}^*}}{\mathrm{d}\boldsymbol{\alpha}^*}\right]^{-1}\frac{\mathrm{d}\langle\mathbf{A}\rangle_{\boldsymbol{\alpha}^*}}{\mathrm{d}t} = \mathbf{C}^{-1}_{\boldsymbol{\alpha}^*}\mathbf{B}_{\boldsymbol{\alpha}^*}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*). \tag{2.15}$$

2.4 A numerical example

To explore the DME method and illustrate the properties of the solutions, let us first introduce an arbitrary example. Let the initial set of forces be $\alpha_0 = \{0.05, 0.005, 1\}$, and after a sudden change (at time 0) their new values become $\alpha_1 = \{0.1, 0.002, 3\}$. The force λ is responsible for the density regulation of the population, yet its role might be more clear in terms of carrying capacity. We can interpret it as the ratio of the growth rate and the carrying capacity, thus our values translate into the initial carrying capacity being around 10, and the final one around 50. Of course, there is incoming migration, therefore the expected population sizes will be higher than these values.

We will show stochastic simulations of the population size changes, the corresponding dynamics of the effective forces predicted by the DME method, compare the time evolution of distributions resulting from various methods, and we will point out some interesting features of the system.

We will use the same parameters throughout the section, so we can better understand how the dynamics of population size translate into the dynamics of the effective forces and vice versa. We keep in this section the color scheme of the plots consistent, so one can easily find the matching values across the figures.

2.4.1 Stochastic simulation of the population size

To evaluate how well the DME method works, we need to compare it to the real solution of the system. Of course, if finding the real solution were without any trouble, we would not indulge in introducing such a complicated method as this. There are several ways to obtain approximating solutions, each introducing its own type of error. We will compare three methods.

We will be using the transition matrix (TM), that describes the process with the most biological feasibility, yet it is different from our process in two ways: it is a discrete process as it considers actual individuals dying or being born, and the variance of the process differs from the variance term in our original equation (see Equation 2.5).

The second method we will compare the DME to is the Euler–Maruyama method (EM), which is a stochastic simulation of Equation 2.5. How close we get to the real solution

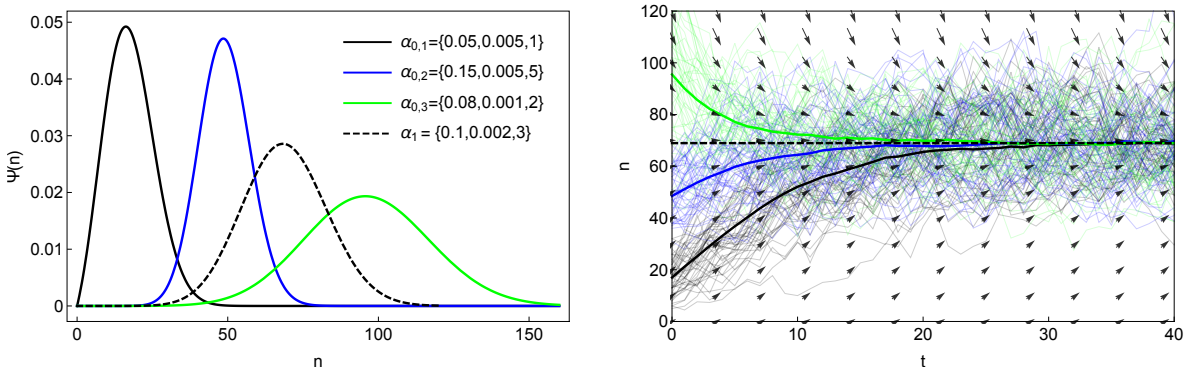


Figure 2.1: Euler-Maruyama simulation of individual trajectories of the process, starting from values randomly drawn from the stationary distribution with parameters $\alpha_{0,1} = \{0.05, 0.005, 1\}$, $\alpha_{0,2} = \{0.15, 0.005, 5\}$, and $\alpha_{0,3} = \{0.08, 0.001, 2\}$. The system is initially at the stationary state and after a sudden change in the forces to $\alpha_1 = \{0.1, 0.002, 3\}$, it starts evolving to a new stationary state. a) The initial distributions corresponding to the α_0s. b) Time evolution of the process corresponding to the different initial states (colorful thin lines: individual trajectories starting from the same initial distribution, thick line: mean of the trajectories, dashed black line: new mean defined by the new equilibrium), and the vector field defined by the original equation.

with this simulation solely depends on the number of replicates we take: the magnitude of stochastic fluctuations decreases with the square root of their number.

The last method is solving numerically the Fokker–Planck Equation (FPE). This introduces errors coming from the numerics itself, furthermore we need to use an additional boundary condition.

Transition matrix (TM). Let us introduce the following Markov-chain, a continuous time birth-death process on the discrete space of non-negative integers. The birth rate is $(1/2 + r)$, the death rate is $(1/2 + \lambda n)$, and the rate of immigration is m. We assume that each time point only one event happens, therefore the population size changes with the following rates:

$$\{-1, 0, +1\} \to \left\{ \left(\frac{1}{2} + \lambda n\right) n, -n(1 + r + \lambda n) - m, \left(\frac{1}{2} + r\right) n + m \right\},$$

where n is the current size of the population. Using the transition rates we just defined, the transition matrix $\underline{T}$ can be constructed so $P(t + \Delta t) = \underline{T}.P(t)$. The stationary state can be found as the left eigenvector corresponding to the biggest eigenvalue, which is 1. The mean and mean square of this process is

$$\mathbb{E}(\Delta n) = m + n(r - \lambda n), \qquad \mathbb{E}(\Delta n^2) = m + n(1 + r + \lambda n).$$

The stationary solution ($\tilde{\Psi}$) of the diffusion that this process converges to can be found as the solution of the following equation:

$$0 = -(m + n(r - \lambda n))\tilde{\Psi} + \frac{1}{2}\partial\left[(m + (1 + r + \lambda n))\tilde{\Psi}\right]. \tag{2.16}$$

Comparing this to Equation 2.6 we notice that their variance terms do not match. However, assuming that $r + \lambda n, m \ll 1$ the previously defined birth-death process is a good approximation for the original diffusion, see Figure 2.2. The transition matrix gives the exact solution in the discrete case, however, the diffusion limit is continuous, therefore a discrepancy arises between the stationary state of the birth-death process and its diffusion limit, which disappears as $n \to \infty$.

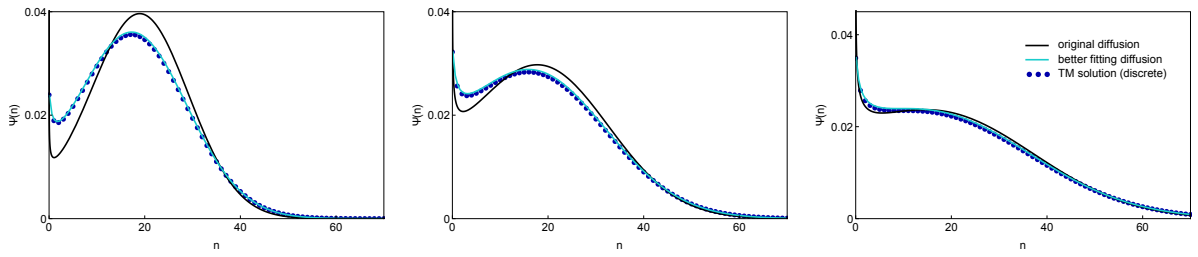


Figure 2.2: Comparison of stationary distributions resulting from the original diffusion (black line), the discrete death-birth process (blue dots), and a diffusion with the variance term matching that of the birth-death process (blue line). There is a small discrepancy between the birth-death process and its diffusion limit, due to one process being discrete whereas the other is continuous. The difference between the original diffusion equation and the birth-death process is larger, however, the distributions converge as $r \to 0$. We keep $K = \frac{r}{\lambda} = 20$, and $m = 0.4$ fixed and decrease the growth rate. a) $r = 0.1, \lambda = 0.005$, b) $r = 0.05, \lambda = 0.0025$, c) $r = 0.025, \lambda = 0.00125$.

Euler–Maruyama method (EM). One way to find a numerical solution for a stochastic differential equation is to generate trajectories according to the equation itself. Let X be a stochastic process such that

$$\mathrm{d}X_t = a(X_t)\mathrm{d}t + b(X_t)\mathrm{d}W_t, \tag{2.17}$$

with initial condition $X_0 = x_0$. Let $0 < \tau_0 < \tau \ldots \tau_n = T$, and $\Delta t = \frac{T}{N}$. The Euler–Maruyama approximation of the true solution X is a Markov chain Y, that can be recur-

sively defined such that:

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_t)\Delta W_n, \tag{2.18}$$

and $Y_0 = x_0$, and $\Delta W_n = W_{\tau_{n+1}-\tau_n}$.

Trajectories from different initial conditions. We use the Euler-Maruyama method to generate trajectories for Equation 2.5, starting from three arbitrary initial conditions: $\alpha_{0,1} = \{0.05, 0.005, 1\}$, $\alpha_{0,2} = \{0.15, 0.005, 5\}$, and $\alpha_{0,3} = \{0.08, 0.001, 2\}$, for their distributions and trajectories see Figure 2.1. Substituting these values into the stationary distribution we find the initial expectations of population sizes to be $17.76, 48.96$, and 96.37. At time 0, after a sudden change in the forces, the system starts to evolve toward the new stationary state and stabilize around the new equilibrium solution.

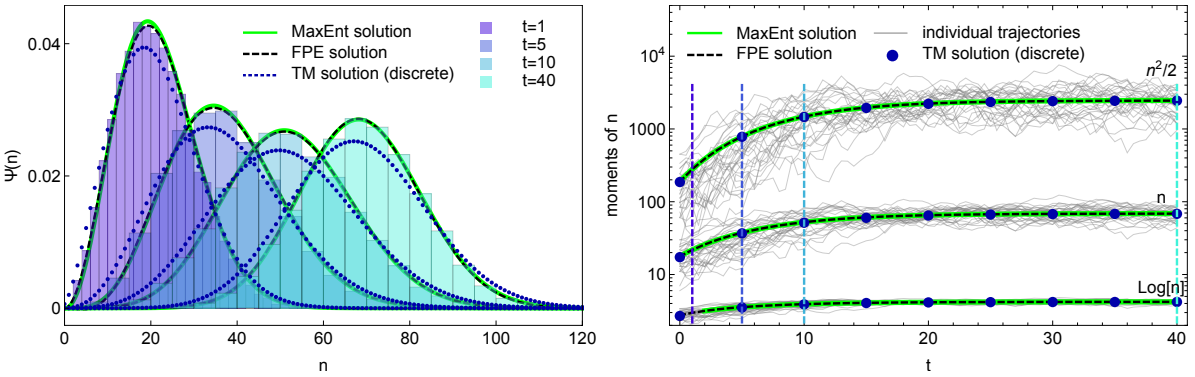


Figure 2.3: Correspondence between the Euler-Maruyama approximation of the ODE, the transition matrix predictions, the numerical solution of the Fokker–Planck Equation, and the distributions predicted by the DME method. The system starts from the equilibrium state with $\alpha_0 = \{0.05, 0.005, 1\}$. At time 0 the forces suddenly change to $\alpha_1 = \{0.1, 0.002, 3\}$, and now the system starts to evolve toward the new equilibrium. In general, the system will not precisely follow the distribution that maximizes entropy, however, the distribution of microscopic variables stays close to the maximum entropy distribution, if the change in the macroscopic variables is small enough. a) Distribution of population sizes at different time points using different approximations. The DME method does not suggest this strong agreement, it only assures that the observables are correctly matched, but their distributions do not necessarily coincide with the DME distributions. b) Time evolution of the observables. Simulations are compared to the changes predicted by DME. (Averages are calculated using 50000 replicates, however, only a few of these trajectories are plotted.) We plotted n on a logarithmic scale with the sole purpose of showing all observables on a single plot.

Let us compare the DME method (here only for one initial condition) to the discrete predictions of the transition matrix method, the numerical solution of the FPE, and

the EM simulations, see Figure 2.3. We find that the histograms of the samples at $t = 1, 5, 10, 40$ are in a good agreement with the DME distributions. Furthermore, the changes occurring in the observables also show a good match between all the solutions. It is important to clarify, that the DME solution is constructed such that only the observables match: the close agreement between the histograms and the distributions themselves is not a requirement.

2.4.2 DME dynamics of the effective forces

Now let us turn our attention to the corresponding dynamics of the effective forces. We saw that the changes in the observables are monotonic so the question naturally arises whether we can say the same about the effective forces as well. The short answer is no, but let us see this in more detail.

We apply the DME method to the logistic growth model that results in a three dimensional dynamics in the space of effective forces. We assumed quasi-stationarity, i. e. at each time point we approximate the real distribution of n with one that is of the stationary form, and its parameters are those effective forces which maximize entropy and produce the correct observables. In Figure 2.4 we show various cross-sections of the 3-dimensional vector field defined by the ODEs (2.15) that governs the forces, and the paths that belong to the same initial and final values as we used earlier. We find that paths that the effective forces follow are complicated, due to their non-linear dependencies, and that they act on different scales. Such behavior was observed for the evolutionary forces on quantitative traits as well (Bod'ová et al., 2016).

2.4.3 Irreversibility

We saw that the paths the effective forces take between two points are non-trivial. Here we show yet another interesting feature, namely that the path between α_0 and α_1 is not the same as the path between α_1 and α_0, see Figure 2.5. The density of the dots shows the speed of the dynamics, as they are equally spaced in time. In the beginning the density is low, meaning that the system changes faster, which slows down getting closer to the final value. This is the consequence of equation 2.15, the force pulling the values close the optimum becomes weaker.

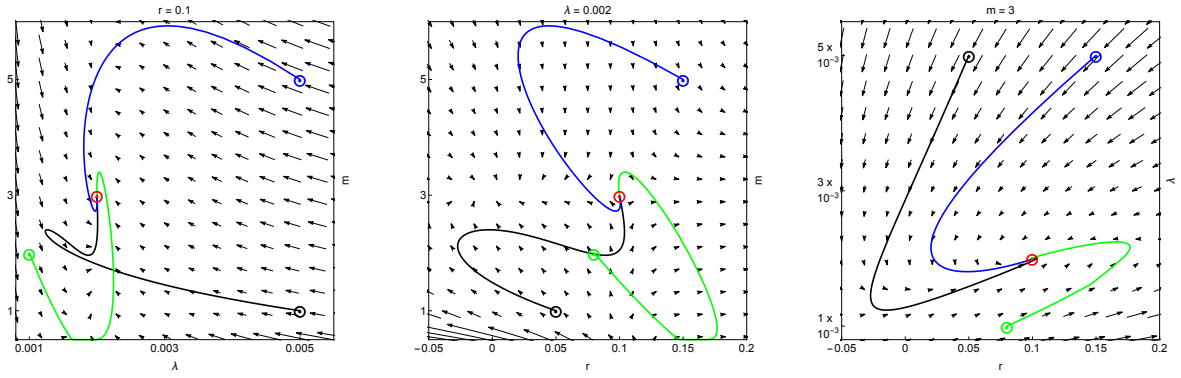


Figure 2.4: DME applied to the logistic growth model leads to a $3D$ dynamics in the space of effective forces $\{r^*, \lambda^*, m^*\}$. At each time point we approximate the distribution of n with the stationary distribution with effective forces predicted by Equation 2.15. The initial conditions shown here are corresponding to the ones used in the previous figures (cf. Figure 2.3): $\alpha_{0,1} = \{0.05, 0.005, 1\}$ (black), $\alpha_{0,2} = \{0.15, 0.005, 5\}$ (blue), and $\alpha_{0,3} = \{0.08, 0.001, 2\}$ (green). Since the dynamics are three dimensional, we show cross-sections of the vector field defined by the ODE system. At each direction we show the section at the final value, $r = 0.1, \lambda = 0.002$, and $m = 3$. The effective forces exhibit non-linear and non-monotonic behavior, yet their combined effect, the behavior of the observables, is monotonic.

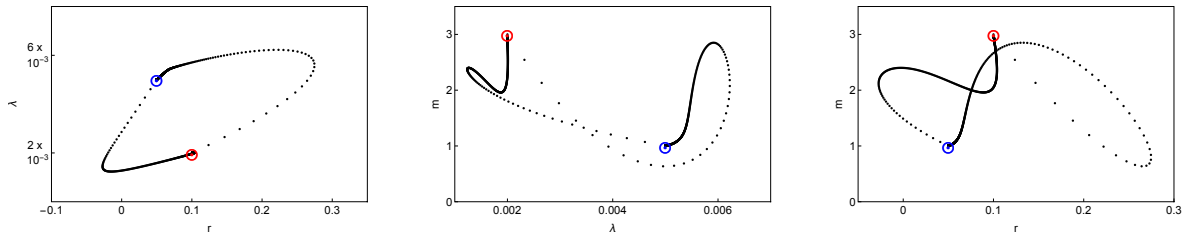


Figure 2.5: Cross-sections of the 3 dimensional space of effective forces and the paths between the points $\alpha_0 = \{0.05, 0.005, 1\}$ (blue circle) and $\alpha_1 = \{0.1, 0.002, 3\}$ (red circle) and reversed. The curves do not intersect themselves in the 3 dimensional space, only in the projection. The black dots are values equally spaced in time. The convergence to the new value is faster in the beginning, as the force pulling it close to the new equilibrium is stronger when they are further apart, see equation 2.15.

2.4.4 Error and accuracy of DME

To investigate in what parameter ranges we can use the method safely, and how the error depends on the parameters, we explored the following parameter space:

$$r \in \{0.01, 0.05, 0.1, 0.2\}, \qquad \lambda \in \{0.0005, 0.001, 0.005, 0.01, 0.05\}, \quad m \in \{1, 3, 5, 8\}.$$

We found that the numerical calculations in the DME approach blow up for very small values of λ. This happens because in Equation 2.15, we are required to invert a matrix,

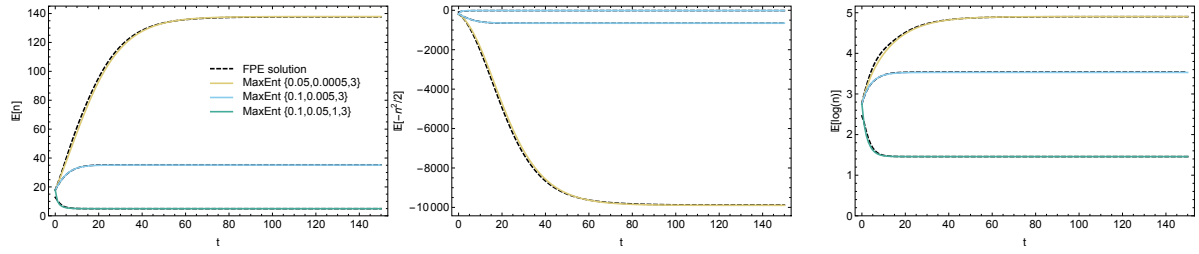


Figure 2.6: Solution of the DME method in comparison to that of the FPE. a) $\mathbb{E}(n)$ b) $-\mathbb{E}(n^2/2)$, c) $\mathbb{E}(\log(n))$

and with these parameters the matrix entries largely differ in magnitudes. This leads to inaccuracies in the resulting matrix, and eventually in the whole dynamics.

As we discussed earlier, the different methods we use to compare the DME introduce different types of errors. In case of the transition matrix approach, the variance depends on all evolutionary parameters instead of only n, as in Equation 2.5. The birth and death rates appearing in the transition matrix could be modified such that the process has the correct variance, nevertheless it would lose the biological meaning, therefore we omit to do so. This process is also a discrete one, therefore we cannot expect it to fully match the continuous version. However, as the growth rate (r) of the process approaches 0, the diffusion fits better.

One can use the Euler–Maruyama simulations: use the discretized stochastic differential equations to obtain sample trajectories and average them in order to obtain the dynamics of the moments. This approach is probably the simplest, however, it requires one to simulate a huge numbers of trajectories to decrease the stochastic scatter, which may be limited by computational power. There are variance reduction techniques which one could utilize to obtain closer estimates with fewer simulations.

The last, and our preferred, method is solving the FPE and calculate the observables from the resulting distributions. This approach also suffers from numerical inaccuracies, and obtaining this solution is not always possible in the aforementioned parameter ranges.

We observed no correlation between the magnitude of errors and the magnitude of changes, or the number of changing parameters, however, the discrepancy between the DME and the FPE remains very small, below 2% of the resulting change in the moments.

2.5 Periodically changing environments

Natural environments are constantly changing, both in space and time, nowadays even more, due to constantly increasing human activity. The strategies populations follow in order to cope with such changes can depend on several factors, such as the population internal ability to adapt, i.e. the amount of genetic variation present in the population, the speed of the environmental change, or its magnitude. Populations may end up migrating to a more suitable environment, or they may adapt to the newly arisen conditions. These sort of changes can facilitate evolution: spatial differentiation in the environment is required for adaptation that later may lead to speciation.

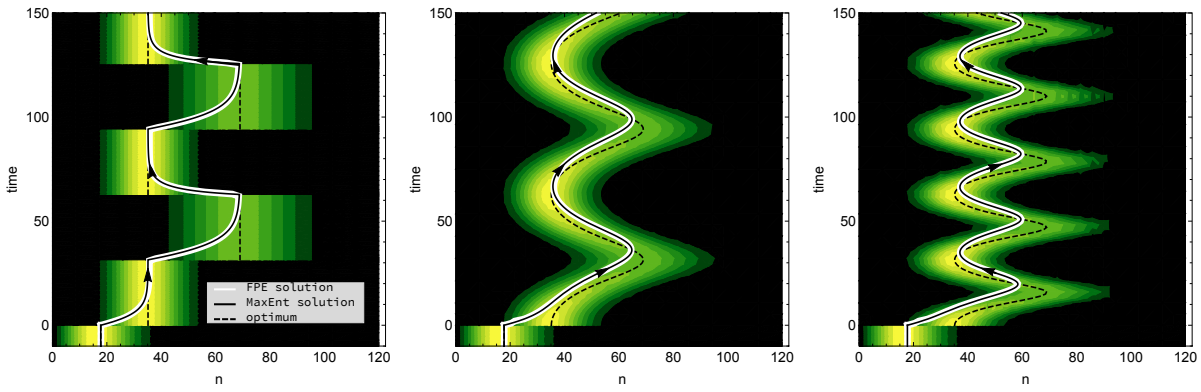


Figure 2.7: Periodic changes in the environment, namely the carrying capacity shifts between 20 and 50. For consistency with the previous examples, the system starts from equilibrium with parameters $\{0.05, 0.005, 1\}$, then a periodic shift between $\{0.1, 0.0005, 3\}$ and $\{0.1, 0.0002, 3\}$ is introduced. In green/yellow the equilibrium distribution of population size is shown as it changes in time. The black dashed line is the expectation of these distributions, the black solid line show the solution of the DME, whereas the white is the solution of the Fokker–Planck Equation. a) abrupt change b) smooth slow change c) smooth fast change

Temporal changes in the environment can be abrupt, causing populations to become maladapted and eventually even extinct, unless it can get out of this declining phase by a process called evolutionary rescue. However, less rapid environmental shifts can be observed yearly: the periodic changes of seasons can drastically change the carrying capacity of the environment, change the migration patterns, or the growth rate of populations.

Describing such eco-evolutionary changes can involve complicated dynamical systems, in which finding solutions, even only numerically, is difficult. In this section, we stay with the process of logistic growth, not considering any evolutionary force, nevertheless

we wish to point out the possibilities that the DME method offers in order to use it for more biologically challenging situations, such as in Chapter 3.

Here we investigate how well the method captures the real dynamics in a situation, where the forces shaping the environment change periodically between two sets of parameters. We compared three scenarios: abrupt change in the forces, and smooth changes considering fast or slow shifts, see Figure 2.7. We start the system from equilibrium, from the same point as earlier, $\{0.05, 0.005, 1\}$. Then we introduce a periodic shift of optima between $\{0.1, 0.0005, 3\}$ and $\{0.1, 0.0002, 3\}$. This change corresponds to a shift between carrying capacities 20 and 50.

We have seen already that the DME method produces rather small errors for large and abrupt changes, therefore in case of smooth shifts we expect the same. As on Figure 2.7 one can see, the difference between the DME solutions and the real solution is negligible in all cases. More interestingly, we see that the system lags behind the optimum of the environment, and that the amount of this lag depends on the speed of change.

2.6 Discussion

We introduced the method of Dynamic Maximum Entropy for an ecological process, namely, to the logistic growth model with immigration. The method originally appeared in thermodynamics (Jaynes, 1957), and it was later applied in many other fields, also in evolutionary biology (de Vladar & Barton, 2011), to predict changes in trait means. This is a powerful method to obtain given moments of a complicated, high-dimensional dynamical system, which is difficult to solve efficiently in other ways. We found that the solution provided by this method remains very close to the real solutions for a wide range of parameters, however, we did not find any consistent dependency of the error on these parameters. In case of low migrations, when extinctions become possible, the method breaks down, which problem (for low mutation rates) was addressed in (Bod'ová et al., 2016) in a mathematically rigorous manner.

There is an interesting analogy to point briefly out here, namely with the original form of the maximum entropy method and the Boltzmann distribution. It describes the probability that a system will be in a certain state as a function of that state's energy (ϵ_i) and the temperature (T) of the system: $\exp\{-\frac{\epsilon_i}{kT}\}$, where k is the Boltzmann constant.

The behavior of this distribution can easily be seen in the opposite limits of temperature: when $T \to \infty$, all states are equally likely to be at (uniform distribution), and when $T \to \infty$, the state with the least energy (ground state) is the most likely. This corresponds to the idea of the particles moving the fastest when the temperature is high creating a low of randomness, whereas at cool temperatures they tend to be more still. When one compares this distribution to the stationary distribution obtained for a quantitative trait by Bod'ová et al. (2016), one can notice the similarity: the temperature acts inversely proportional to the population size. The high temperature scenario corresponds to small population size—exactly when randomness (the genetic drift) is strong, and the low temperature scenario corresponds to large population size, when the allele frequencies are more deterministic.

In our case, N is not a fixed quantity, so the question arises, what is the analogy to temperature in this ecological model. In the original equation (Equation 2.5), the constant γ appeared, which for sake of simplicity, we considered to equal 1. When the individuals have Poisson(1) offspring, then indeed $\gamma = 1$. However, if it differs from this distribution, then γ becomes meaningful: it is the strength of demographic fluctuations—this is the corresponding term of temperature.

This leads to a long standing and important question of ecology: how can one model changes in the environment such that demographic stochasticity (sampling variation in births and deaths) and environmental stochasticity (effect of environmental fluctuations on growth rate) are both simultaneously considered. Branching process are used to model demographic stochasticity but assume constant environment, (although the mathematical framework of time inhomogeneous branching processes exists) (e.g. Uecker et al., 2014), stochastic differential equations are used to model environmental stochasticity (mostly) assuming close to extinctions threshold scenarios (Mangel & Tier, 1993; Engen et al., 1998; Lande et al., 2003) . Understanding how demographic and environmental stochasticity together influences the viability of endangered species is a key challenge in conservation biology (Sæther & Engen, 2019).

As a future direction, the method presented here could be used for such an analysis: demographic stochasticity can be represented with the appropriate scaling of the noise term of the SDE, and the the environmental changes and randomness, as we have seen it for the periodic change, can be enforced by changes on the effective forces.

The ultimate advantage of the DME method lies in its ability to reduce a compli-

cated system to only a few variables, yet containing the information we are interested in. However, deriving the method is not easy, even when the moments of the stationary distribution can be expressed analytically. As we pointed out, the method in its form presented here can not incorporate small migration rates, which is usually the regime where interesting behavior arises, such as extinctions. Of course, there is a way to tackle this problem—but it adds another layer of complication to an already complicated method. Computational power increases, and finding numerical solutions of equations, or even simulating the full dynamics of the systems becomes easier. This raises the question of what is the use of such a model? It may not be a practical one, however, it can give a fundamental understanding of the interplay of the evolutionary and ecological forces.

3 Polygenic local adaptation in metapopulations: a stochastic eco-evolutionary model

This work is a joint work with Himani Sachdeva and Nick Barton. The chapter is presented as the paper submitted to Evolution.

Author contributions: E.S. and N.H.B. designed the study; E.S., H.S. and N.H.B. did the mathematical analysis; H.S. did the simulations; E.S., H.S. and N.H.B. wrote the manuscript.

3.1 Introduction

Adaptation to local environmental fluctuations can be quite rapid (Thompson, 1998; Grant & Grant, 2006; Kokko & López-Sepulcre, 2007; Kinnison & Hendry, 2001), such that the time scale of evolutionary change is comparable to ecological timescales, giving rise to feedback between demography and evolution. Reciprocal interactions between population size and adaptation were described by Fisher (1930), which later led to the notion of *hard selection*, whereby high genetic load and ensuing 'selective deaths' can drive populations to extinction (Haldane, 1956). This is an extreme example of a more general feedback loop: an increase in genetic load due to deleterious variants reduces population size; smaller populations are affected more strongly by drift and gene flow, which increase the fixation of locally deleterious alleles, which further decreases size.

Eco-evolutionary feedbacks of this sort are crucial during evolutionary rescue following a sudden environmental shift (Gomulkiewicz & Holt, 1995; Gonzalez et al., 2013), and also play a key role in the survival of marginal populations (Kawecki, 2008), the colonization of peripheral habitats (Barton & Etheridge, 2018; Sachdeva, 2019), and the emergence of sharp geographic range margins in the absence of environmental discontinuities (Polechová & Barton, 2015; Polechová, 2018).

Eco-evolutionary feedbacks are especially important in fragmented, heterogeneous habitats, where stochastic extinction and recolonization of patches may necessitate recurrent bouts of rapid adaptation, especially if selective pressures vary. This kind of metapopulation structure may arise, for instance, if multiple hosts are available within the same region (Carroll & Boyd, 1992; Dobler & Farrell, 1999); this can favor host-specific adaptations, leading to adaptive divergence between sub-populations that specialize on different hosts. The potential for local adaptation and the stability of sub-populations then depends on the interaction between selection (which is mediated by the genetic architecture of selected traits), dispersal (which protects populations from inbreeding load and stochastic extinction, but may also introduce maladapted phenotypes, thus generating hybridization load) and demography (which is affected by mean genetic fitness, and in turn influences the efficacy of selection).

Previous theoretical work on the persistence of subdivided populations neglects key aspects of this interplay. For instance, Blanquart et al. (2012) analyze conditions for local adaptation in a spatially heterogeneous metapopulation under *soft* selection, thus neglecting feedback between fitness and demography. Ronce & Kirkpatrick (2001) explicitly consider the coupling between fitness and population size in a metapopulation with multiple ecologically distinct habitats, but neglect all sources of stochasticity. Another modeling approach, exemplified by Hanski & Mononen (2011), assumes instant recolonization of patches (at a fixed rate) and subsequent deterministic extinction based on population fitness. However, such approaches do not explicitly consider the coupled *stochastic* dynamics of genotype frequencies and population sizes that are likely to influence extinction thresholds in marginal habitats. Conversely, work on the effects of demographic and environmental fluctuations typically does not consider evolutionary dynamics due to genetic change (e.g. Lande, 1993).

A second challenge is to incorporate realistic assumptions about the genetic architec-

ture of selected traits into eco-evolutionary models. At present, most metapopulation models assume one of two extreme architectures (Govaert et al., 2019; Lion, 2018), either mutations that occur one by one (adaptive dynamics models), or a very large number of infinitesimal effect loci (quantitative genetics). Unlike adaptive dynamics, quantitative genetic models can describe response from standing genetic variation, which often underlies rapid adaptation. However, with a few exceptions (Hanski & Mononen, 2011; Rouhani & Barton, 1993; Ronce & Kirkpatrick, 2001), most such models consider migration into a *single* population (e.g., Barton & Etheridge (2018); Chevin et al. (2017); Tufto (2001)).

Here, we investigate the joint evolution of population size and allele frequencies in an idealized metapopulation, consisting of infinitely many demes that exchange genes with a common migrant pool. We assume that each island belongs to one of several habitats that are characterized by distinct selection pressures, and ask: when can demographically stable, locally adapted populations be maintained within demes despite gene flow? As we argue below, gene flow limits local adaptation not only by opposing selection at individual loci, but also via constraints arising from migration load (Lenormand, 2002). Mean fitness is reduced by at least m (the rate of migration between demes) per locally adapted locus. With hard selection, this constrains the number of locally adapted alleles that can be maintained without population collapse.

An important focus of our study is to clarify the conditions under which demes belonging to a 'rare' or marginal habitat evolve or maintain local adaptation, or instead are reduced to maladapted sink populations. Understanding evolution in marginal habitats has important implications for range limits and the long-term survival of metapopulations, and because local adaptation in marginal habitats may be the first step towards speciation.

A second focus is to understand how local adaptation is influenced by eco-evolutionary feedbacks between demography, dispersal and selection, for *polygenic* traits influenced by many (but not necessarily very many) loci. A central challenge for analyzing polygenic evolution in subdivided populations is to account for statistical associations, i.e., linkage disequilibria (LD) between loci. The theoretical framework developed here neglects such associations by assuming linkage equilibrium (LE) within demes as well as within the migrant pool. This allows us to approximate the evolutionary dynamics of multi-locus genotypes solely by allele frequencies. Assuming LE does *not* imply that loci evolve

independently of each other; as demonstrated below, under hard selection, evolutionary dynamics of different loci become coupled due to their aggregate effects on population size (via average fitness), which in turn influences individual loci via genetic drift.

Our analysis is based on a *diffusion approximation* for the joint stochastic evolution of allele frequencies and population size. While the diffusion approximation has been widely used in population genetics (Fisher, 1922; Kimura, 1955), it remains less prominent in ecology, and has only been used to model stochastic population dynamics, without incorporating genetics (e.g., Lande (1993); Mangel & Tier (1993)). Demographic stochasticity (i.e., stochastic fluctuations in size) has a pronounced effect when populations are small and close to the threshold of extinction. Moreover, selection is less effective in small populations, causing maladaptive alleles to fix due to genetic drift, thus further reducing fitness and size, rendering the population even more vulnerable to stochastic fluctuations. This makes it necessary to account for demographic stochasticity and genetic drift together.

Our framework incorporates both types of stochasticity, following a method presented earlier for asymmetric continent-island migration (Banglawala, 2010). The full model requires a numerical solution, but explicit analytical predictions are possible in various biologically interesting limits. In order to assess the importance of LE and other assumptions underlying the diffusion framework, we compare our analytical predictions against individual-based simulations with a finite number of demes.

3.2 Model and Methods

Consider a metapopulation with infinitely many islands (demes) that exchange genes via a common pool. Individuals are haploid and express an additive trait that is influenced by L unlinked loci. Each locus can be in one of two alternative allelic states, denoted by $x = 0, 1$. The trait value z associated with an individual is the sum of effects of the alleles it carries: $z = \sum_{j=1}^{L} \gamma_j x_j$, where γ_j is the difference between effect sizes of alternative alleles at the j^{th} locus. The additive trait is under environment-dependent directional selection: the fitness $W_i(z)$ of an individual with trait value z on island i, is $e^{\beta_i z + \theta_i}$, where β_i is the strength of selection and θ_i a constant that determines the maximum possible fitness on the i^{th} island.

The life cycle of individuals consists of dispersal, followed by selection and mating. As

our primary focus is on how gene flow influences the maintenance of polymorphisms and local adaptation, we neglect other sources of variation. However, the framework can be easily extended to incorporate mutation. In each generation, a common migrant pool is formed by drawing a fraction m from each island; migrants from this pool are then evenly redistributed across islands. The assumption of infinitely many islands is convenient, since it allows us to treat genotype frequencies in the migrant pool as deterministic (rather than following a distribution). In simulations, we model a large but finite number of islands.

