Local adaptation, genetic load and extinction in metapopulations

by

Oluwafunmilola Olusanya

September, 2023

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

Committee in charge:

Carrie Bernecky, Chair
Nicholas Barton
Jitka Polechova
Himani Sachdeva
Matthew Robinson
Joachim Hermisson

The thesis of Oluwafunmilola Olusanya, titled *Local adaptation, genetic load and extinction in metapopulations*, is approved by:

Supervisor: Prof. Nicholas Barton, ISTA, Klosterneuburg, Austria

Signature: _____

Co-supervisor: Dr. Jitka Polechova, University of Vienna, Vienna, Austria

Signature: _____

Co-supervisor: Dr. Himani Sachdeva, University of Vienna, Vienna, Austria

Signature: _____

Committee Member: Prof. Matthew Robinson, ISTA, Klosterneuburg, Austria

Signature: _____

Committee Member: Prof. Joachim Hermisson, University of Vienna, Vienna, Austria

Signature: _____

Defense Chair: Dr. Carrie Bernecky, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA 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: _____

Oluwafunmilola Olusanya
September, 2023

Signed page is on file

Abstract

In nature, different species find their niche in a range of environments, each with its unique characteristics. While some thrive in uniform (homogeneous) landscapes where environmental conditions stay relatively consistent across space, others traverse the complexities of spatially heterogeneous terrains. Comprehending how species are distributed and how they interact within these landscapes holds the key to gaining insights into their evolutionary dynamics while also informing conservation and management strategies.

For species inhabiting heterogeneous landscapes, when the rate of dispersal is low compared to spatial fluctuations in selection pressure, localized adaptations may emerge. Such adaptation in response to varying selection strengths plays an important role in the persistence of populations in our rapidly changing world. Hence, species in nature are continuously in a struggle to adapt to local environmental conditions, to ensure their continued survival. Natural populations can often adapt in time scales short enough for evolutionary changes to influence ecological dynamics and vice versa, thereby creating a feedback between evolution and demography. The analysis of this feedback and the relative contributions of gene flow, demography, drift, and natural selection to genetic variation and differentiation has remained a recurring theme in evolutionary biology. Nevertheless, the effective role of these forces in maintaining variation and shaping patterns of diversity is not fully understood. Even in homogeneous environments devoid of local adaptations, such understanding remains elusive. Understanding this feedback is crucial, for example in determining the conditions under which extinction risk can be mitigated in peripheral populations subject to deleterious mutation accumulation at the edges of species' ranges as well as in highly fragmented populations.

In this thesis we explore both uniform and spatially heterogeneous metapopulations, investigating and providing theoretical insights into the dynamics of local adaptation in the latter and examining the dynamics of load and extinction as well as the impact of joint ecological and evolutionary (eco-evolutionary) dynamics in the former. The thesis is divided into 5 chapters.

Chapter 1 provides a general introduction into the subject matter, clarifying concepts and ideas used throughout the thesis. In chapter 2, we explore how fast a species distributed

across a heterogeneous landscape adapts to changing conditions marked by alterations in carrying capacity, selection pressure, and migration rate.

In chapter 3, we investigate how migration selection and drift influences adaptation and the maintenance of variation in a metapopulation with three habitats, an extension of previous models of adaptation in two habitats. We further develop analytical approximations for the critical threshold required for polymorphism to persist.

The focus of chapter 4 of the thesis is on understanding the interplay between ecology and evolution as coupled processes. We investigate how eco-evolutionary feedback between migration, selection, drift, and demography influences eco-evolutionary outcomes in marginal populations subject to deleterious mutation accumulation. Using simulations as well as theoretical approximations of the coupled dynamics of population size and allele frequency, we analyze how gene flow from a large mainland source influences genetic load and population size on an island (i.e., in a marginal population) under genetically realistic assumptions. Analyses of this sort are important because small isolated populations, are repeatedly affected by complex interactions between ecological and evolutionary processes, which can lead to their death. Understanding these interactions can therefore provide an insight into the conditions under which extinction risk can be mitigated in peripheral populations thus, contributing to conservation and restoration efforts.

Chapter 5 extends the analysis in chapter 4 to consider the dynamics of load (due to deleterious mutation accumulation) and extinction risk in a metapopulation. We explore the role of gene flow, selection, and dominance on load and extinction risk and further pinpoint critical thresholds required for metapopulation persistence.

Overall this research contributes to our understanding of ecological and evolutionary mechanisms that shape species' persistence in fragmented landscapes, a crucial foundation for successful conservation efforts and biodiversity management.

Acknowledgements

Firstly, I extend my heartfelt gratitude to the funding bodies: The Marie Curie Cofund (co-financed by H2020 Marie Skłodowska-Curie COFUND Action ISTScholar) with grant agreement no. 665385 that funded my PhD for the first two years (2018/09/15 − 2020/01/31); the Austrian Science Fund (FWF) (stand alone grant of Dr. Jitka Polechova) with grant no. P 32896 that provided a 50% funding allocation from 2020/09/15 − 2022/08/31 and the Austrian Academy of Sciences for granting me the DOC fellowship with grant no. 26380 for the period 2022/09/01 − 2024/08/31. Your generous financial support has made the completion of this thesis possible.

I express my sincere appreciation to ISTA as an institution. Also to the grad school, all administrative staff, the HR department, and the IT team for their timely support and assistance and for ensuring a conducive environment.

I am deeply indebted to my supervisors, Prof. Nick Barton, Dr. Jitka Polechova, and Dr. Himani Sachdeva, whose expertise, dedication, and unwavering support have been invaluable over the past few years. Thank you for all the meetings, the insightful feedback, the impromptu drop-bys at your office, and for all the times you graciously entertained the unexpected email musings that crept into your weekend agendas. Beyond your scholarly guidance, your occasional logistical intervention has been truly remarkable. To Nick, thank you for the timely computer rescue after mine staged a dramatic exit. This not only saved my sanity but improved my efficiency and taught me the invaluable lesson of backing everything up. To Jitka, thank you for the Mathematica licenses and for making possible the office in the maths department which shortened my daily pilgrimage to ISTA and boosted my efficiency. Thank you Himani for the occasional guest passwords during my early days in the department.

I say a big thank you to Prof. Joachim Hermisson and Dr. Matthew Robinson, my external and internal thesis committee members for their constructive feedback during progress reviews and for all their support.

I extend my sincere appreciation to all the members of the Barton Group whose friendship and camaraderie I have enjoyed over the past few years. A big thank you to Laura Hayward for all the coffee meetings, useful insights, and discussions on a side project

outside the scope of the thesis. Thank you Dasha for the discussions on butterflies and for your sweet nature. Thank you, Parvathy for your support, humor, and encouragement. Your friendship has made this journey not only manageable but also enjoyable. Thank you Ksennia, a fellow collaborator on one of the papers in the thesis. Also, heartfelt thanks to Ben and Yannic, my office mates in the maths department, your occasional wit and banter made our corner of the department feel less like a labyrinth of theorems and lemmas.

I am forever grateful to my beloved family for their unceasing love and encouragement. Your unwavering support and sacrifices during both challenges and triumphs have been exemplary. Thank you Olaolu for weaving the tapestry upon which the seed of this dream was imprinted.

To my biggest cheerleader, my husband, friend and le compagnon de mon cœur, thank you for your patience, understanding, and unwavering support. Your unconditional love, friendship, encouragement, and belief in my abilities have been a source of strength during this challenging, yet rewarding journey and has made this achievement possible.

To Vienna, the city I have loved and called home for the past five years, I owe a special debt of gratitude. Your rich cultural heritage, vibrant atmosphere, and work cafes have provided an inspiring backdrop for my research. Living amidst the historical grandeur of the city has been a source of motivation and has graced me with beautiful memories I will forever cherish.

Finally, to everyone whose name may not be mentioned here, your contribution, whether big or small, has not gone unnoticed and I say a very big thank you to you all.

About the Author

Olusanya Oluwafunmilola completed a Bachelor's degree in Mathematics at the Obafemi Awolowo University in Nigeria and two master's degrees in Mathematical Sciences at the University of Western Cape, South Africa, and the University of Durham, United Kingdom respectively before joining ISTA in September 2018. Her main research interests revolve around understanding what maintains diversity in populations as well as gaining a theoretical insight into the dynamics of genetic load and local adaptation in metapopulations. During her PhD studies, she worked on several collaborative projects, two of which have been published in the journal *Philosophical Transactions of the Royal Society B*. Oluwafunmilola presented her research results at the Congress of the European Society for Evolutionary Biology in 2022. She also received a two-year funding for her research from the Austrian Academy of Science via the OeAW DOC fellowship in 2022. Outside of science, she enjoys travelling and volunteering. She is also a data science enthusiast and was invited to attend the Erasmus$^+$ SCitrain course (Czech Republic) and the Alan Turing data study group (UK) in 2023.

List of Collaborators and Publications

Chapter 2 (Already published)

Title: The response of a metapopulation to a changing environment

Journal: *Philosophical Transactions of the Royal Society B*

url: https://royalsocietypublishing.org/doi/10.1098/rstb.2021.0009

Author contributions

Barton, N: Conceptualization, formal analysis, methodology, supervision, visualization, writing original draft.

Olusanya, O: Conceptualization, formal analysis, investigation, visualization, writing review and editing.

Chapter 3 (In preparation)

Title: Local adaptation in a metapopulation - a multi-habitat perspective

Author contributions

Olusanya, O: Conceptualization, formal analysis, methodology, investigation, visualization, writing of manuscript.

Barton, N: Guidance and supervision

Polechova, J: Guidance and supervision

Chapter 4 (Already published)

Title: Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity

Journal: *Philosophical Transactions of the Royal Society B*

url: https://royalsocietypublishing.org/doi/10.1098/rstb.2021.0010

Author contributions

Sachdeva, H: Conceptualization, formal analysis, writing original draft.

Olusanya, O: Individual-based analysis, visualization, writing and editing.

Barton, N: Conceptualization, proof and review.

Chapter 5 (bioRxiv)

Title: Genetic load, eco-evolutionary feedback and extinction in a metapopulation

url: https://www.biorxiv.org/content/10.1101/2023.12.02.569702v1

Author contributions

Olusanya, O., Khudiakova, K., and Sachdeva, H: Designed the study and wrote the manuscript.

Olusanya, O: Carried out the hard selection analysis

Sachdeva, H, and Khudiakova, K: carried out the soft selection analysis

Olusanya, O: Carried out allele frequency simulations for hard selection.

Sachdeva, H: Carried out individual-based simulations for soft selection.

Table of Contents

[1]This work can be found online at https://www.biorxiv.org/content/10.1101/2023.12.02.569702v1

List of Figures

Chapter 1

General introduction

Oluwafunmilola Olusanya

1.1 Motivation

Life on earth as we know it has been creatively woven together over billions of years, shaped by the forces of evolution. From the magnificent expanses of tropical rainforests to the majestic elegance of the Artcic tundra, the ecological complexity and range of life forms we observe is testament to the remarkable capacity of species to adapt and thrive in a wide range of environments. However, in recent decades, there has been a rapid decline in biological diversity with a lot of species teetering on the brink of extinction owing to climate change and other anthropogenic factors (McLaughlin et al., 2002; Kaye et al., 2019; Lande, 1998; Lawton et al., 1995).

Pertinent among these factors is habitat fragmentation (Almond et al., 2020; Díaz et al., 2019). As human populations continue to grow and landscapes are transformed to meet their needs, many plant and animal species are increasingly losing their habitat and gradually becoming confined to small and often isolated populations (Fahrig, 2003; Haddad et al., 2015). These populations often act as islands surrounded by uninhabitable patches, making it difficult for them to spread, locate compatible mates, and maintain gene flow (Fahrig, 2003). As a result, many such populations have remained restricted to these isolated pockets of habitats with limited chances for genetic exchange and increased susceptibility to the negative effects of genetic drift and inbreeding (Frankham et al., 2017). The consequences are a diminished capacity to respond to changes in environmental conditions and a higher susceptibility to extinction (Lande, 1998; Gaggiotti, 2003). This vulnerability to extinction is further exacerbated by the complex interplay

between ecological and evolutionary (eco-evolutionary) processes where an ecological change triggers a bottleneck event for example, resulting in a decline in genetic variation, further reducing the capacity of populations to evolve in response to more ecological changes.

As landscapes continue to evolve, questions emerge about how species respond to these challenges. Do they adapt to novel conditions within fragmented landscapes? what role does gene flow play in fostering or inhibiting such adaptations? How do eco-evolutionary processes interact to influence these dynamics? and what implications do these hold for species maintenance and viability?

In the face of these growing concerns and questions, there is an urgent need for more comprehensive studies and theoretical models to provide a deeper understanding of the effects of habitat fragmentation on populations, the maintenance of diversity in fragmented populations, and the response of such populations to changing conditions. Such theoretical frameworks can potentially inform conservation efforts and policy interventions to address the ongoing crisis.

This thesis therefore focuses on a theoretical exploration of the maintenance of genetic diversity and the vulnerability of populations to extinction risk owing to habitat fragmentation in the context of a metapopulation.

Before we elaborate on the thesis outline, we will give a brief introduction to pertinent notions used throughout the thesis such as eco-evolutionary dynamics, local adaptation, and genetic load. We will further introduce the concept of a metapopulation and give a quick review of different types of metapopulation models.

1.2 Eco-evolutionary Dynamics

Ecology and evolution are two fundamental concepts in evolutionary biology that shed light on the distribution and diversity of life on earth. Evolutionary dynamics focuses on changes in the genetic makeup (allele frequency changes) of populations over time due to key processes like natural selection, mutation, genetic drift and gene flow.

Ecological dynamics in contrast deals with the interaction between organisms or species and their environment and how this influences their distribution and abundance. These dynamics encompass various components such as population dynamics[1], community ecology and ecosystem interactions. Population dynamics considers changes in the population size of a focal species due to factors such as birth rate, death rate, immigration and emigration. An example of a classic model used in studying such dynamics is the logistic growth equation (Verhulst, 1838). Community ecology explores the interaction and

[1]We focus on population dynamics in this thesis.

co-existence between different species such as symbiotic (e.g., mutualism, commensalism, parasitism), competitive, or predator-prey relationships (Connell and Slatyer, 1977). Some mathematical frameworks for understanding these interactions include the Lotka-Volterra equations (Lotka, 1925) and the competitive exclusion principle (Gause, 1934; Hardin, 1960). Ecosystem dynamics on the other hand encompasses complex interactions that take place within ecosystems due to the flow of energy (through food chains and food webs (Lindeman, 1942)) and nutrient cycling (such as carbon and nitrogen cycles, see Vitousek et al. (1997) and Falkowski et al. (2000)).

Eco-evolutionary dynamics as a research field focuses on the interplay between ecology and evolution that occurs on similar time scales (i.e., on the order of years to centuries). In other words, it deals with the study of how evolutionary changes affect ecological changes, how ecological changes affect evolutionary changes, and the feedback between the two processes (Pelletier et al., 2009). A link from ecology to evolution would occur when an ecological change (or a combination of biotic and abiotic environmental factors) imposes selection that in turn produces a change in some ecologically important trait. On the other hand, a link from evolution to ecology would occur when an evolved change in a trait causes a change in the ecological dynamics of a population. Examples abound in host-parasite interactions, range expansions, and many other phenomena.

Evolution had been traditionally thought to occur too slowly to affect ecological dynamics. Thus, studies in ecology ignored the underlying genetic differences inherent in populations, and studies on species' adaptation ignored explicit population dynamics. This traditional view continued into the 50s until the emergence of theoretical paradigms (Haldane, 1956; Anderson and May, 1982; Pease et al., 1989; Abrams and Matsuda, 1997) and empirical findings (Hanski, 1998; Hendry and Kinnison, 1999; Hendry et al., 2008) that underscored the interplay between ecological and evolutionary processes on relevant timescales. This spurred a growing interest among biologists and ecologists in understanding how evolutionary and ecological dynamics interact as coupled processes.

Studies of eco-evolutionary dynamics within fragmented landscapes are not only academically stimulating, but hold practical relevance for safeguarding biodiversity and ensuring ecosystem functioning. Such studies are also indispensable because populations are repeatedly affected by both demographic factors like growth rate and density regulation, as well as genetic factors like genetic drift, gene flow[2] and selection.

Small and often isolated populations in fragmented habitats are highly prone to stochastic events. One such event is variability or fluctuations in birth rates, death rates and sex ratios, termed demographic stochasticity (Gilad, 2008), which can have profound impacts on the survival of species. For example, chance events like individuals failing to reproduce

[2]Gene flow can have both genetic effects - causing changes in allele frequency within a population, as well as demographic effects - boosting population numbers (see Sachdeva et al., 2022).

or dying can yield a disproportionately large impact. Small populations may also undergo phases of rapid growth followed by sharp declines which can intensify their susceptibility to extinction.

Another random event peculiar to small population is genetic drift, defined as the random fluctuation in allele frequency over time owing to chance events like the random sampling of alleles during the process of reproduction. Genetic drift can cause certain alleles to become fixed (i.e., emerge as the only variant in the population) or lost by chance in a population leading to a loss in genetic diversity over time.

Gene flow, the exchange of genetic material among different populations, also plays a key role in shaping the evolutionary trajectories of species. Within fragmented landscapes, isolated populations frequently face decreased connectivity or gene flow from other patches (owing to geographic distance, physical barriers, or altered dispersal patterns), resulting in increased genetic divergence and the potential for speciation among these populations (Harrison and Hastings, 1996; Hanski, 1999). Selection, another key genetic factor can be heterogeneous (i.e., vary in pressure) with different populations encountering unique ecological challenges and opportunities thus fostering localized adaptations (Kawecki and Ebert, 2004). In contrast, the extent of such localized adaptation can be reduced by gene flow when it facilitates the spread of beneficial traits throughout the metapopulation (while maintaining genetic diversity). Density-dependent factors like resource competition and predation can have profound effects on these interactions. For example, altered population densities can lead to shifts in the strength and direction of selection, influencing the evolution of traits that are essential for survival and reproduction (Bolnick and Nosil, 2007). Even in situations where selection is uniform across a landscape, these different forces can interact in interesting ways to drive population dynamics.

The interplay between genetic and demographic factors in fragmented landscapes is far from straightforward, requiring a comprehensive theoretical analysis to tease out these complex relationships. Theoretical studies that use either computer simulations or explicit analytical mathematics are crucial, as they elucidate the various possibilities that can arise from varying assumptions. They also help in the development of analytical tools and conceptual frameworks for the study of real organisms. Thus, in this thesis, we will develop a theoretical understanding (using both simulations and analytical approximations) of eco-evo dynamics in metapopulations (more on this in section 1.6) and highlight the usefulness of these to conservation efforts.

We now go on to introduce the concept of local adaptation.

1.3 Local adaptation

In nature, populations evolve under the selective pressure imposed by their environments. Over time, this can lead to adaptation to local environmental conditions (i.e., native individuals having a higher fitness compared to foreign individuals). Such adaptation to spatial environmental heterogeneity is believed to shape much genetic and phenotypic diversity in the wild (see Endler, 1986; Merilä and Crnokrak, 2001).

Gene flow between populations in a heterogeneous environment plays an important role in the process of adaptation. It can either constrain local adaptation through the introduction of maladapted alleles or favor the process by increasing genetic variation. For example, with assisted gene flow, genetic variants that confer adaptive advantages in certain populations or environments can be introduced into other populations facing environmental challenges thus enhancing their adaptive capacity (Aitken and Whitlock, 2013). It is however important to note that while assisted gene flow may help increase genetic variation and facilitate adaptation, it can in some cases result in outbreeding depression (Weeks et al., 2011; Aitken and Whitlock, 2013) when the introduced variants are not well suited to the environment of the recipient populations or when they are incompatible (independent of the environment). Such risk of outbreeding depression can be mitigated by carefully accessing the genetic compatibility between populations. Though gene flow is an important force in its own right, theory suggests that it interacts in intricate ways with other evolutionary factors such as the genetic architecture of traits underlying adaptation (Billiard and Lenormand, 2005; Yeaman and Whitlock, 2011) and genetic drift (Blanquart et al., 2012), as well as ecological factors such as demographic stochasticity (see Kisdi, 2002).

The study of local adaptation is important for a number of reasons. It provides useful insights into the relative strengths and key interactions among evolutionary forces like genetic drift, mutation, migration, and selection, as well as an understanding of the adaptive divergence of populations. Secondly, an understanding of the adaptation of populations to local environmental conditions, or the lack of it, plays an important role in species' range studies - adaptation at species' borders and the expansion and contraction of species ranges (Polechová and Barton, 2015; Polechová, 2018). More importantly, the knowledge of the genetic basis of local adaptation (and understanding of the conditions facilitating such adaptations) can help improve management decisions involving the maintenance of biodiversity through the proper management of the genetic diversity of endangered populations, thus contributing to conservation and restoration efforts. It can further provide new insights into climate change adaptation and help improve breeding programs in agriculture and forestry (Simmonds, 1991).

This thesis extends previous theoretical models of local adaptation to understand how

quickly a species changes its range in response to shifting environmental conditions (Chapter 2). We do this using allele frequency simulations and a 'fixed-state' approximation (an extension of models of 'adaptive walks' (Orr, 1998; Trubenova et al., 2019)). The thesis further extends our current theoretical understanding of local adaptation in the literature which has traditionally been restricted to adaptation in two habitats to a multi-habitat perspective (Chapter 3). We will elucidate more on specific questions in section 1.6.

1.4 Genetic load

Central to conservation and conservation biology is the need to maintain adaptation, preserve genetic variation, and prevent the extinction of endangered species - a noble but difficult endeavor - difficult because genetic drift in small isolated patches (that may result from anthropogenic factors) dramatically reduces overall genetic variation by randomly fixing deleterious mutations which may lead to extinction (Lynch et al., 1991).

Since the majority of mutations occurring within populations are detrimental to their overall fitness (Kimura et al., 1963; Kim et al., 2017) and purifying selection is largely inefficient in small populations, this leads to a build-up of deleterious mutations in the population over time. Hence, mutation, though being an essential material for evolution (i.e., the ultimate source of genetic variation) can also have detrimental effects on populations when it is deleterious and thus, a key force central to discussions on adaptation and the maintenance of variation (Kondrashov, 1988; Lynch et al., 1995).

The term *genetic load* was introduced by Muller in 1950 as a measure of the reduction in the mean fitness of a population (plagued by the presence of deleterious genetic variants or mutations) relative to that of a perfect population with no such mutations. This concept plays a pivotal role in understanding the drivers of population evolution and adaptation, as well as in forecasting potential risks to population viability.

There are different types of genetic load depending on the underlying mechanism driving it. These are the mutation load, drift load, inbreeding load, segregation load, substitution load and lag load. While we have described these individual categories below, it is important to recognize the underlying interconnectedness between these concepts. Mutations, being the ultimate source of genetic variation, whether beneficial or deleterious, underlie all forms of genetic load, blurring the boundaries between these classifications.

1. Mutation load: This is simply the reduction in fitness due to the build-up of deleterious mutations within a population's gene pool over time (Krakauer and Nowak, 2001; Henry et al., 2015). Such mutations could arise as a result of errors occurring during DNA replication, or may be due to exposure to mutagenic agents such as chemicals and radiation. The burden of mutation load typically depends on the rate of appearance of such

mutations, the efficiency of selection in removing them as well as the size of the population. Some examples of mutation load include genetic diseases such as sickle cell anemia and Tay-Sachs disease where individuals homozygous for the mutation (i.e., carrying two copies of the mutated gene) in a population experience severe health complications.

In very large populations, deleterious allele frequencies are typically low and maintained by a balance between mutation and selection. In the absence of epistasis, these mutations unless completely recessive make the same contribution to load, so that load is approximately $2U$ (the gemomewide deleterious mutation rate). However in small populations, selection may be less efficient, resulting in higher deleterious allele frequencies (which can be driven to fixation) and thus an increase in load (called the 'drift load') as discussed below.

2. Drift load: This is the decline in fitness that stems from an increase in frequency or fixation of deleterious alleles owing to drift (Poon and Otto, 2000; Willi et al., 2013). Drift load can be of particular concern to small or isolated populations like those found on islands or among endangered species (Whitlock, 2000; Willi et al., 2013). Over time, the accumulation of harmful mutations resulting from genetic drift may result in a reduction in the overall fitness of populations, potentially leading to their extinction. On the contrary, drift has little or no effect on larger populations since such populations are less prone to chance events.

3. Inbreeding load: Inbreeding load is the reduction in fitness (measured, for example, as reproductive success, survival, e.t.c.) that arises from the increased expression of recessive deleterious mutations due to inbreeding i.e., the mating of close relatives (Morton et al., 1956; Crow, 1958). It can also be defined as the slope of the regression of fitness on F, the inbreeding coefficient (Morton et al., 1956), where F can be estimated from pedigree analysis of the frequency of inbred matings (Ballou, 1983). A positive regression coefficient typically signifies a higher inbreeding load (i.e., a decrease in fitness with increased inbreeding). To reduce the adverse effect of inbreeding, several populations have developed mechanisms to prevent mating of close relatives. A few such mechanisms include mating preference or mate choice and kin recognition.

4. Segregation load: This is the genetic burden borne by a population due to the occurrence of homozygotes with less fitness owing to overdominance or heterozygote advantage (Morton et al., 1956; Crow, 1958). When sexual reproduction takes place, segregation can lead to a conversion of fit heterozygotes into less fit homozygotes, thus diminishing overall mean fitness.

5. Substitution load: Substitution load (also known as 'the cost of substitution') pertains to the rate at which deleterious alleles are exchanged or substituted with new alleles with fewer detrimental effects (Haldane, 1957; Kimura et al., 1960; Kimura, 1960, 1961; Phelps IV, 1991). Substitution load arises because newly introduced beneficial mutations require time to completely replace older, less beneficial ones. During this transitional phase, individuals with the advantageous allele coexist alongside those with the deleterious allele, leading to a temporary reduction in overall fitness. Substitution load depends on several factors including the intensity of selection, the rate of substitution, mutation and the size of the population in question. If the rate of substitution is high, beneficial mutations may more readily be incorporated into a populations' gene pool, potentially lowering the burden of load. Conversely, when the rate of substitution is low, the accumulation of deleterious mutations may surpass the pace of advantageous substitutions, leading to an increase in the substitution load.

6. Lag load: The lag load is a measure of the reduction in fitness due to a mismatch of the mean trait value of a population from the optimum trait value when there is a deviation of the selective environment from an earlier state (see Smith (1976), Lynch et al. (1993), Kirkpatrick (1996) and Lande and Shannon (1996)). For example, when there is a rapid directional or periodic change in environment, populations may be limited in their rate of response to become better adapted to the new conditions, resulting in reduced fitness. Lag load can have important implications for the ability of populations to survive and reproduce in changing environmental conditions.

In this thesis, we focus our attention on genetic load due to the accumulation and fixation of deleterious mutations (i.e., drift load) because of its significant implications for the conservation of fragmented populations. Drift load is particularly relevant in peripheral populations at the edge of a species' range experiencing limited gene flow from the core as well as in highly fragmented populations (metapopulations) connected by occasional dispersers as these populations are especially prone to stochastic fluctuations. In such populations, a change or decrease in population size due to factors like reduced connectivity, reduced carrying capacity or demographic stochasticity, may lead to an increase in the effect of random drift, which then enhances the accumulation of more deleterious mutations, thus increasing their frequency in the population. This will lead to a decline in the fitness of the population and a corresponding increase in load, which further reduces population size. If this feedback continues and is strong enough, the population can essentially be driven to extinction. This contentious relationship between declining population size and subsequent increase in load, which may ultimately trigger mutational meltdown[3], has been of lasting interest among many researchers (Frankham,

[3]defined as the decline in fitness resulting from the accumulation of deleterious mutations which can

2005; Lande, 1994) and has remained a controversial issue (Charlesworth et al., 1993; Lynch et al., 1995).

In what follows, we will introduce the notion of a metapopulation (upon which our analytical framework is built) and review existing metapopulation models in the field.

1.5 Metapopulations and metapopulation models

The ecologist Richard Levins first introduced the term metapopulation in 1969 to describe a population of populations that persist in a balance between local extinctions and recolonizations (Levins, 1969). This concept was further developed by Hanski (1998) among others. Metapopulation models are important theoretical models used to describe population dynamics in population genetics and ecology whenever local populations are governed by local colonizations and extinctions (Bascompte and Solé, 1998; Andrewartha et al., 1954; Hanski et al., 1997). They are built on the premise that the metapopulation is made up of discrete local populations that interact through dispersal and gene flow.

Many species in nature conform to the metapopulation structure due to continued habitat loss and fragmentation. A classic example of this is the Glanville fritillary butterflies on the Åland islands in southwestern Finland (Thomas and Hanski, 1997). A second example is the mountain-dwelling Pika found in North America and Asia (Rodhouse et al., 2010), which inhabits talus slopes and rocky outcrops, creating a patchy distribution of suitable habitats within its range.

Metapopulation models are useful as they offer a structured framework for assessing the consequences of habitat fragmentation and formulating strategies to enhance species persistence in fragmented landscapes. They also provide a holistic approach for exploring local adaptation, accounting for spatial and temporal dynamics, population interactions, and the influence of gene flow. Hence, they are indispensable in the planning and evaluation of conservation programs (Akçakaya and Sjögren-Gulve, 2000).

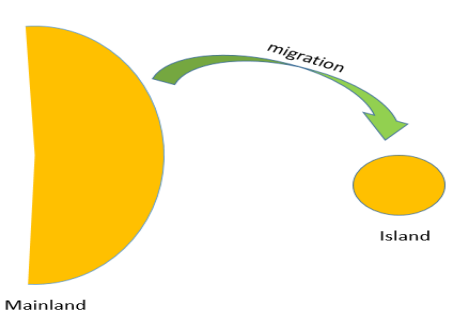
To investigate the interactions, adaptation, and persistence of populations in fragmented landscapes, three distinct metapopulation models can be employed. These are the spatially implicit, spatially explicit, and spatially realistic metapopulation models. Each of these models offer varying degree of detail and realism about the dynamics and spatial distribution of populations within the metapopulation. Below, we give a brief description of these different models.

result in an irreversible downward spiral and eventual extinction of the population (Lynch and Gabriel, 1990).

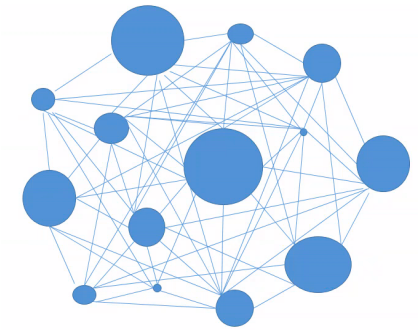
1.5.1 Spatially implicit metapopulation models

As the name depicts, these are metapopulation models that do not consider the physical location, spatial arrangement, or spatial relationship between local populations or patches within the metapopulation. They instead focus on the overall dynamics of the metapopulation ignoring any notion of physical distance between patches. There are two common examples of spatially implicit models, the mainland-island and the many-island model of population structure.

Mainland-Island Models: The mainland-island model (also known as the continent-island model) of population structure was introduced by Haldane in 1930 (fig. 1.1a). It represents the simplest framework for understanding gene flow. In this model, a large and stable population (referred to as the mainland or continent) is connected to a small island population by migration. Migration occurs primarily from the mainland to the island, and the mainland is considered large enough that any reverse migration is assumed to have negligible impact on allele frequency distribution there. Thus, this model is essentially unidirectional, emphasizing migration from the mainland to the island while disregarding reverse migration effects. Consequently, it offers a straightforward means of predicting how gene flow influences allele frequency.



(a) Mainland-island model of population structure



(b) Finite island model with varying deme sizes

Figure 1.1: Spatially implicit metapopulation models

Suppose we assume that an allele has frequency p_m on the mainland and frequency p_i on the island. During each discrete generation, a fraction m of individuals from the mainland arrive on the island such that the island population now consists of a proportion m of immigrant individuals with frequencies p_m and a proportion $(1-m)$ of native individuals whose frequencies were p_i in the last generation. Ignoring drift, the allele frequency on the island follows the deterministic equation,

$$p_{i,t+1} = (1-m)p_{i,t} + m\ p_m, \tag{1.1}$$

which has a stable equilibrium $p_i = p_m$ indicating that over time, with increasing rate, m, of migration, the allele frequency on the island eventually converges to the mainland frequency.

Island Models: This is a generalization of Haldane's model in which there are n islands that exchange migrants randomly with one another at a rate m (Wright, 1931, 1943, 1949). A schematic representation of this is shown in fig. 1.1b (the edges symbolize the exchange of migrants through a common migrant pool). If $\overline{p}$ is the global average allele frequency, then the allele frequency due to gene flow on any island i in generation $t+1$ will be,

$$p_{i,t+1} = (1-m)p_{i,t} + m\overline{p} \tag{1.2}$$

This is very similar to eq. (1.1) with p_m now replaced by $\overline{p} = 1/n \sum_{i=1}^{n} p_i$, implying that the allele frequency in each island[4], i, now approaches the average in the migrant pool as gene flow mixes the populations.

1.5.2 Spatially explicit metapopulation models

Unlike spatially implicit models, these models take into account the spatial structure of the environment as well as the geographic locations of populations. By integrating spatial information, they allow for a more realistic representation of landscapes. They also allow for the investigation of how varying migration patterns (or rules) and genetic interactions influence the long-term evolutionary dynamics of populations.

One important distinction when considering spatially explicit models is whether they assume discrete habitat patches or assume that individuals are dispersing in continuous space. In discrete patch models, landscapes are divided into a set of discrete non-overlapping patches where each patch may represent a different environment or spatial unit and interactions between patches are then described by a set of rules or an arbitrary migration matrix describing the probabilities or rate of movement of individuals from one patch to another. Such models are useful for studying the effects of habitat fragmentation and metapopulation dynamics on genetic diversity and structure. Some examples of discrete patch models include the stepping-stone model (Kimura and Weiss, 1964) and the lattice-based isolation by distance (IBD) model (Malécot, 1951; Guillot et al., 2009).

First introduced by Kimura and Weiss (1964) who coined the term 'stepping-stone', these models simulate the dynamics of populations in discrete habitat patches where individuals migrate predominantly between neighboring populations. They are particularly valuable for studying how limited dispersal or gene flow influences genetic differentiation and the spread of advantageous mutations through populations (Kimura and Weiss, 1964). These

[4]N.B.Throughout the thesis, we use islands, demes, patches, and subpopulations interchangeably.

models have been utilized in studying a wide range of species such as amphibians in wetlands (e.g., populations of choorus frogs *Pseudacris maculata* (Watts et al., 2015)), marine species (e.g., blue mussel, *Mytilus edulis* (Coolen et al., 2020)), etc.

In lattice-based IBD models (Malécot, 1951; Maruyama, 1970), discrete patches or demes are represented using a lattice or grid and genetic similarity is analysed based on the distances between demes. In other words, it assumes that the rate of migration between demes is inversely proportional to the geographical distance between them so that individuals that are in geographical proximity are more likely to be genetically similar due to high gene flow than geographically distant populations. These models are useful for studying the impact of geographic distance on genetic variation in populations.

In spatially explicit continuous space models on the other hand, individuals are allowed to disperse across a spatial gradient or spatially continuous environment rather than being confined to discrete sets. In such models, dispersal may be characterised using a diffusion equation or dispersal kernel to describe the direction of movement or how far an individual can go from its current location (Malécot, 1948; Kot et al., 1996). These models allow for a more realistic representation of populations enabling the study of how landscape features and geographical barriers influence local adaptation and the spatial distribution of genetic variation

1.5.3 Spatially realistic metapopulation models

These are the most detailed and complex of all metapopulation models as they enable the incorporation of precise geometric details of patch networks (e.g., number of patches, their size, quality, and precise locations). They can also go as far as incorporating information about the effects of human activities. Integrating these specifics into the model becomes crucial when seeking to generate accurate quantitative forecasts about the dynamics of actual metapopulations.

One example of a widely used spatially realistic metapopulation model is the Incidence Function Model (IFM) (Hanski, 1994b,a). The IFM considers both occupied and empty patches within its framework, making it particularly suitable for highly fragmented populations or species with patchy distributions across landscapes. The key principle within the IFM is the 'incidence function' - a function that depicts the probability of a successful colonization event in an empty patch after accounting for the spatial distance between patches, the quality of patches, and the dispersal capabilities of the species under consideration. In essence, it provides insights into the probability of immigration from one patch to another while considering both spatial and ecological variables.

Overall, spatially realistic models offer a richer and more nuanced depiction of metapopulation dynamics. Their complexity however often makes them highly computationally

demanding.

Having outlined these different kinds of metapopulation models, it is important to note that a further differentiation might exist based on whether the models assume extinction and colonization events to be instantaneous processes occurring at some rate or whether they explicitly account for population dynamics during extinction and colonization. While the majority of existing literature fall within the former category, a subset of studies (Ronce and Kirkpatrick, 2001; Polechová and Barton, 2015; Polechová, 2018) have contributed to the latter class.

In this thesis, we make use of latter class of models. In particular, we consider spatially implicit models - the mainland-island and island models of population structure (as these are not only simple to work with but are also useful for comprehending fundamental metapopulation dynamics and principles without the need for delving into the particulars of patch configuration) extending these to take into account both ecology and evolution (particularly, the joint dynamics of population size and allele frequency). Our models therefore account for some critical factors that contribute to extinction risk such as genetic stochasticity (mutations and genetic drift) and demographic stochasticity (i.e., random events like the timing and number of births and deaths, as well as variations in individual reproductive success and survival).

1.6 Thesis outline

The Ph.D. thesis is organized into two overarching themes. The first theme focuses on the dynamics of local adaptation and the maintenance of polymorphism in spatially heterogeneous metapopulations. This theme is further divided into two chapters: Chapter 2 and 3.

Chapter 2 investigates how species distributed across various environments adapt to local conditions and how quickly they adjust their range to changing environmental conditions characterised by changes in carrying capacity, selection strength, and migration rate. In particular, it answers the following questions,

1. How does the process of adaptation unfold in a metapopulation when conditions change locally or across the entire metapopulation?

2. How long does it take for a metapopulation to reach a new equilibrium when there are changes in selection or migration rate relative to drift in the metapopulation. How closely does our developed approximation (the 'fixed-state' approximation) approximate this?

Chapter 3 extends previous work on local adaptation and the maintenance of polymorphism (genetic diversity) in two habitats to a multi-habitat perspective. In particular, it considers a metapopulation with three habitats, each adapted to different conditions, and investigates the factors that influence the maintenance of polymorphism in such a metapopulation and the role that chance (genetic drift) plays in the maintenance or loss of genetic variation. Importantly, it provides a theoretical handle on critical (migration and selection) thresholds required for a polymorphism.

In these two chapters (Chapter 2 and 3), we assume a simple metapopulation model where deme sizes remain constant and are not influenced by adaptedness. This is the so-called 'soft selection' model.

The second theme of the thesis offers theoretical insights into the dynamics of genetic load and extinction in metapopulations under spatially uniform selection. This theme is divided into two chapters: Chapter 4 and 5.

Chapter 4 explores the dynamics of load and extinction in a peripheral population that receives migrants from a large mainland. In particular, it provides answers to the following,

1. under what conditions can migration from a large mainland arrest mutational meltdown and prevent the extinction of a peripheral population and what are the demographic and genetic underpinnings of this effect?

2. what are the critical thresholds for extinction and how do these depend on key parameters such as the dominance of deleterious mutations, the selective effect of mutations, the extent of demographic fluctuations and the total genome-wide mutation rate?

3. What role do non-random associations between allelic states at different loci (i.e., linkage disequilibrium, denoted as LD) or between the probability of identity by descent at different loci (i.e., identity disequilibrium, denoted as ID) play?

Chapter 5 extends the analysis in chapter 4 to a metapopulation with infinitely[5] many demes and explores how eco-evo feedback influences extinction risk in such populations using both simulations and analytical approximations. In particular, this chapter investigates

1. the extent of gene flow required to reduce load at a single locus assuming soft selection, and how genomewide heterosis (involving multiple selected alleles) can affect this.

[5]N.B. in simulations, infinitely many demes is approximated by a large but finite number of demes.

2. the effect of migration on equilibrium metapopulation size and load, as well as critical migration thresholds required for the persistence of the metapopulation.

3. how critical thresholds are influenced by the genetic architecture of load (for example, the distribution of selective effects and the dominance coefficients of deleterious mutations) as well as properties of the metapopulation landscape such as the carrying capacities of local demes.

4. What is the role of the hardness of selection[6]?

5. What is the role of mutational bias?

These two chapters contribute to the state of the art in that they go beyond traditional metapopulation models, explicitly accounting for the coupling between population size and allele frequencies in providing a theoretical understanding of the influence of load on the persistence of populations. While there is some exploration of how eco-evo feedback may drive extinction in the context of populations adapting to spatially heterogeneous selection (Ronce and Kirkpatrick, 2001; Szép et al., 2021), this hasn't been explored in the context of drift load in a uniform environment where mutations are deleterious everywhere; this is what is novel here.

In contrast to the first theme, the metapopulation model used here assumes both soft and hard selection. In the hard selection model, population sizes of demes are not fixed but vary due to fluctuations in fitness (driven by stochastic fluctuations in allele frequency) as well as demographic fluctuations.

[6]The 'hardness of selection' here is defined as the extent to which the growth rate of a population is depressed by load.

References

Abrams, P. A. and Matsuda, H. (1997). Prey adaptation as a cause of predator-prey cycles. *Evolution*, 51(6):1742–1750. https://doi.org/10.2307/2410997.

Aitken, S. N. and Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual review of ecology, evolution, and systematics*, 44:367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747.

Akçakaya, H. R. and Sjögren-Gulve, P. (2000). Population viability analyses in conservation planning: an overview. *Ecological bulletins*, pages 9–21. https://www.jstor.org/stable/20113245.

Almond, R. E., Grooten, M., and Peterson, T. (2020). *Living Planet Report 2020-Bending the curve of biodiversity loss*. World Wildlife Fund.

Anderson, R. M. and May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology*, 85(2):411–426. https://doi.org/10.1017/S0031182000055360.

Andrewartha, H. G., Birch, L. C., et al. (1954). *The distribution and abundance of animals*. Number Edn 1. University of Chicago press.

Ballou, J. (1983). Calculating inbreeding coefficients from pedigrees. *Genetics and conservation: a reference for managing wild animal and plant populations*, 509:520.

Bascompte, J. and Solé, R. V. (1998). Modeling spatiotemporal dynamics in ecology. *(No Title)*.

Billiard, S. and Lenormand, T. (2005). Evolution of migration under kin selection and local adaptation. *Evolution*, 59(1):13–23. https://doi.org/10.1111/j.0014-3820.2005.tb00890.x.

Blanquart, F., Gandon, S., and Nuismer, S. (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of evolutionary biology*, 25(7):1351–1363. https://doi.org/10.1111/j.1420-9101.2012.02524.x.

Bolnick, D. I. and Nosil, P. (2007). Natural selection in populations subject to a migration load. *Evolution*, 61(9):2229–2243. https://doi.org/10.1111/j.1558-5646.2007.00179.x.

Charlesworth, D., Morgan, M., and Charlesworth, B. (1993). Mutation accumulation in finite outbreeding and inbreeding populations. *Genetics Research*, 61(1):39–56. https://doi.org/10.1017/S0016672300031086.

Connell, J. H. and Slatyer, R. O. (1977). Mechanisms of succession in natural communities and their role in community stability and organization. *The american naturalist*, 111(982):1119–1144. https://doi.org/10.1086/283241.

Coolen, J. W., Boon, A. R., Crooijmans, R., Van Pelt, H., Kleissen, F., Gerla, D., Beermann, J., Birchenough, S. N., Becking, L. E., and Luttikhuizen, P. C. (2020). Marine stepping-stones: Connectivity of mytilus edulis populations between offshore energy installations. *Molecular Ecology*, 29(4):686–703. https://doi.org/10.1111/mec.15364.

Crow, J. (1958). Some possibilities for measuring selection intensities in man. human biol. 30: 1-13.

Díaz, S. M., Settele, J., Brondízio, E., Ngo, H., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S., et al. (2019). The global assessment report on biodiversity and ecosystem services: Summary for policy makers. https://ipbes.net/system/tdf/inline/files/ipbes_global_assessment_report_summary.

Endler, J. A. (1986). *Natural selection in the wild*. Princeton University Press.

Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual review of ecology, evolution, and systematics*, 34(1):487–515. https://doi.org/10.1146/annurev.ecolsys.34.011802.132419.

Falkowski, P., Scholes, R., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., et al. (2000). The global carbon cycle: a test of our knowledge of earth as a system. *science*, 290(5490):291–296. https://doi.org/10.1126/science.290.5490.291.

Frankham, R. (2005). Genetics and extinction. *Biological conservation*, 126(2):131–140. https://doi.org/10.1016/j.biocon.2005.05.002.

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

Gaggiotti, O. E. (2003). Genetic threats to population persistence. In *Annales Zoologici Fennici*, pages 155–168. JSTOR. https://www.jstor.org/stable/23736522.

Gause, G. (1934). The struggle for existence williams and wilkins. *Baltimore, Maryland*. https://doi.org/10.5962/bhl.title.4489.

Gilad, O. (2008). Spatial distribution models. In Jørgensen, S. E. and Fath, B. D., editors, *Encyclopedia of Ecology*, pages 3311–3314. Academic Press, Oxford. https://www.sciencedirect.com/science/article/pii/B9780080454054006777.

Guillot, G., Leblois, R., Coulon, A., and Frantz, A. C. (2009). Statistical methods in spatial genetics. *Molecular ecology*, 18(23):4734–4756. https://doi.org/10.1111/j.1365-294X.2009.04410.x.

Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., Lovejoy, T. E., Sexton, J. O., Austin, M. P., Collins, C. D., et al. (2015). Habitat fragmentation and its lasting impact on earth's ecosystems. *Science advances*, 1(2):e1500052. https://doi.org/10.1126/sciadv.1500052.

Haldane, J. B. S. (1930). A mathematical theory of natural and artificial selection.(part vi, isolation.). In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 26, pages 220–230. Cambridge University Press. https://doi.org/10.1017/S0305004100015450.

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. https://doi.org/10.1098/rspb.1956.0039.

Haldane, J. B. S. (1957). The cost of natural selection. *Journal of Genetics*, 55:511–524. https://doi.org/10.1007/BF02984069.

Hanski, I. (1994a). Patch-occupancy dynamics in fragmented landscapes. *Trends in Ecology & Evolution*, 9(4):131–135. https://doi.org/10.1016/0169-5347(94)90177-5.

Hanski, I. (1994b). A practical model of metapopulation dynamics. *Journal of animal ecology*, pages 151–162. https://doi.org/10.1007/BF02984069.

Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396(6706):41–49. https://doi.org/10.1038/23876.

Hanski, I. (1999). *Metapopulation ecology*. Oxford University Press.

Hanski, I., Gilpin, M. E., and McCauley, D. E. (1997). *Metapopulation biology*, volume 454. Elsevier.

Hardin, G. (1960). The competitive exclusion principle: an idea that took a century to be born has implications in ecology, economics, and genetics. *science*, 131(3409):1292–1297. https://doi.org/10.1126/science.131.3409.1292.

Harrison, S. and Hastings, A. (1996). Genetic and evolutionary consequences of metapopulation structure. *Trends in Ecology & Evolution*, 11(4):180–183. https://doi.org/10.1016/0169-5347(96)20008-4.

Hendry, A. P., Farrugia, T. J., and Kinnison, M. T. (2008). Human influences on rates of phenotypic change in wild animal populations. *Molecular ecology*, 17(1):20–29. https://doi.org/10.1111/j.1365-294X.2007.03428.x.

Hendry, A. P. and Kinnison, M. T. (1999). Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution*, 53(6):1637–1653. https://doi.org/10.2307/2640428.

Henry, R. C., Bartoń, K. A., and Travis, J. M. (2015). Mutation accumulation and the formation of range limits. *Biology letters*, 11(1):20140871. https://doi.org/10.1098/rsbl.2014.0871.

Kawecki, T. J. and Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology letters*, 7(12):1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x.

Kaye, T. N., Bahm, M. A., Thorpe, A. S., Gray, E. C., Pfingsten, I., and Waddell, C. (2019). Population extinctions driven by climate change, population size, and time since observation may make rare species databases inaccurate. *PloS one*, 14(10):e0210378. https://doi.org/10.1371/journal.pone.0210378.

Kim, B. Y., Huber, C. D., and Lohmueller, K. E. (2017). Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics*, 206(1):345–361. https://doi.org/10.1534/genetics.116.197145.

Kimura, M. (1960). Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. *Journal of Genetics*, 57(1):21–34. https://doi.org/10.1007/BF02985336.

Kimura, M. (1961). Natural selection as the process of accumulating genetic information in adaptive evolution. *Genetics Research*, 2(1):127–140. https://doi.org/10.1017/S0016672300000616.

Kimura, M. et al. (1960). Genetic load and its significance in evolution. *Idengaku Zasshi= Jap. J. Genet.*, 35:7–33. https://doi.org/10.1266/jjg.35.7.

Kimura, M., Maruyama, T., and Crow, J. F. (1963). The mutation load in small populations. *Genetics*, 48(10):1303.

Kimura, M. and Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561. https://doi.org/10.1093/genetics/49.4.561.

Kirkpatrick, M. (1996). Genes and adaptation: A pocket guide. *Adaptation. Academic Press, San Deigo, CA*, pages 125–146.

Kisdi, É. (2002). Dispersal: risk spreading versus local adaptation. *The American Naturalist*, 159(6):579–596. https://doi.org/10.1086/339989.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, 336(6198):435–440. https://doi.org/10.1038/336435a0.

Kot, M., Lewis, M. A., and van den Driessche, P. (1996). Dispersal data and the spread of invading organisms. *Ecology*, 77(7):2027–2042. https://doi.org/10.2307/2265698.

Krakauer, D. and Nowak, M. (2001). Genetic redundancy. In Brenner, S. and Miller, J. H., editors, *Encyclopedia of Genetics*, pages 845–846. Academic Press, New York. https://doi.org/10.1006/rwgn.2001.0519.

Lande, R. (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution*, 48(5):1460–1469. https://doi.org/10.2307/2410240.

Lande, R. (1998). Anthropogenic, ecological and genetic factors in extinction and conservation. *Population Ecology*, 40(3):259–269. https://doi.org/10.1007/BF02763457.

Lande, R. and Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, pages 434–437. https://doi.org/10.2307/2410812.

Lawton, J. H., May, R. M., and Raup, D. M. (1995). *Extinction rates*, volume 11. Oxford University Press Oxford.

Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the ESA*, 15(3):237–240. https://doi.org/10.1093/besa/15.3.237.

Lindeman, R. L. (1942). The trophic-dynamic aspect of ecology. *Ecology*, 23(4):399–417. https://doi.org/10.2307/1930126.

Lotka, A. J. (1925). *Elements of physical biology*. Williams & Wilkins.

Lynch, M., Conery, J., and Burger, R. (1995). Mutation accumulation and the extinction of small populations. *The American Naturalist*, 146(4):489–518. https://doi.org/10.1086/285812.

Lynch, M. and Gabriel, W. (1990). Mutation load and the survival of small populations. *Evolution*, 44(7):1725–1737. https://doi.org/10.2307/2409502.

Lynch, M., Gabriel, W., and Wood, A. M. (1991). Adaptive and demographic responses of plankton populations to environmental change. *Limnology and Oceanography*, 36(7):1301–1312. https://doi.org/10.4319/lo.1991.36.7.1301.

Lynch, M., Lande, R., Kareiva, P., Kingsolver, J., and Huey, R. (1993). Biotic interactions and global change. *Sinauer Associates, Sunderland, MA*, pages 235–250.

Malécot, G. (1948). Les mathématiques de l'hérédité. *(No Title)*.

Malécot, G. (1951). A stochastic treatment of linear problems (mutation, linkage, migration) in population genetics. *Ann. Univ. Lyon. Sci. Dry.*, 14:79–117.

Maruyama, T. (1970). Effective number of alleles in a subdivided population. *Theoretical population biology*, 1(3):273–306. https://doi.org/10.1016/0040-5809(70)90047-X.

McLaughlin, J. F., Hellmann, J. J., Boggs, C. L., and Ehrlich, P. R. (2002). Climate change hastens population extinctions. *Proceedings of the National Academy of Sciences*, 99(9):6070–6074. https://doi.org/10.1073/pnas.052131199.

Merilä, J. and Crnokrak, P. (2001). Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, 14(6):892–903. https://doi.org/10.1046/j.1420-9101.2001.00348.x.

Morton, N. E., Crow, J. F., and Muller, H. J. (1956). An estimate of the mutational damage in man from data on consanguineous marriages. *Proceedings of the National Academy of Sciences*, 42(11):855–863. https://doi.org/10.1073/pnas.42.11.855.

Muller, H. J. (1950). Our load of mutations. *American journal of human genetics*, 2(2):111.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52(4):935–949. https://doi.org/10.2307/2411226.

Pease, C. M., Lande, R., and Bull, J. (1989). A model of population growth, dispersal and evolution in a changing environment. *Ecology*, 70(6):1657–1664. https://doi.org/10.2307/1938100.

