UNIVERSITY OF OXFORD



PROBABILITY MODELS OF POPULATION STRUCTURE AND GENE FLOW

by

Aaron S. A. Smith

Magdalen College, Oxford

This thesis is submitted for the degree of

Doctor of Philosophy

Hilary Term 2021

DECLARATION

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original. This thesis is submitted to the University of Oxford for the degree Doctor of Philosophy, and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Aaron S. A. Smith
February 2021

ABSTRACT

There is a substantial interest in the relationship between population structure and the patterns of genetic variation. In *structured populations*, these patterns heavily depend on *gene flow*, the movement of genetic material from one population to another.

In spatially structured populations, gene flow can be quantified by the spatial motion of *ancestral lineages* — the description of where the ancestors of modern individuals lived at times in the past. Understanding the stochastic motion of these lineages is key to explaining patterns of genetic variability. In Chapter 1 we describe a new model for a population evolving on $\mathbf{Z}^d$ in which the demographic parameters of subpopulations depend on the population density in a local neighbourhood. After suitable scaling, tracing numbers of adults in the population we arrive at a system of interacting SDEs. The motion of a lineage ancestral to an individual sampled from such a population follows a random walk in a random environment and we establish this motion in the scaling limit. In contrast to classical models in which ancestral lineages are drawn into regions of high population density, in our setting, a lineage is attracted to regions of high *fecundity*.

In Chapter 2 we study the movement of ancestral lineages in a model in which local population sizes are determined by the presence of transient, patchy resources. Habitats are either *bad* or *good* and population sizes in good habitats are so large that, given the opportunity, lineages will always move from bad to good habitats. Lineages traced back through time can be explicitly described and we highlight the role of the spatial and temporal autocorrelation structure in determining the behaviour of lineage movement. The typical displacement between an individual and their ancestors may differ substantially from that predicted based on typical offspring dispersal.

Population structure need not be spatial; eg in some species, mating is restricted to occurring between individuals from distinct self-incompatible classes, termed *mating types*. Gene flow between these classes occurs only at the times of rare bouts of sexual reproduction. The evolutionary mechanism that drove the establishment of self-incompatibility is still a debated topic. In a new approach, in Chapter 3 we address this by quantifying the population genetic diversity and the probability of 'effectual sex' as a function of the number of self-incompatible classes. Our results permit a novel understanding of the early origins of sex.

ACKNOWLEDGEMENTS

I owe a great deal of thanks to the many people who have shaped my time at Oxford throughout these last four years.

I would like to extend my deepest gratitude to my supervisor, **Alison Etheridge**, who has been a better supervisor than I could have possibly imagined when starting in Oxford four years ago. I am incredibly grateful to Alison for her ever-present commitment to my education and projects: despite her many, many other commitments and responsibilities, Alison has always been available for guidance, advice and often-needed support. I have learnt a great deal from Alison: not just mathematics, but in how to be a good leader, mentor and friend. I will consider it a success if I can emulate this throughout my own career, even in part. Most importantly, working with Alison has been a huge amount of fun.

It transpired, despite much nervousness, that my viva would be one of the most enjoyable parts of my DPhil. It was a genuine pleasure to have the opportunity to discuss my work with specialists who had truly taken the time to read my thesis and contemplate it in depth. I thank my examiners, **Julien Berestycki** and **Dario Spanò**, for their thoughtful comments, questions, and for taking the time to prepare a list of corrections.

The Oxford Probability Group has been an incredible place to work and, in my view, its commitment to fostering a friendly and encouraging environment with genuine academic exchange between both faculty and students is rivalled by no other. A special thank you goes to officemates and near-officemates past and present — **James Ayre**, **Kimberly Becker**, **Serte Donderwinkel**, **Mitch Gooding**, **Mikołaj Kasprzak**, **Hamish Patten**, **David Rindt**, **Frederik Sørensen**, **Roman Stasiński** and **Dominic Yeo** — who have been constant in their encouragement, humour and, perhaps frustratingly, supply of counter examples to my wishful mathematical thinking. It is a great tragedy that the timing of the COVID-19 pandemic has brought my time in the group to an unfortunate and early close. To James, Kim, Serte, David, Freddie and Romek, I wish you the best of luck with the remainder of your degrees. Further afield, I would also like to thank my collaborators **George Constable**, **Thomas Kurtz**, **Ian Letter**, **Sarah Penington**, **Peter Ralph** and **Daniel Wilson** for their patience, their enthusiasm for our projects, and everything that they have taught me.

That I have made it through this last year of thesis-writing, house-moving, new-job-beginning, viva-preparing and thesis-correcting relatively unscathed — all in the context of the pandemic — is a testament to the unquestioning support and encouragement of **Hugo Paquet**, and, as ever, I am incredibly grateful to him for keeping me happy, healthy and smiling. More than anything, I thank him for his patience, and I am looking forward to being more present now that I have reached the end.

Finally, I thank **Stephen and Suzanne Smith** without whose enduring hard work, sacrifices and commitment to my education I surely would not be writing this thesis today.

Oxford
February 2021

CONTENTS

DECLARATION 3

ABSTRACT 5

ACKNOWLEDGEMENTS 7

INTRODUCTION 13

1 ANCESTRAL LINEAGES IN A POPULATION MODEL WITH NONLOCAL DISPERSAL OF JUVENILES 25

1.1 Introduction 25

1.2 A discrete time and space model for an evolving population 27

1.2.1 The individual based model 27

1.2.2 Main results 30

1.3 Scaling limit for X^n 40

1.3.1 C-tightness of X^n in the product topology 40

1.3.2 Proof of Theorem 1.2.5 52

1.3.3 Proof of Theorem 1.2.6 56

1.3.4 Uniqueness of solutions 58

1.4 Ancestral lineages 59

1.4.1 Construction 60

1.4.2 Approximation of the probabilities s^n 62

1.4.3 Skorokhod representation of (X^n, L^n) 65

1.4.4 Tightness of $\hat{L}^n(\omega, \cdot)$ 66

1.4.5 Proof of Theorem 1.2.10 68

1.5 Central limit theorem for X^n 71

1.5.1 C-tightness of Z^n in the product topology 72

1.5.2 Proof of Theorem 1.2.14 79

1.5.3 Proof of Theorem 1.2.15 84
1.5.4 Proof of Propositions 1.2.16, 1.2.18, and 1.2.19 84
1.5.5 The effect of small fluctuations on ancestral lineages 88
1.6 Appendix 89
1.6.1 Burkholder's predictable square function inequality 89
1.6.2 A discrete local limit theorem 90
1.6.3 Spectra of Laurent operators 90

2 GENE FLOW IN PATCHY HABITATS 91
2.1 Introduction 91
2.2 The model 94
2.2.1 Patchy habitats 94
2.2.2 Temporally uncorrelated patchy habitats 96
2.2.3 A first look at temporally correlated patchy habitats 97
2.2.4 Continuous-time dynamics 98
2.2.5 Temporally correlated patchy habitats 103
2.3 Patch geometry under diffusive rescaling 106
2.3.1 The rescaled parameters 106
2.3.2 Appearance of new patches 108
2.3.3 The geometry of a single patch 110
2.4 Proof of Lemma 2.2.1 112
2.5 Convergence to the continuous-time process 113
2.6 Proof of Proposition 2.3.1 121
2.7 Proof of Theorem 2.3.3 123
2.8 Proof of Theorem 2.3.4 124
2.8.1 The one-dimensional Skorokhod problem 125
2.8.2 Convergence of (M^n, W^n) 127
2.8.3 Convergence of T^n 134
2.8.4 Interactions with neighbouring patches 138
2.9 Effective dispersal distance 144
2.10 Appendix 152
2.10.1 Hitting times of the simple random walk on $\mathbf{Z}$ 152

2.10.2 Random sums 153

3 GENETIC DIVERSITY IN POPULATIONS SUBDIVIDED INTO SELF-INCOMPATIBLE MATING TYPE CLASSES 155

3.1 Introduction 155

3.2 The model 159

3.2.1 Mating rounds 161

3.2.2 Recombination 162

3.2.3 Mutation 162

3.3 Mathematical analysis 163

3.3.1 Allele frequencies at the neutral marker 163

3.3.2 Genealogy 165

3.3.3 Identity by descent 166

3.4 Results 172

3.4.1 Genetic diversity 173

3.4.2 Rate of effectual sex 178

3.5 Discussion 185

3.6 Appendix 190

3.6.1 Diffusion processes 190

3.6.2 The Stroock–Varadhan Theorem 191

3.6.3 The method of duality 193

3.6.4 Uniqueness of solutions to the system of SDEs 194

3.6.5 Proof of the diffusion approximation 197

3.6.6 Proof of the approximation of genealogies 201

3.6.7 Identity by descent in a population with k self-incompatible mating type classes 203

3.6.8 Identity by descent in mixed self-incompatible and unisexual population 204

CONCLUSION AND FURTHER WORK 207

BIBLIOGRAPHY 215

INTRODUCTION

This introduction will outline the biological context and motivation for a mathematical study of our models of population structure and gene flow. No attempt is made to discuss our results in any mathematical detail; we dedicate space to this at the beginning of each chapter.