We assume a model of hard selection, where population size is influenced by genetically determined mean fitness plus density-dependent regulation, which occurs locally within each island. The size n_i^* on island i, after selection and regulation, is a Poisson random variable with mean $n_i \overline{W}_i e^{r_{0,i}(1-n_i/K_i)}$. Here, $r_{0,i}$ is the baseline rate of growth, K_i the carrying capacity, n_i the population size prior to selection, and $\overline{W}_i$ the mean genetic component of fitness on island i. The n_i^* offspring are formed by randomly sampling $2n_i^*$ parents (with replacement) from the n_i individuals in proportion to individual fitness, and then creating offspring via free recombination of each pair of parental genotypes. For simplicity, $r_{0,i}$ and K_i are taken to be the same across all islands: $r_{0,i} = r_0$ and $K_i = K$. However, this can be easily generalized to island-specific growth rates and carrying capacities.

We assume that any island belongs to one of g local environments or habitats, indexed by $\alpha = 1, 2, \ldots g$, such that the environment-dependent selection strength β can take on one of g possible values. We use $\rho_1, \rho_2, \ldots \rho_g$ to denote the fraction of islands that belong to the different habitats ($\sum_i \rho_i = 1$).

When selection is strong relative to migration and drift, and the number of loci not very large (see below), populations can adapt to their local habitat, resulting in LD between alleles favored in a habitat. To describe such a population exactly, we would need to track the frequencies of all genotypes, which becomes cumbersome with large numbers of loci. However, when selection per locus is weak relative to recombination, LD within a deme (generated by immigration of individuals from differently adapted habitats) is rapidly dissipated and can be neglected. Assuming LE allows us to only consider the coupled dynamics of population size and L alleles, rather than 2^L genotypes.

For weak growth, selection and migration (i.e., $r_0, \beta\gamma, m \ll 1$), we can use a continuous time approximation for allele frequency and population size dynamics. The size n_i and

the allele frequency $p_{i,j}$ at the j^{th} locus on the i^{th} island satisfy the following *coupled* equations:

$$\frac{\partial n_i}{\partial t} = \left[r_0(1 - \frac{n_i}{K}) + r_{g,i}\right] n_i + m(\overline{n} - n_i) + \lambda_n(t) \tag{3.1a}$$

$$\frac{\partial p_{i,j}}{\partial t} = p_{i,j}(1 - p_{i,j})\frac{\partial r_{g,i}}{\partial p_{i,j}} + m\frac{\overline{n}}{n_i}[\frac{\overline{np}_j}{\overline{n}} - p_{i,j}] + \lambda_p(t) \tag{3.1b}$$

Here $r_{g,i} = \beta_i\langle z\rangle_i + \theta_i$ is the genetic component of the growth rate (i.e., the log fitness) averaged over all genotypes on island i; it depends on allele frequencies via the trait mean $\langle z\rangle_i = \sum_j \gamma_j p_{i,j}$ on island i. Note that the dynamics of any one deme are coupled to the dynamics of all other demes via the mean number of immigrant individuals $m\overline{n}$ and the mean number of immigrant alleles $m\overline{np_j}$ (at locus j) per unit time (where $\overline{n}$ is the population size and $\overline{np_j}$ the number of allele copies per deme, averaged across all demes in the metapopulation).

Equation (3.1a) describes how population size evolves over time on an island in a given habitat. The first term within the square brackets describes logistic growth, and the second the coupling between size and allele frequencies (via mean fitness): growth rates are reduced relative to the maximum r_0 due to environment-dependent selection on the additive trait. The second term describes the effects of migration, which makes a net positive contribution when the size of the focal deme is less than the average $\overline{n}$ across the metapopulation. The third term $\lambda_n(t)$ is an uncorrelated random process with $\mathbb{E}[\lambda_n] = 0$ and $\mathbb{E}[\lambda_n(t)\lambda_n(t')] = n(t)\delta(t - t')$, where $\mathbb{E}[...]$ denotes an average over independent realizations of the process. This 'noise' term describes fluctuations of population size due to the stochasticity inherent in reproduction and death. In the present model, where the number of offspring is Poisson-distributed, the variance of population sizes is $n(t)$.

Equation (3.1b) describes allele frequency dynamics at locus j: the first term corresponds to the change due to selection on the locally favored allele, while the second term describes the effect of migration, which tends to pull allele frequencies closer to the average amongst migrants (given by $\overline{np}_j/\overline{n}$). Note that islands with larger populations contribute more to the average allele frequency in the migrant pool. Moreover, locally favored alleles are less prone to swamping in larger populations, since the migration term in Equation (3.1b) is proportional to $\overline{n}/n_i$ (a second source of coupling between allele frequency and population size). This results in a positive feedback: better adapted islands are more populous, send out more migrants and are less affected by incoming, maladapted individuals, and thus maintain local adaptation more easily (Haldane, 1956). Fluctuations

about the expected allele frequency are described by $\lambda_p(t)$, which satisfies $\mathbb{E}[\lambda_p] = 0$ and $\mathbb{E}[\lambda_p(t)\lambda_p(t')] = [(p(t)q(t))/n(t)]\delta(t-t')$, as in the haploid Wright-Fisher model.

We restrict our attention to scenarios with just two alternative habitats ($g = 2$), with relative frequencies $\rho_1 = 1-\rho$ and $\rho_2 = \rho$, where $\rho < 1/2$. Thus, ρ denotes the frequency of the rare habitat (denoted by the index 2 hereafter). We assume that directional selection pushes traits towards the two extremes of the phenotypic range in the two habitats: thus, a genotype with all '1' alleles or all '0' alleles has maximum fitness in the first or second habitat respectively. For simplicity, the maximum possible genetic fitness is assumed to be the same in both habitats. Then the population-averaged log fitness r_g in the two habitats is $-\sum_{j=1}^{L} s_{1,j}q_{1,j}$ and $-\sum_{j=1}^{L} s_{2,j}p_{2,j}$, where p indicates the frequency of the '1' allele at any locus, and $s_{1,j}$ (or $s_{2,j}$) the strength of selection against the locally deleterious allele at locus j in habitat 1 (or 2).

These simplifying assumptions allow us to consider a reduced set of parameters (see Key Notation) and focus on key qualitative behaviors. However, equation (3.1) is quite general and applies to more complicated scenarios involving multiple local habitats with heterogeneous patch qualities (e.g., unequal maximum growth rates and carrying capacities).

Equations (3.1a) and (3.1b) can be re-expressed in terms of *dimensionless* parameters, constructed by rescaling population size by the carrying capacity K, and all evolutionary rates by the baseline growth rate r_0. This results in the following re-scaled parameters (denoted by uppercase letters): $T = r_0 t$, $M = m/r_0$, $S = s/r_0$, $N = n/K$, and the new parameter $\zeta = r_0 K$, which represents the number of births per unit time at carrying capacity, and hence governs the magnitude of demographic fluctuations.

Diffusion approximation for the joint distribution of allele frequencies and population size.

We can construct an equation for the time evolution of the joint probability distribution $\Psi(N, p_1, \ldots, p_L, t)$ of allele frequencies $p_1, \ldots, p_L$ and (re-scaled) population size N on any island in the metapopulation using the *diffusion approximation*. This involves approximating the evolution of the distribution by a diffusion that depends only on the mean and variance of the change in N and p per unit time. For ease of notation, the vector $(N, p_1, \ldots p_L)$ is denoted by $\mathbf{x}$. We drop the index specifying the island, since the diffusion

equation has the same form on all islands; the dependence on the local habitat arises only through the population-averaged log fitness R_g, which differs between habitats. Then we have:

$$\begin{aligned}
\frac{\partial}{\partial t}\Psi(\mathbf{x},t) &= -\frac{\partial}{\partial N}\left[A_N(\mathbf{x})\Psi(\mathbf{x},t)\right] + \frac{1}{2\zeta}\frac{\partial^2}{\partial N^2}\left[B_N(\mathbf{x})\Psi(\mathbf{x},t)\right] \\
&\quad - \sum_{j=1}^{L}\frac{\partial}{\partial p_j}\left[A_{p_j}(\mathbf{x})\Psi(\mathbf{x},t)\right] + \frac{1}{2\zeta}\sum_{j=1}^{L}\frac{\partial^2}{\partial p_j^2}\left[B_{p_j}(\mathbf{x})\Psi(\mathbf{x},t)\right] \\
A_N(\mathbf{x}) &= [1 - N + R_g]N + M(\overline{N} - N) \\
A_{p_j}(\mathbf{x}) &= p_{i,j}(1-p_{i,j})\frac{\partial R_g}{\partial p_j} + M\frac{\overline{N}}{N}\left[\frac{\overline{Np_j}}{\overline{N}} - p_{i,j}\right] \\
B_N(\mathbf{x}) &= N \\
B_{p_j}(\mathbf{x}) &= \frac{p_j(1-p_j)}{N}
\end{aligned} \tag{3.2}$$

Here A_N and A_{p_j} specify the expected rate of change of the population size and allele frequencies (see also Equation (3.1)), and B_N and B_{p_j} the variance of the change per unit time (both expressed in terms of re-scaled parameters $S, M, \ldots$). Equation (3.2) involves no mixed derivatives with respect to N and p_j as the covariance of fluctuations of p and N is zero (to first order in $1/N$, s, etc.). The equations also involve the average log fitness R_g given by $-\sum_{j=1}^{L} S_{1,j}q_j$ and $-\sum_{j=1}^{L} S_{2,j}p_j$ in the first and second habitats respectively.

Equation (3.2) describes the stochastic evolution of the joint distribution of population size and allele frequencies and in principle, can be numerically integrated to obtain the expected value, variances and co-variances of N and $\{p_j\}$ through time. However, here we focus on the equilibrium distribution, conditioned on the average number of individuals $\overline{N}$ and allele copies $\overline{Np_j}$ per deme (averaged across demes):

$$\begin{aligned}
\Psi(N,\{p_j\}|\overline{N},\{\overline{Np_j}\}) &= \frac{1}{Z}\,N^{2\zeta M\overline{N}-1}\mathrm{e}^{-\zeta[(1-M)-N]^2} \\
&\quad \cdot \prod_{j=1}^{L}\left(p_j^{2\zeta M\overline{Np}_j-1}(1-p_j)^{2\zeta M(\overline{N}-\overline{Np}_j)-1}\mathrm{e}^{-2\zeta NS_j(1-p_j)}\right)
\end{aligned} \tag{3.3}$$

where Z is a normalization constant. Equation (3.3) specifies the stationary distribution in the first habitat; the distribution in the second (rare) habitat has the same form, except

with the last term replaced by $e^{-2\zeta N S_j p_j}$. Note that an explicit solution (Equation (3.3)) is possible only because the expected change $A_i(\mathbf{x})$ for each variable in Equation (3.2) can be expressed as $B_i(\mathbf{x})$ times the gradient of a (suitably-defined) potential $U(\mathbf{x})$, i.e., $A_i(\mathbf{x}) = B_i(\mathbf{x})[\partial U(\mathbf{x})/\partial x_i]$.

Numerical solution for the equilibrium. The state of each deme is determined by the average number of individuals $m\overline{N}$ and copies of each allele $m\overline{Np_j}$ (for the j^{th} locus) that immigrate per unit time. Given these, we can find the expected numbers and allele frequencies $\{\mathbb{E}_i[N], \mathbb{E}_i[Np_j]\}$ within habitat i, by integrating over the stationary distribution (Equation (3.3)). Here, the expectations $\mathbb{E}[\ ]$ can be thought of as averages for a *given* island, obtained by either averaging over replicate metapopulations (or simulations) or by averaging over measurements at uncorrelated time points at equilibrium within a single simulation. The crucial point is that at equilibrium, the average across all demes in the metapopulation at any instant (denoted by $\overline{\ }$) must be equal to the weighted sum of the expected values across habitats (Rouhani & Barton, 1993; Barton & Rouhani, 1993). Thus:

$$\sum_i \rho_i \mathbb{E}_i[N] = \overline{N}, \qquad \sum_i \rho_i \mathbb{E}_i[Np_j] = \overline{Np_j} \tag{3.4}$$

Equilibria are located by starting at an arbitrary $\left\{\overline{N}, \overline{Np_j}\right\}$, calculating $\{\mathbb{E}_i[N], \mathbb{E}_i[Np_j]\}$ using Equation (3.3), then computing the new $\left\{\overline{N}, \overline{Np_j}\right\}$ using Equation (3.4), and iterating until a fixed point. With this procedure, either a polymorphism is found, or one or other allele is fixed. In principle, this procedure simultaneously yields the equilibrium population size and allele frequencies at all the L loci (which may have different effect sizes and hence attain different frequencies). However, iterating over an $L+1$ dimensional space is computationally intensive. We thus restrict our attention to the case where effect sizes are equal at all loci (in a given habitat), such that $S_{1,j} = S_1$ and $S_{2,j} = S_2$ for all j. Then, we need to find only the fixed point $\left\{\overline{N}, \overline{Np}\right\}$.

The procedure outlined above is exact, *given* the diffusion approximation which, however, relies on three basic assumptions. First, we assume all processes to be sufficiently slow ($r_0, m, s \ll 1$) that a continuous time approximation (Equation (3.1)) is valid. Second, we assume infinitely many demes, such that population size and allele frequency averaged across all demes exhibit negligible fluctuations, even though within any one deme, they follow a distribution (Equation (3.3)). This allows us to treat the migrant

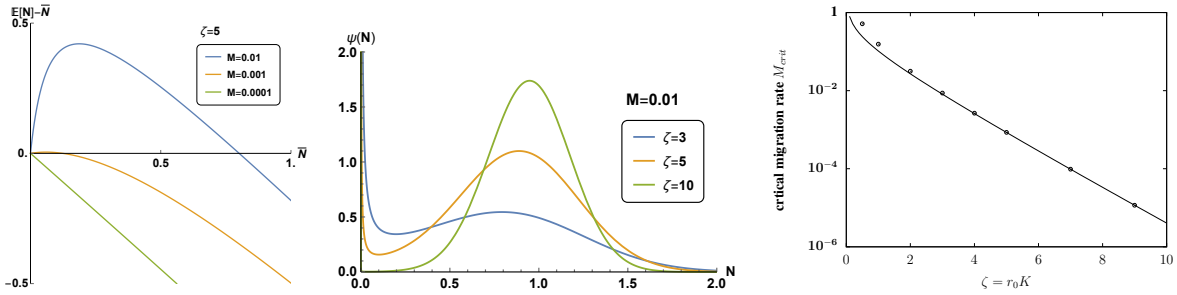
Table 3.1: Key Notation

g	number of distinct habitats in the metapopulation
ρ_α	fraction of islands in habitat α; we focus on two habitats ($g = 2$), frequencies $1 - \rho$ and ρ.
$r_{0,i}$	baseline growth rate on island i; $r_{0,i} = r_0$ for all i.
K_i	carrying capacity of island i; $K_i = K$ for all i.
L	number of loci influencing the trait
$s_{i,j}$	selection coefficient associated with locus j on island i
m	fraction of individuals that migrate from each island
n_i	population size on island i
$p_{i,j}$	frequency (of the '1' allele) at locus j on island i; $q_{i,j} = 1 - p_{i,j}$
$r_{g,i}$	genetic component of growth rate (i.e., log fitness) averaged over all genotypes on island i; $r_{g,i}$ is given by $-\sum_{j=1}^{L} s_{i,j}q_{i,j}$ and $-\sum_{j=1}^{L} s_{i,j}p_{i,j}$ for islands belonging to first and second habitats respectively
$\overline{n}$, $\overline{np_j}$	average population size per deme and average number of '1' alleles per deme at locus j, averaged across all demes contributing to the migrant pool
$\mathbb{E}_i(n)$, $\mathbb{E}_i(np_j)$	expected population size and expected number of '1' alleles at locus j for a deme in habitat i, obtained by integrating over the equilibrium joint distribution for population size and allele frequencies
Scaled parameters	
$N = n/K$	population size scaled by carrying capacity
$S = s/r_0$, $M = m/r_0$	selection coefficient, migration rate scaled by intrinsic growth rate
$\zeta = r_0 K$	average number of births per unit time at carrying capacity; scales inversely with demographic fluctuations.

pool as deterministic, and completely characterized by $\overline{N}$ and $\overline{Np_j}$. Finally, and most critically, we assume that allele frequency evolution at individual loci is unaffected by other loci (except via their joint effects on population size). More specifically, Equation (3.1) (which forms the basis of Equations (3.2) and (3.3)) neglects LD within any deme as well as LD across the whole metapopulation (by neglecting LD within the migrant pool). This final assumption is justified when selection per locus is much weaker than the rate of recombination between loci. We investigate the sensitivity of our results to each of these assumptions using two types of discrete-generation individual-based simulations (SI, section C).

Since the full model involves several parameters, and calculating the joint distribution requires a numerical solution for $\overline{N}$ and $\overline{Np}$, it is useful to consider various limits which allow for simpler approximations. We first consider population dynamics in the absence of selection ($S_1 = S_2 = 0$), and examine how demographic stochasticity and migration affect metapopulation survival. We then introduce selection, assuming it is weak relative to the baseline growth rate, i.e., $Ls_i \ll r_0$ or $LS_i \ll 1$ (and also neglect demographic stochasticity). In this 'soft selection' limit, population dynamics are largely unaffected by local adaptation, and populations are close to carrying capacity in both well-adapted and maladapted demes. We show when individual loci can adapt locally, despite gene flow and drift, and derive explicit expressions for the critical migration rate below which local adaptation is possible, by assuming that loci are close to fixation for one or other allele.

We then consider scenarios where selection across all loci is strong enough to affect population dynamics, i.e., $LS_1, LS_2 \sim 1$ (for equal-effect loci) resulting in 'hard selection', wherein maladaptation leads to extinction. We examine the hard selection model using the numerical solution for the equilibrium distribution of $(N, p_1, \ldots p_L)$, as well as a simpler 'semi-deterministic' approximation which is valid where the population size can be treated as depending deterministically on the expected allele frequencies. In the main paper, we focus on the case where selection per favorable allele is the same in both habitats ($S_1 = S_2 = S$). The more general scenario with $S_1 \neq S_2$ is considered briefly in SI, section B.



(a) (b) (c)

Figure 3.1: Population size distribution, expected population size and critical migration thresholds in the absence of selection. (A) $\mathbb{E}(N) - \overline{N}$ versus $\overline{N}$ for $\zeta = r_0 K = 5$, where $\mathbb{E}(N)$ is the expected population size in a deme, given the average $\overline{N}$ in the metapopulation. The equilibrium solution $\mathbb{E}(N) = \overline{N}$ can be found graphically as the point at which the curve intersects the horizontal zero axis. Stable equilibria are those for which $\mathbb{E}(N) - \overline{N}$ has a negative slope at $\mathbb{E}(N) - \overline{N} = 0$. The metapopulation survives ($\overline{N} > 0$) only above a critical migration rate $M_c = m_c/r_0 \sim 0.00085$ (for $\zeta = 5$). (B) Probability density $\psi[N]$ of the scaled population size $N = n/K$ for various ζ and $M = 0.01$, as obtained from Equation (3.5). (C) The migration threshold M_c (below which the entire metapopulation goes extinct) vs. ζ. Points show exact results and the solid line depicts the large ζ approximation; M_c declines exponentially with increasing ζ.

3.3 Results

Effect of demographic stochasticity and migration in the absence of selection.

Consider a scenario with no selection, such that population sizes are independent of allele frequencies, and only affected by demographic fluctuations and migration. Although individual demes fluctuate, $\overline{N}$ across the entire metapopulation can be treated as deterministic. Demes are coupled through this single variable $\overline{N}$, which determines the expected number of immigrants per deme. From Equation (3.3), it follows that in the absence of selection, the distribution of population size (on any island), conditioned on $\overline{N}$ is:

$$\psi[N|\overline{N}] = \frac{1}{Z_0} N^{2\zeta M\overline{N}-1} \mathrm{e}^{-\zeta[N-(1-M)]^2} \quad \text{where } Z_0 = \int_0^\infty e^{-\zeta[(N-(1-M))]^2} N^{2\zeta M\overline{N}-1}\, dN \tag{3.5}$$

The expected population size, $\mathbb{E}(N)$ can be obtained by integrating, and then equating $\mathbb{E}(N) = \overline{N}$ (Figure 3.1(a)). This yields one or more equilibria for $\overline{N}$, which can be

substituted into (3.5) to obtain the full $\psi[N]$ (as shown in Figure 3.1(b)).

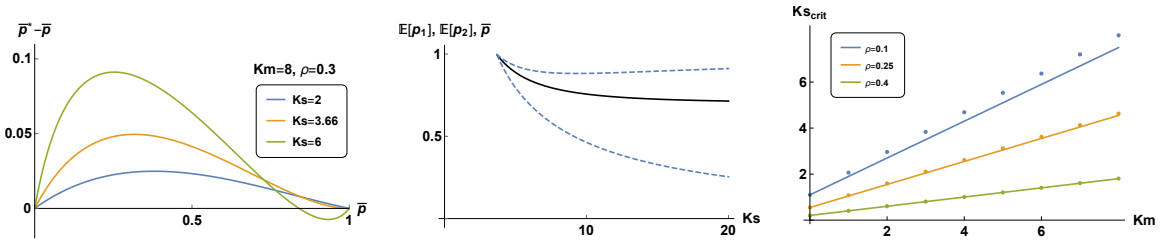
There is always an equilibrium at extinction ($\mathbb{E}(N) = \overline{N} = 0$). Above a critical migration rate M_c, there may also be an equilibrium with $\overline{N} > 0$ (Figure 3.1(a)). Figure 3.1(c) shows that as $\zeta = r_0 K$ increases, the critical rate of migration required to prevent global extinction, decreases exponentially: $M_c \approx e^{-\zeta}/(2\sqrt{\pi\zeta})$. This expression for M_c follows from the fact that it is the migration rate at which the equilibrium $\overline{N} = 0$ becomes unstable (SI, section A).

Above this critical migration threshold, the distribution is bimodal if the number of immigrants is small, i.e., for $2\zeta M\overline{N} = m\overline{n} < 1$, which corresponds to $M \lesssim \frac{1}{2\zeta}\left(1 + \frac{1}{2\zeta}\right)$ (SI, section A). In this case, some populations cluster in a Gaussian distribution around $N = 1 - M$ (i.e., $n = K(1 - m/r_0)$), with variance $1/2\zeta = 1/(2r_0K)$, whilst others are near extinction (Figure 3.1(b)). The parameter ζ thus governs the extent of demographic stochasticity: both the variance of the stationary distribution around carrying capacity and the risk of stochastic extinction fall with increasing ζ. In most of the paper, we will consider growth rates and carrying capacities that are sufficiently high (i.e., $\zeta = r_0 K \gg 1$) that well-adapted populations exhibit essentially deterministic dynamics and are not prone to stochastic extinction.

3.3.1 Soft selection

We now introduce selection, but assume that the evolutionary change it effects is slow compared to population growth (i.e., $\sum_j S_j \ll 1$) and that demographic stochasticity can be neglected ($\zeta = r_0 K \gg 1$). Then the model reduces to the classical infinite island model (Wright, 1932) with *fixed* size $n = K$ on each island and soft selection within demes. Unlike in the general model with hard selection, allele frequencies at different loci evolve independently under soft selection (assuming LE), since genetic drift at any locus just depends on (a fixed) population size, and not on adaptation at other loci. Thus, we need only consider the allele frequency distribution $\psi[p]$ at one locus. This distribution was first derived by Wright (1932), and also emerges from the joint distribution in Equation (3.3) by integrating over N (SI, section A).

The expected allele frequency $\mathbb{E}(p\,|\overline{p}\,)$ in a deme, given the mean $\overline{p}$ in the migrant pool, is obtained by integrating over $\psi[p]$. Allele frequencies in different demes are coupled via



(a) (b) (c)

Figure 3.2: Local adaptation under soft selection. (A) The difference between the weighted average of the expected allele frequency across habitats and the mean in the migrant pool, $(\overline{p}^* - \overline{p})$ vs. $\overline{p}$. The equilibrium solution $\overline{p}^* - \overline{p} = 0$ is where the curve intersects the horizontal axis; equilibria are stable if $\overline{p}^* - \overline{p}$ has a negative derivative at this point. A polymorphic equilibrium $0 < \overline{p} < 1$, exists only above a critical selection strength $Ks_c = 3.66$ (orange). When selection is weaker than Ks_c, the allele favored in the commoner habitat fixes (i.e., the stable equilibrium corresponds to $\overline{p} = 1$), and local adaptation is swamped by gene flow. The frequency of the rare habitat is $\rho = 0.3$, the average number of migrants per generation is $Km = 8$; selection is symmetric: $s_1 = s_2 = s$. (B) Expected allele frequencies in the two habitats (dashed), and the overall mean across the whole metapopulation (solid), vs. Ks for $s > s_c$. (C) The critical selection strength Ks_c above which a polymorphic equilibrium with $0 < \overline{p} < 1$ can be maintained vs. the average number of migrants exchanged between demes Km for various fractions ρ of demes belonging to the rare habitat. Points are obtained by numerically solving for stable equilibria; lines show the approximation $Ks_c = \frac{1}{2}\log\left(\frac{1-\rho}{\rho}\right) + Km(1-2\rho)$.

the mean allele frequency, $\overline{p}$, among migrants: within any deme, migration pulls the expected allele frequency towards $\overline{p}$, whereas selection drives $\mathbb{E}_1[p]$ towards 1 (or $\mathbb{E}_2[p]$ towards 0). Since all demes have equal sizes, they contribute equally to the migrant pool. Defining $\overline{p}^* = \rho\mathbb{E}_1\left(p\,|\overline{p}\,, s\right) + (1-\rho)\mathbb{E}_2\left(p\,|\overline{p}\,, -s\right)$, the difference, $\overline{p}^* - \overline{p}$ must be zero when the metapopulation is at equilibrium. The stability of equilibria can be evaluated using the derivative $\partial\overline{p}^*/\partial\overline{p}$; stable equilibria correspond to $\partial\overline{p}^*/\partial\overline{p} < 1$.

Figure 3.2(a) shows $\overline{p}^* - \overline{p}$ versus $\overline{p}$ for different values of $s_1 = s_2 = s$. Note that there are always equilibria corresponding to $\overline{p} = 1$ or $\overline{p} = 0$ (i.e., when the whole metapopulation is fixed for one or other allele). A polymorphic equilibrium (with $0 < \overline{p} < 1$) can be maintained when selection is sufficiently strong (Figure 3.2(a)), such that alleles favored in the rare habitat can invade. As selection becomes stronger, the different habitats approach fixation for different alleles (Figure 3.2(b)). The critical selection strength s_c, above which a polymorphic equilibrium becomes possible, increases with migration (Figure 3.2(c)).

It is useful to first consider a purely deterministic analysis: this suggests that poly-

morphism can be maintained only above a critical selection strength $s_c = m(1-2\rho)$, where ρ denotes the frequency of the rare habitat (SI, section A). However, in general, we expect the deterministic analysis to break down close to the threshold s_c: allele frequency distributions must necessarily become bimodal as $s \to s_c$, since $Km\overline{p} \to 0$. Thus the deterministic prediction only provides a lower bound on the true s_c, as drift will further inflate the selection threshold.

In the opposite limit of low migration ($Km \to 0$), loci will be close to fixation for one or other allele. The rates of fixation towards and away from an allele with advantage s, which is at frequency $\overline{p}$ in the migrant pool, are in the ratio $\sim (\overline{p}/\overline{q})e^{2Ks}$, such that the expected frequency of the favored allele in the deme is $\overline{p}e^{2Ks}/(\overline{p}e^{2Ks}+\overline{q})$ (SI, section A). Thus, in the symmetric case, $s_1 = s_2 = s$, the metapopulation reaches an equilibrium at:

$$\overline{p}^* = (1-\rho)\frac{\overline{p}e^{2Ks}}{\overline{p}e^{2Ks}+\overline{q}} + \rho\frac{\overline{p}}{\overline{p}+\overline{q}e^{2Ks}} \tag{3.6}$$

A polymorphic equilibrium at $\overline{p} = \frac{(1-\rho)e^{2Ks}-\rho}{e^{2Ks}-1}$ becomes possible if $Ks > Ks_{\mathrm{c}}$ where $Ks_{\mathrm{c}} = \frac{1}{2}\log\left(\frac{1-\rho}{\rho}\right)$ (in the $Km \to 0$ limit). We find that s_{c} increases linearly with m, and approximate this by the corresponding deterministic prediction for linear increase: $Ks_{\mathrm{c}} \approx \frac{1}{2}\log\left(\frac{1-\rho}{\rho}\right) + Km(1-2\rho)$ (solid lines in Figure 3.2(c)). This is reasonably close to the exact results (points).

Thus, both habitats are simultaneously adapted only if $s > s_c$. For $s < s_c$, alleles that confer a selective advantage in the common habitat tend to fix across the entire metapopulation (stable equilibrium at $\overline{p} = 1$ in Figure 3.2(a)). Interestingly, this bias towards alleles favored in the common habitat persists even in the limit of very low migration, for which we would have expected allele frequency dynamics of different demes to decouple and be dominated by drift.

3.3.2 Hard selection

We now consider scenarios where mean fitness has a substantial effect on population size, such that maladapted populations go extinct with high probability. This is the case when net selection against maladapted phenotypes is comparable to the baseline growth rate, i.e., $Ls \gtrsim r_0$ or $LS \gtrsim 1$ (assuming L equal-effect loci). Note that this does not imply that selection at *individual* loci is strong. In fact, typical effect sizes may be small enough

(i.e., $\zeta S = Ks \leq 1$), that drift can significantly degrade adaptation at individual loci. However, the number of loci L affecting fitness is large, so that selection, in aggregate, is strong (i.e., $L > 1/S$). If local adaptation is to be possible even in large populations, then selection must be at least as strong as migration, i.e., $S > M(1-2\rho)$. Further, we focus on parameters for which stochastic extinction of well-adapted populations is extremely improbable (ζ large).

In the following, we use the joint distribution for population size and allele frequencies (Equation 3.3) to identify the conditions under which locally adapted, stable populations are maintained in both habitats. We first analyze one example in detail, and then explore parameter space by investigating how the critical migration (or selection) strength required for local adaptation depends on demographic stochasticity, the number of selected loci, and the frequencies of the two habitats.

Figure 3.3 shows how polygenic adaptation collapses within the rare habitat as migration increases above a critical value, in a scenario with weak coupling between population size and mean fitness, i.e., $LS < 1$ (Figures 3.3(a) and 3.3(b)) and in a strong coupling, i.e., $LS > 1$ scenario (Figures 3.3(c) and 3.3(d)). In both cases, alternative alleles are close to fixation in the two habitats for low migration. As M increases, the frequency of the locally favored allele (Figures 3.3(a) and 3.3(c)) and the expected population size N (Figures 3.3(b) and 3.3(d)) decline in both habitats as a result of migration load. At a critical migration rate, the rarer allele is lost, the population in the rare habitat crashes, and the overall $\overline{N}$ falls to a minimum. As M increases further, the population in the rare habitat starts increasing, signifying that the rare habitat is now a maladapted demographic sink. The emergence of source-sink dynamics at high M causes numbers in the common habitat to show a slight decline with M. This is outweighed by the faster increase in numbers in the rare habitat, resulting in a slight increase in $\overline{N}$, the average population size across the whole metapopulation, at large M. Note that there is another migration threshold below which the whole metapopulation collapses because colonization is too rare; however, this is not visible here, since this threshold is negligibly small ($\sim e^{-\zeta}/(2\sqrt{\pi\zeta})$) for large ζ (Figure 3.1(c)).

Figures 3.3(b) and 3.3(d) also depict how the distribution $\psi[N]$ of the (scaled) population size in the rare habitat changes across the threshold M_c (insets). When net selection against maladapted phenotypes is weak relative to the baseline growth rate r_0, popula-

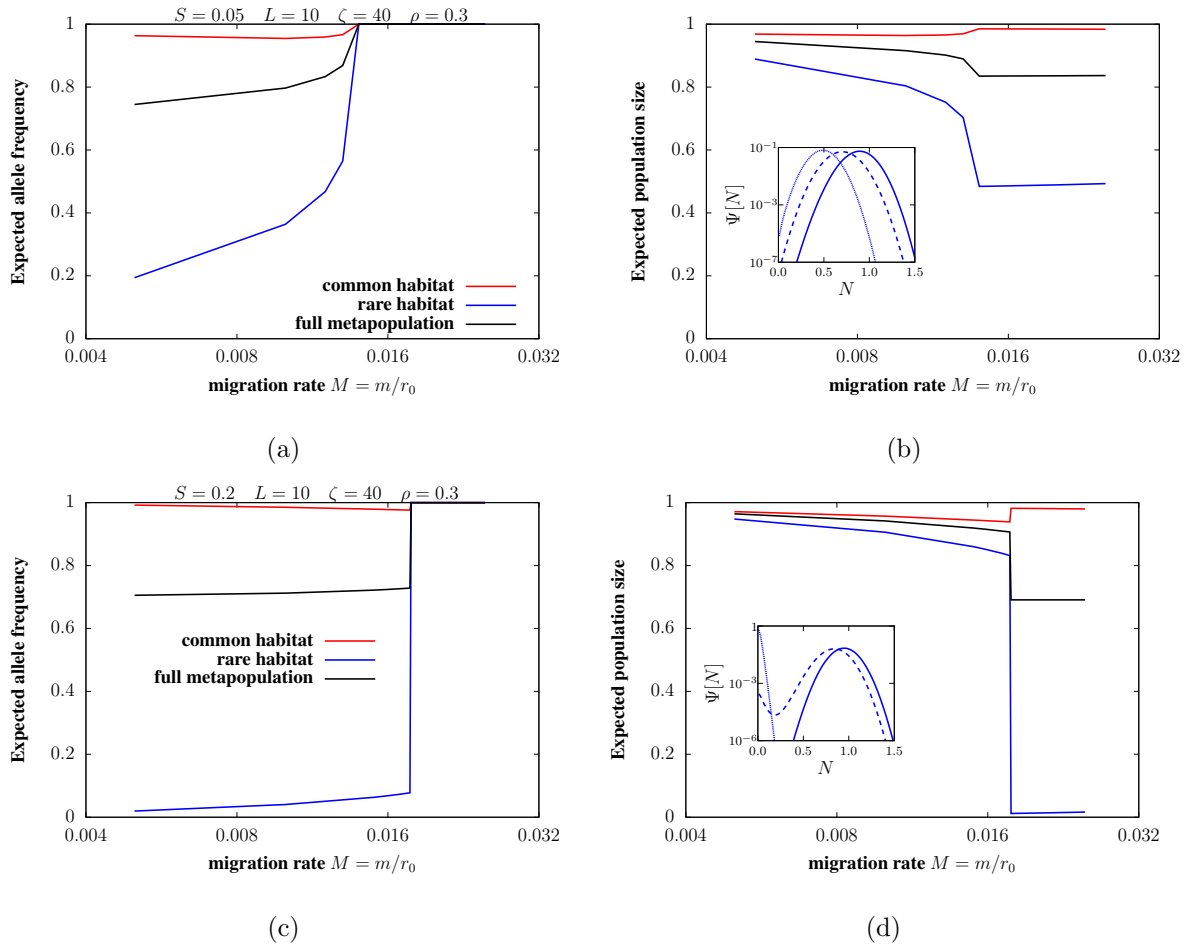


Figure 3.3: Loss of local adaptation at a critical migration rate under hard selection. Expected allele frequencies of the '1' allele (left panels) and expected population sizes (right panels) versus scaled migration rate $M = m/r_0$, for (A)-(B) weak coupling ($S = 0.05$, $LS = 0.5$) and (C)-(D) strong coupling ($S = 0.2$, $LS = 2$) between population size and mean fitness. The number of selected loci is $L = 10$ and selection is symmetric, with $S_1 = S_2 = S = s/r_0$ at each locus; the rare habitat comprises 30% of demes ($\rho = 0.3$) and demographic fluctuations are negligible ($\zeta = r_0 K = 40$). The plots show the expected allele frequencies and sizes in the rare and common habitat (blue, red) as well as the mean $\overline{p}$ and $\overline{N}$ across the whole metapopulation (black). For both weak coupling (i.e., $LS < 1$ in A,B) and strong (i.e., $LS > 1$ in C,D), there is a critical migration threshold M_c above which the allele favored in the rare habitat is lost from the metapopulation. The insets in (B) and (D) depict the probability distribution $\psi[N]$ of population sizes in the rare habitat (integrated over intervals of width $\Delta N = 0.02$) for $M < M_c$ (solid line), $M \sim M_c$ (dashed line), and $M > M_c$ (dotted line). For weak coupling, $\psi[N]$ peaks at some non-zero N, irrespective of M. For strong coupling, $\psi[N]$ peaks close to $N \sim 1$ for $M < M_c$; becomes bimodal, i.e., has peaks at $N = 0$ (corresponding to extinction) and $N \sim 1$ (corresponding to a well-adapted population) at $M \sim M_c$; peaks at $N = 0$ for $M > M_c$. All plots are obtained by numerically determining fixed points (Equations (3.3) and (3.4)) using the joint distribution $\Psi[N, p]$.

tions grow even when completely maladapted. In this weak coupling ($LS < 1$) regime, population sizes are approximately normally distributed about a non-zero expected value $\mathbb{E}[N]$ irrespective of local adaptation, i.e., for both $M < M_c$ and $M > M_c$ (inset, Figure 3.3(b)). Further, $\mathbb{E}[N] \sim 1 - LS\mathbb{E}[p]$, where $\mathbb{E}[p]$ is the expected allele frequency of the locally maladaptive allele in the habitat.