Pelletier, F., Garant, D., and Hendry, A. P. (2009). Eco-evolutionary dynamics. https://doi.org/10.1098/rstb.2009.0027.

Phelps IV, F. (1991). A unifying model for the substitutional genetic load. *Mathematical Medicine and Biology: A Journal of the IMA*, 8(1):31–56. https://doi.org/10.1093/imammb/8.1.31.

Polechová, J. (2018). Is the sky the limit? on the expansion threshold of a species' range. *PLoS biology*, 16(6):e2005372. https://doi.org/10.1371/journal.pbio.2005372.

Polechová, J. and Barton, N. H. (2015). Limits to adaptation along environmental gradients. *Proceedings of the National Academy of Sciences*, 112(20):6401–6406. https://doi.org/10.1073/pnas.1421515112.

Poon, A. and Otto, S. P. (2000). Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution*, 54(5):1467–1479. https://doi.org/10.1111/j.0014-3820.2000.tb00693.x.

Rodhouse, T. J., Beever, E. A., Garrett, L. K., Irvine, K. M., Jeffress, M. R., Munts, M., and Ray, C. (2010). Distribution of american pikas in a low-elevation lava landscape: conservation implications from the range periphery. *Journal of Mammalogy*, 91(5):1287–1299. https://doi.org/10.1644/09-MAMM-A-334.1.

Ronce, O. and Kirkpatrick, M. (2001). When sources become sinks: migrational meltdown in heterogeneous habitats. *Evolution*, 55(8):1520–1531. https://doi.org/10.1111/j.0014-3820.2001.tb00672.x.

Sachdeva, H., Olusanya, O., and Barton, N. (2022). Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity. *Philosophical Transactions of the Royal Society B*, 377(1846):20210010. https://doi.org/10.1098/rstb.2021.0010.

Simmonds, N. (1991). Selection for local adaptation in a plant breeding programme. *Theoretical and Applied Genetics*, 82(3):363–367. https://doi.org/10.1007/BF02190624.

Smith, J. M. (1976). What determines the rate of evolution? *The American Naturalist*, 110(973):331–338. https://doi.org/10.1086/283071.

Szép, E., Sachdeva, H., and Barton, N. H. (2021). Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5):1030–1045. https://doi.org/10.1111/evo.14210.

Thomas, C. D. and Hanski, I. (1997). Butterfly metapopulations. In *Metapopulation biology*, pages 359–386. Elsevier. https://doi.org/10.1016/B978-012323445-2/50020-1.

Trubenova, B., Krejca, M. S., Lehre, P. K., and Kötzing, T. (2019). Surfing on the seascape: Adaptation in a changing environment. *Evolution*, 73(7):1356–1374. https://doi.org/10.1111/evo.13784.

Verhulst, P.-F. (1838). Notice on the law that the population follows in its growth. *Mathematical and physical correspondence*, 10:113–129.

Vitousek, P. M., Mooney, H. A., Lubchenco, J., and Melillo, J. M. (1997). Human domination of earth's ecosystems. *Science*, 277(5325):494–499. https://doi.org/10.1126/science.277.5325.494.

Watts, A. G., Schlichting, P. E., Billerman, S. M., Jesmer, B. R., Micheletti, S., Fortin, M.-J., Funk, W. C., Hapeman, P., Muths, E., and Murphy, M. A. (2015). How spatio-temporal habitat connectivity affects amphibian genetic structure. *Frontiers in Genetics*, 6:275. https://doi.org/10.3389/fgene.2015.00275.

Weeks, A. R., Sgro, C. M., Young, A. G., Frankham, R., Mitchell, N. J., Miller, K. A., Byrne, M., Coates, D. J., Eldridge, M. D., Sunnucks, P., et al. (2011). Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications*, 4(6):709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x.

Whitlock, M. C. (2000). Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution*, 54(6):1855–1861. https://doi.org/10.1111/j.0014-3820.2000.tb01232.x.

Willi, Y., Griffin, P., and Van Buskirk, J. (2013). Drift load in populations of small size and low density. *Heredity*, 110(3):296–302. https://doi.org/10.1038/hdy.2012.86.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97. https://doi.org/10.1093/genetics/16.3.290.

Wright, S. (1943). Isolation by distance. *Genetics*, 28(2):114. https://doi.org/10.1093/genetics/28.2.114.

Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, 15(1):323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x.

Yeaman, S. and Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance. *Evolution*, 65(7):1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x.

Chapter 2

The response of a metapopulation to a changing environment[1]

Nick Barton, Oluwafunmilola Olusanya

Abstract

A species distributed across diverse environments may adapt to local conditions. We ask how quickly such a species changes its range in response to changed conditions. Szep *et al* (2021) used the infinite island model to find the stationary distribution of allele frequencies and deme sizes. We extend this to find how a metapopulation responds to changes in carrying capacity, selection strength, or migration rate when deme sizes are fixed. We further develop a "fixed-state" approximation. Under this approximation, polymorphism is only possible for a narrow range of habitat proportions when selection is weak compared to drift, but for a much wider range otherwise. When rates of selection or migration relative to drift change in a single deme of the metapopulation, the population takes a time of order 1/m to reach the new equilibrium. However, even with many loci, there can be substantial fluctuations in net adaptation, because at each locus, alleles randomly get lost or fixed. Thus, in a finite metapopulation, variation may gradually be lost by chance, even if it would persist in an infinite metapopulation. When conditions change across the whole metapopulation, there can be rapid change, which is predicted well by the fixed-state approximation.

Keywords: metapopulation, local adaptation, species' range, diffusion, adaptive walk, changing conditions, soft selection.

[1]This work has been published at https://royalsocietypublishing.org/doi/10.1098/rstb.2021.0009

2.1 Introduction

Species must adapt to varied environments, whilst drawing on a common pool of genetic variation. Thus, there is a tension between selection that favours different alleles in different places, and the maintenance of diversity across the whole species. Local populations can only sustain themselves if they are sufficiently well-adapted; conversely, adaptation to conditions beyond the current niche can extend the range of the species.

These issues, which lie at the interface between ecology and evolution, have only quite recently attracted sustained theoretical attention. This ranges from studies of "evolutionary rescue", typically of a single isolated deme (Bell and Gonzalez, 2011; Bell, 2017; Uecker et al., 2014), through to analyses of limits to a species' range in one or two spatial dimensions (Case and Taper, 2000; Kirkpatrick and Barton, 1997; Polechová, 2018). Here, we consider an idealised metapopulation; in this island model, there is no explicit spatial structure. Nevertheless, we can ask whether the species' range can extend over a variety of habitats, and examine how it responds dynamically to changing conditions – either in a single deme, or across the whole metapopulation.

We examine a simple model, in which directional selection favours alternative alleles in two different habitats. Provided that selection is stronger than migration, these alternative adaptations can be maintained despite gene flow. There is substantial literature on how heterogeneous selection can maintain diversity, beginning with Levene (1953). However, this is largely deterministic, neglecting random drift in small local populations. Here, we are primarily concerned with the erosion of adaptation by random drift within local demes – which can cause a substantial "drift load" even when the whole metapopulation is very large.

This paper is an extension of Szép et al. (2021), which analysed the joint evolution of allele frequencies and deme sizes, in an island model with explicit density-dependent regulation; a diffusion approximation gave explicit formulae for the stationary distribution of an infinite metapopulation. Here, we extend this treatment to consider the evolution of individual demes, and of the whole metapopulation, as conditions change (directly through changes in selection and gene flow, and indirectly through their effect on the population size); we also consider fluctuations in a metapopulation with a limited number of demes, where variation can be lost by chance. We simplify the problem by assuming that deme sizes are fixed, independent of adaptedness ("soft selection"), but believe that the methods we introduce can be extended to allow density regulation ("hard selection").

In principle, we can calculate the joint distribution of deme size and allele frequencies under the diffusion approximation. However, this is numerically challenging, since it involves a high dimensional partial differential equation; in any case, it can only be done for an infinite metapopulation, where the mean population size and allele frequencies

across the population as a whole change deterministically, even though population sizes and allele frequencies within any deme follow a distribution. In order to go beyond mere simulation, we use the approximation that loci are typically near fixation; this is accurate if the number of non-native alleles that enter per generation is small. It allows us to follow the distribution of states of a finite metapopulation through time, which depends only on the rates of substitutions in either direction. This "fixed-state" approximation is an extension of models of "adaptive walks" (e.g. Orr (1998); Trubenova et al. (2019)) to structured populations.

We first consider an infinite metapopulation, and determine the accuracy of the fixed-state approximation. We then apply the approximation to calculate the dynamics of a finite metapopulation, and to find how its equilibria depend on the number of demes. (In order for a non-trivial equilibrium to exist, we must allow a low rate of mutation to maintain variation in the long term). Finally, we show how metapopulations respond to changing conditions, focusing on changes that take the system between qualitatively different regimes.

2.2 Model and Methods

We simulate a haploid population, assuming linkage equilibrium. Provided that selection is weak, this is accurate, and allows us to efficiently simulate large numbers of loci and demes; Szép et al. (2021, SI C) examined the effects of linkage disequilibrium in this model, using individual-based simulations. We obtain analytical results by taking the diffusion limit, which also assumes weak selection, and then approximate this by assuming that demes are near fixation, which applies when there are few migrants (Nm <1). As is traditional in population genetics, we take the fundamental model to be the diffusion, since this captures the behaviour of a variety of particular life histories, and identifies the key dimensionless parameters.

2.2.1 Simulations

Our baseline island model assumes that demes each have carrying capacity of N haploid individuals, and contribute equally to the migrant pool. A deme of size N is expected to lose a fraction m of individuals by emigration, and receives a Poisson distributed number of migrants, Nm^*, with expectation Nm. There are L biallelic loci, with the two alternative alleles labelled $X_{i,k} = 0$ or 1; i labels the deme, and k the locus. Deme i is described by $\{j_{i,1}, j_{i,2}, \ldots, j_{i,L}\}$, where $0 \leq j_{i,k} \leq N$ is the number of copies of the '1' allele at the k'th locus. That allele is favoured by selection s_i, which we assume to be the same across loci; the marginal relative fitnesses are $1 : e^{s_i}$, and fitnesses multiply across

loci. Under soft selection, loci evolve independently, and so it would be straightforward to extend to allow variation in selection across loci.

We assume linkage equilibrium (LE), and apply the Wright-Fisher model to each locus independently. After selection, allele frequencies are $p^*_{i,k} = j_{i,k}/\left((N - j_{i,k})\, e^{-s_i} + j_{i,k}\right)$, and after migration, $p^{**}_{i,k} = m\,\bar{p}_k + (1-m)p^*_{i,k}$ where $\bar{p}_k$ is the frequency averaged across all demes of the metapopulation. The new population in deme i consists of N individuals, the number of allele copies at locus k being binomially sampled with frequency $p^{**}_{i,k}$. This procedure is accurate provided that s is not too large (< 0.2, say), so that recombination shuffles genes faster than selection, drift, or migration build up associations between them (Szép et al.., 2021, SI C).

A Mathematica notebook containing the simulation code can be found in Barton and Olusanya (2022).

Diffusion approximation

The diffusion approximation to this model describes the evolution of the joint distribution of allele frequencies across different demes, conditional on the mean allele frequency across the metapopulation (Barton and Rouhani, 1993). A single deme follows a stochastic path governed by this distribution, whilst an infinite metapopulation represents the whole distribution, which evolves deterministically at the level of the metapopulation. The diffusion depends only on scaled parameters Ns, Nm.

Wright (1937a,b) gave an explicit solution for the stationary distribution of allele frequencies:

$$\Psi\left[p\,|\overline{p}\right] = \frac{1}{Z}\prod_{k=1}^{L} p_k^{2\,N\,m\bar{p}_k - 1} q_k^{2N\,m\bar{q}_k - 1} e^{2Nsp_k} \tag{2.1}$$

where Z is a normalising constant and $q = 1 - p$. Under this simple model of directional selection, allele frequencies evolve independently across demes and across loci, conditional on the mean allele frequencies, $\bar{p}_k$. Equation (2.1) applies to a single deme; the subscript i was dropped for clarity. All demes that share the same parameters will follow the same distribution, in a given habitat, and so we can integrate over the distribution, and sum over habitats, to find the mean $\bar{p}_k$. This allows us to solve fully for the stationary state.

Fixed-state approximation

If the number of incoming alleles is small ($Nm \ll 1$) then the distribution of allele frequencies will be sharply peaked around 0 and 1. To a good approximation, populations are near fixation for one or other allele, and their state is determined by the rates of

substitution in either direction. Since we will later be considering the stationary state of a finite metapopulation, we must include mutation, which we assume to be symmetric at rate μ. Then, the rate at which demes currently fixed for allele 0 substitute allele 1, $\lambda_{0\to1}$ (or vice versa, $\lambda_{1\to0}$) is the product of the number of '1' (or '0') alleles entering the population, and their individual fixation probability (see SM for details). Thus:

$$\lambda_{0\to1} = \frac{2s\,(N\mu + Nm\bar{p})}{1-e^{-2N\,s}}, \qquad \lambda_{1\to0} = \frac{2s\,(N\mu + Nm\bar{q})}{e^{2N\,s}-1} \tag{2.2}$$

Different loci evolve independently, conditional on the numbers of migrants coming into the deme ($Nm\bar{p}$), ($Nm\bar{q}$). Note that when Ns and Nm are small, these rates are proportional to m in the neutral limit.

For an infinite metapopulation, and two habitats with selection s_1, s_2, with deme sizes fixed at N (i.e., soft selection), we can just follow the proportion of demes fixed for the '1' allele in each habitat, P_1, P_2. Neglecting mutation, the rates given by eq. (2.2) lead immediately to:

$$\begin{aligned}
\partial_t P_1 &= \frac{2s_1\,Nm}{1-e^{-2N\,s_1}}\left(\bar{p}Q_1 - \bar{q}e^{-2Ns_1}P_1\right) \\
\partial_t P_2 &= \frac{2s_2\,Nm}{1-e^{-2Ns_2}}\left(\bar{p}Q_2 - \bar{q}e^{-2Ns_2}P_2\right) \\
\bar{p} &= \rho\,P_1 + (1-\rho)P_2
\end{aligned} \tag{2.3}$$

The first two equations involve the difference in net rates of substitution in each direction, where Q, P are the fraction of loci near fixation for 0, 1; $\bar{p}, \bar{q}$ are the fraction of migrants with allele 1 vs 0, which can contribute to a substitution; and the fixation probabilities in each direction are in the ratio $1 : e^{-2Ns_1}$. Finally, the mean allele frequencies are a weighted average across habitats, which are in the proportions $\rho : 1-\rho$.

These equations apply separately to each locus, but for simplicity, in numerical examples we will assume symmetric initial conditions, so that the proportion of demes fixed for the 1 allele in each habitat, P_1, P_2 are the same for all loci, and correspond to the proportion of loci fixed for the '1' allele in each deme.

If the '1' allele is favoured in habitat 1, but disfavoured in habitat 2 (i.e. $s_2 < 0 < s_1$), and if neither habitat is too rare, then polymorphism is possible, with equilibrium frequency given by:

$$\bar{p} = \frac{\rho\left(e^{2N(s_1-s_2)}-1\right) - \left(e^{-2Ns_2}-1\right)}{\left(e^{2Ns_1}-1\right)\left(e^{-2Ns_2}-1\right)}\ , \quad \frac{e^{-2Ns_2}-1}{e^{2N(s_1-s_2)}-1} < \rho < \frac{\left(e^{-2Ns_2}-1\right)}{e^{2N(s_1-s_2)}-1}e^{2Ns_1}, \tag{2.4}$$

(as derived from eq. (2.2); see SM). If selection is weak relative to drift, polymorphism is possible only for a very narrow range of habitat proportions (see left of fig. A.1 in

Appendix A), whereas if it is strong, polymorphism is possible over a wide range (right of fig. A.1 in Appendix A).

Suppose now that there are a finite number of demes, with d_i having habitat i. At any one locus, the state of the metapopulation is described by the number of demes fixed for the '1' allele, $0 \leq k_i \leq d_i$. For example, with two habitats, there are $(d_1 + 1)(d_2 + 1)$ possible values for the state $\{k_1, k_2\}$. The probability of transitions between these states depends on the mean allele frequency across the metapopulation. With soft selection, where all demes have the same size N, this mean is just $\bar{p} = (k_1 + k_2) / (d_1 + d_2)$. We can therefore calculate the transition matrix that governs the stochastic evolution of the metapopulation; the stationary state is given by the leading eigenvector of this matrix. With soft selection, each locus evolves independently, governed by this matrix, and so we can easily calculate the stochastic evolution of the metapopulation.

In Appendix B, we examine the accuracy of the fixed-state approximation under soft selection. This approximation applies in the limit of low migration, and identifies the failure of adaptation due to random drift.

2.3 Results

2.3.1 Evolution of a single deme

Consider a metapopulation, where Nm is small enough that populations are near fixation. If $Ns_1 = 1$ in a rare habitat, represented in $\rho = 0.2$ of the demes, and $Ns_2 = -2$ in the common habitat, then polymorphism will be maintained with $\bar{p} = 0.079$ overall (eq. (2.4)). We begin by considering how a single deme responds to changes in its local conditions, for fixed $\bar{p}$, and so in fact, all that matters is the value of $\bar{p}$. In the focal deme, allele frequencies will be in the ratio $\bar{q} : \bar{p}e^{2Ns_1}$ when $Nm \ll 1$, since that is the ratio of substitution rates in either direction; hence, the expected allele frequency in the rare habitat is 0.386 (lhs of fig. 2.1a). As Nm increases, the expected allele frequency decreases, approaching $\bar{p} = 0.079$ (rhs of fig. 2.1a). For given Nm, the expected allele frequency in the focal deme increases with Ns_1 from $\bar{p}$ to 1, as selection becomes more effective (fig. 2.1b).

Figure 2.2a shows how the distribution of allele frequencies changes as Nm changes. If all loci start close to the frequency in the gene pool ($\bar{p} = 0.079$) then with a low migration rate ($Nm = 0.05$), even weak selection ($Ns = 1$) can raise the mean substantially, to 0.355. However, this increase is slow, taking ~ 5000 generations, because it occurs through occasional substitutions, at a rate proportional to $m = 5 \times 10^{-4}$ (eq. (2.3)). The population does mostly flip between fixation of one or other allele, giving a U-shaped frequency distribution (e.g. grey trajectory in fig. 2.2), and so the fixed-state

approximation is quite close to the exact mean (orange vs. red at left). However, the average across even 100 loci fluctuates substantially (blue), implying that population fitness will fluctuate randomly, even when adaptation is highly polygenic.

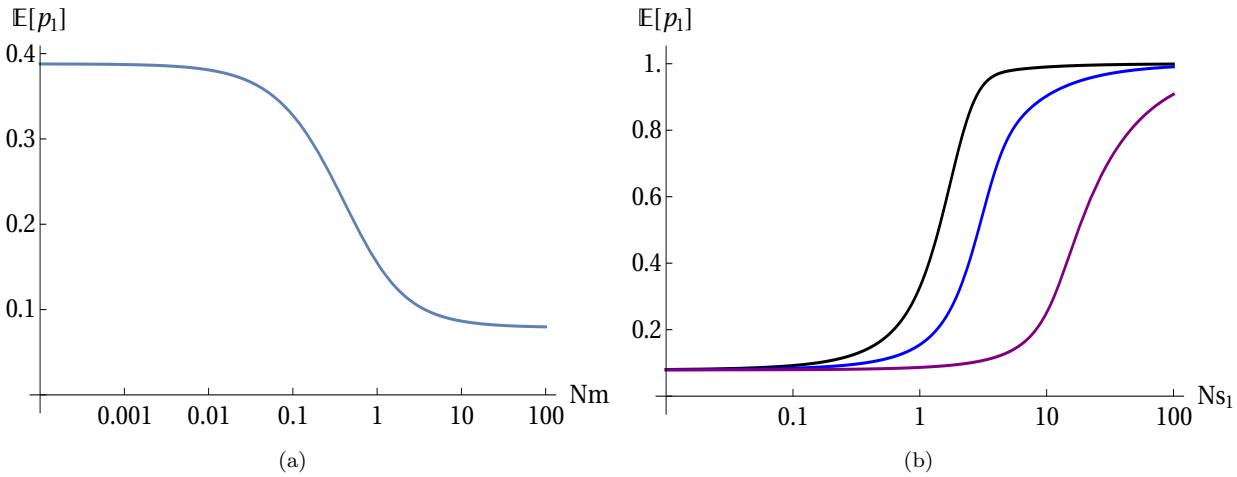
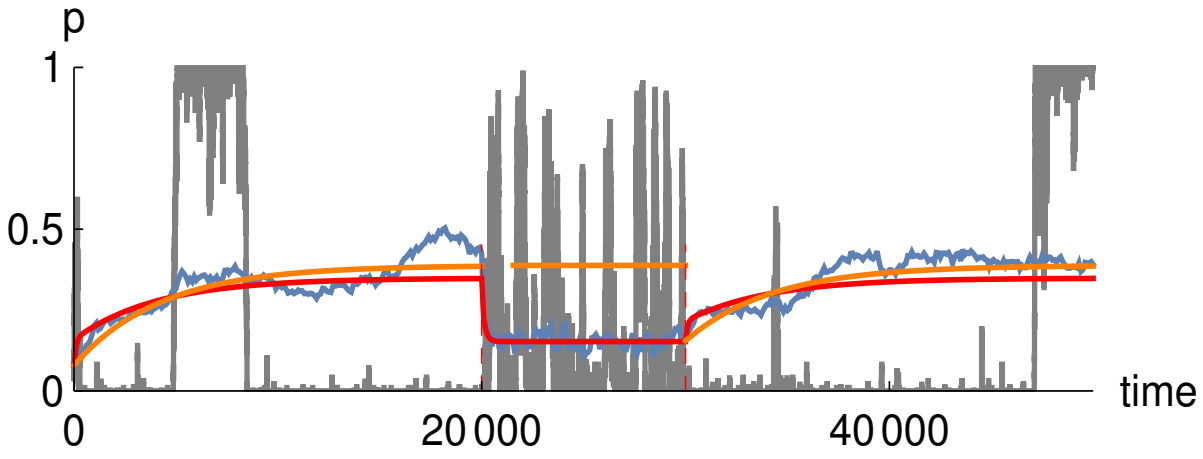


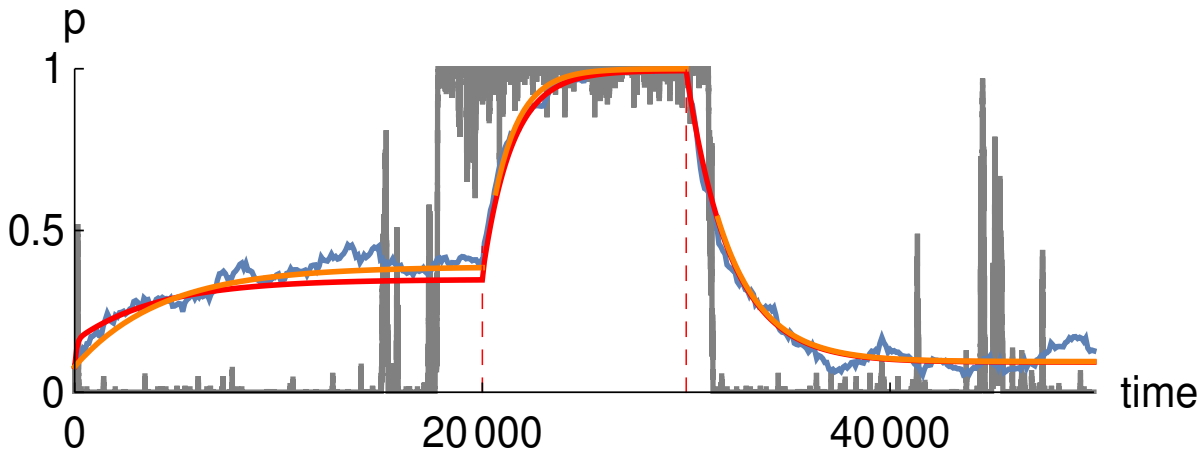
Figure 2.1: (a). Expected allele frequency vs Nm with $\bar{p} = 0.079$, $Ns_1 = 1$. (b). Expected allele frequency vs. Ns, for $Nm = 0.1,\ 1,\ 10$ (black, blue, purple).

At $20,000$ generations, the number of migrants increases to $Nm = 1$, and the mean allele frequency is quickly pulled down towards that in the gene pool, to 0.155. The fixed-state approximation is the limit of low migration, and so is independent of Nm (see fig. A.2). Indeed, allele frequencies are now often intermediate, and so this approximation fails (orange vs. red, fig. 2.2, middle). Nevertheless, it does give the important intuition that rates of change are proportional to migration, which is now $m = 0.01$, implying a $1/m \sim 100$ generation timescale for response of the population mean. In this model, variance is maintained by migration, and so the response to selection is proportional to m: we can see this in eq. (2.3), where rates of change are proportional to m, for given Ns. After Nm returns to the original low value at $30,000$ generations, there is a slow return to the original bimodal distribution, again captured by the fixed-state approximation (orange vs red at right of fig. 2.2).

Figure 2.2b shows the response to changes in Ns, which could arise through changes in selection strength, and/or changes in local effective population size. In this example, $Nm = 0.05$ throughout, and so the fixed-state approximation is accurate (orange vs red curves). The timescale is again set by m, which determines the rate at which variation is introduced into local demes. Since $m = 5 \times 10^{-4}$, it takes thousands of generations for the proportion of loci fixed for the '1' allele to respond to changes in selection strength.



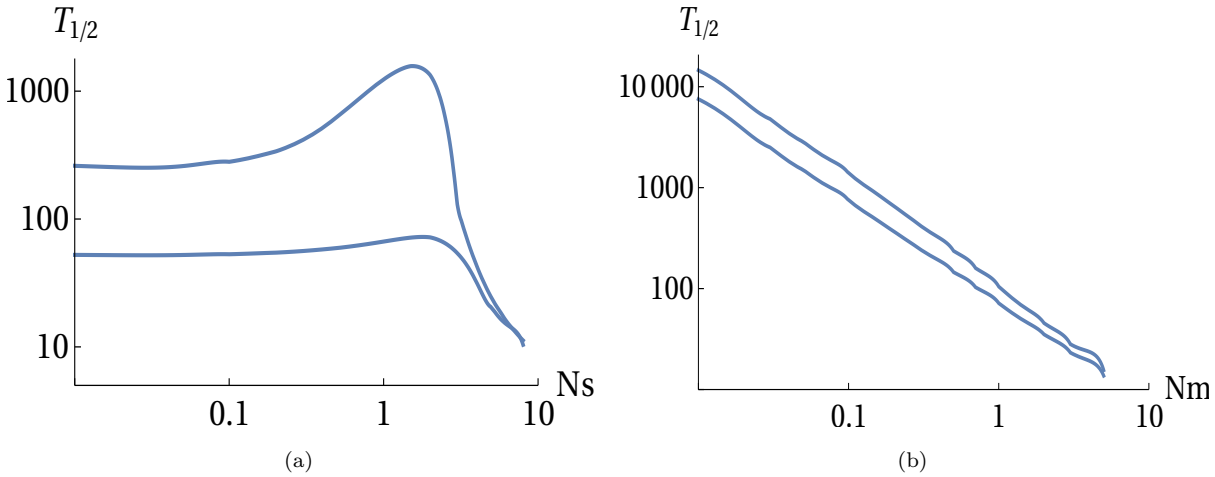
(a)



(b)

Figure 2.2: (a) Evolution of a single deme as Nm changes; $Ns = 1$, $\bar{p} = 0.079$, $L = 100$ loci. Initially, $Nm = 0.05$, and all loci are at $\bar{p}$. After $20,000$ generations, Nm increases to 1, and after another $10,000$ generations, it returns to $Nm = 0.05$. The grey line shows allele frequency at a single locus, and the blue line shows the average over 100 loci. The red curve is the mean of the probability distribution, calculated exactly using the Wright-Fisher transition matrix. The orange curve is the fixed-state approximation (eq. (2.3)), which is accurate only for $Nm \ll 1$. (b) The same as (a), but for Ns changing from 1 to 10 at $20,000$ generations, and then to 0.1 at $30,000$ generations; $Nm = 0.05$ throughout.

Figure 2.3 shows the time taken for a population to respond to changes in Nm (fig. 2.3a) or Ns (fig. 2.3b), as a function of the other parameter. In each figure, the two curves show the time to respond to changes in either direction. As we saw in fig. 2.2a, an increase in Nm (lower plot of fig. 2.3a) causes a much faster response than a decrease (upper plot of fig. 2.3a), simply because high gene flow introduces more genetic variance. However, if selection is very strong, the response time becomes similar in either direction, and decreases in proportion to Ns (right of fig. 2.3a). The response to changes in Ns take somewhat longer for an increase than a decrease (compare upper vs lower plot of fig. 2.3b), but the main pattern here is that the response time decreases in proportion to Nm.



(a) (b)

Figure 2.3: The time to make half of the response to a change in parameters. For both plots, $\bar{p} = 0.079$. Values were calculated using a transition matrix with $N = 100$. Details of the calculation are in the SM. (a) Nm shifts from 0.05 to 1 or from 1 to 0.05 (lower, upper curves, resp.), for varying Ns. (b) Ns shifts from 0.1 to 1 or from 1 to 0.1 (upper, lower curves, resp.) for varying Nm.

2.3.2 Evolution of a metapopulation

We begin by considering the stationary state of a metapopulation, extending Szép et al. (2021) by allowing a finite number of demes – in which case, a low rate of mutation is required to maintain variation in the long term. We then give an example that shows how variation is lost, as loci fix across the whole metapopulation. Finally, we give examples (analogous to fig. 2.2), showing the response when parameters change across the whole metapopulation.

Stationary state of a finite metapopulation in the limit of small Nm

Szép et al. (2021, Fig. 2) show that with soft selection, polymorphism can be maintained in an infinite metapopulation, provided that selection is sufficiently strong. With symmetric selection ($s_1 = s_2$), this requires $Ns > Ns_{crit} = 1/2 \log\left[\frac{1-\rho}{\rho}\right] + Nm(1-2\rho)$; the first term is derived from the fixed-state approximation, in the limit $Nm \ll 1$, and the second from the deterministic model, which requires $s > m(1-2\rho)$ for polymorphism (see Szép et al. 2021, immediately above "Hard selection"). In a metapopulation with a finite number of demes, variation must ultimately be lost: we must include mutation to allow a non-trivial stationary state. In this section, we examine how the outcome depends on the relative rates of selection and drift (Ns) and on the relative rates of mutation and migration (μ/m). In particular, we show that with sufficiently many demes, the outcome is insensitive to the mutation rate.

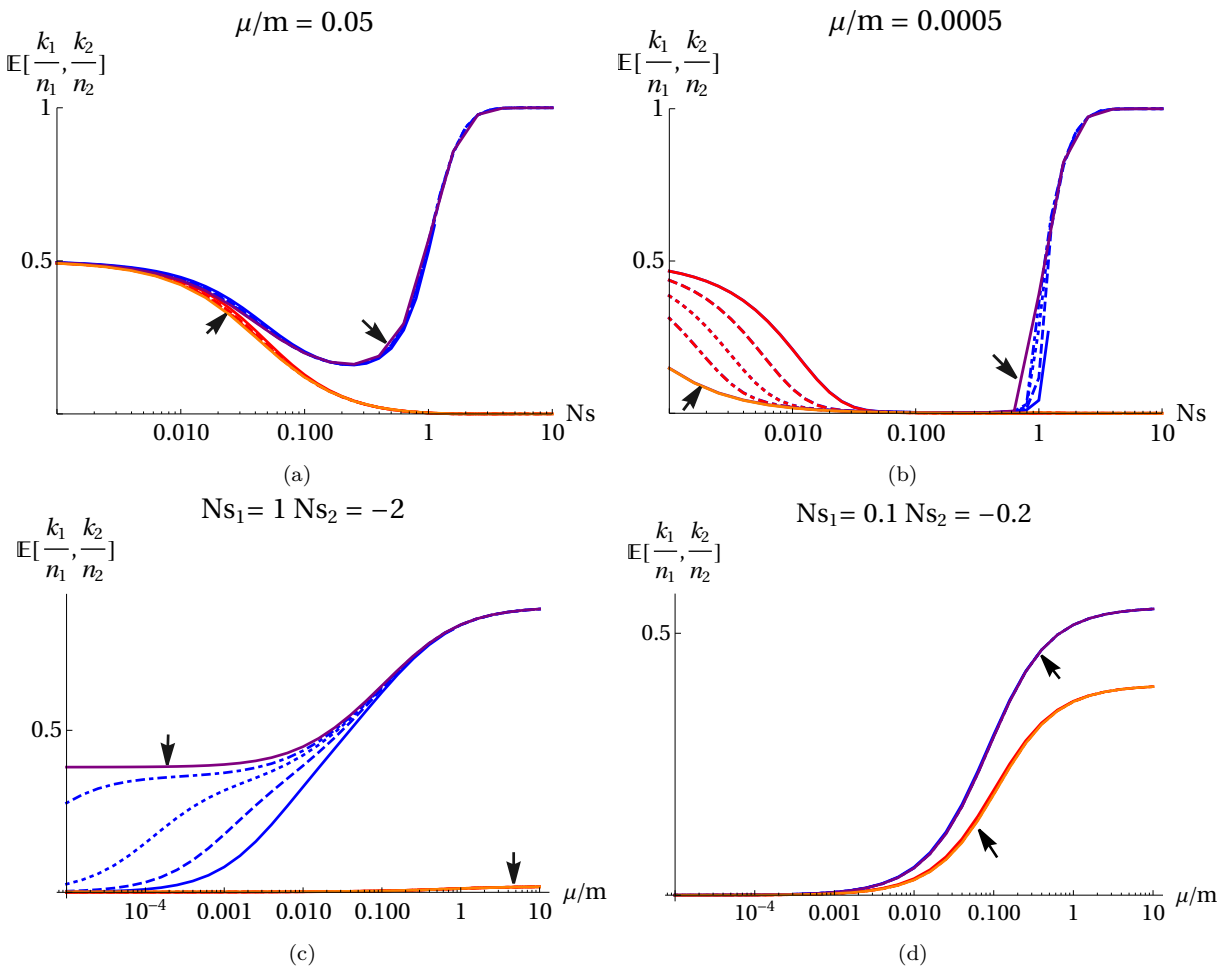


Figure 2.4: The fraction of demes in each of the two habitats (rare habitat – blue, common habitat – red) that are fixed, as a function of selection strength (Ns, top row) and the rate of mutation, relative to migration (μ/m, bottom row). The focal allele is favoured by selection Ns_1 in 20% of demes (i.e. in the rare habitat – blue), and disfavoured by selection $Ns_2 = -2Ns_1$ in the remaining 80% of demes (i.e. in the common habitat – red). In each plot, equilibria for 50, 100, 200 and 400 demes are superimposed (solid, dashed, dotted and dot-dashed lines), together with the limit for an infinite metapopulation (solid purple and orange lines indicated by arrows).

Figure 2.4 shows the stationary state in the limit of small Nm, derived using the fixed-state approximation. The top row of fig. 2.4 shows how the fraction of demes fixed at equilibrium in a rare and common habitat (i.e. $\mathbb{E}[k_1/n_1]$ and $\mathbb{E}[k_2/n_2]$ respectively) depend on the strength of selection when mutation is appreciable (fig. 2.4a) compared to when it is weak (fig. 2.4b), where k_i, $i = 1, 2$, are as defined earlier, n_1 and n_2 are the total number of demes in the rare and common habitats respectively, $n = n_1 + n_2$ is the total number of demes in the metapopulation and mutation is assumed symmetric. Out of all demes, n, in the metapopulation, the focal allele is favoured in 20% of demes (constituting the rare habitat) so that $n_1 = 0.2\, n$, and disfavored twice as strongly in the remaining 80% of demes (constituting the common habitat) so that $n_2 = 0.8\, n$. The blue (red) color represents dynamics in the rare (common) habitat when we have a finite metapopulation (i.e. $n = 50,\ 100,\ 200,$ and 400 demes represented by the solid, dashed,

dotted and dot-dashed lines respectively). Whereas, the purple (orange) color represent dynamics in the rare (common) habitat when we have an infinite metapopulation.

When mutation is appreciable ($\mu/m = 0.05$, fig. 2.4a), the focal allele is unlikely to be lost by chance (blue colors). Furthermore, the equilibrium is insensitive to the number of demes, and close to the solution for an infinite population (as can be seen from the indistinguishability of the various lines for different N). When selection is strong (right of fig. 2.4a, all demes are fixed for the favoured allele, whereas when selection is negligible, on average half of the demes are fixed for each allele. In-between (i.e. $0.1 < Ns < 1$), the allele favoured in the rare habitat becomes rare, being pulled to low frequency by migration from the commoner habitat, where it is more strongly disfavoured. When mutation is weak relative to migration (as is likely in nature, i.e. $\mu/m = 0.0005$; fig. 2.4b), this pattern is exaggerated. Above a critical value, $Ns_{crit} \sim (1/2)\log\left[\frac{1-\rho}{\rho}\right] \sim 0.7$, polymorphism can be maintained by divergent selection, despite drift and gene flow. The equilibrium for an infinite population (purple solid line, indicated by arrow) gives an upper bound, but stochastic loss from a finite set of demes reduces the expected frequency, and increases the critical Ns_{crit} (solid, dashed, dotted and dot-dashed blue lines around $Ns \sim 1$, for $50, 100, 200$ and 400 demes respectively). There is a wide region ($0.03 < Ns < 0.7$) where the allele is almost absent, being swamped by gene flow. However, for very weak selection (fig. 2.4b, lhs), the frequency of the allele increases towards the symmetric neutral equilibrium i.e. 0.5. In this regime, the frequencies in the two habitats are almost identical, and cannot be distinguished in the figure. Furthermore, in this regime (left of fig. 2.4a), although selection is negligible within demes ($Ns < 0.1$), migration is much faster than mutation, and so selection over the whole metapopulation is effective in eliminating the allele that is deleterious in most demes. Therefore in this scenario where mutation is rarer than migration, and selection is weak relative to drift within a single deme (i.e. $\mu \ll m$, $Ns < 0.1$), selection is nevertheless still effective at the level of the whole metapopulation and is especially so in the habitat which has more demes (left of fig. 2.4a).

The bottom row of fig. 2.4 shows the dependence on the relative rates of mutation versus migration, μ/m. With high mutation rates, the equilibrium approaches a fraction $\mathbb{E}[k/n] = 1/(1 + e^{-2Ns})$, given by the ratio of fixation probabilities (eq. (2.2)) in the fixed-state approximation. There is strong divergence when $Ns_1 = 1$ (right of fig. 2.4c), and weaker divergence when selection is weak (fig. 2.4d, $Ns_1 = 0.1$). With moderately strong selection (fig. 2.4c), the allele that is less favoured overall is lost from the common habitat, independent of the number of demes and mutation rate (orange line with arrow head). In the rare habitat, with weak mutation (left of fig. 2.4c), the locally favoured allele can be fixed in nearly half the demes in an infinite metapopulation (purple solid line with arrow head), but tends to be lost by chance from finite metapopulations, even with several

hundred demes (solid, dashed and dotted blue lines). When selection is weak relative to local deme size (fig. 2.4d), selection can still be effective over the whole metapopulation, eliminating the allele that is disfavoured overall (left of fig. 2.4d). However, when mutation becomes comparable with migration, polymorphism is maintained by mutation pressure, with some bias between habitats caused by weak selection (right of fig. 2.4d).

We focus on the regime where selection is comparable to drift ($Ns_1 \sim 1$), and mutation is weak. This corresponds to the right half of fig. 2.4b ($0.1 < Ns_1$), and the middle of fig. 2.4c ($10^{-4} < \mu/m < 0.1$). Then, as long as mutation is not extremely small, and there are enough demes, the stationary state is close to that in an infinite metapopulation (compare blue dashed and dotted lines with purple line in fig. 2.4c). However, note that with weak mutation ($\mu/m \sim 10^{-4} - 10^{-3}$, say), the locally favoured allele tends to be lost even when there are several hundred demes.

Loss of diversity in a finite metapopulation

When deme sizes are fixed, and numbers of migrants are low enough that loci are typically fixed for one or other allele, the state of the metapopulation at each locus can be described by the number of demes, k_i, in each habitat, i, that are fixed for the '1' allele. The distribution of k_i evolves according to a transition matrix, and each locus follows an independent realisation of the same stochastic process. Figure A.3 and A.4 in Appendix C compares the dynamics of the fixed-state approximation with simulations, to illustrate its accuracy. For $Nm = 0.05$ (fig. A.3), there is reasonable agreement between simulations and the fixed state approximation and for lower Nm (i.e. $Nm = 0.01$), agreement is even more close (fig. A.4). In both cases, variation is lost faster than predicted by the fixed-state approximation (compare red and black lines of fig. A.3a and A.4a), because migration tends to swamp adaptive divergence. The timescale is inversely proportional to m, and is therefore slow. Here, we are focusing on the slow loss of adaptation through random drift in small populations; with higher migration rates, swamping by gene flow causes additional, faster, degradation.

Note that because the number of demes is limited, and because each deme flips between fixation for alternative alleles, there is substantial variability in average allele frequency between loci (grey lines). Therefore, adaptation is lost slightly faster in a finite than in an infinite metapopulation (compare black and magenta lines in both fig. A.3a and A.3b, which both derive from the fixed-state approximation). Nevertheless, the overall mean, averaged over 40 loci, changes smoothly and predictably (red curves in fig. A.3 and A.4). We assume no mutation, and so all variation will inevitably be lost. However, because the total population is large (i.e. 100 demes $\times$ 50 individuals per deme $=$ 5000 individuals), and because very low migration rate increases the effective size of the whole metapopulation, loss across the whole metapopulation is extremely slow: none of the 40

loci fix during the 10^4 generations shown in fig. A.3 and A.4 (grey lines).

Response to changing conditions across the metapopulation

Figure 2.5 shows some examples of the response to a change in conditions when parameters change gradually in all demes of the rare habitat. We focus on the rarer habitat since we are mostly concerned with adaptation in the rare habitat and its degradation by gene flow.

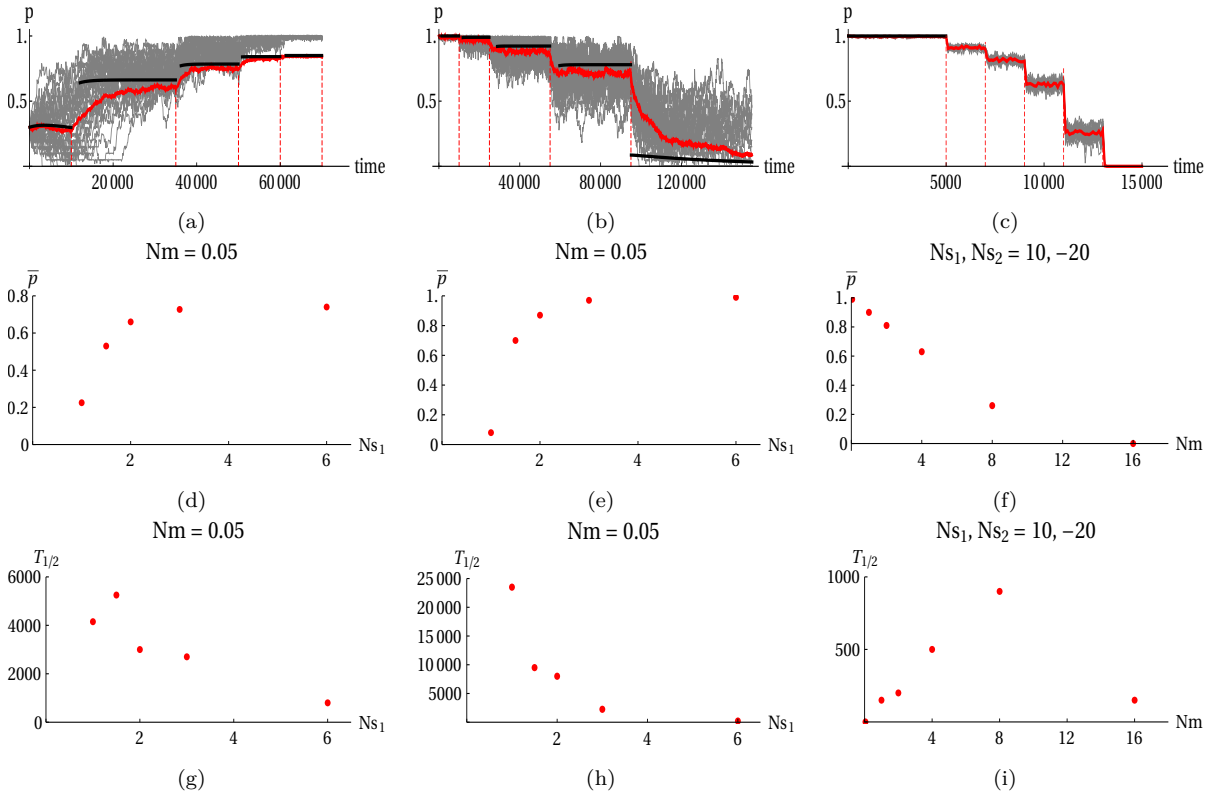


Figure 2.5: (A)-(C) Response of a metapopulation to changing conditions. Grey lines show the allele frequencies, averaged over the 20 demes in the rare habitat, at each 40 loci; the red lines show the overall mean in the rare habitat. The black line shows the prediction in the limit of small Nm (i.e. the fixed state approximation). In figure A. Nm is kept fixed at 0.05 and Ns_1 is increased gradually from $1 \to 1.5 \to 2 \to 3 \to 6$ with Ns_2 fixed at 2. In figure B., we have the reverse of figure A. where Nm is again kept fixed at 0.05, Ns_2 kept fixed at 2 and Ns_1 is now decreased gradually from $6 \to 3 \to 2 \to 1.5 \to 1$. In figure C. Ns_1 and Ns_2 are kept fixed (at 10 and -20 respectively) and Nm is gradually increased from $0.05 \to 1 \to 2 \to 4 \to 8 \to 16$. (D)-(F) Plot of $\overline{p}$ against the changing parameter. (G)-(I) Plot of $T_{1/2}$ (i.e. half time to reach the new equilibrium $\overline{p}$) as a function of the changing parameter. Each point in fig. 2.5d – 2.5i is based on a single replicate. For all plots, simulations are run with 100 demes, $N = 50$, $L = 40$ loci and $\rho = 0.2$.

Figure 2.5a, 2.5d and 2.5g respectively show the consequence of a gradual increase in Ns_1 on the distribution of allele frequencies in the rare habitat, the equilibrium allele frequency, $\overline{p}$ (averaged across all demes in the rare habitat) and the half time, $T_{1/2}$, taken to reach this equilibrium mean frequency. We begin initially with selection comparable to drift ($Ns = 1$) and with a fraction of demes fixed for the locally favored allele, in the

proportions predicted by an infinite metapopulation. After several thousand generations when the local allele has reached an equilibrium, we gradually increase Ns_1 until a new equilibrium is obtained and do this continuously until all loci in the rare habitat reach (near) perfect adaptation (fig. 2.5a, grey lines). With selection comparable to drift, there is a substantial variation across loci with the 40 individual loci following markedly different paths and a chance loss of 6 of the loci (due to the effect of drift). As selection becomes stronger, the remaining polymorphic loci rise in frequency. Although there is still considerable variation in the rates of increase across loci (20,000 – 50,000 generations, fig. 2.5a), the overall mean, $\overline{p}$ is approached quite smoothly (red lines). These new equilibrium mean values have a positive dependence on Ns_1 (fig. 2.5d) and it takes a shorter time to reach a new equilibrium with stronger selection (fig. 2.5g).

Figure 2.5b, 2.5e and 2.5h correspondingly show a similar scenario as the above but with Ns_1 now changing in the opposite direction (i.e. from high to low value). Initially, with selection much stronger than drift, all 40 loci are at a considerably high frequency (near 1) with less variability amongst them (lhs of fig. 2.5b, grey lines) so that the equilibrium mean allele frequency ~ 1 (fig. 2.5e, rhs). Also, because of this strong selection, the new equilibrium is approached rather fast (fig. 2.5h, rhs). As Ns_1 declines, there is an apparent increase in the variability amongst loci so that with $Ns_1 = 1$ drift becomes sufficient enough to cause the loss of three quarters of the loci (rhs of fig. 2.5b, grey lines). Because only 10/40 of the loci remain polymorphic, the overall mean is ~ 0.1 (fig. 2.5e, lhs).

Figure 2.5c, 2.5f and 2.5i show the response (of the distribution of allele frequency, the equilibrium mean frequency, $\overline{p}$ and the half time to $\overline{p}$) to an increase in Nm with $Ns_1,\ Ns_2 = 10,\ -20$ throughout. Since selection is generally strong relative to drift ($Ns \gg 1$) and gene flow is initially low ($Nm = 0.05$, fig. 2.5c, lhs), the rare habitat is initially perfectly adapted with $\overline{p} \sim 0.99$ (fig. 2.5f, lhs). However, as Nm increases, gene flow gradually degrades adaptation as the focal allele is rapidly swamped until it is lost from the population after $\sim 13,000$ generations (fig. 2.5c, rhs). Since 0 loci remain polymorphic at this point, $\overline{p} \sim 0$ (fig. 2.5f, rhs). The half time to reach the new equilibrium values depend non-monotonically on Nm.

In fig. 2.5a – 2.5c, the dynamics are closely predicted by the small Nm limit (compare red and black curves) which is based simply on the rates of substitution in both direction.

2.4 Discussion

Our analysis uses simulation, the diffusion approximation and the "fixed-state" approximation to understand how drift degrades adaptation in a finite metapopulation as well as how a finite metapopulation changes through time, as it responds to changes in both local

and global conditions. The "fixed state" approximation applies either where variation is due to mutation (when it is plausible that $N\mu < 1$ within local demes, or even at the level of the whole population), or when variation is maintained by divergent selection across the whole metapopulation, but migration is low relative to drift ($Nm < 1$).

Using the fixed-state approximation, it was shown that when selection is weaker than drift (i.e. $Ns < 1$), polymorphism can only be maintained for a very narrow range of habitat proportions (fig. A.1). However with strong selection, this range becomes much wider. Also, using both simulations and the fixed state approximation, we showed that when conditions in a single deme of the metapopulation change, the population responds on a short time scale of order $1/m$, simply because in the regime we study, local genetic variance is maintained by migration. Variation may be temporarily lost as local conditions change, but can quickly be recovered. On the other hand, when conditions change across the whole metapopulation, variation that was maintained by divergent selection can be permanently lost, and is only slowly recovered by mutation. Even under constant conditions, variation at a locus can be lost by chance, unless there are a very large number of demes.

To simplify our analysis, we assumed an island model, with a large number of spatially equivalent demes (i.e. soft selection). This is unlikely to be the case in nature, but may nevertheless capture the behaviour of spatially extended populations if there is long-range migration, which can introduce locally adaptive alleles from a distant habitat. It may be that a leptokurtic dispersal distribution can allow efficient adaptation, if locally favoured alleles are not swamped, and yet can be recovered by occasional long-range migration (Atkinson et al., 2002; Fric and Konvicka, 2007).

Our analysis can however be further extended to hard selection, by including explicit density regulation; Szép et al. (2021) showed that one can still apply the diffusion approximation, provided that growth rates are not too high. With hard selection, substitution rates now depend on deme size through Ns, and through the number of immigrant alleles, $m\overline{Np}$, $m\overline{Nq}$. This dependence can be approximated by assuming that the population size is determined by the genetic load. Sachdeva et al. (2022) and Szép et al. (2021) refer to this as the "semi-deterministic" approximation which is accurate when demographic stochasticity is weak. One can apply the "fixed-state" approximation by further assuming that there are enough loci that the mean load is proportional to the mean across loci of the number of demes fixed for one or the other allele. The transition matrix can then be calculated as before, but is now a function of the population sizes in the two habitats, $\{N_1, N_2\}$ which both depend on the current state via the load. The key assumption here is that with enough loci, the population sizes change almost deterministically, following the distribution of states across loci. One complication with hard selection however is the existence of multiple stable equilibria: changing conditions would not just cause equilibria to shift but also changes the rates of transitions between equilibria.

A key assumption in our analysis is that selection is directional: in a given environment, alleles experience a fixed selection pressure, which tends to drive out variation. More often, selection may favour an intermediate optimum for a quantitative trait (i.e., stabilizing selection), such that when the mean is well-adapted, alleles are close to neutral. Our modelling framework can describe this case, but it is much more complex, since many different allele combinations can achieve the same optimum. However, if selection on each allele is weak ($Ns < 1$), then the infinitesimal model (Barton et al., 2017) applies, and can also describe the population dynamics (Barton and Etheridge, 2018). Local adaptation may be possible under higher migration rates in such a regime.