The fundamental aim of theoretical population genetics is to understand the complex patterns of genetic variation that we observe in the world around us. The central thesis underpinning this theory is that of Mendelian inheritance. In the mid-nineteenth century, Mendel [95] described a mechanism through which patterns of genetic variation are passed from generation to generation in a population. An individual's observable characteristics or traits are determined by discrete units of genetic material called *genes*. Genes occur in finitely many types, termed *alleles*, and different alleles may be expressed differently. Offspring appear similar to their ancestors because they inherit genes from their parents. Patterns in genetic variation, then, are determined by the exact mechanisms through which offspring inherit genes from their parents. Of course, the story is not so simple. Exactly how individuals inherit their genes from the previous generation is incredibly complicated and determined by a huge number of competing factors. It remains to this day a hugely contested topic.

Genetic inheritance.

At the lowest level, one is interested in the exact biological mechanisms which determine how the complete collection of an offspring's genes, its *genome*, is constructed from that of its parents. In most living organisms, an individual's genome is encoded by a chemical structure called deoxyribonucleic acid, or DNA. In humans, for example, the complete genome is determined by twenty-three pairs of separate strands of DNA called chromosomes. An offspring's genome is a hybrid of that of its mother and father. Roughly speaking, a complete set of twenty-three pairs of chromosomes is constructed by inheriting one of each pair from the mother and one from the father. The inherited chromosomes are not an exact replica of those of the parents, however, instead a mosaïc of the two chromosomes in each pair. This is because of a process called *genetic recombination*. Understanding

exactly how these chromosomes are copied, a process called *DNA transcription*, is fundamental to understanding patterns of genetic variation. Although we will encounter genetic recombination again in Chapter 3, the precise biological mechanisms of inheritance are not the central focus of this thesis.

How individuals inherit their genes is also determined in a fundamental way by the dynamics of the population and its environment. By this we are referring to a very broad set of characteristics. For example, we might consider fluctuations in population size, how the population is distributed across space, how it interacts with its environment or other species, the impact of extreme extinction or recolonisation events triggered, for example, by climate change, or how frequently individuals may reproduce and, if they may reproduce sexually, how potential mates are selected. Of course, the list of factors influencing the dynamics of the population — and thus genetic inheritance and patterns of genetic variation — is potentially endless and a great deal of work has gone into assessing their relative importance.

In this thesis we will focus on the demography of the individuals that make up the population: fecundities, mortalities, movement rates, and so forth. Stochastic population models can lead to complex behaviour, especially across continuous geography. Inferring these demographic parameters of populations can be of substantial practical importance; for instance, when working to conserve an endangered species or to control an invasive species or disease vector. Although reproduction is a fundamental way by which populations grow, if an individual's genotype does not affect its reproductive success, the genealogical connections between parents and offspring can be ignored if one is only interested in population size changes. This observation has motivated a huge literature on branching, or Galton–Watson, processes and their spatial analogues. However, these relationships and the patterns of genetic variation determined by them clearly contain important information about population demography, and so there is substantial interest in using the patterns of relatedness inferred from genetic data to learn about the demographic dynamics and history of real populations. For instance, genome sequencing has provided surprising information about the history of our own species both in the recent and distant past (see, for example, [89]). However, the relationship between population demographic dynamics and the resulting patterns of genealogical relationships have only been described in the simplest possible models.

Random genetic drift and natural selection.

One of the earliest and simplest models of genetic inheritance is the celebrated Wright–Fisher model (see, for example, Crow and Kimura [25], Etheridge [37], Durrett [32] or Ewens [41] for an introduction and thorough review of theoretical population genetics,

including the Wright–Fisher and related models). In its simplest form, it is a model for the fluctuations of gene frequencies in a finite population of fixed size. Consider a single gene on each member of the population's genome. It is assumed that this gene may take one of two potential types, denoted A and a, and each individual in the population carries one copy of this gene inherited from a single parent. The population evolves in discrete non-overlapping generations; the genetic type of each individual in the present is determined by sampling uniformly with replacement from the types of all individuals alive in the previous generation and so on. Assuming that neither carrying the type A or the type a allele confers a reproductive advantage to the individual, the long-term dynamics of the Wright–Fisher model are rather boring: eventually all individuals will carry either the type a or the type A allele. We say that the allele has *fixed* within the population. The probability that the a allele fixes is given by its frequency in the present day population.

The Wright–Fisher model as stated fits into a so-called *neutral theory* of evolution originally proposed by Kimura [77]. Kimura's hypothesis is that all genetic types or alleles are *selectively neutral*, meaning that they are neither beneficial nor detrimental to an organism's ability to survive or reproduce. Fluctuations in gene frequencies within a population, then, are solely due to the fact that ancestors are randomly sampled from the previous generation and because the population is finite. This phenomenon is referred to as *random genetic drift*. Although Kimura's neutral theory acts as a suitable null model for evolution, such a theory is, of course, a gross oversimplification of the true dynamics. In fact, the Wright–Fisher model vastly over-predicts the amount of genetic variation within a population [3]; random genetic drift alone is insufficient for explaining the patterns of genetic variation we see.

One explanation is that a neutral theory of evolution, that all alleles are selectively neutral, is far too restrictive. To fully understand patterns of genetic variation we need to take into account the fact that organisms carrying different alleles may have a different *fitness*, where by the fitness of an individual we mean the number of offspring it leaves after one generation. This is observation is the central premise of Charles Darwin's [30] theory of evolution by natural selection, which may be succinctly described:

> Heritable traits that increase reproductive success will become more common in a population.

In other words, populations evolve so that the average fitness of individuals within them increases. This phenomenon is called *natural selection*.

Sewall Wright developed a useful caricature for understanding evolution through both random genetic drift and natural selection. In his 1932 paper [140], Wright proposed that evolution refers to a population's trajectory (in time) through an abstract state space. A particular state in this abstract space corresponds to the whole population's genetic con-

figuration — that is, the genotype of all individuals in the population — and movement through it represents a change in gene frequencies. The population's fitness is given by some function of its current state in this space, referred to as the *fitness function*; each value in this state space may correspond to high or low fitness for the population. Wright referred to this abstract state space, together with the fitness function, as the *evolutionary landscape*. The landscape is made up of *fitness peaks*, local maxima of the fitness function, and *fitness valleys*, local minima of the fitness function. On the one hand, natural selection acts to push the population along the lines of the steepest gradient of the fitness function, toward fitness peaks; on the other, random genetic drift allows the population to wander around the abstract state space. If the effect of random genetic drift is strong enough, the population may be pushed through fitness valleys allowing natural selection to push the population toward new peaks. Wright referred to this as a *shifting balance theory*.

Population structure.

It is clear from Wright's shifting balance theory that the relative strength of random genetic drift and natural selection is important in determining the population's trajectory through the evolutionary landscape and thus in explaining patterns of genetic variation. When natural selection is strong relative to random genetic drift there will be large temporal correlations in gene frequencies; the population will become trapped in fitness peaks. When selection is weak, the population will wander around the evolutionary landscape relatively unconstrained and there will be high variability in gene frequencies.