By contrast, for strong coupling (i.e., $LS > 1$), the distribution of sizes is approximately normal only when the population is locally adapted, i.e., for $M < M_c$. Close to the threshold for loss of adaptation ($M \sim M_c$), the distribution $\psi[N]$ becomes bimodal, with one peak at $N \approx 1$ and the other at $N = 0$, implying that a finite fraction of demes in the rare habitat is nearly extinct, while remaining demes support well-adapted populations. For $M > M_c$, the distribution is peaked at $N = 0$ (i.e., most demes are extinct) and decays exponentially with N. The threshold for loss of local adaptation is sharper for larger LS— a finding that we clarify below.

Semi-deterministic approximation. The fact that population sizes are approximately normally distributed about $\mathbb{E}(N)$ for $LS \leq 1$, suggests that in this 'weak coupling' regime, a simpler approximation, in which population size is assumed to depend *deterministically* on the expected log fitness $\mathbb{E}[R_g]$, may suffice. This *semi-deterministic* approximation (details in SI, section A) assumes that fluctuations in average log fitness R_g about the expected value $\mathbb{E}[R_g]$ are negligible. Since the magnitude of fluctuations, given by $\Delta R_g \sim S[LVar(p)]^{1/2}$, must fall with number of loci L (for fixed LS), we expect the semi-deterministic approximation to become more accurate when L is large, and selective effects correspondingly small. The approximation further assumes that ζ is so large that demographic fluctuations (in well-adapted populations) are negligible, but ζS small enough that drift has an appreciable effect at individual loci. As shown below, the semi-deterministic approximation accurately predicts the threshold for loss of local adaptation if $LS \leq 1$ (such that the distribution of N is unimodal about the expected population size), $L \gg 1$ (fluctuations in mean population fitness are negligible), and $\zeta \gg 1$ (negligible demographic fluctuations).

In these regimes (i.e., when the semi-deterministic approximation is accurate), the outcome is governed by three parameters. For a given rare habitat frequency ρ, and assuming symmetric selection $S_1 = S_2 = S$, the three parameters are: $\zeta S = Ks$, which

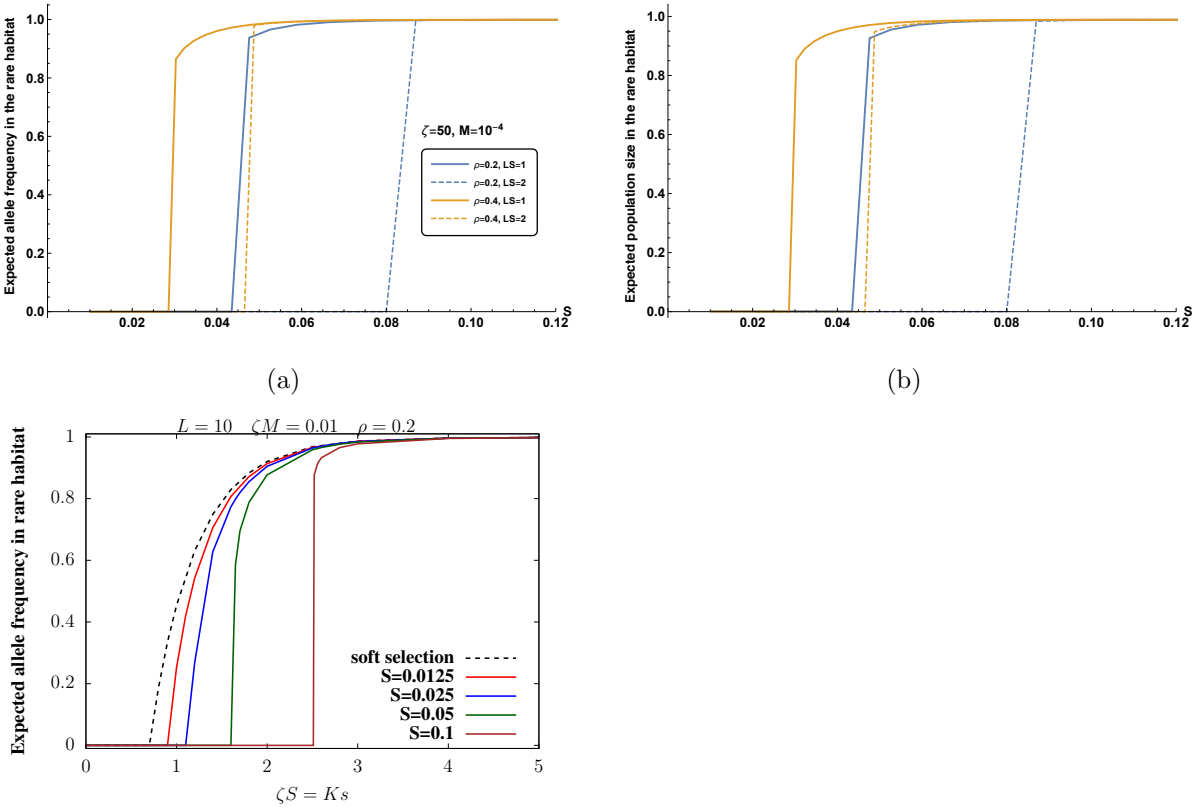


Figure 3.4: Local adaptation in the rare habitat, for weak migration. (A) Expected frequency of the locally favored allele and (B) expected population size in the rare habitat vs. selective effect per locus $S = s/r_0$ (for symmetric selection across the two habitats $S_1 = S_2 = S$) for $M = 0.0001$ and $\zeta = 50$. In each plot, the selective effect S and the number of loci L are changed simultaneously such that maximum possible genetic load LS is constant. Different colors correspond to different frequencies ρ of the rare habitat. Solid vs dashed lines correspond to different LS. Local adaptation in the rare habitat is lost (fig A) and populations go extinct (fig B) when selective architectures are highly polygenic with weak selective effect per locus (high L, low S). (C) Expected frequency of the locally favored allele in the rare habitat versus $\zeta S = Ks$ for different S, for $\zeta M = Km = 0.01$. The parameter ζS is varied by varying ζ (for a given S); ζM is held constant by varying M. The dashed line shows the corresponding prediction for allele frequencies under soft selection. For a fixed ζS, populations approach the soft selection prediction as S decreases and ζ increases. All solid-line plots are obtained by determining fixed points numerically (Equations (3.3) and (3.4)) using the joint distribution $\Psi[N, p]$; dashed line in (c) is obtained by determining exact fixed points under soft selection (SI, section A).

governs the strength of drift relative to selection in a population at carrying capacity, $\zeta M = Km$, which determines the average number of migrants exchanged between demes at carrying capacity, and LS, which determines the extent to which population sizes are reduced below carrying capacity due to maladaptation. Below we clarify the roles of these parameters in the low migration limit $\zeta M \ll 1$, which is most conducive to local adaptation.

Low migration limit. When migration is rare, loci are close to fixation for one or other allele (within a deme). As with soft selection, this implies that the fixation rates of alternative alleles (at a given locus) on island i are in the ratio $\approx (\overline{Np}/\overline{N})e^{2\zeta S_i N_i} : 1 - (\overline{Np}/\overline{N})$, where $\zeta = r_0 K$, and S_i is the (rescaled) selective advantage of the locally favored allele at that locus on island i. Further, $\overline{Np}/\overline{N}$ is the frequency (within the migrant pool) of alleles favored on island i. A comparison of this heuristic (for fixation rates) under hard selection with the analogous approximation under soft selection (Equation (3.6)) highlights two important features of allele frequency evolution under hard selection.

First, the rate of fixation and hence the frequency of the favored allele at any locus depends on the degree of maladaptation at all other loci via the population size N_i. In particular, locally deleterious alleles at very many loci, at even modest frequencies, can have substantial effects (in aggregate) on mean fitness, thus reducing size. This further accentuates drift at individual loci, causing locally deleterious alleles to increase or even fix, further reducing population size, thus generating a positive feedback between loss of fitness and decline in numbers.

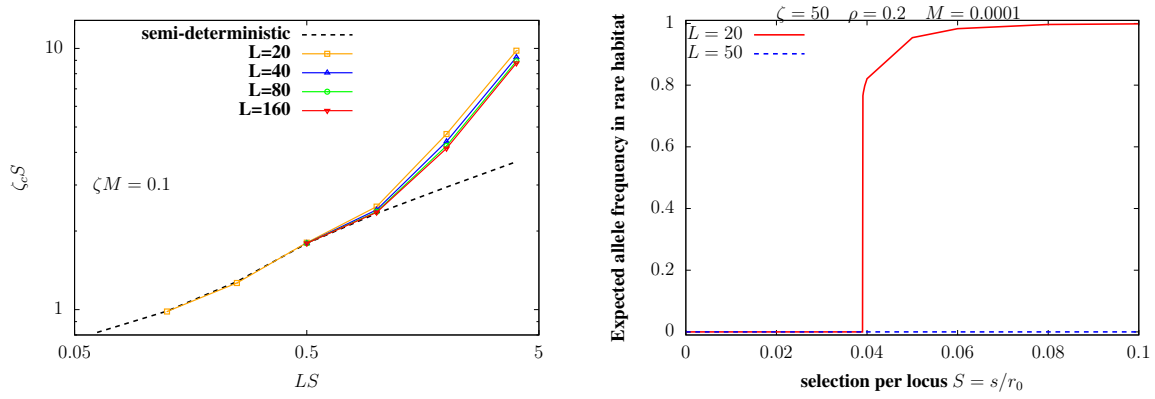
Second, any island contributes to the allele frequency $\overline{Np}/\overline{N}$ in the migrant pool in proportion to its size, which depends on the fitness of the island. Since locally adaptive alleles are at slightly lower frequency in the rare as opposed to the common habitat (even when both are locally adapted), the average population size is also somewhat lower in the rare habitat. Thus, an island belonging to the rare habitat contributes less to the allele frequency in the migrant pool than an island within the common habitat by a factor proportional to the ratio of population sizes of the two islands. This causes allele frequency in the migrant pool to shift further towards the optimal frequency in the common habitat, which increases migration load and reduces numbers in the rare habitat. This in turn reduces the contribution of the rarer habitat to the migrant pool further below

ρ, generating a second positive feedback loop. Crucially, both kinds of feedback depend on the strength of coupling between population size and mean fitness, and are thus stronger for larger $LS = L(s/r_0)$. Here, we focus on how these feedback affects influence local adaptation when selection is symmetric across habitats ($S_1 = S_2 = S$) and then briefly discuss asymmetric selection in SI, section B.

First, consider how local adaptation depends on the the selective effect per locus $S = s/r_0$ or alternatively, the number of selected loci L, for a *fixed* LS under very weak migration ($\zeta M = 0.005$ in Figures 3.4(a), 3.4(b)). These plots thus reveal how local adaptation is influenced by the genetic architecture of (i.e., the number and selective effects of loci contributing to) genetic load, for a given (maximum possible) total load LS in the population. Figure 3.4(a) shows that local adaptation in the rare habitat is possible only above a critical S_c per locus. For $S < S_c$, drift overpowers selection at individual loci, causing alleles favored in the common habitat to fix across the entire metapopulation despite very low genetic exchange, as with soft selection (Figure 3.2(c)). As a result, only very small maladapted populations are maintained in the rare habitat for $S < S_c$ (Figure 3.4(b)). Alternatively, given a certain (maximum) load LS, local adaptation is possible only if the selected trait is determined by a modest number of loci (i.e., for $L < L_c$, where $L_c = L(S/S_c)$), and fails for highly polygenic traits.

Further, local adaptation requires stronger selection per locus when the total cost of maladaptation, LS, is higher (solid vs. dashed lines in Figures 3.4(a) and 3.4(b)). In other words, the critical selection threshold S_c increases as the number of loci under divergent selection increases. In fact, for sufficiently large L, local adaptation is not possible for any S (see also Figure 3.5(b)). As we argue below, higher S increases the efficacy of selection at individual loci (via ζS), but also results in stronger coupling between allele frequencies and population size (via LS) which has the potential to degrade local adaptation due to the two kinds of feedback described above.

From these arguments, it follows that as one considers models of hard selection with lower and lower LS (corresponding to weaker and weaker coupling between population size and genetic fitness) while keeping ζS and ζM fixed, the behaviour of the metapopulation should converge to the predictions under soft selection with $Ks = \zeta S$ and $Km = \zeta M$. This is indeed what we see (Figure 3.4(c)): the frequency of the locally favored allele increases towards the soft selection prediction on approaching lower S and consequently



(a) (b)

Figure 3.5: Exact model vs. semi-deterministic approximation in the weak migration limit. (A) The critical threshold $\zeta_c S$ for local adaptation in the rare habitat versus LS for different values of L (depicted by different symbols), for $\zeta M = 0.1$ and $\rho = 0.2$. For given L, we vary LS by changing S, and then compute the critical ζ_c for each S. The migration rate M is always varied along with ζ such that ζM is constant at 0.1. The symbols and solid lines represent predictions of the full model (obtained using Equations (3.3) and (3.4)), while the dashed line represents predictions of the semi-deterministic approximation which treats population size as being determined by the expected allele frequencies. There is good quantitative agreement between the full model and the semi-deterministic approximation for $LS \lesssim 1$, but not for larger LS. (B) Expected frequency of the locally favored allele in the rare habitat vs. S for $L = 20$ (red) and $L = 50$ (blue) for $\rho = 0.2$, $\zeta = 50$, $M = 0.0001$. Local adaptation in the rare habitat is not possible for any S for $L = 50$.

LS, as long as ζS is held constant (by simultaneously increasing ζ). Note however that weaker selection S by itself (for fixed ζ and L) would make local adaptation more difficult, as the increase in genetic drift (relative to selection) at low ζS damages local adaptation—an effect not compensated by the more modest increase in numbers.

Finally, note that S_c is lower for larger ρ (orange vs. blue plots in Figure 3.4(a)), i.e., if the rare habitat encompasses a larger fraction of demes. In this case, the rare habitat is subject to a lower migration load (since allele frequencies in the migrant pool tend to be more intermediate), resulting in a weaker reduction in population size as well as weaker swamping at individual loci.

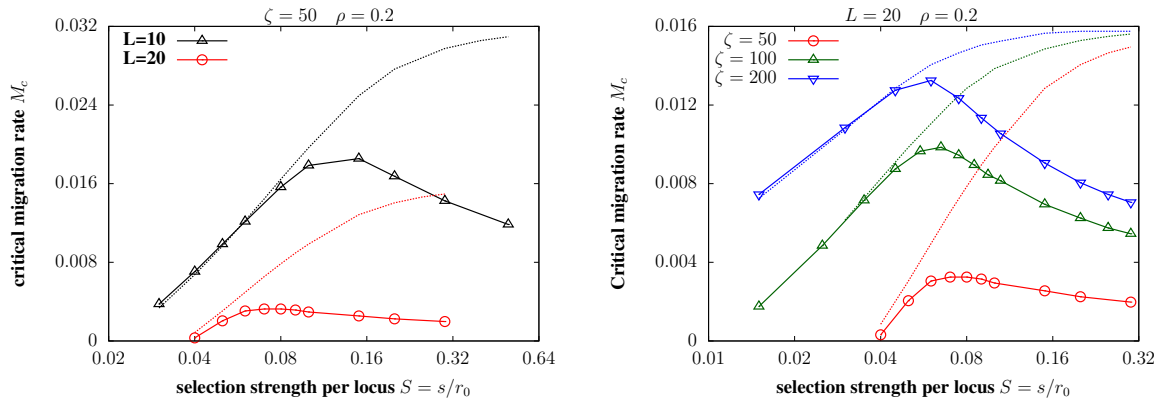
We now ask: for a given ζM, (when) does local adaptation in the rare habitat depend only on the two composite parameters ζS (which determines the strength of drift relative to selection per locus at carrying capacity) and LS (which governs how population sizes change due to fixation of deleterious alleles)? Recall that these parameters completely

determine the state of the metapopulation in the semi-deterministic ($LS \leq 1$) regime, in which sizes depend deterministically on the *expected* mean fitness. To investigate this, we determine the threshold ζ_c (such that local adaptation occurs for $\zeta > \zeta_c$), as a function of S, for various L, for fixed ζM. Here, ζM is held constant by reducing the rescaled migration rate $M = m/r_0$ as $\zeta = r_0 K$ increases, such that the average number of migrants (between demes at carrying capacity) remains unchanged.

Figure 3.5(a) shows that the semi-deterministic prediction for $\zeta_c S$ (dashed line) is extremely accurate for $LS \lesssim 1$: in this regime, the threshold $\zeta_c S$ for local adaptation in the rare habitat is independent of the number of selected loci, for a given LS. Moreover, this threshold only increases sub-linearly with LS for $LS \lesssim 1$. By contrast, for $LS \gtrsim 1$, the semi-deterministic approximation fails: the critical $\zeta_c S$ threshold increases much faster (nearly linearly) with LS, than predicted by the semi-deterministic approximation. However, even in this regime, the threshold for adaptation $\zeta_c S$ depends weakly on the number of loci, and is essentially governed by LS.

Increasingly stringent thresholds for local adaptation at large LS imply that when the number of loci underlying local adaptation is sufficiently high, simultaneous local adaptation across the two habitats is not possible for *any* selection strength (even for moderate ζ and very small M). For instance, with $L = 50$ loci, local adaptation cannot be maintained in the rare habitat, even for ζS as large as 10, for $\zeta = 50$ and $M = 10^{-4}$ (blue line in Figure 3.5(b)). Further, for $LS \gtrsim 1$, the whole metapopulation may converge towards the alternative fixed point $\overline{p} = 1$ (i.e., fix the allele favored in the rare habitat), if the initial allele frequency is close to $\overline{p} = 1$ and migration sufficiently high (results not shown).

Loss of local adaptation in the rare habitat: critical migration rates. We now consider cases where selection is strong relative to drift (so that both habitats are locally adapted under low genetic exchange), and ask: how high can migration be while still allowing local adaptation in the rare habitat and polymorphism at the level of the metapopulation? Figure 3.6(a) shows M_c, the critical migration rate above which polymorphism collapses, as a function of S for the case with $S_1 = S_2 = S$ for two different L. The points represent results of the full model (based on the joint distribution of N and p), while dashed lines represent the predictions of the semi-deterministic approximation.



(a) (b)

Figure 3.6: Critical migration rates for loss of local adaptation in the rare habitat. (A) Critical migration rate $M_c = m_c/r_0$ versus selection $S = s/r_0$ per locus for $L = 10$ (triangles) and $L = 20$ (circles), assuming $S_1 = S_2 = S$, $\zeta = 50$ and $\rho = 0.2$. (B) M_c versus S for different values of ζ for $L = 20$. Symbols depict M_c obtained via fixed point estimation from the joint distribution of population size and allele frequencies; dashed lines represent the predictions of the semi-deterministic approximation. M_c falls with S for large S, which is not captured by the semi-deterministic approximation. For any S, M_c increases with increasing ζ (in (B)), and appears to be approaching the semi-deterministic prediction.

Note that for both values of L, the critical migration rate increases with S when S is small, but then starts declining as selection increases beyond a certain threshold (which corresponds approximately to $LS \sim 1$). Thus, the range of migration rates allowing local adaptation in the rare habitat is widest (i.e., M_c largest), for intermediate selection. As in Figure 3.5(a), the semi-deterministic approximation is accurate for small S, but fails to (even qualitatively) predict large S behaviour.

The semi-deterministic approximation (dashed lines) predicts that M_c should saturate to a constant, independent of S, for large S. Such saturation is also predicted by a deterministic analysis (which should be accurate as $\zeta \to \infty$, when drift is negligible). This is most easily demonstrated in the $\rho \to 0$ limit, when allele frequencies in the common habitat are unaffected by immigration, such that $N_2 = \overline{N} = 1$ and $N_2 p_2 = \overline{Np} = 1$. Then one can show (using Equation (3.1)) that the rare habitat sustains local adaptation only below a critical migration rate, $M_c \sim 1/(4L)$ (for $L \gg 1$ and large S). Such a selection-independent threshold M_c emerges (at large S) simply due to the constraint that under hard selection, a population is viable only while its total migration load is less than its intrinsic growth rate. Since genetic load per locus is *at least* m (and typically greater than m under hard selection), this sets a limit on the number of polymorphic loci that

can be maintained without extinguishing the population.

For moderately large ζ, the actual M_c (points in Figure 3.6(a)) is several times less than this deterministic threshold, and *declines* with increasing S. As before, this is due to stronger coupling between population size and mean fitness (at large LS), which accentuates both drift and swamping in the rare habitat. However, even for large LS, the critical migration rate is expected to approach the deterministic prediction with increasing ζ (where higher ζ corresponds to weaker stochastic fluctuations in both population size and allele frequencies). This general expectation is supported by Figure 3.6(b), which shows M_c versus S for different ζ. Note, however, that M_c converges very slowly towards the deterministic threshold with increasing ζ: for instance, in Figure 3.6(b), even for ζ as high as 200 (i.e., 200 births per generation in a well-adapted population), M_c is approximately half the corresponding deterministic threshold for $LS \geq 2$.

3.3.3 Individual-based simulations.

Our analysis is based on the diffusion framework, which involves three approximations (see also Model and Methods). First, we approximate discrete generation dynamics by the continuous time evolution of population size and allele frequencies. Second, we assume infinitely many demes such that the migrant pool is deterministic. Third, we neglect LD within any deme as well as across the whole metapopulation. The validity of each assumption is tested by comparing with individual-based simulations (SI, section C).

These comparisons reveal that the most drastic approximation is due to neglecting LD amongst migrants, which can be substantial when demes are close to fixation for alternative alleles at multiple loci. Then, immigrant genotypes entering a deme are either perfectly adapted (if they originate from the same habitat) or severely unfit (if they originate from the alternative habitat). Consequently, an immigrant allele experiences a first-generation disadvantage proportional to the aggregate effect of all alleles it is associated with; this selective disadvantage is halved in each subsequent generation (Robertson, 1961). Thus the *effective* immigration rate of an allele is lower than the raw migration rate m— an effect not captured by equation (3.1) or (3.2). As a result, the migration threshold for loss of local adaptation, as observed in simulations, is significantly higher than the M_c predicted by the diffusion approximation. This discrepancy between simulations and

the diffusion approximation becomes weaker on approaching smaller growth rates r_0 and larger carrying capacities K, while holding the scaled parameters ζ, S, M constant: this corresponds to making selection weaker (relative to recombination), resulting in weaker LD.

In SI, section C, we show how the effects of LD can be partially accounted for by assuming that immigrants are drawn from two distinct pools corresponding to the two habitats (rather than a single well-mixed pool), and then considering the effective number of immigrant individuals and alleles from each pool to be weighted by their first-generation selective disadvantage in the recipient deme. However, a complete analysis along these lines (c.f. Barton & Bengtsson, 1986) is not attempted here.

3.4 Discussion

Metapopulation models have attracted much interest as idealized settings for understanding how population structure affects neutral diversity (Wright, 1932; Slatkin, 1977; Whitlock & Barton, 1997). Another line of research concerns rapid adaptation and extinction in fragmented habitats, leading to eco-evolutionary models of metapopulations (Hanski & Gilpin, 1991). Patches may have heterogeneous carrying capacities and/or different micro-environments. Metapopulation models thus also help us understand local adaptation, survival and extinction in marginal or novel habitats (i.e., "evolutionary rescue"). Yet, they have only recently been investigated in this context.

An important limitation of existing metapopulation models is that very few incorporate realistic assumptions about genetics within an eco-evolutionary framework, or explicitly consider the coupled stochastic dynamics of population size and allele frequencies. Various models do include some form of stochasticity: stochastic colonization-extinction dynamics (Hanski & Mononen, 2011), genetic drift in structured populations (Whitlock & Barton, 1997), or demographic fluctuations in the absence of selection (Mangel & Tier, 1993; Lande et al., 2003; Black & McKane, 2012). However, none consider different sources of stochasticity together within a common framework.

Our modeling framework is based on a diffusion approximation for the joint evolution of population sizes and allele frequencies (Banglawala, 2010; Barton & Etheridge, 2018), which we extend here to a metapopulation with multiple ecological niches. It assumes

a polygenic architecture for local adaptation, and accounts for both genetic drift and demographic stochasticity. It predicts the full stationary distribution of population sizes and allele frequencies in different habitats, thus yielding the conditions under which local adaptation is maintained simultaneously across habitats under divergent selection. The approximations underlying our theoretical framework (including that the migrant pool is in LE) can be formally justified as $r_0 \to 0$. Thus, they may not be accurate in typical populations, where growth rates may be high. Similarly, fluctuations in population size are assumed to be only due to demographic stochasticity, and so may be greatly underestimated. Nevertheless, our modeling approach captures key processes involved in local adaptation, and the approximations presented here apply over a broader range. We aim at understanding the fundamental processes, rather than precise prediction.

We identify two qualitatively distinct reasons why local adaptation fails within a rare habitat. First, if selection on locally favored alleles is weak relative to drift, then alleles favored in the common habitat tend to fix across the metapopulation, even when migration is extremely rare (Figure 3.4). This drift-dominated regime is also predicted under soft selection, where we obtain an explicit expression for the critical selection threshold below which polymorphism is lost: this threshold depends on the relative proportions of the two habitats, and increases as the rare habitat becomes more marginal (i.e., ρ decreases; Figure 3.2(c)). Interestingly, the critical selection threshold remains non-zero, implying that the bias towards alleles favored in the common habitat persists (at low $S = s/r_0$), even as $M \to 0$. In practice, we expect the time scale over which swamping of the rare habitat (and loss of polymorphism) occurs to increase as $M \to 0$. Thus, under weak migration and selection, local adaptation in the rare habitat may be metastable and loss of polymorphism extremely slow: this is consistent with behaviour observed in individual-based simulations (results not shown).

Second, migration (beyond a critical rate) may swamp local adaptation even when selection per locus is strong relative to drift: this kind of migration threshold emerges quite generally even in single locus models under soft selection when loci are subject to gene flow from a differently adapted population (Haldane, 1956). In the present model (with hard selection on multiple loci), the total migration load sets a more severe constraint: it must be sufficiently low that the population can still grow. Since migration load scales with the number of loci under divergent selection, moderate maladaptation at very many loci

is sufficient to cause the population to crash. Declining population size further reduces the efficacy of selection at individual loci via increased drift, but also results in stronger swamping, causing a positive feedback that extinguishes populations in the rare habitat. This feedback sets an upper limit on the migration rate or alternatively (given M) on the number of loci that can be divergently selected across the two habitats, while still allowing local adaptation in both. In the deterministic limit (i.e., as $\zeta = Kr_0 \to \infty$, such that both genetic drift and demographic stochasticity can be neglected), this migration threshold becomes independent of selection for large $S = s/r_0$, and depends solely on the number of loci and the habitat frequencies (Figure 3.6). However, for moderate $\zeta S = Ks$, the critical migration threshold decreases with increasing $S = s/r_0$ for $LS \gtrsim 1$.

A key result is that hard selection renders local adaptation in the rare habitat more difficult. The extent to which selection is hard is governed by $LS = L(s/r_0)$. We identify two qualitatively distinct regimes demarcated by $LS \sim 1$. For $LS < 1$ (and $\zeta = Kr_0 \gg 1$), the effects of hard selection can be encapsulated by assuming that population size is reduced in proportion to the expected genetic load; the strength of drift and swamping at individual loci is then governed by this reduced size. In the $LS < 1$ regime, an increase in selection allows populations to better withstand maladaptive gene flow and drift (Figure 3.4). By contrast, for $LS > 1$, our semi-deterministic approximation that treats size as determined by expected allele frequencies fails: in this regime, the population size distribution is bimodal (Figure 3.3), such that there is a small probability of extinction, even when expected frequencies of locally adaptive alleles are high. This results in a somewhat paradoxical situation: the conditions for stable local adaptation become *more* restrictive as selection per locus increases in the $LS > 1$ regime (Figure 3.6), because maladaptation then affects population size more strongly. Our results show that hard selection and random drift can substantially increase the damage that gene flow may cause - as, for example, when farmed fish escape into wild populations (Glover et al., 2017).

In our model, polygenic adaptation that depends on many loci, L, is difficult because both migration load and the coupling between population size and allele frequencies increases with L (for fixed S). This is because we assume an extreme form of environmental heterogeneity, in which any allele has opposite effects on fitness in the two habitats. In an alternative model with stabilizing selection (towards habitat-specific optima) on a quan-

titative trait, the deviation of the trait mean from the optimum and the trait variance within any deme are expected to be independent of genetic architecture in the infinitesimal $L \to \infty$ limit– provided optima lie near the center of the phenotypic range. Then, migration load would be independent of L, and can be calculated even in the infinitesimal limit (Barton & Etheridge, 2018). This relaxes the constraint that migration load places on the number of polymorphisms that can be maintained under hard selection.

We focus here on the case where locally adapted populations are demographically stable. However, the joint distribution derived in Equation 3.3 can be used to explore alternative regimes. For instance, we might consider a metapopulation with many very small demes and frequent extinction (i.e., $\zeta = r_0 K \sim 1$). The whole metapopulation can still adapt (if migration is sufficiently high), even when selection within each deme is weaker than local drift. Indeed, Wright (1932) argued that such a 'shifting balance' allows efficient search across alternative adaptive peaks (see Rouhani & Barton, 1993; Coyne et al., 1997). However, it would not be possible for populations to adapt to local variations in environment between demes in this regime.

The framework presented here is quite general, and can be applied to metapopulations with multiple niches, or differences in patch quality. While we have focused on local adaptation, the framework can be applied to other questions in metapopulation biology. For example, we find that in the neutral case, F_{ST} depends only on the number of incoming migrants, $\zeta M\overline{N} = m\overline{n}$, regardless of the size of the focal deme (Equation 3.3). The model can also extend to include dominance, and so could be used to understand heterosis and inbreeding depression within a metapopulation, and their interaction with population dynamics.

Our main analysis neglects linkage disequilibrium, which arises when allele frequencies differ between habitats at multiple loci. Associations between locally adaptive alleles allow simultaneous local adaptation over a wider range of parameters than predicted by the diffusion (Figures S2C, S2D, SI), because sets of introgressing alleles from differently adapted populations are eliminated together, thus reducing the effective rate of gene flow (Barton & Bengtsson, 1986). In SI, section C, we outline a heuristic which could capture this effect. It may also be possible to estimate the extent of local adaptation, and the extent to which it reduces effective gene flow, by observing how divergence and LD vary along the genome (cf. Aeschbacher et al., 2017).

Local adaptation in a metapopulation may lead to parapatric speciation, despite gene flow: as populations diverge, selection against introgressing alleles increases, reducing effective migration, and allowing further divergence. A key issue here is whether a heterogeneous environment will lead to distinct clusters, separated by strong barriers to gene flow, which eventually become good biological species. This may depend on the distribution of available habitats. If these are broadly continuous, and select along multiple environmental dimensions, then there may be substantial local adaptation without clusters being apparent. However, with distinct environments, local adaptation may lead to strong isolation, as multiple divergent loci become coupled together (Barton & De Cara, 2009; Barton, 2010). The framework developed here may be used to investigate how the distribution of selective challenges influences whether populations evolve as generalists, adapting to a range of local environments, or split into distinct and well-isolated species.

4 Spatially explicit coalescent simulations with demographic histories

This work is a joint work with Barbora Trubenová and Katalin Csilléry. The chapter is based on our manuscript which is planed for submission in the early fall.

Author contributions: E.S. B.T. and K.C. designed the study, E.S. did the simulations, E.S. B.T. and K.C. worked on the analysis, E.S. B.T. and K.C. wrote the manuscript.

4.1 Introduction

The distribution and dynamics of genetic diversity within species are shaped by a myriad of evolutionary and ecological processes acting across different spatial and temporal scales (Ellegren & Galtier, 2016). Although the role of space and, in particular, spatial autocorrelation in allele frequencies has been recognized from the dawn of population genetics (Wright, 1943; Malécot, 1948; Felsenstein, 1976), disproportionately more theoretical and methodological developments focused on understanding the effect of temporal changes in population size and gene flow between spatially non explicit populations (e.g. Hey & Nielsen, 2007). Statistical methods have been developed to detect past population size changes either by testing different hypotheses such as exponential growth and bottleneck (e.g. Excoffier et al., 2013) or using Bayesian methods to detect arbitrary population size changes from whole genome sequencing data (e.g. Drummond et al., 2005). In contrast, the much younger field of landscape genetics, set the ambitious aim of integrating pop-

ulation genetics and spatial statistics. In their foundation paper, Manel et al. (2003) argued that landscape genetics can overcome the limitation of population genetic methods that rely on the assumption of non-spatial and discrete populations. The field was more influenced by the metapopulation models than population genetics. Hanski & Gilpin (1991) provided a mathematical framework for describing the distribution and movement of species in space. However, much of the landscape genetic literature uses spatial statistical and statistic principles and has no link to population genetic models (e.g. Guillot et al., 2005; Smouse et al., 2008; Forester et al., 2016), or in fact, relies on non-spatial population genetic null models, such as the Wright-Fisher or island models (Meirmans, 2012).

The Wright-Fisher model has been the most widely used null model in population genetics, proposed for studying the effect of genetic drift. Under this model, the level of genetic diversity is proportional to the population size (Hartl et al., 1997). The infinite island model adds a next step of complexity by relaxing the assumption of random mating. It considers infinitely many Wright-Fisher populations (or islands) that receive migrants at rate m chosen at random from the other subpopulations (Wright, 1931). A spatially explicit version of the island model is the stepping-stone model (Kimura, 1953). The two dimensional version assumes that populations are situated on a rectangular (or other) lattice and an exchange of migrants with the four neighboring demes. However, real populations are not arranged at regular distances, but in continuous space. This most realistic view is captured by Wright's isolation by distance model (Wright, 1943) (also referred to as Wright-Malécot model (Malécot, 1948)). In this model individuals mate with neighboring individuals within radius that is dependent on the dispersal rate (σ) and the population density (d). Several equivalences have been shown across these models assuming infinite or finite populations and in the absence or presence of mutations (Felsenstein, 1976; Malécot, 1975; Slatkin, 1985).

In spatially subdivided populations the genetic variance is also partitioned. From studying this variation one can infer demographic structure and gene flow. One of the first—and ever since most popular—statistical measure to describe genetic differentiation is the fixation index, F_{ST}. Originally it was defined as the correlation between genotypes chosen randomly from the same subpopulation relative to the entire population (Wright,

1949). The popularity of F_{ST} originates from the simple formula given by Wright:

$$F_{ST} = \frac{1}{1 + 4Nm}, \tag{4.1}$$

where Nm is the number of migrants per deme. This connection is only true when the assumptions of the island model hold, however it became a general tool to infer gene flow and estimate population differentiation.