We have shown that it is very difficult for directional selection to maintain local adaptations when selection is weak, relative to deme size ($Ns \sim 1$). Migration must also be weak if it is not to swamp adaptation, in which case alleles are typically near to fixation, limiting the genetic variance available for adaptation, and preventing the recovery of variation lost to random drift. This contrasts with *global* adaptation: selection can be effective across the whole metapopulation, even if selection is weaker than drift within local demes ($Ns < 1$), provided that there is sufficient gene flow ($Nm > 1$) (Barton and Rouhani, 1993). Thus, we expect local adaptation to depend on relatively more strongly selected alleles than global adaptations (Yeaman, 2015).

In this work, we have also introduced a novel approach to understanding the dynamical evolution of metapopulations. Although the full behaviour requires simulation, the diffusion approximation allows the stationary state to be calculated, and identifies the key dimensionless parameters. Moreover, when migration is rare, we can use a fixed state approximation that connects population genetics with models of adaptive walks (Orr, 1998).

Our work suggests several open questions that invite theoretical study. First, although we show that local adaptation requires that directional selection be stronger than drift within demes, that may not be the case with stabilizing selection. Under the infinitesimal model (Barton et al., 2017), genetic variance may be due to weakly selected alleles ($Ns < 1$), and yet can still sustain adaptation of polygenic traits. Second, local adaptation may be greatly impeded by hard selection: if maladapted populations collapse, they cannot be the site of future adaptation. Testing the theory in nature is challenging, because it requires measurement of fitness as well as genetic data. Nevertheless, it might be possible to find how manipulation of local deme size and gene flow (or natural variation in these parameters) alters fitness. If we know of loci responsible for local adaptation (e.g Pfeifer et al. (2018); Jones et al. (2012)), then the theory developed here can be applied more directly, though in practice it would need to be modified to account for actual spatial structure.

In conclusion, the methods introduced in this study allow us to explore how species adapt to diverse environments: we can find how organisms expand their range to a wider span of environments, through local adaptation that is sustained by variation maintained across the whole metapopulation. Many questions remain to be studied within this framework: for example, how populations adapt to large numbers of diverse habitats; whether population regulation ("hard selection") leads to a feedback that impedes adaptation; and whether genetic variation can be better sustained under stabilizing selection towards a varying optimum, rather than the directional selection studied here. We believe that, despite the absence of explicit spatial structure, this approach will be a fruitful way to better understand what limits the range of environments that a species can occupy.

References

Atkinson, R., Rhodes, C., Macdonald, D., and Anderson, R. (2002). Scale-free dynamics in the movement patterns of jackals. *Oikos*, 98(1):134–140. https://doi.org/10.1034/j.1600-0706.2002.980114.x.

Barton, N. and Olusanya, O. (2022). The response of a metapopulation to a changing environment. https://git.ista.ac.at/oolusany/the-response-of-a-metapopulation-to-changing-conditions.

Barton, N. and Rouhani, S. (1993). Adaptation and the 'shifting balance'. *Genetics Research*, 61(1):57–74. https://doi.org/10.1017/S0016672300031098.

Barton, N. H. and Etheridge, A. (2018). Establishment in a new habitat by polygenic adaptation. *Theoretical Population Biology*, 122:110–127. https://doi.org/10.1016/j.tpb.2017.11.007.

Barton, N. H., Etheridge, A. M., and Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical population biology*, 118:50–73. https://doi.org/10.1016/j.tpb.2017.06.001.

Bell, G. (2017). Evolutionary rescue. *Annual Review of Ecology, Evolution, and Systematics*, 48:605–627. https://doi.org/10.1146/annurev-ecolsys-110316-023011.

Bell, G. and Gonzalez, A. (2011). Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. *Science*, 332(6035):1327–1330. https://doi.org/10.1126/science.1203105.

Case, T. J. and Taper, M. L. (2000). Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. *The american naturalist*, 155(5):583–605. https://doi.org/10.1086/303351.

Fric, Z. and Konvicka, M. (2007). Dispersal kernels of butterflies: power-law functions are invariant to marking frequency. *Basic and Applied Ecology*, 8(4):377–386. https://doi.org/10.1016/j.baae.2006.06.005.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392):55–61. https://doi.org/10.1038/nature10944.

Kirkpatrick, M. and Barton, N. H. (1997). Evolution of a species' range. *The American Naturalist*, 150(1):1–23. https://doi.org/10.1086/286054.

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, 87(836):331–333. https://doi.org/10.1086/281792.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, 52(4):935–949. https://doi.org/10.2307/2411226.

Pfeifer, S. P., Laurent, S., Sousa, V. C., Linnen, C. R., Foll, M., Excoffier, L., Hoekstra, H. E., and Jensen, J. D. (2018). The evolutionary history of nebraska deer mice: local adaptation in the face of strong gene flow. *Molecular Biology and Evolution*, 35(4):792–806. https://doi.org/10.1093/molbev/msy004.

Polechová, J. (2018). Is the sky the limit? on the expansion threshold of a species' range. *PLoS biology*, 16(6):e2005372. https://doi.org/10.1371/journal.pbio.2005372.

Sachdeva, H., Olusanya, O., and Barton, N. (2022). Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity. *Philosophical Transactions of the Royal Society B*, 377(1846):20210010. https://doi.org/10.1098/rstb.2021.0010.

Szép, E., Sachdeva, H., and Barton, N. H. (2021). Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5):1030–1045.

Trubenova, B., Krejca, M. S., Lehre, P. K., and Kötzing, T. (2019). Surfing on the seascape: Adaptation in a changing environment. *Evolution*, 73(7):1356–1374. https://doi.org/10.1111/evo.13784.

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

Wright, S. (1937a). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences*, 23(6):307–320. https://doi.org/10.1073/pnas.23.6.307.

Wright, S. (1937b). The distribution of gene frequencies in populations. *Science*, 85(2212):504. https://doi.org/10.1126/science.85.2212.504-a.

Yeaman, S. (2015). Local adaptation by alleles of small effect. *The American Naturalist*, 186(S1):S74–S89. https://doi.org/10.1086/682405.

Chapter 3

Local adaptation in a metapopulation - a multi-habitat perspective

Oluwafunmilola Olusanya, Nick Barton, Jitka Polechova

Abstract

This research extends previous soft selection models of local adaptation in metapopulations containing two habitats to a multi-habitat scenario where each habitat experiences distinct selection pressures. In particular, we focus on a three-habitat example where selection favors an allele A in habitat 1, disfavors it in habitat 3 and A is either favored or disfavored in habitat 2 depending on the locus under consideration. Employing the diffusion and fixed state approximations while assuming linkage equilibrium, we explore the conditions for the persistence of a polymorphism and derive analytical formulae for critical thresholds required for such persistence. Our findings reveal that the range of polymorphism is notably wider under conditions of restricted gene flow between demes and this is more pronounced when the habitat where the allele is favored is similar in proportion to the habitat where it is maladaptive. Furthermore, we demonstrate that the maintenance of polymorphism is significantly shaped by the balance between selection and migration, with strong selection relative to migration being a key driver for continued persistence. Overall, this study sheds light into how the interplay between migration, selection, drift and habitat proportions shape polymorphism in metapopulations, providing insights into evolutionary mechanisms that drive and sustain genetic diversity in fragmented populations.

Keywords: metapopulation, local adaptation, polymorphism, selection, gene flow, migration load, drift.

3.1 Introduction

In evolutionary biology, local adaptation refers to the process by which populations evolve unique traits and genetic variations to better suit their respective habitats (Williams, 2018). These adaptations can encompass changes in an organism's physiology (e.g., metabolic rate), morphology (e.g., body size) or other attributes. Local adaptation plays an important role in the evolution and survival of populations in heterogeneous environments. It can be a driver for speciation (i.e., the formation of new species) when populations become increasingly specialized to their local conditions and lose their capacity to interbreed with individuals from other populations (Gavrilets, 2003). On the contrary, its loss can have serious consequences for species and for the ecosystem they inhabit (Walters and Berger, 2019). For example, it can lead to increased vulnerability to environmental changes and extinction risks as well as the disruption of ecological interactions (Frankham et al., 2017; Urban, 2015). Therefore, understanding the dynamics and factors that promote or constrain such adaptations is crucial for conservation and management efforts. This understanding can help in designing strategies to preserve genetic diversity within metapopulations and further help to predict the response of populations to changing environmental conditions such as climate change and habitat fragmentation.

Local adaptation typically emerges from complex interactions between ecological and evolutionary processes. Ecological processes constitute the interactions between organisms and their environment. When populations are distributed through space (as a result of either natural or human-induced activities, like habitat fragmentation), they are exposed to different environmental conditions such as variations in abiotic factors like temperature, resource availability, climate, etc., or biotic factors like predation and competition. These environmental differences can engender variation in the selective pressures faced by these populations, promoting the emergence of local adaptations and genetic differentiation among them. A typical example of this are metapopulations of pocket gopher (*Thomomys spp*) in the Great Basin region of North America (Rogers, 1991) where populations residing in meadow habitats characterised by depressions (or valleys) with high water availability, have evolved narrower skulls and longer claws for burrowing through soft, moist soils. Whereas, those residing in the sagebrush steppe in more arid regions, have evolved broader skulls and shorter claws to help them traverse drier, more compact soils. These adaptations have enhanced survival and reproduction within each local gopher population.

Evolutionary forces also greatly influence local adaptation. These forces include natural selection, mutation, recombination, gene flow (due to migration) and genetic drift. Strong selection pressures enhance the spread and accumulation of locally beneficial mutations within sub-populations, increasing fitness as a whole. Mutations generate the novel genetic

variants needed for adaptation, and recombination rearranges genetic material, providing novel allelic combinations that can contribute to such adaptations. Migration plays an integral role as it connects demes or sub-populations within a metapopulation. It can however have discordant consequences (Blanquart et al., 2012; Sachdeva et al., 2022) – it can engender genetic diversity thereby facilitating adaptation through the proliferation of advantageous traits across the metapopulation, yet can nevertheless also lead to the introduction of maladapted alleles to an otherwise perfectly adapted deme, thus disrupting the local gene pool, and reducing fitness. This "migration load" may drastically reduce adaptation, increasing the risk of extinction (Holt and Gomulkiewicz, 1997). Finally, genetic drift, the random fluctuation in allele frequency within a population also interferes with local adaptation. This is particularly true in small populations (LaBar and Adami, 2017; Whitlock, 2000) that are more susceptible to chance events which causes the loss or fixation of particular alleles despite their adaptive value (Blanquart et al., 2012).

Metapopulation models of local adaptation typically distinguish between two modeling paradigms: hard and soft selection. The hard selection model considers an explicit feedback between ecological processes (in particular, population size) and evolutionary processes (allele frequencies at different loci) in shaping metapopulation dynamics and considers the possibility of both local and global extinction resulting from maladaptation within the metapopulation (Haldane, 1956; Szép et al., 2021). Whereas, the soft selection model, which is a useful simplification, ignores this feedback and assumes a constant population size for each subpopulation over time despite the level of maladaptation. In this work, we explore the latter kind of model, as this allows us to disentangle the dynamics of local adaptation and the maintenance of genetic variation from the confounding effects of demographic fluctuations, thus providing us with a basic understanding of the evolutionary processes at play as well as insights into the stability of the metapopulation.

An important concern when exploring these models are the assumptions about the metapopulation landscape. Theoretical models exploring the dynamics of local adaptation with constant deme sizes have historically focused on the interactions between two patches or habitats (Hoekstra et al., 1985; Barton and Whitlock, 1997; Lenormand, 2002; Blanquart et al., 2012; Bolnick and Otto, 2013; Szép et al., 2021; Barton and Olusanya, 2022). However, since nature is more intricate, many species experience environmental gradients that span more than two distinct environments, which calls for the expansion of theoretical models to account for more than two habitats simultaneously. This will not only enable us better capture the more complex nature of adaptation across a heterogeneous landscape but will also provide insights into the relative significance of local adaptation versus gene flow in shaping population divergence and maintaining genetic diversity.

In this study, we therefore focus our attention on a metapopulation with more than two habitats. Specifically, we concentrate on simple soft selection models, aiming to

identify the key factors or conditions that facilitate the persistence of genetic variation (or polymorphism) in such metapopulations. To achieve this, we rely solely on mathematical theory formulated based on the diffusion approximation. This approach allows us to describe the dynamics of allele frequency under soft selection.

3.2 Model and Methods

We consider an infinite-island metapopulation model (Wright, 1931) where the metapopulation comprises a large number of demes spread across n habitats. Each habitat ($i = 1, \ldots, n$) exhibits local adaptation via a polygenic trait, Z_i (where the subscript, i is used here to indicate that the same genotype can be associated with different trait values in the different habitats) and Z_i is influenced by l loci. The demes are connected by migration so that every generation, they each send out migrants at a rate m into a migrant pool from which these migrants are subsequently redistributed (with equal probability) to the different demes. Within each deme, individuals are randomly mating haploids, implying two potential genotypes: A and a per locus, and loci contribute multiplicatively to individual fitness. We also ignore both forward and backward mutations between A and a so that populations rely on existing genetic variation for adaptation. This is valid in the limit where mutations are rare and where there are very many demes (Wright, 1931) so that polymorphism can be maintained by other processes such as migration and selection even in the absence of new mutations.

Our model incorporates an antagonistic environmental effect, where selection favors different allelic variants depending on the specific environmental condition. Put simply, selection pressure varies across the metapopulation, leading to distinct fitness advantages of $Z_i := \sum_{j=1}^{l} s_{i,j}$ in i. Consequently, what may enhance fitness in one habitat could potentially reduce fitness in another habitat. In particular, we assume that fitness of the A/a allele at any given locus j ($j = 1 \ldots, l$) follows the ratio $e^{s_{i,j}} : 1$, where $s_{1,j} > 0$ (i.e., A is always favored) in habitat 1, $s_{n,j} < 0$ (A is always disfavored) in habitat n, and may have positive or negative values in the other $n-2$ habitats so that A can either be favored or disfavored depending on the locus and habitat.

While our model generalizes to any number of habitats, n, to reduce the complexity of our analytical derivations, we focus primarily on a specific example involving three habitats (i.e., $i = 1, 2, 3$) where each habitat is in proportion, α_1, α_2 and α_3 respectively. First, we consider a scenario where $s_{1,j} = s_{3,j}$ (in magnitude) and $s_{2,j} = 0$ (so that selection is neutral in habitat 2). This implies that the fitness of the A allele at any locus j in the three habitats are respectively in the ratio $e^{s_{i,j}} : 1 : e^{-s_{i,j}}$ as illustrated in fig. 3.1 (scenario 1).

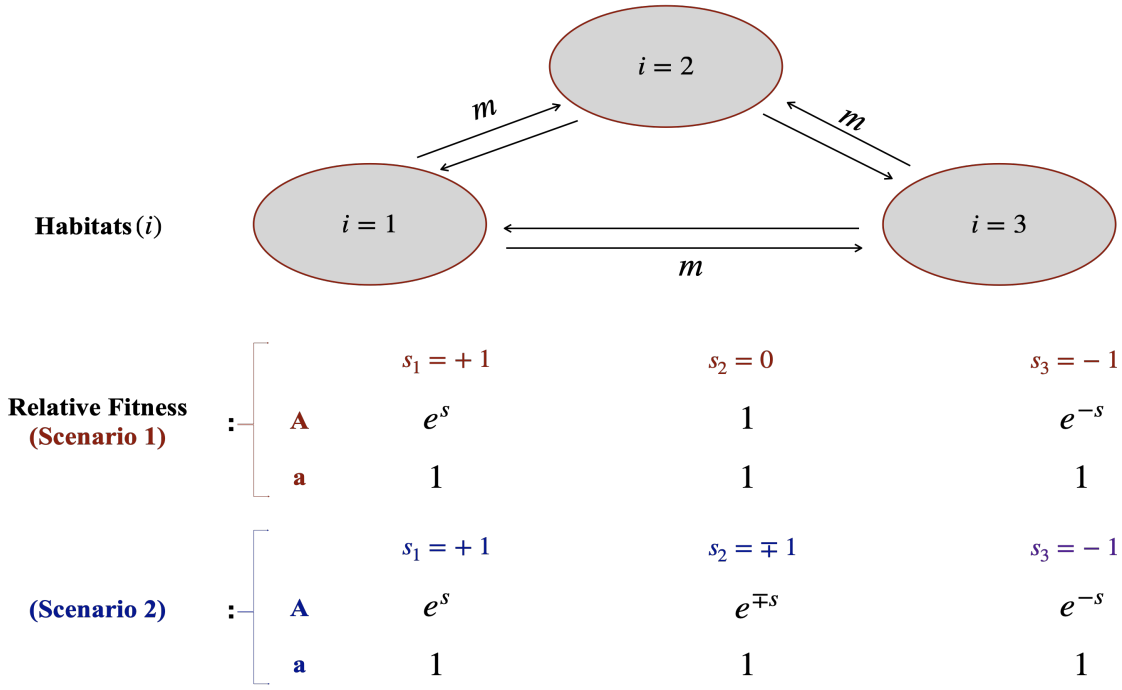


Figure 3.1: Antagonistic environmental effect for the allele A i.e., the strength and direction of selection for the A allele changes across the metapopulation as we go from one extreme habitat to the other.

Next, we consider a second scenario where $s_{1,j} = s_{3,j}$ (in magnitude) but $\{s_{2,1}, \cdots, s_{2,j}\} = \{-1, -1, \cdots, +1, +1\}$ i.e., where the A allele is selectively favored at half the loci in habitat 2 and maladaptive at the remaining half. This is done to check whether there exists an advantage to having an actual intermediate habitat. The assumption of symmetry (i.e., $s_{1,j} = s_{3,j}$) is later on relaxed.

Within this framework, we adopt a model of soft selection where the population size in each deme of the metapopulation is equal and constant over time (we denote this as N). To simplify the model further, we assume linkage equilibrium (LE), disregarding any nonrandom associations or correlations between loci. This assumption is valid when evolutionary processes occur slowly relative to the rate of recombination. Under these assumptions of constant size and LE, allele frequencies at each locus can be assumed to evolve independently and the problem reduces to that of a single locus, where we need only know the average allele frequency across all demes at a given locus. Hence, going forward, we can drop the locus index j and $s_{i,j}$ reduces to s_i.

Diffusion approximation (single locus dynamics - soft selection)

Population genetics relies extensively on the diffusion approximation which establishes a framework for understanding the distribution of allele frequency across a range of models which are equivalent if $s, 1/N \ll 1$.

In continuous-time and assuming linkage equilibrium, the rate of change in frequency of the A allele due to selection, migration and drift at any haploid locus, and in any deme in

habitat i can be written as,

$$\frac{dp_i}{dt} = p_i q_i \frac{\partial \log(\overline{W}_i)}{\partial p_i} + m(\overline{p} - p_i) + \sigma_{p_i} \tag{3.1}$$

where the direction of evolution as determined by the slope of $\overline{W}_i$ can be obtained using fig. 3.1. σ_{p_i} symbolizes the effect of drift and is a real-valued stochastic process with zero mean and covariance $\left\langle \sigma_{p_i}(t)\sigma_{p_i}(t^*) \right\rangle = \frac{p_i(t)q_i(t)\delta(t-t^*)}{N}$. Finally, $\overline{p}$ is the frequency of the A allele averaged across all demes of the metapopulation, is equal to the average in the migrant pool since migration is uniform, and depends both on the mean frequency of the A allele in the different habitats as well as the proportion of demes in these habitats i.e.,

$$\overline{p} \equiv \mathbb{E}[p] = \alpha_1 \overline{p}_1 + \alpha_2 \overline{p}_2 + \alpha_3 \overline{p}_3 \quad \text{with,} \quad \alpha_1 + \alpha_2 + \alpha_3 = 1 \tag{3.2}$$

According to Wright (1937) and Kimura (1955), the stationary distribution of allele frequency $\psi(p)$ at any locus and in any i can be written as,

$$\psi_i[p] = \frac{1}{C_i} p_i^{2Nm\overline{p}-1}(1-p_i)^{2Nm\overline{q}-1}\overline{W}_i^{2N}$$

where C_i is the normalization constant and $\overline{q} = 1 - \overline{p}$. Using the above, the stationary distribution in the three habitats (i.e., $\psi_1[p], \psi_2[p]$ and $\psi_3[p]$) can be obtained by substituting in i and the corresponding $\overline{W}_i$ (using the values in fig. 3.1) and this will depend on the parameters Nm and Ns_i. One can then numerically integrate over these distributions to obtain the expectations, $\overline{p}_1, \overline{p}_2$ and $\overline{p}_3$, which can now be substituted into eq. (3.2) to get the mean frequency of the A allele in the metapopulation (using a self-consistent iterative process).

Fixed-state approximation (limit of low migration - $Nm << 1$)

The fixed state approximation is a simplification which assumes gene flow is limited (i.e., $Nm \ll 1$) among the different habitats (see Barton and Olusanya, 2022) so that any deme is "nearly fixed" for one or other allele with stochastic transitions between them. Under soft selection, this allows us to characterize the genetic state of habitat i by the rate of transition (i.e., the rate at which one allele replaces the other in the population) from a to A (or A to a). The transition rate from a to A (A to a) is simply the product of the fixation probability of the A (a) allele times the number of new A (a) alleles entering the population (see equation 2.2 of Barton and Olusanya (2022) where $1 \equiv A$ and $0 \equiv a$ here). Using this approximation, one can then estimate the equilibrium expectation of A (a) in i by the equilibrium proportion of demes fixed for A (a).

As in Barton and Olusanya (2022), suppose we represent the proportion of demes fixed for the A allele in habitat i by P_i and that fixed for the a allele by Q_i then, focusing on

the A allele, the evolution of P_{i} in any i can be expressed as,

$$\partial_t P_{i} = \frac{2s_{i}Nm}{1-e^{-2Ns_{i}}}\left(\overline{p}Q_{i} - \overline{q}e^{-2Ns_{i}}P_{i}\right), \quad i = 1,2,3 \tag{3.3}$$

$$\overline{p} = \sum_{i=1}^{3}\alpha_{i}P_{i}, \quad 1 = \sum_{i=1}^{3}\alpha_{i} \tag{3.4}$$

This will be used later on to analyze the conditions for a polymorphism.

3.3 Results

We begin by considering the scenario where the A allele is favored in habitat 1 (i.e., $\{s_{1,1},\cdots,s_{1,\mathrm{j}}\} = \{+1,\cdots,+1\}$), selectively neutral in habitat 2 (i.e., $\{s_{2,1},\cdots,s_{2,\mathrm{j}}\} = \{0,\cdots,0\}$) and maladaptive in habitat 3 ($\{s_{3,1},\cdots,s_{3,\mathrm{j}}\} = \{-1,\cdots,-1\}$). Since alleles evolve independently, we drop the j subscript and focus on dynamics at a single locus taking $s_1 = 1$, $s_2 = 0$ and $s_3 = -1$ respectively. Our interest is in determining the conditions that favour the persistence of a polymorphism under such a scenario. In particular, we derive critical selection and migration thresholds that allows for such persistence. But first, let us explore the role of gene flow.

3.3.1 Role of gene flow

Figure 3.2 shows how the expected frequency of the A allele in habitats $1,2$ and 3 respectively (i.e., $\mathbb{E}_1[p]$, $\mathbb{E}_2[p]$ and $\mathbb{E}_3[p]$) depend on the average allele frequency, $\overline{p}$, in the migrant pool given different levels of gene flow, Nm. The results are obtained numerically using the diffusion approximation.

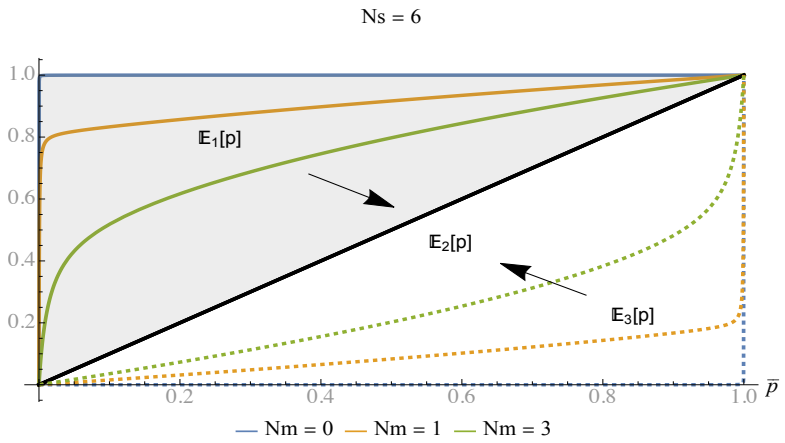


Figure 3.2: Dependence of the expected frequencies of the A allele in the three habitats on $\overline{p}$ for different levels of gene flow (Nm) and with fixed Ns. Black arrows pointing towards $\mathbb{E}_2[p]$ indicate that gene flow pushes the allele frequencies in the extreme habitats (i.e., in $i = 1$ and $i = 3$) towards $\mathbb{E}_2[p]$ which coincides with the average, $\overline{p}$ in the migrant pool. Results are obtained numerically using the diffusion approximation.

In the absence of gene flow (i.e., $Nm = 0$), the extreme habitats are fully well adapted to their local environmental conditions with the frequency of the A allele being 1 in $i = 1$ (solid blue line) and 0 in $i = 3$ (dashed blue line; so that the a allele has frequency 1 here). Increasing gene flow however introduces maladaptation, reducing the frequency of the favored allele in both habitats and pushing them towards the average in the migrant pool.

3.3.2 Deterministic equilibria.

We obtain the equilibria[1] for the system numerically by plotting $\overline{p} - \mathbb{E}[p]$ against $\overline{p}$ using eq. (3.2).

We find (fig. 3.3a) that a stable polymorphism ($0 \leq \overline{p} \leq 1$) always exists for any combination of $\alpha_1, \alpha_2, \alpha_3$ provided that selection is strong relative to migration (i.e., $Ns \gtrsim Nm$, blue line) as can also be seen in fig. 3.3b (black dots). This means that if Ns is less than a given critical threshold value, which we will denote by Ns_{cr} (in this example $Ns_{\text{cr}} = 3.1$, fig. 3.3b), polymorphism will be lost and one of the alleles (in this case the A allele) would fix throughout the metapopulation (see lhs of fig. 3.3b).

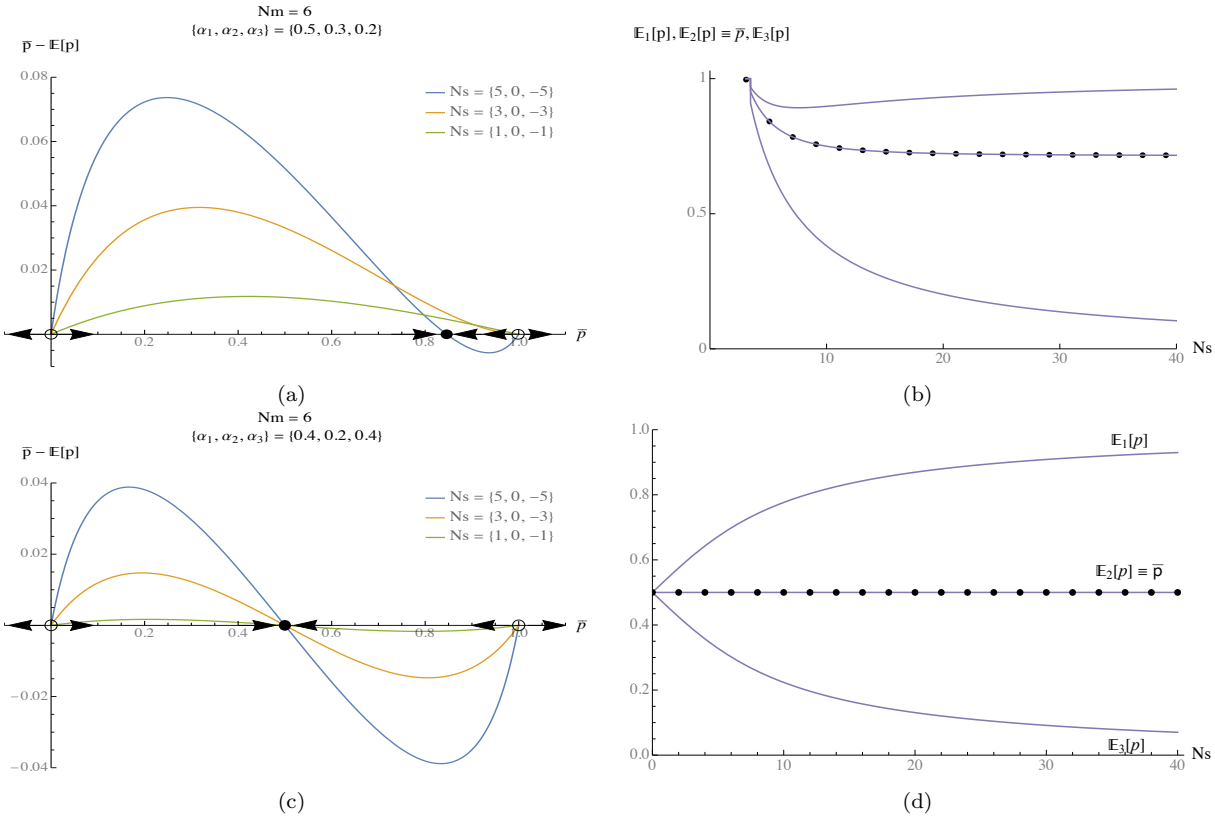


Figure 3.3: (a.) The three possible equilibria $0 \leq \overline{p} \leq 1$ for the A allele for a given combination $\{\alpha_1, \alpha_2, \alpha_3\}$. (b.) The stable polymorphism $\overline{p}$ only exists past a critical threshold, $Ns_{cr} = 3.1$ (c.)-(d.) Trivial case where the stable polymorphism always exists at $\overline{p} = 0.5$ independent of Nm and Ns. Black dots in figs. (b.) and (d.) represent the average allele frequency $\overline{p}$ across the metapopulation. Results are obtained numerically using the diffusion approximation.

[1] Equilibria are points where the curves intersect the $x - axis$ (see figs. 3.3a and 3.3c).

A trivial case occurs when the two extreme habitats (α_1 and α_3) are precisely balanced (i.e., are equally common or rare), then there would always exist a stable polymorphism at $\overline{p} = 0.5$ independent of Nm and for all values of Ns (as can also be seen in fig. 3.3d).

In this study, we are interested in cases where $0 \leq \overline{p} \neq 0.5 \leq 1$ as this provides a basis for exploring critical thresholds. In particular, we consider situations where $\alpha_1 > \alpha_3$ independent of α_2 (i.e., the habitat where the A allele is favored has a larger proportion of demes compared to the habitat where it is disfavored).

3.3.3 Maintenance of polymorphism and critical thresholds

Here, we consider how the persistence of a polymorphism is influenced by factors such as migration, selection, demic proportion and drift as well as provide an analytical handle on thresholds for persistence.

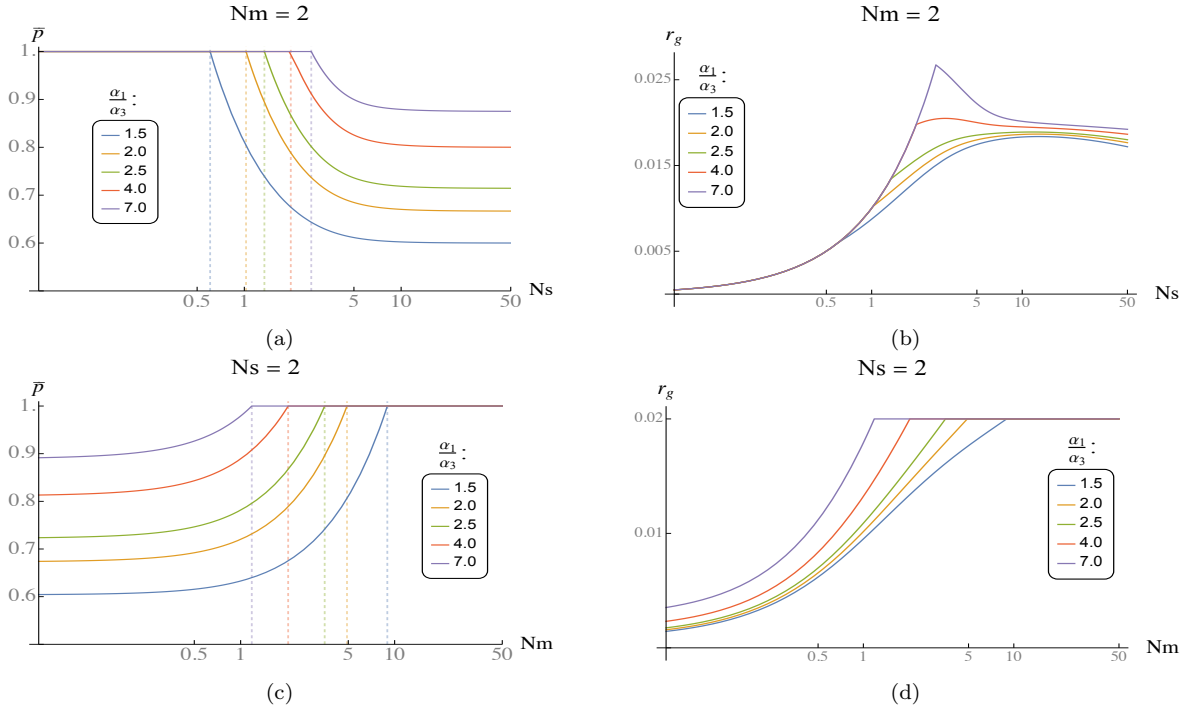


Figure 3.4: Dependence of the equilibrium average allele frequency on Ns (a.) and Nm (c.). Dashed lines in (a.) represent the critical selection thresholds, Ns_{cr}, above which a polymorphisn is possible and those in (c.) represent the critical migration threshold, Nm_{cr} below which a polymorphism is possible. (b.) and (d.) show the mean load in the metapopulation and how it depends on Ns and Nm respectively. The $x - axis$ is plotted on a log scale to better visualise behaviour at longer ranges. Results are obtained numerically using the diffusion approximation.

Figure 3.4a shows that when selection is weak (or comparable) relative to migration, allele A eventually fixes across the metapopulation (hence an initial increase in load fig. 3.4b) and this load decreases as the intensity of selection increases. Figure 3.4c further shows that limited migration favors the maintainance of polymorphism as the migration load in the population is lower in this case (lhs. of fig. 3.4d) due to the lower homogenizing (and

hence deleterious) effect of migration. With increasing Nm however, the A allele begins to invade and fixes across the metapopulation with the load converging to the maximum possible ($\sim sp$). Interestingly, all that matters here is the ratio of the proportion of demes where the A allele is favored to that where it is maladaptive i.e., α_1/α_3, and not the actual proportions α_1 and α_3. We call this ratio β (i.e., $\beta = \alpha_1/\alpha_3$). The lower the value of β, the longer the polymorphism persists, despite increasing Nm, and thus, the higher the critical migration threshold below which polymorphism is possible. This makes sense as the potential for swamping increases the more dissimilar the extreme habitats are.

We now investigate whether or not there is an advantage to having a truly intermediate habitat[2]. Does this for instance make it easier to maintain polymorphism? In other words, we consider a scenario where $\{s_{2,1}, \cdots, s_{2,j}\} = \{-1, -1, \cdots, +1, +1\}$. However since loci are decoupled under our model, we can simply follow the dynamics at a single locus conditioned on $\overline{p}$ at that locus. Hence, considering the dynamics at locus 1 and dropping the j index, we have $s_1 = +1$, $s_1 = -1$ and $s_3 = -1$.

From fig. B.1a (and B.1b) in appendix B.0.1, we notice that the critical threshold, Ns_{cr} (and Nm_{cr}) as well as the dynamics past this threshold, i.e., $Ns > Ns_{cr}$ (and $Nm < Nm_{cr}$) are exactly the same as can be seen in fig. 3.4a (and 3.4c) provided that β is the same in both scenarios (compare blue and red curves in fig. B.1a and fig. 3.4a as well as in B.1b and 3.4c). In fig. 3.4a (blue curve), the actual demic combination used to obtain the plot was $\{\alpha_1, \alpha_2, \alpha_3\} = \{0.3, 0.5, 0.2\}$ meaning that the A allele was favored in 30% of demes, selectively neutral in 50% of demes and disfavored in 20% of demes so that $\beta = 1.5$. However, in fig. B.1a, the combination used was $\{\alpha_1, \alpha_2, \alpha_3\} = \{0.6, 0.2, 0.2\}$ meaning that the A allele was favored in 60% of demes and disfavored in 40% of demes so that β is again 1.5. In essence, we see that what really matters for the overall dynamics is not the individual proportion of demes but the value of β. So for a single locus, independent of whether selection is neutral or disadvantageous in the intermediate habitat, we will obtain similar dynamics overall provided that β is the same.

In fact, for a single locus, this second scenario considered above, i.e., with $s_1 = 1$, $s_2 = -1$ and $s_3 = -1$ reduces to the two habitat case considered in Szép et al. (2021) where we only need to know the proportion of demes where the allele is favored i.e., α_1 (since the proportion where it is disfavored can be obtained simply as $1 - \alpha_1$).

The trivial behaviour observed above can be attributed to the assumptions of our model. With hard selection (not assumed here), where alleles co-evolve with each other coupled via N, we would expect to see a non-trivial dynamics.

Next, we direct our attention towards obtaining analytical formulae for critical thresholds

[2]truly intermediate meaning that half of the loci in this habitat favors the A allele, whereas, it is disfavored in the remaining half.

(focusing strictly on the case $s_1 = 1$, $s_2 = 0$, and $s_3 = -1$). To do this, we employ the fixed state approximation introduced in section 3.2. First, we focus on the critical selection threshold, Ns_{cr}, above which a polymorphism is possible. This can be split into threshold values when gene flow among the different habitats is limited ($Nm \ll 1$) and when it is abundant ($Nm \gg 1$). In the limit of low migration where allele frequency distribution are bimodal with loci nearly fixed for the A or a allele, the equilibrium mean allele frequency $\overline{p}$ across the metapopulation can be obtained by first setting the lhs of eq. (3.3) to 0 and solving for P_1, P_2 and P_3 respectively. These can now be substituted into eq. (3.4) to obtain $\overline{p}$ as,

$$\overline{p} \;=\; \frac{\alpha_{_1}\, e^{2Ns} - \alpha_{_3}}{(e^{2Ns} - 1)\,(\alpha_{_1} + \alpha_{_3})}, \quad \alpha_{_1} + \alpha_{_2} + \alpha_{_3} = 1 \tag{3.5}$$

Equation (3.5) can then be solved for Ns yielding $Ns = \frac{1}{2} log\left[\frac{\alpha_1}{\alpha_3}\right]$. So, in the limit of low Nm, we require a selection strength above $((1/2N) log(\alpha_1/\alpha_3))$ to maintain a polymorphism. To find the threshold value for larger values of Nm, we use a deterministic analysis (see also soft selection analysis in Szép et al. (2021)). Just above the critical selection threshold, $\overline{p}_1$ will be close to 1, $\overline{p}_2 = \overline{p}$ and the difference between $\overline{p}_1$ and $\overline{p}_3$ will be very slight (see fig. 3.3b for example). Thus, we can set $\overline{p}_1 = 1 - \epsilon_1$, $\overline{p}_1 - \overline{p}_3 = \epsilon_2$ and $\overline{p}_2 = \frac{\alpha_1(1-\epsilon_1)+\alpha_3(1-\epsilon_1-\epsilon_2)}{1-\alpha_2}$. Substituting these into eq. (3.1), replacing p_i with $\overline{p}_i$ (where $i = 1, 2, 3$) and setting $\sigma_{p_i=0}$ (since we're dealing with a deterministic analysis), we obtain differential equations, $\frac{d\overline{p}_1}{dt}$, $\frac{d\overline{p}_2}{dt}$ and $\frac{d\overline{p}_3}{dt}$ respectively. Consequently, solving $\frac{d\overline{p}_i}{dt} = 0$ for $\overline{p}_i$ (retaining only lower order terms) gives the deterministic threshold as $s_{cr} = \frac{1-\alpha_2-2\alpha_3}{1-\alpha_2} m \;\equiv \frac{\alpha_1-\alpha_3}{\alpha_1+\alpha_3}\, m$. Hence, we have,

$$Ns_{cr} \approx \left(\frac{\frac{\alpha_1}{\alpha_3} - 1}{\frac{\alpha_1}{\alpha_3} + 1}\right) Nm + \frac{1}{2} log\left[\frac{\alpha_1}{\alpha_3}\right] \equiv \left(\frac{\beta - 1}{\beta + 1}\right) Nm + \frac{1}{2} log(\beta) \tag{3.6}$$

In a similar fashion, using eq. (3.6), we obtain an analytical expression for the critical migration threshold, Nm_{cr} as,

$$Nm_{\text{cr}} \approx \left(\frac{\beta + 1}{\beta - 1}\right) \left(Ns - \frac{1}{2} log(\beta)\right). \tag{3.7}$$

Equations (3.6) and (3.7) are approximate equations obtained by adding together the $Nm \to 0$ (i.e., fixed state) threshold and the deterministic threshold.

Interestingly, both equations (i.e., eq. (3.6) and (3.7)) depend only on β $(i.e., \alpha_1/\alpha_3)$ and not on α_2 implying that in this case, having an intermediate habitat makes no difference to critical selection and migration thresholds.

A comparison of these two equations (i.e., eq. (3.6) and (3.7)) with numerically obtained values from the diffusion approximation (see fig. B.4b in appendix B.0.4) shows quite a good fit between our derived formulae and the numerical expectation. In particular,

as the rate of gene flow (Nm) increases (fig. B.4a), we observe a corresponding rise in Ns_{cr} (the critical selection threshold) due to heightened migration load within the population resulting from increased gene flow. Consequently, a greater selection strength is necessary to counteract this effect. Ns_{cr} is also higher with higher $\beta := \alpha_1/\alpha_3$ allowing for polymorphism in a restricted range of selection intensity when the two extreme habitats are more dissimilar in proportion. Furthermore, with increasing Ns, (fig. B.4b) we observe an increase in the critical migration threshold, Nm_{cr} as stronger selection reduces the tolerance for the coexistence of both alleles (the A and a allele). Consequently, higher levels of migration are needed to counteract the loss of genetic diversity and maintain polymorphism within the metapopulation. This effect is amplified when the proportion of demes favoring the A allele is lower than the proportion disfavoring it as in this case, the A allele faces a greater challenge in spreading and persisting across the metapopulation because it encounters more demes where it is at a disadvantage.

So far, we have established that in the symmetric case ($s_1 = s_3 = s$), the intermediate habitat (i.e., where $s_2 = 0$) makes no difference to critical migration and selection thresholds (see also (3.6) and (3.7)). To check whether this habitat matter in any other way (i.e., past critical thresholds), we compare the dynamics of $\overline{p}$ past the threshold values i.e., at $Nm < Nm_{cr}$ ($Ns > Ns_{cr}$) for two metapopulations with equal β and different α_2. In particular, for the two metapopulations, we use the demic proportions $\{0.2, 0.7, 0.1\}$ and $\{0.4, 0.4, 0.2\}$ respectively. Our results, fig. B.2a (fig. B.3), show similar dynamics (divergence) for $\overline{p}$ for $Nm < Nm_{cr}$ ($Ns > Ns_{cr}$) for both metapopulations suggesting the independence of these results on α_2. We furthermore check if this conclusion holds true with asymmetric selection (i.e., with $s_1 \neq s_3$) and find that even under the assumption of asymmetry, α_2 has no influence on critical migration thresholds for polymophism or on the divergence of $\overline{p}$ at $Nm < Nm_{cr}$ (see fig. B.2b).

Finally, we quantify the effect of drift (finite deme sizes) on the maintenance of polymorphism and how this depends on habitat proportions (fig. 3.5a and 3.5b) while also relaxing the assumption of symmetric selection. To do this, we consider the s_1, s_3 region within which a polymorphism persists and explore its dependence on different deme sizes N (starting from larger N to lower N). Although what really matters for polymorphism are the scaled parameters Ns_1, Ns_3 and Nm, plotting this way allows us to easily draw comparison and identify whether for a given s_1, s_3 value, polymorphism is better maintatined in a metapopulation with larger or smaller deme sizes.

We see from fig. 3.5a and 3.5b that increasing drift (i.e., increasing $1/N$) reduces the overall genetic diversity in the metapopulation and constrains the region for which polymorphism is possible. This is because with increasing drift, certain alleles become more or less common purely by chance, leading to a decrease in the overall number of the two alleles segregating in the population. This reduction in genetic diversity consequently limits the

potential for polymorphism. We also observe that this behaviour is more pronounced with higher β (i.e., with $\alpha_1 \gg \alpha_3$).

We further explore the effect of drift on the critical migration threshold. We see from fig. 3.5c that genetic drift substantially reduces the critical migration threshold required for a polymorphism so that relatively low levels of gene flow can have a large impact in preventing genetic differentiation resulting from drift.

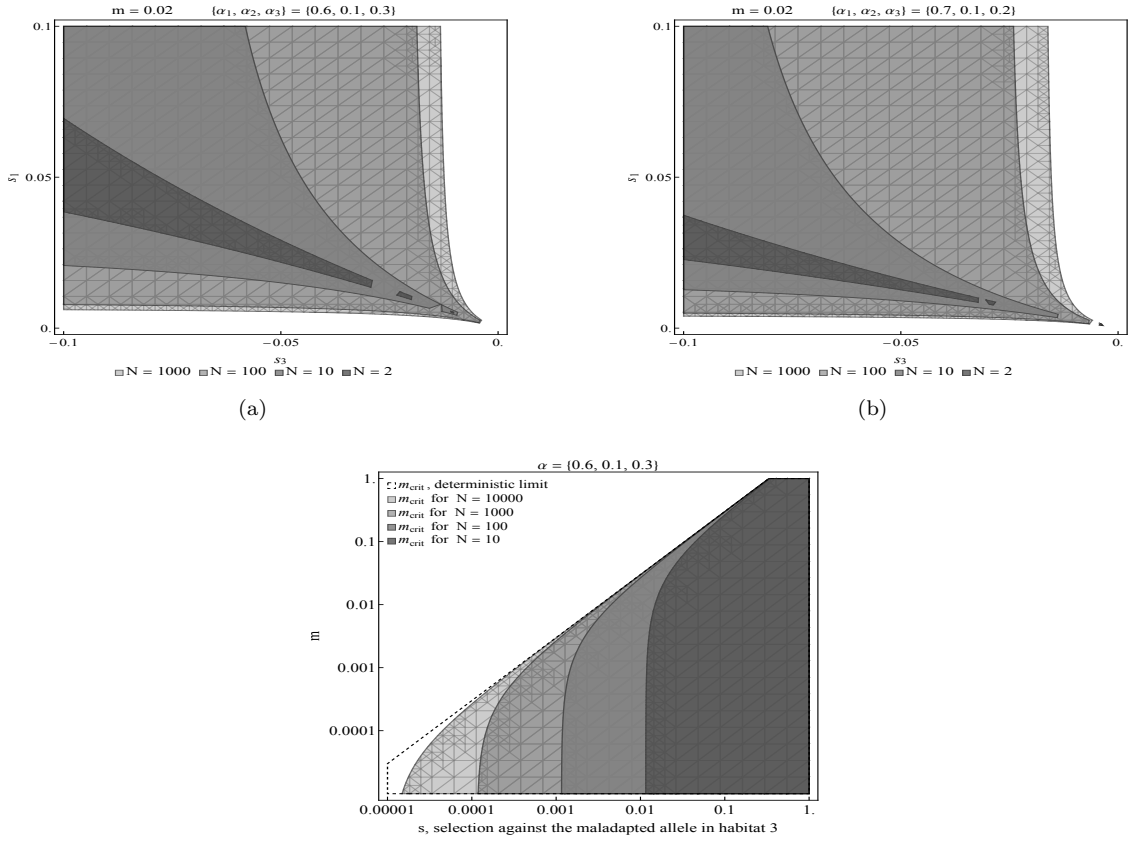


Figure 3.5: (a.)-(b.) Effect of drift on s_1/s_3 region for which polymorphism is possible (s_1 and s_3 are the strength of selection for and against the A allele in habitats 1 and 3 respectively). We have also relaxed the assumption of symmetric selection here (c.) Effect of drift on the critical migration threshold below which polymorphism is possible. (a).-(b.) are obtained from the diffusion approximation. (c.) is obtained using eq. (3.7).

3.4 Discussion

The preservation of genetic ploymorphisms within metapopulations has been a topic of significant interest in population genetics. Our work builds upon previous research on local adaptation in a metapopulation involving two habitats, extending the analysis to a three-habitat scenario. Our findings offer insights into the interplay between selection, migration, drift and habitat proportions in maintaining genetic diversity within metapopulations.

One key finding of our study is the notable increase in the range of polymorphism

with limited migration between demes. This is consistent with previous research (albeit involving two habitats) where reduced gene flow is found to enhance local adaptation and promote genetic diversity within metapopulations. Constrained gene flow fosters the persistence of polymorphism by preventing the rapid homogenization of alleles across habitats. This reinforces the fact that in nature, factors such as dispersal barriers or other mechanisms that hinder gene flow could be vital in preserving genetic diversity within metapopulations. For a given Nm, we also find that the range of polymorphism is further increased if the habitat where the allele is well adapted and maladaptive respectively are in roughly equal proportions. This suggests that when selective pressures in these habitats are evenly matched, this creates an environment where polymorphism can thrive. This may arise from a dynamic equilibrium in which opposing selection strengths sustain a stable polymorphic state, thereby preventing the complete fixation of one allele over the other.

Our study also highlights the crucial role of selection relative to migration ($Ns \gtrsim Nm$) in driving the persistence of polymorphism. Strong selection counteracts the homogenizing effect of gene flow ensuring the coexistence of both alleles in the metapopulation. Conversely, genetic drift constrains the region within which polymorphism is possible.

Under our (single locus) model of soft selection, we found no clear advantage for the maintenance of polymorphism when there is an intermediate habitat where alleles are selectively neutral and this holds true away from critical thresholds (i.e., for $Ns > Ns_{cr}$ and $Nm < Nm_{cr}$) and with asymmetric selection (i.e., with $s_{1,j} \neq s_{3,j}$). Instead, our findings emphasize that what really matters for the persistence of polymorphism is the relative balance of favorable and maladaptive habitats. Specifically, the ratio of the proportion of demes where the allele is favored to where it is maladaptive (i.e., β) emerges as an important parameter driving the dynamics of polymorphism. This result however holds under the infinite island model assumed in this work. With finite islands, genetic variation (and hence polymorpgism) could be lost by chance (see chapter 2) and a low amount of mutation may therefore be required to maintain variation in the long term.

Finally, we derived analytical formulas for the critical selection and migration thresholds for polymorphism in the three-habitat case. These formulas can provide useful insights into the parameter ranges where a polymorphism is possible.

While this study represents a step forward in understanding local adaptation and the maintenance of polymorphism in metapopulations with more than two habitats, several avenues for future research remain open. Investigating the dynamics of adaptation under a model of hard selection (where we account for changes in population size via allele frequency changes and vice versa i.e., explicit eco-evo feedback) are promising directions for further exploration. In this case having an intermediate habitat where the A allele

is at a selective advantage at half the loci and at a disadvantage at the remaining half (or any other complex architecture) could produce interesting dynamics and reveal novel insights into local adaptation and the maintenance of polymorphism.

Secondly, relaxing the assumption of linkage equilibrium and exploring the role of non-random associations or interference between loci could provide a more realistic depiction of genetic interactions within metapopulations. In this case, the net effect of other loci on any selected locus can be captured using the effective migration approximation[3] (Sachdeva, 2022)).

Another compelling avenue for future exploration would be to consider explicit spatial structure. Incorporating explicit spatial configuration or arrangement of populations into our framework may enable a more precise investigation into the role of the parameter β for the maintenance of variation and how this is influenced by habitat connectivity. It may also provide a more nuanced understanding of how gene flow, genetic drift and spatial heterogeneity influences the stability and maintenance of polymorphism, offering a more realistic perspective on the mechanisms that shape the genetic composition of natural populations.

Finally, exploring the impact of various ecological factors such as variation in carrying capacities across the different habitats could offer a richer understanding of the interplay between genetic and ecological dynamics in shaping and sustaining genetic diversity within fragmented landscapes.

3.5 Funding

Recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Institute of Science and Technology, Austria. The research was also funded by the Austrian Science Fund (FWF) [FWF P-32896B].

[3]The effective migration approximation accounts for how multilocus heterosis influence allele frequencies by assuming that it essentially alters the effective migration rate of deleterious alleles, so that the allele frequency distribution at any given locus can still be obtained using the diffusion approximation with the the rate of migration, m replaced by an effective migration rate, m_e.

References

Barton, N. and Olusanya, O. (2022). The response of a metapopulation to a changing environment. *Philosophical Transactions of the Royal Society B*, 377(1848):20210009. https://doi.org/10.1098/rstb.2021.0009.

Barton, N. H. and Whitlock, M. C. (1997). The evolution of metapopulations. In *Metapopulation biology*, pages 183–210. Elsevier. https://doi.org/10.1016/B978-012323445-2/50012-2.

Blanquart, F., Gandon, S., and Nuismer, S. (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of evolutionary biology*, 25(7):1351–1363. https://doi.org/10.1111/j.1420-9101.2012.02524.x.