Wright observed a population's size is key in determining this relative effect. This is easy to see after making the following observation. In a small population, an individual member's genotype contributes much more to the total proportion of the population carrying the same type than it would in a large population. That being the case, small populations are much more susceptible to random genetic drift and they are more easily driven through fitness valleys to explore new fitness peaks. Because of this, we expect higher genetic variability and more frequent fluctuations in gene frequencies in populations which are divided into largely non-interacting subpopulations.

We refer to the division of populations in this way as *population structure*. Such division may take many forms, the most obvious being when a population is divided into groups separated by physical barriers such as rivers, mountain ranges, oceans or even man-made structures such as roads, border fences or oil pipes. Portions of the population could even be isolated simply due to separation by vast distances. For example, the species of stick insect *Clitarchus hookeri* provides evidence of geographically isolated populations undergoing repeated and rapid changes to their reproductive strategy in response to local conditions. *Clitarchus hookeri* is native to New Zealand, residing on both the North and South Islands,

but a population has been accidentally introduced to the Isles of Scilly in the United Kingdom. Although the population introduced to the United Kingdom was originally sexual, it has since lost its ability to reproduce sexually and the population now consists exclusively of asexual females [97]. Population structure created by physical barriers of this kind is referred to as *allopatric structure*. The term comes from the Greek *allos*, meaning 'other' and *patra* meaning 'fatherland'.

Population division is not only caused by physical barriers however; populations may exhibit division into subpopulations (identified by significant variation in genetic variability between groups) within the same geographic region. Such structure is referred to as *sympatric structure* and can take many forms. Many species are subdivided by a set of reproductive rules. For example, the yeast species *Saccharomyces cerevisiae* may be one of two types, denoted a and α, identified by its ability to produce or detect a signal used to search for a mate when initiating sexual reproduction. Cells of the a class signal their presence by producing a signal, or mating pheromone, detectable only by α cells and, *vice versa*, cells of the α class signal their presence by producing a mating pheromone detectable only by a cells. *Saccharomyces cerevisiae* largely reproduces asexually, and the two subpopulations of a and α cells evolve independently of one another except at the times of rare bouts of sexual reproduction [122]. During these events, cells are only able to detect the presence, and thus mate with, cells of the opposite class. In this example, the frequency of sex between these otherwise independently evolving subpopulations determines the strength of of the population structure or subdivision.

Gene flow.

If allopatric and sympatric barriers create population structure, *gene flow* erodes it. Gene flow is the movement of genetic material from one population to another. In geographically isolated populations, this might be the occasional migration of an individual between the separated groups. In the example of *Saccharomyces cerevisiae*, occasional bouts of sexual reproduction create offspring with genetic material from parental cells of each class. The presence of gene flow between populations ensures that we are in fact referring to a single population rather than different species altogether; in its absence random genetic drift would cause the populations to diverge. In fact, the amount of gene flow to needed to prevent this can be very small indeed. Wright showed that it is sufficient for there to be just one migrant per generation to prevent significant genetic variation between groups. This is often dubbed 'Wright's rule of thumb' [118]. In contrast to settings where gene flow between subpopulations is very low, if gene flow is very high then one sees a *collapse of structure*: the population structure is eroded and the otherwise separated groups act as a homogeneously mixing population [106].

The relationship between population structure, gene flow and patterns of genetic variation is the primary focus of this thesis. The literature on the role of gene flow and population structure is vast. The earliest and simplest models of spatially structured populations are Wright's finite island model [139] and its natural extension, the matrix migration model of Bodmer and Cavalli-Sforza [14]. Understanding these models in the limit of fast migration was developed by Nordborg and Krone [106] who showed that population structure can be completely eroded if gene flow between subpopulations is strong enough. Wakeley [131], Wakeley and Aliacar [133] and Wilkins and Wakeley [136] consider these models in the limit of the number of subpopulations growing without bound. Kimura and Weiss [78, 79] developed a stepping-stone model for subpopulations located at the vertices of a regular lattice; migration between subpopulations is restricted to occurring between those at neighbouring sites. Such models have received considerable attention [24, 33, 45, 80, 96, 98, 114, 119, 142]. The stepping-stone model captures what Wright terms *isolation by distance* [141], the idea that if dispersal is restricted to geographically close subpopulations then genetic dissimilarity between subpopulations increases with the distance between them. Lattice-based models of this kind serve as an approximation for populations distributed continuously across space. Developing models for populations distributed across a spatial continuum has long presented analytic challenges, as identified by Felsenstein [43], but they have recently been overcome in a series of papers by Barton et al. [4, 6].

Such studies necessitate a means of quantifying patterns of genetic variation in subdivided populations. Typically, one wants to look at correlations in genetic type between individuals sampled at random from the whole population. Focusing on a single position on the genome, this is well-captured by the probability that two individuals sampled from the population carry the same allele at this location. This probability is referred to as the probability of *identity by descent* when the two individuals carry the same allele because they have inherited it from a common ancestor (and there has been no mutation on either line of descent since), or the probability of *identity in state* when they carry the same allele but have inherited it from different ancestors. Assuming an 'infinite alleles' model of mutation, where each allele mutates to a new allele never before seen in the population with some non-zero probability in each generation, these two notions coincide. Originally developed by Malécot [92, 93], the concept of identity by descent has been a key tool for understanding and quantifying patterns of genetic variation in structured populations [102, 118]. Depending on the level of gene flow, two individuals sampled from the same subpopulation may have a significantly higher probability of identity by descent than two individuals sampled from distinct subpopulations. In fact, we expect this probability to be an increasing function of the amount of gene flow between subpopulations.

Populations with spatial structure and ancestral lineages.

In models for populations with spatial structure, in particular when individuals in the population are isolated by distance, the probability of identity by descent is given by the celebrated Wright–Malécot formula [92, 141]. Their expression is for the probability that two individuals separated by a distance d in the present are identical by descent as a function of the population's demographic parameters. For stepping-stone models, the probability of identity by descent has been calculated by Slatkin [119]; in the continuous-space models of Barton et al. [4, 6] the same probability has been recently computed by Kelleher et al. [74, 75]. In all cases, the computation of this probability depends on the exact specification of the model and the techniques do not extend generally. However, whilst this is true, the key to calculating such a probability is apparent: if we begin with a single individual from some population alive today and choose a particular bit of her genome, then, looking back through history, we can imagine recording the geographic location of the ancestor from whom that individual inherited that bit of genome. This spatial motion, as a function of amount of time in the past, we call a *lineage* — the description of where the ancestors of modern individuals lived at times in the past. Again, assuming an 'infinite alleles' model of mutation as above, where each allele mutates to a new allele never before seen in the population with some non-zero probability in each generation, the probability of identity by descent will be a decreasing function of the time it takes for two lineages to meet in the same location.

The necessity of understanding the spatial motion of these lineages is thus clear. In the simplest models, such as those with homogeneously distributed populations, these lineages will behave as a simple random walk, the behaviour of which is well understood. In greater generality, the demographic details of an individual-based stochastic population model determine not only the time evolution of population sizes, but also the dynamics of lineages in a fundamental way. For example, suppose we follow the inheritance of the chloroplast, and thus the maternal lineage, in a population of plants. Suppose we know the birth location of a given ancestor at a particular time. Roughly speaking, the location of the next ancestor might be chosen from amongst the possible ancestors, weighted by the reproductive output of those ancestors that established at that location at that time. On a broad scale, therefore, lineages when traced back through time should tend to follow regions of high population density, if the regions have correspondingly high total reproductive output. More generally, tracing the ancestral lineage into the past, the probability that a particular individual living at some point in the past is an ancestor to a given individual in the present is proportional to the amount of genetic material that individual has contributed to the present-day population. If we are looking one generation back, this is proportional to how many surviving offspring they had; that is, their fitness [16]. Lineages, then, are driven in the direction of

increasing fitness.