Another way of thinking about the effect of spatial structure on genetic differentiation is to use idea of isolation by distance (IBD) presented by Wright (1943). He defined the genetic neighborhood (the number of inbreeding individuals in the local neighborhood) as $NS = 4\pi\sigma^2 D$, where σ is the mean squared parent-offspring dispersal distance along one axis in a two-dimensional habitat and D is the ideal population density. It was shown by Rousset (1997) that the rate of decay in genetic relatedness (e.g. in pairwise F_{ST}) can be used to infer the neighborhood size, NS. One way is to fit IBD models and use the approximately linear relationship between the logarithm of spatial distance and genetic distance between individuals (Shirk & Cushman, 2014). Meirmans (2012) showed however, that isolation by distance can lead to false positives when detecting hierarchical population structure and loci under selection. Battey et al. (2020) used spatially continuous individual based simulations to draw the attention to the fact that ignoring space can severely bias common population genetic summary statistics, especially, when the neighborhood size is small. Finally, in this context, Joseph et al. (2016) proposed a spatially explicit simulator connected with an Approximate Bayesian Computation algorithm to estimate the neighborhood size from genomic data.

Population genetics, a century old field of evolutionary biology by now, has always used rigorous mathematical tools to describe how evolutionary processes shape genetic diversity and to provide proper frameworks for inference methods. Genetic data became widely available and mathematical tools that can connect theory to empirical data are valuable. Such a method is the coalescent theory, which is the basis of much of the modern population genetics as it can also be used for the analysis of DNA sequence data. Furthermore, it can be simulated simulated efficiently. Kingman (1982a,b) showed that the coalescent process is the ancestral limiting process for many discrete population genetics models, including the Wright-Fisher or Moran models. The coalescent describes the history of n samples in a well-mixed population by identifying $(n-1)$ coalescent events

and the time at which they happened: this process results in a bifurcating tree. The relationship between the coalescent process and the diversity follows the idea presented by Kimura (1969), namely that under the infinite sites model, where all mutations are unique and neutral, the expected number of segregating sites in a sample is proportional to the expected branch length of the tree of that sample (Hey, 1991).

The standard model of coalescent assumes that the samples are exchangeable, which means that the samples are identically distributed but not independent. Relaxing this assumption leads to the notion of structured coalescent. This model corresponds to Wright's island model. Formally, this limit exists when $N \to \infty$ and Nm is finite. It is a continuous-time Markov process, where the lineages either coalesce or migrate with small probabilities, and the time is scaled by N.

Biological models dealing with spatial structure existed long before the structured coalescent process was mathematically formalized (Takahata, 1988; Notohara, 1990). Previous theoretical work includes that Wright (1950) discovered the fundamental importance of Nm determining the extent of population differentiation, Malécot (1951); Maruyama (1977) calculated probabilities of identity by descent in a subdivided population, and Slatkin (1987) and Strobeck (1987) worked out the expected numbers of pairwise differences between sequences. After the formal introduction of the structured coalescent, it was quickly extended with various features such as selfing or background selection (Nordborg, 1997).

Now that we introduced the coalescent process, we can turn our attention back to the fixation index, F_{ST}. It can be expressed in many different forms Nei (1972); Weir & Cockerham (1984), however, the modern population genetics approach is to use the definition involving coalescence times (Slatkin, 1991):

$$F_{ST} = \frac{T_T - T_W}{T_T}, \tag{4.2}$$

where T_T is the average total coalescence time and T_W is the average within coalescence time across the subpopulations. This is the form that we will mainly use in this chapter, alongside of mean coalescence times and a pairwise genetic measure (more details can be found in the Methods section), to describe diversity, and explore whether we can explain patterns of genetic diversity better by taking into account known population structure.

Although it appears necessary to account for space in future studies, spatial population genetic models are the least mathematically tractable and theoretical predictions are valid only under limited conditions (Slatkin, 1985; Barton et al., 2002; Bradburd & Ralph, 2019). This is particularly true for the continuous space model, which is also hardly tractable in a forward time simulation setting (Battey et al., 2020). The stepping stone model can provide a more pragmatic solution. Indeed, it can approximate well the decrease of genetic correlation with distance in continuous space, which is essentially the same phenomenon as isolation by distance (Malécot, 1955; Kimura & Weiss, 1964).

Two basic approaches exist to simulate the coalescent process: continuous time approximation and a generation-by-generation approach. The first approach was developed by Hudson (1983) and is implemented in the software *ms* (Hudson, 2002). Events occur at a rate that depends only on the state of the extant ancestors, and the time to coalescent events can be simulated without explicitly considering the generations. The second approach was proposed by Hudson et al. (1987) and first implemented in the software *Simcoal* (Excoffier et al., 2000; Laval & Excoffier, 2004) This methods is more flexible, because demographic events can be defined at a generation level, nevertheless, only one coalescent event is allowed per deme per generation. If a coalescent event happens, two lineages of the deme are chosen at random to coalesce. More recently, an efficient algorithm has been proposed to produce genealogies by (Kelleher et al., 2016), and implemented in the software *msprime*. Events occur at rates that depend on the state of ancestral lineages existing at a given time point: with these rates the waiting times until the next even occur are generated. This feature makes the algorithm much faster compared to other simulations that are considering each generation separately.

The first spatially explicit coalescent simulator under the stepping stone model was IBDSim (Leblois et al., 2004). However, it has not been used extensively and had a limited ability to define spatial and temporal heterogeneity with only three time steps allowed. The most used spatially explicit coalescent simulator is SPLATCHE, which was initially designed to study the impact of spatial and ecological information on molecular diversity (Ray et al., 2010; Currat et al., 2019). It consists in a forward demographic simulation of population demography and migration, followed by a backward coalescent simulation step. In the coalescent step, the ancestry of a sample of gene lineages taken from one or several populations is simulated until the most recent common ancestor of

these lineages. Then, genetic diversity of the sample is generated by adding mutations over the simulated coalescent tree. SPLATCHE can handle spatial and environmental complexity through the use of population carrying capacities (e.g. linked to available environmental parameters), migration rates (i.e. directional gene flow) and frictions (i.e. dispersal constraints in different environments) based on user-specified raster maps that can change over time.

In this work, we develop a spatially explicit coalescent simulation tool, *gridCoal*, using functions from *msprime* (Kelleher et al., 2016), in which demographic histories can be defined using a two-dimensional stepping stone model. *gridCoal* is faster than previous simulation tools because it uses the algorithm of *msprime*, and because it bypasses the simulation of genetic data. Simulating genetic data while considering a full spatio-temporal demography can be extremely time consuming (Battey et al., 2020; Ray et al., 2010; Currat et al., 2019), and may prohibit users from considering the effects of space. *gridCoal* uses a coalescence time approximations of spatial diversity and divergence, and thereby, could allow for a more widespread use of spatially explicit null models in population genetic inference.

We test *gridCoal* for various scenarios of spatial and temporal changes in population size. First, we consider variation in space, but constant population sizes in time, and compare these results with theoretical expectations of the island and stepping stone models. We also compare isolation by distance patterns under these scenarios using a measure of genetic distance derived from F_{ST}. Although F_{ST} may be used as a measure of gene flow only under the infinite island model and neutral alleles, or when $Nm > 1$ (Slatkin, 1985; Whitlock & McCauley, 1999). Nevertheless, genetic distance measures derived from F_{ST} (pairwise F_{ST}) have been extensively used in empirical studies (e.g. Kitada et al., 2020). Second, we investigate the pattern of gene diversity and genetic distances when the population sizes vary in space and time, and notably, under conditions when spatial and temporal autocorrelation are decoupled. Duforet-Frebourg & Slatkin (2016) developed theory for isolation by distance and time, so we compare our results to their findings.

4.2 Methods

4.2.1 Simulation tool: *gridCoal*

We developed a spatially explicit coalescent simulation tool, *gridCoal*, based on *msprime* (Kelleher et al., 2016). Populations are distributed across an arbitrary, finite, two dimensional rectangular grid. Each grid cell represents a single panmictic population of defined size. The demographic history is specified by providing multiple corresponding grids defining the deme sizes at equally spaced time points in the past. The time step can be chosen arbitrary.

The spatial structure is further defined by a forward migration matrix, and can be asymmetric (between neighboring cells) and heterogeneous. This migration matrix defines the fraction of individuals that migrate from one cell to another. Here we use and analyze a $2D$ stepping stone model, where each cell is connected to its four direct neighbors with which it shares an edge. The forward migration rate is constant across time and homogeneous across the grid. The backward migration matrix, that is required for the coalescent simulations by *msprime*, contains elements that specify the fraction of individuals in a given cell that have parent in another deme. This value is calculated for each timestep based on the population sizes and the forward migration matrix. For detailed description of the simulator inputs and methods, see Appendix B.1.

4.2.2 Simulated scenarios

Several sets of simulations were carried out throughout this study, initially to validate the results against theoretical predictions (which exist only under very specific assumptions), and investigate its limitations, and after that to study how the results change if certain assumptions, like the heterogeneity of the populations in space and time, are relaxed.

Static populations

The first set of simulations assumed individual demes with fixed population sizes over time. The simulations were run in a factorial design on a square grid of size 30×30, with average population size of N, with migration rate m between neighboring cells (see Table 4.1). We run 1000 simulations on each combination of the following parameters

described in this section. Different maps (spatial grids) were simulated, with increasing spatial variance between the individual deme sizes:

- *homogeneous* map with equal-sized demes of size N,
- *low spatial variance* map with deme sizes drawn from Poisson distribution with mean N,
- *high spatial variance* map with deme sizes drawn from uniform distribution with range from 0 to 2N,
- *clustered* map, with randomly generated clusters where neighboring deme sizes were correlated and an average deme size across the whole map was N (see Figure 4.1).

In all scenarios, individual deme sizes were fixed over time, with 5 identical lines describing demographic history and the time step being 8 million generations (200000000 years, 25 years per generation).

Variable	Symbol	Values
Map type		homogeneous, poisson, uniform, clusters
Average cell population size	N	10, 50, 100,250, 500
migration rate	m	0, 10^{-8}, 10^{-5}, 10^{-3},10^{-2},10^{-1}, 10^{0}

Table 4.1: Summary of parameter values used in the simulations of static populations

Effect of time until ancestral state and grid sizes. To address how the summary statistics are affected by the size of the grids, whether we observe any edge effects, or to assess how long we need to run the simulations to obtain useful information we carried out some further tests.

First, we used the earlier defined homogeneous map with population size 100 on a 30×30 grid with time steps of $2 \cdot 10^2, 2 \cdot 10^4, 2 \cdot 10^6, 2 \cdot 10^8$ years. (Generation time is assumed to be 25 years.) We used the same migration rates as before: $m = 0, 10^{-8}$, 10^{-5}, 10^{-3},10^{-2},10^{-1}, 10^{0}. Second, we tested homogeneous maps on a 10×10 and a 50×50 grids, taking 10 or 50 samples, respectively. This was done for three values of migration rates, $m = 10^{-5}$, 10^{-3}, and 10^{-1} and was running up to 8 million generations.

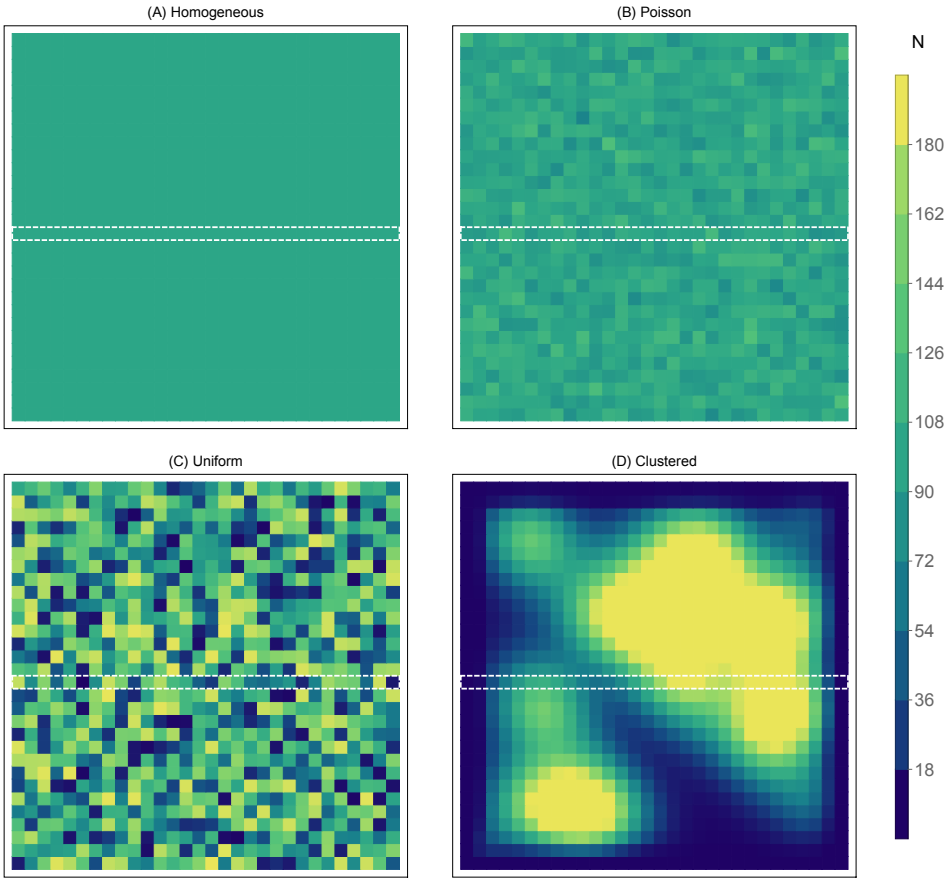


Figure 4.1: Simulated grids with no temporal changes. They express different amounts of spatial variation: (A) no variation (Homogeneous), (B) low (Poisson) and (C) high (Uniform) variation, and spatially correlated (Clustered). In white the sampled row: 30 grid cells, is marked.

Populations with simple demographic history

The second set of simulations was carried out on populations with variable demographic history. To allow comparison across different demographic histories, all scenarios assume exactly the same population structure with high spatial variance (individual deme sizes drawn from a uniform distribution $U_{final} = \mathcal{U}(0, 2N)$ in the final (most recent) time-point.

- *Linear expansion* was simulated for different maps by supplying T data points describing a population with linearly increasing overall size. Individual demes change linearly between their original sizes drawn from a uniform distribution $U_{startLinE} = \mathcal{U}(0, N)$ and their final sizes U_{final}. Note that while the whole population was expanding, some demes became smaller.

- *Exponential expansion* was simulated for different landscapes by supplying T data points describing a population with exponentially increasing overall size. Individ-

ual demes change exponentially between their original sizes drawn from a uniform distribution $U_{startExpE} = \mathcal{U}(0, N)$ and their final sizes U_{final}. Note that while the whole population was expanding, some demes became smaller.

- *Population decline* was simulated by supplying T data points describing linearly decreasing population. Individual demes change linearly between their original sizes drawn from a uniform distribution $U_{startDec} = \mathcal{U}(2N, 3N)$ and their final sizes U_{final}. Note that while overall population was declining, some demes became larger.

- *Bottleneck* was simulated by supplying T data points for a population linearly declining from $U_{startBot} = \mathcal{U}(0, 2N)$ to $U_{midBot} = \mathcal{U}(0, 0.4N)$ then expanding to U_{final}. Note that while the mean deme size changed from N to 0.2N to N again, individual demes may have experienced different demographic histories.

- *Stable* population was simulated by supplying T data points for a stable population (no change in overall population size), but with individual demes changing between their original sizes $U_{startStable} = \mathcal{U}(0, 2N)$ to U_{final}.

All the above described scenarios were simulated with 1000 replicates for $N = 100$, $T = 30$, with time steps of 2000 generations (50000 years, 25 years per generation), unless stated otherwise, and migration rates of $m \in (10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1})$. See Figure 4.2 for illustration of various demographic histories.

Variable	Symbol	Values
Map type		lin. and exp. increases, decline, bottleneck
Average cell population size	N	100
Migration rate	m	10^{-5}, 10^{-4}, 10^{-3},10^{-2}, 10^{-1}
Time of known demographic history	T	30

Table 4.2: Summary of parameter values used in the simulations of simple demographic histories

Furthermore, we run a set of simulations where instead of the uniform map we used a spatially autocorrelated one (clustered). We run similar demographic histories: no change in time, bottleneck, decline, or expansion (see Figure B.3 in Appendix). We used the same parameters of m, T as before. The time between two steps when the population sizes are defined. is 50000 years.

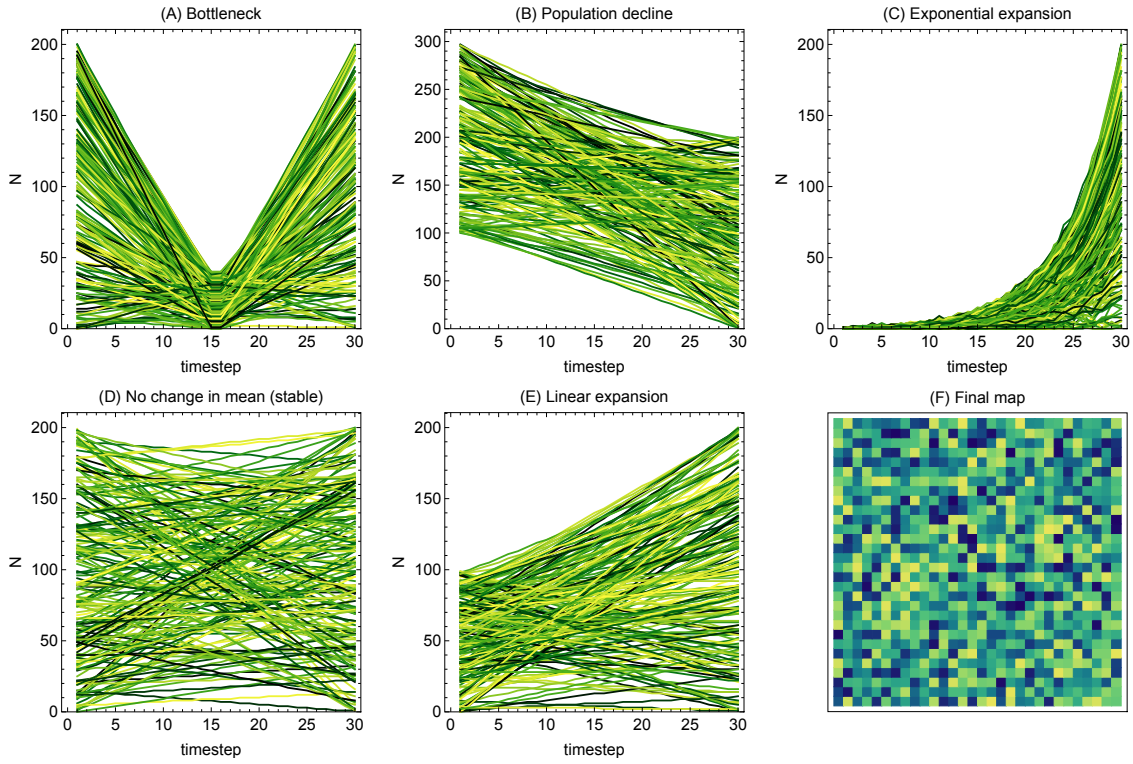


Figure 4.2: Simulated simple demographic histories (A-E) and their final spatial state (F). The different colors represent individual grid cells (200 displayed out of the 900). All histories converge to the exact same final map, which is sampled from a uniform distribution.

Biologically realistic populations

The third set of simulations was carried out on more complex populations, including different colonization scenarios, see Figure 4.3 and Table 4.3. We used a 30×30 grid over the course of 30 steps distributed 50000 years apart.

- **Front colonization (side colonization):** to simulate a population invading a new territory we simulated a population entering the grid from one side, populating one row of cells in each time-step.

- **Seed colonization (growing cluster)**: to simulate populations that colonized a new territory from a small number of immigrants, we populated the grid with a small number of 'seed' populations that grow in time, as well as spread from the seeds to neighboring cells.

- **Migrating and expanding colony (migration cluster)**: To simulate a population that moves across a space, colonizing its surroundings, we simulated a kernel

population of 20x10 cells entering and moving across the grid. In each step, expansion and migration occurred, so each sub-population size was also increased by populations in neighboring cells multiplied by factor of 0.2. This resulted in a kernel populations slowly colonizing larger part of the grid.

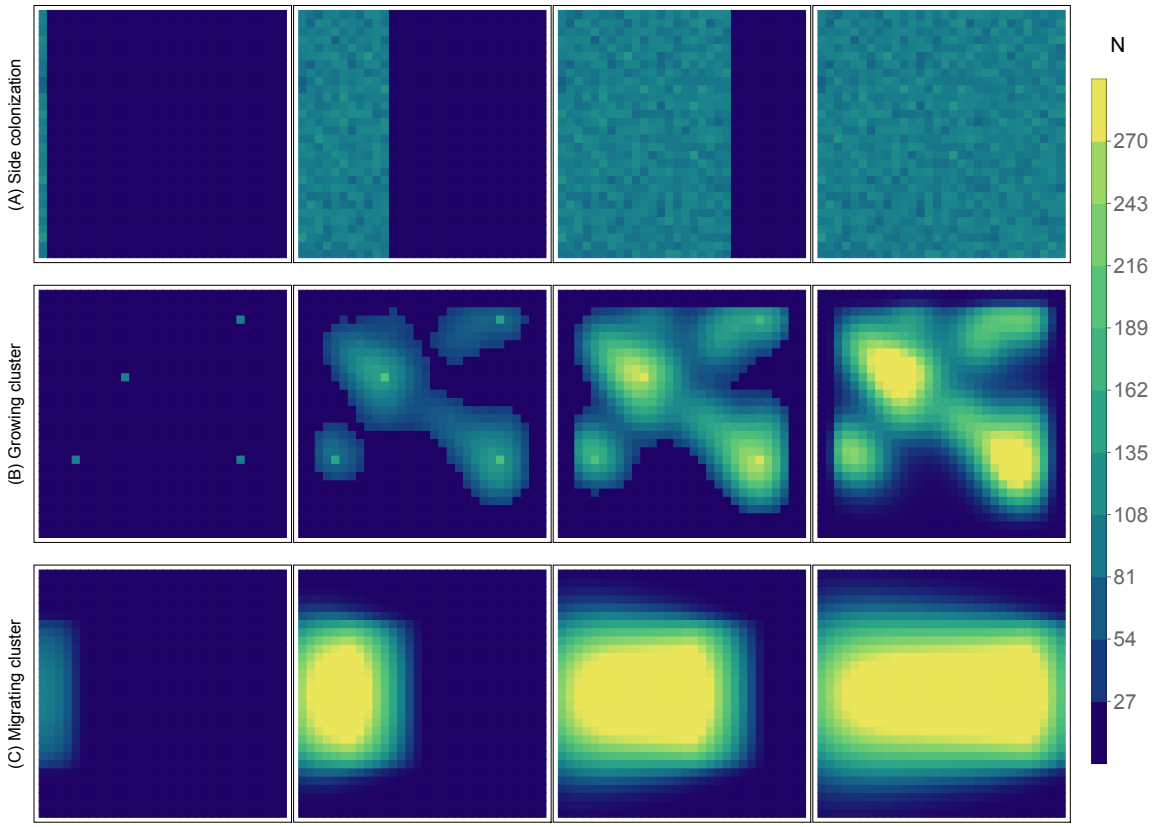


Figure 4.3: (A) colonization from a side where the whole map becomes occupied in the end. The final map (current time) has low variance, sampled from a Poisson distribution. (B) the colonization starts from seeds and converges to a clustered map. (C) migrating and expanding cluster. The four columns correspond to four time points: $30, 20, 10$ steps ago, and the final (current) map.

Variable	Symbol	Values
demographic history		side colonization, expanding clusters
Migration rate	m	10^{-5}, 10^{-4}, 10^{-3},10^{-2}, 10^{-1}

Table 4.3: Summary of parameter values used in the simulations of biologically inspired scenarios.

Demographies inspired by silver fir populations

The last set of simulations involved demographic histories that were inspired by the LPX-Bern model predictions for silver fir distributions across Europe (Ruosch et al., 2016). The LPX-Bern is a dynamic global vegetation model that simulates species distributions under various climate estimates of past 21000 years. We used this silver fir dataset to create some further input files for our tests, as it adds a realistic touch: we found difficult to create artificial datasets that are neither too regular nor too random.

The size of the spatial grid (1272 grid cells, 24×53 grid) and the number of time steps (220) were kept identical to the LPX-Bern model. We defined four demography types that are detailed below, also see Figure 4.4. As population size we used the best-estimate of N_e (the original dataset defined the distribution of species in foliar protective cover (FPC), that we turned into N_e based on various assumptions, these details can be found in Appendix Section B.2.) at particular time points to stay close to the model predictions and simplified the events in between these time points (see Table 4.4). Furthermore, to explore the effect of N_e, we scaled the population sizes of one scenario (A-Shape) by 10, 100 and 1000.

To investigate whether the demographic changes on a short timescale can have an influence on current diversity patterns despite the enormous effective population sizes, we used very short time steps: 100 and a 1000 years, equivalently, 4 and 40 generations (assuming 25 years as generation time). This first option of 220 timesteps of 100 years corresponds to 22000 years, which marks the beginning of the deglaciation of Europe after last ice age.

1. Expansion: we simulated mostly expanding populations. As the initial value at 22 kyr BP we used the best-estimate N_e when the overall population size was the smallest (17kyr BP), and, as the final point (0 kyr BP), we used the best-estimate N_e at 0kyr BP. Populations of all grid cells changed linearly between these two set points. Note that while the overall population size linearly increased, the initial population sizes in some grid cells (approximately 20% of all grid cells and 32% of grid cells with non-zero N_e) were larger than those predicted for today, so in those cases, the population size was slowly decreasing.

2. Decline: we simulated shrinking populations. As the initial value at 22 kyr BP we

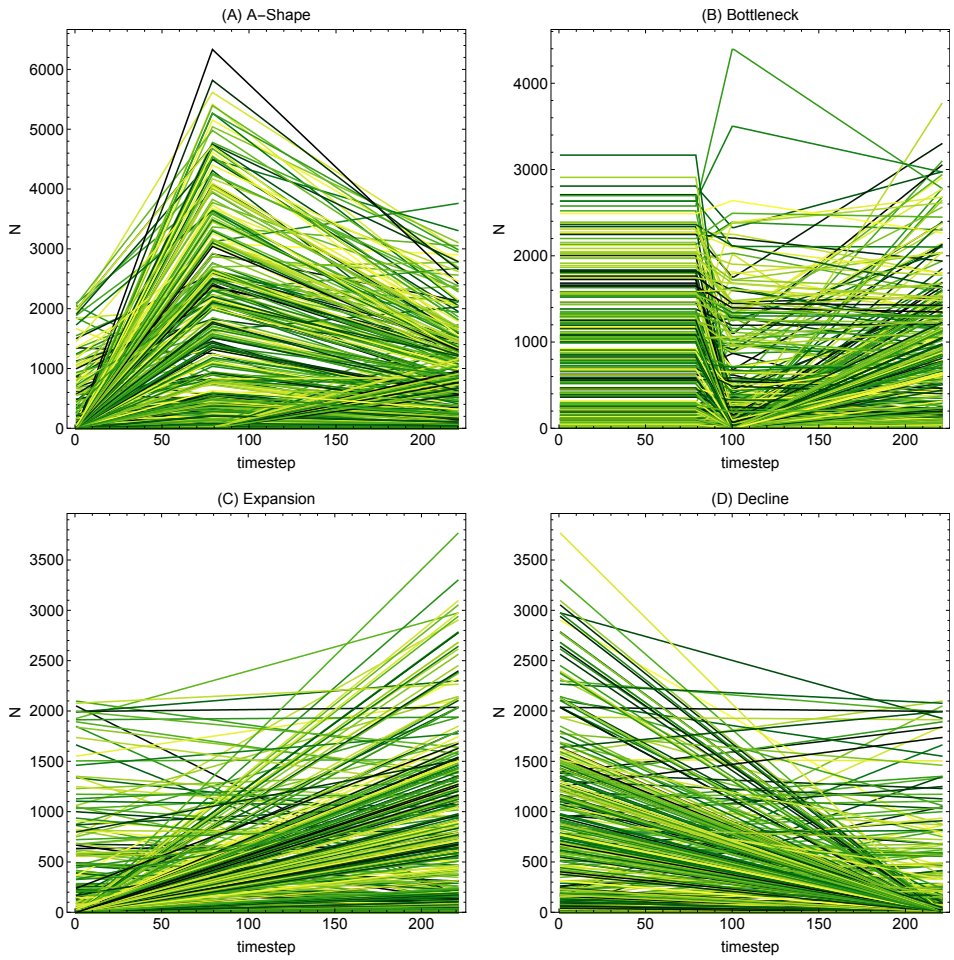


Figure 4.4: The following histories are defined on a 24×53 grid over 220 steps. The colors represent individual grid cells (not all of them shown here). The mean of the population sizes across the grid follows the described pattern over time ((A) A-Shape, (B) Bottleneck, (C) Expansion, (D) Decline), however, the individual cells can follow different courses. The data was created to preserve some level of spatial autocorrelation.

used the best-estimate N_e when the overall population size was the smallest (17 kyr BP) multiplied by 7, and, as the final point (0 kyr BP) we used the best-estimate N_e at 0 kyr BP. Populations of all grid cells changed linearly between these two set points. Note that while the overall population size linearly decreased, the population sizes in some grid cells (about 62% of grid cells with non-zero N_e) were smaller at the beginning than today, so in these cases, the population size was linearly increasing.

3. Bottleneck: we simulated a bottleneck similar to that observed in LPX-Bern data. The 220 time steps were divided into three parts:

   (a) constant population size with best-estimate N_e at 14.2 kyr BP, from 22 kyr to

14.2 kyr BP;

(b) linear decrease until 12 kyr BP, to best-estimate N_e at that time;

(c) linear increase until presence, to best-estimate N_e at 0 kyr BP.

4. A-shape: we simulated populations expanding starting with the best-estimate N_e at 14.2 kyr BP (smallest overall population size) to the best-estimate N_e at 12 kyr BP multiplied by 2. Then, the populations declined linearly until present time best-estimate N_e (0 kyr BP).

Scenario	Population size expressed as best-estimateN_e at [kyr BP]			
Time of events	22 kyr BP	14.2 kyr BP	12 kyr BP	0 kyr BP
Expansion	Ne[17]			Ne[0]
Decline	7*Ne[17]			Ne[0]
Bottleneck	Ne[14.2]	Ne[14.2]	Ne[12]	Ne[0]
A-shape	Ne[17]	2*Ne[14.2]		Ne[0]

Table 4.4: Events of the four simplified demographic histories based on the LPX-Bern model.

4.2.3 Summary statistics

While population allele frequencies can be simulated, in cases with large effective population sizes, long demographic history, and large spatial extent it is more feasible to simulate only the genealogies of two lineages per population to obtain coalescent approximations of population diversity and divergence (Slatkin, 1991).

For several parameter sets, simulation time was not sufficient for all lineages to coalesce in all simulations. Therefore, at the end of the simulation, all populations are pulled into a single panmictic, spatially non-explicit population of size 1, effectively stopping the simulations. The resulting coalescence times from the simulations can be corrected by the weighted size of the ancestral population and the mean of the simulation. Further details can be found in the next subsections.

Mean coalescence time

Calculating the mean within and between-coalescence times for the individual grid cells and for the pairs of grid cells sounds like a triviality, however, it raises some questions—both biological and mathematical ones. Depending on the migration rates and the time up until we run our simulations, or better yet as far back in time as we have data, some lineages may not coalesce. The question arises: what happened before. As we may have no further information about the history of the populations from before the data ends, we can assume either that they all come from an ancestral population and within that there is no spatial structure and the lineages follow the standard coalescent process, or that the population structure changes no more and the lineages will coalesce depending on the population sizes and the connectivity of the grid cells.

We assume in the simulations that all the lineages that did not coalesce in the time interval from the present back to a given time will be moved into a single ancestral population. This is a reasonable assumptions when one considers separations of timescales: the *scattering* and *collecting* phase of the coalescent process in a structured population (Wakeley, 1998, 1999). This means that the genealogy can be separated into two parts: a short scattering phase (more recent events) and a long collecting phase (ancestral events). The more recent events happen in the deme: lineages either coalesce or move away. In the second phase, the lineages diffuse over the whole grid before they coalesce, thus the population will behave as a single ancestral population and follow a standard coalescent process. The expected coalescence time for a pair of randomly chosen individuals is the sum of the expected times of these two phases: coalescence in the first phase is proportional to the neighborhood size ($2\rho L^2$, where ρ is the density) and the time spent in the second phase is proportional to L^2/σ^2 (time until lineages diffuse across the range). (Charlesworth et al., 2003, Equation 9.).

In some cases, we know that the coalescence time distribution is exponential, when migration is very low or very high and the population sizes are constant over time - in these cases one can calculate the mean from the median, but in most cases this approach will not work. In case of exponential distributions the following holds:

$$\mathbb{E}(T) = \mathrm{med}(T)/\ln 2. \tag{4.3}$$

We propose to calculate a corrected mean coalescent time in the following way. Let T^* be the time point when all lineages are pulled into an ancestral population, let P be the fraction of simulations that coalesced before T^*, and Q the fraction that did not. The mean coalescence time of the lineages that coalesced is $t_0 < T^*$, and the mean of the rest is $T^* + 2N_a$, where $2N_a$ is the effective size of the ancestral population, without further knowledge, assumed equal to the total population size of the grid at time 0.

$$\mathbb{E}(T_{coal}) = P \cdot t_0 + Q(T^* + N_a) \tag{4.4}$$

When the simulation is long enough, most lineages coalesce therefore there is no difference between the mean and the corrected mean. When dt, the time between two time points at which the population is updated, is too short, the corrected mean is dominated solely by the size of the ancestral population. On Figure B.2 in the Appendix the difference between the mean of the simulated data and the corrected mean is shown for four values of dt.

Global $\mathbf{F_{ST}}$ and pairwise $\mathbf{F^*}$

To assess how differently populations evolve in face of gene flow, we calculate the fixation index, also known as F-statistics. We will use this statistical measure to obtain isolation by distance patterns and observe how it changes under different demographic histories. F_{ST} was first introduced by Wright (1943) to express expected heterozygosities at various levels of the population structure. It can be interpreted as a measure of the amount of differentiation among subpopulations relative to the limiting amount under complete fixation. F_{ST} is a widely used measure in population genetics, both in theoretical and empirical studies. It can be expressed in terms of many biologically important quantities, such as gene frequencies, identity by descent probabilities, or coalescence times, we will use this latter one as we said earlier. A large amount of equivalent formulas of statistics estimating F_{ST} can be found in the literature, partially due to the different simplifying assumptions introduced in different models (e.g. Nei, 1972; Slatkin, 1993; Hudson, 2002).

The original formula can written in terms of coalescent times, namely:

$$F_{ST} = \frac{T_T - T_W}{T_T}, \tag{4.5}$$

where T_T is the average total coalescence time and T_W is the average within coalescence time across the subpopulations. We refer to this as the global F_{ST} that is meant to measure the strength of population structure. Using this formula allows us to consider only the mean coalescence times instead of the full distribution.

Furthermore, we are interested in calculating pairwise measures as well, in order to investigate genetic differentiation between subpopulations. If we were to have only two subpopulations, then Equation 4.6 would transform into

$$F^*_{i,j} = \frac{T_{ij} - \frac{T_{ii}+T_{jj}}{2}}{T_{ij} + \frac{T_{ii}+T_{jj}}{2}} = \frac{T_B - T_W}{T_B + T_W}, \tag{4.6}$$

where T_{ij} denotes average coalescence time of two genes drawn from demes i and j, T_B is the mean coalescence time for two lineages sampled from different demes, and T_W is for lineages sampled from the same place. This formula is correct if (and only if) two subpopulations are involved, nevertheless, as Slatkin (1993) suggests, this may not be a correct way to assess the strength of population structure in general, but it may be a good measure to estimate genetic distances between subpopulations.

We will use two approximations of F_{ST} in the island model and the $2D$ stepping stone model in order to verify the results of the simulations and draw further conclusions about more complicated scenarios. These approximating formulas have relatively simple forms since they rely on strict assumptions. We will investigate how different the simulated F_{ST} values are from the theoretical predictions when certain assumptions are relaxed.