Bolnick, D. I. and Otto, S. P. (2013). The magnitude of local adaptation under genotype-dependent dispersal. *Ecology and evolution*, 3(14):4722–4735. https://doi.org/10.1002/ece3.850.

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

Gavrilets, S. (2003). Perspective: models of speciation: what have we learned in 40 years? *Evolution*, 57(10):2197–2215. https://doi.org/10.1111/j.0014-3820.2003.tb00233.x.

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. https://doi.org/10.1098/rspb.1956.0039.

Hoekstra, R. F., Bijlsma, R., and Dolman, A. (1985). Polymorphism from environmental heterogeneity: models are only robust if the heterozygote is close in fitness to the

favoured homozygote in each environment. *Genetics Research*, 45(3):299–314. https://doi.org/10.1017/S001667230002228X.

Holt, R. D. and Gomulkiewicz, R. (1997). How does immigration influence local adaptation? a reexamination of a familiar paradigm. *The American Naturalist*, 149(3):563–572. https://doi.org/10.1086/286005.

Kimura, M. (1955). Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences*, 41(3):144–150. https://doi.org/10.1073/pnas.41.3.144.

LaBar, T. and Adami, C. (2017). Evolution of drift robustness in small populations. *Nature Communications*, 8(1):1012. https://doi.org/10.1038/s41467-017-01003-7.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in ecology & evolution*, 17(4):183–189. https://doi.org/10.1016/S0169-5347(02)02497-7.

Rogers, M. A. (1991). Evolutionary differentiation within the northern great basin pocket gopher, thomomys townsendii. i. morphological variation. *The Great Basin Naturalist*, pages 109–126. https://www.jstor.org/stable/41712630.

Sachdeva, H. (2022). Reproductive isolation via polygenic local adaptation in sub-divided populations: Effect of linkage disequilibria and drift. *PLoS Genetics*, 18(9):e1010297. https://doi.org/10.1371/journal.pgen.1010297.

Sachdeva, H., Olusanya, O., and Barton, N. (2022). Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity. *Philosophical Transactions of the Royal Society B*, 377(1846):20210010. https://doi.org/10.1098/rstb.2021.0010.

Szép, E., Sachdeva, H., and Barton, N. H. (2021). Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5):1030–1045. https://doi.org/10.1111/evo.14210.

Urban, M. C. (2015). Accelerating extinction risk from climate change. *Science*, 348(6234):571–573. https://doi.org/10.1126/science.aaa4984.

Walters, R. J. and Berger, D. (2019). Implications of existing local (mal) adaptations for ecological forecasting under environmental change. *Evolutionary Applications*, 12(7):1487–1502. https://doi.org/10.1111/eva.12840.

Whitlock, M. C. (2000). Fixation of new alleles and the extinction of small populations: drift load, beneficial alleles, and sexual selection. *Evolution*, 54(6):1855–1861. https://doi.org/10.1111/j.0014-3820.2000.tb01232.x.

Williams, G. C. (2018). *Adaptation and natural selection: A critique of some current evolutionary thought*, volume 61. Princeton university press.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97. https://doi.org/10.1093/genetics/16.2.97.

Wright, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences*, 23(6):307–320. https://doi.org/10.1073/pnas.23.6.307.

Chapter 4

Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity[1]

Himani Sachdeva, Oluwafunmilola Olusanya, Nick Barton

Abstract

We analyse how migration from a large mainland influences genetic load and population numbers on an island, in a scenario where fitness-affecting variants are unconditionally deleterious, and where numbers decline with increasing load. Our analysis shows that migration can have qualitatively different effects, depending on the total mutation target and fitness effects of deleterious variants. In particular, we find that populations exhibit a genetic Allee effect across a wide range of parameter combinations, when variants are partially recessive, cycling between low-load (large-population) and high-load (sink) states. Increased migration reduces load in the sink state (by increasing heterozygosity) but further inflates load in the large population state (by hindering purging). We identify various critical parameter thresholds at which one or other stable state collapses, and discuss how these thresholds are influenced by the genetic versus demographic effects of migration. Our analysis is based on a 'semi-deterministic' analysis, which accounts for genetic drift but neglects demographic stochasticity. We also compare against simulations which account for both demographic stochasticity and drift. Our results clarify the importance of gene flow as a key determinant of extinction risk in peripheral populations, even in the absence of ecological gradients.

[1]This work has been published at https://royalsocietypublishing.org/doi/10.1098/rstb.2021.0010

Keywords: deleterious variants, genetic load, extinction, migration, demographic stochasticity, semi-deterministic approximation

4.1 Introduction

Most outcrossing populations carry a substantial masked mutation load owing to recessive variants, which can contribute significantly to inbreeding depression in peripheral isolates or after a bottleneck. The extent to which the increased segregation (or fixation) of deleterious mutations due to drift (drift load) exacerbates extinction risk in isolated populations has been a subject of long-standing interest (Lande, 1994; Lynch et al., 1995; Frankham, 2005; O'Grady et al., 2006). Theory predicts that moderately deleterious mutations contribute the most to genetic load and extinction in small populations (Lynch et al., 1995; Higgins and Lynch, 2001); however, the prevalence of such deleterious variants of mild or moderate effect and their dominance values remain poorly characterised, except for a few model organisms (Agrawal and Whitlock, 2011; Huber et al., 2018).

The relative risks posed by mutation accumulation and demographic stochasticity to a population depend crucially on its size, with some theory suggesting that these may be comparable for populations in their thousands (Lande, 1994). Additionally, environmental stochasticity — catastrophic events, as well as fluctuations in growth rates and carrying capacities, may dramatically lower extinction times (Lande, 1993). Both demographic and environmental fluctuations, in turn, reduce the effective size of a population, making it more prone to fix deleterious alleles; the consequent reduction in fitness further depresses size, pushing populations into an 'extinction vortex', which is often characterised by a complex interaction between the effects of genetic drift, demographic stochasticity and environmental fluctuations (Lande, 1988).

Peripheral populations at the edges of species' ranges receive dispersal from the core to an extent which varies over space and time. Moreover, ranges may be fragmented owing to habitat loss and individual sub-populations connected to each other via low and possibly declining levels of migration. Under what conditions are such extinction vortices arrested by migration, and what are the genetic and demographic underpinnings of this effect, when it occurs?

Migration boosts numbers, mitigating extinction risk owing to demographic and environmental stochasticity, or at the very least, allows populations to regenerate after chance extinction. The demographic consequences of migration are especially important in fragmented populations with many small patches (Ovaskainen and Hanski, 2003): above a critical level of migration, the population may survive as a whole over long timescales even if individual patches frequently go extinct (Gyllenberg and Hanski, 1992; Lande et al., 1998).

Migration also influences extinction risk by shifting the frequencies of fitness-affecting variants: the resultant changes in fitness may decrease or increase population size, thus further boosting or depressing the relative contribution of migration to allele frequency changes within a population, setting in motion a positive feedback which may culminate in extinction (when gene flow is largely maladaptive; e.g., (Ronce and Kirkpatrick, 2001; Szép et al., 2021)) or evolutionary rescue (e.g., if gene flow supplies variation necessary for adaptation to local conditions, or reduces inbreeding load; e.g., (Higgins and Lynch, 2001; Bell and Gonzalez, 2011; Uecker et al., 2014)).

The maladaptive consequences of migration have largely been explored for extended populations under spatially varying environments: here, gene flow typically hinders local adaptation, especially at range limits, leading to 'swamping' and extinction (Bridle and Vines, 2007). However, the consequences of gene flow for fitness, and consequently survival, are not always intuitive when the fitness effects of genetic variants are uniformly deleterious (or beneficial) across populations. For example, while gene flow may alleviate inbreeding load by preventing the fixation of deleterious alleles in small populations, it may also render selection against recessive mutations less effective by increasing heterozygosity. A striking consequence is that under a range of conditions, the fitness of metapopulations is maximised at intermediate levels of migration (Whitlock, 2002) and more generally, at intermediate levels of population structure (Roze, 2015).

A key consideration is whether or not gene flow is symmetric, i.e., whether some sub-populations are merely influenced by the inflow of genes from the rest of the habitat or if all sub-populations influence the genetic composition of the population as a whole (Lenormand, 2002). Asymmetric dispersal is common at the geographic peripheries of species' ranges or on islands. Moreover, populations occupying small patches within a larger metapopulation with a wide distribution of patch sizes, or sub-populations with lower-than-average fitness (and consequently, atypically low numbers) may also experience predominantly asymmetric inflow of genes. Asymmetric gene flow allows for allele frequency differences across the range of a population even in the absence of environmental heterogeneity, e.g., when population sizes (and hence the efficacy of selection relative to drift) vary across the habitat. This, in turn, may generate heterosis or outbreeding depression across *multiple* loci, when individuals from different regions hybridize.

From a conceptual viewpoint, the consequences of asymmetric gene flow are typically simpler to analyse as we can focus on a single population, while taking the state of the rest of the large habitat as 'fixed'. Such analyses are key to understanding more general scenarios where genotype frequencies and population sizes across different regions co-evolve.

Here, we analyse the eco-evolutionary dynamics of a single island subject to migration from a larger mainland population in a scenario with uniform selection across the two populations, i.e. where fitness is affected by a large number of variants that are unconditionally deleterious. We ask: under what conditions can migration from the mainland alleviate inbreeding load, thus preventing 'mutational meltdown' and extinction of the island population? Further, how are the effects of migration mediated by the genetic architecture of load, i.e., by the genome-wide mutation target and fitness effects of deleterious variants? A key focus is to understand the coupled evolution of allele frequencies (across multiple loci) and population size: to this end, we consider an explicit model of population growth with logistic regulation, where growth is reduced by an amount equal to the genetic load.

While the effects of maladaptive gene flow on marginal populations have been studied under various models (Kawecki, 2008; Barton and Etheridge, 2018), there has been little work (under genetically realistic assumptions) on the (possibly) beneficial effects of migration on inbreeding load and survival. In particular, modelling the polygenic nature of fitness variation is crucial, as changes in load (e.g. owing to migration) at any locus can affect all other loci by effecting changes in population size, which in turn influences the efficacy of selection across the genome.

4.2 Model and Methods

Consider a peripheral island population subject to one-way migration from a large mainland. Individuals are diploid and carry L biallelic loci that undergo bidirectional mutation between the wild-type and deleterious state at rate u per generation per individual per locus in either direction. Mutations have the same fitness effects on the mainland and island, i.e. there is no environment-dependent fitness component.

Deleterious variants at different loci affect fitness multiplicatively (no epistasis): individual fitness is given by $e^{-\sum_{j=1}^{L} s_j(X_j+h_jY_j)}$, where X_j (Y_j) equals 1 if the j^{th} locus is homozygous (heterozygous) for the deleterious allele, and is zero otherwise. Here, s_j is the homozygous selective effect and h_j the dominance coefficient for the deleterious allele at locus j. We assume $0 \leq h_j \leq 1/2$, so that deleterious alleles are always (partially) recessive. In individual-based simulation, we use the form $(1-hs)^n(1-s)^m 1^{m'}$ (which is equivalent to the above fitness function for small s) where n represents the number of heterozygous loci, m represents the number of loci homozygous for the deleterious allele and m' represents the number of homozygous loci for the wild-type allele with $n+m+m'=L$.

In each generation, a Poisson-distributed number of individuals (with mean m_0) migrate from the mainland to the island. For simplicity, we assume that the mainland population is large enough that deleterious allele frequencies among migrants (denoted by $\{p_j^{(m)}\}$)

are close to the deterministic predictions for a single locus under mutation-selection equilibrium.

We assume density-independent selection on individuals on the island and density-dependent population growth with logistic regulation, where the baseline growth rate is reduced by the genetic load: the population size n_t in any generation t is then Poisson-distributed with mean equal to $n_{t-1}\exp[r_0(1-n_{t-1}/K)]\overline{W}$, where n_{t-1} is the population size in the previous generation, r_0 the baseline growth rate, K the carrying capacity, and $\overline{W}$ the mean genetic fitness. The n_t individuals in the t^{th} generation are formed by randomly sampling $2n_t$ parents (with replacement) from the n_{t-1} individuals in the previous generation with probabilities equal to their relative fitnesses, followed by free recombination between parental haplotypes to create gametes. Diploid offspring are then formed by randomly pairing gametes.

Simulations. We carry out two kinds of simulations: individual-based simulations that explicitly track multi-locus genotypes of all individuals on the island, and simulations that assume LE (linkage equilibrium) and neglect inbreeding. The latter kind of simulations are computationally less intensive as they only track allele frequencies at the L loci and the size of the population. However, they make two simplifying assumptions: first, that genotypes at any locus are in Hardy–Weinberg proportions (i.e. no inbreeding); second, that any statistical associations, e.g., between the allelic states of different loci (linkage disequilibria or LD) or between the probability of identity by descent, and consequently homozygosity, at different loci (identity disequilibria or ID) — are negligible. Then, individual genotypes are simply random assortments of deleterious and wildtype alleles, and can be generated, e.g., in a simulation, by independently assigning alternative allelic states to different loci with probabilities equal to the allele frequencies. Details of the two kinds of simulations are provided in the supplementary material, Appendix A.

Because selection pressures are identical on the mainland and island, any systematic differences in allele frequencies or homozygosity between the two populations across multiple loci (which would generate LD and ID respectively) must arise solely due to differences in population size, which would cause the efficacy of selection to be different on the mainland and island. In general, we expect LD and ID to be negligible when all ecological and evolutionary processes are slower than recombination (Szép et al., 2021): this may not hold, however, when populations are small (and drift significant), making it necessary to evaluate how statistical associations between deleterious variants affect extinction thresholds. In the rest of the paper, we will only show results of the allele frequency simulations (assuming LE and zero inbreeding); we compare these with individual-based simulations and discuss how LD and ID affect population outcomes in the supplementary material, Appendix B.

Joint evolution of allele frequencies and population size. Assuming LE and no inbreeding, the evolution of the island population is fully specified by how allele frequencies $\{p_j\}$ and the population size n co-evolve in time t. If all evolutionary and ecological processes (except recombination) are slow, we can describe this co-evolution in continuous time. We rescale all rates by the baseline growth rate r_0 and population sizes by the carrying capacity K. This yields the following dimensionless parameters: $\tau = r_0 t$, $S = s/r_0$, $M_0 = m_0/(r_0 K)$, $U = u/r_0$, $N_t = n_t/K$ and an additional parameter $\zeta = r_0 K$ which governs the strength of demographic fluctuations. The joint evolution of allele frequencies $\{p_j\}$ and the scaled population size N in continuous time is described by the following equations:

$$\begin{aligned} \frac{dp_j}{d\tau} &= -\frac{p_j q_j}{2}\frac{\partial R_g}{\partial p_j} + U(q_j - p_j) + \frac{M_0}{N}(p_j^{(m)} - p_j) + \lambda_{p_j} \\ \frac{dN}{d\tau} &= N\left[1 - N - R_g\right] + M_0 + \lambda_N \qquad\qquad \text{where} \quad R_g = \sum_{j=1}^{L} S_j p_j (2h_j q_j + p_j) \end{aligned} \tag{4.1}$$

The random processes λ_N and λ_{p_j} satisfy $\mathbb{E}[\lambda_{p_j}] = \mathbb{E}[\lambda_n] = 0$, $\mathbb{E}[\lambda_N(t)\lambda_N(t')] = \frac{N(t)}{\zeta}\delta(t-t')$ and $\mathbb{E}[\lambda_{p_j}(t)\lambda_{p_j}(t')] = \frac{1}{\zeta}\frac{p_j(t)q_j(y)}{N(t)}\delta(t-t')$.

The four terms in the first equation correspond (in order of appearance) to changes in allele frequency due to selection, mutation, migration and drift. Note that the strength of migration is inversely proportional to population size N, reflecting the stronger (relative) effect of migration on the genetic composition of smaller, as opposed to larger, island populations. The second equation describes the evolution of population size N: the first term describes changes in N under logistic growth, where the growth rate is reduced by a factor proportional to the log mean fitness (i.e., the genetic load); the second term captures the effect of migration; the third term corresponds to demographic fluctuations (whose variance is proportional to N, the size of the population). These equations captures a key feature of polygenic eco-evolutionary dynamics — namely, that the evolution of allele frequencies at different loci is *coupled* via their dependence on a common N, which in turn is influenced by the degree of maladaptation at all loci via R_g. Thus, allele frequencies do *not* evolve independently, even though allelic states at different loci are statistically independent at any instant (under LE).

For *fixed* N and under LE and IE, the joint distribution of allele frequencies at mutation-selection-migration-drift equilibrium is a product of the single-locus distributions. This was given by Wright (1937), and for a given locus j is (in terms of scaled parameters):

$$\psi(p_j|N) \;\propto\; p_j^{4\zeta NU + 4\zeta M_0 p_j^{(m)} - 1}(1-p_j)^{4\zeta NU + 4\zeta M_0(1-p_j^{(m)}) - 1} e^{-2\zeta N S p_j[p_j + 2h(1-p_j)]} \tag{4.2}$$

Integrating over this distribution yields the expected allele frequency $\mathbb{E}(p_j|N)$ and the expected heterozygosity $\mathbb{E}(2p_jq_j|N)$ at any locus, and thence the expected total load $\mathbb{E}(R_g|N) = \sum_j S_j[\mathbb{E}(p_j|N) - (1/2 - h)\mathbb{E}(2p_jq_j|N))]$ (scaled by the baseline growth rate r_0) for fixed N.

However, in reality, N is not fixed and will fluctuate — both due to random fluctuations in fitness (due to the underlying stochastic fluctuations of $\{p_j\}$ at equilibrium) as well as in the reproductive output of individuals (demographic stochasticity). Thus, N itself follows a distribution. While we can write down an equation for the stochastic co-evolution of N and $\{p_j\}$, no explicit solution for the joint equilibrium distribution is possible unless mutation rates are strictly zero (as assumed by Szép et al. (2021)). Thus, we must employ various approximations to describe the coupled dynamics of N and $\{p_j\}$.

Approximate semi-deterministic analysis. We expect the population size distribution to be sharply peaked around one or more values $\{N_*\}$ if demographic fluctuations are weak (i.e., for high values of $\zeta = r_0K$) and if fluctuations in mean fitness are also weak (i.e., for large L). Such sharply-peaked distributions correspond to populations that transition only rarely between alternative peaks, i.e., where the alternative peaks $\{N_*\}$ of the size distribution represent 'metastable' states, allowing allele frequencies sufficient time to equilibrate at any N_*. Then the genetic load would be close to the expected value $\mathbb{E}(R_g|N_*)$ under mutation-selection-drift-migration balance (*given* N_*) when populations are in one of the alternative metastable states, though not necessarily while they transition between states. In order to determine $\{N_*\}$, we postulate that these must represent stable equilibria of the population size dynamics, neglecting demographic stochasticity and assuming that mean fitness (given N_*) is close to the expectation under mutation-selection-drift-migration balance for that N_*. Then we have:

$$0 = N_*\left(1 - N_* - \mathbb{E}[R_g|N_*]\right) + M_0 \qquad 1 - 2N_* - \mathbb{E}[R_g|N_*] - N_*\left.\frac{d\mathbb{E}[R_g]}{dN}\right|_{N=N_*} < 0 \quad (4.3)$$

The equality follows from eq. (4.1), by setting the third (noise) term to zero and assuming that R_g is close to its *equilibrium* expectation $\mathbb{E}(R_g|N_*)$ given N_*. The second inequality is the condition for N_* to be a stable equilibrium: this means that populations starting at an arbitrary N in the vicinity of N_* would evolve towards this equilibrium size. Equation (4.3) can be solved numerically to obtain the equilibria $\{N_*\}$, which, under the above assumptions, will be close to the peaks of the population size distribution. As we see below, depending on the parameter regime, there may be one or two stable equilibria of eq. (4.3), corresponding to population size distributions that are unimodal or bimodal. Bifurcations (i.e., critical parameter thresholds where one of the two stable equilibria vanishes) thus correspond to qualitative transitions in the state of the population.

We will refer to the analysis above as a 'semi-deterministic analysis' as it accounts for the stochastic effects of genetic drift on allele frequencies and load (via the expectation $\mathbb{E}[R_g|N_*]$) but neglects demographic stochasticity. In general, we expect the semi-deterministic analysis to become more accurate for large enough L (so that fluctuations in the load about the deterministic expectation can be neglected) and for larger $\zeta = r_0 K$ (see also Appendix C.2), since increasing ζ, which implies higher number of births per generation, results in weaker demographic fluctuations (Szép et al., 2021).

4.3 Results

4.3.1 Metastable populations and extinction times in the absence of migration

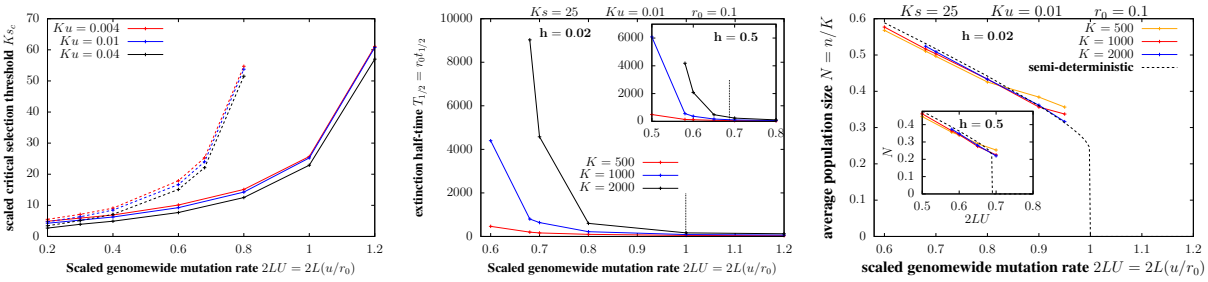
We first consider peripheral populations in the absence of migration. Such populations necessarily become extinct in the long run: however, depending on the mutation target for deleterious mutations (relative to the baseline growth rate r_0), the fitness effects of mutations and the carrying capacity of the island, extinction times may be very long and populations metastable.

We can use eq. (4.3) to gain intuition for the conditions for metastability under zero migration. Setting $M_0 = 0$, it follows that there is always an equilibrium at $N = 0$ (corresponding to extinction): this equilibrium is stable for $L(S/2) > 1$ where $S = s/r_0$, and unstable otherwise (see appendix C.2 for details). There may exist another equilibrium at $N_* = 1 - \mathbb{E}[R_g|N_*, M_0 = 0]$; the population size N_* is positive (i.e., the population is not extinct) only if $\mathbb{E}[R_g|N_*, M_0 = 0] < 1$, i.e., the equilibrium genetic load is lower than the baseline growth rate.

Because the equilibrium load $\mathbb{E}[R_g|N_*, M_0 = 0]$ depends on four independent parameters (which we can choose as $\zeta S = Ks$, $\zeta U = Ku$, h and $2LU = 2L(u/r_0)$), mapping the conditions for metastability boils down to asking: in the absence of demographic fluctuations, where in this four-dimensional parameter space, can we find populations with a non-zero equilibrium size or sufficiently low load (fig. 4.1a)? In reality, there is a fifth parameter $\zeta = r_0 K$, which governs demographic stochasticity: thus, all other parameters being equal, extinction times will be longer when demographic stochasticity is weaker, i.e., $r_0 K$ larger (fig. 4.1b).

For simplicity, we assume that deleterious alleles anywhere on the genome have the same selective effects and dominance coefficients; this assumption is relaxed in Appendix C.4. We use Ks_c to denote the critical selection strength per homozygous deleterious allele (scaled by the carrying capacity K), such that a non-zero equilibrium N_* exists for $Ks > Ks_c$ but not for $Ks < Ks_c$. Figure 4.1a shows Ks_c as a function of $2LU = 2L(u/r_0)$, the

total mutation rate relative to the baseline growth rate, for different Ku (different colors) for nearly recessive ($h = 0.02$; solid lines) and additive ($h = 0.5$; dashed lines) alleles.



(a) (b) (c)

Figure 4.1: Population outcomes with zero migration. (A) Scaled critical selection threshold Ks_c, above which populations are metastable, as a function of the scaled mutation target $2LU = 2L(u/r_0)$ for different values of Ku (different colors) for nearly recessive ($h = 0.02$; solid lines) and additive ($h = 0.5$; dashed lines) alleles. A non-zero equilibrium population size $N_* = 1 - \mathbb{E}[R_g|N_*]$ exists for $Ks > Ks_c$ but not for $Ks < Ks_c$. This selection threshold is calculated using eq. (4.3) by neglecting demographic stochasticity, and thus strictly provides a criterion for stable populations in the limit $\zeta = r_0 K \to \infty$. (B) The scaled extinction half-time $T_{1/2} = r_0 t_{1/2}$ (see text for definition) as a function of $2LU$ for various K (different colors) for nearly recessive ($h = 0.02$; main plot) and additive alleles (inset). (C) The average scaled population size $N = n/K$ of metastable populations versus $2LU$ for various K (different colours) along with the semi-deterministic prediction $N_* = 1 - \mathbb{E}[R_g|N_*]$ (dashed black line) for nearly recessive ($h = 0.02$; main plot) and additive alleles (inset). In both (B) and (C), the carrying capacity is increased while proportionately decreasing s, u and increasing L, such that $Ks = 25$, $Ku = 0.01$ and $2LU = 2L(u/r_0)$ remain unchanged; increasing K thus has the sole effect of weakening demographic stochasticity. Extinction times and the average population sizes in the metastable state are computed from allele frequency simulations (under LE and IE) of 1000 replicates with $r_0 = 0.1$.

For populations to be metastable, the total genetic load must be less than the baseline growth rate. The total load scales with the mutation target L; the load per locus is approximately $2u$ for strongly deleterious variants ($u/hs \ll 1$ and $Ks \gg 1$). For smaller Ks, drift will typically inflate load above this deterministic expectation when deleterious alleles are additive ($h \sim 1/2$) but may also reduce load (via more efficient purging) when alleles are recessive ($h \sim 0$). This is reflected in figure 4.1a: the threshold Ks_c required to maintain metastable populations is higher for additive deleterious alleles (dashed lines) than recessive alleles (solid lines). Moreover, because genetic load in this drift-dominated regime depends only weakly on the mutational input Ku, the threshold Ks_c is largely independent of Ku (different colors). Finally, for all parameter combinations, the threshold Ks_c increases as the mutation target becomes larger: this simply reflects the fact that for the total load to be less than the baseline growth rate, the load per locus must be lower (requiring stronger selection) if deleterious variants segregate at a greater number of loci. Accordingly, for very large mutation targets $2LU = 2L(u/r_0) \gtrsim 1$, the total load will exceed r_0 (and populations will fail), irrespective of the strength of selection against deleterious mutations.

The critical selection thresholds for metastability shown in fig. 4.1a are computed by neglecting demographic stochasticity, i.e., by assuming $\zeta = r_0 K$ to be very large. However,

for moderate ζ, stochastic fluctuations in reproductive output from generation to generation may accelerate extinction: this effect can be especially significant in smaller populations as these tend to fix more deleterious alleles, which further reduces fitness and size, thus rendering populations even more vulnerable to stochastic extinction.

To investigate how demographic stochasticity contributes to extinction, we simulate populations residing on islands with different carrying capacities K, but characterized by the same values of Ks, Ku, $2LU = 2L(u/r_0)$ and h. Mathematically, this involves taking the limit $s \to 0$, $u \to 0$, $K \to \infty$, $L \to \infty$, while holding Ks, Ku, $2LU$ constant: then, increasing K has the sole effect of weakening demographic stochasticity (by increasing $r_0 K$). Populations are initially perfectly fit, but accumulate deleterious variants over time, eventually becoming extinct owing to the combined effects of genetic load and demographic stochasticity. Figure 4.1b shows the extinction 'half-time' $T_{1/2} = r_0 t_{1/2}$ (scaled by the baseline growth rate r_0) — the time by which precisely half of all 1000 simulation replicates are extinct, as a function of $2LU$ for various K for nearly recessive (main plot) and additive alleles (inset). All results are from allele frequency simulations (assuming LE and zero inbreeding). Dashed vertical lines indicate the threshold $2LU$ above which metastable populations (with $N_* = 1 - E[R_g|N_*] > 0$) cannot exist (even for large $r_0 K$).

We find that extinction times increase with increasing K for all parameters. However, for parameter combinations that correspond to extinction in the large $r_0 K$ limit (i.e., to the right of the dashed lines), this increase is approximately linear in K, while for parameters leading to metastability in the large $r_0 K$ limit (left of dashed lines), this increase is faster than linear. In fact, in the metastable regime, even with a carrying capacity K of a few thousand individuals, demographic stochasticity will be sufficiently weak and extinction times large enough that isolated populations can persist over geological timescales: for these parameter regimes, it is environmental fluctuations (which our model ignores), rather than mutation load or demographic stochasticity, that will be the main determinant of extinction risk.

We also compute the average (scaled) population size $N = n/K$ in the metastable state (4.1c). This declines with increasing $2LU$, i.e., as we approach the threshold for loss of metastability, and is close to the semi-deterministic prediction (dashed black lines) for all values of K.

4.3.2 Effect of migration on equilibrium population sizes and genetic load

We now consider how migration from a large mainland influences population dynamics on the island, for different genetic architectures of load, i.e., given certain selective effects

and dominance coefficients of deleterious alleles. As before, we assume that all deleterious alleles have equal selective effects and dominance values, relegating the discussion of more general scenarios, where alleles with different fitness effects segregate, to Appendix C.4.

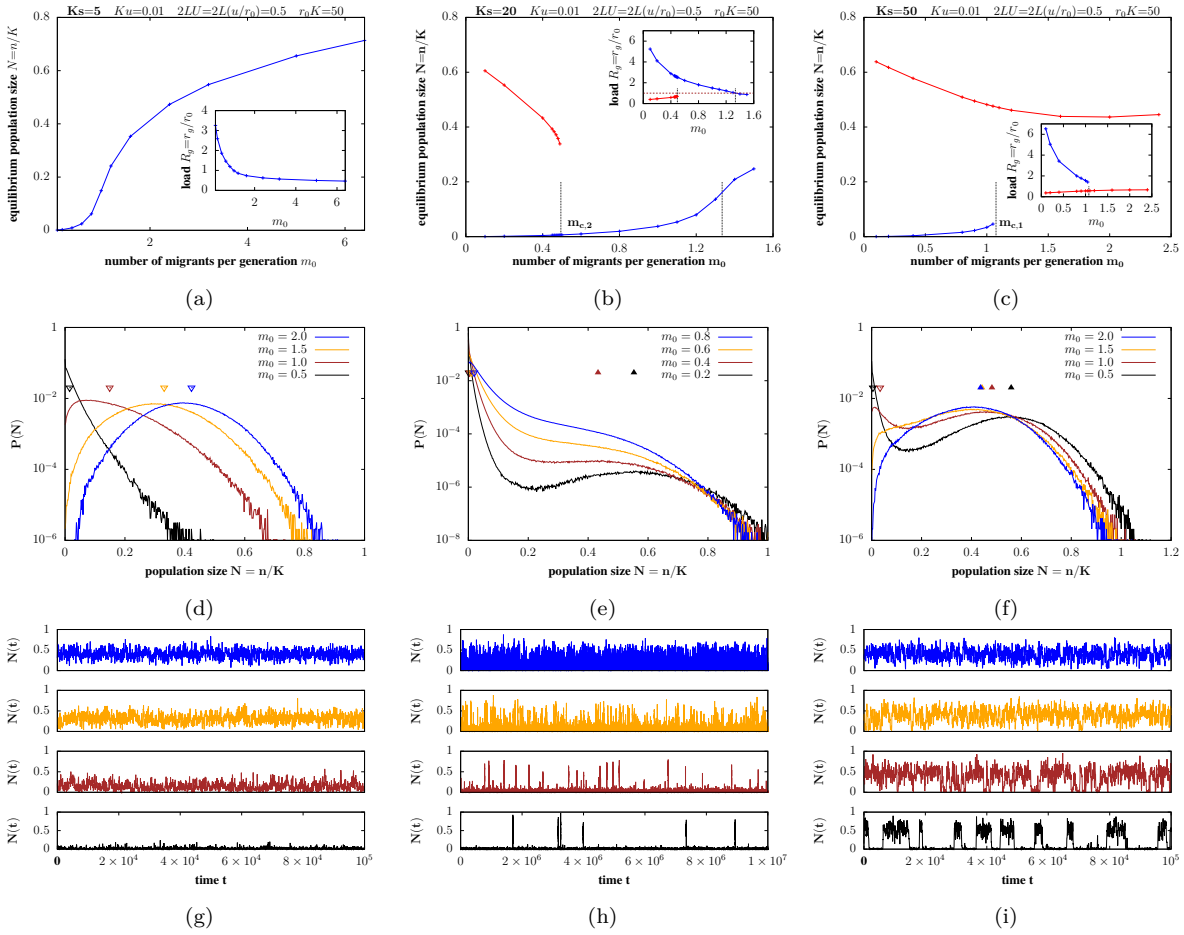


Figure 4.2: Effect of migration on equilibrium population sizes and load for weakly deleterious ($Ks < Ks_c$; left column), moderately deleterious ($Ks \gtrsim Ks_c$; middle) and strongly deleterious ($Ks \gg Ks_c$; right) alleles. (A)-(C) Population size (main plots) and genetic load (inset), corresponding to one or more stable equilibria, vs. m_0, the number of migrants per generation. Equilibria are obtained by numerically solving eq. (4.3) (semi-deterministic predictions). Blue lines represent the sink equilibrium (with an associated population size that tends to zero as $m_0 \to 0$); red lines represent the large-population equilibrium (with non-zero population size in the $m_0 \to 0$ limit). The thresholds $m_{c,2}$ (2B) and $m_{c,1}$ (2C) represent critical migration thresholds at which the large-population or the sink equilibrium vanishes. The threshold $m_{c,3}$ (2B) is the critical migration level at which the (scaled) load associated with the sink state becomes less than 1. (D)-(F) Equilibrium probability distributions of the scaled population size $N = n/K$ for various values of m_0, as obtained from simulations (under LE and IE) in the three parameter regimes. The filled and empty triangles indicate the population sizes corresponding to alternative equilibria as predicted by the semi-deterministic analysis. (G)-(I) Time series $N(t)$ vs. t over an arbitrary period after equilibration in the three regimes. All plots show results for: $Ku = 0.01$, $2LU = 2L(u/r_0) = 0.5$, $\zeta = r_0 K = 50$, **h=0.02** and $Ks = 5$, 20, 50 for the left, middle and right columns respectively (the critical threshold is $Ks_c \sim 7.65$). In addition, for the simulations (plots D-I), we use $r_0 = 0.1$.

We first identify critical parameter thresholds associated with qualitative changes in population outcomes for the two extremes: nearly recessive ($h = 0.02$ in fig. 4.2) and additive ($h = 0.5$) alleles. We then consider how these thresholds depend on the level of dominance of deleterious alleles (fig. 4.3).

Figure 4.2 illustrates the effect of migration on population sizes and load on the island for three parameter combinations, corresponding to weakly deleterious ($Ks < Ks_c$; left column), moderately deleterious ($Ks \gtrsim Ks_c$; middle) and strongly deleterious ($Ks \gg Ks_c$; right) nearly recessive alleles ($h = 0.02$). Recall that Ks_c is the selection threshold (obtained by neglecting demographic stochasticity) such that populations rapidly go extinct in the absence of migration for $Ks < Ks_c$, but can be metastable (even without migration) for $Ks > Ks_c$ and $r_0 K \to \infty$ (fig. 4.1).

Figure 4.2a-4.2c show population sizes corresponding to stable equilibria of eq. (4.3) versus m_0, the number of migrants per generation, for $Ks < Ks_c$, $Ks \gtrsim Ks_c$ and $Ks \gg Ks_c$. The insets show the equilibrium load versus m_0. Since these equilibria are calculated by numerically solving eq. (4.3), they represent *semi-deterministic* predictions, i.e., they neglect demographic stochasticity but account for the effects of drift, assuming that populations spend enough time in any metastable state that the distribution of allele frequencies can equilibrate. Figure 4.2d-4.2f shows the stochastic distribution of the scaled population size $N = n/K$, as obtained from allele frequency simulations (assuming LE and zero inbreeding) after the population has equilibrated, for various m_0, in the three parameter regimes. Figures 4.2g-4.2i show the corresponding time series $N(t)$ over an arbitrary period, after equilibration, for a single randomly chosen stochastic realization.

Weakly deleterious recessive alleles. For $Ks < Ks_c$, there exists a single stable equilibrium at all migration levels (fig. 4.2a), with the corresponding population size approaching zero as migration becomes rarer. Accordingly, the stochastic distribution of N exhibits a single peak for all m_0 (fig. 4.2d). An increase in migration causes the genetic load to decrease (inset, fig. 4.2a) and the population size to increase. The stochastic distribution of population sizes is approximately exponentially distributed for low values of m_0 (corresponding to a sink state), but shifts towards higher values of N as m_0 increases, and is approximately normally distributed about the deterministic equilibrium N_* (indicated by empty triangles in fig. 4.2d) for large m_0.

Increasing migration has two effects in this case: it reduces genetic load, typically by reducing homozygosity— a *genetic* effect, but also increases population numbers— a *demographic* effect. The genetic effect of migration, i.e., the dependence of the expected load $\mathbb{E}(R_g|N) = \sum_j S_j[\mathbb{E}(p_j|N) - (1/2 - h)\mathbb{E}(2p_j q_j|N)]$ on m_0, can be further decomposed into effects on the expected frequencies $\mathbb{E}(p_j|N)$ and expected heterozygosity $\mathbb{E}(2p_j q_j|N)$ of deleterious alleles. For weakly deleterious recessive alleles, the expected frequency decreases with increasing migration for low m_0, is minimum at intermediate m_0, and then increases as m_0 increases further (not shown). However, genetic load still decreases monotonically with increasing migration (inset, fig. 4.2a), due to the much sharper increase in heterozygosity with migration.

Strongly deleterious recessive alleles. For $Ks \gg Ks_c$, there exist *two* stable equilibria of the semi-deterministic population size dynamics (eq. (4.3)) at low levels of migration: population sizes and values of load corresponding to these two equilibria are shown in blue and red in fig. 4.2c. One equilibrium (blue lines in fig. 4.2c) corresponds to a sink state with the associated population size approaching zero as $m_0 \to 0$. The other equilibrium (depicted in red) corresponds to a large population, which can persist at finite fraction of carrying capacity even as migration becomes exceedingly rare, i.e., has size N_*, which approaches $1 - \mathbb{E}[R_g|N_*, M_0 = 0] > 0$ as $m_0 \to 0$.

The existence of two stable equilibria at low m_0 translates into *bimodal* population size distributions (in black and brown in fig. 4.2f), with one peak close to extinction and the other at the semi-deterministic N_* (depicted by filled triangles). Population size $N(t)$ also exhibits characteristic dynamics (see time series in black and brown in fig. 4.2i): populations tend to remain 'stuck' in either the sink state or the large-population state for long periods of time, typically exhibiting rather small fluctuations about the characteristic sizes associated with either state, and only transitioning rarely between states.

The sink state and large-population state can also be thought of as "high-load" (i.e., genetic load greater than the baseline growth rate r_0) and "low-load" (genetic load less than r_0) states. Thus, in this parameter regime, populations exhibit a genetic Allee effect, wherein load is sufficiently low and net growth rates positive only above a threshold population size: therefore starting from an initially empty island, populations cannot grow deterministically (and persist instead as migration-fed sinks), even though large populations can maintain themselves, at least in the absence of demographic fluctuations. Transitions between the sink (high-load) and large-population (low-load) states are thus inherently stochastic, arising owing to demographic fluctuations (which are aided by higher levels of migration) rather than owing to a systematic drive towards the alternative state.

Changes in m_0 have qualitatively different effects on the two states: at low m_0, increasing migration reduces load and thus increases numbers in the sink state (blue curve), but has the *opposite* effect on the large-population equilibrium (red curves). This is due to the non-monotonic dependence of the frequency of deleterious recessive alleles on population size: increasing size causes drift to become weaker relative to selection but also reduces homozygosity, so that fewer deleterious alleles are exposed to selection; frequencies are thus minimum at intermediate population sizes, reflecting the tension between these two opposing effects. In small sink populations, migration from the continent reduces homozygosity as well as deleterious allele frequencies, thus reducing load. However, in larger populations, increasing migration reduces the homozygosity but raises the frequency of deleterious alleles (since deleterious alleles are purged less efficiently in very large mainland populations than in intermediate-sized or large island populations); the overall effect of migration is thus to increase load in the large-population state.

Above a critical migration threshold, which we denote by $m_{c,1}$, the sink equilibrium vanishes. Thus, for $m > m_{c,1}$, populations always have a positive growth rate and reach a finite fraction of carrying capacity, regardless of starting size. As before, this qualitative change in the (semi-deterministic) population size dynamics at a critical migration level has its analog in a qualitative change in the stochastic distribution of population sizes (fig. 4.2f): increasing migration causes the population size distribution to change from bimodal to unimodal (with a sole peak at the large-population equilibrium $N_* = 1 - \mathbb{E}[R_g|N_*])$). At higher migration rates, i.e., as we approach the critical migration threshold $m_{c,1}$ starting from low m_0, turnover between the sink and large-population states becomes more rapid—note the shorter intervals between transitions in the brown vs. black time series in fig. 4.2i.

Moderately deleterious recessive alleles. Consider now the case where selection is stronger than the threshold Ks_c but not too strong, i.e., $Ks \gtrsim Ks_c$ (middle column in fig. 4.2). As in the $Ks \gg Ks_c$ case, there exist two alternative stable equilibria at low migration levels, one corresponding to the sink state (blue lines in fig. 4.2b) and the other to the large-population state (red lines). As before, increasing migration alleviates inbreeding load and increases population size in the sink state but elevates load (by hindering purging) and decreases population size in the large-population state. Unlike in the $Ks \gg Ks_c$ state, the latter effect is much stronger: thus, above a critical migration threshold, which we denote by $m_{c,2}$, it is the large-population equilibrium that vanishes, so that there is only a single (sink) equilibrium for $m > m_{c,2}$.

The analogous change in the population size distribution with increasing migration (fig. 4.2e) is somewhat subtle: for the lowest value of m_0 (black), the distribution is clearly bimodal, with most of the weight close to $N = 0$ (sink state) and a very small peak at the large-population equilibrium. As migration increases, the distribution of sizes associated with the sink state widens, while the peak corresponding to the large-population equilibrium shifts towards lower N (e.g., brown plot). At high migration levels (orange and blue plots), there is no longer a distinct second peak and only the sink state persists. In this state, the genetic load exceeds r_0 *on average*; however, its distribution and the corresponding distribution of N is quite wide.

If migration increases further, then the average load associated with the sink state continues to fall, until at a third threshold (denoted by $m_{c,3}$, here ~ 1.33), the load again becomes lower than the baseline growth rate r_0 or, alternatively, scaled load $R_g = r_g/r_0$ less than 1 (depicted by a dashed line in fig. 4.2b). Thus, for $m_0 > m_{c,3}$, populations can grow, starting from small numbers, and reach a finite fraction of carrying capacity. However, unlike the large-population state that emerges at higher Ks, in this case, populations are highly dependent on migration and would rapidly collapse (owing to the fixation of

deleterious alleles) if cut off from the mainland population.

Population size dynamics (fig. 4.2h) are characterised, as in the $Ks \gg Ks_c$ case, by occasional transitions between the sink state and the large-population state for low values of m_0 (black), with transitions becoming more frequent with increasing m_0 (orange). Transition times are much longer than in the $Ks \gg Ks_c$ case (note the values on the x-axis); thus transitions are unlikely to occur over realistic timescales, and populations will typically be observed in the sink state. At higher migration levels, there are no obvious transitions, with population sizes and load fluctuating (with some skew) about a mean value.

Additive alleles. With additive effects ($h \sim 0.5$), any deleterious allele experiences the same selective disadvantage, irrespective of whether it appears in the heterozygous or homozygous state: thus, there is no purging in smaller populations (which have higher homozygosity) and allele frequencies decrease monotonically with increasing population size. As a result, migration from the larger mainland population always decreases load (by decreasing the deleterious allele frequency) and increases the size of the island population.

This implies that there are only two qualitatively distinct regimes: with weakly deleterious alleles (Ks less than the corresponding Ks_c; see fig. 4.1a), populations tend to extinction as migration declines. There is a single equilibrium of the semi-deterministic population size dynamics for all m_0 or analogously, a single peak of the stochastic population size distribution, which shifts towards higher sizes as m_0 increases.

For strongly deleterious alleles ($Ks > Ks_c$), population size N and load R_g are largely insensitive to migration, since populations can always grow, starting from small numbers, at least when demographic stochasticity is unimportant (i.e., for $r_0 K \gg 1$). With smaller $r_0 K$, population sizes and load depend weakly on migration since demographic stochasticity may depress N and inflate the effects of drift (even with $Ks > Ks_c$). Moreover, in this regime, populations are also prone to stochastic extinction for $m_0 < 1/2$ (Szép et al., 2021), such that the distribution of N is inherently bimodal. Thus, where demographic stochasticity is significant, a low level of migration may be needed for stable populations even with strong selection against additive alleles.

This analysis outlined here (based on eq. (4.3)) also applies to a distribution of effects across loci. In Appendix C.4, we consider examples where a fraction of deleterious mutations are additive and the remaining fraction recessive. We also compare the results of allele frequency simulations with those of individual-based simulations with unlinked loci (Appendix C.3). These show fairly close agreement, suggesting that LD and ID do not significantly affect allele frequency dynamics, at least for the typical parameter values considered here: this is also consistent with earlier work, which suggests only a

modest effect of disequilibria on background selection in sub-divided populations under soft selection (Roze, 2015).

4.3.3 Disentangling genetic vs. demographic effects of migration on recessive alleles.

In summary, given a mutation target $2LU \lesssim 1$ and assuming equal-effect loci, there is a critical selection threshold Ks_c such that large populations are metastable in $m_0 \to 0$ for $Ks > Ks_c$ but not for $Ks < Ks_c$. For $Ks > Ks_c$, there is a genetic Allee effect at low migration levels and with partially recessive alleles: populations cannot grow after recolonization of an initially empty island, and persist only as demographic sinks until a chance fluctuation increases numbers sufficiently that load can be purged and the alternative (large-population) equilibrium attained; such large populations can then be maintained over long periods of time (fig. 4.2i). For very low dominance values ($h \lesssim 0.15$), there is a second threshold $Ks_{c,2}$ with $Ks_{c,2} > Ks_c$ (see below), which separates parameter regimes characterised by qualitatively different effects of migration on population outcomes: for $Ks > Ks_{c,2}$, increasing migration destabilizes the sink state, so that only the large-population state persists above a migration threshold $m_{c,1}$, whereas for $Ks_c < Ks < Ks_{c,2}$, increasing migration destabilizes the large-population state, so that only the sink state persists above a threshold $m_{c,2}$. However, such migration-fed sink populations can be quite large: above a third threshold $m_{c,3}$, populations may even have sufficient heterozygosity to again attain a low-load ($r_g < r_0$) state.

To what extent can we attribute such qualitative changes in population outcomes to the genetic versus demographic effects of migration? As before, one approach is to compare critical thresholds for populations with the same scaled parameters Ks, Ku, $2LU = 2L(u/r_0)$ and h, while increasing the carrying capacity K (simultaneously increasing L and lowering u, s). Then as K increases, the demographic effects of migration (which depend on the dimensionless parameter $M_0 = m_0/(r_0 K)$; see eq. (4.1)) can be neglected, while its genetic effects on load (which depend on Ks, Ku and m_0) remain important.

Figure 4.3 shows these comparisons for two dominance values — $h = 0.02$ (figs. 4.3a, 4.3b) and $h = 0.1$ (figs. 4.3c, 4.3d). Note that we only consider predictions of the semi-deterministic analysis (eq. (4.3)), which assumes that population sizes are sharply clustered around the peaks of the distribution and transitions between peaks are infrequent. This assumption clearly breaks down close to transition thresholds (e.g., see figs. 4.2e, 4.2f); thus, thresholds observed in simulations may differ somewhat from semi-deterministic predictions (details in Appendix C.2). Moreover, the semi-deterministic analysis neglects all demographic stochasticity, and thus does not account for the fact that changes in K will affect not just the (systematic) demographic effect of migration but also the (stochastic)

effect of demographic fluctuations. Nevertheless, the semi-deterministic analysis is useful as it allows us to explore qualitative dependencies of critical thresholds on the underlying parameters without resorting to time-consuming simulations.

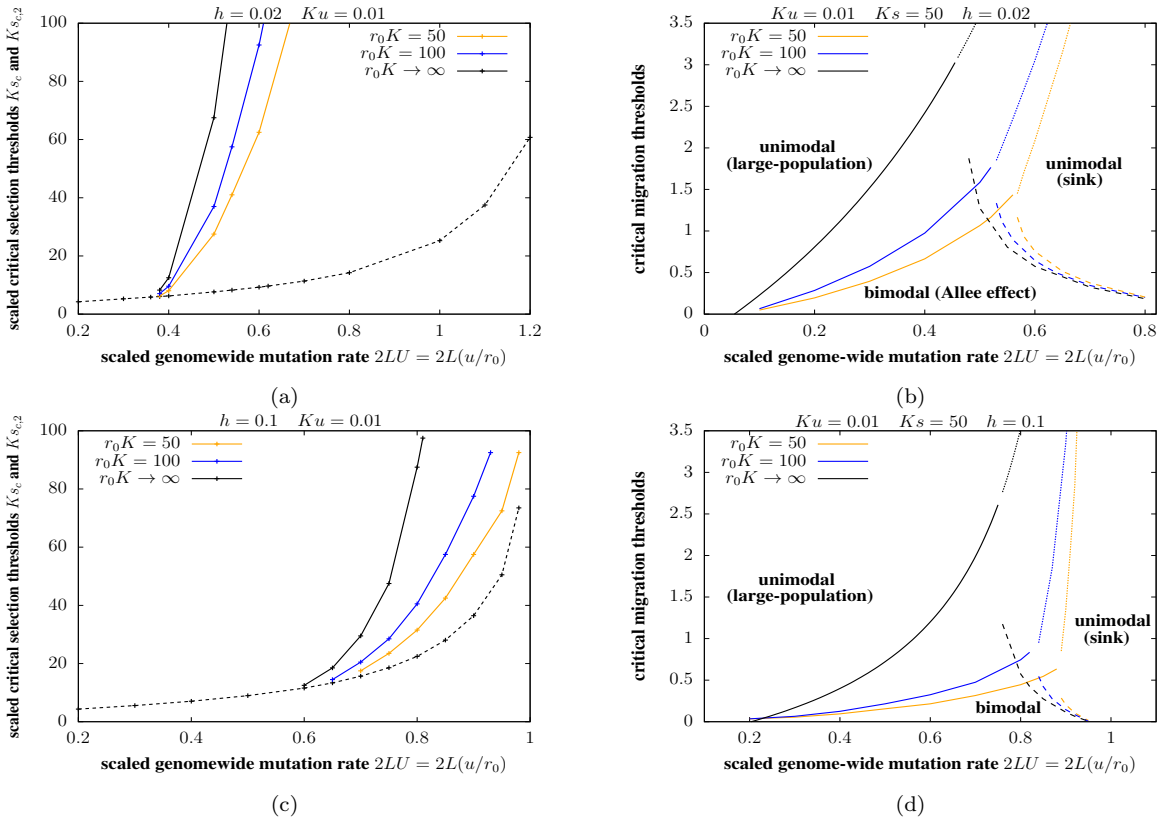


Figure 4.3: Semi-deterministic predictions for critical thresholds with and without accounting for demographic effects of migration. (A) and (C): Critical selection thresholds Ks_c (dashed line) and $Ks_{c,2}$ (solid lines) as a function of the scaled mutation target $2LU = 2L(u/r_0)$ with (A) $h = 0.02$ and (C) $h = 0.1$, for $r_0K = 50$, $r_0K = 100$ and for $r_0K \to \infty$ (in which limit migration has negligible demographic effects). The parameter r_0K, which governs the magnitude of the demographic effect of migration (via the term $M_0 = m_0/(r_0K)$) is varied by changing K, while simultaneously varying s, u and L such that Ks, Ku and $2LU = 2L(u/r_0)$ are constant (so that the genetic effects of migration remain unchanged). The threshold Ks_c is such that populations rapidly go extinct for $Ks < Ks_c$ in the limit of zero migration, but are metastable for $Ks > Ks_c$, if starting from large but not small sizes (so that there is a genetic Allee effect). The second threshold $Ks_{c,2}$ is such that increasing migration destabilizes the sink state for $Ks > Ks_{c,2}$, but destabilizes the large-population state for $Ks_c < Ks < Ks_{c,2}$. The threshold Ks_c is independent of migration, but $Ks_{c,2}$ increases as r_0K increases, i.e., as the demographic effect of migration becomes weaker. (B) and (D) The critical migration thresholds $m_{c,1}$ (solid lines), $m_{c,2}$ (dashed lines) and $m_{c,3}$ (dotted lines) vs. $2LU = 2L(u/r_0)$ with (B) $h = 0.02$ and (D) $h = 0.1$, for various r_0K for $Ks = 50$. The threshold $m_{c,1}$, which separates parameter regimes with bimodal population size distributions (characterised by alternation between the sink state and the large-population state) and unimodal size distributions (populations always in the large-population state) increases with increasing r_0K. The threshold $m_{c,2}$, which separates parameter regimes with bimodal size distributions and unimodal distributions (populations always in the sink state) decreases with increasing r_0K. The threshold $m_{c,3}$, which separates parameter regimes with load greater than or less than the baseline growth rate r_0, increases with increasing r_0K. All predictions are for $Ku = 0.01$.