However, we are not just interested in the direction of the lineages but also the speed at which they are moving across the environment. Unsurprisingly, this will depend in a fundamental way on the *dispersal rate*, σ, of the population, characterising the typical distance between a parent and its offspring. The relationship between dispersal and patterns of genetic variation in spatially structured populations has recently been reviewed by Cayuela et al. [19]. Specifically, the dispersal rate is given by the standard deviation of the displacement between parent and offspring along an arbitrary axis. Imagining a lineage as a random walk, one might guess that in a homogeneous population, the distance from a given individual to their single ancestor t generations ago is proportional to $\sigma\sqrt{t}$. Although this may be close in some cases, it can be dramatically wrong in others. To capture this, one defines the *effective dispersal rate*, σ_e, the corresponding quantity obtained by averaging back along a lineage. If L_t is the distance from an individual to their ancestor living t generations ago, then $\mathrm{var}[L_t] \approx t\sigma_e^2$ in the large t limit. For instance, if there is significant competition between siblings, individuals that disperse further from their parents may be more successful. Since these long dispersers are more successful, they contribute more to the modern-day gene pool, and hence are more likely to be seen when looking back along a lineage. In this case, we would expect $\sigma_e > \sigma$. Conversely, if habitat quality is highly heterogeneous, so that longer-distance dispersers are likely to end up in an inhospitable region, then σ_e may be substantially less than σ. Lineages in populations with a small effective dispersal rate will take a long time to meet in the same location and moreover any processes that reduce the effective dispersal rate between populations will generate patterns of greater genetic differentiation. When the effective dispersal rate is high, the correspondingly high amount of gene flow erodes the effect of the population subdivision.

The motivation for developing a rigorous mathematical framework in which to understand the spatial motion of the ancestral lineages is clear. Whilst it is evidently possible to measure the dispersal rate σ by taking empirical measurements in the field, measurements of the effective dispersal rate σ_e are much more difficult to obtain. Making estimates of such a quantity would involve the collection of vast historical data sets recording all parent-child relationships of all individuals in the present and in the past, and their locations. This presents obvious practical difficulties. Even the numerical simulation of these models presents substantial problems due to the necessity to manipulate huge data sets. Although recent developments are beginning to enable progress in this direction [16, 55, 56, 76], a mathematical analysis, when it is feasible, is nevertheless preferred since it highlights the dependence of the observed quantities on the underlying parameters.

Structure of the thesis.

This thesis consists of three separate papers, labelled as chapters here; in each, we define a novel model for a population exhibiting structure. The models of Chapters 1 and 2 are for spatially distributed populations. In the former, the dynamics of the ancestral lineages are driven by the population's reproductive mechanism; in the latter, the dynamics of the ancestral lineages are driven by spatial and temporal autocorrelations in the population's environment. In the final chapter, Chapter 3, we consider a model for a population, not structured by space, but by its own reproductive strategy — much like *Saccharomyces cerevisiae*. We outline these chapters in a little more detail below.

Chapter 1.

Chapter 1 is joint work with Alison Etheridge and Peter Ralph. Our goal is to understand the spatial motion of the lineages as determined by the specific details of the population dynamics. In this chapter we consider a new model for a spatially distributed population where an individual's fitness, and thus also fluctuations in local subpopulation sizes, are determined by some function of the population sizes within some local neighbourhood. The novelty of the models that we consider is that, in the reproduction mechanism, we distinguish between the production of juvenile offspring, which will depend on the local crowding at the location of the adult, and the probability of them successfully reaching maturity, which will depend upon the adult population in the location in which they are trying to establish. As mentioned, in classical models, tracing ancestral lineages, one typically sees them drawn towards regions of high population density. This is no surprise, since such models are generally based on branching processes in which the total number of offspring produced in a given location is proportional to the population size there; no distinction is made between juveniles and adults. In our model, the motion of a lineage ancestral to an individual sampled from such a population follows a random walk in a random environment. We establish the motion of such a lineage in the scaling limit. Perhaps not surprisingly, we shall see that the rate at which ancestral lineages jump is driven by a combination of the fecundity at neighbouring sites and the establishment probability of juveniles relative to population size. We also find that lineages drift towards regions of higher 'total fecundity', which may or may not correspond to higher population density. For example, one can imagine scenarios in which high local population density is beneficial to the production of juveniles, perhaps because of ease of finding a mate; on the other hand competition for scarce resources may have the opposite effect. Similarly, a juvenile's chances of reaching maturity could benefit from increased cooperation in a large adult population, or be decreased by crowding.

Chapter 2.

Chapter 2 is also joint work with Alison Etheridge and Peter Ralph. In contrast to Chapter 1, in Chapter 2 we study the movement of lineages in a population whose local population sizes are entirely determined by exogenous factors: in our model, by the presence of transient, patchy resources in the habitat. Models of this kind capture the concept of *isolation by environment* (see, for example, [135] for a review) as opposed to isolation by distance in Wright and Malécot's earlier work. As an example, imagine that we are tracking populations of a soil microbe across a grassland, and the microbe is tightly associated with a particular type of shrub. A low density of microbes is found in the soil away from shrubs, but much higher densities live in the root zones of shrub patches. Looking into the past, these shrubs expand, contract, appear, and disappear over time and, in doing so, exogenously specify the dynamics of the population sizes of the soil microbe underneath. New patches of shrubs appear only rarely, but when they do, the lucky microbes living near to the location where the patch originated reproduce rapidly and, since the area surrounding the patch produces very few offspring, the microbes found in that patch are descendants of those initially lucky individuals for as long as the patch exists. This means that as one traces back the lineage of a microbe found in a patch today, the lineage must remain in that patch until its origination — the motion of lineages is strongly constrained by the dynamics of the patches.

In our model, the habitat at each location of space may take one of two values, *good* or *bad*, indicating the presence or absence of such a resource. In fact, we will assume that good habitats are so favourable that population sizes in good habitats will be large enough so that, when given the opportunity, lineages will always move from bad to good habitats and remain within them until their origination, as in the motivating example. By varying the spatial and temporal autocorrelation of habitat quality in several models, we can highlight how these determine the long-term behaviour of lineage movement, and hence spatial patterns of relatedness and gene flow. In spite of its apparent simplicity, our model allows us to explore a variety of scenarios, while retaining some analytical and numerical tractability. Moreover, we are able to make novel contributions regarding the speed and direction of ancestral lineage in populations exhibiting isolation by environment.

Chapter 3.

In our two models of populations with spatial or environmental structure above, the spatial motion of the lineages is fundamental for accessing the patterns of genetic variability in the populations. However, recall that population structure need not be engendered by spatial isolation, physical barriers or the environment. In the example above, the yeast *Saccharomyces cerevisiae* was divided into two classes by its reproductive strategy. These

classes evolve independently and genetic material is only exchanged at the time of rare bouts of sexual reproduction. Furthermore, each cell may only mate with another of the opposite class. In fact, amongst most fungi, ciliates and algae, mating is similarly restricted to occurring between individuals from distinct self-incompatible classes, termed mating types [51]. However, the number of mating types is not necessarily restricted to two, with as many as 23,328 in fungi [82] and 100 in ciliates [109]. The amount of gene flow between these otherwise independently evolving populations is determined by the frequency of sexual reproduction, when two individuals of distinct classes come together and exchange genetic material.

In joint work with George Constable, Ian Letter, Sarah Penington and Daniel Wilson, in Chapter 3 we address the question of when self-imposed population structure of this kind would be favourable and which evolutionary forces govern the number of mating types maintained by the population. Many hypotheses exist (see, for example, [9] for a review) but our focus is on the so-called 'sex advantage enhancer model'. This model suggests that self-incompatibility amplifies the benefits of genetic recombination; that is, the mixing of genetic material from both parents so that offspring develop a combination of traits seen in neither parent. Since genetic recombination can only be advantageous if parents are genetically distinct from one another, population structure which increases the genetic dissimilarity between potential mates should be positively selected for.

However, with rare exceptions [26, 101], the sex advantage enhancer model has received little theoretical attention. The reason for this is clear. Whilst the hypothesis is straightforward to formulate verbally, it is significantly harder to formulate mathematically; the precise mechanisms that drive selection for genetic recombination are themselves still an intensely debated subject [1]. The novelty in our methods is that we abandon any attempt to model selection for recombination and instead take a coarser-grained perspective. Our approach is to quantify the patterns of genetic variability by computing the probability of identity by descent for individuals sampled from either the same or distinct mating types. For our model, we are able to express this probability as a function of the number of mating types and the frequency of sex.