For neutral alleles in the infinite island model we will use the following well-known formula:

$$F_{ST} = \frac{1}{4Nm + 1}. \tag{4.7}$$

In Kimura's two dimensional stepping stone model, it is possible to derive F_{ST} as a function of migration rate and population density in a homogeneous environment (Cox et al., 2002):

$$F_{ST} = \frac{\frac{L^2 logL}{2\pi\nu\sigma^2}}{\frac{L^2 logL}{2\pi\nu\sigma^2} + 2NL^2} \tag{4.8}$$

where $\sigma = 1/2$ is the standard deviation of parent-offspring distance along an axis, $\nu = 4m$ (ν is the total rate of migration to all possible demes: in our case it is the 4 neighbors, each with rate m) and L is the grid size.

4.3 Results

4.3.1 Static populations

Comparing theoretical predictions and simulation results

Several tests were carried out to assess how well the simulations match various analytical predictions. First, we compared simulation results with predictions given for the standard coalescent process in panmictic populations, then with predictions given by the $2D$ stepping stone, and infinite island models in structured populations.

Standard coalescent process Theory predicts that mean coalescence time between samples taken from a single panmictic population of size N is equal to $2N$. On a grid, this predictions can be tested by either defining no migration, leading to each deme representing an isolated panmictic population, or by choosing a large migration rate, leading to a single panmictic population across the whole grid. In the first case, the average coalescence time of two lineages taken from the i-th deme is expected to be $2N_i$. In the second case (large migration), the average coalescence time between two genes taken from any demes should be $2N_{total} = 2L^2\overline{N}$, that is twice the total population size of the whole $L \times L$ grid. In both cases, the coalescence times follow and exponential distribution with the aforementioned means.

Figure 4.5 shows the average coalescence time between two lineages taken from the same deme, for all scenarios described in Section 4.2.2, with no migration. As predicted by the theory, average coalescence time depends solely on the population size of each individual deme. In comparison, Figure 4.6 shows the average coalescence time between two lineages taken from the same deme, for the same scenarios, with high migration rate ($m = 1$). In this case, the whole grid represents a single, (almost) panmictic population. The figure shows that the coalescence time is independent of the individual deme size, rather, it is determined by the total population size of the whole grid, as expected. We

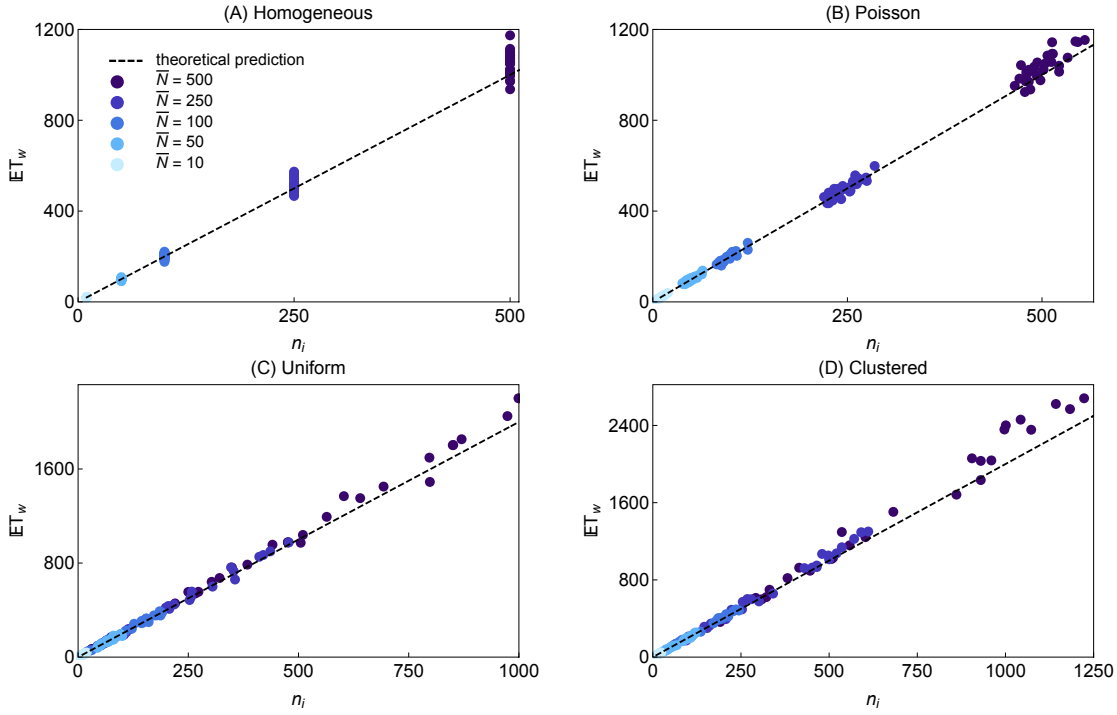


Figure 4.5: Within cell coalescence time (on y-axis) depends solely on the population size of each individual sampled cell (x-axis), if no migration occurs. The mean was estimated from the observed median using the formula $M = \lambda \log(2)$ under the assumptions that coalescence time distribution is exponential. Different underlying maps are shown here: (A) no variation (Homogeneous), (B) low (Poisson) and (C) high (Uniform) variation, and spatially correlated (Clustered). Different colors represent simulations with different average population sizes ranging from 10 to 500.

observe a slight discrepancy between the simulation results and the theoretical prediction for the mean coalescence time ($2L^2\overline{N}$). Even though the migration rate used here is high, the spatial structure does not fully vanish: the probability of lineages located physically closer to each other coalescing first is higher than that of lineages further apart. Figure B.4 shows that lineages taken from different demes coalesce, on average, at the predicted time ($2L^2\overline{N}$). If the population structure would fully disappear, we would see no difference between the within and between coalescence times.

To calculate the means we used the fact that the mean coalescence time distributions are exponential in these special cases, so we used the median formula. Note, that since we run the simulations long enough, the mean of the data would give (almost) the same results, for the number of lineages that did not coalesce please refer to Table B.1.

Structured populations Predictions for two theoretical models, the 2D stepping stone (Equation (4.8)) and the infinite island models (Equation (4.7)) were compared with the

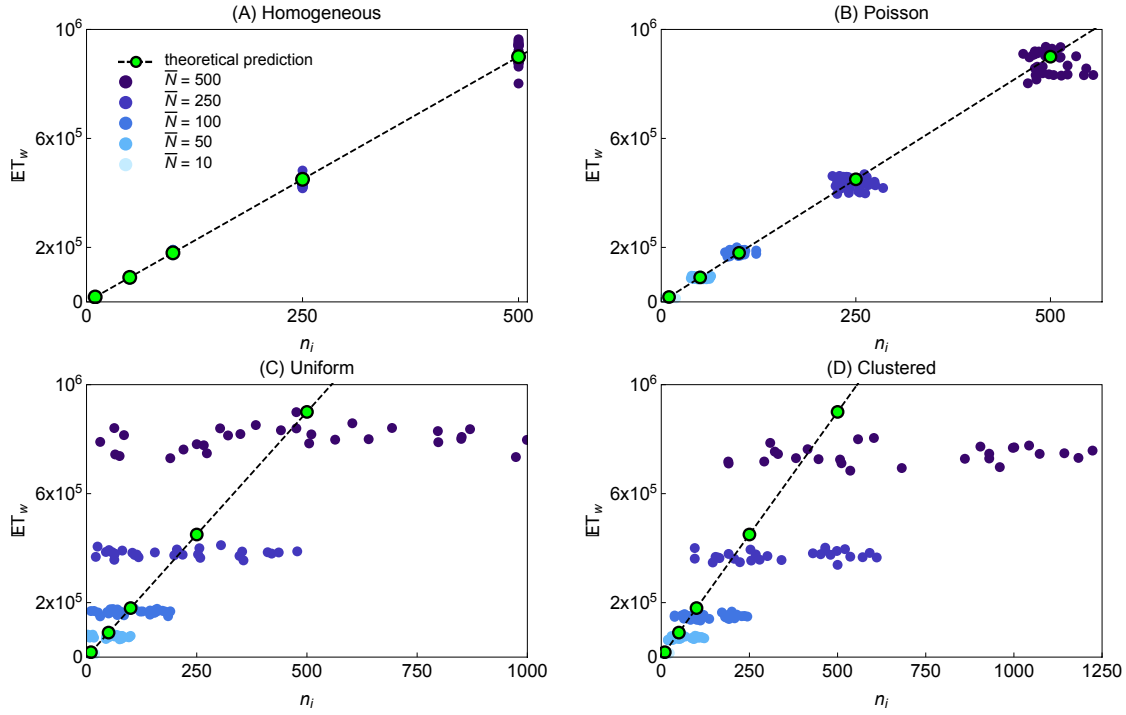


Figure 4.6: Within cell coalescence time ($\mathbb{E}T_W$) against the individual population sizes. $\mathbb{E}T_W$ depends on the mean population size across all cells if the migration rate is high ($m = 1$) for the four types of spatial structure. The mean was estimated from the observed median using formula $M = \lambda \log(2)$ under the assumptions that coalescence time distribution is exponential. In this case, the demes form one single population across the whole grid together, hence the theoretical prediction for the mean coalescence time is $2L^2N_e$ (shown in dashed with green dots), twice the total size of the grid. Different colors represent simulations with different average population sizes. The green dots mark the exact function values of the theoretical expectation for the five simulated $\overline{N} = 10, 50, 100, 250, 500$.

results of the simulations of all scenarios described in table 4.2.2. We calculated global F_{ST} values for all parameter combinations using Wright's formula (Equation 4.6)

Figure 4.7 shows that the predictions of the 2D stepping stone model fit well with the simulation results of the homogeneous environment. On the other hand, F_{ST} in simulations of an environment with large variance and no correlation between deme sizes across the grid (uniform map) is best predicted by the infinite island model. See Figure B.5 for other scenarios.

Effect of spatial variation, migration rate, and deme size

Below we investigate the effects of spatial variance, the average deme sizes, and migration rate on the expected coalescence times, global F_{ST} values and isolation by distance patterns on genetic distance F^*.

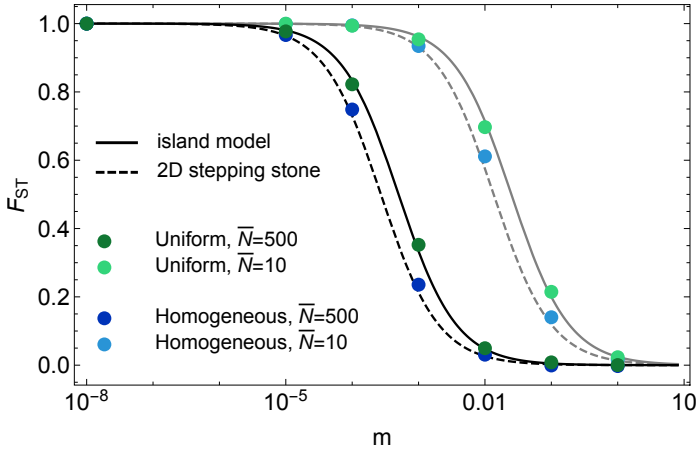


Figure 4.7: Global F_{ST} values against migration rates, as predicted by the theory (Equation (4.7) and (4.8)) and calculated from simulations. The green dots represent simulations with high spatial variance between population sizes and the blue ones correspond to no spatial variance. The darker dots belong to larger population sizes ($\overline{N} = 500$) and lighter ones to smaller population sizes ($\overline{N} = 10$).

Dependence of mean coalescence time on m and $\overline{N}$ As expected, the average coalescence time is influenced both by the deme size and the migration rate. Simulation time (total time of defined demographic history) was limited to 40000000 generations, and not all pairs of lineages coalesced within this time. (The total population size was $900 \cdot N$, and was taking values in $\{10, 50, 100, 250, 500\}$.) Therefore, to illustrate the effect of migration rate in a least biased way, Figure 4.8 shows median coalescence time for various scenarios described in table 4.2.2. Figure 4.8 reveals that different scenarios provide most contrasting results for intermediate migration rate values. For very large or very small migration rates, coalescence times for lineages taken from the same deme are well estimated by standard coalescence, as described above. However, at intermediate migration rates, migration between the demes enhances differences between the demes.

Isolation by distance patterns To analyze the effect of geographic distance on genetic distance between the subpopulations, we calculated pairwise genetic distance coefficient F^* for pairs of samples taken from a row of demes. The effect of mean deme size, variance in deme size, and migration rate is shown in Figure 4.9. The figure reveals that larger mean deme size (panel (A)), as well as larger migration rate (panel(B)), lead to lower values of F^*, meaning weaker differentiation between the populations.

The amount of spatial variance of a given map also contributes to the genetic differen-

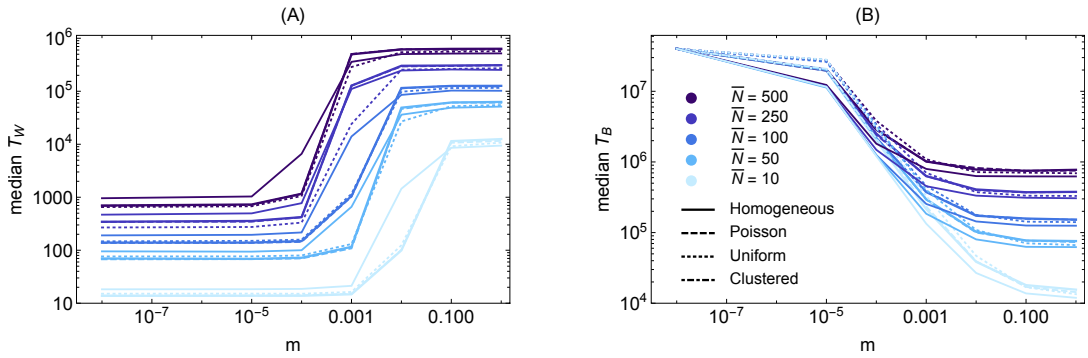


Figure 4.8: Median coalescence time for lineages taken from the same (A) or from different (B) demes as a function of migration rate m. Different colors represent different average deme size, while different line types correspond to different spatial variance (map type). The difference in median coalescence times created by the different spatial arrangements (maps) is neglectable compared to the differences due to different mean population sizes.

tiations between populations (panel (C)). One can consider the no or low variance maps as a baseline value: these result in similar values to each other in case of each distance class.

The uniform map (high spatial variance) has a higher mean F^* across all the distance classes: the compared demes in any distance class are just as likely to have similar or different size, which pulls the averages up.

The clustered map (with spatial autocorrelation), however, in case of small distance classes results in a lower mean, as the pairs of populations located close to each other tend to have similar sizes, thus being less differentiated. On the other hand, these distance classes consist of pairs of small populations and pairs of big populations. Their contribution to F^* thus has a large variance: pairs of big populations have low F^*, and the pairs of small populations have high F^* values, as we have seen on panel (A) of Figure 4.9. In case of populations further apart, most of the comparisons will be made between a large and a small population, resulting in increased genetic differentiation.

4.3.2 Populations with variable demographic history

After exploring the effect of spatial heterogeneity on diversity, i.e. the mean within coalescence time, we proceed with adding temporal variation to the simulated scenarios. In real biological situations, spanning long periods of time, we expect natural populations

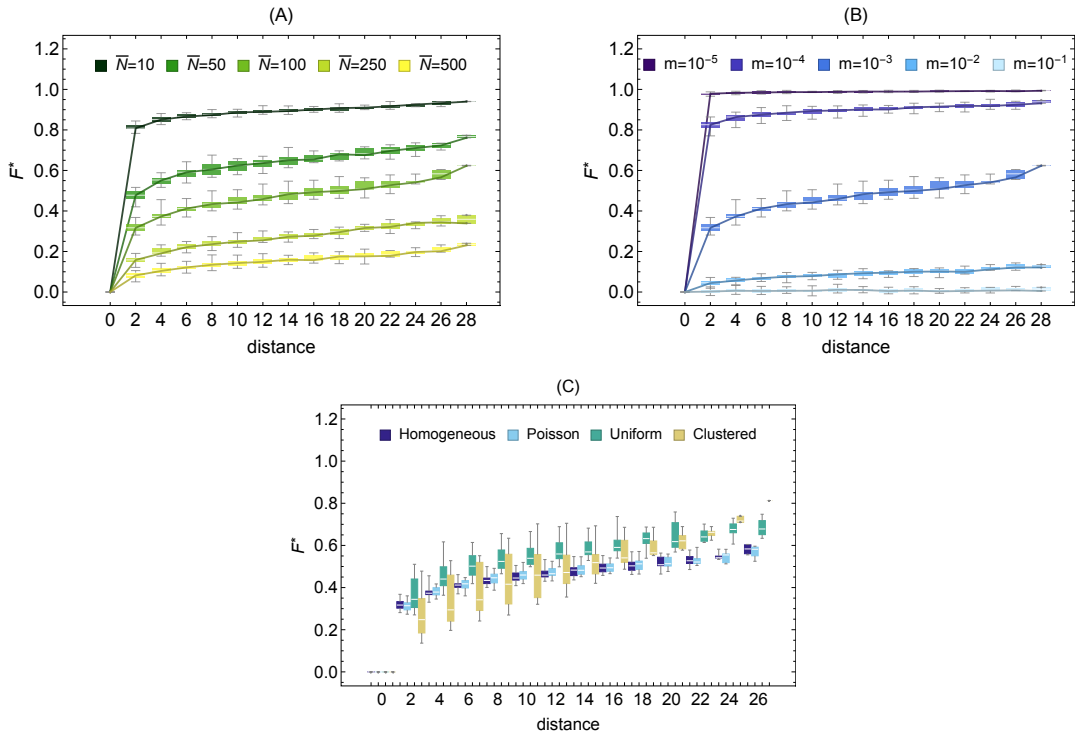


Figure 4.9: Pairwise genetic distance F^* as function of distance between the sampled demes. (A) Effect of population size: bigger populations differentiate less as they are less affected by genetic drift. (B) Effect of migration rate: more migration results in more admixture, making populations more similar to each other thus decreasing F^*. (C) Effects of maps exhibiting different amounts of spatial variance: the homogeneous and the Poisson maps (no and low variance) can be considered as a baseline value. Compared to this, spatial autocorrelation (clustered map) results in lower mean and bigger variance in case of small distance classes, and higher mean and smaller variance in case of large distance classes. In the case of small distance classes, the pairs close to each other tend to have similar sizes. As panel (A) suggests, the genetic difference, the F^* values produced by a pair of large populations on the one hand and a pair of small populations on the other hand can be substantial, hence the high variance. In the case of large distance classes, most of the comparisons will be made between a large and a small population, also resulting in an increased genetic differentiation. The uniform map has a higher mean across all the distance classes for the same reason: two demes in any distance class are just as likely to have similar or different size.

to go through various phases, including expansions, decreases, bottlenecks, colonizations, or extinctions. Their combined effects influence the diversity we observe in the current populations. However, disentangling these effects, finding their separate signatures on diversity, and assessing their relative importance is a complex problem.

In this section we investigate the effects of temporal variability on population diversity and isolation by distance patterns. Furthermore, by correlating the diversity pattern with

historical population sizes, we assess the effects of past and recent demographic history on observed diversity. The simulation inputs are described in Section 4.2.2 in more details.

Simple demographic histories on uniform maps

The first set of simulations was meant to shed light on whether we can distinguish between different simple demographic histories, such as increase, decrease, or a bottleneck, assuming that they have the same current population distribution on a grid, as shown in Figure 4.2. Of course, F_{ST} is known to be robust under population size changes, whereas other genetic measures, like *Tajima's D* (Tajima, 1989; Ross-Ibarra et al., 2008), or the *PSMC* model (Li & Durbin, 2011) can pick up signals of demographic events. However, to use these inference methods we would need to simulate whole genomes, which is currently not possible with *gridCoal*. Nevertheless, we calculate the global F_{ST} values and after we compare the diversity (mean coalescence time) to the historical population sizes.

First, we calculate the global F_{ST} values of the five types of simulations for different migration rates. A relatively long time frame was used: the number of generations until the ancestral state is reached is 240000. This made sure that most lineages coalesced, thus we did not need to use the correction formula. Figure 4.10 shows that in comparison to a baseline value ("No change in mean"), the declining populations produce smaller F_{ST} values, and the populations going through linear increase, bottleneck, or exponential increase produce larger ones. Their order is the same for all migration rates: this reflects the rate of increase in the population sizes, that is, the higher the rate of increase, the higher F_{ST} it produces. Note, that the bottleneck increases to the same map as the linear increase but over half of the time, after its minimal size was reached. That being said, the migration rate has a more profound influence on the F_{ST} values and the differences between them produced by the different demographic histories are not significant, partially due to the large time-scales (many coalescent events happening in the last timestep), and due to the correlation between space and time. Thus, differentiating between these histories merely based on the global F_{ST} is impossible.

To find out which time point in the past has the largest influence on the current diversity, we correlated the historical population sizes to the obtained diversity (mean within coalescence time for each currently occupied cell), see Figure 4.12. (The obtained diversities for all scenarios can be see on Figure B.7 in the Appendix.) In case of high

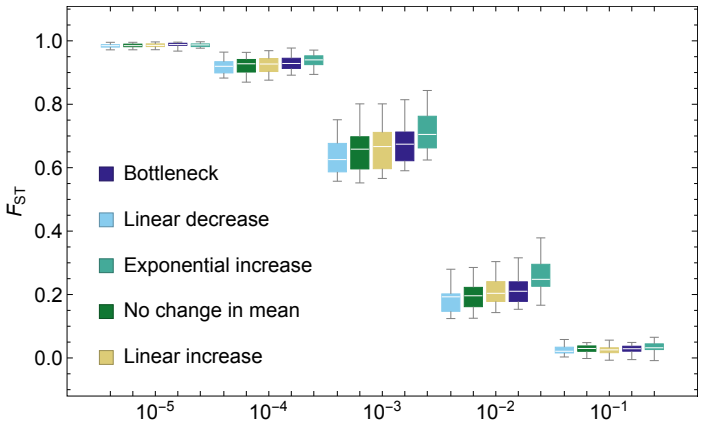


Figure 4.10: Global F_{ST} as a function of migration rate for different demographic histories. For each migration rate, the declining population gives the smallest value, whereas the exponential the biggest one. The population with no change in mean can be considered as the baseline value. The order of the three populations producing the highest F_{ST} values reflect on the rate of increase in population size. The migration rate has a more significant effect on the F_{ST} than the demography types.

migration (panel (A) in Figure 4.12) we do not observe strong correlations as there is no spatial structure on the grid and because high migration has a strong homogenizing effect among the cells. This produces slightly negative correlations because it enhances the effect of the neighboring demes. On a uniform map, on average, small demes are surrounded by big demes and big demes by small ones, thus the negative correlation. The only stronger pattern is produced by the bottleneck: its minimal population size at half time of the simulation shows the strongest correlation, however, it is still very small. For low migration rates (panel (B)) the correlations behave differently. The maps are the same, however, each cell is less connected to its neighbors thus being less influenced by what happens at the other parts of the map: the coalescence time depends more on the individual cell. The highest correlation is reached at the current time for all scenarios, and the convergence to this value reflects the rate of increase in the population size.

The declining population shows a very similar pattern to that of the increase. This is the side effect of not every individual grid cells following the same behavior, only on average. Many declining demes vanish completely by the current time, making it impossible to sample, therefore their effects are only indirect on the coalescence times. In fact, what is called decrease is not a very good representation of this sort of demographic history. However, in this chapter we will not try to find more suitable ways of describing

population declines and we will include decline in its current form into our analysis.

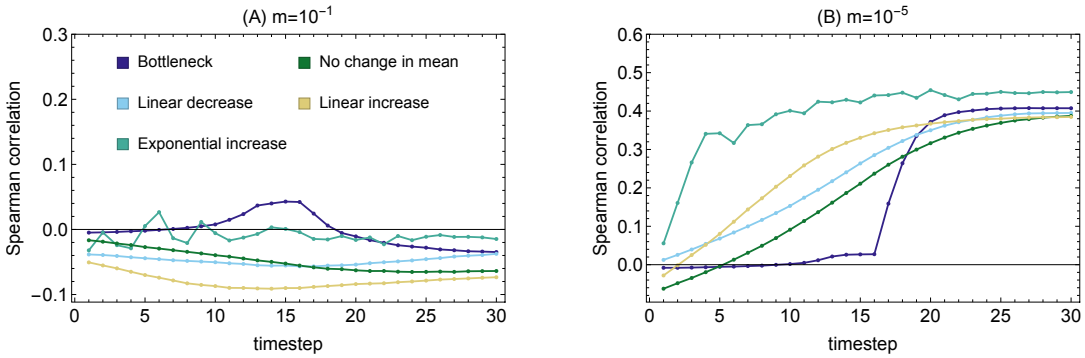


Figure 4.11: Correlations between historical population sizes and observed diversity (mean within coalescence times) through time. Timestep refers to the time points at which the population sizes got redefined. The time between two timesteps is what we denoted with dt: the actual years or generation time between two updates (here dt=2000 generations). Comparison of five demographic histories (described in Section 4.2.2). They all converge to the same current uniform map. (A) high migration scenario: $m = 10^{-1}$, (B) low migration scenario: $m = 10^{-5}$. For high migration rates the correlations are slightly negative and the different demographic histories make little difference due to the homogenizing effect of migration. The correlations are negative because at high migration rates the neighborhood size becomes more important: in case of a uniform map, small demes are more likely to be surrounded by big ones and vice versa (see Figure 4.1), leading to a negative tendency in the correlations. In case of low migration rates, the current population sizes are the best predictors of diversity as the grid cells are not spatially correlated and the coalescence times are more dependent on the individual cells themselves.

Simple demographic histories on clustered maps

To obtain stronger correlations between the diversity and the population sizes we created datasets where the populations are distributed in clusters and go through similar demographic events that we had before, namely a bottleneck, expansion, or decline, see Figure B.3 in the Appendix.

The correlation between the observed diversity (shown in Figure B.8) and the population sizes are shown in Figure 4.11 for high (A) and low (B) migration rates.

In this cases much higher values of correlations can be observed due to the strong spatial autocorrelation of the maps. However, there is a correlation between space and time as well: at most time steps, the spatial arrangement, the number of occupied cells are not changing. Only the population sizes change, however, their relative size compared to each other remains the same, thus the flat correlation. The only non-flat correlation

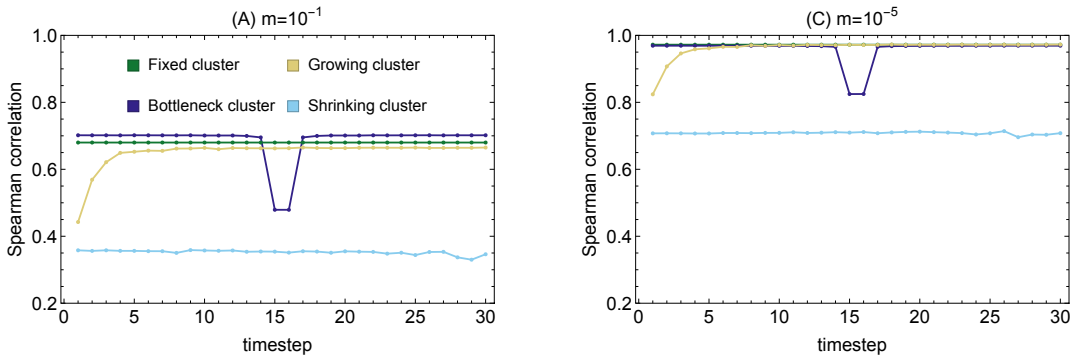


Figure 4.12: Correlations between historical population sizes and observed diversity (mean within coalescence times). Spatially autocorrelated (clustered) maps go through various demographic histories, similar to Figure 4.11. (A) high migration scenario: $m = 10^{-1}$, (B) low migration scenario: $m = 10^{-5}$. We observe higher correlations than in case of the uniform maps, however, there are no changes in the correlations over time due to the correlation between space and time. (The grids at different time points have similar spatial arrangements and they only differ by a scaling factor.) The only change occurs in case of the bottleneck and the early steps of the growing cluster. At these time points different cells are occupied than in the rest of the time (cf. Figure 4.3). Changes in migration rates only shifts the correlations up or down: weak migration enhances the effect of the individual cells as they are less connected to their neighbors.

is produced at the beginning phase of the growing cluster and at the middle phase of the bottleneck. These are exactly the time points, where the actual spatial arrangement is different compared to the rest of the time (cf. Figure 4.3 in the Appendix). We find that in case of a bottleneck, the lowest correlation is produced when the population size is minimal, which is the opposite compared to what we have observed in case of a bottleneck with a uniform map. (cf. Figure 4.11 (A)).

Biologically inspired demographic histories

In order to break the temporal and spatial correlations we created datasets in which the populations are changing their sizes and locations over time. These examples are inspired by real biological situations, such as a colonization dynamics, where the front edge of the population constantly pushes into new territories, or when populations are colonizing their empty neighboring cells, expanding to all available directions, merging and forming new clusters, or when a single cluster moves along an environmental gradient, and expands along the way, for our examples see Figure 4.3. We created demographic histories corresponding to the reversed events: declining and receding clusters, however

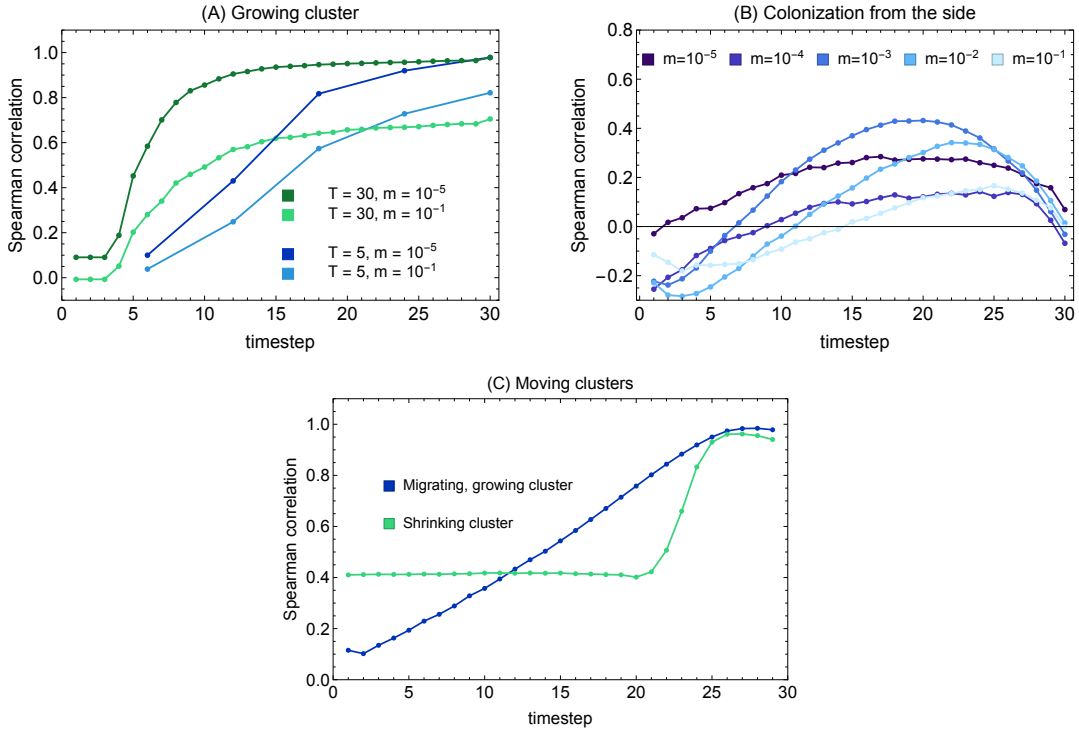


Figure 4.13: Correlations between historical population sizes and observed diversity (mean within coalescence times). (A) Comparison of different time resolutions in case colonization happens from seeds. We reach the same final map in 5 timesteps (blue) or 30 step (green). Two results for two migration rates are shown here, a smaller one 10^{-5} (dark) and a bigger one 10^{-1} (light). (B) Comparison of different migration rates in case colonization happens from the side. The biggest correlation is observed when the migration rate is intermediate. (C) Comparison of expanding and migrating cluster to a receding one. The spatial and temporal correlation is broken: the correlation between the population sizes and the diversity increases as the cluster moves into its current location. In case of the receding one we only sampled the current population, which is small. The correlation is the strongest when the population is only present at the sampled locations.

these exhibit the same flat correlation patterns as we observed before.

The first panel (A) in Figure 4.13 shows the effect of having 5 or 30 timesteps spanning the same total time up until we run the simulations. We find that the finer resolution reaches the maximum faster, whereas the coarser resolution takes longer time. This can be explained by the fact that the first steps of the finer resolution is to colonize the cells that are occupied in the end, whereas after it only increases in size leaving the correlation unchanged. The diversity patterns generated by the 30 steps demography can be seen in Figure B.9, they reflect the current population sizes.

The second panel (B) in Figure show the correlation between the diversity (Figure 4.14) and the population sizes in case of the population colonizing the grid from the side.

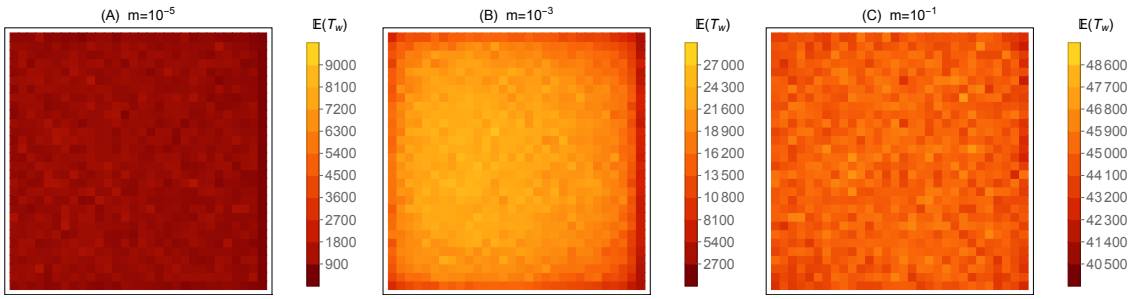


Figure 4.14: Observed diversity (corrected mean within coalescence time) in case of colonization from the side. (A) Low migration rate 10^{-5}, (B) medium migration rate 10^{-3}, and (C) high migration rate 10^{-1}. Note the different scales. We see the strongest effect at the intermediate migration rate.

The high ($m = 10^{-1}$)and low ($m = 10^{-5}$) migration rates do not result in any spatial pattern in the diversity, however, in case of an intermediate value (10^{-3}) we find that the last colonized region has a much lower coalescence time than the rest of the grid. This is responsible for the highest correlation around timestep 20, when the right side of the map is yet to be occupied.

Lastly, panel (C) on the same figure shows the correlations for moving and expanding, shrinking, and receding clusters. The moving and growing cluster produces increasing correlation over time, as the cluster slowly moves into its currently occupied location. In case of a shrinking cluster, we could only sample the grid cells on the left side, so the correlation is bigger at the time points where the occupied grid cells are the same as the sampled ones, meaning the last few timesteps.

Histories based on the LPX-Bern dataset

The last set of simulations we discuss in this chapter is the one created based on the LPX-Bern dataset, see Figure 4.4. We used a much shorter time scale (220 steps, each 4 generations long) so we can ask questions about silver fir populations that expanded their range across Europe since the last ice age ended. Effective population sizes of trees can be immense, corresponding to extremely long coalescence times. (Estimating the actual number of trees is challenging but to have a rough idea of the magnitude in question: a study by Crowther et al. (2015) suggested that there are more than 3 trillion of trees in the world. Note, that this estimate considers all the different species across all the world,

and it is the census size, rather than the effective population size.) However, since the last ice age, due to swift shifts in temperature, the populations of silver firs went through various phases of bottlenecks, extinctions, and colonizations. To assess, whether despite the short time period of 22000 years we can identify some signals of the demographic history, we run further simulations as described in Section 4.2.2. This dataset is less artificial than the previous ones, as some patches are more correlated with each other, yet different parts of the maps may go through different events:.