Figures 4.3a and 4.3c show the critical selection thresholds Ks_c (as in fig. 4.1a) and $Ks_{c,2}$ versus the scaled mutation target $2LU = 2L(u/r_0)$ for the two values of h, for various carrying capacities (with Ks, Ku, $2LU = 2L(u/r_0)$ held constant, as K is varied). The threshold Ks_c, which relates to population outcomes in the zero migration limit,

is (by definition) independent of K in the semi-deterministic setting, as increasing K merely weakens the demographic effects of migration. The threshold $Ks_{c,2}$ decreases as K decreases, i.e., as the demographic effects of migration become stronger. This means that when alleles are moderately deleterious, i.e., characterised by a certain intermediate value of Ks, populations are stabilized (i.e., rescued from recurrent collapse into the high-load sink state) by increasing migration more easily on *smaller* islands, where the demographic effects of migration are stronger.

Figure 4.3b and 4.3d show the critical migration thresholds $m_{c,1}$, $m_{c,2}$ and $m_{c,3}$ vs. $2LU$ for a given Ks value (chosen to be 50). Here, with $h = 0.02$ for example (fig. 4.3b), increasing migration causes the sink state to vanish for $2LU \lesssim 0.5$, but degrades the large-population (low-load) state for $2LU \gtrsim 0.5$. The threshold $m_{c,1}$ (solid lines) is highly sensitive to K: populations can be stabilized at a finite fraction of carrying capacity at much lower levels of migration on smaller islands (given Ks, Ku), suggesting a key role of the demographic effects of migration. We can obtain an explicit expression for the threshold $m_{c,1}$ in the limit $K \to \infty$ (so that $M_0 = m_0/(r_0 K)$), which governs the demographic effects of migration, is negligible):

$$m_{c,1} = \frac{LSp^{(m)} - 1}{4[1 - LSp^{(m)}(p^{(m)} + 2h(1 - p^{(m)}))]} \qquad \text{as} \quad K \to \infty \tag{4.4}$$

Interestingly, this threshold depends only on the load, $LSp^{(m)}(p^{(m)} + 2h(1 - p^{(m)}))$, and breeding value, $LSp^{(m)}$, among migrants, and is independent of Ks and Ku: this simply reflects the fact that the growth rate of a very small population (just after recolonization) depends primarily on the genetic composition of founders, (and not selection on their subsequent descendants). Thus, the critical level of migration required to prevent a genetic Allee effect in the limit of very large carrying capacities is also independent of Ks and Ku.

The threshold $m_{c,2}$, which signals the collapse of the large-population state, is less sensitive to K: this is consistent with our expectation that demographic effects of migration should be less important when numbers are larger. There is, nevertheless, a moderate decrease in $m_{c,2}$ with K, which can be rationalised as follows: the (detrimental) genetic effects of migration on the large-population state are more effectively compensated by its demographic effects when islands are smaller (lower carrying capacity), allowing the large-population state to persist despite higher levels of migration on such islands. Finally, the third threshold $m_{c,3}$, which signals the emergence of a migration-dependent low-load state (at large $2LU$) is also highly sensitive to the demographic effects of migration, and becomes unrealistically high in the large K limit, where increasing migration can only shift allele frequencies but has little effect on population numbers (relative to carrying capacity).

A comparison of the top and bottom rows in fig. 4.3, which correspond respectively to $h = 0.02$ and $h = 0.1$, shows that as deleterious alleles become less recessive (larger h), the parameter regime in which increasing migration eliminates the large-population equilibrium shrinks drastically, and only emerges for very high genome-wide mutation rates — with $h = 0.1$, the second threshold $Ks_{c,2}$ exists only for $2LU \gtrsim 0.8$ in fig. 4.3c. This is consistent with the fact that purging of recessive mutations in smaller populations (which allows them to evolve significantly lower load than mainland populations) is only effective for very low h; thus, gene flow from the mainland becomes less detrimental to the large-population equilibrium in island populations for less recessive alleles. In fact, for $h \gtrsim 0.15$, increasing migration always causes the sink equilibrium to vanish, irrespective of Ks. Moreover, the migration threshold $m_{c,1}$ at which the sink state vanishes falls with increasing h (compare figs. 4.3b and 4.3d). Thus, we observe a genetic Allee effect only at very low migration rates for moderately recessive alleles. This is consistent with the fact that inbreeding load in small populations (owing to excess homozygosity) becomes lower as alleles becomes less recessive, and is alleviated by even low levels of migration.

4.4 Discussion

A key parameter governing the fate of peripheral populations is $2LU = 2L(u/r_0)$, the genome-wide deleterious mutation rate relative to the baseline rate of population growth: low-load (large-population) states are possible only for $2LU \lesssim 1$, provided selection against deleterious variants is sufficiently strong and/or migration high. Conversely, for $2LU \gtrsim 1$, populations exist only as demographic sinks, irrespective of selection strength. The parameter $2LU$ is a measure of the 'hardness' of selection and can be small either if the total mutation rate $2Lu$ (which determines total load in the absence of drift) is small or, more realistically, if the growth rate r_0 i.e., the logarithm of the baseline fecundity is high (corresponding to the soft selection limit).

Our analysis highlights qualitatively different effects of migration on population outcomes, depending on fitness effects and the total mutation target of deleterious variants (figs. 4.2 and 4.3). For example, with $2LU = 0.5$ (which corresponds to a $\approx 50\%$ reduction in growth rate owing to genetic load in a deterministic population), typical values of Ks must be at least approximately 5 for recessive and approximately 10 for additive alleles (reflecting the twofold difference in load between recessive vs additive alleles), if populations are to be metastable in the absence of migration. For weaker selection, gene flow from the mainland is beneficial, and aids population survival by hindering the fixation of deleterious alleles. However, for stronger selection and with recessive alleles, the fitness and size of 'low-load' island populations actually declines with increasing migration from the mainland (due to higher deleterious allele frequencies in the latter). In the

most extreme scenario, where load is primarily owing to segregation of recessive alleles of moderate effect, intermediate levels of migration may actually increase load so much that populations degenerate into high-load demographic sinks (fig. 4.2b and fig. C.4 (in appendix C.4)).

We identify two regimes in which peripheral populations can maintain stable numbers at a substantial fraction of carrying capacity, with qualitatively different roles of migration in the two. When selection is strong, i.e., $Ks \gg Ks_c$ (for additive alleles) or $Ks > Ks_{c,2}$ (for recessive alleles) for a given $2LU$, genetic load is low and populations stable, largely independent of migration. In this case, low levels of migration (typically $\lesssim 1$ migrant per generation; note typical values of $m_{c,1}$ in fig. 4.3) are sufficient to prevent a genetic Allee effect, should demographic stochasticity or chance fluctuations in load drive population numbers down. On the other hand, with weakly or moderately deleterious alleles, stable populations rely on rather high levels of migration ($\gg 1$ migrant per generation; note typical $m_{c,3}$ in fig. 4.3) and are only weakly differentiated with respect to the mainland. Here, migration is essential for maintaining heterozygosity and preventing fixation of deleterious alleles, even though numbers are relatively large.

Both classical quantitative genetics and analyses of allele frequency spectra suggest that most mutation load is owing to weakly deleterious alleles (Charlesworth, 2015, 2018). The weak selection on deleterious mutations may be strong relative to random drift in the species as a whole (Charlesworth, 2015), but is likely to be dominated by random drift within local demes (i.e., $Ks < 1$ in our notation). If this is so, then extinction can be avoided only if many migrants enter demes in each generation. Fortunately, this is generally the case: F_{ST} is typically small (King et al., 2001), implying that m_0 is between 0.5 and 5 (note that m_0 is the number of migrant genes, corresponding to $2Nm$ in the usual diploid notation). Thus, while selection can act effectively to suppress the mutation load in a well-connected metapopulation, demes that receive few migrants will be vulnerable to the accumulation of load due to weakly selected mutations. Moreover, when fitness has additional environment-dependent components, local adaptation must depend on alleles of intermediate effect and is hindered by high migration, especially for marginal habitats within the metapopulation. (Szép et al., 2021).

What implications might our work have for the conservation of natural populations? Provided that several migrants are exchanged per generation, selection against deleterious alleles can be effective across the whole population. Indeed, subdivision into small subpopulations can help purge deleterious recessives, making selection more effective than with panmixia. Thus, random drift would lead to a severe load only if local demes are highly isolated–in which case environmental fluctuations are more likely to cause extinction than the gradual accumulation of weakly deleterious mutations (Lande, 1988). Our work implies that an intermediate rate of migration minimizes mutation load, by

preventing extinction of local populations, and yet still allowing some purging. However, the extinction risk arising from environmental fluctuations (which we underestimate by including only demographic stochasticity) favours higher migration. Conversely, local adaptations require selection that is stronger than drift within local demes ($Ks > 1$) (Szép et al., 2021); if this is a concern, then substantial deme sizes are required in the long term.

Our model makes various assumptions: first, we take mutation rates to and from the deleterious state to be equal. Asymmetry in mutation rates would not qualitatively alter our conclusions as long as there is even weak migration ($m_0 \gtrsim 0.1$), as load is then alleviated primarily by migration rather than reverse mutation. However, with zero migration and no reverse mutation, selection must be strong enough to prevent long-term ratchet-like accumulation of deleterious variants, resulting in a much higher threshold Ks_c for metastability.

Second, we assume deleterious allele frequencies on the mainland to be close to deterministic. This assumption is not crucial: our qualitative conclusions remain unaltered as long as mainland frequencies are much lower than typical island frequencies. However, if mainland populations are small enough to harbour deleterious alleles at high frequencies at a subset of loci (which would, in general, be different from the loci fixed for deleterious alleles on the island), then we expect heterosis and the beneficial effects of migration to be weaker (with one-way migration between the mainland and island). More generally, extending this analysis to the co-evolution of load and population sizes in a metapopulation, where each sub-population may be close to fixation for different deleterious alleles, remains an interesting direction for future work.

Third, we assume a rather simple genetic architecture of load: loci are assumed to be unlinked, and load additive across loci. Deviations from additivity, e.g. synergistic epistasis between deleterious variants can lower load (Kimura and Maruyama, 1966; Kondrashov, 1988), and may arise, for example, if multiple traits are under stabilizing selection. However, linkage between deleterious variants can inflate load by compromising selection efficacy at individual loci via Hill–Robertson interference (Hill and Robertson, 1966), making it difficult to arrive at general predictions for the effects of selective interference (owing to linkage and epistasis between selected variants) on the eco-evolutionary dynamics of marginal populations. Fourth, we ignore environmental stochasticity and demographic Allee effects, which may strongly influence outcomes (Lande, 1993; Courchamp et al., 2008), especially in parameter regimes where mutation accumulation and demographic stochasticity in themselves are unlikely to cause extinction.

Finally, our ananlysis relies on a semi-deterministic analysis, which accounts for genetic drift but neglects demographic stochasticity: while this approximation captures various qualitatively different population states across parameter space, it gives little insight

into the dynamics. In particular, where alternative — low-load and high-load states are possible, the key assumption underlying the semi-deterministic analysis — namely, that allele frequencies have sufficient time to equilibrate at any given population size, is satisfied only when populations are in one or other state, and not while they transition between states. This makes it challenging to describe the rather complex co-evolution of load and population size during transitions and arrive at a complete understanding of the factors governing transition timescales.

Our aim has been to base our analysis on as few parameters as possible, in the hope that these can be related to observations from nature. We have reasonably good estimates of the fitness effects of deleterious mutations, and their degree of dominance – albeit largely from *Drosophila* (Charlesworth, 2015). We also now have accurate measures of the total mutation rate; the total rate of deleterious mutations is still uncertain (Graur et al., 2013), but may be substantial in complex organisms (Böndel et al., 2019). Population structure is less well understood: we have very many estimates of F_{ST} (Morjan and Rieseberg, 2004), which reflects the numbers of incoming migrants, but local effective deme size is harder to estimate, even if demes can be defined at all. However, the common observation of heterosis implies that different deleterious recessives are common in different populations, suggesting a substantial drift load.

The rather complex effects of migration that emerge even in this relatively simple model with unconditionally deleterious alleles suggest that a comprehensive understanding of the effects of gene flow on eco-evolutionary dynamics at range limits must account for both environment-dependent (local) and environment-independent (global) components of fitness. These may be influenced by (partially) overlapping sets of genetic variants, so that genetic load is shaped fundamentally by pleiotropic constraints. Such extended models are key to understanding when, for example, assisted gene flow is beneficial, and whether its mitigatory effect on inbreeding depression may be outweighed by any outbreeding depression that it might generate.

From a conceptual viewpoint, our analysis highlights the importance of considering explicit population dynamics when analysing the influence of gene flow on the efficacy of selection in small or sub-divided populations. Simple predictions, e.g., that a sub-divided population under hard selection should behave as a single population with an inbreeding coefficient equal to F_{ST} (Caballero et al., 1991; Whitlock, 2002), may break down when sub-populations can undergo local extinction. In this case, purging may be ineffective and the efficacy of selection reduced relative to undivided populations, in contrast to standard predictions that subdivision always reduces load when selection is hard. We regard the framework presented here as a starting point for detailed studies in specific metapopulations, which take into account the joint evolution of population size and mutation load.

References

Agrawal, A. F. and Whitlock, M. C. (2011). Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics*, 187(2):553–566. https://doi.org/10.1534/genetics.110.124560.

Barton, N. H. and Etheridge, A. (2018). Establishment in a new habitat by polygenic adaptation. *Theoretical Population Biology*, 122:110–127. https://doi.org/10.1016/j.tpb.2017.11.007.

Bell, G. and Gonzalez, A. (2011). Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. *Science*, 332(6035):1327–1330. https://doi.org/10.1126/science.1203105.

Böndel, K. B., Kraemer, S. A., Samuels, T., McClean, D., Lachapelle, J., Ness, R. W., Colegrave, N., and Keightley, P. D. (2019). Inferring the distribution of fitness effects of spontaneous mutations in chlamydomonas reinhardtii. *PLoS biology*, 17(6):e3000192. https://doi.org/10.1371/journal.pbio.3000192.

Bridle, J. R. and Vines, T. H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in ecology & evolution*, 22(3):140–147. https://doi.org/10.1016/j.tree.2006.11.002.

Caballero, A., Keightley, P., and Hill, W. (1991). Strategies for increasing fixation probabilities of recessive mutations. *Genetics Research*, 58(2):129–138. https://doi.org/10.1017/S0016672300029785.

Charlesworth, B. (2015). Causes of natural variation in fitness: evidence from studies of drosophila populations. *Proceedings of the National Academy of Sciences*, 112(6):1662–1669. https://doi.org/10.1073/pnas.1423275112.

Charlesworth, B. (2018). Mutational load, inbreeding depression and heterosis in subdivided populations. *Molecular ecology*, 27(24):4991–5003. https://doi.org/10.1111/mec.14933.

Courchamp, F., Berec, L., and Gascoigne, J. (2008). *Allee effects in ecology and conservation*. OUP Oxford.

Frankham, R. (2005). Genetics and extinction. *Biological conservation*, 126(2):131–140. https://doi.org/10.1016/j.biocon.2005.05.002.

Graur, D., Zheng, Y., Price, N., Azevedo, R. B., Zufall, R. A., and Elhaik, E. (2013). On the immortality of television sets:"function" in the human genome according to the evolution-free gospel of encode. *Genome biology and evolution*, 5(3):578–590. https://doi.org/10.1093/gbe/evt028.

Gyllenberg, M. and Hanski, I. (1992). Single-species metapopulation dynamics: a structured model. *Theoretical Population Biology*, 42(1):35–61. https://doi.org/10.1016/0040-5809(92)90004-D.

Higgins, K. and Lynch, M. (2001). Metapopulation extinction caused by mutation accumulation. *Proceedings of the National Academy of Sciences*, 98(5):2928–2933. https://doi.org/10.1073/pnas.031358898.

Hill, W. G. and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, 8(3):269–294. https://doi.org/10.1017/S0016672300010156.

Huber, C. D., Durvasula, A., Hancock, A. M., and Lohmueller, K. E. (2018). Gene expression drives the evolution of dominance. *Nature communications*, 9(1):2750. https://doi.org/10.1038/s41467-018-05281-7.

Kawecki, T. J. (2008). Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, 39:321–342. https://doi.org/10.1146/annurev.ecolsys.38.091206.095622.

Kimura, M. and Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics*, 54(6):1337. https://doi.org/10.1093/genetics/54.6.1337.

King, T., Kalinowski, S. T., Schill, W., Spidle, A., and Lubinski, B. (2001). Population structure of atlantic salmon (salmo salar l.): a range-wide perspective from microsatellite dna variation. *Molecular Ecology*, 10(4):807–821. https://doi.org/10.1046/j.1365-294X.2001.01231.x.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, 336(6198):435–440. https://doi.org/10.1038/336435a0.

Lande, R. (1988). Genetics and demography in biological conservation. *Science*, 241(4872):1455–1460. https://doi.org/10.1126/science.3420403.

Lande, R. (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist*, 142(6):911–927. https://doi.org/10.1086/285580.

Lande, R. (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution*, 48(5):1460–1469. https://doi.org/10.1111/j.1558-5646.1994.tb02188.x.

Lande, R., Engen, S., and Sæther, B.-E. (1998). Extinction times in finite metapopulation models with stochastic local dynamics. *Oikos*, pages 383–389. https://doi.org/10.2307/3546853.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in ecology & evolution*, 17(4):183–189. https://doi.org/10.1016/S0169-5347(02)02497-7.

Lynch, M., Conery, J., and Bürger, R. (1995). Mutational meltdowns in sexual populations. *Evolution*, 49(6):1067–1080. https://doi.org/10.2307/2410432.

Morjan, C. L. and Rieseberg, L. H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular ecology*, 13(6):1341–1356. https://doi.org/10.1111/j.1365-294X.2004.02164.x.

Ovaskainen, O. and Hanski, I. (2003). Extinction threshold in metapopulation models. In *Annales Zoologici Fennici*, pages 81–97. JSTOR. https://www.jstor.org/stable/23736517.

O'Grady, J. J., Brook, B. W., Reed, D. H., Ballou, J. D., Tonkyn, D. W., and Frankham, R. (2006). Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological conservation*, 133(1):42–51. https://doi.org/10.1016/j.biocon.2006.05.016.

Ronce, O. and Kirkpatrick, M. (2001). When sources become sinks: migrational meltdown in heterogeneous habitats. *Evolution*, 55(8):1520–1531. https://doi.org/10.1111/j.0014-3820.2001.tb00672.x.

Roze, D. (2015). Effects of interference between selected loci on the mutation load, inbreeding depression, and heterosis. *Genetics*, 201(2):745–757. https://doi.org/10.1534/genetics.115.178533.

Szép, E., Sachdeva, H., and Barton, N. H. (2021). Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5):1030–1045. https://doi.org/10.1111/evo.14210.

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

Whitlock, M. C. (2002). Selection, load and inbreeding depression in a large metapopulation. *Genetics*, 160(3):1191–1202. https://doi.org/10.1093/genetics/160.3.1191.

Wright, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences*, 23(6):307–320. https://doi.org/10.1073/pnas.23.6.307.

Chapter 5

Genetic load, eco-evolutionary feedback and extinction in a metapopulation[1]

Oluwafunmilola Olusanya, Khudiakova Kseniia, Himani Sachdeva

Abstract

Fragmented landscapes pose a significant threat to the persistence of species as they are highly susceptible to heightened risk of extinction due to the combined effects of genetic and demographic factors such as genetic drift and demographic stochasticity. This paper explores the intricate interplay between genetic load and extinction risk within metapopulations with a focus on understanding the impact of eco-evolutionary feedback mechanisms. We distinguish between two models of selection: soft selection, characterised by subpopulations maintaining carrying capacity despite load, and hard selection, where load can significantly affect population size. Within the soft selection framework, we investigate the impact of gene flow on genetic load at a single locus, while also considering the effect of selection strength and dominance coefficient. We subsequently build on this to examine how gene flow influences both population size and load under hard selection as well as identify critical thresholds for metapopulation persistence. Our analysis employs the diffusion, semi-deterministic and effective migration approximations. Our findings reveal that under soft selection, even modest levels of migration can significantly alleviate the burden of load. In sharp contrast, with hard selection, a much higher degree of gene flow is required to mitigate load and prevent the collapse of the metapopulation. Overall, this study sheds light into the crucial role migration plays in shaping the dynamics of

[1]This work can be found online at https://www.biorxiv.org/content/10.1101/2023.12.02.569702v1

genetic load and extinction risk in fragmented landscapes, offering valuable insights for conservation strategies and the preservation of diversity in a changing world.

Keywords: population fragmentation, genetic load, extinction, migration, soft selection, hard selection.

5.1 Introduction

The long held belief that ecological (eco) and evolutionary (evo) processes occur on too different timescales to influence each other has fallen under scrutiny as many studies (Thompson, 1998; Hairston Jr et al., 2005) have documented that these timescales can be comparable: this can be the case when selection varies sharply over space and time, or when populations are too small to adapt efficiently, resulting in rapid loss in fitness (e.g., due to the accumulation of deleterious alleles).

Such eco-evolutionary feedback is particularly important in fragmented landscapes with small-sized patches, where mating between genetically related individuals within patches (i.e., inbreeding) as well as stochastic changes in genetic composition (due to genetic drift) can compromise the efficiency of natural selection. As a consequence, populations in such landscapes may become more susceptible to fixing deleterious mutations, resulting in an increase in genetic load and a decline in population numbers, which can further exacerbate both genetic drift and inbreeding. The resultant positive feedback loop between increasing load and falling population sizes may drive populations towards extinction through a process termed *mutational meltdown* (Lynch et al., 1995b). Evidence of such extinction events have been found in plant populations (Matthies et al., 2004) and in some vertebrate (Fagan and Holmes, 2006) and invertebrate (see *M. cinxia*) populations (Saccheri et al., 1998). This kind of feedback may be further aggravated by demographic and environmental stochasticity, both of which can limit the potential of populations to adapt to changing conditions. However, despite some understanding of individual factors contributing to extinction risk, a comprehensive picture of how deterministic and stochastic processes together structure genetic diversity and influence the persistence of populations remains lacking. This understanding is crucial for formulating effective conservation strategies that can safeguard biodiversity and promote long-term survival of endangered species (Lande, 1993; Frankham, 1998) particularly as human activities continue to exert significant pressures on natural environments. Gaining such quantitative understanding requires us to consider how fragmented the landscape is: how many local populations they contain, what their sizes are, to what extent they are connected by gene flow, etc.

Migration can affect various aspects of the eco-evo feedback loop, in essence homogenizing both genetic composition and population density across fragmented populations. Investi-

gations of natural populations yield examples of both the beneficial and deleterious effects of migration. For instance, Finger et al. (2011) demonstrated how augmented gene flow reduced the negative effects of inbreeding in a critically endangered and isolated jellyfish tree (*Medusagyne oppositifolia*) population. Similarly, Land and Lacy (2000) showed that introducing eight wild-caught Texan female panthers into a small isolated Florida panther population caused a tripling of its size, bringing it back from near extinction within just 12 years of re-introduction. In contrast, Herman et al. found that gene flow between surface and cave populations of Mexican tetra, *Astyanax mexicanus*, had a swamping effect on cave-related traits. These examples show how migration can have very complex effects - they can have a 'rescue effect' when migrants are introduced into a vulnerable population or they can have negative consequences or generally engender a tension between inbreeding and outbreeding depression, in particular, when migration occurs between populations that are adapted to different environmental conditions or happens between distinct species (Frankham et al., 2011; Edmands, 2007; Templeton, 1986)). These varied effects of migration introduce complexities to practical decisions in conservation such as those concerning assisted gene flow (Aitken and Whitlock, 2013), underscoring the need for a more quantitative and theoretical understanding of gene flow in fragmented landscapes.

Migration can have rather intricate effects on genetic diversity even when there is no local adaptation i.e., when different local populations are subject to uniform selection pressures, e.g., due to purifying selection against unconditionally deleterious alleles. Small and somewhat isolated local populations may be nearly fixed for deleterious alleles at different loci; migration between such populations can result in hybrid offspring of increased vigour due to masking of recessive (deleterious) mutations. Migration can also increase fitness variation, which has two competing effects - it may alleviate drift load by preventing fixation of deleterious alleles (especially at low levels of migration), but also prevent purging of recessive alleles by increasing heterozygosity (at higher levels of migration) (Glémin et al., 2003).

The effect of migration on genetic diversity and adaptation in structured populations depends crucially on whether selection is "soft" or "hard", i.e., whether different local populations contribute equally to the next generation regardless of fitness, or if the contribution of fitter populations is higher (e.g., when these are larger and send out more migrants). In the latter case, adaptation is expected to be most efficient at intermediate migration rates (Uecker et al., 2014; Gomulkiewicz et al., 1999), i.e., when gene exchange between demes is sufficiently low so as not to displace exceptionally fit local populations from the highest "adaptive peaks", but high enough to eventually cause the entire population to move towards these peaks (Wright, 1931; Rouhani and Barton, 1993).

However, most work on hard vs. soft selection in subdivided populations assumes a rather

specialised life cycle in which all (adult or juvenile) individuals across all demes join a common pool prior to reproduction, followed by uniform redistribution of zygotes back into demes (Levene, 1953; Christiansen, 1975; Ravigné et al., 2004), making it difficult to assess the role of limited dispersal. In particular, models of hard selection typically assume that while different local demes can vary in size and thus contribute differentially to the common pool of reproducing adults, the metapopulation as a whole is under global density regulation and is moreover at carrying capacity (see, e.g., Whitlock (2002)). This makes it difficult to use such models to assess how genetic load can decrease local populations over multiple generations due to the positive feedback between declining population size and increasing load (or vice versa) and to what extent this kind of meltdown may be arrested by (limited) dispersal between populations– questions that are crucial for a quantitative understanding of how migration ameliorates extinction risk in fragmented populations. This therefore calls for more realistic models that allow for variation in both local and global population sizes and explicitly incorporate eco-evolutionary feedback.

Lynch et al.(1995a; 1995b) considered such a model where load due to deleterious mutation accumulation leads to increased extinction risk. Using extensive computer simulations, they showed that populations with low effective size (typically less than 100 individuals) are at a significant risk of extinction via mutational meltdown within only about 100 generations (Lynch et al., 1995a); a conclusion with potential implications for management programs such as captive breeding programs. However, their analysis was limited to single randomly mating populations. Higgins and Lynch (2001) went beyond this to consider the dynamics of a metapopulation. Using metapopulation simulations with global and nearest-neighbour dispersal, they demonstrated that the eventual fate of a metapopulation (as measured by the median time to extinction) is critically dependent on the number of demes or metapopulation patches (see also Lande et al. (2003), Ch 4, Table. 4.1) as well as on the dispersal neighbourhood. Their analysis however primarily focused on extinction times and was restricted to scenarios where patch sizes were relatively small resulting in early population extinctions. Furthermore, their study was entirely simulation-based, making it hard to generalize their conclusions.

Going beyond simulations, Szép et al. (2021) introduced a stochastic polygenic model, that explicitly captures the coupling between population size and allele frequency at multiple loci and investigated how this influences local adaptation and extinction in metapopulations, as well as the role of gene flow and stochastic events i.e., demographic stochasticity and drift. They found that the extinction risk of metapopulations is higher when population sizes are small and the coupling between population size and allele frequency is strong. They also showed that local adaptation is more difficult under hard selection and if locally adaptive traits are more polygenic, causing populations to become extinct under much lower levels of gene flow than would be expected from single-locus

theory. However, their conclusions beg the question of how these results change when selection is uniform across space. Using a similar theoretical model, but with spatially uniform selection, Sachdeva et al. (2022) investigated whether or not asymmetric gene flow (from a mainland to an island) can help arrest mutational-meltdown due to eco-evo feedback and thus prevent extinction of the island population. They found that migration can have qualitatively different effects on the island; having a positive effect (i.e., reducing load and thus the risk of extinction) when the island population is small and isolated (i.e., in a sink state) and having an opposite effect (i.e., increasing load by hindering purging) when it is large and connected.

However, beyond marginal populations that are plagued by asymmetric gene flow from central populations, it is important to understand how eco-evo feedback influences extinction risk and load in a metapopulation where migration is random and where the different islands (demes) can fix for different subsets of deleterious alleles that mask one another. We therefore extend the work of Sachdeva et al. (2022) to explore the eco-evolutionary dynamics of a metapopulation made up of a large number of demes exchanging migrants at random (i.e., under the island model). Our goal is to investigate the effect of gene flow on equilibrium load and population size, as well as the critical thresholds required for the persistence (i.e., for non-extinction) of the metapopulation. We further analyse how these thresholds are influenced by the architecture of load (the genomewide deleterious mutation rate and the distribution of selective effects and dominance coefficients of deleterious mutations) as well as features of the metapopulation landscape (carrying capacities and baseline growth rate of local demes). As we see below, a key parameter that determines the fate of the metapopulation is the total mutation rate relative to the baseline growth rate, which determines the extent to which genetic load depresses population growth and can thus be considered a proxy for the "hardness of selection". Besides the advantage of explicitly incorporating the relationship between population size and load, our theoretical approach explicitly accounts for the indirect coupling between the dynamics of different loci: a small increase in drift load at very many weakly selected loci can have rather strong effects on total load which can further exacerbate the effects of drift per locus, resulting in an indirect coupling between different loci even in the absence of epistasis. This indirect coupling adds more complexity to the eco-evo dynamics and can have significant implications for population fitness and evolutionary outcomes in the metapopulation. We later go on to discuss the application of our results to conservation issues.

5.2 Model and Methods

We consider a metapopulation with n_D ($i = 1 \cdots, n_D$) local populations or patches all connected to each other via migration. In each patch i, mating occurs randomly and adult

individuals are allowed to undergo only one breeding season per generation. Individuals are diploid and subject to deleterious mutations at L unlinked loci at a rate u per haploid locus, resulting in a genome-wide deleterious mutation rate of $2Lu$. We also allow for mutations in the reverse direction at rate v. The fitness of an individual is multiplicative across loci where at any locus the wild-type has fitness 1, the heterozygote has fitness e^{-hs} and the mutant has fitness e^{-s} respectively, with h representing the dominance coefficient and s, the strength of selection against the mutant allele. We assume here that s and h are the same across all loci, a gross simplification that is relaxed in appendix D.0.7. Patches are also allowed to have different population sizes and mean fitness.

In each generation, a fraction m of individuals in each patch i migrate into a common pool and individuals from this pool are then evenly dispersed back to the patches. We assume that all patches experience the same environment, so that the fitness of any genotype is independent of the patch. We further assume a logistic model of population growth in each i i.e., $n_i(t+1) = n_i(t)\ \mathrm{Exp}[r_0(1 - n_i(t)/K))]\overline{W_i}$, where the population size, $n_i(t+1)$ in i after reproduction and density regulation depends on its previous size $n_i(t)$, its carrying capacity K, the baseline growth rate, r_0 as well as on the mean fitness, $\overline{W_i}$ of individuals (hard selection) in i which holds true in a number of natural populations. Note that the carrying capacity K and the baseline growth rate, r_0 are assumed to be the same across all patches.

The parameters of our model are therefore, the number of loci L, the rate of mutation $v = g\ u$ (where g represents the degree of mutational bias; we mostly assume $g = 1$), the strength of selection s, the dominance coefficient, h, the migration rate, m, the baseline growth rate, r_0 in each patch and the carrying capacity K. As will be argued below, the behaviour of our model is governed largely by composite parameters Km (which represents the strength of migration scaled by the carrying capacity), Ks (which represents the per locus strength of selection scaled by the carrying capacity), Ku (which is the deleterious mutation rate scaled by the carrying capacity), h, $2LU = 2Lu/r_0$ (which is a measure of the hardness of selection) and r_0K (whose inverse represents the strength of demographic fluctuations).

To guide the choice of parameter ranges, we look to some estimates in the literature. For example, estimates of the frequency at which deleterious mutations appear in the genome per individual per generation has been documented to be of the order of $0.1 - 1$ in multicellular eukaryotes (Lynch et al., 1999), $1.2 - 1.4$ in *Drosophila* (Haag-Liautard et al., 2007), $0.25 - 2.5$ for *C. elegans* (Denver et al., 2004) and 2.2 for humans (Keightley, 2012), the latter being an unusually high value. Similarly, data on the distribution of fitness effects remain controversial as different studies have generated varied selection estimates on phenotypic traits. Although some experimental studies have shown that a sizeable fraction ($< 15\%$) of mutations are likely to be lethal (Mukai et al., 1972;

Eyre-Walker et al., 2006), there is consensus that small effect mutations (with effect $< 5\%$ on quantitative traits) have higher densities (Mukai, 1964; Mukai et al., 1972; Lyman et al., 1996; Lynch et al., 1999). Dominance coefficients are harder to estimate for weakly deleterious mutations but may be ~ 0.2 for moderately deleterious variants (Charlesworth and Charlesworth, 1987). In addition, several studies point to a negative relationship between s and h, with mutations of small effect being almost additive and those of large effect being almost recessive (Greenberg and Crow, 1960; Simmons and Crow, 1977; Lynch et al., 1999; Gillespie, 2004). With regards to the baseline growth rate, r_0, this is a very challenging parameter to estimate as it can fluctuate due to a multitude of interrelated factors. Statistical and mathematical models that incorporate factors like the carrying capacity of the population in question, birth and death rates as well as other ecological parameters are often used to predict r_0 but they can only provide an approximation at best due to the complexities involved.

Finally, indirect measures (e.g, F_{ST}) have reported moderate to high levels of gene flow in plants and mammals with typical values between closely related species being of the order $0.05-0.2$. For example, F_{ST} is ~ 0.099 between North American and Eurasian populations of gray wolf and ~ 0.018 between gray and red wolf populations which corresponds roughly to 2 and 14 migrants respectively per generation between local populations. Similarly, average F_{ST} was found to be ~ 0.07 in a study of 12 populations of *Camelina sativa* and ~ 0.077 in 337 species of seed plants (Gamba and Muchhala, 2020). Studies have however found that these F_{ST} values are higher in plants with limited dispersal mechanism such as those that are self-pollinated or wind-dispersed. For instance, Hamrick et al. (1990) found F_{ST} values to be of the order of 0.32 in a population of trees that dispersed their seeds by gravity and of the order of 0.15 in those that dispersed their seeds by wind; corresponding to approximately 5 and 14 migrants every ten generations. It is important to note that in populations that are on the verge of extinction, gene flow may be much less than this (Frankham et al., 2002; Casas-Marce et al., 2013; Szczecińska et al., 2016) and indeed one of the goals of this study is to identify critical levels of migration required to prevent population collapse. Thus, most of the study will focus on low to moderate levels of gene flow. In addition, we will consider a genome-wide mutation rate (scaled relative to the growth rate), $2LU < 1$ (which corresponds to $< 63\%$ reduction in fitness), otherwise, populations will not grow. We will also consider both recessive and additive alleles and growth rates of the order of 0.1.

Given this set-up, we are generally interested in understanding the varied effects migration has on populations outcomes and the role parameters such as dominance, selective effects and mutational bias play in this. Using simulations and analytical approximations, we explore this problem from two different angles.

First, we consider a soft selection model where the maximum possible genetic load, $2Lu$

in any deme is much less than the intrinsic growth rate (i.e., $2Lu/r_0 \ll 1$) so that demic population size (and hence that of the metapopulation) is always constant over time (even at high levels of maladaptation) thus ignoring the possibility of local extinctions. With this model, we explore the effect of gene flow on load due to a single locus and how this depends on the selection strength and dominance coefficient at the given locus.

Secondly, we also consider a hard selection model where the maximum possible genetic load in a population or deme is appreciable relative to the intrinsic growth rate (i.e., $2Lu/r_0 \gtrsim 1$) so that population size now depends on the degree of maladaptation through load. This model therefore takes into account the interplay between population size and allele frequency, thus accounting for the possibility of local (and by extension global) extinction. Also, since population size differences exist among demes under this kind of model, larger and more fit islands would contribute more to the migrant pool and would be less influenced by incoming migrants. Using this model, we focus on understanding how migration influences population size and load (and hence metapopulation persistence); how it impacts critical extinction thresholds; and finally what the role of the hardness of selection is (where hardness of selection is simply a measure of the reduction in the growth rate of a population due to load and is quantified by the value of $2LU = 2Lu/r_0$; the larger this value, the more strongly is growth rate reduced, resulting in harder selection).

To gain insight into the above questions, we make use of a number of analytical approximations which are described below. First, we assume the diffusion approximation upon which we base our soft selection analysis. This will be used to obtain the equilibrium distribution of allele frequency at any given locus in any deme conditional on the mean allele frequency across the entire metapopulation (Wright, 1937).

In using the diffusion approximation (Wright, 1937), we make a simplifying assumption that alleles at different loci evolve independently (which is valid when all evolutionary and ecological processes are much weaker than the strength of recombination). However, to account for the effect of multilocus LD on load, we will consider a diffusion approximation with effective migration rates m_e called the 'effective migration approximation' (Sachdeva, 2022).

Under hard selection, population sizes can vary between demes and also within a deme over time, so that the metapopulation is characterised by a distribution of population sizes and we hence need to follow the joint distribution of population size and allele frequencies in any given deme. It turns out that the diffusion approximation cannot be used to obtain an explicit solution for such a joint distribution in the presence of mutation. Hence, to make analytical progress we look to a different approximation - *the semideterministic approximation* (see also section 4.2). This assumes that the population size of any deme is reduced with respect to carrying capacity by an amount that depends on the genetic load

and is thus *determined* by load, rather than following a distribution. It further assumes that the load is that expected at mutation-selection-migration-drift equilibrium for the population size, allowing us to solve self-consistently for the population size (and load). Over the time scale required to reach such equilibrium, population size remains relatively constant due to weak demographic fluctuations and thus the approximation accounts for genetic drift but not demographic stochasticity.

The different approximations are discussed below. First we describe the dynamics of the joint evolution of population size and allele frequencies.

5.2.1 Evolution of allele frequencies and population sizes in continuous time

If ecological and evolutionary processes are weak relative to recombination so that they cause only minor changes per generation, and if in addition, we ignore non-random associations and correlations among loci, then one can describe the dynamics of allele frequency, $p_{i,j}$, at any locus j and population size, N_i, in any deme i in continuous time as,

$$\frac{dp_{i,j}}{d\tau} = -\frac{p_{i,j}q_{i,j}}{2}\frac{\partial R_{g,i}}{\partial p_{i,j}} + U_i(q_{i,j} - p_{i,j}) + M\frac{\overline{N}}{N_i}\left(\frac{\overline{Np_j}}{\overline{N}} - p_{i,j}\right) + \lambda_{p_{i,j}} \qquad q_{i,j} = 1 - p_{i,j} \tag{5.1}$$

$$\frac{dN_i}{d\tau} = N_i(1 - N_i - R_{g,i}) + M(\overline{N} - N_i) + \lambda_{N_i} \quad \text{where} \quad R_{g,i} = \sum_{j=1}^{L} S\ p_{i,j}(2h_{i,j}q_{i,j} + p_{i,j}) \tag{5.2}$$

where $R_{g,i}$ is the load in deme i obtained as the sum of the load due to all j loci.

Equation (5.1) and (5.2) are non-dimensional equations that have been obtained by re-scaling population size, n relative to carrying capacity, K (i.e., $N = n/K$), re-scaling load, R_g relative to the rate of increase in the population from low numbers - the characteristic growth rate, r_0 (i.e., $R_g = r_g/r_0$) as well as re-scaling all other parameters relative to r_0 i.e., $S = s/r_0,\ U = u/r_0,\ M = m/r_0$ and $\tau = r_0\ t$. $q_{i,j} = 1 - p_{i,j}$, $\overline{N}$ is the population size averaged across all demes of the metapopulation and $\overline{Np_j}$ is the allele copy number at locus j averaged across all demes of the metapopulation. Finally $\lambda_{p_{i,j}}$ and λ_{N_i} are the stochastic parts of the equations. $\lambda_{p_{i,j}}$ represents the processes of drift (i.e., the stochastic fluctuations in allele frequencies due to random mating) and has mean, $\mathbb{E}[\lambda_{p_{i,j}}] = 0$ and variance, $\mathbb{E}[\lambda_{p_{i,j}}(t)\ \lambda_{p_{i,j}}(t')] = ((p_{i,j}(t)q_{i,j}(t))\,\delta(t-t'))\,/\,(\xi\ N_i(t))$. λ_{N_i} represents demographic stochasticity (population size fluctuations due to randomness in birth and death) and has mean, $\mathbb{E}[\lambda_{N_i}] = 0$ and variance, $\mathbb{E}[\lambda_{N_i}(t)\ \lambda_{N_i}(t')] = (N_i(t)\delta(t-t'))\,/\xi$. Note that $\xi = r_0\ K$ (its inverse determines the strength of demographic fluctuations).

5.2.2 The diffusion approximation for allele frequencies

The diffusion approximation has a long history in population genetics, being an effective mathematical tool for understanding how the long term behaviour of populations are shaped by evolutionary (ecological) processes.

Under soft selection, each deme is at carrying capacity such that population size in each deme is $n = K$ and is not impacted by load, hence, we can ignore eq. (5.2). The third term in eq. (5.1) then depends only on the scaled migration rate, M, the allele frequency at locus j and the mean allele frequency at j (i.e., averaged across all demes of the metapopulation). We can now write down equations for the time evolution of the joint distribution, $\Psi(p_1, \ldots, p_L|\overline{p})$ of allele frequencies for a fixed N (i.e., under soft selection). In the spirit of Wright, and under soft selection one can find an equilibrium solution for $\Psi(p_1, p_2, \ldots, p_L|\overline{p})$,

$$\Psi\left(p_1, p_2, \ldots, p_L|\overline{p}\right) = \prod_{j=1}^{L} \psi\left(p_j|\overline{p}_j\right) \qquad \text{with,} \tag{5.3}$$

$$\psi\left(p_j|\overline{p}_j\right) \propto p_i^{4Ku+4Km\overline{p}_j-1}(1-p_j)^{4Ku+4Km\overline{q}_j-1}e^{-2Ks(p_j^2+2hp_jq_j)};\ q_j = 1-p_j,\ \overline{q}_j = 1-\overline{p}_j \tag{5.4}$$

where ψ is the marginal distribution at j. The expected allele frequency $\mathbb{E}[p_j|\overline{p}_j]$ and heterozygosity, $\mathbb{E}[p_jq_j|\overline{p}_j]$ at any locus j in the population can be obtained by numerically integrating over eq. (5.4) i.e.,

$$\mathbf{E}[p_j|\overline{p}_j] = \frac{\int_0^1 p_j\ \psi(p_j|\overline{p}_j)dp}{\int_0^1 \psi(p_j|\overline{p}_j)dp}, \qquad \mathbb{E}[p_jq_j|\overline{p}_j, N] = \frac{\int_0^1 p_jq_j\ \psi(p_j|\overline{p}_j)dp}{\int_0^1 \psi(p_j|N, \overline{p}_j)dp}. \tag{5.5}$$

The expected load, $\mathbb{E}[R_g]$ in the population can then be obtained as the sum,

$$\mathbb{E}[R_g] = \sum_{j=1}^{L} S\left(\mathbb{E}[p_j] + (2h-1)\mathbb{E}[p_jq_j]\right) \tag{5.6}$$

Under hard selection, however, we cannot find a direct solution for the joint distribution of population size and allele frequency with mutation. Hence, we resort to a semi-deterministic approximation as described below.

5.2.3 The semi-deterministic approximation

As stated earlier, under the model of hard selection and in the presence of mutation, the diffusion approximation introduced above does not give an explicit formula for the equilibrium population size and load in any deme of a metapopulation. Hence, we introduce here a new approximation - the semi-deterministic approximation (Szép et al.,

2021; Sachdeva et al., 2022).

This allows us to get straightforward and accurate solutions for the expected size and load in any population at equilibrium. It is called semi-deterministic because it takes into consideration one kind of stochasticity (genetic drift) but ignores the other (demographic stochasticity). It assumes that allele frequencies in a population change rather slowly and that demographic fluctuations in the population are very weak (implying a fairly steady size), so that at equilibrium, the load can be approximated by the expectation at mutation-migration-selection-drift balance and population size can then be calculated by a simple expression that depends on the load.

Mathematically, this can be obtained by setting the l.h.s. of eq. (5.2) to zero and assuming that at selection-mutation-migration-drift equilibrium, $R_g \sim \mathbb{E}[R_g|N_*]$ where N_* represents the equilibrium population size. If we further assume that population sizes are roughly similar across patches (so that no island acts as a demographic source or sink), then population size on any island is just reduced (w.r.t. carrying capacity) by an amount proportional to load, i.e., $N_* \sim 1 - \mathbb{E}[R_g|N_*]$. This together with the equilibrium expression for $\mathbb{E}[R_g|N_*]$ (given in eq. (5.5)) allows us to solve for N_*.

The above assumptions imply that we account for the genetic effects of migration on load but ignore its demographic effects which can be important in source-sink dynamics (i.e., when population sizes differ across demes). Our approximation is thus limited in this respect. The situation is however not too dire as we shall later on see in the result section, even in scenarios with source-sink dynamics, our approximation accurately predicts the population sizes of the non-extinct patches, though it fails to predict the proportion that are extinct.

There are five important parameters that arise from our semi-deterministic approximation. They are, the dominance coefficient, h, that determines the degree of dominance or recessiveness of an allele; the selection strength scaled by the carrying capacity, Ks, that determines the strength of selection against deleterious homozygote at a locus relative to local drift when demes are at carrying capacity; the mutation rate scaled by the carrying capacity, Ku; the migration rate scaled by the carrying capacity, Km, that determines the degree of subdivision and finally, the geneomewide mutation rate relative to the baseline growth rate, $2LU = 2Lu/r_0$, that determines the hardness of selection (see Sachdeva et al. (2022)). $2LU = 0$ is equivalent to pure soft selection and $2LU = 1$ represents selection that is very hard.

We now describe our simulation setup.

5.2.4 Simulations

We run individual-based simulations only for soft selection to identify regimes in which multilocus interactions are important and also to test the validity of approximations based on effective migration rates, m_e (i.e., the effective migration approximation). However, for hard selection, we perform allele frequency simulations (that assume linkage equilibrium) since in general, multilocus interactions only have limited effects which are captured reasonably well by simple extensions of the diffusion approximation, that include m_e. We describe the allele frequency simulation below.

Under soft selection, since population size is always constant, we only follow allele frequencies, $\mathbf{p_{i,j}}$ at all L loci (i.e., $j = 1, \ldots, L$) and in all n_D demes (i.e., $i = 1, \ldots, n_D$) of the metapopulation. Whereas, under hard selection, where population size vary across patches, we follow both the size, N_i and allele frequencies $\mathbf{p_{i,j}}$ in all patches of the metapopulation.

The simulation is initialized such that the patches comprising the metapopulation are initially perfectly fit (i.e., the mutant allele is assumed to be absent so that its frequency in each patch is initially zero) and individuals then gradually accumulate deleterious mutations over time. In each generation or time step, allele frequencies (under soft and hard selection) and population size (under hard selection) undergo changes due to processes of migration, mutation, selection, and stochastic events (i.e., drift and demographic stochasticity) as described in Appendix D.0.1. The metapopulation is allowed to equilibrate and the equilibrium load and population size (under hard selection) are then computed.

The simulation was implemented in Fortran (see Olusanya et al. (2023) for code).

5.3 Soft selection

We first consider soft selection, where each subpopulation has a fixed number K of individuals, regardless of mean fitness. Assuming loci evolve independently (i.e., under linkage and identity equilibrium), the distribution $\psi[p|\overline{p}]$ of allele frequencies (at a given locus) within subpopulations at mutation-drift-selection-migration equilibrium depends only on the scaled parameters Ks, Ku, Km and the dominance coefficient h, and is given by eq. (5.4) (where $n \equiv K$), conditional on $\overline{p}$, the mean allele frequency across all subpopulations (where $\overline{p}$ can be obtained self-consistently by numerically solving $\overline{p} = \int dp\, p\, \psi[p|\overline{p}]$).

The soft selection model has been analysed in earlier studies that assume either very weak or strong selection, i.e., $Ks \ll 1$ (Whitlock, 2002), or $Ksh + Km$ at least 5 (Glémin et al., 2003; Roze, 2015). These limits are useful to consider as they yield simpler intuition and

allow for explicit analytical results: for instance, when selection on deleterious homozygote is much weaker than local drift (i.e., $Ks \ll 1$), probabilities of identity by descent involving two or more genes can be approximated by their expectation under the *neutral* island model, allowing one to solve explicitly for $\overline{p}$ in terms of (essentially) neutral higher cumulants of the allele frequency distribution (Whitlock, 2002). In this limit, selection is only efficient at the level of the population as a whole and not within local demes, some of which may be nearly fixed for the deleterious allele (if Km is small enough). By contrast, for large Ksh, deleterious alleles segregate at low frequencies within any deme and the allele frequency distribution is concentrated about $\overline{p}$. Thus, in this limit, gene flow only serves to further narrow the distribution of allele frequencies, i.e., further increase the effective size of local demes (Glémin et al., 2003; Roze, 2015). Here we focus on *moderately* deleterious alleles with $Ks \sim 1$ (which contribute the most to drift load in a single isolated deme). As we argue in Appendix D.0.2, in this regime, selection has more subtle effects, which are not captured by either intuitive description above.

Beyond these more conceptual issues, the concrete question we address here is: how much gene flow between subpopulations is required to alleviate load at a single locus under soft selection, and how does this depend on the (scaled) selection strength Ks and dominance coefficient h at the locus? We then ask: to what extent is load per locus affected by multilocus associations when multiple unlinked loci across the genome are under selection? In the next section, we build upon this to investigate similar questions for the case of hard selection, where positive feedback between increasing drift load at large numbers of loci and declining local deme sizes can drive the entire metapopulation to extinction.

To begin with, consider the case of fully isolated demes ($Km = 0$). For $Ku \ll 1$, any deme will be close to fixation for one or other allele at any locus. Under this "fixed state approximation" (see also Szép et al. (2021)), the probability that the deme is nearly fixed for the wildtype allele is proportional to: $\int_{p=0}^{1/2} dp\ p^{4Ku-1} \approx \frac{1}{4Ku}$, while the corresponding probability for the deleterious allele is proportional to: $\int_{p=1/2}^{1} dp\ e^{-2Ks}\ (1-p)^{4Ku-1} \approx \frac{e^{-2Ks}}{4Ku}$, regardless of h. Thus, for $Ku \ll 1$, the expected deleterious allele frequency is $\mathbb{E}[p] \approx \frac{e^{-2Ks}}{1+e^{-2Ks}}$, and the expected load G (scaled by the deterministic expectation $2u$) is: $\frac{G}{2u} = \frac{Ks}{2Ku}\left(\mathbb{E}[p] - (1-2h)\mathbb{E}[pq]\right) \approx \frac{Ks}{2Ku}\ \frac{e^{-2Ks}}{1+e^{-2Ks}}$, assuming $\mathbb{E}[pq] \approx 0$. It follows from this that the maximum contribution to load is from loci with $Ks \approx 0.64$, independent of h, with load per locus being $\frac{G}{2u} \approx \frac{0.139}{2Ku}$ for this value of Ks: thus drift can inflate the load associated with moderately selected loci by a factor of several hundreds or thousands in an isolated population (see also Kondrashov (1995)).

Let us now consider how gene flow changes these simple expectations: fig. 5.1a shows load per locus (scaled by $2u$) as a function of Km for $Ks = 0.64$ for various values of h; the

inset shows the expected deleterious allele frequency vs. Km. Symbols depict predictions of the full diffusion approximation (eqs. (5.4)-(5.6), $j = 1$). In addition, we also show predictions of a simpler 'moderate selection' approximation (lines), details of which are outlined in Appendix D.0.2.