We can reconcile our findings with Wright's shifting balance theory and his metaphor of the evolutionary landscape. Small populations are more susceptible to random genetic drift and are thus more easily driven across the landscape, allowing the exploration of new fitness peaks. Thus populations divided into subpopulations with limited gene flow between them exhibit greater genetic variability. There are two reasons why a population structure engendered by reproductive strategies of this kind may be positively selected for then. First, by increasing the overall genetic variability of the population, the population is much more likely to survive extreme or rapid changes in its environment. Secondly, with a reproductive strategy which only permits mating between pairs of individuals selected from

distinct subpopulations, individuals are much more likely to mate with another individual genetically distinct from themselves. These effects and their implications for the origin and maintenance of sex are discussed in significant detail in the chapter.

Although these three chapters are concerned with very different scenarios, our three models have been deliberately chosen in order to capture different forms of population structure, and by contrasting these we highlight the complex relationship between population structure, gene flow and patterns of genetic variation.

CHAPTER 1

ANCESTRAL LINEAGES IN A POPULATION MODEL WITH NONLOCAL DISPERSAL OF JUVENILES

Joint work with
ALISON M. ETHERIDGE AND PETER L. RALPH

We describe a model for a population evolving on $\mathbf{Z}^d$ in which reproduction is as a result of a two-stage process in which fecundity of individuals depends on the local population density at the parental location, and the establishment probability of the juvenile offspring depends on the local population density at the location where they fall. After suitable scaling, tracing numbers of adults in the population we arrive at a system of interacting SDEs. Both migration and growth rates depend on local population densities. In a special case they reduce to a discrete and stochastic version of a porous medium equation with an additional reaction term.

The motion of a lineage ancestral to an individual sampled from such a population follows a random walk in a random environment. We establish the motion of such a lineage in the scaling limit. In contrast to classical models in which ancestral lineages are drawn into regions of high population density, in our setting, a lineage is attracted to regions of high *fecundity*, which may or may not correspond to the most densely populated regions of space.

Under conditions that guarantee that the deterministic version of the limiting system of equations has a stable fixed point, we investigate the fluctuations about that fixed point, which are determined by an infinite system of weakly coupled Ornstein–Uhlenbeck processes.

1.1. Introduction.

This paper is concerned with a family of discrete time and space population models. In a parallel paper [38], we shall derive a lookdown construction of a continuous time and space analogue. On the one hand, the discrete setting facilitates the analysis of the motion of ancestral lineages; on the other hand, it also allows direct comparison with the corresponding motion under classical models of evolving populations. The novelty of the

models that we consider is that, in the reproduction mechanism, we distinguish between the production of juvenile offspring, which will depend on the local crowding at the location of the adult, and on the probability of their successfully reaching maturity, which will depend upon the adult population in the location in which they are trying to establish.

In classical models, tracing ancestral lineages, one typically sees them drawn towards regions of high population density. This is no surprise, since such models are generally based on branching processes in which the total number of offspring produced in a given location is proportional to the population size there; no distinction is made between juveniles and adults. Here, also perhaps not surprisingly, we shall see that the rate at which ancestral lineages jump is driven by a combination of fecundity at neighbouring sites and the establishment probability of juveniles relative to population size, and that lineages drift towards regions of higher 'total fecundity', which may or may not correspond to higher population density. For example, one can imagine scenarios in which high local population density is beneficial to production of juveniles, perhaps because of ease of finding a mate; on the other hand competition for scarce resources may have the opposite effect. Similarly, a juvenile's chances of reaching maturity could benefit from increased cooperation in a large adult population, or be decreased by crowding.

In §1.2 we shall describe our individual based model and the weighted ℓ_p-spaces in which it takes its values, before giving a careful statement of our results. The first (for which the proof is in §1.3) concerns a scaling limit of the model. The limit will be described by a system of interacting SDEs with Feller noise. This is the natural discrete space analogue of the generalised superprocesses that arise in the continuum setting. We emphasise that the techniques that we use to analyse the continuum model do not easily generalise to the model considered here. Instead we draw heavily on Perkins [108], who constructs the Dawson–Watanabe superprocess as a limit of branching random walks (except that here we do *not* rescale space), while also borrowing ideas from Sturm [120, 121], who considered scaling limits of systems of interacting populations taking values in the same discrete spaces as those we use here.

Our second topic, for which proofs can be found in §1.4, is the motion of ancestral lineages, conditional on the history of the vector of population sizes. Although qualitatively interesting, even in the scaling limit, we do not know how to reverse the population process with respect to time and so quantitative statements are out of reach. However, as we prove in §1.5, under certain assumptions, in a high density regime, the fluctuations in population density about a stable limit can be characterised as a system of Ornstein–Uhlenbeck processes (coupled through a 'migration' term), whose (reversible) stationary distribution is easily established. This leads the way in §1.5.5 to a brief discussion of the first order effect of the fluctuations in the population on the motion of ancestral lineages.

1.2. A discrete time and space model for an evolving population.

In this section we describe our individual based models and state our main results.

Let us begin by introducing the weighted ℓ_p-spaces in which our population models will take their values. These are a generalisation of the weighted ℓ_2-spaces used by Shiga and Shimizu [115] as state spaces for infinite-dimensional systems of stochastic differential equations. Such generalisations were used, for example, in [96, 120, 121].

DEFINITION 1.2.1 (The space $\ell_{\Gamma,p}$). Let $\Gamma : \mathbf{Z}^d \to \mathbf{R}_+$ satisfy

$$\sum_{x\in\mathbf{Z}^d} \Gamma(x) < \infty,$$
$$\frac{\Gamma(x)}{\Gamma(y)} < f_\Gamma(\|x-y\|),$$

for some continuous function $f_\Gamma : \mathbf{R}_+ \to \mathbf{R}_+$. We define

$$\ell_{\Gamma,p} = \left\{X \in \mathbf{R}^{\mathbf{Z}^d} : \|X\|_{\Gamma,p} < \infty\right\},$$

where

$$\|X\|_{\Gamma,p}^p = \sum_{x\in\mathbf{Z}^d} \Gamma(x)|X(x)|^p.$$

We can, for example, take $\Gamma(x) = \exp(-\|x\|)$, and then $f(x-y) = \exp(\|x-y\|)$.

For a topological space E, we will write $D([0,T],E)$ for the space of right-continuous paths $\omega : [0,T] \to E$ equipped with the Skorokhod topology.

1.2.1. The individual based model. We are going to define a sequence of individual based models, indexed by $n \in \mathbf{N}$. We suppose that we have two sequences τ_n, η_n, of non-negative real numbers. Individuals live in *demes* indexed by $\mathbf{Z}^d$. The population evolves in discrete generations with inter-generation time τ_n. Each individual in the population is assigned 'mass' η_n. We write $X_t^n(x)$ for the size of the population in deme $x \in \mathbf{Z}^d$ at time t, that is

$$X_t^n(x) := \eta_n \times \text{number of particles alive at } x \text{ at time } t.$$

Reproduction is described as follows. For each $t \in \{\tau_n, 2\tau_n, \ldots, \lfloor T/\tau_n\rfloor\tau_n\}$,

1. at each point $x \in \mathbf{Z}^d$, each individual, independently, produces a Poisson number of juvenile offspring with parameter $\gamma_n(X_{t-}^n(x))$;

2. at each point $x \in \mathbf{Z}^d$, after production of juveniles, each adult at x, independently, dies with probability $\mu_n(X_{t-}^n(x))$;

3. with probability $m\tau_n$ each juvenile, independently, migrates to a neighbouring deme (chosen with equal probability from among its nearest neighbours), otherwise it stays in its parental deme;

4. after migration, for each $x \in \mathbf{Z}^d$, each juvenile at x, independently, establishes (and so becomes an adult in the population) with probability $r_n(X_{t-}(x))$. We treat this maturation as instantaneous. Juveniles which do not establish do not survive to adulthood are removed from the system. Once more, we treat this removal as instantaneous.