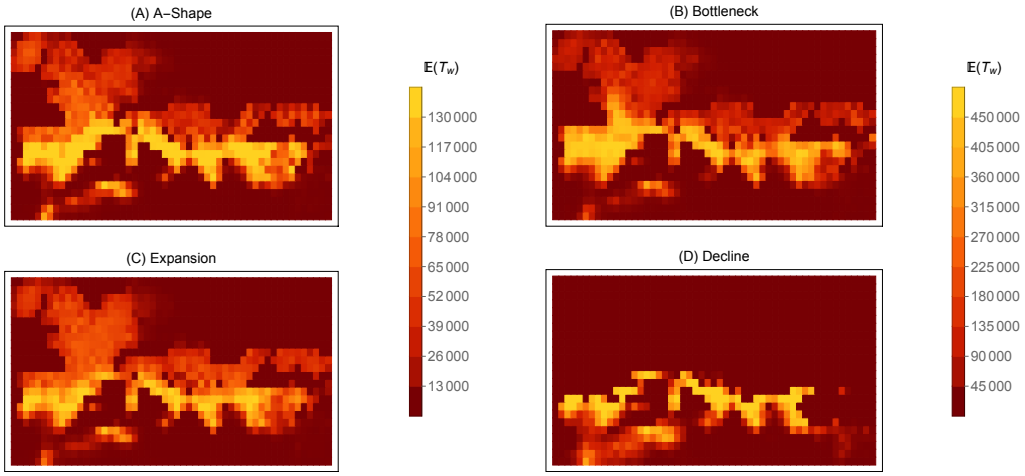


Figure 4.15: Observed diversity (corrected mean) for four types of demographic histories motivated by the LPX-Bern dataset. (A) A-Shape, (B) Bottleneck, (C) Expansion, and (D) Decline. For panels (A-C) the same cells are sampled, for panel (D) we had much fewer cells that are populated at the current time. We found that in 25% of the simulations lineages did not coalesce. We used the correction formula to account for this. The ancestral population sizes were defined as the total population size on the grid at time 0. In case of (A) and (C), the ancestral population size is $9.6 \cdot 10^4$, in case of (B) it is $2.89 \cdot 10^5$ and for (D) it is $3.8 \cdot 10^5$. Note that these values are higher than the number of generation up until the simulation was run, which is only 880.

In Figure 4.16, we can see that this dataset indeed breaks the strong spatial and temporal correlations and we can actually observe patterns changing through time. We find that in the simulations a quarter of the lineages did not coalesce, therefore we used the correction term. In panel (A) we compare the different demography types and find that the recent population sizes are the most informative in all cases. This should not be directly compared to the previous results, since the effective population sizes and the time scales are vastly different in this case. Panel (B) may shed some light on this issue: we compare the correlations for four scalings of population sizes. We find that the population size affects not only the height of the correlation curve but its shape too. In

case of bigger population sizes, the older time points are more descriptive of the current diversity, whereas when the population sizes are rather small then recent event influence it more.

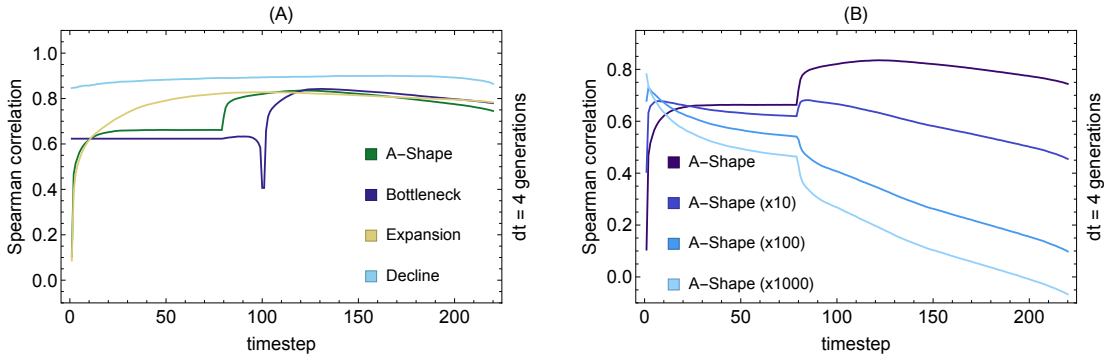


Figure 4.16: Correlations between historical population sizes and observed diversity (mean within coalescence times). (A) Population histories were generated based on the LPX-Bern dataset and run for a much shorter time (as it meant to simulate the effect of demographic changes since the last glaciation). This data breaks spatial and temporal correlations we have seen earlier, producing more clear patterns in the correlations. However, the declining population still is a problem as many of them disappear and thus cannot be sampled. B) Different scaling of population sizes in the LPX-Bern motivated A-Shape demographic history. The same input file was used but at each multiplied with a scaling factor of $1, 10, 100,$ and 1000. For bigger population sizes the older events are more important and for smaller population sizes the more recent ones.

4.3.3 Addressing limitations

Effect of grid size and edges

Most of our simulations were carried out on a square grid of $L \times L$, meaning that the demes residing at the edges were experiencing different conditions than those in the middle. To investigate the effect of grid size on the IBD patterns, we simulated different grid sizes of $L \in 10, 30, 50$ under the same conditions, with deme sizes drawn from uniform distribution, fixed in time.

Figure 4.17 A shows IBD patterns for different grid sizes. In all cases we see increase of genetic differentiation with increasing geographic distance. However, in order to disentangle the effect of the grid edges from the grid size itself, we calculated pairwise F^* for a set of 30 or 10 demes in the middle of the larger grids (4.17 (C-D). This shows that

rather than grid size, proximity of the edges increases the genetic distance between the demes.

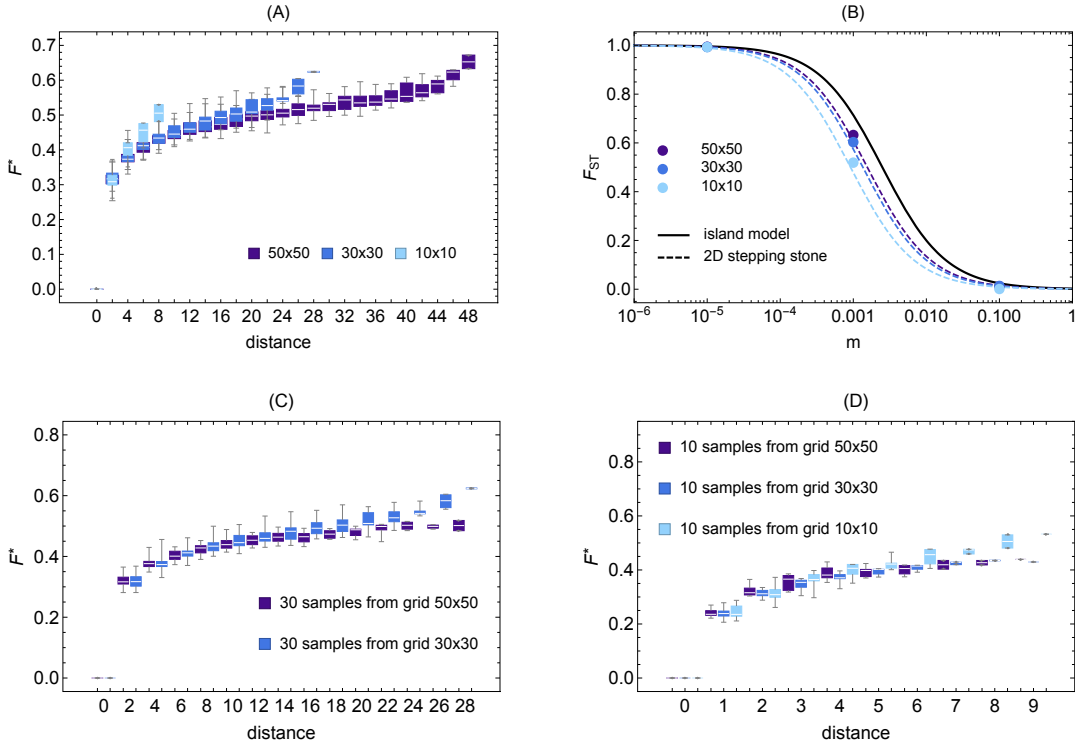


Figure 4.17: (A) Pairwise F^* on different sized grids: 10×10, 30×30, 50×50. (B) Global F_{ST} values together with the theoretical predictions for the different sized grids. (C-D) The genetic distance for the largest distance classes shows an elevation. To confirm that it is indeed an edge effect, we calculated this pairwise measure for the same amount of samples taken from the middle of a larger grid and found that it did not appear there.

Figure 4.17 B shows that the size of the grid determines the suitability of the model predictions for the simulated scenario. In all cases, the 2D stepping stone model is a better approximation of the simulated scenarios than infinite island model, and larger grids fit better to prediction of both models than small grids.

Temporal resolution, time step size and limited historical information

Our results suggest that temporal resolution (number of supplied time points over the same period of time) plays a minor role for mean coalescence time estimates as well as genetic distance patterns, whether it is between pairs of lineages taken from the same, or from different demes (see Figure B.6). We find that in case of a linear increase a coarser resolution results in a somewhat higher differentiation between cells. This happens

because the increase in population size at a time is bigger than in case of a finer resolution, hence the already existing differences become more pronounced.

On the other hand, limited knowledge of demographic history, with respect to the total simulated time, imposes a major limitation on the simulated scenarios. When this time is too short compared to the effective population sizes (either small number of time points, or short time step), lineages do not coalesce within this time and the coalescence time is determined mainly by the non-spatial (standard) coalescent process in a panmictic ancestral population (see (4.4). Figure 4.18 shows that insufficient simulation time (short time between time points) results in wrong estimates. We compared the theoretical predictions of effective population sizes with the mean coalescence times. The means used here were calculated directly from the simulations. N_e is estimated as half of the mean coalescence time. According to Cox et al. (2002), the effective population sizes in a a $2D$ stepping stone model and in an island model are:

$$N_e = \frac{L^2 \log(L)}{4\pi\sigma^2\nu}, \qquad N_e = Ns\left(1 + \frac{(s-1)^2}{4N\nu s^2}\right), \tag{4.9}$$

where $\nu = 4m$ (the total migration rate for each grid cell), and s is the number of demes in the island model. In Figure 4.18 one can see that to obtain good estimates of F_{ST} values one can run the simulations much shorter than to obtain correct N_e values. This is the result of F_{ST} being calculated as a ratio, therefore the problematic part simply cancels out. One could use formula 4.4 to correct for the large amount of simulations that did not finish in case of short time steps, however, it does retain any additional information for us about whether we see any effect of the spatial structure, even if it persisted only for a short time. The formula would simply predict that the effective population size is the census size at time 0, which we know if assuming that the population spent a much larger time in a state of being a single panmictic population.

4.4 Discussion

4.4.1 gridCoal

Here we presented an efficient coalescent simulator for a two dimensional stepping stone model. Individual based, forward time simulation models exist that allow for performing

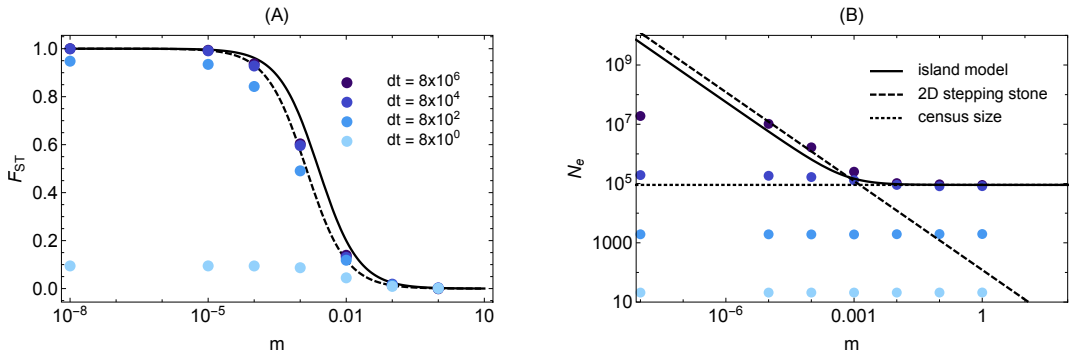


Figure 4.18: A) Global F_{ST} values calculated for a 30×30 grid, in case of different times at which the lineages are pulled into a non-spatial ancestral population. B) Effective population size (N_e) as predicted for the 2D stepping stone and island models, and from the simulations, as half of the mean coalescence time. Note, that we used the mean of the simulations, regardless how many lineages did not coalesce before the ancestral state started. Correcting for this would only increase N_e to the census size for the two shorter values of dt, thus would be still uninformative. This shows that F_{ST}-s can be estimated relatively well, even for short dt values, while getting better estimates for the effective population size requires longer runs.

spatially explicit simulations (such as Easypop, Nemo, CDpop, SLIM). Even though it is possible to simulate the spatial dependence between individuals in a forward time simulation framework, such an approach is not practically feasible for two main reasons. First, due to computational reasons. Incorporating realistic demographic and spatial processes is inevitably limited by the computational burden of tracking a large number of individuals in every generation. This is particularly true for large populations and long dispersal distances (Battey et al., 2020; Ray et al., 2010). Second, forward simulators are useful, when sufficient background knowledge is available about the biological system, such as population census size, mating system, life history, and population allele frequencies.

gridCoal may be particularly useful for simulating spatially explicit null models when it is sufficient to simulate summary statistics that can be derived from coalescent times, i.e. gene diversity (within deme coalescence time or T_w), the strength of population structure (F_{ST}) and genetic similarity between pairs of demes (F^*). The efficiency of gridCoal stems from the optimized continuous time approximation of the coalescent process in *msprime*, as well as from the fact that only coalescence times and no genetic data are simulated. gridCoal may be used for similar questions as SPLATCHE3 (Currat et al., 2019). The following options are available in SPLATCHE3, but not implemented yet in gridCoal: (i)

simulation of long-distance dispersal events could be implemented by directly modifying the migration matrix, (ii) simulation of heterogeneous migration rates in space and time, and (iii) simulation of population extinction and colonization of new patches. One main shortcoming of gridCoal is that coalescence times are closely related to gene diversity and F_{ST} only when the mutation rate is low, and migration is possible to neighboring demes only (Slatkin, 1985). The assumption about the mutation rate can be a problem for practical applications using microsatellites. The assumption about the migration to neighboring cells may also be a limitation for some organisms, such as migrating animals or wind-dispersed plants. In these cases, it is preferable to simulate genetic data and calculate the necessary summary statistics from these (see for example the Appendix of Battey et al. (2020)).

From the point of view of a gridCoal end user the choice of several parameters has to be carefully to considered. First, the choice of dt (time between two time points when the population size is specified) is one of the most important practical consideration. Too short dt mean that lineages would not have time to coalesce in the spatially explicit phase of the simulations, and too long dt would imply an unnecessarily long running time. We suggest users to perform test simulations to choose dt if it is necessary for the research question that all lineages coalesce during the spatially explicit phase. In contrast, if the question concerns a particular organism with a given generation time and at a particular time period, the choice of dt should be adjusted accordingly. For example, for the LPX-Bern motivated scenarios (Figure 4.4), we used a dt=100, 220 time steps, and 25 years of generation time, which takes the ancestral population back to the Last Glacial Maximum (LGM, 21 kya). Under these conditions, roughly 25% of the lineages did not coalesce before the LGM, which is potentially partly reliable for the similarities between the diversity maps (Figure 4.15). Second, two lineages per deme were sufficient to estimate the summary statistics based on coalescence times, however, for simulating genetic data, it is desired to simulate as many lineages as sampled individuals. We did not test gridCoal using more than two lineages or for simulating genetic data. Third, grid size is an important practical limitation for spatially explicit simulations. Many theoretical studies use a torus to avoid edge effects. 50×50 grid appears sufficient to approximate well a continuous space process (Battey et al., 2020).

4.4.2 The role of spatial and temporal autocorrelation in genetic data

Climatic oscillations of the Quaternary have substantially changed the distribution and demographic history of most species (de Lafontaine et al., 2018). During warm glacial interstadials, and, in particular, since the LGM, most temperate species colonized the northern hemisphere via successive foundation events. Thus, spatially explicit null models would be required for landscape genetic studies of most species. However, inference of the full demographic history of a species in space and time solely from genetic data is extremely challenging (e.g. Beichman et al., 2018). Due to the infinitely many possible scenarios, even simulation-based likelihood-free approaches, such as spatially explicit coalescent simulations (Ray et al., 2010) coupled with Approximate Bayesian Computation become computationally challenging and may lead to inaccurate inferences (Beaumont, 2011). One of the current motivations to study and develop methods to make inferences across both space and time is the widespread development of ancient DNA techniques, so samples are available both across space and time.

We found that scenarios involving spatial structure, including those with high spatial variance (Figure 4.7), provided a relatively good fit to infinite island model prediction. F_{ST} showed the strongest deviation from the island model in the case the homogeneous map, and the tightest match in the case of the uniform map (Figure 4.7). This is because high spatial variance in the local effective population size breaks the spatial correlation between demes, so the within deme coalescence times are spatially uncorrelated. However, this scenario is unlikely to be realistic, and this result does not mean that the infinite island model is a good approximation for actual biological samples (Meirmans, 2012; Wang & Whitlock, 2003). In contrast, the stepping stone model accounts for the correlation between neighboring demes, thus F_{ST} from the homogeneous map provided the closest fit, and F_{ST} from the uniform map the worst fit to the theoretical predictions (Fig 4.7). Most biologically realistic scenarios are situated in between these two theoretical models, with a non uniform degree of correlation between the neighboring subpopulations. Meirmans (2012) pointed out that many of the commonly performed analyses assume spatial independence of the data, or are based on a null model that is inherently non-spatial. An exception is the study of Duforet-Frebourg & Blum (2014). The authors developed

a testing procedure to evaluate if population density or locally limited migration rate generated nonstationary of isolation by distance patterns, and use simulated data under the stepping stone model to generate a null distribution for their test statistics.

4.4.3 Spatial null models for detecting loci under selection?

Understanding the genetic architecture of ecologically important traits and detecting loci under selection has become a major objective of evolutionary biology with the wide-spread appearance of population-level genomic data (Vitti et al., 2013). The ability to detect selection at individual loci is heavily dependent on the assumptions of a neutral model, such as the population size and its historical fluctuations, mating system, and also the number and genetic architecture of the loci involved (François et al., 2016; Hoban et al., 2016). The non-spatial, infinite island model has almost exclusively been used as a biological null model, notably, to detect isolation by distance (Wright, 1943) and to detect loci under selection (Lewontin & Krakauer, 1973; Beaumont & Balding, 2004). Indeed, a wide-range of F_{ST}-based tests of selection have been developed during the past two decades. The main motivation for developing new methods has been accounting for the confounding signal of demography, thus trying to develop more biologically meaningful null models than the island model used by Beaumont & Nichols (1996). Authors either suggested a more complex demographic model (e.g. Foll & Gaggiotti, 2008; Excoffier et al., 2009), or, proposed estimating population structure from all loci to generate a null expectation (e.g. Frichot et al., 2013; Duforet-Frebourg et al., 2014). However, assuming a specific, and incorrect, model of demographic history can result in extremely high false positive rates for detecting loci under selection (Lotterhos & Whitlock, 2015). We encourage the more widespread use of spatially continuous and stepping stone models for simulating more biologically appropriate null models that incorporate the effects of spatial and temporal autocorrelation in allele frequencies when detecting the signature of selection from genomic data (Battey et al., 2020). Exploring the precise connection between continuous space forward-time demographies and stepping stone backward-time models also remains a task for the future (Kelleher et al., 2014).

5 Future directions

This thesis describes three projects exploring how evolutionary and ecological processes shape patterns of diversity and evolutionary outcomes. The three projects attack this central theme by focusing on different questions and tackling them with different mathematical tools. Each project raises new questions and suggests directions for future research. We formulate such directions for the project described in Chapter 4 in Section 5.1 and the projects from Chapters 2 and 3 in Section 5.2.

5.1 Population structure

Understanding the role of explicit spatial structure and demographic changes in time is important for modelling real biological organisms (Battey et al., 2020). They are (mostly) distributed across wider ranges of habitats instead of forming a single panmictic population, and experience environmental changes resulting in population declines, expansions, extinctions, and recolonizations. Such effects are known to have a profound effect on diversity, yet most models used for statistical inference of demography rely on non-spatial null-models such as the island model (e.g. Leffler et al., 2012; Sousa & Hey, 2013; Rousseau et al., 2017).

In Chapter 4, we presented a coalescent-based simulation tool that can incorporate realistic spatial structure and temporal changes in demography. Most of my work here explores how demographic processes and different spatial arrangements affect current levels of diversity. The long-term goal of this project, however, is to provide a more

realistic null-model that can be used for inference methods and to detect loci under selection. This goal is ambitious and the work presented here is not complete, yet it adds a step towards a more realistic combined treatment of space and time.

One major shortcoming of the simulation presented in this thesis is the way it treats colonization. The grid cell that is to be colonized needs to neighbor an occupied cell, which is an unrealistic restriction if one considers species that can disperse far distances. Currently, an empty cell gets colonized from the neighboring cells, moving lineages proportionately to theirs sizes. However, one may wish to use a model where colonization happens from a single cell, which at this time is not possible within the used software. We also did not use any other form of migration than to the immediate neighbors, however, different types of migrations schemes can lead to different patterns in diversity (Ibrahim et al., 1996).

Slatkin (1977) suggested an extinction-colonization model that considers n demes connected by migration. The demes go extinct or get colonized with a given rate. He suggested two different models of colonization, the *migrant-pool model*, where the founders arrive from different demes, and the *propagule-pool model*, where they originate from the same deme. Pannell & Charlesworth (2000) pointed out that the total and within coalescence time in the propagule-pool model can be significantly reduced compared to the simple island model. This effect is less severe for the migrant-pool model. They also found that the rate demes go extinct with has a different effect on F_{ST} depending on the type of colonization model. Which colonization model to use may depend on the physical sizes of the grid or the given species, nevertheless, adding different colonization schemes to our simulation too would be important. Extinction-colonization is known to reduce diversity, within coalescence times, and total coalescence times. However, skewed offspring numbers can lead to similar conclusions in the absence of such demographic history (Eldon & Wakeley, 2009). How can we distinguish between scenarios with similar effects? Skewed offspring numbers can be observed in many viruses, fungi, and plant species. Modelling such organisms may require the usage of a multiple-merger coalescence process instead of the standard coalescence model we used here (Sackman et al., 2019).

The fixation index, F_{ST}, is robust to many demographic effects (Beaumont, 2005), and this is consistent with our observations. This can be explained by the separation-of-timescales approximation used in structured coalescence models (Wakeley, 1999). The

genealogy is separated into a short scattering phase (more recent history) and a long collecting phase (ancestral state, which behaves as a single population). There are relatively few mutations in the scattering phase, therefore F_{ST} depends on the collecting-phase allele frequencies. This suggests that simulating genetic data in addition to coalescence times may provide statistics that are more sensitive to demographic changes than F_{ST}.

5.2 Local adaptation

Metapopulation models have long been used to understand how population subdivision affects neutral diversity (Wright, 1931; Whitlock & Barton, 1997; Charlesworth et al., 2003) and to study extinction-colonization dynamics of a fragmented habitat (Hanski & Mononen, 2011). Assuming heterogeneous environments, such models can also be used to understand the process of evolutionary rescue (Uecker et al., 2014). These models usually incorporate dependencies between population sizes and the mean fitness (Barton & Rouhani, 1993; Hanski & Mononen, 2011), however, a more realistic treatment of these variables is called for. Another important feature of these models is stochasticity: while most of these models incorporate some stochasticity: drift (Whitlock & Barton, 1997) or population fluctuation (Lande et al., 2003) they do not consider them together.

To remedy this shortcoming, we used a model based on the diffusion approximation (Rouhani & Barton, 1993; Banglawala, 2010; Barton & Etheridge, 2018) to describe the joint stochastic behavior of demography and genetics in a heterogeneous environment. The model accounts for drift and population size fluctuations, and assumes polygenic architecture for local adaptation. Finding the joint stationary distribution allowed us to identify the conditions under which local adaptation under divergent selection in a rare habitat is possible. The specific details and results of this work can be found in Chapter 3. Without going into details, we will mention a few results and insights we gained that lead to new research questions and possible extensions of the model.

The most important conclusion is that hard selection makes local adaptation in the rare habitat more difficult than soft selection. This effect becomes more pronounced the more loci are involved: the number of selective deaths depends on the total effect of selection (LS). Due to hard selection, the population size is reduced by the mean genetic load, therefore the amount of drift in the population will be defined by this reduced

population size, making the individual loci more prone to swamping. We found that when selection is weak ($LS < 1$), an increase in the strength of selection helps the population to better withstanding maladapted gene flow, while in case of stronger selection ($LS > 1$), extinctions are possible even when the frequencies of locally adapted alleles have high expectation: even a little maladaptation can result in high load causing the population to collapse.

As the number of loci increases, the coupling between population size and allele frequency increases, making adaptation less likely. This happens due to the extreme form of the environmental heterogeneity: each allele is either advantageous or disadvantageous. To mediate this effect, instead of using directional selection, one could introduce the model with stabilizing selection acting on a quantitative trait. In this case, the migration load would not depend on the number of loci, allowing adaptation to occur for a wider range of parameters.

Our model assumes linkage equilibrium (requiring weak selection compared to recombination) that enables us to trace only allele frequencies and neglect associations arising due to allele frequencies being different at different loci across the habitats. However, in the case of strong selection (compared to migration and drift), linkage disequilibrium can build up in the population requiring a more realistic treatment of such associations. We also have seen in individual based simulations that adaptation is possible for a wider range of parameters than suggested by the analysis based on the diffusion approximation. Incorporating linkage disequilibrium offers another possible future extension of our model.

We observed in our simulations that the time until equilibrium is reached can be long, therefore understanding how the dynamics evolve through time could provide interesting further insights. For example, how much time does it take until local adaptation is built or completely destroyed? Knowledge of the joint probability distribution of population sizes and allele frequencies at a given time would also allow us to explore how populations react to shifts in the environment: both random or deterministic. Of course, if one can solve the Fokker–Planck Equation of the problem, all these questions can be immediately answered, however, this is in general not possible. A logical extension could therefore be to apply the Dynamic Maximum Entropy method (presented in Chapter 2) to this eco-evolutionary problem and explore the temporal changes both in population size and allele frequency distributions.

Bibliography

Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, *114*(27), 7061–7066.

Antonovics, J., Bradshaw, A. D., & Turner, R. (1971). Heavy metal tolerance in plants. In *Advances in Ecological Research*, vol. 7, (pp. 1–85). Elsevier.

Bairey, E., Kelsic, E. D., & Kishony, R. (2016). High-order species interactions shape ecosystem diversity. *Nature Communications*, *7*(1), 1–7.

Banglawala, N. (2010). *Local adaptation under demographic and genetic fluctuations*. (Doctoral dissertation), The University of Edinburgh.

Barton, N. (2010). What role does natural selection play in speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1547), 1825–1840.

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, *57*(20), 357–376.

Barton, N., & Etheridge, A. (2018). Establishment in a new habitat by polygenic adaptation. *Theoretical Population Biology*, *122*, 110–127.

Barton, N., & Rouhani, S. (1993). Adaptation and the 'shifting balance'. *Genetics Research*, *61*(1), 57–74.

Barton, N. H., & De Cara, M. A. R. (2009). The evolution of strong reproductive isolation. *Evolution*, *63*(5), 1171–1190.

Barton, N. H., & de Vladar, H. P. (2009). Statistical mechanics and the evolution of polygenic quantitative traits. *Genetics*, *181*(3), 997–1011.

Barton, N. H., Depaulis, F., & Etheridge, A. M. (2002). Neutral evolution in spatially continuous populations. *Theoretical Population Biology*, *61*(1), 31–48.

Barton, N. H., & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics*, *3*(1), 11–21.

Battey, C., Ralph, P. L., & Kern, A. D. (2020). Space is the place: Effects of continuous spatial structure on analysis of population genetic data. *Genetics*, *215*(1), 193–214.

Beaumont, M. A. (2005). Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution*, *20*(8), 435–440.

Beaumont, M. A. (2011). Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 379–406.

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, *13*(4), 969–980.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *263*(1377), 1619–1626.

Beichman, A. C., Huerta-Sanchez, E., & Lohmueller, K. E. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 433–456.

Bialek, W., Cavagna, A., Giardina, I., Mora, T., Silvestri, E., Viale, M., & Walczak, A. M. (2012). Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences*, *109*(13), 4786–4791.

Black, A. J., & McKane, A. J. (2012). Stochastic formulation of ecological models and their applications. *Trends in Ecology & Evolution*, *27*(6), 337–345.

Blanquart, F., Gandon, S., & Nuismer, S. L. (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology*, *25*(7), 1351–1363.

Bod'ová, K., Tkačik, G., & Barton, N. H. (2016). A general approximation for the dynamics of quantitative traits. *Genetics*, *202*(4), 1523–1548.

Boltzmann, L. (1872). Weitere studien über das wörmegleichgewicht unter gasmolekülen. *Sitzungsberichte Akad. Wiss., Vienna, part II*, *66*, 275–370.

Bradburd, G. S., & Ralph, P. L. (2019). Spatial population genetics: It's about time. *Annual Review of Ecology, Evolution, and Systematics*, *50*, 427–449.

Carroll, S. P., & Boyd, C. (1992). Host race radiation in the soapberry bug: Natural history with the history. *Evolution*, *46*(4), 1052–1069.

Caswell, H. (2001). *Matrix population models*. Sunderland, MA.

Charlesworth, B., Charlesworth, D., & Barton, N. H. (2003). The effects of genetic and geographic structure on neutral variation. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 99–125.

Chevin, L.-M., Cotto, O., & Ashander, J. (2017). Stochastic evolutionary demography under a fluctuating optimum phenotype. *The American Naturalist*, *190*(6), 786–802.

Cox, J. T., Durrett, R., et al. (2002). The stepping stone model: New formulas expose old myths. *Annals of Applied Probability*, *12*(4), 1348–1377.

Coyne, J. A., Barton, N. H., & Turelli, M. (1997). Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution*, *51*(3), 643–671.

Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., Thomas, S. M., Smith, J. R., Hintler, G., Duguid, M. C., Amatulli, G., et al. (2015). Mapping tree density at a global scale. *Nature*, *525*(7568), 201–205.

Currat, M., Arenas, M., Quilodran, C., Excoffier, L., & Ray, N. (2019). SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal. *Bioinformatics*.

de Lafontaine, G., Napier, J. D., Petit, R. J., & Hu, F. S. (2018). Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology*, *99*(7), 1530–1546.

de Vladar, H. P., & Barton, N. H. (2011). The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift. *Journal of The Royal Society Interface*, *8*(58), 720–739.

Débarre, F., & Gandon, S. (2011). Evolution in heterogeneous environments: between soft and hard selection. *The American Naturalist*, *177*(3), E84–E97.

Dixit, P. D., Wagoner, J., Weistuch, C., Pressé, S., Ghosh, K., & Dill, K. A. (2018). Perspective: Maximum Caliber is a general variational principle for dynamical systems. *The Journal of Chemical Physics*, *148*(1), 010901.

Dobler, S., & Farrell, B. (1999). Host use evolution in Chrysochus milkweed beetles: evidence from behaviour, population genetics and phylogeny. *Molecular Ecology*, *8*(8), 1297–1307.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, *22*(5), 1185–1192.

Duforet-Frebourg, N., Bazin, E., & Blum, M. G. (2014). Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution*, *31*(9), 2483–2495.

Duforet-Frebourg, N., & Blum, M. G. (2014). Nonstationary patterns of isolation-by-distance: inferring measures of local genetic differentiation with Bayesian kriging. *Evolution*, *68*(4), 1110–1123.

Duforet-Frebourg, N., & Slatkin, M. (2016). Isolation-by-distance-and-time in a stepping-stone model. *Theoretical Population Biology*, *108*, 24–35.

Edelaar, P., & Bolnick, D. I. (2012). Non-random gene flow: an underappreciated force in evolution and ecology. *Trends in Ecology & Evolution*, *27*(12), 659–665.

Eldon, B., & Wakeley, J. (2009). Coalescence times and F_{ST} under a skewed offspring distribution among individuals in a population. *Genetics*, *181*(2), 615–629.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422.

Engen, S., Bakke, Ø., & Islam, A. (1998). Demographic and environmental stochasticity-concepts and definitions. *Biometrics*, (pp. 840–846).

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, *9*(10), e1003905.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285.

Excoffier, L., Novembre, J., & Schneider, S. (2000). SIMCOAL: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, *91*, 506–509.

Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, *10*(1), 253–280.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*(594-604), 309–368.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*(2), 977–993.

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, *25*(1), 104–120.

François, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Molecular Ecology*, *25*(2), 454–469.

Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M., Dudash, M. R., Fenster, C. B., Lacy, R. C., & Sunnucks, P. (2017). *Genetic management of fragmented animal and plant populations*. Oxford University Press.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*, 1687–1699.

Gause, G. F., et al. (1934). Experimental analysis of Vito Volterra's mathematical theory of the struggle for existence. *Science*, *79*(2036), 16–17.

Gavrilets, S. (2003). Perspective: models of speciation: what have we learned in 40 years? *Evolution*, *57*(10), 2197–2215.

Ghosh, K., Dixit, P. D., Agozzino, L., & Dill, K. A. (2020). The maximum caliber variational principle for nonequilibria. *Annual Review of Physical Chemistry*, *71*, 213–238.

Glover, K. A., Solberg, M. F., McGinnity, P., Hindar, K., Verspoor, E., Coulson, M. W., Hansen, M. M., Araki, H., Skaala, Ø., & Svåsand, T. (2017). Half a century of genetic interaction between farmed and wild Atlantic salmon: Status of knowledge and unanswered questions. *Fish and Fisheries*, *18*(5), 890–927.

Gomi, T., & Takeda, M. (1996). Changes in life-history traits in the fall webworm within half a century of introduction to japan. *Functional Ecology*, (pp. 384–389).

Gomulkiewicz, R., & Holt, R. D. (1995). When does evolution by natural selection prevent extinction? *Evolution*, *49*(1), 201–207.

Gomulkiewicz, R., Holt, R. D., & Barfield, M. (1999). The effects of density dependence and immigration on local adaptation and niche evolution in a black-hole sink environment. *Theoretical Population Biology*, *55*(3), 283–296.

Gonzalez, A., Ronce, O., Ferriere, R., & Hochberg, M. E. (2013). Evolutionary rescue: an emerging focus at the intersection between ecology and evolution.

Govaert, L., Fronhofer, E. A., Lion, S., Eizaguirre, C., Bonte, D., Egas, M., Hendry, A. P., De Brito Martins, A., Melián, C. J., Raeymaekers, J. A., et al. (2019). Eco-evolutionary feedbacks – Theoretical models and perspectives. *Functional Ecology*, *33*(1), 13–30.

Grant, P. R., & Grant, B. R. (2006). Evolution of character displacement in Darwin's finches. *Science*, *313*(5784), 224–226.

Guillot, G., Estoup, A., Mortier, F., & Cosson, J. F. (2005). A spatial statistical model for landscape genetics. *Genetics*, *170*(3), 1261–1280.

Haldane, J. B. S. (1956). The relation between density regulation and natural selection. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, *145*(920), 306–308.

Hanski, I., & Gilpin, M. (1991). Metapopulation dynamics: brief history and conceptual domain. *Biological journal of the Linnean Society*, *42*(1-2), 3–16.

Hanski, I., & Mononen, T. (2011). Eco-evolutionary dynamics of dispersal in spatially heterogeneous environments. *Ecology Letters*, *14*(10), 1025–1034.

Hanski, I. A., Gilpin, M. E., & McCauley, D. E. (1997). *Metapopulation biology*, vol. 454. Elsevier.

Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics*, vol. 116. Sinauer Associates Sunderland, MA.

Hassell, M., & Comins, H. (1976). Discrete time models for two-species competition. *Theoretical Population Biology*, *9*(2), 202–221.

Hendry, A. P. (2016). *Eco-evolutionary dynamics*. Princeton University Press.

Henle, K., Sarre, S., & Wiegand, K. (2004). The role of density regulation in extinction processes and population viability analysis. *Biodiversity & Conservation*, *13*(1), 9–52.

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, *405*(6789), 907–913.

Hey, J. (1991). A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. *Theoretical Population Biology*, *39*(1), 30–48.