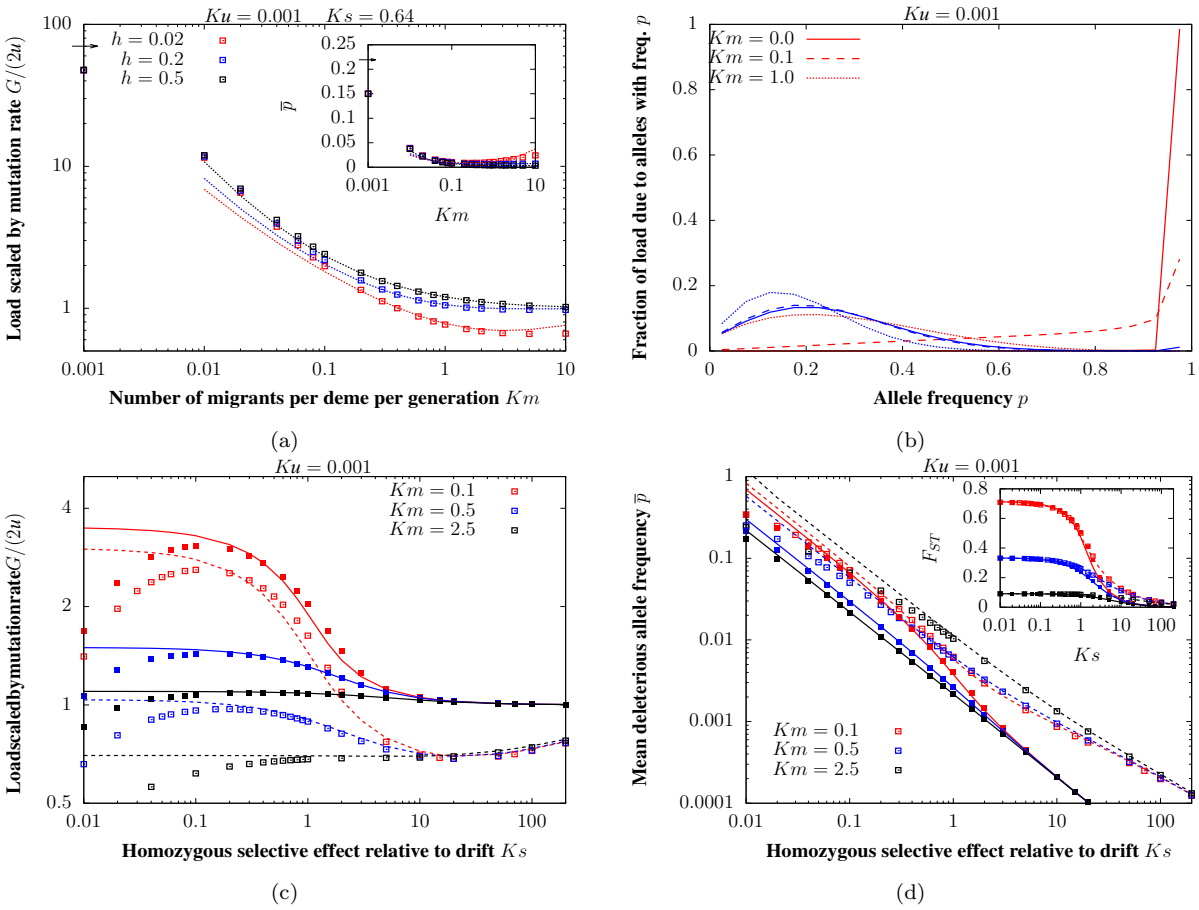


Figure 5.1: Mutation-selection-drift-migration equilibrium at a single locus under the infinite-island model with soft selection. (a) Main plot and inset show respectively the expected per locus load G (scaled by $2u$) and the mean deleterious allele frequency $\overline{p}$ vs. Km, the number of migrants per deme per generation, for various values of h, for $Ks = 0.64$ (which is the Ks value for which load is maximum in an isolated population). Symbols depict results of the diffusion approximation (eq. (5.6), $j = 1$), while lines represent predictions of the 'moderate selection' approximation (eq. (D.7a), Appendix D.0.2). Arrows on the y-axis represent the corresponding diffusion predictions for $Km = 0$. (b) The fraction of total load that is due to alleles with frequency in the interval between p and $p + \Delta p$ (with $\Delta p = 0.05$) for various values of Km (solid, dashed and dotted lines), for weakly selected ($Ks = 0.64$; red) and strongly selected ($Ks = 6.4$; blue) loci. Predictions are based on the diffusion approximation, i.e., obtained by integrating over the allele frequency distribution in eq. (5.4). (c) Expected per locus load G (scaled by $2u$) vs. Ks, the homozygous selective effect scaled by drift, for various values of Km (different colours) for $h = 0.5$ (filled symbols) and $h = 0.02$ (open symbols). (d) Mean deleterious allele frequency $\overline{p}$ (main plot) and the expected F_{ST} at the selected locus (inset) vs. Ks, for various Km (different colours) for $h = 0.5$ (filled symbols) and $h = 0.02$ (open symbols). In both (c) and (d), symbols depict results of the diffusion approximation (eq. (5.6)) and lines represent the approximate 'moderate selection' predictions. All plots are with $Ku = 0.001$.

Figure 5.1a shows that a very low level of gene flow ($Km \sim 0.1$) is enough to dramatically reduce drift load, regardless of dominance. For instance, $G/(2u)$ falls from 140 at $Km = 0$ to $4 - 5$ at $Km = 0.1$ for both nearly recessive ($h = 0.02$) and co-dominant ($h = 0.5$)

deleterious alleles, for $Ku = 0.001$. In both cases, the expected allele frequency $\overline{p}$ declines from 0.22 at $Km = 0$ to ~ 0.01 at $Km = 0.1$ (inset), while heterozygosity hardly increases (from ~ 0.002 to $0.003 - 0.004$). Thus the reduction in load at low levels of migration is almost entirely due to the decline in the number of fixed deleterious alleles, with little to no change in the number of segregating alleles.

A further increase in migration further reduces load, though in the case of recessive alleles, this is partially offset by an increase in heterozygosity (resulting in less efficient purging). Thus, for recessive alleles, load is minimum at $Km \approx 4$ and rises again as Km increases further, approaching the level expected in a panmictic population at large Km. For the parameters shown in fig. 5.1a, the reduction in load due to purging at intermediate Km is rather modest and hardly visible on the scale of the plot; purging has a more substantial effect when $u/(hs)$ and Ks are smaller, and provided $h < 1/4$ (Whitlock, 2002).

Increasing gene flow also shifts the frequency spectrum of alleles that contribute to load – for $Ks \lesssim 0.1$, the predominant contribution at $Km = 0.1$ is from fixed or nearly fixed alleles; however, one migrant per deme per generation ($Km = 1$) is already enough to prevent fixation of deleterious alleles, so that load is entirely due to alleles segregating at intermediate frequencies at this higher migration level (dashed vs. dotted red curves in fig. 5.1b). As expected, gene flow has a much weaker effect at strongly selected loci: for $Ks = 6.4$, load is entirely due to segregating alleles regardless of Km (blue curves). Thus, increasing gene flow reduces load only very weakly at strongly selected loci (fig. 5.1c), and may even be detrimental if it hinders purging (in the case of recessive deleterious alleles).

Moreover, even low levels of gene flow tend to "even out" the contributions of alleles with different selection coefficients to load, so that moderately deleterious alleles no longer contribute disproportionately (as in isolated populations). For instance, with $Km = 0.1$, load is maximum for $Ks \approx 0.1$ (regardless of h), for which it is only about ~ 3 times larger than the deterministic expectation (fig. 5.1c); by contrast, in isolated populations, it may be several hundred or thousand times larger (see above). With modest levels of gene flow ($Km = 0.5$), there is an even weaker inflation of load for intermediate Ks, while $\gtrsim 2$ migrants per generation are enough for load to be more or less independent of Ks as in a large, essentially panmictic population.

One can ask further: are the processes underlying this dramatic reduction in load (even with low levels of gene flow) qualitatively different for alleles with different selective effects? More concretely, expressing load as $\frac{G}{2u} = \frac{Ks}{2Ku}\left[\overline{p} - (1-2h)\ \overline{pq}\right] \equiv \frac{Ks}{2Ku}\left[\overline{p} - (1-2h)\overline{p}\ \overline{q}(1-F_{ST})\right]$, it follows that for recessive alleles ($h < 1/2$), load declines if the mean deleterious allele frequency $\overline{p}$ and/or F_{ST} at the selected locus decrease. A decline in F_{ST} (for a fixed $\overline{p}$), in turn, reflects a change in the allele frequency distribution from a more U-shaped (wherein a fraction $\overline{p}$ of demes are nearly fixed for the deleterious allele) to a more unimodal or

concentrated distribution (wherein the deleterious allele segregates at a low frequency close to $\overline{p}$ in almost all demes). Thus, in effect, we ask: to what extent is the reduction in the mean allele frequency across the entire population (as measured by $\overline{p}$) and/or the reduction in probability of local fixation of the deleterious alleles (as measured by F_{ST}) sensitive to the selection coefficient of the deleterious allele?

Figure 5.1d shows $\overline{p}$ vs. Ks (main plot) and F_{ST} vs. Ks (inset), for various values of Km, for co-dominant and nearly recessive alleles (filled vs. open symbols). Note that while selection is quite efficient in reducing the mean number of deleterious alleles already for $Ks \gtrsim 0.05$ (say, under low gene flow, i.e., $Km = 0.1$), it has little effect on F_{ST} (i.e., on the probability of local fixation within demes, given $\overline{p}$) unless $Ks \gtrsim 0.5$. For example, F_{ST} is reduced by only about 15% with respect to its neutral value for alleles with $Ks = 0.5$, but by $> 50\%$ for $Ks = 2$, for both values of h. Thus, selection for increased heterozygosity within demes (as reflected in a decrease in F_{ST}) does not markedly influence the evolutionary dynamics of alleles with $Ks \lesssim 1$ (which suffer the most severe inflation of load in isolated populations) but can be important for moderately deleterious alleles with $1 \lesssim Ks \lesssim 5$ (which may still contribute substantially to load).

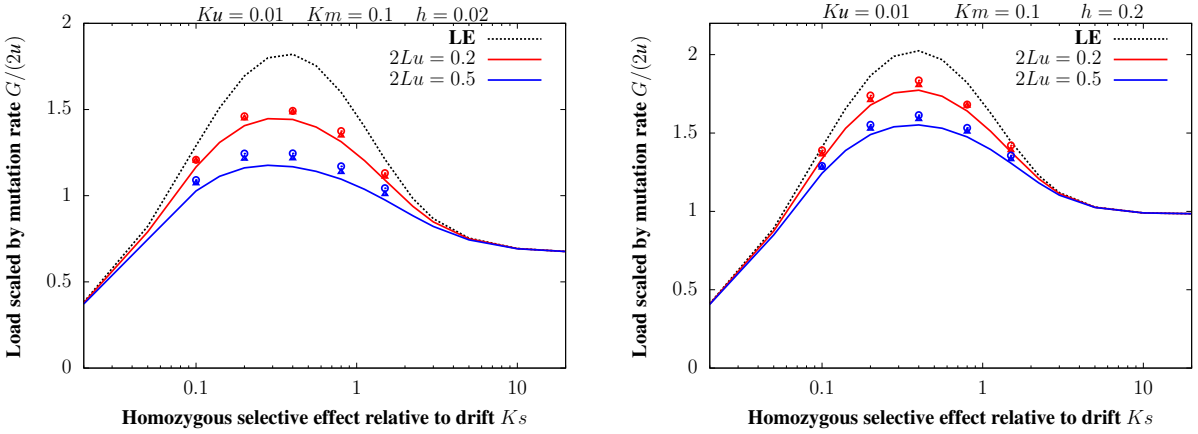
In Appendix D.0.2, we introduce a 'moderate selection' approximation, which allows F_{ST} to be affected by selection, but assumes that the relationship between (appropriately scaled) higher cumulants of the allele frequency distribution and F_{ST} is the same as under the neutral island model. The predictions of this approximation are shown by lines in fig. 5.1a, 5.1c and 5.1d, and appear to match the full diffusion (symbols) quite well for $0.5 \lesssim Ks \lesssim 5$. This suggests that moderate selection essentially changes pairwise coalescence times (to different extents within and between demes), without appreciably affecting other statistical properties of the genealogy, e.g., the degree to which branching is skewed or asymmetric, at loci subject to purifying selection (see also Appendix D.0.2).

Effect of multilocus interactions on load. So far, we have considered the equilibrium load at a *single* locus, neglecting linkage and identity disequilibrium between deleterious alleles segregating across multiple loci in the genome. These may be substantial if, for instance, different subpopulations are nearly fixed for partially recessive deleterious alleles at very many different loci, so that $F1$ offspring of migrants and residents (and their descendants) have higher fitness than residents, giving rise to heterosis. In such a scenario, the extent of gene flow at any one locus depends not only on Km, the average number of migrants exchanged between demes, but also on their relative fitness, which may differ significantly from that of residents if deleterious alleles segregate at multiple loci in the population as a whole (i.e., if $2Lu$ is large and Ks small), allele frequencies differ markedly between subpopulations (migration is low or F_{ST} high), and alleles are recessive (h is small).

More specifically, heterosis implies that deleterious alleles that enter a given deme on a migrant genetic background are more likely to be transmitted to the next generations than deleterious alleles on resident backgrounds, resulting in an *effective* migration rate that is higher than the raw migration rate (see also Ingvarsson and Whitlock (2000)). For a genome with L equal-effect unlinked loci with selective effect s and dominance h per locus, and assuming weak selective effects ($s \ll 1$), the effective migration rate at any locus is shown by Zwaenepoel et al. (2023) to be approximately:

$$\frac{m_e}{m} \approx e^{2Ls(1-2h)\overline{p}\,\overline{q}F_{ST}} \tag{5.7}$$

Following Sachdeva (2022) and Zwaenepoel et al. (2023), we can incorporate the effects of multi-locus heterosis by assuming that allele frequencies at any locus follow the equilibrium distributon in eq. (5.4), but with the raw migration rate replaced by an effective migration rate which itself depends on the expected allele frequency $\overline{p}$ and the expected heterozyosity $\mathbb{E}[pq]$ within demes (or equivalently, $\overline{p}$ and F_{ST}). As before, this allows for a numerical solution, yielding the theoretical predictions (solid lines) in fig. 5.2.



(a) (b)

Figure 5.2: Effects of selective interference on load. Load per locus (scaled by $2u$) vs. Ks in a population with recurrent deleterious mutations at L biallelic equal-effect loci with (a) $h = 0.02$ and (b) $h = 0.2$, for $Km = 0.1$. Symbols depict results of individual-based simulations, which are carried out for two values of total mutation rate $2Lu$ and two values of L for each total mutation rate: $L = 1000$, $u = 0.0001$ (red circles) and $L = 2000$, $u = 0.00005$ (red triangles) both of which correspond to $2Lu = 0.2$; $L = 2500$, $u = 0.0001$ (blue circles) and $L = 5000$, $u = 0.00005$ (blue triangles) which correspond to $2Lu = 0.5$. For each value of L and u, the carrying capacity K and migration rate m are chosen such that scaled parameters are $Ku = 0.01$ and $Km = 0.1$; for a given K, the scaled selective effect Ks is varied by varying s. Dashed curves depict single-locus predictions in the absence of multilocus interactions (obtained using eq. (5.6)). Solid curves show predictions that account for interference via effective migration rates; these are independent of L for a fixed Lu, u/s, Ks and Km. The reduction in load per locus due to selective interference between loci is most significant for intermediate Ks, small h and large values of Lu.

Figure 5.2 shows load per locus scaled by $2u$ as a function of Ks for low migration ($Km = 0.1$), for two values of the total mutation rate ($2Lu = 0.2$ in red and $2Lu = 0.5$ in blue) and two dominance coefficients ($h = 0.02$ in fig. 5.2a and $h = 0.2$ in fig. 5.2b). For

each value of $2Lu$, we further simulate with two values of L, scaling down s, u and m, and scaling up K as we increase L, so that Ks, Km, Ku and Ls remain constant. Symbols depict results of individual-based simulations for a metapopulation with 100 demes; dashed curves show single-locus predictions (obtained using eq. (5.6)) that do not account for multilocus heterosis; solid curves represent predictions that account for interference via effective migration rates: note that these depend on L only via the combination Ls (or alternatively, Lu for a given u/s). As expected, there is better agreement between simulations and theory for smaller values of s (or alternatively, larger L), for a given total mutation rate, as the expression for effective migration rate in eq. (5.7) becomes more accurate as $s \to 0$. As can be seen in fig. 5.2, load per locus is most strongly reduced by multilocus heterosis when $2Lu$ is large, and for loci with small h and intermediate Ks. Moreover, the effects of multilocus heterosis become weaker with increasing migration (which corresponds to lower F_{ST}) which reduces allele frequency differences across demes and consequently the extent of heterosis (see also Roze, 2015).

5.4 Hard selection

In our soft selection analysis, we showed that the load in a metapopulation can be quite significant (l.h.s. of fig. 5.1a) when migration is limited, in particular, when $Km < 1$). We also demonstrated that a minimal amount of gene flow (e.g., $Km = 0.1$ in fig. 5.1a) is enough to purge or alleviate such load (by about 96% in fig. 5.1a) irrespective of the value of h. This holds true even for moderately deleterious alleles that would otherwise contribute disproportionately to load. This purging advantage however decreases as Km becomes very large (for example, $Km > 4$ in fig. 5.1a). Here, we would like to understand how this changes under hard selection where there is a feedback between population size and genetic load so that population sizes are not fixed at K but decline with increasing load, placing populations burdened by high load at an increased risk of extinction.

More concretely, we aim to understand the impact of gene flow on drift load and purging in the context of hard selection and what the attendant consequences of these are for metapopulation outcomes, for example, on critical extinction thresholds. We also aim to understand the role of dominance and how critical thresholds are influenced by the strength of eco-evo feedback. The strength of such feedback depends largely on $2LU = 2L(u/r_0)$ which is the expected load (in the absence of drift) divided by the baseline growth rate. If $2LU = 0.5$ for example, this indicates a 50% reduction in growth rate due to genetic load (assuming load is primarily deterministic), we therefore take $2LU$ as a measure of the hardness of selection.

For simplicity, our analysis will focus on scenarios where loci have equal effects; more complicated but realistic scenarios involving a distribution of fitness effects are explored

in Appendix D.0.7. As introduced in section 5.2.3, the scaled effect per locus is denoted by Ks and gives the strength of selection at a deleterious homozygous locus relative to local drift, when population sizes are at carrying capacity K. However in general, under hard selection and depending on parameter values, population sizes will be less than K, so that drift will usually be stronger relative to selection, than indicated by the value of Ks (or Ksh).

The rest of this section is organised as follows, first we will identify critical migration thresholds for extinction by exploring how metapopulation outcomes as measured by the mean equilibrium average population size per island scaled by the carrying capacity K (i.e., N) depend on the magnitude of gene flow (as measured by Km). We will then test the validity of the semi-deterministic approximation by contrasting semi-deterministic results of N, (under different Km values) with outcomes derived from our simulations, distinguishing between two simulation runs, each with different strengths of demographic fluctuations, while keeping other (scaled) parameters, Ks, Km, Ku fixed. We do this to see how well the results from our simulations converge to the semideterministic results as demographic fluctuations become weaker (i.e., as r_0K increases, since the strength of demographic fluctuation scales as $1/(r_0K)$); our expectation is that the semi-deterministic approximation should be good enough in the limit $r_0K \to \infty$. We will later build on this in a more general way to explore how the thresholds depend on homozygous selective effects relative to drift, dominance and the 'hardness' of selection. Finally, we will explore parameter regimes which allow for the metapopulation to persist (i.e., avoid extinction) and explore how equilibrium load and population size per deme in this regime depend on various parameters.

5.4.1 Extinction thresholds with low gene flow

Figure 5.3a shows the mean population size across the metapopulation vs. Km (the average number of migrants exchanged between any deme and the metapopulation, when all demes are at carrying capacity). This is shown for $Ks = 1$ and $Ks = 10$ (fig. 5.3a and 5.3b respectively) and for both recessive and additive alleles assuming fairly hard selection (i.e., $2LU = 0.6$). The solid lines are results from the semi-deterministic approximation and the symbols connected by dashed lines are results from simulations. Keeping all other scaled parameters fixed, simulations are run for various values of r_0K (where r_0 is fixed at 0.1 and K is varied) to see the impact of the strength of demographic stochasticity. Circular symbols denote simulations run with a higher K (here, $K = 3000$) and triangles denote simulations run with a lower K (here, $K = 500$). Note that to keep Ks fixed, increasing (decreasing) K would mean reducing (increasing) s.

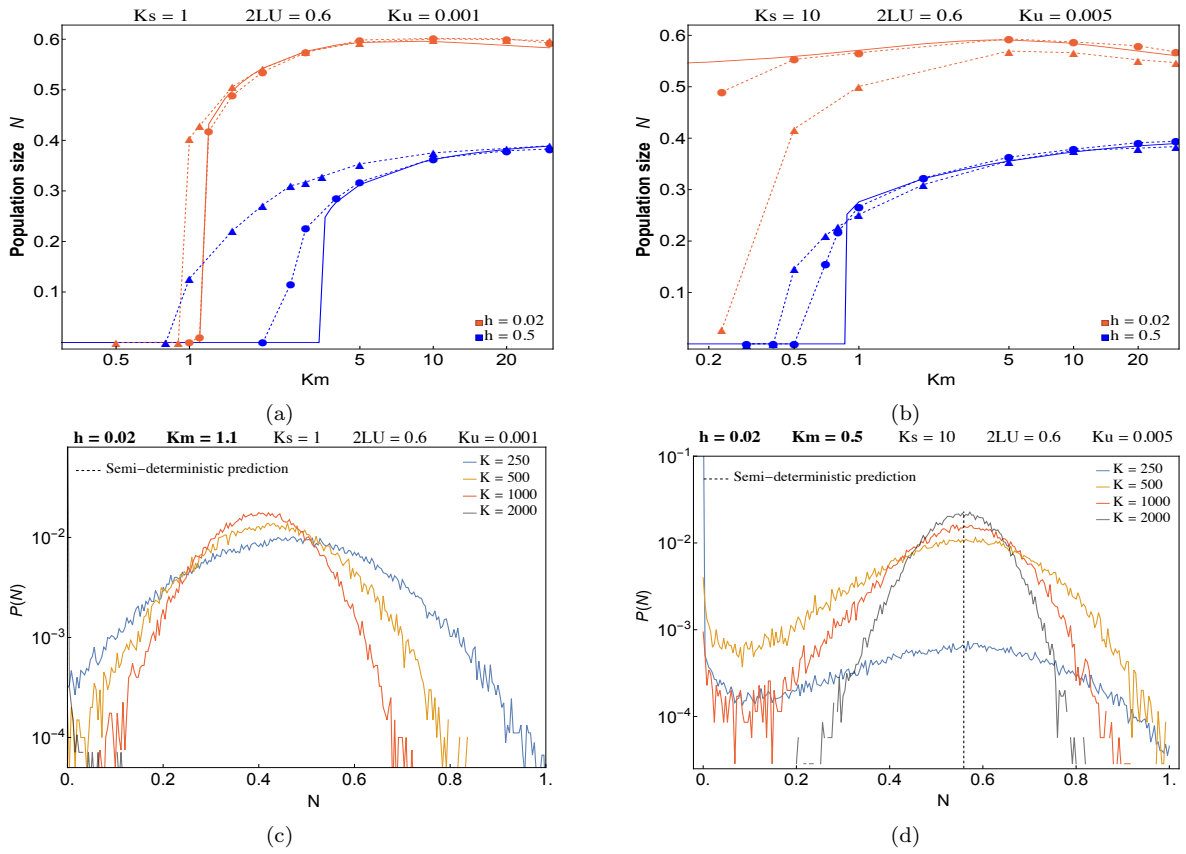


Figure 5.3: Mean population size across the metapopulation plotted against Km for (a.) $Ks = 1$ and (b.) $Ks = 10$ and for different dominance levels. Solid lines are results from the semi-deterministic approximation. Symbols connected by dashed lines represent simulation results (using allele frequency simulations) - triangles and circles represent simulation results with carrying capacity per island, K equal to 500 and 3000 respectively. (c.)-(d.) Equilibrium distribution of population sizes (with $h = 0.02$) for different values of K and at low Km. (c.) is a plot for $Ks = 1$ and with $Km = 1.1$ and (d.) is a plot for $Ks = 10$ and with $Km = 0.5$. The black vertical dashed line represents the semi-deterministic prediction. Simulation results were obtained using 100 demes and all plots are obtained using $r_0 = 0.1$.

Let us first concentrate on the analytical semi-deterministic prediction (solid lines). Figure 5.3a and 5.3b show that there exist a critical Km threshold (which we will call Km_c) below which the metapopulation collapses (i.e., below which $N = 0$) and this threshold is highest for additive alleles (blue vs red colors). We call this a critical threshold because it represents a tipping point in the fate of the metapopulation wherein a slight variation (for example the introduction of an additional migrant) can change the expected metapopulation outcome. The extinction of the metapopulation occurs at very low Km values because such low values typically imply very little gene flow among the different patches. As such, these isolated patches rapidly dwindle in size (due to inbreeding) at a faster rate than can be rescued by migration, thus leading to their extinction and the collapse of the metapopulation. However, for $Ks = 1$ fig. 5.3a, the situation is less gloomy for patches with recessive alleles (red solid line) because they are are able to purge some of their load, thereby requiring less migration to escape extinction; we will come to this later in more detail.

An important aspect to focus on in figs. 5.3a and 5.3b is the disparity between the critical threshold, Km_c, observed in simulations compared to our analytical approximation (symbols vs solid lines) – Km_c from simulations are lower compared to semi-deterministic outcomes. However, it is worth noting that these thresholds approach the semi-deterministic expectation as the parameter K (and consequently r_0K (for fixed r_0)) increases, suggesting that our analytical approximation remains valid when demographic fluctuations become less pronounced.

Interestingly, we observe two qualitatively different behaviours for $Ks = 1$ (fig. 5.3a) and $Ks = 10$ (fig. 5.3b). In the case of $Ks = 1$ (fig. 5.3a), we observe that for low Km, near the semi-deterministic threshold, the mean population size, N is higher and the critical migration threshold lower when K is smaller ($K = 500$ here; represented by triangles with dashed lines) compared to when we have a higher value of K ($K = 3000$ - represented by circles with dashed lines). To illustrate this point, fig. 5.3c shows the equilibrium distribution of deme sizes for different values of K just at the semi-deterministic threshold ($Km = 1.1$). At this value of Km, we see that large-sized patches ($K = 2000$; black line) are extinct whereas patches with smaller sizes (e.g., $K = 250$; blue line) are still able to avoid extinction. This observation is puzzling as it indicates that smaller patches are more stable, requiring fewer migrants per generation (i.e., lower Km) to prevent extinction (all other parameters in particular Ks, the strength of selection relative to drift, remaining constant). Usually, we would expect an opposite outcome since larger patches would be imagined to be less prone to the effects of demographic stochasticity (hence more stable than small-sized patches). This unusual result however highlights the role of the demographic effect of migration in small-sized patches where a few migrants (say, 1 per generation) can go a long way in preventing extinction (see also figs. 4.3b and 4.3d).

In contrast, for larger Ks, we observe the more expected behaviour where mean sizes are higher for higher K (compare red triangles and circles). Figure 5.3d further illustrates this point, where at low K (e.g. $K = 250$ (blue line)) we have a bimodal distribution of population sizes with a fraction of demes close to extinction and the remaining fraction peaked at a stable non-zero N. With increasing K however, there is a reduction in the weight of the distribution near extinction due to a decline in demographic fluctuations which then further narrows the distribution around N. With $K = 2000$, we obtain a unimodal distribution where the population peaks at $N = 0.55$ which coincides with the semi-deterministic approximation (dashed vertical black line). Similar figures are shown in Appendix D.0.3 for $h = 0.5$ (fig. D.1a and D.1b). Hereon we will only show results from the semideterministic approximation with the understanding that these should be accurate if r_0K is sufficiently large.

Now we investigate what determines the thresholds discussed above. In particular, we explore how the selection strength per locus influences the critical migration rates, Km_c,

below which the entire metapopulation collapses, starting from a state in which it is stable, i.e., where all demes are at the stable population equilibrium. We then build on this to understand the role of dominance and the sensitivity of the thresholds to the hardness of selection. It is important to note that the corresponding thresholds for the population to grow starting from a state where there are only a small number of individuals could be much more stringent because small populations suffer from genetic Allee effects (see also Sachdeva et al., 2022).

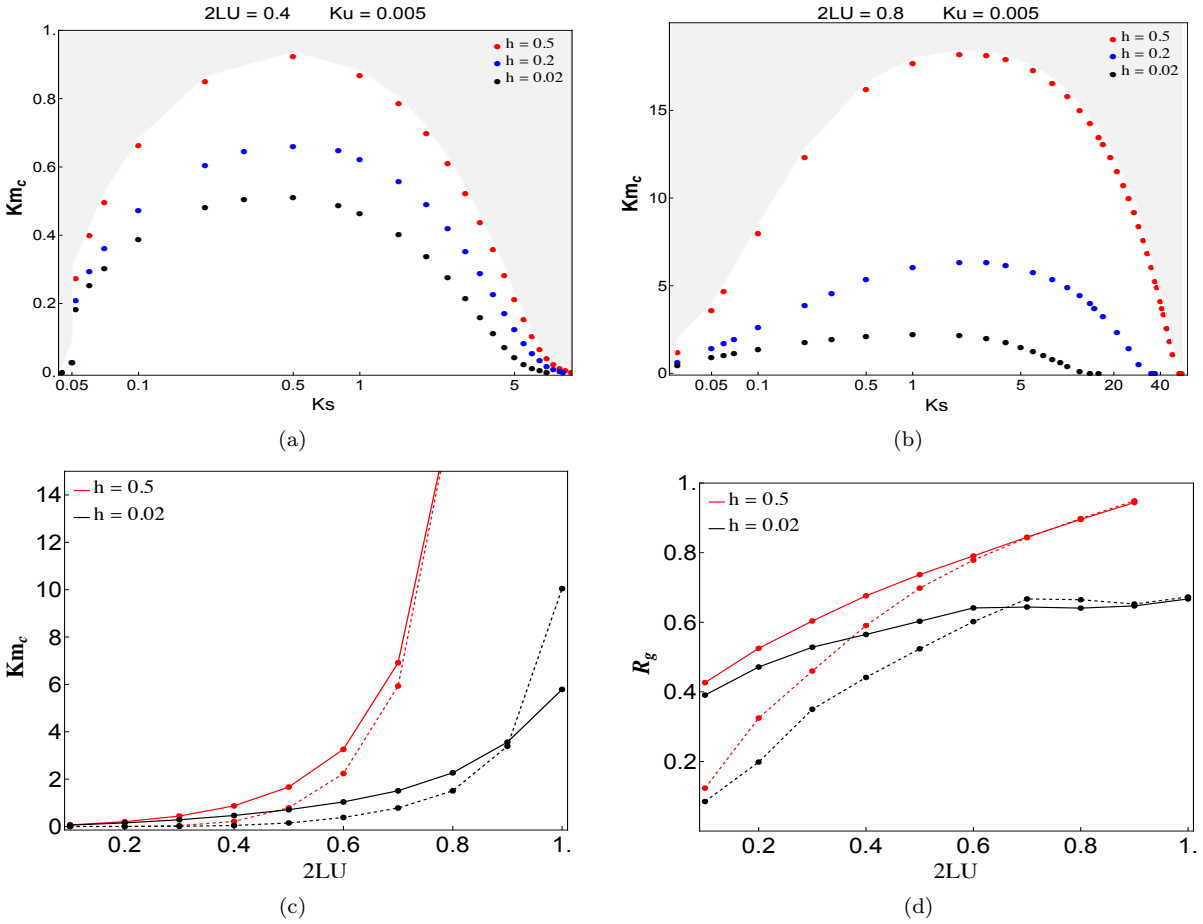


Figure 5.4: (a.)-(b.) Critical migration thresholds (below which the metapopulation goes extinct) as a function of Ks as obtained from the semi-deterministic approximation. Figure 5.4 (a) shows plot with moderately hard selection i.e, $2LU = 0.4$ and 5.4 (b) shows plot with much harder selection, $2LU = 0.8$. (c.) Critical migration threshold below which the metapopulation collapses as a function of the hardness of selection $2LU$. (d.) Load at Km_c for different $2LU$. Solid and dashed lines in 5.4 (c.) and figure 5.4 (d.) represent results with $Ks = 1$ and $Ks = 5$ respectively.

For simplicity, we begin by assuming equal effect loci; scenarios involving a distribution of fitness effects are explored in Appendix D.0.7. Figure 5.4a and 5.4b show that there is a non-monotonic relationship between Km_c (the critical migration threshold required to prevent a complete meltdown of the metapopulation) and the scaled strength of selection, Ks. In particular, when Ks is low, the fitness differences between genotypes in each subpopulation is minimal so that low rates of migration are sufficient to maintain diversity and prevent extinction. As Ks increases from low values, the fitness differences between

genotypes become more pronounced and the fitness cost of carrying harmful mutations becomes higher (as was also seen in the soft selection results; figs. 5.2a and 5.2b). Consequently, a higher migration is necessary to counteract this accumulation of load and prevent extinction leading to a rise in Km_c until a maximum value at intermediate Ks. Beyond this peak, as Ks increases further, selection becomes strong enough to eliminate individuals with high load so that we again require less migration to prevent extinction.

When alleles are at least (partially) recessive, it is much easier to maintain the metapopulation as load in each subpopulation is lower, due to more efficient purging of deleterious recessives alleles. Furthermore, we see that when selection is moderately hard (i.e., $2LU = 0.4$ in fig. 5.4a), critical migration thresholds are much lower compared to when selection is much harder emphasizing the impact of the strength of the coupling between population size and load. In addition, we see that the metapopulation never collapses beyond $Km_c = 1$ (corresponding here to $NKm \sim 1$ i.e., one migrant per generation). On the other hand, with harder selection (i.e., $2LU = 0.8$ in fig. 5.4b), we require very high migration between subpopulations to prevent the metapopulation from going extinct. In essence, we see that the conditions for metapopulation persistence for both recessive and additive alleles are much more stringent as selection gets harder (compare gray region in 5.4a and 5.4b). It is important to emphasize that these results however assume equal effect loci and equal forward and backward mutation rates between deleterious and wild-type alleles. We explore the role of asymmetric mutation rates in Appendix D.0.6.

The relatively small difference in load per locus between co-dominant and recessive alleles that we observe at a given level of migration under soft selection (fig. 5.1c) can translate into rather different critical migration thresholds (depending on whether load is primarily due to recessive or additive alleles) for metapopulation persistence under hard selection. This is especially marked when selection is harder (fig. 5.4b) as a small increase in load (e.g., due to a small decline in gene flow) can set in motion a very strong positive feedback between increasing load and declining population size (which results in stronger drift), culminating in extinction.

To further quantify the role of the hardness of selection, we look at the dependence of the critical migration threshold on the type of selection going from very soft to very hard selection (i.e., as $2LU$ increases). We see clearly from fig. 5.4c that the critical migration threshold above which populations go extinct due to high load (fig. 5.4d) is always higher when selection is hard. This holds true for both additive and recessive alleles with the threshold for recessive alleles lower than that for additive alleles as drift always increases load with additive alleles. In addition, when selection is soft and per locus strength of selection is stronger (compare dashed lines corresponding to $Ks = 5$ with solid lines corresponding to $Ks = 1$), populations are a bit more stable, having a lower load (lhs of fig. 5.4d) and hence, requiring a lower Km_c (lhs of fig. 5.4c) to prevent extinction.

Finally, for a metapopulation to persist (i.e., for $N = 1 - \mathbb{E}[R_g] > 0$), fig. D.6 (in appendix D.0.8) show that we need a reasonable amount of gene flow ($Km > 0.2$) and under this level of gene flow, moderately deleterious alleles contribute most to load. Albeit, this is numerically only a very modest ($5 - 10\%$) effect (see also fig. 5.2). This makes sense due to the simple fact that large effect alleles, though having the potential to drastically increase load, are less likely to fix. Small effect alleles on the other hand, which are more likely to fix, have less effect on load when they do fix.

5.4.2 Load and population size under high gene flow

Having analysed the low Km behaviour, we now consider the behaviour of large Km. Going back to figs. 5.3a and 5.3b, we observe that above the critical migration threshold, Km_c and in particular for large Km, our simulation results depend only weakly on K. For example, for $Ks = 1$ (fig. 5.3a) and with $h = 0.02$ and $Km = 2$, $K = 500$ is already sufficient for our simulations to match the semi-deterministic approximation. To further investigate this, we explore the behaviour of the equilibrium distribution of deme sizes for different values of K and for $h = 0.02$. We do this with $Ks = 1$ (and $Km = 2.0$; fig. D.2a in appendix D.0.4) as well as with $Ks = 10$ (and $Km = 5.0$; fig. D.2b in appendix D.0.4). The peaking of these distributions at the semideterministic expectation for $K \sim 250 - 500$ further confirms the weak dependence of our simulation results on K.

In addition to the above, beyond Km_c, there exists only a weak dependence of population size on Km, especially when alleles are recessive (fig. 5.3a), with load close to the deterministic prediction above Km_c (fig. 5.5b). This is contrary to previous findings (Whitlock, 2002) where low gene flow is thought to substantially reduce load, due to increased expression and consequently stronger purging of deleterious alleles in more isolated populations. To better understand our results, we look at how the equilibrium allele frequency in the metapopulation (relative to that in an undivided population) as well as how the equilibrium mean load (relative to the deterministic load) depend on the level of gene flow (i.e., Km) as shown in figs. 5.5a and 5.5b respectively. The solid lines are results from the semi-deterministic approximation and colored circles represent simulation results. Here, we have obtained the deterministic expectation of allele frequency, p_{det}, at any given locus and load, $R_{g,det}$, in an undivided population by respectively solving eq. (5.1) at equilibrium and plugging the result into $R_{g,i}$ in eq. (5.2) to get $R_{g,det}$; this is as opposed to assuming $p_{det} \sim u/hs$ and $R_{g,det} \sim 2Lu$ which both overestimate the expectation when alleles are nearly fully recessive.

We see from fig. 5.5a that when alleles are (partially) recessive, there is a decrease in the frequency of the deleterious allele (as migration decreases) due to purging. However, fig. 5.5b shows that this purging effect is not strong enough to counter the negative effect

of inbreeding as load on the whole increases[2] with decreasing migration (solid lines and circles), hence the low population size observed at low Km in fig. 5.3a. Similar analysis for strongly selected alleles ($Ks = 10$) are shown in Appendix D.0.5.

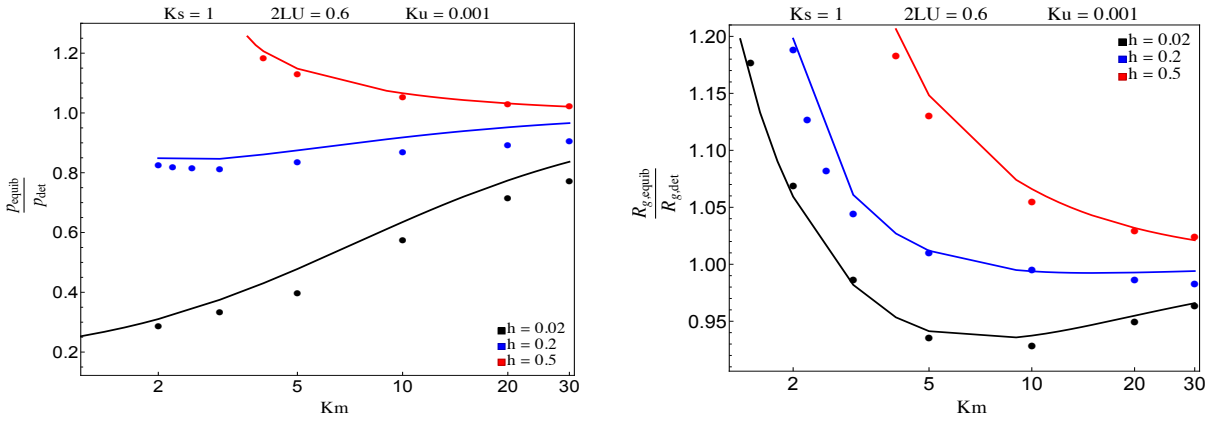


Figure 5.5: (a.) and (b.) are respectively the equilibrium allele frequency and load (relative to that in an undivided population) plotted against Km. Filled circles show simulation results (using allele frequency simulations) and solid lines show results from our semi-deterministic approximation. Simulations are run with 100 demes, with $K = 3000$, $L = 90000$ and $r_0 = 0.1$.

5.5 Discussion

In this study, we have explored the varied roles of gene flow, selective effects and dominance on load and extinction dynamics in a metapopulation, explicitly capturing the interaction between population size (in each patch) and allele frequencies at multiple loci. In particular, we distinguished between two models of selection: a soft selection model, which holds true when the total deleterious mutation rate in a population is much less than the baseline growth rate (i.e., $2LU \ll 1$) and size of each subpopulation is assumed to always be at carrying capacity, and a hard selection model (with $2LU \sim 1$), that captures the feedback between population size and allele frequency. Our results provide useful insights into the intricate relationship between genetic diversity, eco-evolutionary processes and the long-term persistence of populations, thus contributing to our understanding of biodiversity and conservation in fragmented landscapes.

Under the soft selection model, we showed that independent of the dominance of deleterious alleles, very little migration (as little as one migrant per approximately ten generations, $Km = 0.1$) between a set of interconnected patches is enough to reduce the load due to slightly deleterious mutations (i.e., those with $Ks < 1$) by a factor of 100, bringing it close to the deterministic expectation. We show that this reduction happens due to decrease in the fixation probability of deleterious alleles. With one migrant per generation ($Km = 1$) and for partially recessive mutations, load is reduced below this deterministic

[2]We see a slight non-monotonic behaviour with $h = 0.02$ but this is essentially negligible.

limit and is only due to alleles segregating at intermediate frequencies. Our findings are consistent with those of Whitlock (2002) who demonstrated that at an intermediate level of migration, characterized by medium variance among local populations, load would be lowest under soft selection when mutations are partially recessive and exhibit mildly deleterious effects. This is also corroborated by Roze and Rousset (2004), who similarly observed minimum load at intermediate migration rates due to the purging of weakly deleterious and partially recessive mutations (see also Zhou and Pannell (2010)).

It has long been recognized that the reduction in the mean population fitness due to weakly deleterious mutatations ($Ks \sim 1$) reaching high frequencies or even fixation just by chance greatly exceeds the reduction due to strongly deleterious mutations that are efficiently kept at deterministic mutation-selection balance frequencies (Kimura et al., 1963; Kondrashov, 1995). Our result suggests that as long as soft selection operates, in order for drift load to be a real issue, populations have to be very strongly fragmented ($Km < 0.1$).

Soft selection models are interesting for two reasons. Firstly, some fitness components affect population size more than others (adult viability versus male mating success, say). Secondly, the estimates of load obtained under the soft selection model can be an upper bound to the load in real populations, answering the following question: if the effective population size stays the same, what fraction of offspring will fail to survive under the predicted burden of deleterious mutations?

In reality however, the accumulation of deleterious mutations is likely to decrease the size of the population, or even lead to its extinction (Kondrashov, 1995). Soft selection models do not take into account the complete positive feedback loop when the population size decreases due to genetic load, which in turn leads to further increase in genetic load. Therefore, extinction due to drift load is not possible under soft selection (Charlesworth 2013; Keightley and Eyre-Walker 2010), and we must explicitly model hard selection (where population size declines with increasing load) to investigate how much gene flow is required to prevent metapopulation collapse. Such a feedback loop has been studied in the mutational meltdown literature (e.g., Lynch et al. 1995a,b). In the second part of the paper, we therefore extended our results by considering hard selection.

Under hard selection (i.e., when the intrinsic growth rate is comparable to the total deleterious mutation rate) and when local deme sizes (and hence typical values of Ks) are small, we find that much more gene flow is required to ensure persistence. In this case, one may need as many as ~ 2 and 5 migrants per generation for recessive and additive alleles respectively to ensure metapopulation persistence. These thresholds are highly sensitive to $2LU$ (a proxy for the hardness of selection) with higher values of $2LU$ necessitating higher thresholds. Such higher migration thresholds, Km_c, are typical of metapopulations

with small local population sizes as dynamics in these populations are largely driven by stochastic events (drift and demographic stochasticity) resulting in a positive feedback between population size and load (via allele frequency changes at different loci) which exacerbates extinction risk. What insights might we glean from these high threshold values? These underscore the importance of preserving connectivity (and hence genetic diversity) within metapopulations and the necessity of increased conservation measures to ensure/maintain the viability of such populations.

A second interesting result we found was that the range of Km_c below which extinction is possible is highest for intermediate values of Ks (~ 1). Moreover, with much harder selection, the condition for the metapopulation to persist becomes more restrictive as total load (due to a large number of loci) is higher in this case and hence a higher rate of migration is required to counteract the negative effect of deleterious mutation accumulation.

Bimodal distributions (see figs. 5.3c and 5.3d) where some demes teeter on the brink of extinction while others thrive with larger populations often emerge naturally close to critical migration thresholds even in the absence of heterogeneity (i.e., when patch qualities and carrying capacities are uniform) due to the stochasticity inherent in our model. Such bimodality can serve as an important signal of impending population collapse, where minor variations in migration rate or external factors can push some demes past the point of no return, causing local extinctions and possibly cascading towards a collapse of the metapopulation (Scheffer et al., 2001; Drake and Griffen, 2010).

Overall, we identified several parameters that govern the fate of a metapopulation. One such key parameter is the genome-wide mutation rate scaled by the intrinsic growth rate, $2Lu/r_0$ which is a measure of the hardness of selection. How might one estimate such a parameter? To do this, we would need to know the total deleterious mutation rate, $2Lu$ in the population (some estimates of which exist in the literature) as well as the intrinsic growth rate r_0, which is a much harder parameter to estimate.

Our findings, while intuitive, often become obscured by the prevailing confusion surrounding the concept of "hard selection". This term is invoked in two distinct classes of eco-evo models. The first category of models meticulously track the coevolution of population size and allele frequencies, and account for the positive feedback between increasing load and declining numbers (Szép et al. (2021)) while the second class of models primarily focus on the consequences of frequency and/or density-dependent selection for the preservation of genetic diversity (Whitlock, 2002). In the latter class of models, hard vs. soft selection (which refers to whether or not local populations contribute to the next generation in proportion to their fitness) is often confounded with local vs. global density regulation (which determines, among other factors, whether changes in deleterious allele frequency can

accumulate over multiple generations within local populations). Moreover, these models assume that populations are consistently at their carrying capacity (Whitlock, 2002), thereby neglecting any influence of genetic load on total population size and consequently, on the extent of drift. It is these complexities that underscore the need for models that explicitly incorporate eco-evolutionary feedback and allow for changes in both local and global population sizes. These models not only provide a fresh perspective but also yield distinct results that diverge from the existing body of research. For example, we showed that under such a model with explicit regulation, beyond a critical threshold, reduced Km has only a slight effect on metapopulation outcomes (causing a minor decline in average size); a result contrary to results from constant metapopulation size models where reduced gene flow is thought to be beneficial for metapopulations (Whitlock, 2002).

How might the results of our analysis compare with results from models that account for the explicit arrangement of patches (and hence different migration patterns) such as the one and two dimensional stepping-stone models? Bascompte and Solé (1996) posited that metapopulations with explicit spatial considerations might exhibit heightened vulnerability to habitat fragmentation compared to their spatially implicit counterparts suggesting that non-spatial models such as the one considered in this study may tend to underestimate critical thresholds and the impact of fragmentation on metapopulation persistence. This begs the question of the role of different spatial configurations or arrangements on critical thresholds for persistence; are there specific arrangements that prove to be more efficient than others? In general, since the amount of habitat loss that a population can withstand may depend on its spatial distribution, it will be interesting in future to extend our model to incorporate more complex effects of landscape structure and habitat configurations (such as habitat corridors and stepping stone habitats (Bennett, 1999)). Such models may not only provide a more realistic depiction of how fragmentation affect populations but may also play a crucial role in moderating its impacts on the long-term persistence of metapopulations.

What are some of the limitations of our study? Our theoretical framework is based on the infinite island model. Although this is a useful simplification that ensures analytical tractability, in reality, no natural population is truly infinite. Our analysis may therefore benefit from considering finite size populations as this may provide us with better theoretical insights (see Barton and Olusanya (2022)).

Secondly, in our analysis, we make the simplifying assumption that all demes of the metapopulation are at one equilibrium state $N*$ so that $\overline{N} = N*$ in eq. (5.2). We therefore fail to account for scenarios where the different demes can end up in different equilibrium states so that $\overline{N} - N* \neq 0$ (the bimodality we saw in fig. 5.3c and fig. 5.3d were obtained in the absence of this heterogeneity and such bimodality occur close to critical threshold for extinction). In addition to this, we defined load as $R_g = 1 - \overline{W}$

where we have essentially assumed that the optimal genotype in a population is one with fitness 1. Although, this is a theoretically sound way to think about the problem, however, in reality, we cannot always measure the perfect genotype.

There are other extensions to our model that can be considered in the future. For one, we can explore the role of negative epistasis which may help reduce load (Kimura and Maruyama, 1966). The idea behind this being that with such negative epistasis, the more deleterious mutation an individual has, the less and less fit it becomes just because each mutation becomes worse in the presence of the other. Thus, as individuals die out, they take with them a lot of deleterious mutations which consequently helps to reduce load.

Secondly, in this work, we have assumed that offspring gametes are formed merely by freely recombining parental gametes i.e., assuming a recombination rate, $r = 0.5$. Such high rates of recombination can be more potent at disrupting the linkage between genetic loci, fostering the formation of novel allelic combinations. This can lead to a more rapid reduction in load, as deleterious mutations are more likely to be separated and ultimately purged from the population (Kimura and Maruyama, 1966; Crow, 1970, 2017). How might this change with lower rates of recombination or if recombination is allowed to vary across the genome and how much more will load be exacerbated in this case? Providing answers to these questions may further contribute to our understanding of the maintenance of genetic variation and the adaptation of populations to changing environmental conditions.

In addition, we can also extend the investigation of the role of multilocus interactions (i.e., linkage and identity disequilibrium) on load to the case of hard selection using the effective migration approximation. However, validating this with individual-based simulations would be computationally intensive.

Finally, we have assumed a metapopulation landscape where selection is uniform across space. It would be interesting to extend our framework to explore what happens when some loci are under spatially heterogeneous selection (as in Szép et al. (2021)) while others are subject to unconditionally deleterious mutation. What would be the impact of gene flow under such scenarios? While gene flow may potentially swamp local adaptation, it could also concurrently alleviate the burden of deleterious mutations. Determining the relative strengths of these contrasting effects under realistic parameter regimes remains a crucial endeavor worth exploring.

References

Aitken, S. N. and Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual review of ecology, evolution, and systematics*, 44:367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747.

Barton, N. and Olusanya, O. (2022). The response of a metapopulation to a changing environment. *Philosophical Transactions of the Royal Society B*, 377(1848):20210009. https://doi.org/10.1098/rstb.2021.0009.

Bascompte, J. and Solé, R. V. (1996). Habitat fragmentation and extinction thresholds in spatially explicit models. *Journal of Animal ecology*, pages 465–473. https://doi.org/10.2307/5781.

Bennett, A. F. (1999). *Linkages in the landscape: the role of corridors and connectivity in wildlife conservation*. Number 1. Iucn.

Casas-Marce, M., Soriano, L., López-Bao, J. V., and Godoy, J. A. (2013). Genetics at the verge of extinction: insights from the iberian lynx. *Molecular ecology*, 22(22):5503–5515. https://doi.org/10.1111/mec.12498.

Charlesworth, B. (2013). Why we are not dead one hundred times over: Brief communication. *Evolution*, 67(11):3354–3361. https://doi.org/10.1111/evo.12195.

Charlesworth, D. and Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual review of ecology and systematics*, 18(1):237–268. https://doi.org/10.1146/annurev.es.18.110187.001321.

Christiansen, F. B. (1975). Hard and soft selection in a subdivided population. *The American Naturalist*, 109(965):11–16. https://doi.org/10.1086/282970.

Crow, J. (1970). Genetic loads and the cost of natural selection. In *Mathematical topics in population genetics*, pages 128–177. Springer. https://doi.org/10.1007/978-3-642-46244-3_5.

Crow, J. F. (2017). *An introduction to population genetics theory*. Scientific Publishers.

Denver, D. R., Morris, K., Lynch, M., and Thomas, W. K. (2004). High mutation rate and predominance of insertions in the caenorhabditis elegans nuclear genome. *Nature*, 430(7000):679–682. https://doi.org/10.1038/nature02697.

Drake, J. M. and Griffen, B. D. (2010). Early warning signals of extinction in deteriorating environments. *Nature*, 467(7314):456–459. https://doi.org/10.1038/nature09389.

Edmands, S. (2007). Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular ecology*, 16(3):463–475. https://doi.org/10.1111/j.1365-294X.2006.03148.x.

Eyre-Walker, A., Woolfit, M., and Phelps, T. (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics*, 173(2):891–900.

Fagan, W. F. and Holmes, E. (2006). Quantifying the extinction vortex. *Ecology letters*, 9(1):51–60. https://doi.org/10.1111/j.1461-0248.2005.00845.x.

Finger, A., Kettle, C., Kaiser-Bunbury, C., Valentin, T., Doudee, D., Matatiken, D., and Ghazoul, J. (2011). Back from the brink: potential for genetic rescue in a critically endangered tree. *Molecular Ecology*, 20(18):3773–3784. https://doi.org/10.1111/j.1365-294X.2011.05228.x.