REMARK 1.2.2. Evidently we could take a more general mechanism for migration of juveniles. Our results will involve taking scaling limits, and both the proof that our sequence of scaled population models is tight in $\ell_{\Gamma,p}$, and the proof of convergence of the corresponding sequence of scaled ancestral lineages exploit the fact that the juveniles can only jump to finitely many different locations.

In order to obtain a sensible scaling limit, we make some assumptions on the parameters τ_n, η_n, r_n,γ_n, μ_n.

ASSUMPTIONS 1.2.3. We suppose that $\tau_n, \eta_n \in (0,1)$ with

$$\tau_n \to 0, \quad \eta_n \to 0, \quad \eta_n/\tau_n := \delta_n \to \delta \in [0,\infty), \qquad \text{as } n \to \infty.$$

Defining the function $g_n : \mathbf{R}_+ \to \mathbf{R}$ by $g_n : u \mapsto g_n(u)$, where

$$g_n(u) = \frac{r_n(u)\gamma_n(u) - \mu_n(u)}{\tau_n}, \tag{1.2.1}$$

we assume that as $n \to \infty$, $r_n, \gamma_n, \mu_n, g_n$ converge to Lipschitz continuous functions r, γ, μ, g, where the convergence is uniform on compact subsets and bounded on $\mathbf{R}_+$. In particular, there exist finite constants $\overline{\gamma}$, $\overline{g}$ such that

$$\sup_{n\in\mathbf{N},u\in\mathbf{R}_+} |\gamma_n(u)| \le \overline{\gamma},$$
$$\sup_{n\in\mathbf{N},u\in\mathbf{R}_+} |g_n(u)| \le \overline{g}.$$

We will also assume that the death probabilities $\mu_n(u)$ satisfy

$$\inf_{n\in\mathbf{N},u\in\mathbf{R}_+} \mu_n(u) > 0$$

and, writing $\Phi_n(u) = u\gamma_n(u)$, there is a finite constant $\overline{\Phi}$ such that

$$\sup_{n\in\mathbf{N},u\in\mathbf{R}_+} |\Phi_n(u)| \le \overline{\Phi}.$$

The (biologically natural) assumption that $\mu_n(u)$ is uniformly bounded away from zero is only actually used in the proof of Lemma 1.3.12. We shall sometimes refer to $\Phi_n(u)$ as the total fecundity when the population density is u.

Following [108, 134] individuals will be labelled by multi-indices

$$\alpha \in I := \bigcup_{k\in\mathbf{N}} \mathbf{N}^{\{0,\dots,k\}} = \{(\alpha_0,\dots,\alpha_k) : \alpha_i \in \mathbf{N}, k \in \mathbf{N}\}.$$

Denote by $k = |(\alpha_0,\dots,\alpha_k)|$ the generation of α. Let N^α denote the number of (mature) offspring of an individual with label α. Individuals are labelled as follows: each individual alive at time $t = 0$ (of which there are at most countably many) is given a label $\alpha_0 \in \mathbf{N}$. At time $t = \tau_n$, the individual α_0 has a random number of surviving offspring. Denote by $\widetilde{N}^{\alpha_0}$ this number and give each of these offspring a distinct label (α_0,α_1), $\alpha_1 \in \{1,\dots,\widetilde{N}^{\alpha_0}\}$. If α_0 itself survives until the next generation, it is treated as a new individual and labelled $(\alpha_0, 0)$. We then proceed inductively.

Omitting a subscript n, at the nth stage of rescaling we shall write

$$\underline{t} := \lfloor t/\tau_n \rfloor \tau_n$$

to denote the last generation time before t. Observe that a multi-index $\alpha \in I$ can label a particle at time t if and only if $|\alpha| = \underline{t}/\tau_n$. Any such particle is alive during the time interval $[|\alpha|\tau_n, (|\alpha|+1)\tau_n)$. We write $\alpha \sim t$ if and only if α labels a particle alive at time t. Note that $\alpha \sim t$ if and only if $\alpha \sim \underline{t}$.

Let us denote by $Y^\alpha \in \mathbf{Z}^d$ the position of the particle with label α (which is constant on the interval $[|\alpha|\tau_n, (|\alpha|+1)\tau_n)$ on which α is alive). With this notation, we may write

$$X_t^n(x) = \eta_n \sum_{\alpha\sim\underline{t}} \mathbf{1}_x(Y^\alpha).$$

We now use the Poisson structure of the reproduction mechanism to break down the population in generation $|\alpha|+1$ into components that are conditionally independent given the population in generation $|\alpha|$. We write $\{e_i\}_{i=1}^d$ for the unit vectors along each of the axes in $\mathbf{Z}^d$. For each α define

$$\begin{aligned}
\mathscr{L}(N^\alpha_{+e_i}) &= \text{Poisson}\big(m\tau_n r_n(X^n_{|\alpha|\tau_n}(Y^\alpha + e_i))\gamma_n(X^n_{|\alpha|\tau_n}(Y^\alpha))/2d\big),\\
\mathscr{L}(N^\alpha_{-e_i}) &= \text{Poisson}\big(m\tau_n r_n(X^n_{|\alpha|\tau_n}(Y^\alpha - e_i))\gamma_n(X^n_{|\alpha|\tau_n}(Y^\alpha))/2d\big),
\end{aligned}$$

$$\mathscr{L}(N_{=}^{\alpha}) = \text{Poisson}\big((1-m\tau_n)r_n(X^n_{|\alpha|\tau_n}(Y^\alpha))\gamma_n(X^n_{|\alpha|\tau_n}(Y^\alpha))\big),$$
$$\mathscr{L}(N_0^{\alpha}) = \text{Bernoulli}\big(1-\mu_n(X^n_{|\alpha|\tau_n}(Y^\alpha))\big),$$
$$N^\alpha := N_0^\alpha + \sum_{i=1}^d N^\alpha_{-e_i} + N^\alpha_{=} + \sum_{i=1}^d N^\alpha_{+e_i}.$$

In this notation, $N^\alpha_{+e_i}$ (respectively $N^\alpha_{-e_i}$, $N^\alpha_{=}$) is the number of surviving offspring of α which establish at the site $Y^\alpha + e_i$ (respectively $Y^\alpha - e_i$, Y^α), $\widetilde{N}^\alpha = N^\alpha_{=} + \sum_{i=1}^d \big(N^\alpha_{-e_i} + N^\alpha_{+e_i}\big)$ is the total number of surviving offspring of the individual with label α, N_0^α is the indicator of the event that individual α survives until the next generation, and N^α is the total number of descendants (including itself if it survives) from individual α in generation $|\alpha|+1$.

ASSUMPTION 1.2.4. To avoid a plethora of subscripts, in all our proofs, we are going to specialise to $d = 1$, and we shall write N^α_+, N^α_- in place of $N^\alpha_{+e_1}$, $N^\alpha_{-e_1}$. The proofs carry over with only notational differences to higher dimensions.

We define a filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$ by

$$\mathcal{F}^n_t := \sigma\left(\left\{Y^\alpha : |\alpha|\tau_n \le \underline{t}\right\} \cup \left\{N_0^\alpha, N_-^\alpha, N_=^\alpha, N_+^\alpha : |\alpha|\tau_n < \underline{t}\right\}\right),$$

ie the natural filtration generated by the position and number of offspring of all individuals that have existed up to time t. By the thinning of Poisson random variables, N^α_+, $N^\alpha_=$, N^α_- and N^α_0 can be assumed to be conditionally independent given $\mathcal{F}_{|\alpha|\tau_n}$. Also, note that, because $\{Y^\alpha : |\alpha|\tau_n \le \underline{t}\}$ is $\mathcal{F}^n_{\underline{t}}$-measurable, so is X^n_s for all $s < \underline{t} + \tau_n$.

Existence of the process follows by comparison with a branching random walk in which individuals never die, and they give birth in each generation to a Poisson number of offspring with parameter $\bar{\gamma}$, all of which successfully establish.