Hey, J., & Nielsen, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, *104*(8), 2785–2790.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379–397.

Holt, R. D., & Gomulkiewicz, R. (1997). How does immigration influence local adaptation? A reexamination of a familiar paradigm. *The American Naturalist*, *149*(3), 563–572.

Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology*, *23*, 183–201.

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337–338.

Hudson, R. R., Kreitman, M., & Aguadé, M. (1987). A test of neutral molecular evolution. *Genetics*, *116*, 153–159.

Ibrahim, K. M., Nichols, R. A., & Hewitt, G. M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, *77*(3), 282–291.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620.

Jaynes, E. T. (1980). The minimum entropy production principle. *Annual Review of Physical Chemistry*, *31*(1), 579–601.

Joseph, T., Hickerson, M., & Alvarado-Serrano, D. (2016). Demographic inference under a spatially continuous coalescent model. *Heredity*, *117*(2), 94–99.

Karlin, S., & Taylor, H. E. (1981). *A second course in stochastic processes*. Elsevier.

Kawecki, T. J. (2000). The evolution of genetic canalization under fluctuating selection. *Evolution*, *54*(1), 1–12.

Kawecki, T. J. (2008). Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 321–342.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241.

Kelleher, J., Etheridge, A., & Barton, N. H. (2014). Coalescent simulation in continuous space: Algorithms for large neighbourhood size. *Theoretical Population Biology*, *95*, 13–23.

Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Computational Biology*, *12*(5), e1004842.

Kimura, M. (1953). "Stepping-Stone" model of population. *Annu. Rep. Natl. Inst. Genet*, *3*, 62–63.

Kimura, M. (1955). Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences of the United States of America*, *41*(3), 144.

Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, *61*(4), 893.

Kimura, M., & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, *49*(4), 561.

Kingman, J. F. (1982a). On the genealogy of large populations. *Journal of Applied Probability*, (pp. 27–43).

Kingman, J. F. C. (1978). A simple model for the balance between selection and Mutation. *J. Appl. Prob.*, *15*, 1–12.

Kingman, J. F. C. (1982b). The coalescent. *Stochastic processes and their applications*, *13*(3), 235–248.

Kinnison, M. T., & Hendry, A. P. (2001). The pace of modern life ii: From rates of contemporary microevolution to pattern and process. *Genetica*, *112*(20), 145–164.

Kitada, S., Nakamichi, R., & Kishino, H. (2020). Population-specific F_{ST} and pairwise F_{ST}: History and environmental pressure. *bioRxiv*.

Kokko, H., & López-Sepulcre, A. (2007). The ecogenetic link between demography and evolution: can we bridge the gap between theory and data? *Ecology Letters*, *10*(9), 773–782.

Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetics Research*, *26*(3), 221–235.

Lande, R. (1988). Genetics and demography in biological conservation. *Science*, *241*(4872), 1455–1460.

Lande, R. (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist*, *142*(6), 911–927. URL `http://www.jstor.org/stable/2462690`

Lande, R., Engen, S., & Saether, B.-E. (2003). *Stochastic population dynamics in ecology and conservation*. Oxford University Press on Demand.

Lande, R., Engen, S., & Sæther, B.-E. (2017). Evolution of stochastic demography with life history tradeoffs in density-dependent age-structured populations. *Proceedings of the National Academy of Sciences*, *114*(44), 11582–11590.

Lande, R., & Orzack, S. H. (1988). Extinction dynamics of age-structured populations in a fluctuating environment. *Proceedings of the National Academy of Sciences*, *85*(19), 7418–7421.

Laval, G., & Excoffier, L. (2004). {SIMCOAL} 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, *20*(15), 2485–2487.

Leblois, R., Rousset, F., & Estoup, A. (2004). Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. *Genetics*, *166*(2), 1081–1092.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Segurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology*, *10*(9), e1001388.

Leinonen, P. H., Remington, D. L., & Savolainen, O. (2011). Local adaptation, phenotypic differentiation, and hybrid fitness in diverged natural populations of Arabidopsis lyrata. *Evolution*, *65*(1), 90–107.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183 – 189.

Leslie, P. (1966). The intrinsic rate of increase and the overlap of successive generations in a population of guillemots (uria aalge pont.). *Journal of Animal Ecology*, *35*(2), 291–301.

Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist*, *15*(3), 237–240.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, *74*, 175–195.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496.

Lion, S. (2018). Theoretical approaches in evolutionary ecology: environmental feedback as a unifying perspective. *The American Naturalist*, *191*(1), 21–44.

Lotka, A. J. (1924). Elements of mathematical biology.

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*(5), 1031–1046.

Louthan, A. M., Doak, D. F., & Angert, A. L. (2015). Where and when do species interactions set range limits? *Trends in Ecology & Evolution*, *30*(12), 780–792.

Malécot, G. (1948). Mathématiques de l'hérédité.

Malécot, G. (1951). Un traitement stochastique des problems lineaires (mutation, linkage, migration) en genetique de population. *Ann. Univ. Lyon. Sci. Sec.*, *14*, 79–117.

Malécot, G. (1955). Remarks on decrease of relationship with distance, following paper by M. Kimura. *Cold Spring Harbor Symposium Quantitative Biololgy*, *20*, 52–53.

Malécot, G. (1975). Heterozygosity and relationship in regularly subdivided populations. *Theoretical Population Biology*, *8*(2), 212–241.

Malthus, T. R. (1878). *An essay on the principle of population: Or, a view of its past and present effects on human happiness, with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions*. London, Reeves and Turner.

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, *18*(4), 189–197.

Mangel, M., & Tier, C. (1993). Dynamics of metapopulations with demographic stochasticity and environmental catastrophes. *Theoretical Population Biology*, *44*(1), 1–31.

Maruyama, T. (1977). Maruyama lecture notes in biomathematics. 17. Stochastic problems in population genetics.

Mateo, L., Rech, G. E., & González, J. (2018). Genome-wide patterns of local adaptation in western european drosophila melanogaster natural populations. *Scientific Reports*, *8*(1), 1–14.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, *261*(5560), 459–467.

Mayr, E. (1942). Systematics and the origin of species–Columbia Univ. *Press, New York*.

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846.

Mora, T., Walczak, A. M., Bialek, W., & Callan, C. G. (2010). Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences*, *107*(12), 5405–5410.

Nagylaki, T. (1980). The strong-migration limit in geographically structured populations. *Journal of Mathematical Biology*, *9*(2), 101–114.

Nåsell, I. (2001). Extinction and quasi-stationarity in the Verhulst logistic model. *Journal of Theoretical Biology*, *211*(1), 11–27.

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283–292.

Nordborg, M. (1997). Structured coalescent processes on different time scales. *Genetics*, *146*(4), 1501–1514.

Notohara, M. (1990). The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology*, *29*(1), 59–75.

Ovaskainen, O., & Meerson, B. (2010). Stochastic models of population extinction. *Trends in Ecology & Evolution*, *25*(11), 643–652.

Pannell, J. R., & Charlesworth, B. (2000). Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *355*(1404), 1851–1864.

Polechová, J. (2018). Is the sky the limit? on the expansion threshold of a species' range. *PLoS Biology*, *16*(6), e2005372.

Polechová, J., & Barton, N. H. (2015). Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences*, *112*(20), 6401–6406.

Prügel-Bennett, A., & Shapiro, J. (1997). An analysis of genetic algorithms using statistical mechanics. *Physica D: Nonlinear Phenomena*, *104*, 75–114.

Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, *26*(23), 2993–2994.

Robertson, A. (1961). Inbreeding in artificial selection programmes. *Genetical Research*, *2*(2), 189–194.

Ronce, O., & Kirkpatrick, M. (2001). When sources become sinks: migrational meltdown in heterogeneous habitats. *Evolution*, *55*(8), 1520–1531.

Ross-Ibarra, J., Wright, S. I., Foxe, J. P., Kawabe, A., DeRose-Wilson, L., Gos, G., Charlesworth, D., & Gaut, B. S. (2008). Patterns of polymorphism and demographic history in natural populations of arabidopsis lyrata. *PloS One*, *3*(6), e2411.

Rouhani, S., & Barton, N. (1993). Group selection and the 'shifting balance'. *Genetics Research*, *61*(2), 127–135.

Rousseau, E., Moury, B., Mailleret, L., Senoussi, R., Palloix, A., Simon, V., Valière, S., Grognard, F., & Fabre, F. (2017). Estimating virus effective population size and selection without neutral markers. *PLoS Pathogens*, *13*(11), e1006702.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, *145*(4), 1219–1228.

Ruosch, M., Spahni, R., Joos, F., Henne, P. D., van der Knaap, W. O., & Tinner, W. (2016). Past and future evolution of *Abies alba* forests in Europe - comparison of a dynamic vegetation model with palaeo data and observations. *Global Change Biology*, *22*, 727–740.

Sachdeva, H. (2019). Effect of partial selfing and polygenic selection on establishment in a new habitat. *Evolution*, *73*(9), 1729—1745.

Sackman, A. M., Harris, R. B., & Jensen, J. D. (2019). Inferring demography and selection in organisms characterized by skewed offspring distributions. *Genetics*, *211*(3), 1019–1028.

Sæther, B.-E., & Engen, S. (2019). Towards a predictive conservation biology: the devil is in the behaviour. *Philosophical Transactions of the Royal Society B*, *374*(1781), 20190013.

Salguero-Gómez, R., Jones, O. R., Archer, C. R., Buckley, Y. M., Che-Castaldo, J., Caswell, H., Hodgson, D., Scheuerlein, A., Conde, D. A., Brinks, E., et al. (2015). The COMPADRE Plant Matrix Database: an open online repository for plant demography. *Journal of Ecology*, *103*(1), 202–218.

Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, *440*(7087), 1007–1012.

Schoener, T. W. (2011). The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science*, *331*(6016), 426–429.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shirk, A. J., & Cushman, S. A. (2014). Spatially-explicit estimation of Wright's neighborhood size in continuous populations. *Frontiers in Ecology and Evolution*, *2*, 62.

Shoemaker, L. G., Sullivan, L. L., Donohue, I., Cabral, J. S., Williams, R. J., Mayfield, M. M., Chase, J. M., Chu, C., Harpole, W. S., Huth, A., et al. (2020). Integrating the underlying structure of stochasticity into community ecology. *Ecology*, *101*(2), e02922.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., et al. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, *9*(2), 161–185.

Slatkin, M. (1977). Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology*, *12*(3), 253–262.

Slatkin, M. (1985). Gene flow in natural populations. *Annual Review of Ecology and Systematics*, *16*(1), 393–430.

Slatkin, M. (1987). The average number of sites separating dna sequences drawn from a subdivided population. *Theoretical Population Biology*, *32*(1), 42–49.

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. *Genetics Research*, *58*(2), 167–175.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, *47*(1), 264–279.

Smouse, P. E., Peakall, R., & Gonzales, E. (2008). A heterogeneity test for fine-scale genetic structure. *Molecular Ecology*, *17*(14), 3389–3400.

Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, *14*(6), 404–414.

Strobeck, C. (1987). Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics*, *117*(1), 149–153.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics*, *123*(3), 585–595.

Takahata, N. (1988). The coalescent in two partially isolated diffusion populations. *Genetics Research*, *52*(3), 213–222.

Thompson, J. N. (1998). Rapid evolution as an ecological process. *Trends in Ecology & Evolution*, *13*(8), 329–332.

Tufto, J. (2001). Effects of releasing maladapted individuals: a demographic-evolutionary model. *The American Naturalist*, *158*(4), 331–340.

Tuljapurkar, S. (1989). An uncertain life: demography in random environments. *Theoretical Population Biology*, *35*(3), 227–294.

Tuljapurkar, S. (1990). Delayed reproduction and fitness in variable environments. *Proceedings of the National Academy of Sciences*, *87*(3), 1139–1143.

Tuljapurkar, S. (2013). *Population dynamics in variable environments*, vol. 85. Springer Science & Business Media.

Turelli, M. (1977). Random environments and stochastic calculus. *Theoretical Population Biology*, *12*(2), 140–178.

Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology*, *25*(2), 138–193.

Turelli, M., Barton, N. H., & Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology & Evolution*, *16*(7), 330–343.

Uecker, H., Otto, S. P., & Hermisson, J. (2014). Evolutionary rescue in structured populations. *The American Naturalist*, *183*(1), E17–E35.

Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.*, *10*, 113–126.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, *47*, 97–120.

Wakeley, J. (1998). Segregating sites in Wright's island model. *Theoretical Population Biology*, *53*(2), 166–174.

Wakeley, J. (1999). Nonequilibrium migration in human history. *Genetics*, *153*(4), 1863–1871.

Wakeley, J. (2009). Coalescent theory. *Roberts & Company*.

Wallace, B. (1975). Hard and soft selection revisited. *Evolution*, *29*(3), 465–473.

Wang, J., & Whitlock, M. C. (2003). Estimating effective population size and migration rates from genetic samples over space and time. *Genetics*, *163*(1), 429–446.

Waples, R. S., Do, C., & Chopelet, J. (2011). Calculating n_e and n_e/n in age-structured populations: a hybrid Felsenstein-Hill approach. *Ecology*, *92*(7), 1513–1522.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, (pp. 1358–1370).

Whitlock, M. C., & Barton, N. H. (1997). The effective size of a subdivided population. *Genetics*, *146*(1), 427–441.

Whitlock, M. C., & Gomulkiewicz, R. (2005). Probability of fixation in a heterogeneous environment. *Genetics*, *171*(3), 1407–1417.

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4\text{nm} + 1)$. *Heredity*, *82*(2), 117–125.

Wilson, E. O., & MacArthur, R. H. (1967). The theory of island biogeography.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*, *1*, 356—-366.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114.

Wright, S. (1946). Isolation by distance under diverse systems of mating. *Genetics*, *31*(1), 39.

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*(1), 323–354.

Wright, S. (1950). Genetical structure of populations. *Nature*, *166*(4215), 247–249.

A

Polygenic local adaptation in metapopulations: a stochastic eco-evolutionary mode

A.1 Miscellaneous analytical results

Population size distribution with no selection. In the absence of selection, the joint distribution $\Psi(N, \{p_j\})$ in equation 3 (main text) can be written as the product of independent distributions for N and each of the p_j. At each locus, we simply have Wright's distribution for $\psi[p]$ (Wright, 1932) under the infinite island model. The population size distribution $\psi[N]$ is given by Equation 5 (main text). Integrating over $\psi[N]$ yields the expected population size $\mathbb{E}(N)$ in any deme, given the mean $\overline{N}$ across the metapopulation.

The equilibrium can be found by equating $\mathbb{E}(N)$ with $\overline{N}$; there is always a solution $\mathbb{E}(N) = \overline{N} = 0$, which corresponds to global extinction. Above a critical migration rate M_c, this solution becomes unstable, and a second solution, corresponding to a non-zero population size emerges. The threshold M_c is that migration rate for which the solution $\overline{N} = 0$ just becomes unstable, i.e., where $\partial\mathbb{E}(N)/\partial\overline{N}|_{\overline{N}=0} = 1$. This yields the following equation for M_c: $e^{(1-M_c)^2\zeta}\sqrt{\pi\zeta}(1 - Erf[-(1-M_c)\sqrt{\zeta}])M_c = 1$. For large ζ, the threshold M_c must be correspondingly small, such that we need only retain first order terms in M_c, and can further approximate $Erf[-(1-M_c)\sqrt{\zeta}] \approx 1$. This yields $M_c \approx e^{-\zeta}/(2\sqrt{\pi\zeta})$ for large ζ. Only for $M > M_c$ is the expected population size non-zero.

The shape of the distribution $\psi[N]$ is governed by the parameter $2\zeta M\overline{N}$, which is the

average number of immigrants per generation per deme. For $2\zeta M\overline{N} < 1$, the distribution is bimodal: a fraction of demes support large populations with $N \sim 1 - M$, while the remaining fraction is close to extinction. Conversely, for $2\zeta M\overline{N} > 1$, the distribution is unimodal: all demes support large populations, i.e., are close to carrying capacity.

We obtain an explicit expression for the migration threshold M_* for which $2\zeta M_*\overline{N}(M_*)$ equals 1. Substituting the condition $2\zeta M_*\overline{N}(M_*) = 1$ into the expression for $\mathbb{E}(N|\overline{N})$ and then equating with $\overline{N}$ gives us an explicit expression for $\overline{N}(M_*)$. For large ζ, this is simply $\overline{N} \approx 1 - M_*$. Substituting this into $2\zeta M_*\overline{N}(M_*) = 1$, yields $M_* \approx \frac{1}{2\zeta}\left(1 + \frac{1}{2\zeta}\right)$. Thus, below this second threshold, i.e., for $M_c < M < M_*$, the stationary distribution of population sizes is bimodal, and there is a turnover of occupied vs. extinct demes due to frequent extinction and recolonization events, even at equilibrium.

Allele frequency distribution and local adaptation under soft selection. In the limit $\zeta \to \infty$, $S, M \to 0$ (with $\zeta S = Ks$ and $\zeta M = Km$ held constant), we recover the soft selection model, in which each deme is at carrying capacity ($N = n/K = 1$), irrespective of adaptation. Allele frequencies now evolve independently of each other (assuming LE). In a deme where the favored allele has selective advantage s, the allele frequency distribution is:

$$\psi[p] = \frac{1}{Z_1} p^{2Km\overline{p}-1} q^{2Km\overline{q}-1} \exp(2Ksp), \tag{A.1}$$

$$Z_1 = \Gamma\left(2Km\overline{p}\right)\Gamma\left(2Km\overline{q}\right)\ {}_1\tilde{F}_1\left(2mK\overline{p}; 2mK; 2Ks\right)$$

$$\mathbb{E}\left(p\,|\overline{p}\,, s\right) = 2mK\overline{p}\ \frac{{}_1\tilde{F}_1\left(2mK\overline{p}+1; 2mK+1; 2Ks\right)}{{}_1\tilde{F}_1\left(2mK\overline{p}; 2mK; 2Ks\right)}$$

This distribution (Wright, 1932), can also be obtained by integrating the joint distribution in Equation 3 (main text) over N, and evaluating this integral as $\zeta \to \infty$, $S, M \to 0$ for fixed $\zeta S = Ks$ and $\zeta M = Km$. Note that the above distribution is expressed in terms of unscaled parameters s, m (since scaling by r_0 is not meaningful under soft selection). One can now calculate the equilibrium frequency by equating the average allele frequency in the migrant pool $\overline{p}$ with the expected allele frequency (average over habitats) $\overline{p}^* = \rho\mathbb{E}_1\left(p\,|\overline{p}\,, s\right) + (1-\rho)\mathbb{E}_2\left(p\,|\overline{p}\,, -s\right)$, and solving numerically for the fixed point.

However, we can obtain approximate analytical solution in various limits. First, consider the deterministic limit ($Ks \gg 1$ and $Km \gg 1$), in which allele frequencies are

tightly clustered about the expected value. Then the equilibrium can be found directly from Equation 1B (main text) by setting $\lambda_p = 0$, $n_i = \overline{n} = K$ and $\overline{np} = K\overline{p}$. We can also obtain a prediction for s_c, the critical selection required for a polymorphic equilibrium in the deterministic limit, by noting that just above the threshold s_c (i.e, when a polymorphic equilibrium first appears), the difference between allele frequencies in the two habitats must be very small, and the allele frequency in the common habitat close to 1. Thus retaining only lowest order terms in $1 - p_1$ and $p_1 - p_2$ in Equation 1B (main text) and solving for p_1, p_2 gives $s_c \approx m(1 - 2\rho)$. Note that this prediction just provides a lower bound on s_c, as we do not expect the deterministic analysis (which ignores variance of allele frequencies) to be accurate close to the threshold. Just above the threshold s_c, we have $\overline{q} \to 0$, such that the distribution of allele frequencies is necessarily bimodal, making it necessary to account for drift. In general, we expect drift to further degrade local adaptation in the rare habitat, thus pushing the selection threshold s_c above the deterministic prediction.

The opposite limit is that of weak migration $Km \to 0$, in which any locus (within a deme) are nearly fixed for one or other allele. The rates of fixation towards and away from an allele with advantage s, which is at frequency $\overline{p}$ in the migrant pool, are in the ratio $\int_0^{1/2} \psi[p]dp \Big/ \int_{1/2}^{1} \psi[p]dp$ (for a deme in the rare habitat). Since most of the weight of the distribution $\psi[p]$ is concentrated near $p = 0$ and $p = 1$ (for $Km \ll 1$), we can approximate the integrand in the numerator (or in the denominator) by Taylor expanding $\psi[p]$ near $p = 0$ (or $p = 1$). It then follows that the rates of fixation towards and away from the favored allele are in the ratio $\sim (\overline{p}/\overline{q})e^{2Ks}$. The expected frequency of the locally favored allele is simply the (normalized) probability of fixation of this allele (Equation 7 in the main text).

Semi-deterministic approximation for $LS \lesssim 1$. Under weak coupling ($LS \lesssim 1$), population sizes are approximately normally distributed about the expected value $\mathbb{E}(N)$, irrespective of local adaptation (inset, Figure 3A, main text). This allows us to approximate the stationary distribution using a simpler 'semi-deterministic' approximation: we treat population size as being determined by the expected frequencies (i.e, neglect fluctuations in N), and assume that allele frequencies are distributed according to this deterministic population size.

In practice, we replace the population-averaged log fitness r_g by its expected value $\mathbb{E}[r_g]$ in Equation 1A (main text), neglect the stochastic term λ_n, and solve for the equilibrium size as a function of $\mathbb{E}[r_g]$. Further assuming that migration does not significantly affect population size, so that terms proportional to m in Equation 1A (main text) are also negligible, the scaled population sizes in the two habitats are $N_1^{det} \sim \text{Max}[1 - LS_1\mathbb{E}_1[q], 0]$ and $N_2^{det} \sim \text{Max}[1 - LS_2\mathbb{E}_2[p], 0]$: population sizes are depressed relative to carrying capacity in proportion to the genetic load. To obtain $\mathbb{E}_1[p]$ and $\mathbb{E}_2[p]$, we assume that allele frequencies are distributed as in a model with fixed and deterministic population sizes N_1^{det} and N_2^{det}. Then:

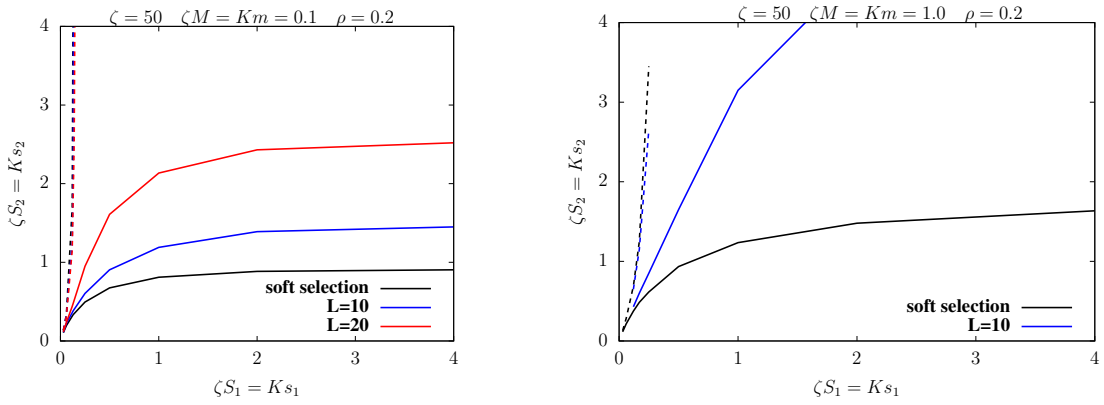
$$\begin{aligned} &\mathbb{E}_i\left(p \,\middle|\, \overline{Np}\,, N_i^{det}, S_i\right) = 2\zeta M\overline{Np}\ \frac{{}_1\tilde{F}_1\left(2\zeta M\overline{Np}+1,\ 2\zeta M\overline{N}+1,\ 2\zeta S_i N_i^{det}\right)}{{}_1\tilde{F}_1\left(2\zeta M\overline{Np},\ 2\zeta M\overline{N},\ 2\zeta S_i N_i^{det}\right)} \\ &N_1^{det} = \text{Max}[1 - LS_1\mathbb{E}_1[q], 0] \quad \text{and} \quad N_2^{det} = \text{Max}[1 - LS_2\mathbb{E}_2[p], 0] \end{aligned} \tag{A.2}$$

As before, the equilibrium $\overline{Np}$ and consequently the expected allele frequencies can be determined iteratively from: $\overline{Np} = (1-\rho)\mathbb{E}_1[p]N_1^{det} + \rho\mathbb{E}_2[p]N_2^{det}$.

A.2 Asymmetric selection across habitats

In the main text, we considered scenarios where alternative alleles are favored in the different habitats, but selection per locally favored allele is the same in each habitat ($S_1 = S_2$). Here, we briefly consider the more general scenario where alternative alleles at a given locus may be selected more or less strongly in one of the two habitats ($S_1 \neq S_2$). For simplicity, we still focus on the case where selective effects are equal across loci within a given habitat (i.e, $S_{1,j} = S_1$ and $S_{2,j} = S_2$ for all j).

Figure A.1 depicts (S_1, S_2) combinations that allow simultaneous local adaptation (i.e., a polymorphic equilibrium) across the two habitats. For a given selection strength per locus, S_1, in the common habitat, we determine two selection thresholds, $S_{2,a}$ and $S_{2,b}$, such that polymorphism is only possible for S_2 (i.e., selection strengths per locus in the rare habitat) lying between these thresholds. For $S_2 < S_{2,a}$, alleles favored in the common habitat fix, while for $S_2 > S_{2,b}$, alleles favored in the rare habitat tend to fix. Note that the threshold $S_{2,b}$ exists, i.e., strong selection in favor of alleles in the rare habitat can drive fixation of such alleles across the entire metapopulation, only if selection per locus in the common habitat is sufficiently weak.



(a) (b)

Figure A.1: Local adaptation in scenarios with asymmetric selection across habitats. Critical selection thresholds $\zeta S_{2,a}$ (solid lines) and $\zeta S_{2,b}$ (dashed lines) in the rare habitat versus selection per locus, ζS_1, in the common habitat for (A) low migration ($\zeta M = 0.1$) and (B) intermediate migration ($\zeta M = 1.0$). Alleles favored in the common habitat fix across the entire metapopulation for $S_2 < S_{2,a}$ (below solid line); alleles favored in the rare habitat fix across the entire metapopulation for $S_2 > S_{2,b}$ (above dashed line); a polymorphic equilibrium corresponding to local adaptation across the two habitats is possible for $S_{2,a} < S_2 < S_{2,b}$, i.e., for parameter combinations lying between the solid and dashed lines. The different colors correspond to different degrees of coupling between population size and mean fitness: soft selection, i.e., no coupling (black), hard selection with $L = 10$ (blue) and $L = 20$ (red) loci; larger L corresponds to stronger coupling (for fixed ζS_1). The region in parameter space allowing for simultaneous local adaptation shrinks with increasing L; at sufficiently high migration, there is no (S_1, S_2) combination for which polymorphism is possible with $L = 20$ loci in (B). Selection thresholds under hard selection are obtained by determining fixed points numerically (Equations 3,4 in main text) using the joint distribution $\Psi[N, p]$; selection thresholds under soft selection are obtained via fixed point determination using the allele frequency distribution $\psi[p]$ (Equation (A.1)).

Figure A.1 shows the two selection thresholds for soft selection (black), and hard selection involving 10 (blue) or 20 (red) divergently selected loci. As before, soft selection is most conducive to simultaneous local adaptation. Under hard selection, the parameter combinations allowing for polymorphism become more restrictive with increasing L: this is consistent with the fact that larger L corresponds to stronger coupling between mean fitness and population size, which increases extinction probabilities in one or other habitat. This effect is exacerbated at higher migration rates (see Figure A.1(b)): in this case, there is no (S_1, S_2) combination for which polymorphism is possible with $L = 20$ loci (for $\zeta = 50$).

A.3 Individual-based simulations

The theoretical predictions detailed above are based on the diffusion framework, which involves three approximations, namely, continuous time, an infinite number of demes and linkage equilibrium (see also Model and Methods). We test the validity of each assumption by comparing analytical predictions with two types of individual-based simulations.

In simulations of the first kind, we simulate a single focal deme belonging to the rare habitat by drawing the number of immigrants per generation from a Poisson distribution with mean $m\overline{N}$, and assigning the '1' allele at (any) locus j in each migrant genome independently with probability $\overline{Np_j}/\overline{N}$. Here, $\overline{Np_j}$ and $\overline{N}$ are the means across the whole metapopulation, and are obtained numerically from the diffusion approximation (as described above). These simulations thus respect the infinite island assumption, since migrants are drawn from an effectively infinite pool, characterized by deterministic $\overline{N}$ and $\overline{Np}$, which are unaffected by stochastic fluctuations within the focal deme. Moreover, choosing the state of each locus in each migrant genotype independently ensures LE within the migrant pool. This set of simulations thus allows us to test the consequences of discrete time and LD *within* demes. For a given set of scaled parameters (ζ, S_1, S_2, M), we expect the results of these simulations to approach the predictions of the diffusion approximation as r_0 is reduced while holding the scaled parameters constant.

In simulations of the second kind, we simulate the full metapopulation consisting of n_D demes, of which $(1-\rho)n_D$ and ρn_D belong to the first and second habitat respectively. These simulations relax the assumption that the migrant pool is infinite and in LE. We test the importance of the infinite deme assumption by varying n_D. We also test the effects of LD within the migrant pool, by comparing simulations with the same scaled parameters but different r_0. LD in the migrant pool (which reflects LD across the whole metapopulation, when demes belonging to different local habitats are simultaneously adapted at many loci) is expected to be insignificant when $s = r_0 S$ is weak relative to recombination. Thus, as in the first set of simulations, we expect the behavior of individual-based simulations to converge to the diffusion approximation (which assumes LE) as r_0 is reduced (for a given combination of scaled parameters). However, the rate of convergence as $r_0 \to 0$ can be quite different for the two set of simulations, for instance, if our analytical results are more sensitive to some assumptions than to others.

Since individual-based simulations of large numbers of demes (e.g., in simulations of the second kind) with large carrying capacities ($\zeta \gg 1$) and polygenic architectures are computationally intensive, we focus on one set of parameters and explore how the critical migration threshold for loss of local adaptation is affected by deviations from these three assumptions for this set.

Continuous time approximation. To test the validity of the continuous time approximation, we simulate a single focal deme in the rare habitat subject to immigration from a migrant pool whose state is assumed to follow the predictions of the diffusion framework (simulations of the first kind). Figure A.2(a) and A.2(b) show the average allele frequency (averaged across all loci) and the average population size (averaged over 100 replicates) from individual-based simulations for various r_0. The unscaled parameters s and m are decreased and the carrying capacity K increased as we reduce r_0, such that the scaled parameters $\zeta = r_0 K$, $S = s/r_0$ and $M = m/r_0$ remain constant. Figure A.2(a) and A.2(b) show that even for $r_0 \sim 0.5$, individual-based simulations only deviate moderately from the diffusion prediction (solid line), with the agreement between the two improving for lower r_0.

The slightly higher frequencies of locally adaptive alleles in simulations (than predicted theoretically) at larger r_0 may reflect the importance of $\mathcal{O}(s^2)$ terms (which the continuous time approximation neglects) but also that the focal deme is not strictly in LE, as individuals with more recent immigrant ancestry will typically carry deleterious alleles at multiple loci (even when deleterious allele frequency is low). This kind of LD can often be adequately accounted for by assuming that immigrant alleles are introduced at an *effective* migration rate that is lower than the actual migration rate, due to their association with low-fitness genetic backgrounds (Barton & Bengtsson, 1986; Banglawala, 2010).

Finite number of demes. We also carry out individual-based simulations of metapopulations with a finite number n_D of demes (simulations of the second kind). Figures A.2(c), A.2(d)) show simulations with 100 versus 500 demes (circles versus triangles) of which 20% belong to the rare habitat, along with theoretical predictions for an infinite number of demes. For any given set of parameters, the average frequency of locally favorable alle-

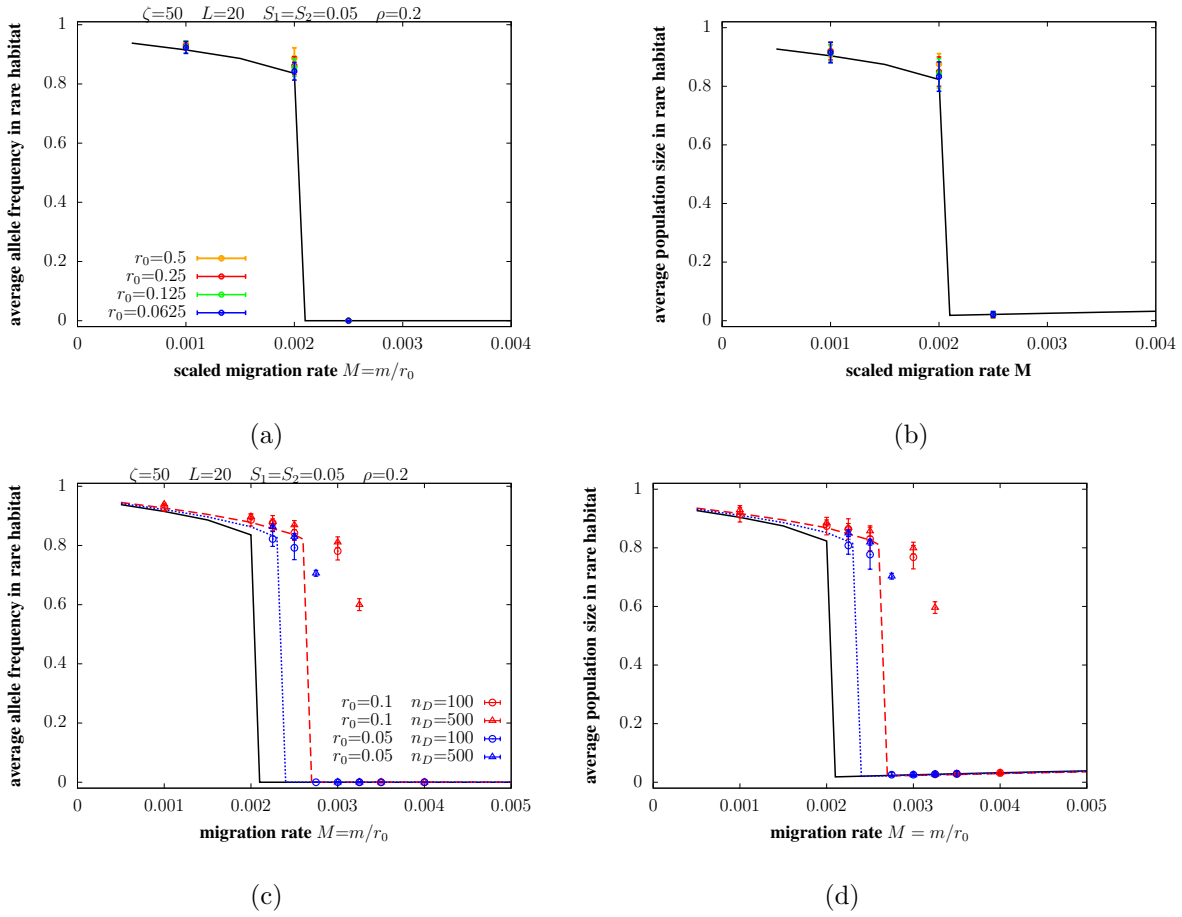


Figure A.2: Comparison of individual-based simulations with predictions of the diffusion approximation (Equations 3 and 4 in the main text). (A)-(B) Average allele frequency and population size in the rare habitat versus scaled migration rate $M = m/r_0$ from individual-based simulations of a focal deme which receives a Poisson-distributed number of immigrants (on average $m\overline{N}$) in each generation. Migrants are sampled from an infinite pool in LE; allele frequencies in this pool are equal to $\overline{Np}/\overline{N}$ (at each locus). The quantities $\overline{N}$ and $\overline{Np}$ are calculated using the diffusion approximation for the infinite island model (Equations 3,4 in the main text). Results of individual-based simulations are shown by circles, the different colors corresponding to different values of r_0. All other unscaled parameters are varied as we vary r_0 such that the scaled parameters remain constant at $\zeta = 50$, $S_1 = S_2 = 0.05$. Solid line shows predictions of the diffusion approximation. (C)-(D) Average allele frequency and population size in the rare habitat versus scaled migration rate $M = m/r_0$ from individual-based simulations of a metapopulation with n_D demes, where circles and triangles represent results for $n_D = 100$ and $n_D = 500$. For each n_D and M, we simulate two sets of unscaled parameters, one with $r_0 = 0.1$ (red) and the other with $r_0 = 0.05$ (blue), both corresponding to the same scaled parameters: $\zeta = 50$ and $S_1 = S_2 = 0.05$. The solid black line shows predictions of the diffusion approximation (assuming LE and infinitely many islands). The dashed lines show the predictions of a more elaborate approximation that partially relaxes the LE assumption by accounting for the first generation selective disadvantage of immigrant alleles originating from different habitats due to association with other negatively selected alleles (see text). Red and blue dashed lines correspond to $r_0 = 0.1$ and $r_0 = 0.05$ respectively. The frequency of the rare habitat is $\rho = 0.2$ and the number of selected loci is $L = 20$ in all simulations.

les in the rare habitat (and consequently population size) is slightly higher for $n_D = 500$, compared to $n_D = 100$. Moreover, the migration threshold M_c itself is also somewhat higher for larger n_D. However, these effects are modest in magnitude.