Frankham, R. (1998). Inbreeding and extinction: island populations. *Conservation biology*, 12(3):665–675. https://doi.org/10.1046/j.1523-1739.1998.96456.x.

Frankham, R., Ballou, J. D., Eldridge, M. D., Lacy, R. C., Ralls, K., Dudash, M. R., and Fenster, C. B. (2011). Predicting the probability of outbreeding depression. *Conservation Biology*, 25(3):465–475. https://doi.org/10.1111/j.1523-1739.2011.01662.x.

Frankham, R., Briscoe, D. A., and Ballou, J. D. (2002). *Introduction to conservation genetics*. Cambridge university press.

Gamba, D. and Muchhala, N. (2020). Global patterns of population genetic differentiation in seed plants. *Molecular Ecology*, 29(18):3413–3428. https://doi.org/10.1111/mec.15575.

Gillespie, J. H. (2004). *Population genetics: a concise guide*. JHU press.

Glémin, S., Ronfort, J., and Bataillon, T. (2003). Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics*, 165(4):2193–2212. https://doi.org/10.1093/genetics/165.4.2193.

Gomulkiewicz, R., Holt, R. D., and 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. https://doi.org/10.1006/tpbi.1998.1405.

Greenberg, R. and Crow, J. F. (1960). A comparison of the effect of lethal and detrimental chromosomes from drosophila populations. *Genetics*, 45(8):1153. https://doi.org/10.1093/genetics/45.8.1153.

Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D. L., Charlesworth, B., and Keightley, P. D. (2007). Direct estimation of per nucleotide and genomic deleterious mutation rates in drosophila. *Nature*, 445(7123):82–85. https://doi.org/10.1038/nature05388.

Hairston Jr, N. G., Ellner, S. P., Geber, M. A., Yoshida, T., and Fox, J. A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecology letters*, 8(10):1114–1127. https://doi.org/10.1111/j.1461-0248.2005.00812.x.

Hamrick, J. L., Godt, M. W., et al. (1990). Allozyme diversity in plant species. *Plant population genetics, breeding, and genetic resources.*, pages 43–63.

Herman, A., Brandvain, Y., Weagley, J., Jeffery, W. R., Keene, A. C., Kono, T. J., Bilandžija, H., Borowsky, R., Espinasa, L., O'Quin, K., et al. (2018). The role of gene flow in rapid and repeated evolution of cave-related traits in mexican tetra, astyanax mexicanus. *Molecular ecology*, 27(22):4397–4416. https://doi.org/10.1111/mec.14877.

Higgins, K. and Lynch, M. (2001). Metapopulation extinction caused by mutation accumulation. *Proceedings of the National Academy of Sciences*, 98(5):2928–2933. https://doi.org/10.1073/pnas.031358898.

Ingvarsson, P. K. and Whitlock, M. C. (2000). Heterosis increases the effective migration rate. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 267(1450):1321–1326. https://doi.org/10.1098/rspb.2000.1145.

Keightley, P. D. (2012). Rates and Fitness Consequences of New Mutations in Humans. *Genetics*, 190(2):295–304. https://doi.org/10.1534/genetics.111.134668.

Keightley, P. D. and Eyre-Walker, A. (2010). What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1544):1187–1193. https://doi.org/10.1098/rstb.2009.0266.

Kimura, M. and Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics*, 54(6):1337. https://doi.org/10.1093/genetics/54.6.1337.

Kimura, M., Maruyama, T., and Crow, J. F. (1963). The mutation load in small populations. *Genetics*, 48(10):1303–1312. https://doi.org/10.1093/genetics/48.10.1303.

Kondrashov, A. S. (1995). "contamination of the genome by very slightly deleterious mutations: Why have we not died 100 times over?". *Journal of Theoretical Biology*, 175(4):583–594. https://doi.org/10.1006/jtbi.1995.0167.

Land, E. D. and Lacy, R. C. (2000). Introgression level achieved through florida panther genetic restoration. *Endangered Species Update*, 17(5):100–100.

Lande, R. (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist*, 142(6):911–927. https://doi.org/10.1086/285580.

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

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, 87(836):331–333. https://doi.org/10.1086/281792.

Lyman, R. F., Lawrence, F., Nuzhdin, S. V., and Mackay, T. F. (1996). Effects of single p-element insertions on bristle number and viability in drosophila melanogaster. *Genetics*, 143(1):277–292. https://doi.org/10.1093/genetics/143.1.277.

Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L., and Willis, J. (1999). Perspective: spontaneous deleterious mutation. *Evolution*, 53(3):645–663. https://doi.org/10.2307/2640707.

Lynch, M., Conery, J., and Burger, R. (1995a). Mutation accumulation and the extinction of small populations. *The American Naturalist*, 146(4):489–518. https://doi.org/10.1086/285812.

Lynch, M., Conery, J., and Bürger, R. (1995b). Mutational meltdowns in sexual populations. *Evolution*, 49(6):1067–1080. https://doi.org/10.2307/2410432.

Matthies, D., Bräuer, I., Maibom, W., and Tscharntke, T. (2004). Population size and the risk of local extinction: empirical evidence from rare plants. *Oikos*, 105(3):481–488. https://doi.org/10.1111/j.0030-1299.2004.12800.x.

Mukai, T. (1964). The genetic structure of natural populations of drosophila melanogaster. i. spontaneous mutation rate of polygenes controlling viability. *Genetics*, 50(1):1. https://doi.org/10.1093/genetics/50.1.1.

Mukai, T., Chigusa, S. I., Mettler, L., and Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability in drosophila melanogaster. *Genetics*, 72(2):335–355. https://doi.org/10.1093/genetics/72.2.335.

Olusanya, O., Khudiakova, K., and Sachdeva, H. (2023). Genetic load, eco-evolutionary feedback and extinction in a metapopulation. https://git.ista.ac.at/oolusany/genetic-load-and-extinction-in-a-metapopulation.

Ravigné, V., Olivieri, I., and Dieckmann, U. (2004). Implications of habitat choice for protected polymorphysms.

Rouhani, S. and Barton, N. H. (1993). Group selection and the 'shifting balance'. *Genetics Research*, 61(2):127–135. doi.org/10.1017/S0016672300031232.

Roze, D. (2015). Effects of interference between selected loci on the mutation load, inbreeding depression, and heterosis. *Genetics*, 201(2):745–757. https://doi.org/10.1534/genetics.115.178533.

Roze, D. and Rousset, F. (2004). Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. *Genetics*, 167(2):1001–1015. https://doi.org/10.1534/genetics.103.025148.

Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., and Hanski, I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, 392(6675):491–494. https://doi.org/10.1038/33136.

Sachdeva, H. (2022). Reproductive isolation via polygenic local adaptation in sub-divided populations: Effect of linkage disequilibria and drift. *PLoS Genetics*, 18(9):e1010297. https://doi.org/10.1371/journal.pgen.1010297.

Sachdeva, H., Olusanya, O., and Barton, N. (2022). Genetic load and extinction in peripheral populations: the roles of migration, drift and demographic stochasticity. *Philosophical Transactions of the Royal Society B*, 377(1846):20210010. https://doi.org/10.1098/rstb.2021.0010.

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., and Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, 413(6856):591–596. https://doi.org/10.1038/35098000.

Simmons, M. J. and Crow, J. F. (1977). Mutations affecting fitness in drosophila populations. *Annual review of genetics*, 11(1):49–78. https://doi.org/10.1146/annurev.ge.11.120177.000405.

Szczecińska, M., Sramko, G., Wołosz, K., and Sawicki, J. (2016). Genetic diversity and population structure of the rare and endangered plant species pulsatilla patens (l.) mill in east central europe. *PloS one*, 11(3):e0151730. https://doi.org/10.1371/journal.pone.0151730.

Szép, E., Sachdeva, H., and Barton, N. H. (2021). Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5):1030–1045. https://doi.org/10.1111/evo.14210.

Templeton, A. R. (1986). Coadaptation and outbreeding depression. *Conservation biology: the science of scarcity and diversity*.

Thompson, J. N. (1998). Rapid evolution as an ecological process. *Trends in ecology & evolution*, 13(8):329–332. https://doi.org/10.1016/S0169-5347(98)01378-0.

Uecker, H., Otto, S. P., and Hermisson, J. (2014). Evolutionary rescue in structured populations. *The American Naturalist*, 183(1):E17–E35. https://doi.org/10.1086/673914.

Whitlock, M. C. (2002). Selection, Load and Inbreeding Depression in a Large Metapopulation. *Genetics*, 160(3):1191–1202. https://doi.org/10.1093/genetics/160.3.1191.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97. https://doi.org/10.1093/genetics/16.2.97.

Wright, S. (1937). The distribution of gene frequencies in populations. *Proceedings of the National Academy of Sciences of the United States of America*, 23(6):307. https://doi.org/10.1073/pnas.23.6.307.

Zhou, S.-R. and Pannell, J. R. (2010). Inbreeding depression and genetic load at partially linked loci in a metapopulation. *Genetics Research*, 92(2):127–140. https://doi.org/10.1017/S0016672310000133.

Zwaenepoel, A., Sachdeva, H., and Fraïsse, C. (2023). Polygenic barriers to gene flow: the role of dominance, haploid selection and heterogeneous genetic architectures. *bioRxiv*. https://www.biorxiv.org/content/early/2023/09/25/2023.09.24.559235.

General discussion

A comprehension of the profound impact of evolutionary processes on genetic diversity is of paramount importance when seeking to unravel the mysteries of the natural world and predict how species might respond to changes in their environment. This principle, which lies at the heart of evolutionary biology, holds important implications for biodiversity and conservation. Over the last century, climate change and human-induced activities such as pollution, habitat destruction, and modifications in land use have had profound effects on the demographics of many species (Shivanna, 2022). While some have already gone extinct, some others are teetering on the brink due to declining population sizes and shrinking habitats. If species cannot keep pace with shifting climate and evolving land use patterns, the future prospects of many such species become yet bleak.

This threat to biodiversity is a stark reminder of the complex interplay between ecology, evolution, and environmental shifts. We find ourselves in a compelling race between the rate at which environments transform and the ability of species to adapt genetically. It is a race with incredibly high stakes as species extinction resulting from failure to adapt can reverberate through ecosystems, impacting essential functions like nutrient cycling and pollination.

Efforts to understand how evolutionary and ecological forces interact to shape genetic diversity have been a subject of inquiry for several decades. Ecologists and geneticists have employed various approaches including phylogenetics, population genetics, and genomics to unravel these intricate relationships. Recent advancements in genomics (and computational approaches) have for example improved our capacity to explore these dynamics, shedding light on the genetic consequences of changing conditions (Matthews et al., 2018; Rodríguez-Verdugo et al., 2017). For instance, studies have revealed how certain genes associated with response to pollutants and temperature tolerance have evolved in response to different environmental stressors (Fisher and Oleksiak, 2007; Chen et al., 2018; Elakhdar et al., 2023). Through whole-genome sequencing, one can now explore genetic variation at unprecedented scales, uncovering insights into selected genomic

regions and genes (Lappalainen et al., 2019; Satam et al., 2023). While these approaches have offered valuable insights, a comprehensive theoretical understanding is still lacking and many questions remain unanswered.

The work done in this thesis seeks to fill this gap by providing a theoretical understanding of how ecological (population dynamics) and evolutionary processes shape genetic diversity and eventual outcomes in fragmented populations (modelled here as metapopulations consisting of many connected subpopulations).

This theoretical exploration begins in chapter 2, where we investigate how well a population inhabiting a heterogeneous landscape adapts to changes in both local and global conditions. We assumed a soft selection metapopulation model where the metapopulation is made up of two habitats each adapted to differing local conditions and each having different proportions of demes (ρ and $1-\rho$ respectively). Our modelling framework was based on the diffusion approximation (Barton and Rouhani, 1993; Banglawala, 2010; Szép et al., 2021) and fixed state approximation (which holds true in the limit of limited migration i.e., $Nm \ll 1$). Our results revealed that when selection is weak relative to drift, polymorphism under this framework is only possible for a narrow range of habitat proportions. We further found that when conditions change locally (i.e., in a single deme), it takes the population a time of order $\sim m^{-1}$ to reach a new equilibrium. However, this rate of convergence is much faster when conditions change across the metapopulation.

In chapter 3, we extended the two-habitat soft selection metapopulation model to a multi-habitat case, focusing on a three-habitat example. We explored the conditions for the maintenance of polymorphism under this framework and how this depends on factors such as the rate of migration, selection, drift, and habitat proportions. Our findings showed that strong selection and limited migration favor the maintenance of a wider range of polymorphism (in particular $Ns \gtrsim Nm$). We also found that an important determinant for the maintenance of polymorphism is the ratio, β, of the proportion of demes that favor an allele to the proportion where it is disfavored. The proportion of demes in an intermediate habitat where alleles are neither favoured nor disfavoured makes little difference.

In chapter 4 and 5, we shifted our attention from spatially heterogeneous populations (considered in 3) to one where selection is uniform across space. In chapter 4 we explored how asymmetric migration from a large mainland to an island influences load and the eventual outcome of the island population as well as the impact of eco-evo feedback. Our analysis was based on the semi-deterministic approximation which ignores fluctuations in population size due to randomness in birth and death (i.e., demographic stochasticity) but accounts for genetic drift. The important conclusion here is that migration can have different effects on the outcome of peripheral isolates - it can result in an increase in load

in a large-population state (by preventing purging) and reduce the load in a sink state (through an increase in heterozygosity). The actual effect of migration however depends crucially on the specific parameters involved such as the total genomic mutation rate as well as the fitness effects of deleterious variants.

Chapter 5 wraps up the thesis by extending the mainland-island analysis of chapter 4 to a metapopulation with very many demes connected by migration and explores the role of gene flow on load (under both soft and hard selection) and metapopulation persistence (under hard selection). We again make use of the diffusion and semi-deterministic approximation for our analysis. However, in addition, we also utilize the effective migration approximation to examine the role of multilocus effects (i.e., selective interference) on load due to a single locus (and under soft selection). Our results showed that similar to previous work (Whitlock, 2002; Roze and Rousset, 2004), a little migration is enough to substantially offset load under soft selection. Interestingly, with hard selection, one would require much more gene flow to ensure metapopulation persistence. We further identified critical thresholds for metapopulation persistence and demonstrated that thresholds are more stringent as selection becomes harder.

The studies conducted in this thesis are not without their limitations. One such limitation is our focus on simple spatially implicit models - the mainland-island and many island model of population structure. Although these are useful simplifications of spatially extended populations, nature is far more complex. In light of this, future studies should employ more sophisticated models of subdivision and explore how different patterns of migration and interactions within the metapopulation landscape might influence the conclusions of this studies. For example, we could consider the one (or two) dimensional model where demes are connected linearly (or on a lattice) and migration occurs between neighboring demes (stepping stone models) or is defined via some Gaussian dispersal kernel, approximating diffusive dispersal (see Polechová, 2018). We can also look into models with truly continuous space, e.g., models of isolation by distance (IBD), which assume that genetic differentiation among populations increases with increasing dispersal distance. These models may better capture the complexities inherent in real-world systems and could therefore provide a better understanding into the dynamics of adaptation, genetic load and extinction in fragmented populations.

A second limitation of the study is that the models considered here assume linkage equilibrium, LE (i.e., that alleles at different loci are associated randomly with each other), while this holds true in the limit of weak selection relative to the rate of recombination. With strong selection however, gene combinations that offer a certain fitness advantage can buildup in the population at a faster rate than can be broken down by recombination, leading to a departure from LE and the need to account for linkage disequilibrium (LD). Hence, future exploration should incorporate such non-random associations (as well as

varying recombination rates). The effect of LD at a single locus can for instance be captured using the effective migration approximation as discussed in Zwaenepoel et al. (2023) (also see Petry (1983); Sakamoto and Innan (2019); Sachdeva (2022)) which accounts for how the net effect of selection at other loci influence gene flow.

Thirdly, in our analysis, we assumed simple patterns of selection. For example in chapter 2, we assumed directional selection favoring alternative alleles in the two habitats considered and in chapter 3, we assumed an antagonistic environmental effect where selection for a given trait varies in magnitude and direction as we move from one habitat to another across the metapopulation. However, genetic variation underlying many traits is often governed by stabilizing selection (Kingsolver et al., 2001; Hunt et al., 2007; Sanjak et al., 2018; Holand et al., 2020), in which close adaptation of a population to a given optimum is achieved by different allelic combinations. Such selection, when it occurs, is expected to decrease a population's genetic variance as it favors only average phenotypes and selects against extreme variants. Future research can therefore extend the current framework to stabilizing selection and explore the dynamics of adaptation and maintenance of genetic variation under such a model. We would expect this to be an interesting albeit challenging endeavor because of the presence of multiple adaptive peaks under such a model.

Furthermore, chapter 4 and 5 explored the effects of selection on multiple loci. Extending this to a highly polygenic architecture in terms of the infinitesimal model (Barton and Keightley, 2002; Barton et al., 2017) where load depends on infinitely many loci each with small effect would be an important contribution. Not only might this allow for a more accurate reflection of the complexity of real-world genetic systems, it may also provide a more comprehensive understanding of the impact of deleterious mutation accumulation on metapopulation viability and long-term persistence.

Despite the above limitations, this research provides us with a fundamental understanding of eco-evolutionary dynamics in fragmented landscapes and can help in predicting the long-term effects of habitat fragmentation on the adaptability and continued existence of species. For instance, the theoretical understanding of critical thresholds for persistence provided in chapter 5 may be useful in identifying tipping points in ecosystems, thereby providing crucial insights into the management and protection of vulnerable species across fragmented habitats. Secondly, an understanding of the factors that promote the persistence of polymorphisms can inform strategies to enhance genetic variation within populations and the ability of such populations to thrive despite changing conditions. Our study therefore has an important bearing to conservation and biodiversity efforts.

References

Banglawala, N. (2010). Local adaptation under demographic and genetic uctuations. *(Doctoral thesis))*. The university of Edinburgh.

Barton, N. and Rouhani, S. (1993). Adaptation and the 'shifting balance'. *Genetics Research*, 61(1):57–74. https://doi.org/10.1017/S0016672300031098.

Barton, N. H., Etheridge, A. M., and Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical population biology*, 118:50–73. https://doi.org/10.1016/j.tpb.2017.06.001.

Barton, N. H. and Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics*, 3(1):11–21. https://doi.org/10.1038/nrg700.

Chen, B., Feder, M. E., and Kang, L. (2018). Evolution of heat-shock protein expression underlying adaptive responses to environmental stress. *Molecular ecology*, 27(15):3040–3054. https://doi.org/10.1111/mec.14769.

Elakhdar, A., Slaski, J. J., Kubo, T., Hamwieh, A., Hernandez Ramirez, G., Beattie, A. D., and Capo-Chichi, L. J. (2023). Genome-wide association analysis provides insights into the genetic basis of photosynthetic responses to low-temperature stress in spring barley. *Frontiers in Plant Science*, 14:1159016. https://doi.org/10.3389/fpls.2023.1159016.

Fisher, M. A. and Oleksiak, M. F. (2007). Convergence and divergence in gene expression among natural populations exposed to pollution. *BMC genomics*, 8(1):1–10. https://doi.org/10.1186/1471-2164-8-108.

Holand, H., Kvalnes, T., Røed, K. H., Holand, Ø., Sæther, B.-E., and Kumpula, J. (2020). Stabilizing selection and adaptive evolution in a combination of two traits in an arctic ungulate. *Evolution*, 74(1):103–115. https://doi.org/10.1111/evo.13894.

Hunt, J., Blows, M. W., Zajitschek, F., Jennions, M. D., and Brooks, R. (2007). Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (teleogryllus commodus). *Genetics*, 177(2):875–880. https://doi.org/10.1534/genetics.107.077057.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C., Hoang, A., Gibert, P., and Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, 157(3):245–261. https://doi.org/10.1086/319193.

Lappalainen, T., Scott, A. J., Brandt, M., and Hall, I. M. (2019). Genomic analysis in the age of human genome sequencing. *Cell*, 177(1):70–84. https://doi.org/10.1016/j.cell.2019.02.032.

Matthews, B., Best, R. J., Feulner, P. G., Narwani, A., and Limberger, R. (2018). Evolution as an ecosystem process: insights from genomics. *Genome*, 61(4):298–309. https://doi.org/10.1139/gen-2017-0044.

Petry, D. (1983). The effect on neutral gene flow of selection at a linked locus. *Theoretical population biology*, 23(3):300–313. https://doi.org/10.1016/0040-5809(83)90020-5.

Polechová, J. (2018). Is the sky the limit? on the expansion threshold of a species' range. *PLoS biology*, 16(6):e2005372. https://doi.org/10.1371/journal.pbio.2005372.

Rodríguez-Verdugo, A., Buckley, J., and Stapley, J. (2017). The genomic basis of eco-evolutionary dynamics. *Molecular ecology*, 26(6):1456–1464. https://doi.org/10.1111/mec.14045.

Roze, D. and Rousset, F. (2004). Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. *Genetics*, 167(2):1001–1015. https://doi.org/10.1534/genetics.103.025148.

Sachdeva, H. (2022). Reproductive isolation via polygenic local adaptation in sub-divided populations: Effect of linkage disequilibria and drift. *PLoS Genetics*, 18(9):e1010297. https://doi.org/10.1371/journal.pgen.1010297.

Sakamoto, T. and Innan, H. (2019). The evolutionary dynamics of a genetic barrier to gene flow: From the establishment to the emergence of a peak of divergence. *Genetics*, 212(4):1383–1398. https://doi.org/10.1534/genetics.119.302311.

Sanjak, J. S., Sidorenko, J., Robinson, M. R., Thornton, K. R., and Visscher, P. M. (2018). Evidence of directional and stabilizing selection in contemporary humans. *Proceedings*

of the National Academy of Sciences, 115(1):151–156. https://doi.org/10.1073/pnas.1806837115.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., et al. (2023). Next-generation sequencing technology: Current trends and advancements. *Biology*, 12(7):997. https://doi.org/10.3390/biology12070997.

Shivanna, K. (2022). Climate change and its impact on biodiversity and human welfare. *Proceedings of the Indian National Science Academy*, 88(2):160–171. https://doi.org/10.1007/s43538-022-00073-6.

Szép, E., Sachdeva, H., and Barton, N. H. (2021). Polygenic local adaptation in metapopulations: A stochastic eco-evolutionary model. *Evolution*, 75(5):1030–1045. https://doi.org/10.1111/evo.14210.

Whitlock, M. C. (2002). Selection, Load and Inbreeding Depression in a Large Metapopulation. *Genetics*, 160(3):1191–1202. https://doi.org/10.1093/genetics/160.3.1191.

Zwaenepoel, A., Sachdeva, H., and Fraïsse, C. (2023). Polygenic barriers to gene flow: the role of dominance, haploid selection and heterogeneous genetic architectures. *bioRxiv*. https://www.biorxiv.org/content/early/2023/09/25/2023.09.24.559235.

Appendix A

The response of a metapopulation to a changing environment

A.1 Range of habitat proportions for which polymorphism is possible

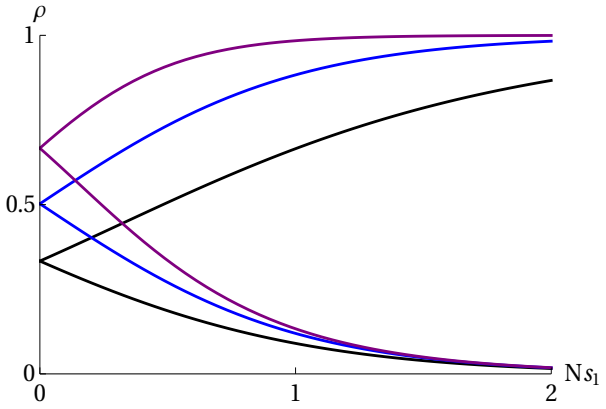


Figure A.1: Bounds on the proportions of habitat 1, ρ, between which polymorphism is possible, as a function of the strength of selection in that habitat, Ns_1. For a given Ns_1, ρ has to lie between the two curves for polymorphism to be maintained. The three sets of bounds correspond to $Ns_2/Ns_1 = 0.5, 1, 2$ (black, blue and purple respectively). These results apply in the limit of low migration, and soft selection.

A.2 Accuracy of the fixed-state approximation

Here, we compare the the mean allele frequency in an infinite metapopulation under the diffusion approximation with the fixed-state approximation for different Nm values. As expected, the accuracy of the fixed state approximation holds only for small Nm.

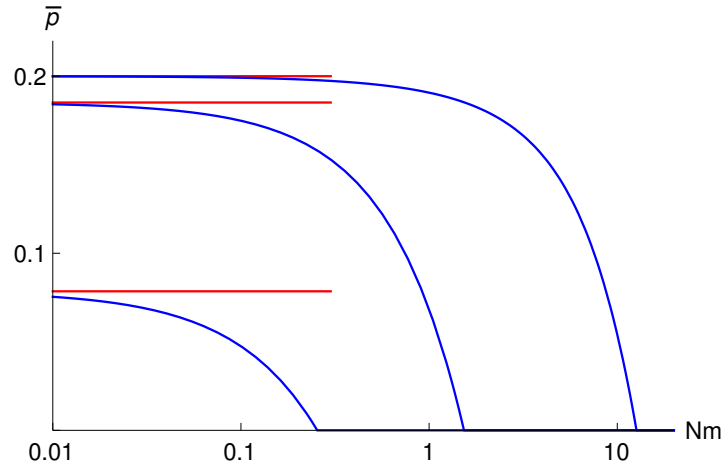
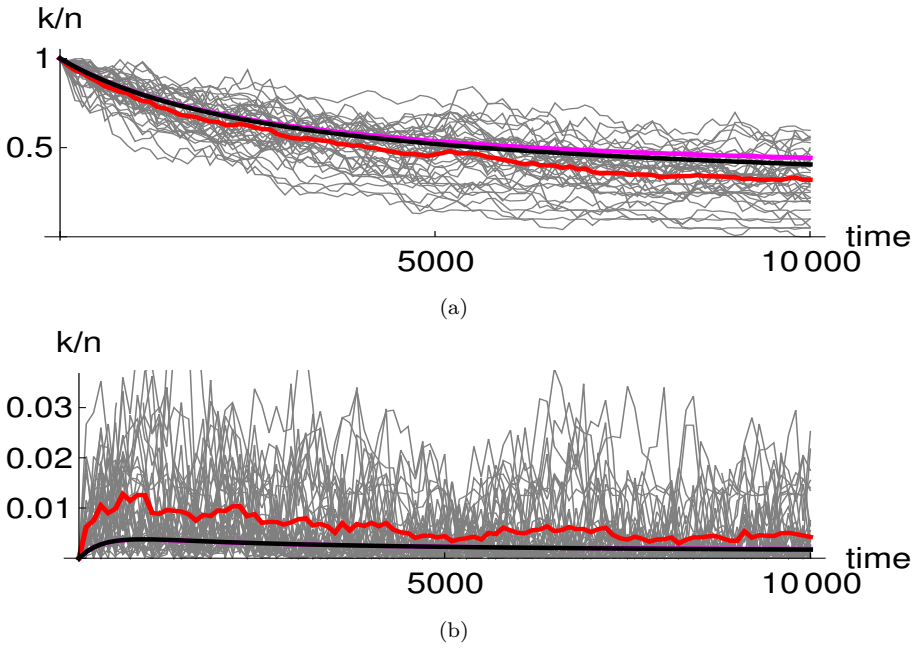


Figure A.2: The mean allele frequency in an infinite metapopulation, plotted against Nm; $\rho = 0.2$, Ns_1, $Ns_2 = 1, -2$ (lower curve) 2, -4 (middle curve) or 10, -20 (upper curve). The fixed-state approximation, which applies for small Nm, is shown by the red lines.

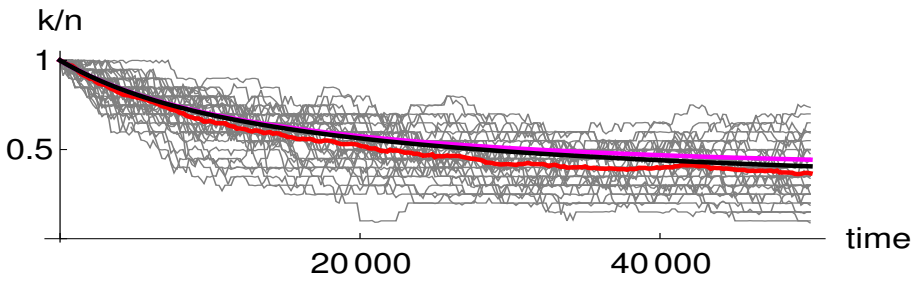
A.3 Loss of diversity in a finite population



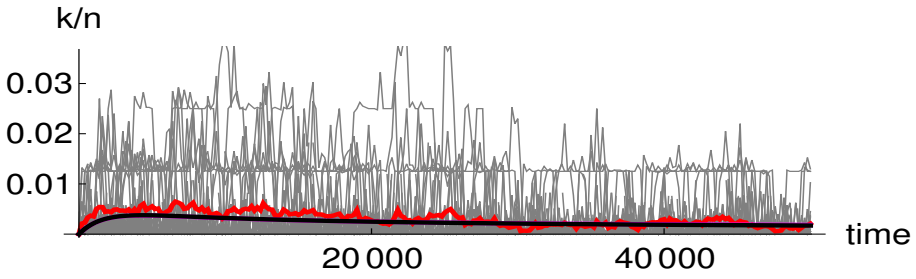
(a)

(b)

Figure A.3: Loss of diversity in a metapopulation of 100 demes, which is initially perfectly adapted. (a) Mean allele frequency is plotted against time, in the 20 demes where the focal allele is favoured and, (b) in the 80 demes where it is not. Thin grey lines show allele frequencies at 40 loci, averaged over demes; the red line shows the overall mean. The black curve shows the fixed-state approximation, for a finite metapopulation, and the magenta line, for an infinite metapopulation. Simulations are for $N = 50$, $Nm = 0.05$, $s_{1,2} = \{0.02, -0.04\}$; thus, $Ns_{1,2} = \{1, -2\}$, so that selection and drift are of similar magnitude.



(a)



(b)

Figure A.4: This is identical to fig. A.3, except that $Nm = 0.01$, and the timescale is correspondingly longer. The fixed-state approximation is more accurate with a lower number of migrants.

A.4 Distribution of mean allele frequency

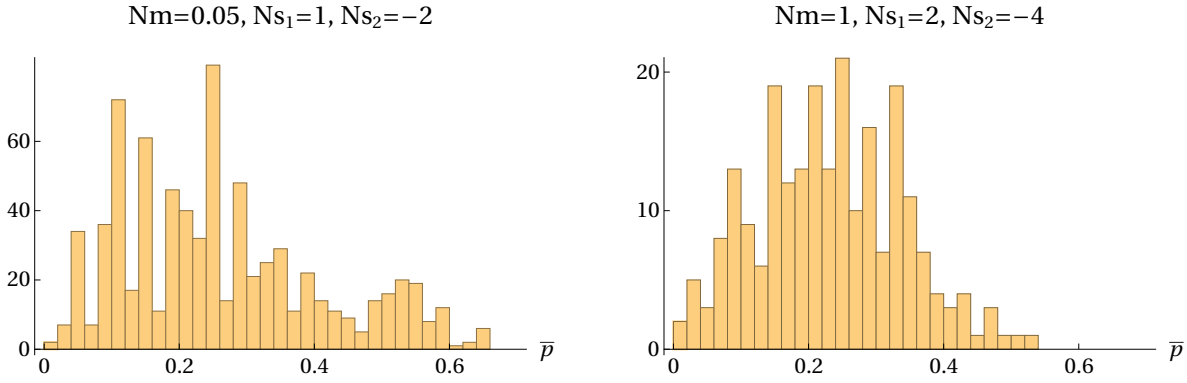


Figure A.5: The distribution of allele frequencies, averaged over the 20 demes in the rare habitat, conditional on polymorphism, and accumulated over generations 8,000 and 8,100, to 10,000.

Appendix B

Local adaptation in a metapopulation - a multi-habitat perspective

B.0.1 Maintenance of polymorphism and critical thresholds for the case $X_1 = 1$, $X_2 = -1$, and $X_3 = -1$

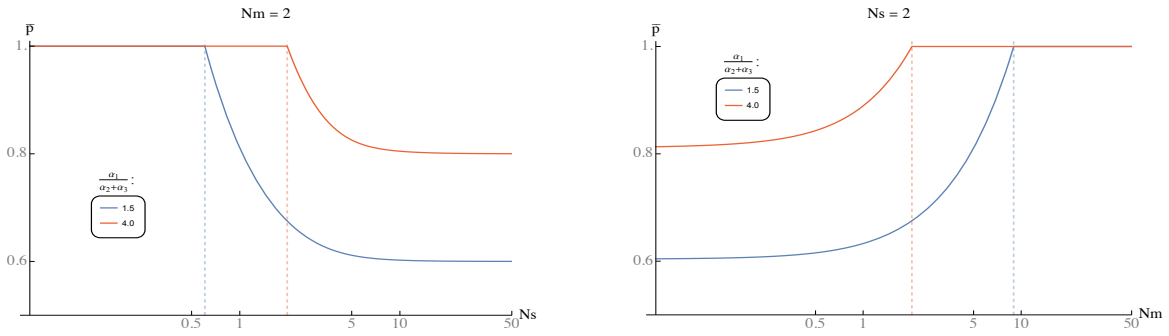


Figure B.1: Critical: (a.) Ns threshold above which and (b.) Nm threshold below which a polymorphism is possible.

B.0.2 Critical migration threshold for polymorphism with similar β but different α_2 values. $X_1 = 1$, $X_2 = 0$, $X_3 = -1$

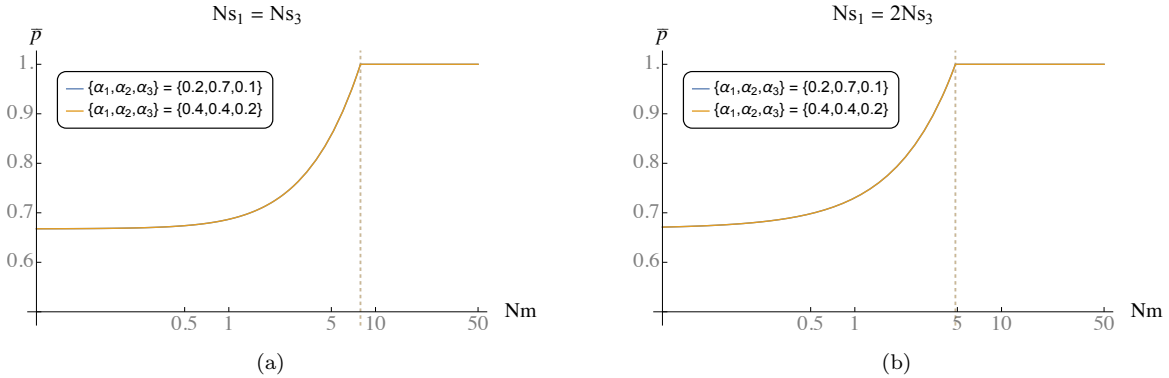


Figure B.2: (a.) Symmetric selection $Ns_1 = Ns_3$ (in magnitude). (b.) Asymmetric selectiom.

B.0.3 Critical selection threshold for polymorphism with similar β but different α_2 values. $X_1 = 1$, $X_2 = 0$, $X_3 = -1$

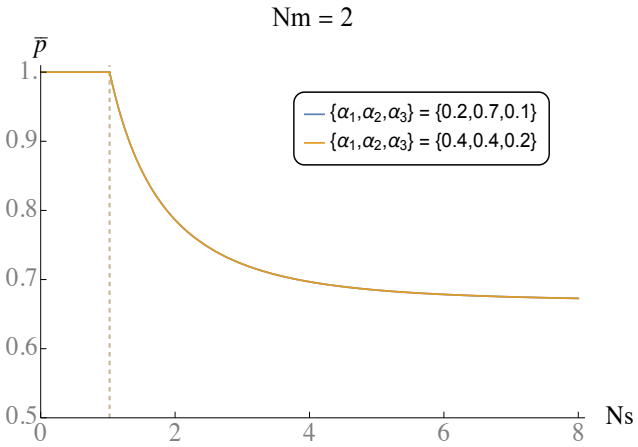
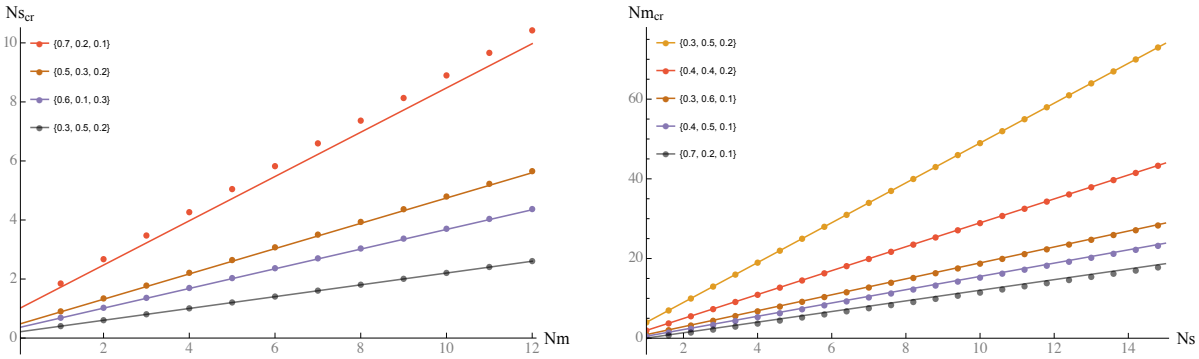


Figure B.3

B.0.4 Comparing Ns_{cr} and Nm_{cr} with numerical solution from the diffusion approximation



(a) (b)

Figure B.4: Critical: (a.) Ns threshold above which and (b.) Nm threshold below which a polymorphism is possible. The different colors represent different $\{\alpha_1, \alpha_2, \alpha_3\}$ combinations. Dotted lines are numerical solutions from the diffusion approximation and solid lines results from eq. (3.6).

B.0.5 Effect of drift on the maintenance of a polymorphism

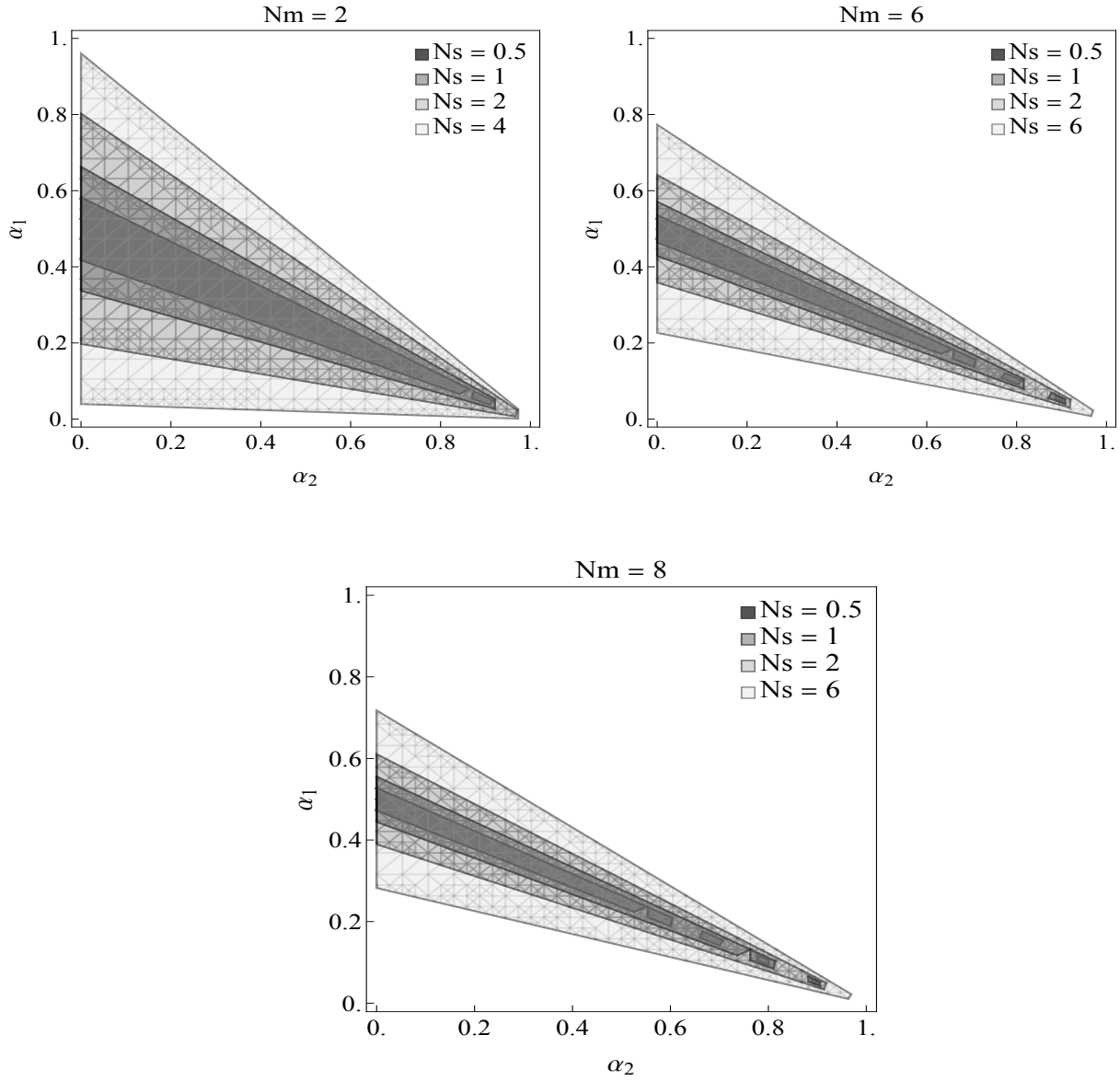
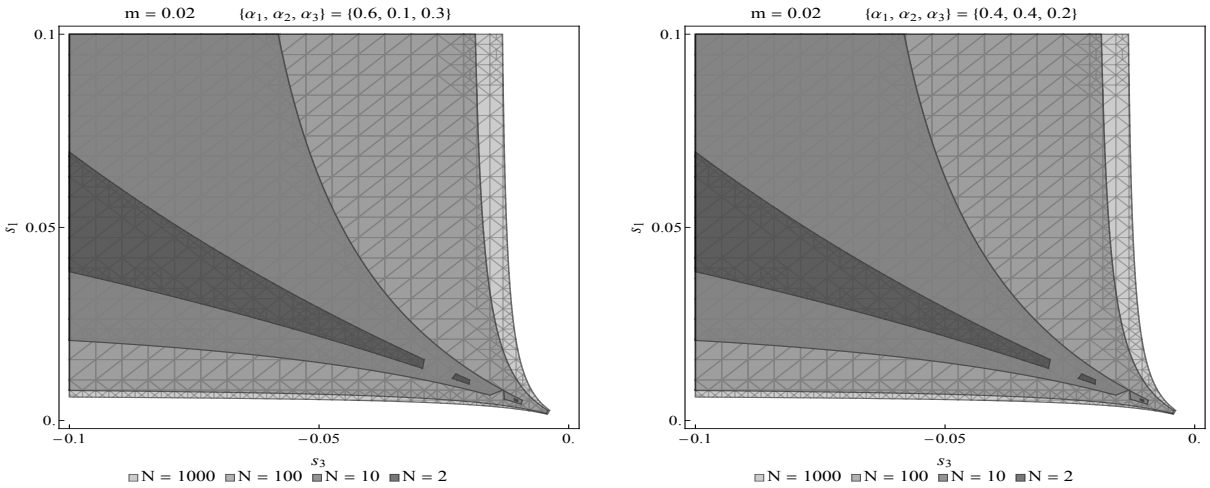


Figure B.5: Drift constrains the region within which a polymorphism is possible.

B.0.6 Further plots showing α_2 does not matter for a polymorphism even with drift.



(a) (b)

Figure B.6: (a.) and (b.) show the effect of drift for the maintenance of a polymorphism using similar values of β but different α_2 values: (a.) $\beta = 2$, $\alpha_2 = 0.1$; (b.) $\beta = 2$, $\alpha_2 = 0.4$.

Appendix C

Genetic load and extinction in marginal populations: the role of migration, drift and demographic stochasticity

C.1 Description of simulations

We carry out two kinds of simulations: simulations assuming LE (linkage equilibrium) and zero inbreeding, which only track allele frequencies at the L loci and population size N as a function of time; and individual-based simulations, which make no simplifying assumptions and track whole diploid genomes (with L loci) of all individuals present in the island population.

Simulations assuming LE and zero inbreeding. If recombination is faster than all ecological and evolutionary processes, then statistical associations between allelic states at different loci (linkage disequilibria or LD) can be neglected. In addition, if there is no significant inbreeding, then the probability of identity by descent at a locus, and correlations between identity by descent across loci (identity disequilibria or ID) are also negligible. Then, individual genotypes are simply random assortments of deleterious and wildtype alleles, and can be generated by independently assigning alternative allelic states to different loci with probabilities equal to the allele frequencies, allowing us to only track allele frequencies (instead of genotypes).

We initiate simulations by assuming that there are K individuals on the island and that the deleterious allele at each locus is absent. We then evolve the population in discrete time by updating allele frequencies and population size in each generation to reflect the effects of migration, mutation, selection and reproduction, and genetic drift. Starting

with frequencies $\{p_j(t)\}$ and size $n(t)$ at the end of generation t, we first implement migration in generation $t+1$ by sampling the number of migrants m from a Poisson distribution with mean m_0. The population size and allele frequencies are then updated as: $p'_j = \frac{n(t)\,p_j(t)+m\,p_j^{(m)}}{n(t)+m}$ and $n' = n(t)+m$. Here, $p_j^{(m)}$ are the corresponding frequencies on the mainland.

Mutation has no effect on population size: $n'' = n'$; allele frequencies are changed as: $p''_j = (1-\mu)p'_j + \mu(1-p'_j)$.

The effects of selection and reproduction are then captured by changing allele frequencies according to: $p'''_j = p''_j + \frac{p''_j[p''_j e^{-s} + q''_j e^{-hs}]}{\overline{w}_j}$, where $\overline{w}_j = {q''_j}^2 + 2p''_j q''_j e^{-hs} + {p''_j}^2 e^{-s}$ and $q''_j = 1 - p''_j$. This is the standard equation relating allele frequencies before and after selection; it assumes that loci evolve independently (no indirect selection due to LD and ID), and that there is no inbreeding. The new population size n''' after selection and reproduction is generated by drawing a Poisson-distributed random variable with mean $n'' e^{r_0(1-n''/K)}\overline{W}$, where the mean population fitness $\overline{W}$ is calculated by multiplying the marginal mean fitnesses of all loci: $\overline{W} = \prod_{j=1}^{L} \overline{w}_j$. Note that sampling the population size from a Poisson distribution (rather than simply choosing it to be equal to the mean of the distribution) introduces demographic stochasticity into the simulation.

The new population size at the end of generation $t+1$ is just: $n(t+1) = n'''$. The new allele frequencies $\{p_j(t+1)\}$ are generated by drawing Binomially distributed random variables $\{X_j\}$ with corresponding parameters $n(t+1)$ and $\{p'''_j\}$ and then setting $\{p_j(t+1)\} = \{X_j/n(t+1)\}$. This last step, which amounts to standard Wright-Fisher sampling of the new allele frequencies based on the new population size $n(t+1)$ and the deterministic allele frequencies p'''_j, 'adds' the random effects of genetic drift on to the systematic effects of migration, mutation and selection (which are already captured by p'''_j).

The procedure described above updates $\{p_j\}$ and n over discrete generations by implementing the effects of migration, mutation, selection and drift in each generation sequentially. An alternative would be to numerically integrate the continuous time equations (eq. 1 in the main text). This is, however, prone to numerical error, and is computationally far more intensive. Both procedures are expected to yield very similar results if $m_0/n, \mu, s, r_0 \ll 1$.

While the allele frequency simulations described here are fast, a fundamental limitation is that the underlying assumptions (i.e., no significant inbreeding, and effects of drift and migration weaker than those of recombination) must break down close to extinction thresholds. Thus, we also carry out individual-based simulations (described below), that make no such assumptions and capture the evolution of all multi-locus associations.

Individual-based simulations. In this bottom-up simulation approach, we track eco-evolutionary dynamics in peripheral populations by explicitly following individuals,

recording their allelic states and population numbers from one generation to another. Besides the advantage that comes from the inclusion of individual variations and interactions, this modeling framework allows us track the evolution of each locus as well as all the associations among loci (i.e. LD and ID).

We simulate a sexually reproducing population with random mating and non-overlapping discrete generations. At the start of the simulation, the island population is assumed to consist of N diploid individuals where the fitness of each individual is determined by l loci with two possible allelic state per locus coded by 1's and 0's. The 1 allele is considered the wildtype and the 0 allele the recessive deleterious variant so that the fitness of the three possible genotypes (11, 10 and 00) at each locus are respectively in the ratio $1 : 1 - hs : 1 - s$ where s is the homozygous selective effect and h is the dominance coefficient affecting the expression of the deleterious allele. We assume multiplicative fitness across loci (i.e. no epistasis) so that the fitness of an individual with n heterozygous loci, m homozygous loci for the deleterious allele and m' homozygous loci for the wild type allele can be expressed as $(1 - hs)^n(1 - s)^m$ where $n + m + m' = l$. The simulation is initialized by assuming that individuals are initially perfectly fit (i.e are composed of only the 0 allelic state) so that the frequency of the deleterious variant at each locus is 0. The order of events in the life cycle of individuals in each generation is taken as, mutation $\rightarrow$ migration $\rightarrow$ reproduction (meiosis and genetic recombination)+ density-dependent survival. These would be explained in a little more detail below.

Mutation: We model both forward and backward mutation rates i.e. with probability μ, there is a shift from an allelic state 1 to 0 and with probability ν there is a shift in the opposite direction. In this work, we assumed $\mu = \nu$ although this is not true in general.

Migration: the migration step is implemented by choosing a random number, of individuals (drawn from a Poisson distribution with rate m_0) to be migrants from a fixed mainland pool assumed to be in deterministic mutation-selection balance. Allelic states are then randomly assigned to each migrant locus such that the frequency of the deleterious variants (among migrants) at each locus are close to the deterministic predictions for a single locus under mutation-selection balance. Following migration, both the frequency at each locus as well as the population size on the island change.

Reproduction+density dependent survival: This phase of the simulation involves gamete formation (through the process of meiosis and genetic recombination) as well as a decision as to how many offspring survive to be parents in the next generation. The latter is done by sampling an integer number, l say, from a Poisson distribution with rate $N'\overline{W}e^{r_0\left(1-\frac{N'}{K}\right)}$ where K is the carrying capacity of the population, N' is the population size after migration, $\overline{W}$ is the mean fitness of the population and r_0 the population growth rate. Once this is determined, l pairs of parents are then selected (with replacement) from the

population in proportion to their fitness to form offspring individuals. Each offspring gamete is formed by assuming that the allelic state contributed to the gamete at any given locus is sampled from the same or alternative chromosome in both parents with probability 0.5 (i.e. free recombination).

C.2 Semi-deterministic approximation

The semi-deterministic approximation outlined in the main text applies when the distribution of population sizes is sharply peaked around one or more values. For a given set of parameters Ks, Ku, h, $2LU = 2L(u/r_0)$, and m_0, we expect the peaks of the distribution to become sharper and the predictions of the semi-deterministic approximation to become more accurate for larger r_0K (corresponding to weaker demographic fluctuations) and larger L (corresponding to weaker fluctuations in genetic load). Figure C.1a illustrates this by plotting the distribution $P(N)$ (as obtained from allele frequency simulations) for various values of r_0K (which is changed by changing K). Note that in increasing K, while holding Ks, Ku and $2L(u/r_0)$ constant, we must simultaneously lower s and u, while increasing L proportionately.