1.2.2. Main results. Having defined the model, we record our main results.

Scaling limits. Recall that a sequence of processes is C-tight if it is tight and every limit point is continuous.

We use '$\Rightarrow$' to denote weak convergence; ie for a topological space $(V,\mathcal{T})$ we write $X^n \Rightarrow X$ in $(V,\mathcal{T})$ to mean that $\mathbf{E}[f(X^n)] \to \mathbf{E}[f(X)]$ for each $f : V \to \mathbf{R}$ bounded and continuous (with respect to $\mathcal{T}$). Finally, we will denote by Δ_h the discrete Laplacian acting on lattice functions $f : h\mathbf{Z} \to \mathbf{R}$ defined by

$$\Delta_h f(x) := \frac{f(x+h) + f(x-h) - 2f(x)}{h^2},$$

so that, in particular,

$$\Delta_1 f(x) = f(x+1) + f(x) - 2f(x)$$

for all $f : \mathbf{Z} \to \mathbf{R}$.

THEOREM 1.2.5. *Let $\mathbf{R}_+^{\mathbf{Z}^d}$ be endowed with the product topology. Consider a sequence X^n of populations evolving according to the model of §1.2.1 with parameters satisfying Assumption 1.2.3. Moreover, suppose that $X_0^n \Rightarrow u$ in the space $\ell_{\Gamma,p}$ for some $p \geq 2$. Then as $n \to \infty$, the sequence of processes X^n is C-tight in $D([0,T], \mathbf{R}_+^{\mathbf{Z}^d})$ and any weak limit point X is a weak solution to the infinite system of stochastic differential equations*

$$\begin{aligned} X_0(x) &= u(x), \\ X_t(x) &= X_0(x) + \int_0^t X_s(x) g(X_s(x))\,\mathrm{d}s + \int_0^t \frac{m}{2} r(X_s(x)) \Delta_1(\Phi \circ X_s)(x)\,\mathrm{d}s \\ &\qquad + \int_0^t \sqrt{\delta \mu(X_s(x))(2 - \mu(X_s(x)))X_s(x)}\,\mathrm{d}W_s(x), \end{aligned} \tag{1.2.2}$$

for all $x \in \mathbf{Z}^d$, where $\{W(x)\}_{x \in \mathbf{Z}^d}$ denotes a countable collection of independent standard Brownian motions. Moreover, any solution to (1.2.2) *is $\ell_{\Gamma,p}$-valued.*

The form of the quadratic variation seems more natural if one decomposes it into two parts: that coming from the variation in both the number of surviving adults from one generation to the next, and that from the Poisson number of offspring born in each generation. This can be seen as follows. Recall that since the function g_n defined in equation (1.2.1) is convergent, it must necessarily be the case that $r_n\gamma_n - \mu_n \to 0$ as $n \to \infty$ (as $\tau_n \to 0$). Equivalently, the limiting functions r, γ and μ satisfy $r\gamma = \mu$. We then recognise the quadratic variation as the limit of $(\mu_n(1-\mu_n) + r_n\gamma_n)\eta_n^2/\tau_n$ times the number of individuals in site x in the parental generation (ie $X^n(x)/\eta_n$), where the first term comes from the Bernoulli variables that determine whether an adult survives to the subsequent generation, and the second is the variance of the Poisson number of surviving offspring of an individual that establish in the same deme as the parent (terms in the quadratic variation corresponding to migrating offspring vanish in the limit). The exact details of this decomposition will be spelled out in the calculation (1.3.2).

The nonlinear migration term appearing in this limit is a result of our two-stage reproduction mechanism and novel to our class of models. In certain parameter regimes it takes a more familiar form. For example, taking $r(u) \equiv 1$ and $\Phi(u) \equiv u^\beta$ for some $\beta > 1$, the limiting system reduces to a discrete and stochastic version of porous medium equation plus an additional reaction term; and taking $g(u) \equiv u(1-u)$, $r(u) \equiv 1$ and $\Phi(u) \equiv u$ it reduces to a discrete and stochastic version of the Fisher–Kolmogorov-–Petrovsky-–Piskunov (or Fisher–KPP) equation. We can also recover the stepping-stone model with varying pop-

ulation size; that is, an independent critical branching process at each deme with nearest neighbour offspring dispersal. This limit is achieved by setting $g(u) \equiv 0$, $r(u) \equiv 1$, $\mu(u) \equiv 1$ and $\Phi(u) \equiv u$.

If we make stronger assumptions on the initial condition, then we can strengthen this result to tightness in $\ell_{\Gamma,p}$.

THEOREM 1.2.6. *Under the conditions of Theorem 1.2.5, suppose further that*

$$\sup_{n\in\mathbf{N}} \mathbf{E}\left[\|X_0^n\|_\infty^p\right] = \sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{x\in\mathbf{Z}^d} |X_0^n(x)|^p\right] < \infty, \tag{1.2.3}$$

then the sequence of processes X^n is C-tight in $\ell_{\Gamma,p}$ and any weak limit point X is a weak solution to (1.2.2).

REMARK 1.2.7. Whilst the condition (1.2.3) may seem arbitrary, it is in fact (biologically speaking) rather mild. In words, it stipulates that the population, distributed across the one-dimensional lattice, does not become arbitrarily large as $x \to \infty$ or $x \to -\infty$. We would not expect this to be the case in any biological relevant applications.

In order to deduce convergence of the scaled processes, we need uniqueness of the limit.

PROPOSITION 1.2.8 (Uniqueness). *In the notation above, let*

$$\begin{aligned} b(X,x) &:= g(X(x))X(x) + \frac{m}{2} r(X(x))\Delta_1(\Phi\circ X)(x), \\ \sigma(X(x),x) &:= \delta\mu(X(x))\big(2-\mu(X(x))\big)X(x). \end{aligned}$$

Suppose that

$$\begin{aligned} \|b(X,\cdot)\|_{\Gamma,p} &\le C\big(1+\|X\|_{\Gamma,p}\big), \\ \|b(X,\cdot)-b(Y,\cdot)\|_{\Gamma,1} &\le C\|X-Y\|_{\Gamma,1}; \qquad (1.2.4) \\ |\sigma(X(x),x)| &\le C(1+|X(x)|), \\ |\sigma(y,x)-\sigma(z,x)| &\le C|y-z|, \qquad (1.2.5) \end{aligned}$$

then the solutions to (1.2.2) *are pathwise unique and hence unique in law.*

The proof of this result, which will be an easy consequence of Theorem 5.2.1 of Sturm [121], is in §1.3.4. The linear growth conditions are automatically satisfied under Assumptions 1.2.3.

COROLLARY 1.2.9. *Under the additional conditions* (1.2.4) *and* (1.2.5)*, we have convergence of the processes X^n in Theorems 1.2.5 and 1.2.6.*

Ancestral lineages. We would like to understand the effect of our two-step reproduction mechanism on the motion of ancestral lineages. In other words, conditional on knowing the size of the population in each deme at all times in $[0, T]$, if we sample an individual from deme $x \in \mathbf{Z}$ (which without loss of generality we may take to be the origin) at time T, and trace back its ancestry, what is the motion of the ancestral lineage? In particular, we wish to establish this motion in the scaling limit. Since the forwards in time migration probabilities scale with τ_n, the lineage will almost surely only make a finite number of jumps over the time interval $[0, T]$, and in the scaling limit we expect to recover a continuous time Markov chain. This is indeed the case.