Linkage equilibrium. For a given n_D, we further compare populations with different values of the baseline growth rate: $r_0 = 0.05$, $K = 1000$ vs. $r_0 = 0.1$, $K = 500$, (red vs. blue symbols), both of which are described by the same scaled parameters ($\zeta = 50$, $S_1 = s_1/r_0 = s_2/r_0 = S_2 = 0.05$) and should thus exhibit the same (scaled) migration threshold M_c for loss of local adaptation (provided that the assumptions underlying the diffusion approximation are valid). However, as seen in figure A.2(c), M_c differs significantly between the two set of simulations and is also much higher than the prediction of the diffusion approximation (solid black line).

As discussed above, our theoretical framework ignores LD within demes as well as LD within the migrant pool. Neglecting LD within demes (i.e., neglecting the fact that allele frequencies among residents of a deme may be different from average allele frequencies within the migrant pool at multiple loci) appears to make only a modest difference to the results (Figures A.2(a) and A.2(b)). However, LD within the migrant pool has a far stronger effect: neglecting such LD significantly underestimates the parameter ranges over which local adaptation can be simultaneously maintained across habitats (Figures A.2(c), A.2(d)).

LD within the migrant pool essentially implies that immigrant genotypes entering a particular deme may either be perfectly adapted (if they originate from other demes belonging to the same habitat) or severely unfit (if they emerge from demes belonging to the other habitat). Consequently, an immigrant allele experiences an immediate (first-generation) selective disadvantage which depends on the fitness of its genetic background in the recipient deme; this selective disadvantage is halved in each subsequent generation (Robertson, 1961). Thus, the effective migration rate of alleles differs from m and moreover, depends on both the habitat from which the allele originates and the habitat into which it migrates.

As a first approximation, we assume that immigrants entering any deme are drawn from two distinct pools (corresponding to the two habitats) with relative contributions $1-\rho$ and ρ. The 'effective' number of migrants from the i^{th} pool to a deme within the common

(or rare) habitat is assumed to be $m\overline{N_i}e^{-r_0SL[1-\overline{N_ip_i}/\overline{N_i}]}$ (or $m\overline{N_i}e^{-r_0SL[\overline{N_ip_i}/\overline{N_i}]}$) respectively. Similarly, the average number of immigrant alleles (at a given locus) from the i^{th} pool to a deme within the common (or rare) habitat is assumed to be $m\overline{N_ip_i}e^{-r_0S(L-1)[1-\overline{N_ip_i}/\overline{N_i}]}$ and $m\overline{N_ip_i}e^{-r_0S(L-1)[\overline{N_ip_i}/\overline{N_i}]}$. Thus, in this formulation, the effective number of immigrants from each habitat is weighted by the first-generation fitness of immigrants (from that habitat) in the recipient deme. Similarly, the effective number of immigrant alleles at a particular locus is weighted by the marginal fitness of this allele, which depends on the frequencies of immigrant alleles at the other $L-1$ loci as well as their selective effects in the recipient deme. Note also that as we consider smaller and smaller unscaled parameters $r_0, s, \ldots$ (red vs blue symbols), while holding scaled parameters $\zeta, S, \ldots$ constant, the weights $e^{-r_0SL[1-\overline{N_ip_i}/\overline{N_i}]}$, $e^{-r_0SL[\overline{N_ip_i}/\overline{N_i}]}$ approach 1, so that this approximation converges to the theory in the main paper (which assumes LE).

Assuming that immigrants come from two distinct pools (corresponding to the two distinct habitats) results in a slightly more complicated theoretical approximation. Fixed points can be obtained numerically as before, but now by equating the expected values in each habitat to the corresponding average in the migrant pool derived from that habitat (i.e., by solving for $E(N_1) = \overline{N_1}$, $E(N_2) = \overline{N_2}$, $E(N_1p_1) = \overline{N_1p_1}$, $E(N_2p_2) = \overline{N_2p_2}$). The results of this approximation are depicted by dashed lines in Figures A.2(c) and A.2(d). Note that the critical migration rates obtained via this approximation (dashed lines) are significantly higher than those predicted by ignoring LD altogether (solid lines), but still somewhat lower than M_c found in individual-based simulations (points). A more accurate approximation can be obtained by accounting for the fact that immigrant alleles originating from differently adapted demes experience a selective disadvantage that persists over multiple generations (and not just the first generation, as assumed above). We defer details to future work.

B
gridCoal

B.1 Simulation inputs

In order to run the simulations, it is necessary to define the following input files and parameters.

Demographic history The demographic history of the collection of subpopulations distributed on a grid is represented by a matrix of size $T \times n$, T being the number of time points one wishes to define the population sizes at, and n being the number of grid cells. The matrix contains the population sizes of the grid cells at given time points. Each row is the flattened two dimensional grid, indexed from 0 to $n-1$, defining the sizes of the subpopulations. The first line shall be the oldest time point. In msprime, a population is not allowed to be of size 0. In our case, however, we do not want to exclude the possibility that populations can go extinct and get recolonized, even repeatedly, therefore we define 0 populations to have a technically 0 size, 10^{-10}.

List of sampled cells. A list of cells from which the samples are taken is required. [1]. These cells must not be empty at the final time point (presence), but could be empty in the past. For efficiency, two samples are taken from each sampled deme. It is more efficient to run more replicates with fewer samples than fewer replicates with more samples. In the coalescent process, the waiting time until the next coalescence event happens is

[1]Mind the indexing of the sampled cells, this also starts from 0.

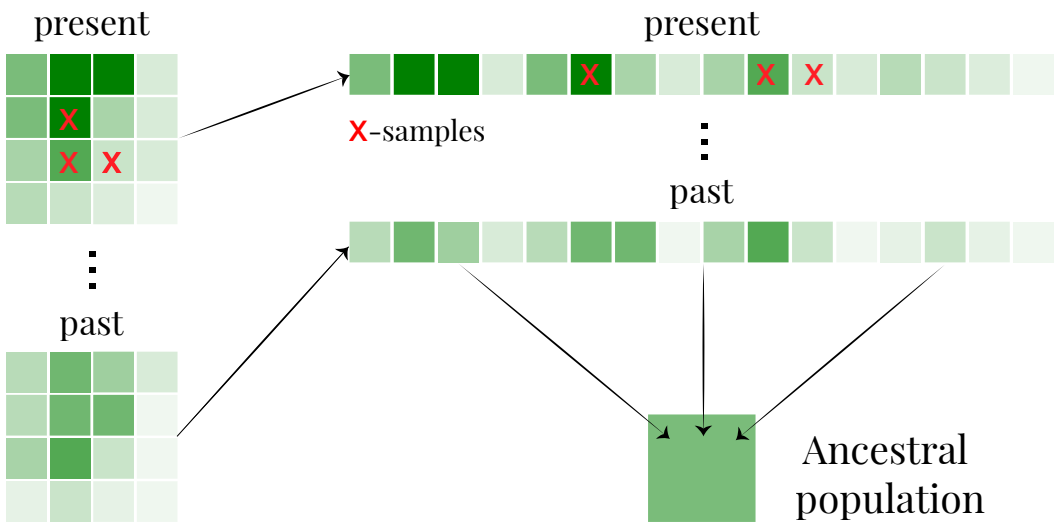


Figure B.1: Input data preparation.

exponentially distributed with mean proportional to the number of lineages. Thus, in the beginning several coalescence events happen in quick succession, yet the mean coalescence time of the deme will be largely dominated by the times that the last remaining few samples took to coalesce.

Migration matrix A two-dimensional migration matrix capturing backward migration is needed for running the simulation itself, however, as an input file we only need the forward migration matrix, M [2]. We assume that M remains constant in time because it depends on the dispersal ability of the species. The element $(M_{i,j})$ defines the fraction of the lineages in population i that migrates to population j. However, the backward migration matrix BM changes through time, as it depends on the actual population sizes of the neighboring cells. $BM_{i,j}(t)$ defines the fraction of lineages in population i at time t that has parents in population j, thus can be calculated as

$$BM_{i,j} = \frac{M_{j,i}N_j(t-1)}{N_i(t)}. \tag{B.1}$$

The fact that the population size changes and migration matrices are not updated at every generation, but at arbitrary time steps, represents a significant gain in computing time in comparison to other tools such as SPLATCHE (Ray et al., 2010).

[2]The diagonal elements of M must be zero, required by msprime.

Adjacency matrix This matrix has the same format as the migration matrix and it defines which cells are connected to which other ones by migration. It is needed so that the migration matrix can be defined as a 0 matrix and we still have this spatial information.

Time periods and generation times Time is measured in generations in msprime, therefore we need to specify the generation time of the population at hand. We define the generation time as the time at which the species comes to a reproductive age. This factor sets a scale for the results. Note, that dt and generation time are both given in actual years. The value dt expresses how coarse our data is in time.

Ancestral populations At the time point, beyond which the demography is unknown, all lineages are merged into spatially non-explicit ancestral populations where they follow the standard coalescent process. We assume either a single, or multiple panmictic ancestral populations with specified sizes and a very low rate of migration between them (10^{-8}). Furthermore, it is necessary to specify which of the cells originate in which ancestral population.

Basic parameters One can define the recombination and mutation rates, however, we chose these values always to be 0.

Other In order to avoid errors later in the simulations, we implemented a few checks in the beginning about the sizes of the matrices, the occupancy of sampled cells, and so on. For this, the number of rows in the grid also needs to be specified. To allow easier parallel runs, one can set a scaling parameter that scales the ancestral population sizes accordingly, and each run has its own serial number.

Demographic events

All demographic changes, including population size and migration rate changes, need to be defined as a demographic events at a given time point, going backwards in time, and collected to a list, which is used by *msprime* for the coalescent simulation. This is automatically done by gridCoal provided the file containing the demographic history.

At each generation, we calculate the backward migration matrix as described earlier - if it differs from the migration matrix of the previous generation, we update it.

Similarly, we update the population sizes, with one additional constraint. If a deme has individuals at a given time point t but was empty in the previous time step $t-1$, we need to define the source of those individuals. We do it such that we define a mass migration event: from this cell to its neighboring 4 cells, proportional to their sizes in the previous generation. Thinking forward in time, this corresponds to the idea of an empty cell being colonized by its neighbors, proportional to their sizes.

B.2 Additional information, plots and tables

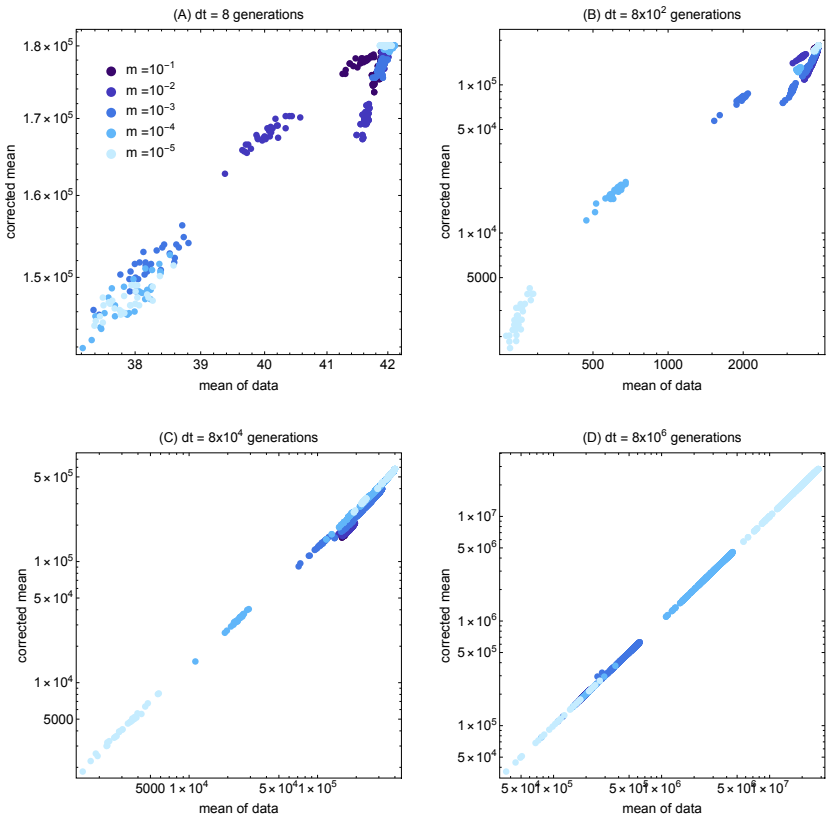


Figure B.2: Mean of the simulations against the corrected mean (Equation 4.4) for different step length and various migration rates. Note the differing axes: they are on a logarithmic scale and have very different ranges (especially for smaller *dt* values). In case of small time steps (A-B), most simulations do not coalesce, therefore in Equation 4.4 the second term dominates, while for large time steps (C-D) it is the first term.

Best-estimate N_e. The demographic history of silver fir in the past 22,000 years was obtained from the LPX-Bern dynamic global vegetation model with a resolution of 1° by 1° Lat/Lon (Ruosch et al., 2016; Sitch et al., 2003). The model predictions cover the time period from the Last Glacial Maximum (LGM) to today. The output of LPX-Bern is the foliar projective cover (FPC), which is the fraction of a grid cell that is covered by the species. The model was forced with climate anomalies and included competition between common tree species and plant functional types. In the following we shall refer to one grid cell as one population. We estimated the population census size from FPC assuming that a mature tree occupies 40m^2. Census size was used to estimate N_e using the hybrid approach of Waples et al. (2011) implemented in the software *AgeNe*. This approach takes into account life-history characteristics of the species estimated from a population matrix. We calculated an average population matrix from 32 available at least 20 years long studies from related *Abies* species (*A. concolor*, *A. magnifica*, and *A. sachalinensis*) in the COMPADRE data base (Salguero-Gómez et al., 2015). The generation time was 42 years and net reproductive rate 1.16 in the average population matrix. *AgeNe* estimated 0.038 as a scaling factor to obtain N_e from census size. We assumed that this scaling factor was constant in space and time. In the following, we shall refer to these N_e values of the populations as best-estimate N_e.

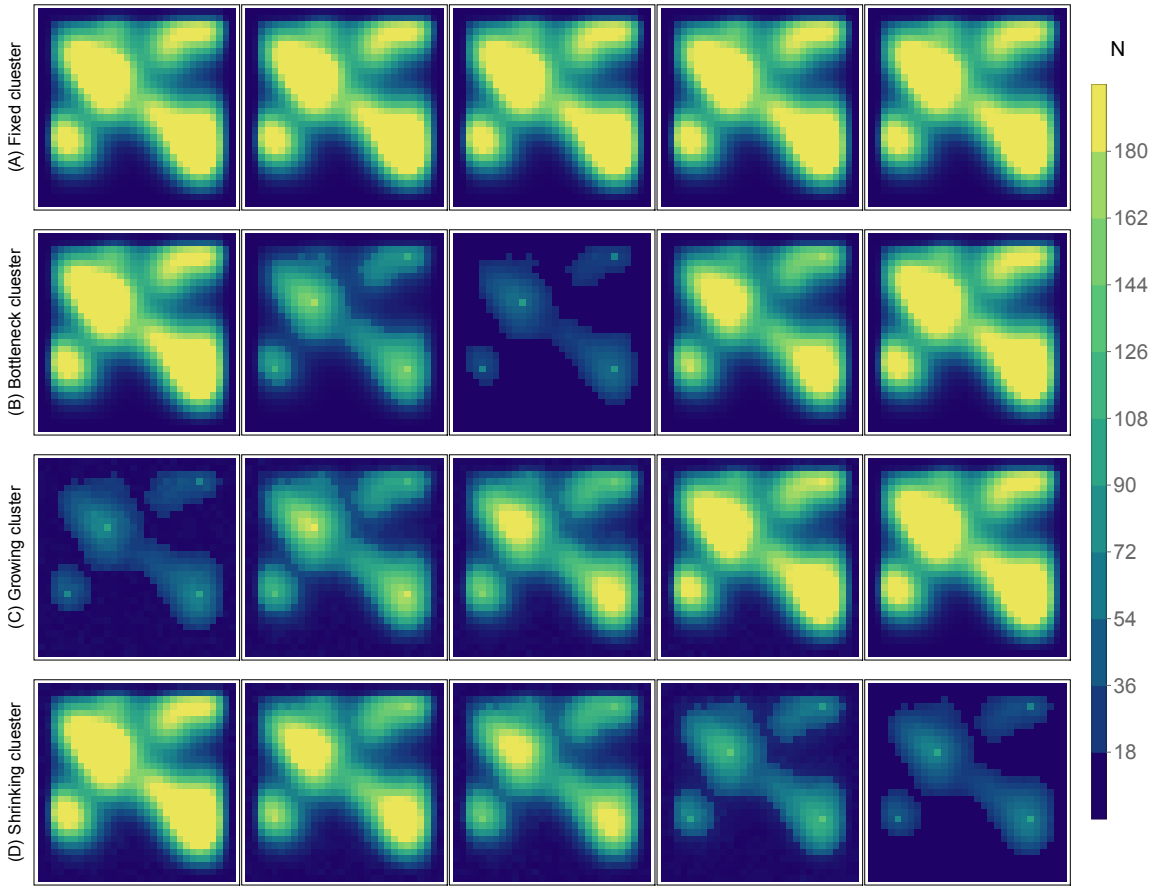


Figure B.3: Clustered maps going through various demographic histories. The columns are 30, 23, 15, 7 steps back in time and the current one on the right. The rows are: fixed cluster, bottleneck, expansion, decline.

Second row:

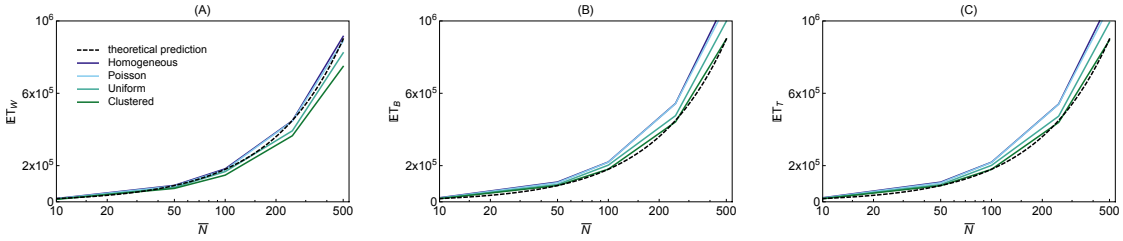


Figure B.4: Within (A), between (B), and average mean coalescence times (C), when migration rate is high ($m = 1$). Different colors represent different underlying maps: no variance, low variance, high variance, and clustered. Note that the mean coalescence time fits the prediction well, whereas the within coalescence time is slightly below the expectation. This is due to the fact that although the migration rate is very high, considering only lineages in the same cell retains the spatial structure.

Homogeneous map							
$\overline{N}$ and m	10^{-8}[%]	10^{-5}[%]	10^{-4}[%]	10^{-3}[%]	10^{-2}[%]	10^{-1}[%]	10^{0}[%]
10	0.	0.02	0.	0.	0.	0.	0.
50	0.0033	0.12	0.	0.	0.	0.	0.
100	0.	0.193	0.	0.	0.	0.	0.
250	0.00667	0.4233	0.	0.	0.	0.	0.
500	0.013	0.9967	0.	0.	0.	0.	0.

Poisson map							
$\overline{N}$ and m	10^{-8}[%]	10^{-5}[%]	10^{-4}[%]	10^{-3}[%]	10^{-2}[%]	10^{-1}[%]	10^{0}[%]
10	0.	0.0167	0.	0.	0.	0.	0.
50	0.	0.09	0.	0.	0.	0.	0.
100	0.	0.2	0.	0.	0.	0.	0.
250	0.	0.4867	0.	0.	0.	0.	0.
500	0.0033	0.9867	0.	0.	0.	0.	0.

Uniform map							
$\overline{N}$ and m	10^{-8}[%]	10^{-5}[%]	10^{-4}[%]	10^{-3}[%]	10^{-2}[%]	10^{-1}[%]	10^{0}[%]
10	0.	0.01072	0.	0.	0.	0.	0.
50	0.	0.1	0.	0.	0.	0.	0.
100	0.003572	0.178571	0.	0.	0.	0.	0.
250	0.00333	0.37	0.006667	0.	0.	0.	0.
500	0.01667	0.88	0.07667	0.	0.	0.	0.

Clustered map							
$\overline{N}$ and m	10^{-8}[%]	10^{-5}[%]	10^{-4}[%]	10^{-3}[%]	10^{-2}[%]	10^{-1}[%]	10^{0}[%]
10	0.	0.00741	0.	0.	0.	0.	0.
50	0.	0.03704	0.	0.	0.	0.	0.
100	0.0037037	0.0703704	0.	0.	0.	0.	0.
250	0.00371	0.1778	0.	0.	0.	0.	0.
500	0.01852	0.3778	0.	0.	0.	0.	0.

Table B.1: Percentage of simulations for which the within coalescence time is higher than the time at we pull all lineages together into a single non-spatial ancestral population. It is the average of 1000 independent runs and over all the sampled cells. At each simulation we used the same row to sample, which results in 30, 28, or 27 cells, depending on the number empty cells. The time at which the lineages are pulled into a single population of size 1 is 10^9 years, or 40000000 generations.

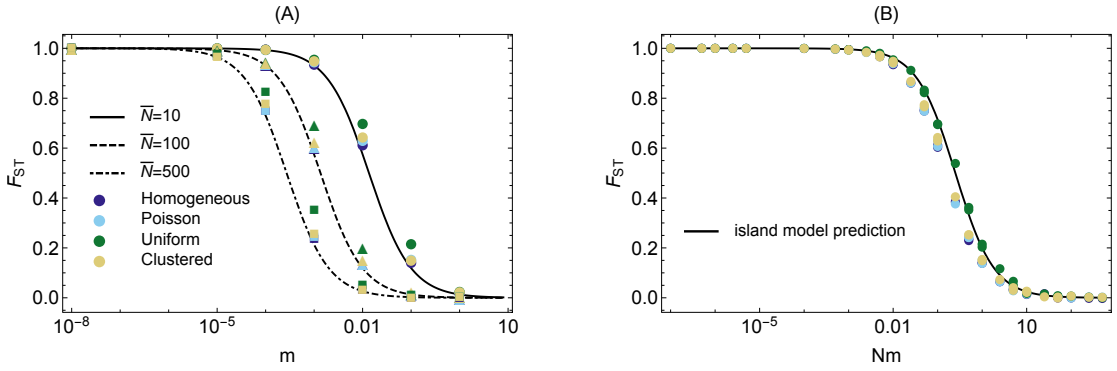


Figure B.5: Global F_{ST} values against migration rates, as predicted by the theory (Equation (4.7) and (4.8)) and calculated from simulations.

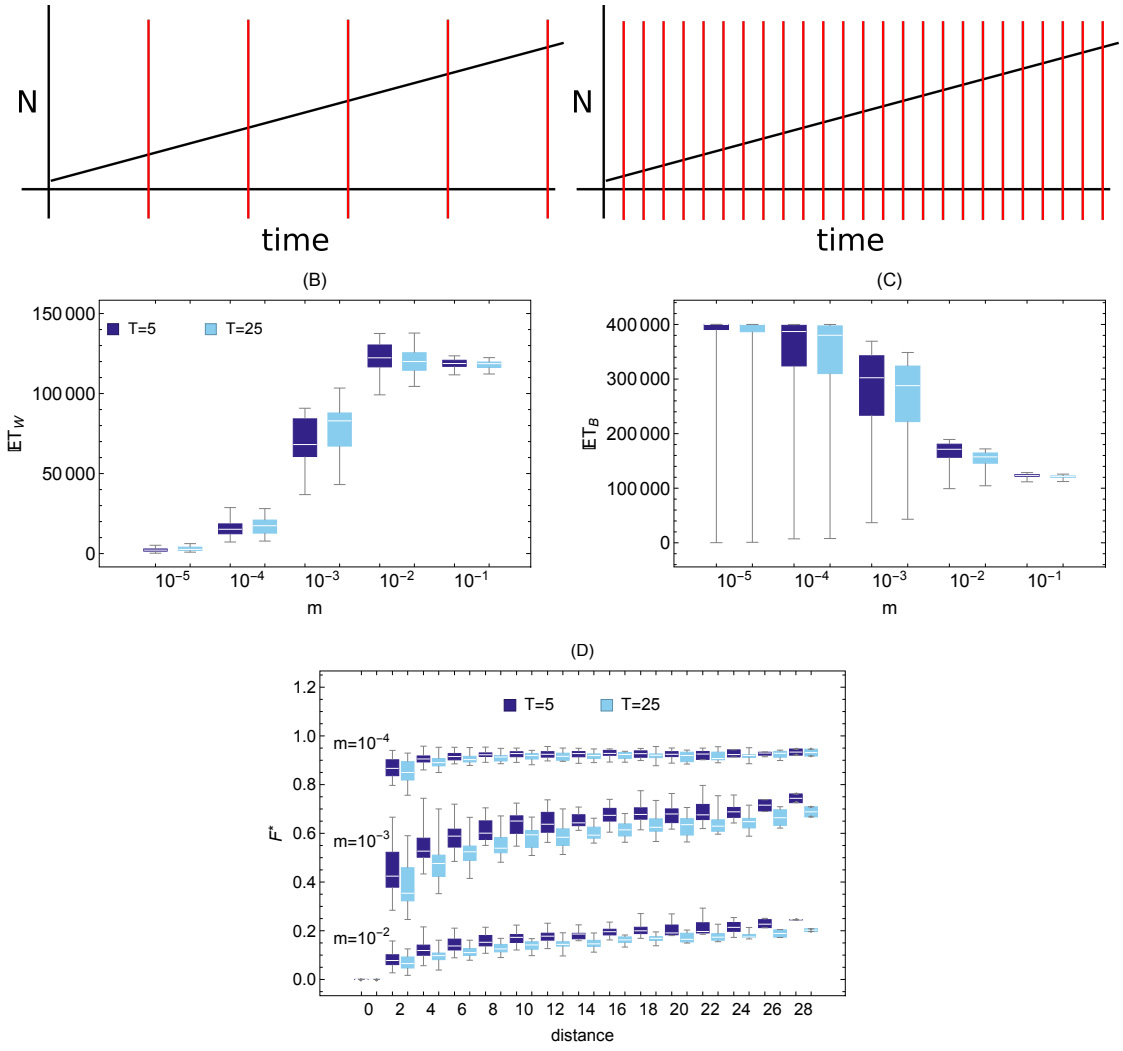


Figure B.6: Linear increase for two different time resolution: demographic changes were defined at $5, 25$ steps apart, spanning over the same total time. The means are calculated directly from the simulations, as we have run it long enough.

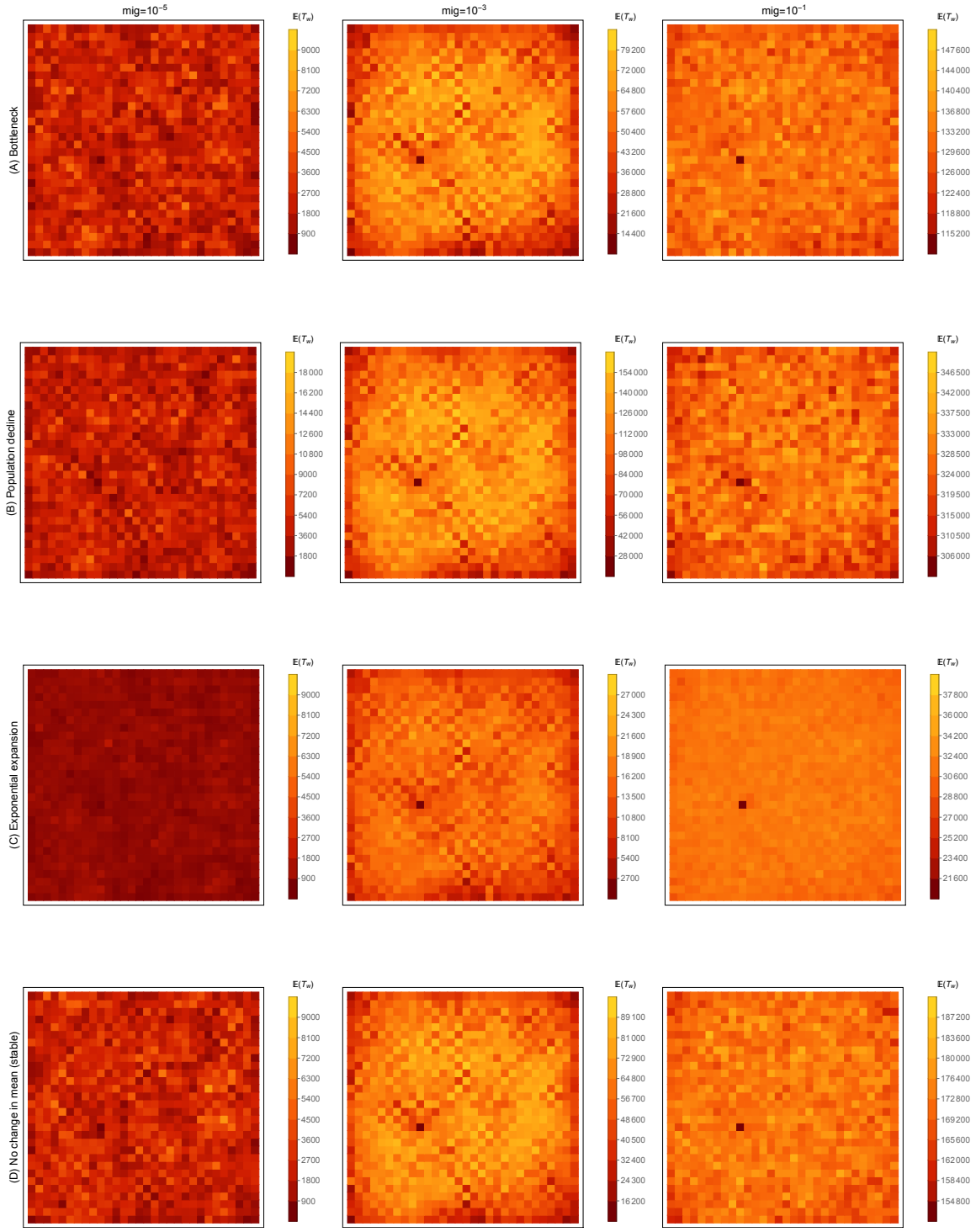


Figure B.7: Observed diversity, i.e. corrected mean within coalescence time for the grid. The columns correspond to migration rates $10^{-5}, 10^{-3}$ and 10^{-1} and the rows to the different demography types.

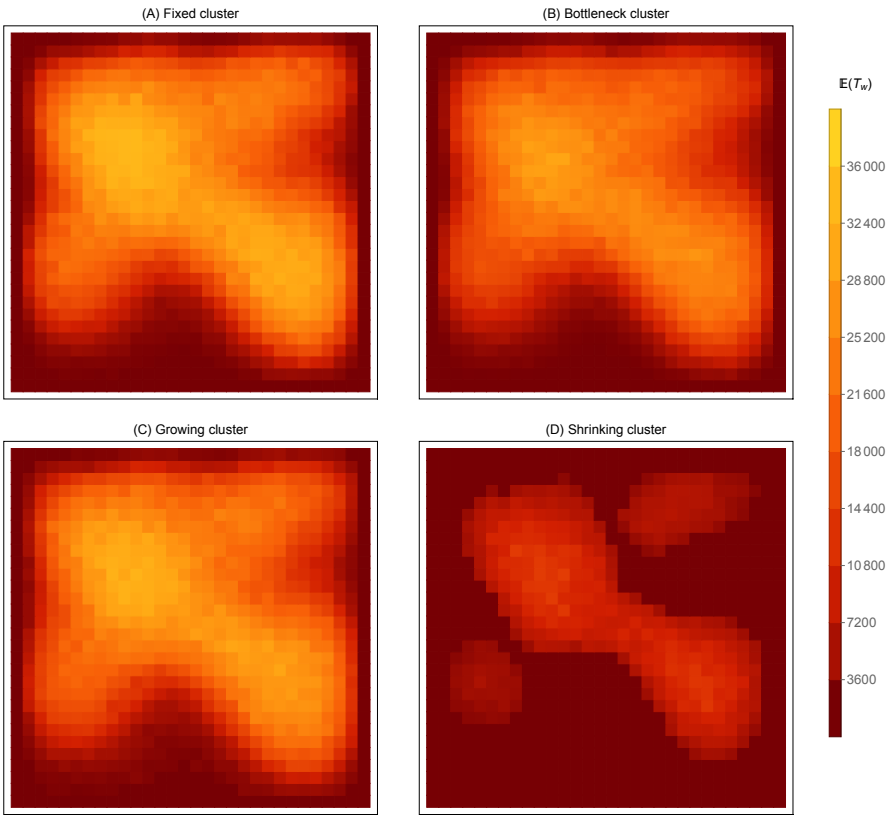


Figure B.8: Corrected mean within coalescence time for (A) Cluster that is fixed in time, (B) Cluster experiencing bottleneck](C) Expanding cluster, and (D) Shrinking cluster. The most obviously different is the declining one, as the numbered of sampled demes is smaller and there is very little difference between neighboring cells, whereas for the other cases one can observe more differentiation across the panels.

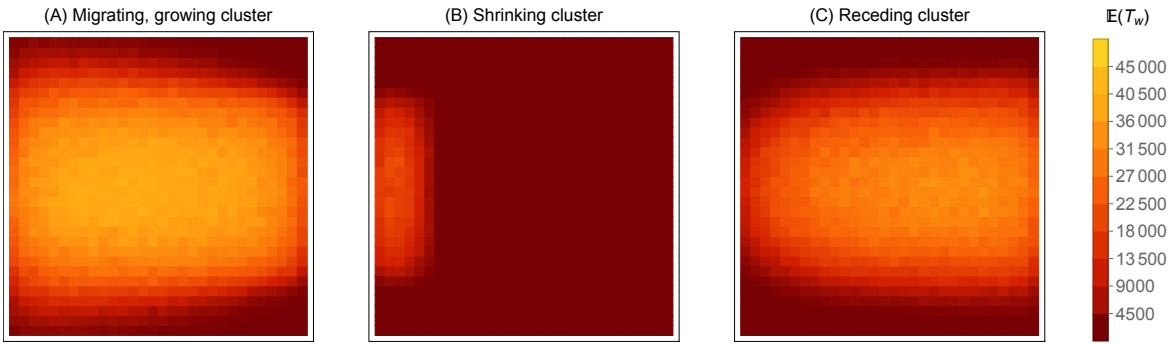


Figure B.9: Corrected mean within coalescence time for (A) Migrating and growing cluster, (B) Shrinking cluster, and (C) receding cluster. The pattern produced is very similar to the current distribution of the species.