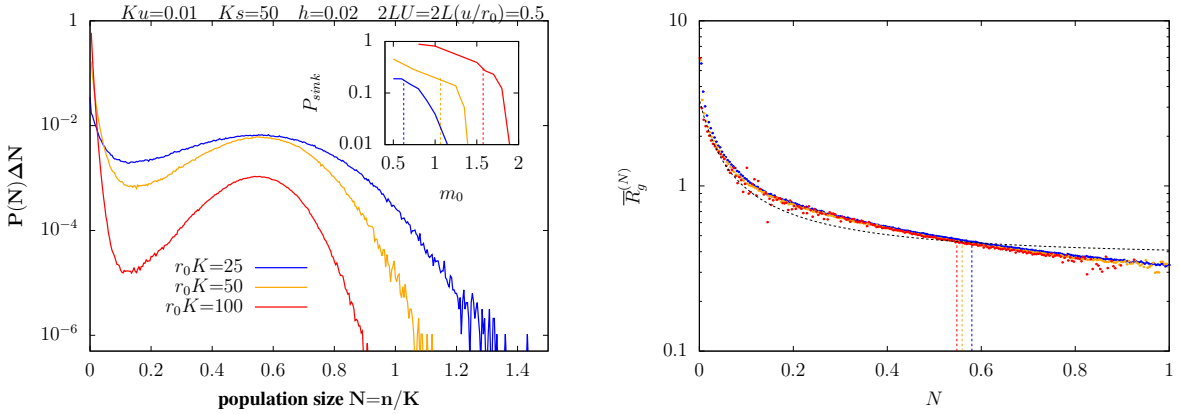


Figure C.1: (A) Distribution of population sizes $P(N)$ integrated over intervals of size $\Delta N = 0.004$ for $r_0K = 25, 50, 100$ (blue, orange, red), where r_0K is varied by changing K (along with s, u, L), while keeping Ku, Ks, $2LU = 2L(u/r_0)$ and r_0 constant. Inset: P_{sink}, the probability that the population is in the sink state, vs. m_0, the number of migrants per generation. Vertical dashed lines represent semi-deterministic predictions for the critical threshold $m_{c,1}$, at which the sink equilibrium vanishes. P_{sink} is calculated by integrating the distribution $P(N)$ from 0 to the minimum of the distribution (which lies between the sink and large-population equilibrium). (B) Average load $\overline{R_g^{(N)}} = \int R_g P[R_g, N]\, dR_g$ at a given N (points; different colours correspond to different r_0K) and the expected load under mutation-selection-drift equilibrium for that N (dashed line) as a function of N. Average load equals the equilibrium expectation at the predicted large-population equilibrium (indicated by vertical dashed lines) for all r_0K; the two quantities also match closely near $N = 0$ for larger values of r_0K. All results (solid lines in A and points in B) are from allele frequency simulations; dashed lines indicate various semi-deterministic predictions (as described above). Parameter values: $Ku = 0.01$, $Ks = 50$, $h = 0.02$, $2LU = 2L(u/r_0) = 0.5$ and $r_0 = 0.1$.

As expected, the peak corresponding to the large-population equilibrium becomes narrower,

i.e., typical fluctuations in N about the average (in that state) become smaller, as r_0K increases. The distribution of population sizes in the sink state also becomes more sharply clustered around 0, and the valley separating the large-population equilibrium from the sink state becomes deeper as r_0K increases.

We then ask whether the basic approximation underlying the semi-deterministic analysis– that genetic load is close to its equilibrium expectation under drift-mutation-migration-selection balance when populations are in one or other metastable state– becomes increasingly accurate for larger r_0K. Figure C.1b shows the average load $\overline{R_g^{(N)}} = \int R_g P[R_g, N]\, dR_g$ at a given N (as obtained from allele frequency simulations), as a function of N for various r_0K. The average load (points) is different from the load expected under mutation-selection-drift-migration balance $\mathbb{E}[R_g|N]$ (black line) except under two conditions: first, the average load equals the equilibrium expectation near the large-population equilibrium (indicated by vertical dashed lines) for all values of r_0K (different colours); second, the average load also matches the equilibrium expectation in the vicinity of $N = 0$ for large r_0K.

These observations are consistent with the general expectation that allele frequencies and genetic load will equilibrate only if population sizes remain roughly constant over the time scales required to reach mutation-selection-drift-migration balance: this condition can only be satisfied close to the peaks of the distribution (which correspond to equilibria of the semi-deterministic population size dynamics) and will typically not hold while populations transition between equilibria. A priori, it is unclear whether this condition is even satisfied in the sink state, in which N exhibits fluctuations that are large (relative to the mean) and characterised by significant skew towards small sizes. Figure C.1b suggests that it may nevertheless be reasonable to approximate the average load in the sink state by the equilibrium expectation, at least for large values of r_0K.

Accordingly, we note that the accuracy of the semi-deterministic prediction for the critical migration threshold, $m_{c,1}$, at which the sink state vanishes, improves with increasing r_0K. This is illustrated in the inset of fig. C.1a, which shows P_{sink}, the probability that the population is in the sink state (as observed in simulations) against m_0. The semi-deterministic prediction for $m_{c,1}$ (vertical dashed lines) approaches the corresponding threshold in simulations (where P_{sink} goes to 0) as r_0K increases.

Semi-deterministic prediction for equilibrium population size in the absence of migration. In the absence of migration (i.e., with $M_0 = 0$), there is always an equilibrium at $N = 0$. From eq. (4.3) of the main text, it follows that this equilibrium is stable of $1 - \mathbb{E}[R_g|N_*, M_0 = 0] < 0$. To compute the expected load in the $N \to 0$ limit, note that the expected frequency and the expected heterozygosity are respectively $1/2$

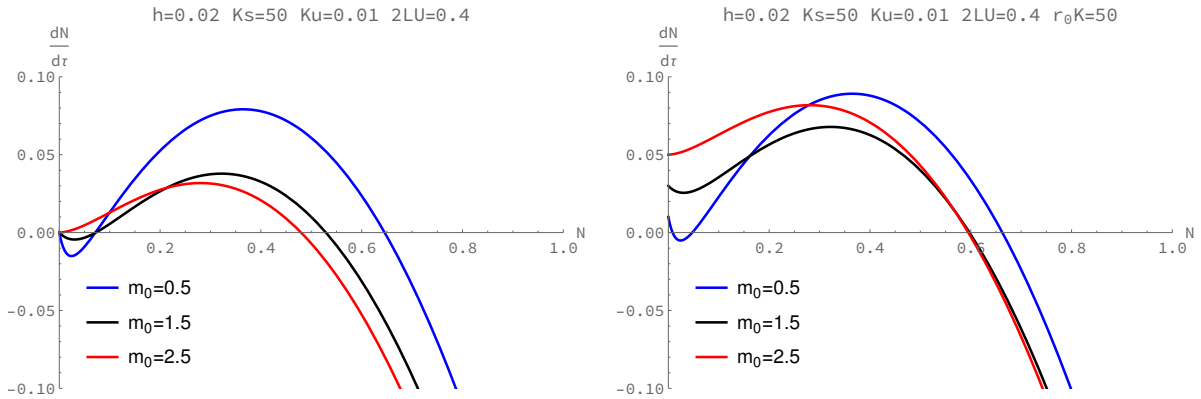
and 0 in this limit, so that $\mathbb{E}[R_g|N_*, M_0 = 0] = L(S/2)$. Thus, the extinction equilibrium is stable if $L(S/2) > 1$

There may exist a second equilibrium at $N_* = 1 - \mathbb{E}[R_g|N_*, M_0 = 0]$; the population size N_* is positive (i.e., the population not extinct) only if $\mathbb{E}[R_g|N_*, M_0] < 1$ for some $N_* > 0$, i.e., if the equilibrium genetic load is lower that the baseline growth rate.

Semi-deterministic prediction for $m_{c,1}$ in the $r_0K \to \infty$ limit. In order to obtain the semi-deterministic prediction for the population size, genetic load and expected allele frequencies at one or more equilibria, we numerically solve for the $\{N_*\}$ satisfying:

$$N_* \left(1 - N_* - \mathbb{E}[R_g|N_*]\right) + M_0 = 0 \text{ where } \mathbb{E}[R_g|N_*] = \sum_j S_j[\mathbb{E}(p_j|N_*) + (2h_j - 1)\mathbb{E}(p_j q_j|N_*))] \tag{C.1}$$

where the expectations are obtained by integrating over the equilibrium allele frequency distributions (eq. (4.2) of the main text). The equation above is the same as (4.3) of the main text. An equilibrium N_* is a stable equilibrium if the function $N\left(1 - N - \mathbb{E}[R_g|N]\right) + M_0$ has a negative derivative with respect to N at $N = N_*$.



(a) (b)

Figure C.2: (A)-(B) Rate of change of population size N under the semi-deterministic approximation (eq. (C.1)) as a function of N (A) in the $r_0K \to \infty$ limit (where the demographic effects of migration, represented by the M_0 term, can be neglected) (B) for finite r_0K (i.e., including the M_0 term). Equilibria correspond to points at which the curves cross the horizontal (zero growth rate axis); stable equilibria are those for which the curves have a negative slope at the point of zero crossing. Parameter values ($Ks = 50$, $Ku = 0.01$, $2LU = 0.4$ and $h = 0.02$ in A and B; $r_0K = 50$ in B) correspond to a regime where increasing migration causes the extinction fixed point to become unstable (in the $r_0K \to \infty$ limit; fig. A) or the sink fixed point to vanish (for finite r_0K; fig B).

In the limit $r_0K \to \infty$, $L \to \infty$, $s \to 0$, $u \to 0$, with m_0, Ks, Ku and $2LU = 2L(u/r_0)$ constant, the demographic contributions of migration, represented by the $M_0 = m_0/r_0K$ term in the above equation, can be neglected, yielding $N_*\left(1 - N_* - \mathbb{E}[R_g|N_*]\right) = 0$. This function (which must be zero at equilibrium) is plotted as a function of population size in fig. C.2a for three different levels of migration (which influence the function via the term $\mathbb{E}[R_g|N_*]$). There is always an equilibrium ($N = 0$) at extinction; this equilibrium

is stable if the curve is downward sloping at $N = 0$, i.e., if $1 - \lim_{N_* \to 0} \mathbb{E}[R_g|N_*] < 0$. In addition, there may be two other equilibria— one stable and the other unstable— satisfying $N_* = 1 - \mathbb{E}[R_g|N_*]$. For low $2LU$ and/or h not too low, an increase in migration causes the sink equilibrium to become unstable, so that above a critical migration threhsold $m_{c,1}$, only a single, stable 'large-population' equilibrium exists.

In the limit $r_0 K \to \infty$, the change in the stability properties of the sink equilibrium occurs at the migration rate for which $\lim_{N_* \to 0} \mathbb{E}[R_g|N_*] = 1$. Since $\lim_{N_* \to 0} \mathbb{E}[R_g|N_*]$, the expected load in the limit $N_* \to 0$, depends only on the neutral allele frequency and neutral heterozygosity under migration-drift balance, we can obtain an explicit expression that depends on LS, h, $p^{(m)}$ and m_0. Finally, setting $\lim_{N_* \to 0} \mathbb{E}[R_g|N_*] = 1$ and solving for the number of migrants per generation yields:

$$m_{c,1} = \frac{LSp^{(m)} - 1}{4[1 - LSp^{(m)}(p^{(m)} + 2h(1 - p^{(m)}))]} \tag{C.2}$$

which is eq. (4.4) of the main text.

In order to illustrate the effects of the demographic term on equilibria and their stability properties, we also plot $N\,(1 - N - \mathbb{E}[R_g|N]) + M_0$ (eq. (C.1) including the M_0 term) as a function of N (fig. C.2b). At low migration levels, there is a sink equilibrium (with N_* close to but not equal to zero) and a large-population equilibrium. Increasing migration now causes the sink equilibrium to vanish, rather than rendering it unstable. The critical threshold $m_{c,1}$ in this case is significantly lower than the $r_0 K \to \infty$ prediction above (see also fig. 4.3b of the main text), highlighting how, migration can influence population outcomes via both genetic and demographic effects.

C.3 Comparison of individual-based simulations with simulations under LE and IE

Here we compare results from the two simulation approaches by looking at the effect of migration on the mean population size and mean genetic load at equilibrium (fig. C.3a – C.3c) as well as on the stochastic distribution of population size (fig. C.3d – C.3f) on the island under different genetic architectures i.e. for weakly deleterious alleles ($Ks < Ks$; left column), mildly or moderately deleterious alleles ($Ks \gtrsim Ks_c$; middle column) and strongly deleterious alleles ($Ks \gg Ks_c$; right column).

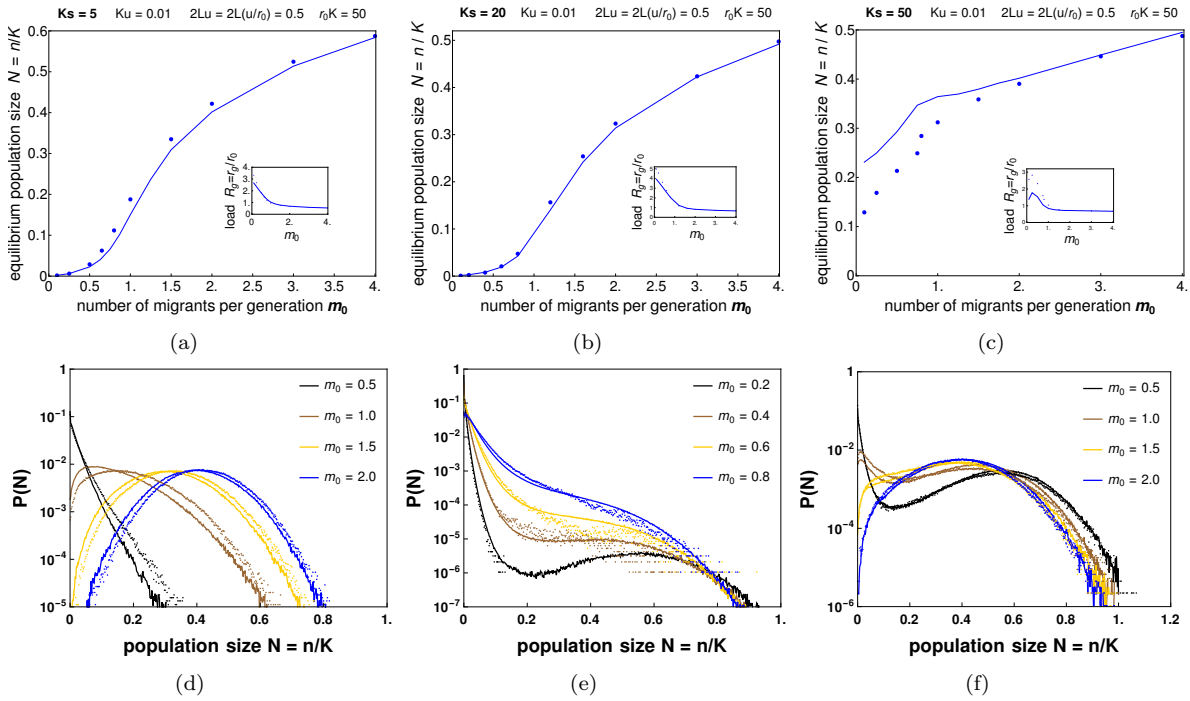


Figure C.3: (A)-(C) Mean population size (main figure) and mean genetic load (inset) at equilibrium plotted against m_0 (the number of migrants per generation) for weakly deleterious ($Ks < Ks_c$; left column), mildly deleterious ($Ks \gtrsim Ks_c$; middle column) and strongly deleterious ($Ks \gg Ks_c$; right column) nearly recessive alleles. Solid lines represent result obtained from allele frequency simulations (assuming LE and zero inbreeding) while dots represent result from individual-based simulations. (D)-(F) Equilibrium probability distribution of scaled population sizes $N = n/K$ for various values of m_0, as obtained from simulations assuming LE and IE (solid lines) as well as from individual based simulations (dots) under the three parameter regimes.

Weakly deleterious nearly recessive alleles (left column of fig. C.3): For $Ks < Ks_c$, the population approaches extinction for low migration. However, increasing migration causes an increase in the size of the island population (main plot of fig. C.3a) as well as a corresponding decrease in the genetic load (inset). This is true for both types of simulations although the individual based simulation (dots) produces slightly higher population numbers with increasing m_0. Looking at the stochastic distribution of population size (fig. C.3d), populations exhibit a single stable equilibrium for all migration rates and shift to the right towards increasing N (following a Gaussian distribution centered around the stable equilibrium) as m_0 increases. As noticed in fig. C.3a, there exists a slight difference between both simulation approaches, namely that the individual based simulation produces slightly higher numbers (see dotted plots in fig. C.3d). This suggests that LD can cause a somewhat higher increase in population size by its purging effect on sets of locally maladapted alleles.

Moderately deleterious nearly recessive alleles (middle column of fig. C.3): Just as with weakly deleterious alleles, an increase in migration increases the population size (main plot of fig. C.3b) and reduces the genetic load (inset) on the island, although we observe a higher load for low values of migration in the individual based simulation.

Looking at fig. C.3e, the individual based simulation breaks down for very low migration rates as it requires extremely long simulation runs (both time and memory consuming) to observe the peak at the large population state equilibrium (see LE simulations - solid lines).

Strongly deleterious nearly recessive alleles (right column of fig. C.3): For $Ks \gg Ks_c$, we observe an initial sharp increase in population size with m_0. Beyond $m_0 = 1$, the population size increases further but with a less steep slope. Similarly, for very low m_0, the genetic load in the population increases initially and then falls sharply with increasing m_0 after which it approaches a steady value. Interestingly, for low migration rates, individual based simulations slightly diminish the population size and increase the load (see blue dots) suggesting a negative effect of LD on strongly deleterious alleles when migration is low. From fig. C.3f, we observe a bimodal distribution of population size (with one peak close to extinction and the other close to the deterministic equilibrium) when migration is rare. However, with increasing migration, there is a shift from the bimodal to a unimodal population size distribution.

C.4 Evolutionary outcomes with a distribution of fitness effects

In the main text, we analyse scenarios where all deleterious variants have equal selective effects and dominance coefficients, and illustrate how migration may have qualitatively different effects, depending on the magnitude of the homozygous selection coefficient, dominance coefficient and the total mutation rate.

Here, we ask: what is the effect of migration on population outcomes when the genetic load is due to loci with a distribution of fitness effects? For the purposes of illustration, we consider a somewhat artificial distribution where a fraction α of loci are subject to deleterious mutations that are nearly recessive ($h_R = 0.02$) with scaled selection coefficient Ks_R, and the remaining fraction $1 - \alpha$ are additive ($h_A = 0.5$) with scaled selective coefficient Ks_A. Thus, the mutation targets for the two types of deleterious variants are $2L\alpha U$ and $2L(1 - \alpha)U$ respectively.

If recessive variants are weakly deleterious (Ks_R less than the corresponding critical threshold Ks_c), then their contribution to genetic load will decrease with increasing migration. Since migration always reduces additive load (though only marginally for $Ks_A \gg 1$), the net effect of increasing migration in this case is to reduce load and increase the equilibrium population size (results not shown). We thus focus on recessive alleles with moderate or strongly deleterious effects ($Ks_R > Ks_c$), where an increase in recessive load with increasing migration can potentially counteract a reduction in additive load.

Figures C.4a-C.4c show semi-deterministic predictions for equilibrium population sizes versus m_0 for various combinations of Ks_A and Ks_R for $\alpha = 0.1$ (predominantly additive load), 0.5 (nearly equal contributions of additive and recessive alleles to load) and 0.9 (load primarily due to recessive alleles). Where two stable equilibria exist, these are shown by solid and dashed lines (of the same color).

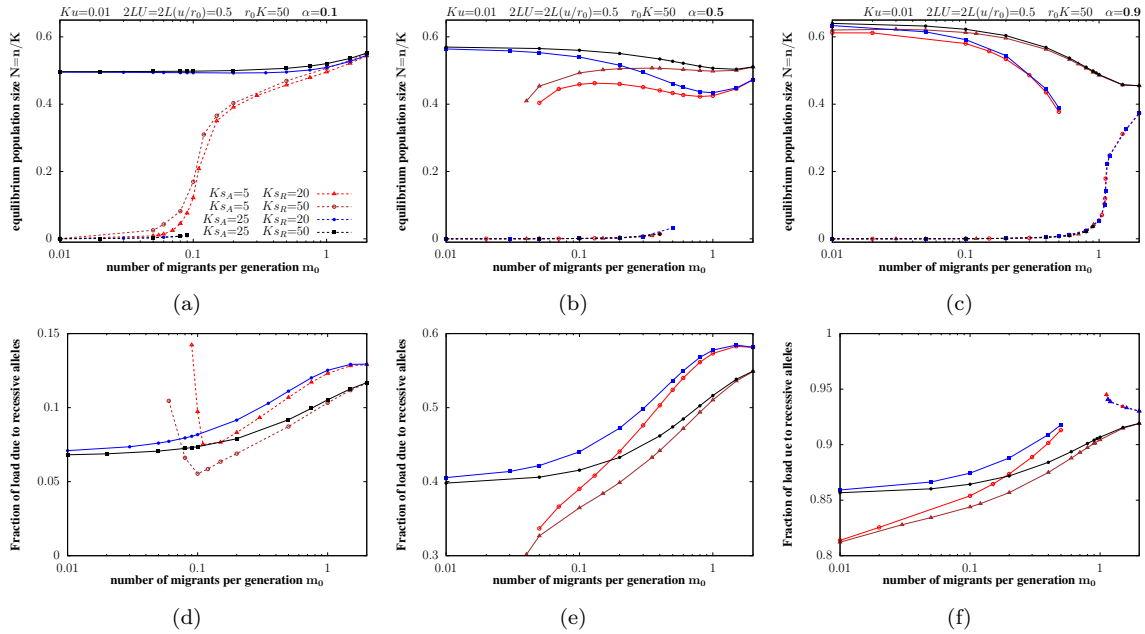


Figure C.4: (A)-(C) Semi-deterministic predictions for population size(s) at equilibrium vs. m_0, the number of migrants per generation. Different colors correspond to different values of Ks_A and Ks_R, the (scaled) homozygous selection coefficients for the additive ($h = 0.5$) and nearly recessive ($h = 0.02$) alleles (see legend of fig. A). We depict cases where a fraction α of alleles have nearly recessive effects and the remaining fraction $1 - \alpha$ additive effects, for $\alpha = 0.1$ (left), $\alpha = 0.5$ (middle) and $\alpha = 0.9$ (right). Where two alternative equilibria exist, these are shown by solid and dashed lines (of the same color). (D)-(F) Semi-deterministic predictions for the fraction of the total load that is due to recessive alleles vs. m_0, for parameter combinations where the total load is less then r_0, i.e., where the population is not a sink. Where two equilibria with load less than r_0 exist, the fractions corresponding to each are depicted by solid and dashed lines. All figures show results for: $Ku = 0.01$, $2LU = 0.5$, $r_0K = 50$.

We observe a genetic Allee effect (characterised by the co-existence of alternative 'sink' and 'large-population' equilibria) at low migration rates for all parameter combinations, except when load is primarily due to weakly deleterious additive alleles (red and brown curves in fig. C.4a). Increasing migration tends to destabilize the sink state in all cases, except where load is largely due to recessive alleles with moderately deleterious effects (red and blue curves in fig. C.4c): in this extreme limit, it is the large-population state that vanishes at high migration levels (see also fig. 4.2b, main text). The critical migration threshold at which the sink state vanishes (and Allee effects no longer occur) is higher when the contribution of recessive alleles to total load is higher (α larger) and homozygous effects associated with recessive alleles more moderate. The population size associated with the large-population equilibrium increases marginally with increasing migration for small α (as expected when load is predominantly additive), decreases with increasing

migration for large α (as expected when load is predominantly recessive), and exhibits a weak, non-monotonic dependence on m_0 for intermediate α (reflecting the opposing effects of migration on additive and recessive alleles).

We also track how the relative contributions of additive and recessive alleles to total load change with increasing migration. Figures C.4d-C.4f shows the fraction of total load that is due to recessive alleles vs. m_0, for parameter combinations where the population is not a sink, i.e., where the total load is less than the baseline growth rate r_0. This fraction increases with increasing m_0 in all cases (except when the population is close to a sink state, i.e., $R_g \sim 1$), with the most significant increase occurring for $\alpha = 0.5$, and where additive alleles have weak effects and recessive alleles moderate effects. Under these conditions, smaller isolated populations (with low migration) can more efficiently purge recessive load, but maintain higher levels of additive load. An increase in migration tends to decrease the additive load and increase the recessive load by pushing the frequencies of deleterious additive and recessive alleles closer to the mainland values (that are respectively lower and higher than those on the island).

Note that in this case, the total load and population size do not change significantly with migration (because of the opposing effects of migration on the two components of load). The change in the relative contributions of different kinds of alleles may nevertheless have significant consequences, by making populations with higher levels of migration (and accordingly, a greater number of segregating recessive alleles) more vulnerable to future extinction, e.g., in the event of a bottleneck.

Appendix D

Genetic load, eco-evolutionary feedback and extinction in a metapopulation

D.0.1 How events in the life cycle of individuals change patch size and load.

To account for the effect of migration on population size, we first determine the net number of migrants, $netM(t)$ (i.e., the difference between the total number of immigrants and emigrants in a given patch) and add this to the existing patch size. Under soft selection, we assume a zero net migration rate (i.e., a balance between the number of immigrants and emigrants) so as to keep the patch size fixed. However, under hard selection, the net number of migrant is assumed non-negative. Migration therefore changes the population size according to $N_i'(t) = N_i(t)$ (under soft selection) and $N_i'(t) = N_i(t) + netM(t)$ (under hard selection). Similarly, to account for the effect of migration on allele frequencies in any patch, we add the net number of migrant alleles, $netp(t)$ to the existing allele copy number in the patch and divide this by the population size $N_i'(t)$. In essence, migration changes allele frequencies according to $p_{i,j}'(t) = (N_i p_{i,j}'(t) + netp(t))/N'(t)$.

With regards to the effect of mutation, we assume equal mutation rate to and from the deleterious allele so that mutation changes allele frequency according to $p_{i,j}''(t) = p_{i,j}'(t) + \mu(1 - 2p_{i,j}'(t))$. However, mutation has no effect on population size (under hard selection) so that $N_i''(t) = N_i'(t)$.

Following mutation, adults in each patch mate to produce offspring that survive to be next generation parents. Under soft selection, the mating process in each patch involves sampling with replacement, $N_i(t)$ pair of individuals based on their fitness and freely recombining their gametes to form offspring gametes. This means that we do not explicitly

distinguish between the male and female sexes and there is also the possibility of self mating. Under hard selection on the other hand, load and density-dependent regulation changes population size according to $N_i'''(t) = N_i''(t)e^{1-N_i''(t)-R_g}$ and demographic stochasticity (randomness in birth and death) is further imposed on the population to determine how much survivable offspring, $N_i''''(t)$ can be formed. The latter is achieved by sampling from a Poisson distribution with parameter $N_i'''(t)$. Offspring are then formed by choosing with replacement $N_i''''(t)$ pairs of parents also based on their fitness and freely recombining their gametes. For both soft and hard selection, selection changes allele frequency according to $p_{i,j}'''(t) = (p_{i,j}''(t)\ e^{-s} + q_{i,j}''(t)\ e^{-hs})/\overline{w}$ where $\overline{w} = p_{i,j}''(t)^2\ e^{-s} + 2p_{i,j}''(t)\ q_{i,j}''(t)\ e^{-hs} + q_{i,j}''(t)^2$ is the mean fitness in the i^{th} patch and $q_{i,j}''(t) = 1 - p_{i,j}''(t)$.

Finally, the allele frequency at the end of the generation, i.e., $p_{i,j}''''(t)$ is obtained by sampling $p_{i,j}'''(t)$ from a Binomial distribution with parameters, $N''''(t)$, $p_{i,j}'''(t)$ thus accounting for genetic drift.

D.0.2 'Moderate selection' approximation under soft selection

In principle, any quantity such as load, F_{ST}, etc. for a single locus can be computed from the equilibrium allele frequency distribution $\psi[p|\overline{p}]$, by first numerically solving $\overline{p} = \int p\ \psi[p|\overline{p}]\ dp$ to obtain the mean allele frequency $\overline{p}$, then plugging this into the equilibrium distribution and finally integrating over the distribution to obtain higher moments. However, it is also useful to consider an alternative approach based on equations for moments (or cumulants) of the allele frequency distribution.

In general, any moment will depend on higher moments, resulting in a set of recursions that is not closed. Approximations thus rely on closing this set of recursions in different ways, depending on assumptions about the relative magnitudes of Ks (or Ksh) and Km (Whitlock (2002); Glémin et al. (2003); Roze (2015)). Here, we introduce another such moment closure approximation, which applies also for recessive ($h = 0$) alleles and intermediate selection coefficients ($Ks \sim 1$). We also attempt to discuss the biological meaning of the underlying assumptions.

As in the main text, let p denote the frequency of the deleterious allele at a given locus in a given deme and $\overline{p}$ the mean across all demes in the population. We denote the expected change in allele frequency per unit time by $M(p)$ and the variance of the change by $V(p)$. These are:

$$M(p) = -s(hp + (1-3h)p^2 - (1-2h)p^3) + u(1-2p) + m(\overline{p} - p) \tag{D.1a}$$

$$V(p) = \frac{p(1-p)}{2K} \tag{D.1b}$$

where K is the number of individuals per deme. Under the diffusion approximation, the

expectation (denoted by $\mathbb{E}_\psi[...]$) of any function $f(p)$ of the allele frequency p over the allele frequency distribution $\psi[p]$ satisfies (see also Ohta and Kimura (1971)):

$$\frac{d\,\mathbb{E}_\psi[\,f(p)\,]}{d\,t} \;=\; \mathbb{E}_\psi\left[M(p)f'(p) + \frac{1}{2}V(p)\,f''(p)\right] \tag{D.2}$$

Setting $f(p)$ to be p and p^2 yields the following equations for the first and second moments of the allele frequency distribution respectively (see also Whitlock (2002); Glémin et al. (2003)):

$$\frac{d\,\mathbb{E}[\,p\,]}{d\,t} \;=\; -s\left(h\mathbb{E}[p] + (1-3h)\mathbb{E}[p^2] - (1-2h)\mathbb{E}[p^3]\right) + u\,(1-2\mathbb{E}[p]) \tag{D.3a}$$

$$\begin{aligned}\frac{d\,\mathbb{E}[\,p^2\,]}{d\,t} \;=\; -s\left(h\mathbb{E}[p^2] + (1-3h)\mathbb{E}[p^3] - (1-2h)\mathbb{E}[p^4]\right) + u\left(\mathbb{E}[p] - 2\mathbb{E}[p^2]\right)\\ + m\left(\mathbb{E}[p]\,\overline{p} - \mathbb{E}[p^2]\right) + \frac{\mathbb{E}[p] - \mathbb{E}[p^2]}{4K}\end{aligned} \tag{D.3b}$$

It will be useful to express the above equations in terms of appropriately scaled cumulants (rather than moments) of the allele frequency distribution:

$$F_{ST} = \frac{\mathbb{E}[p^2] - \overline{p}^2}{\overline{p}(1-\overline{p})} \quad \gamma \;=\; \frac{\mathbb{E}[p^3] - 3\overline{p}\,\mathbb{E}[p^2] + 2\overline{p}^3}{\overline{p} - 3\overline{p}^2 + 2\overline{p}^3} \quad \kappa \;=\; \frac{\mathbb{E}[p^4] - 4\overline{p}\,\mathbb{E}[p^3] + 6\overline{p}^2\,\mathbb{E}[p^2] - 3\overline{p}^4}{\overline{p} - 4\overline{p}^2 + 6\overline{p}^3 - 3\overline{p}^4} \tag{D.4}$$

Here, F_{ST}, γ and κ denote respectively the variance, skew and kurtosis of the frequency distribution within any deme, scaled by the corresponding cumulant calculated for the distribution of alleles across the entire population.

At equilibrium, moments of the allele frequency distribution are constant in time, i.e., $d\,\mathbb{E}[\,p\,]/dt \;=\; d\,\mathbb{E}[\,p^2\,]/dt \;=\; 0$; further, $\mathbb{E}[p] = \overline{p}$. Combining equations (D.3) and (D.4), and assuming deleterious alleles to be sufficiently rare *overall* in the population that $\mathbb{O}[\overline{p}^2]$ terms can be neglected, we have at equilibrium:

$$0 \;=\; \frac{u}{s} - \left[\,h + F_{ST}(1-3h) - \gamma(1-2h)\,\right]\overline{p} \tag{D.5a}$$

$$0 \;=\; [1 - F_{ST}(1+4Km) - 4Ks(hF_{ST} + (1-3h)\gamma - (1-2h)\kappa]\overline{p} \tag{D.5b}$$

This *pair* of equations is underdetermined as it involves *four* variables $\overline{p}$, F_{ST}, γ and κ. To obtain an approximate solution, we further assume that third and higher cummulants of the allele frequency distribution, i.e., the skew and kurtosis are related to F_{ST} as in the *neutral* infinite-island model. Note that F_{ST} itself is not taken to be neutral (or unaffected

by selection), as is assumed by Whitlock (2002), but only that:

$$\gamma = \frac{2F_{ST}^2}{1+F_{ST}} \qquad \kappa \approx \frac{6F_{ST}^3}{(1+F_{ST})(1+2F_{ST})} + \mathbb{O}(\overline{p}) \tag{D.6}$$

Substituting into eq. (D.5) and expressing load as $G = s[\overline{p} - (1-2h)(1-F_{ST})\overline{p}\,\overline{q}] \approx s\overline{p}[2h + (1-2h)F_{ST}]$, we obtain:

$$\overline{p} = \frac{u}{s}\frac{1+F_{ST}}{(1-F_{ST})(h+(1-h)F_{ST})} \qquad G = u\frac{(1+F_{ST})(2h+(1-2h)F_{ST})}{(1-F_{ST})(h+(1-h)F_{ST})} \tag{D.7a}$$

$$\text{where} \quad (1+F_{ST})(1+2F_{ST})(1-F_{ST}(1+4Km)) - 4KsF_{ST}(1-F_{ST})(h+2F_{ST}(1-h)) = 0 \tag{D.7b}$$

Equation (D.7b) is cubic in F_{ST} and can be solved (e.g., numerically); this solution for F_{ST} can then be substituted into eq. (D.7a) to obtain the average deleterious allele frequency $\overline{p}$ and load G. Thus, in essence, our approximation allows for selection that is strong enough to change F_{ST} but not the relationship between higher cummulants (γ, κ etc.) and F_{ST}.

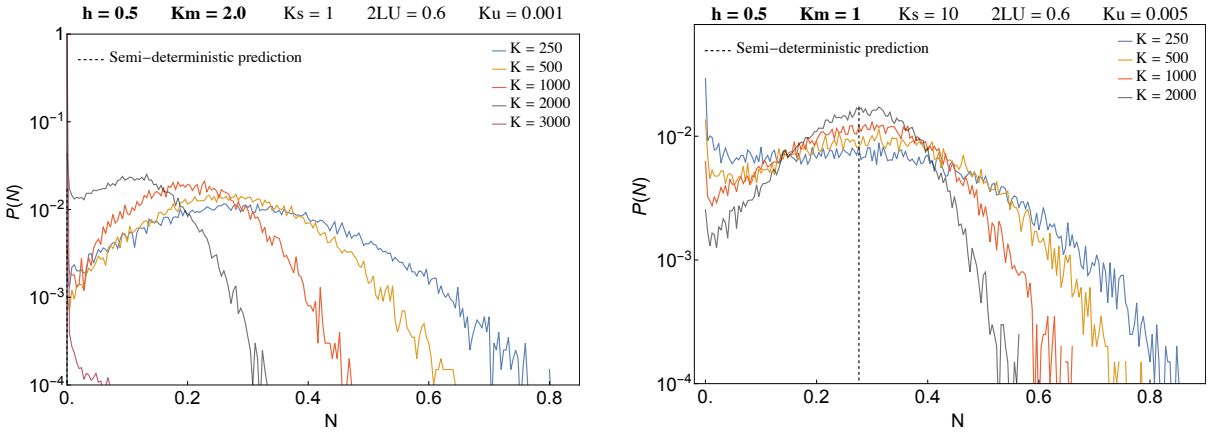
It is useful to juxtapose this approximation with those used in earlier work. Whitlock (2002) assumes that not just the relationship between higher cummulants (γ, κ etc.) and F_{ST}, but also F_{ST} is unaffected by selection and is essentially neutral, i.e., equal to $1/(1+4Km)$, where K is the population size per deme. Thus, his is a *weak selection* approximation and applies when $Ks \ll 1$, so that the term involving Ks in eq. (D.7b) (or equivalently in eq. (D.5b)) can be neglected. On the other extreme, Roze (2015) and Glémin et al. (2003) consider a parameter regime where the local deme size K is large enough that allele frequency distributions are essentially concentrated around the mean $\overline{p}$, which is relatively low. In practice, this means that F_{ST} is small enough that $\mathcal{O}(F_{ST}^2)$ terms in eq. (D.7b) can be neglected, which gives: $F_{ST} \approx \frac{1}{4Km+4Khs}$. Theirs is thus a *strong selection, strong migration* approximation; in practice, it applies for $Ksh + Km$ greater than 5. By contrast, the approximation for F_{ST} introduced above applies also for intermediate Ks (see fig. 5.1d) and is thus a *moderate selection* approximation: in essence, it captures the key effect of selection which is to change coalescence times within and between demes (to *different* extents) without necessarily changing the extent to which the branching structure of genealogies is (a-)symmetric. This is, however, only a rough interpretation, and a more rigorous analysis will be required to fully justify such approximations and more generally understand intermediate Ks regimes where neither selection nor drift can be treated as minor perturbations (to neutral or deterministic predictions respectively).

Relaxing the low mean allele frequency ($\overline{p} \ll 1$) assumption. All the approximations described above assume that the expected deleterious allele frequency is sufficiently low (i..e, deleterious alleles are very far from global fixation) that $\mathcal{O}(\overline{p}^2)$ terms can be neglected. When this is no longer true (but if (D.6) still applies, i.e., if third and higher cumulants depend on F_{ST} as in the neutral model), equations (D.3), (D.4) and (D.6) together yield a cubic equation for $\overline{p}$:

$$\frac{u}{s}\left(\frac{1+F_{ST}}{1-F_{ST}}\right) = \left[F_{ST} + h(1-F_{ST}) + \frac{2u}{s}\left(\frac{1+F_{ST}}{1-F_{ST}}\right)\right]\overline{p} + [1 - 2F_{ST} - 3h(1-F_{ST})]\overline{p}^2 - (1-F_{ST})(1-2h)\overline{p}^3 \tag{D.8}$$

Since deleterious allele frequencies are expected to be high only for very weak selection, we can typically neglect the effects of selection on F_{ST} in this parameter regime, and simply solve equation (D.8) for $\overline{p}$ under the assumption $F_{ST} \approx \frac{1}{1+4Km}$.

D.0.3 Demographic effect of migration in small and large patches and dependence on Ks

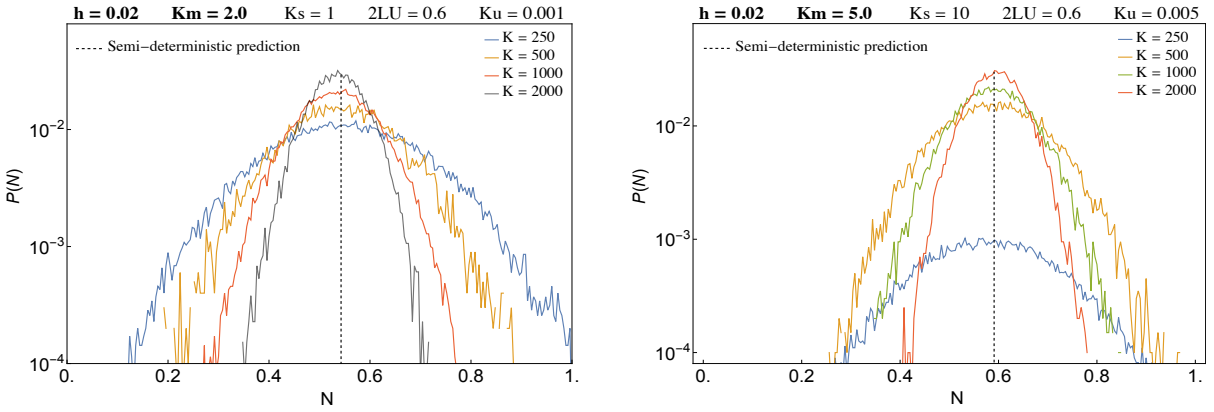


(a) (b)

Figure D.1: Equilibrium distribution of population sizes with additive alleles ($h = 0.5$) for different values of K and near Km_c. (a.) represents $Km = 2.0$ and $Ks = 1$ (b.) represents $Km = 1.0$ and $Ks = 10$. The black vertical line represents the semi-deterministic prediction.

Just like for recessive alleles (i.e., $h = 0.02$ in fig. 5.3c and fig. 5.3d), we see that with $Ks = 1$ (fig. D.1a, small-sized patches (e.g., $K = 250$) benefit from the demographic effect of migration and are more stable. With increasing K, we observe a bimodal distribution of population sizes with most of the weight close to $N = 0$ (e.g., see $K = 2000$ (black line)) as many of the patches are already extinct and the remaining non-extinct patches lie at a fraction of carrying capacity. With much larger K however, e.g., $K = 3000$ (brown line) all the demes already go extinct. In the case of $Ks = 10$, larger populations are more stable.

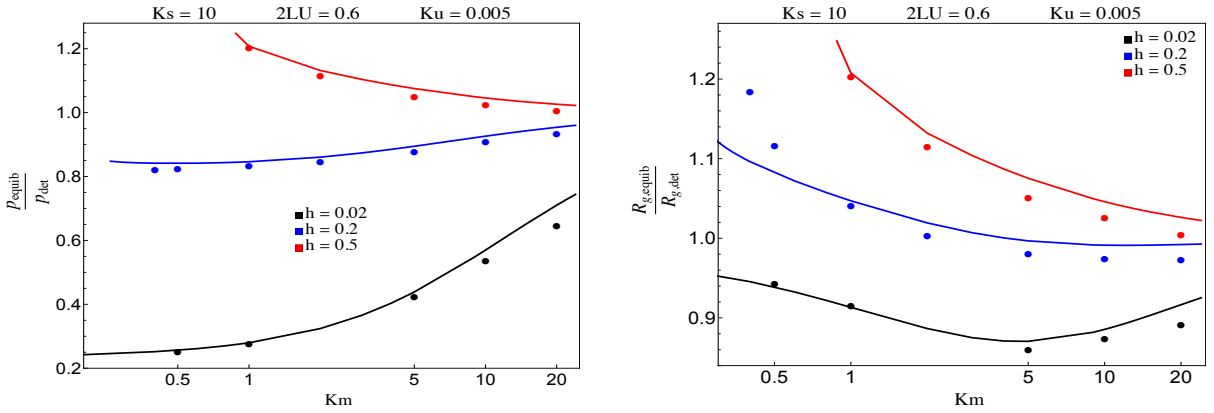
D.0.4 Weak dependence of simulation results on K beyond Km_c



(a) (b)

Figure D.2: Beyond the critical migration threshold Km_c, in particular for large Km population size distribution peaks at the semideterministic expectation even for low K. (a.) $Ks = 1$ and $Km = 2.0$ (b.) $Ks = 10$ and $Km = 5.0$.

D.0.5 Effect of gene flow on equilibrium allele frequency and equilibrium mean load for $Ks = 10$.



(a) (b)

Figure D.3: (a.) and (b.) are respectively the equilibrium allele frequency and load (relative to that in an undivided population) plotted against the level of gene flow, Km. Filled circles show our simulation results (using allele frequency simulations) and solid lines show results from our semi-deterministic approximation. Simulations are run with 100 demes, with $K = 2000$, $L = 12000$ and $r_0 = 0.1$.

D.0.6 Exploring the role of asymmetric mutation on critical migration thresholds

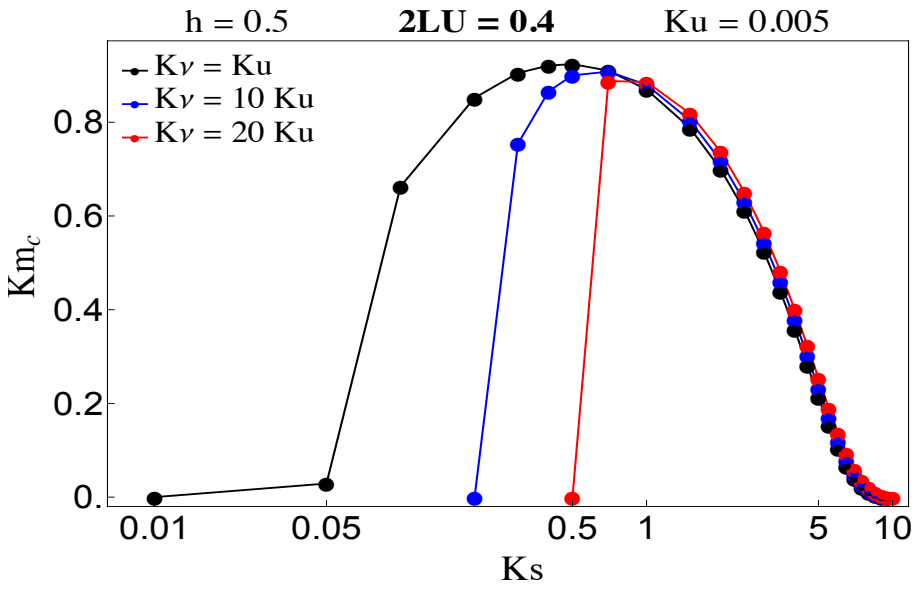


Figure D.4: Critical migration thresholds below which the metapopulation goes extinct accounting for asymmetric mutation rates and for $2LU = 0.4$.

Here, we explore the effect of asymmetric mutation (i.e., mutation biased towards generating more deleterious alleles) on the critical migration threshold necessary to prevent the extinction of the metapopulation. We do this for additive alleles (i.e., $h = 0.5$) and with $2LU = 0.4$. Looking at fig. D.4, we observe some interesting dynamics. First, we see that for weak to mildly deleterious effects ($Ks \leq 1$), the critical migration threshold is higher with increasing degree of mutational bias (lhs of fig. D.4). This makes sense as the more mutation is biased towards the formation of deleterious alleles, the more the number of deleterious alleles we have segregating in the population and the higher the load. Thus, a higher migration rate is needed to bring in new variation and alleviate this burden of load. In contrast, with moderate to strong selective effects, this difference in the critical migration threshold owing to mutational bias disappears.

D.0.7 Distribution of fitness effects and dominance coefficient

Here, we explore the impact of gene flow on the mean population size in the metapopulation when we have a distribution of fitness effects and dominance coefficients. To do this, we assume that a fraction a of loci are additive and have fitness effects Ks_A and the remaining fraction, $(1 - a)$ of loci are recessive with fitness effects Ks_R as shown below,

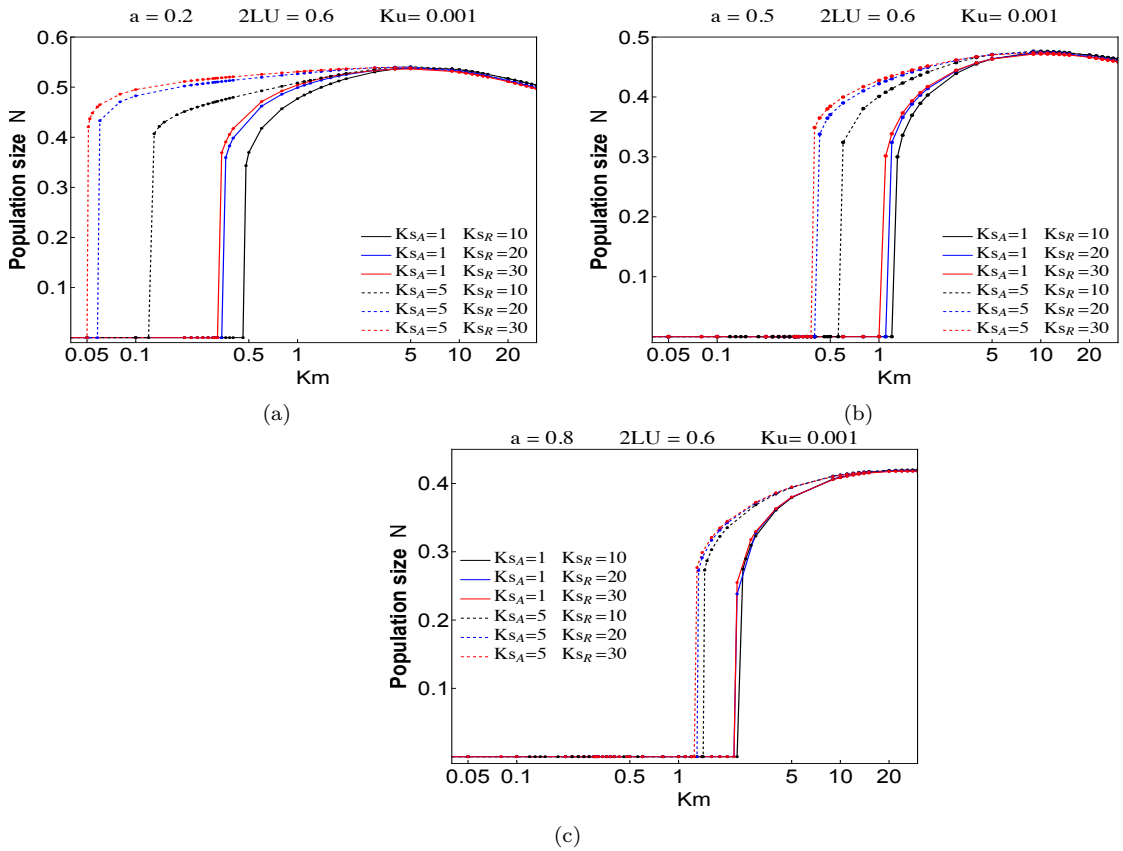
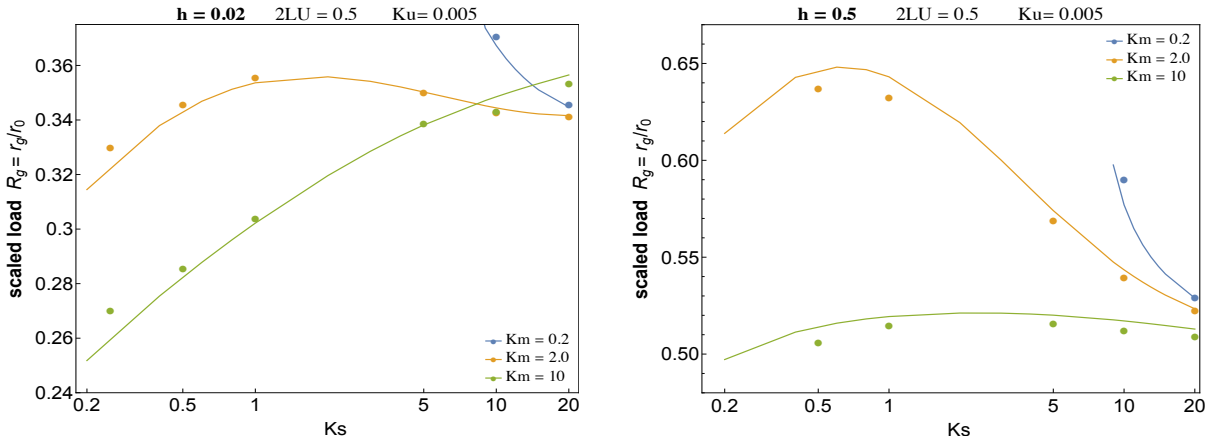


Figure D.5: Average population size across the metapopulation plotted against Km with (a.) $a = 0.2$ i.e., 20% of loci are additive and 80% are recessive (b.) $a = 0.5$ i.e., with equal proportion of additive and recessive alleles (c.) $a = 0.8$ i.e., with 80% of loci being additive and 20% recessive.

We see from fig. D.5a - D.5c that independent of the value of a, when additive alleles are nearly neutral and recessive alleles are non-neutral, increasing the selective effect of the recessive alleles, i.e., making them more strongly deleterious has little or no effect on the critical migration threshold above which the metapopulation survives. This also holds for the case where the additive alleles are mildly deleterious and occupy a higher proportion of loci (dashed lines in fig. D.5c). On the other hand, when additive alleles occupy a lower or equal proportion of loci as the recessive alleles, we see a somewhat different dynamics. We observe a (slightly) higher critical migration threshold when additive and recessive alleles are mildly deleterious (black dashed line in fig. D.5a and D.5b). As the recessive alleles become moderately deleterious, the threshold reduces and increasing the selective effect further makes little or no effect on the threshold migration rate (compare blue and red dashed lines in both fig. D.5b and D.5c). Finally, for all Ks_A, Ks_R combinations considered, we see the critical migration threshold at least doubling as a increases.

D.0.8 Alleles that contribute most to load



(a) (b)

Figure D.6: Plot of scaled load against Ns for (a.) recessive alleles (b.) additive alleles. Solid lines represent the semi-deterministic prediction and filled circles represent simulation results.

The above plot shows that under our model of hard selection, when we have intermediate levels of connection in the metapopulation (i.e., with $Km = 2$, orange color), alleles with mildly deleterious effects contribute most to load.

References

Glémin, S., Ronfort, J., and Bataillon, T. (2003). Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics*, 165(4):2193–2212. https://doi.org/10.1093/genetics/165.4.2193.

Ohta, T. and Kimura, M. (1971). Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics*, 68(4):571. https://doi.org/10.1093/genetics/68.4.571.

Roze, D. (2015). Effects of interference between selected loci on the mutation load, inbreeding depression, and heterosis. *Genetics*, 201(2):745–757. https://doi.org/10.1534/genetics.115.178533.

Whitlock, M. C. (2002). Selection, Load and Inbreeding Depression in a Large Metapopulation. *Genetics*, 160(3):1191–1202. https://doi.org/10.1093/genetics/160.3.1191.