We need some notation. Consider a realisation of $(X_t^n)_{0\le t\le T}$. Since X^n evolves in discrete generations, $X_T^n = X^n_{\lfloor T/\tau_n\rfloor\tau_n}$. We write

$$\vec{t} := \lceil T/\tau_n\rceil\tau_n - t.$$

Then in our previous notation, as we trace backwards in time, the generation times for our population in $[0, T]$ are of the form $\underline{\vec{t}}$. We shall use $\{L_t^n : 0 \le t \le T\}$ to denote the process that records the location at time t before the present of the lineage ancestral to an individual sampled from the origin (so that $L_0^n = 0$). The process L^n can only jump at the discrete times $\underline{t}$, $0 \le t \le T$. If we write $\mathbf{P}[\cdot|X^n]$ for the law of the process L^n conditional on $(X_t^n)_{0\le t\le T}$, then our goal will be establish the one-step transition probabilities

$$p^n(X^n; \underline{t}, x, y) := \mathbf{P}[L^n_{\underline{t}+\tau_n} = y | L^n_{\underline{t}} = x, X^n], \qquad x, y \in \mathbf{Z};$$

that is the probability that an individual alive at x at time $\underline{\vec{t}} = \lceil T/\tau_n\rceil\tau_n - \underline{t}$ is descended from an individual at y at time $\overrightarrow{\underline{t} + \tau_n}$, conditional on X^n. In particular, we shall be interested in the behaviour of the lineage process L^n as $n \to \infty$.

THEOREM 1.2.10. *Fix $c > 0$ and define $T^{n,c} := \inf\{\underline{t} \ge 0 : X^n_{\underline{\vec{t}}}(L^n_{\underline{t}}) < c\}$, the first time at which the population density at the site of the lineage falls below c. Suppose further that $\tau_n^2/\eta_n \to 0$ as $n \to \infty$. Under the assumptions of Corollary 1.2.9, the process $(X^n, L^n_{\cdot\wedge T^{n,c}}) \Rightarrow (X, L)$ in $D([0, T], \mathbf{R}_+^{\mathbf{Z}} \times \mathbf{Z})$ as $n \to \infty$. The law of L is given by its law conditional on X. If $X_T(0) = 0$ then $L_t = \varnothing$ for all $0 \le t \le T$. If $X_T(0) > 0$ then L is the unique process such that $L_0 = 0$ and*

$$f(L_t) - f(L_0) - \int_0^{t\wedge T^c} A_s^X f(L_s)\,\mathrm{d}s$$

is a $\mathbf{P}[\cdot|X]$-martingale for each $f \in B(\mathbf{Z})$, where

$$A_s^X f(x) := \frac{1}{2}\frac{r(X_{T-t}(x))}{X_{T-t}(x)}\Big(\Phi(X_{T-t}(x+1))\big(f(x+1)-f(x)\big) + \Phi(X_{T-t}(x-1))\big(f(x-1)-f(x)\big)\Big).$$

Recalling that r is the establishment probability, and Φ the total fecundity, the form of the generator is easy to interpret. The instantaneous jump rate of a lineage at x at time t is determined by the probability that it was an immigrant from a neighbouring deme in the previous infinitesimal time interval. Since individuals within any given deme are exchangeable, this is just the proportion of the population that are immigrants, so, using that $X_{\bar{t}}$ is continuous, the jump rate is

$$\frac{m}{2}\frac{r(X_{\bar{t}}(x))\Phi(X_{\bar{t}}(x+1)) + r(X_{\bar{t}}(x))\Phi(X_{\bar{t}}(x-1))}{X_{\bar{t}}(x)}$$

If a jump occurs at time t, when the particle is at site x, then it jumps left or right in proportion to the rate at which the neighbouring demes are producing migrants. That is, conditional on a jump taking place, it is to $x+1$ with probability

$$\frac{\Phi(X_{\bar{t}}(\omega, x+1))}{\sum_{y=x\pm1}\Phi(X_{\bar{t}}(\omega, y))}.$$

In other words, the direction of the jump is determined by the total fecundity in each of the neighbouring demes, rather than the total population size.

Let us briefly compare this to the spatial motion of ancestral lineages in the stepping-stone model, when $g(u) \equiv 0$, $r(u) \equiv 1$, $\mu(u) \equiv 1$ and $\Phi(u) \equiv u$. In this case, as we hypothesised above, since the total number of offspring produced in a given location is proportional to the population size there, and no distinction is made between juveniles and adults, the direction of the jump is determined by the total population size in each of the neighbouring demes.

REMARK 1.2.11. It seems reasonable to hope that under an additional scaling of space, one could recover nonlinear superprocesses that are the natural analogue of these models in a spatial continuum. Proceeding informally, if one defines a sequence of processes Ξ^h by setting

$$\Xi_t^h(x) := X_t(\lfloor x/h \rfloor),$$

for all $t \in \mathbf{R}_+$ and $x \in \mathbf{R}$, then by way of substitution into equation (1.2.2),

$$\begin{aligned}\Xi^h_t(x) = \Xi^h_0(x) &+ \int_0^t \Xi_s(x) g(\Xi^h_s(x))\,\mathrm{d}s \\ &+ \int_0^t \frac{m}{2} r(\Xi^h_s(x))\Delta_1(\Phi \circ X_s)(\lfloor x/h \rfloor)\,\mathrm{d}s \\ &+ \int_0^t \sqrt{\delta\mu(\Xi^h_s(x))(2-\mu(\Xi^h_s(x)))\Xi^h_s(x)}\,\mathrm{d}W_s(\lfloor x/h \rfloor).\end{aligned}$$

One may readily check that

$$\Delta_1(\Phi \circ X_s)(\lfloor x/h \rfloor) = h^2 \Delta_h(\Phi \circ \Xi^h_s)(x).$$

Let us additionally assume that $mh^2 \to v$ as $h \to 0$. Then for any smooth and compactly supported $\varphi : \mathbf{R} \to \mathbf{R}$, passing to the limit $h \to 0$ in the sum $\sum_{x \in h\mathbf{Z}} \phi(x)\Xi^h_t(x)$, one may readily believe that if there were to be enough regularity in the solutions, then Ξ^h possesses a non-trivial limit Ξ, given by a (weak) solution to the stochastic partial differential equation

$$\begin{aligned}\int_{-\infty}^{\infty} \varphi(x)\Xi_t(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \varphi(x)\Xi_0(x)\,\mathrm{d}x &+ \int_{-\infty}^{\infty}\int_0^t \varphi(x)\Xi_s(x) g(\Xi^h_s(x))\,\mathrm{d}s\,\mathrm{d}x \\ &+ \int_{-\infty}^{\infty}\int_0^t \varphi(x)\frac{v}{2} r(\Xi_s(x))\Delta(\Phi \circ \Xi_s)(x)\,\mathrm{d}s\,\mathrm{d}x \\ &+ \int_{-\infty}^{\infty}\int_0^t \varphi(x)\sqrt{\delta\mu(\Xi_s(x))(2-\mu(\Xi_s(x)))\Xi_s(x)}\,W(\mathrm{d}s,\mathrm{d}x),\end{aligned}$$

where W is a space-time white noise.

This actually seems to be analytically challenging. In one spatial dimension, there is a body of work that characterises scaling limits of branching particle systems as solutions to stochastic partial differential equations (in the topology of uniform convergence on compact sets), see eg Müller and Tribe [98], Kliem [80] and Durrett and Fan [33]. However, key steps in the proofs rest on estimates of the semigroup generated by the discrete Laplacian; those steps do not transfer to our nonlinear setting.

Moreover, proving convergence in a space of measure- or distribution-valued processes, is insufficient for analysis of the backwards-in-time ancestral lineages. The form of the generator in Theorem 1.2.10 suggests that in the limit, motion of ancestral lineages should involve a drift given by $r\gamma$ times $\nabla \log(\Phi)$ (which is consistent with the results of our forthcoming paper on continuum models[1]). Even if we take $\delta = 0$, so that the limit is a solution to a deterministic p.d.e., it is not obvious that solutions of the prelimiting models

[1]This work is joint work with an additional author, Thomas G. Kurtz, and its content is not discussed in this thesis.

will be sufficiently regular that the discrete gradients of total fecundity converge to the expected gradient term in the the limit. By remaining in discrete space, we are able to use the pointwise convergence of the population size at each site.

Fluctuations about a stable equilibrium. One of the difficulties that we have in assessing the effect of the fluctuations in fecundity on the population dynamics, is that we don't know how to reverse the system (1.2.2) with respect to time (or indeed conditions under which it has a stationary distribution). On the other hand, one can write down conditions under which the deterministic equation obtained by setting $\delta = 0$ has a stable steady state, and if δ is sufficiently small, one can expect the fluctuations about that steady state to be small.

Our next result establishes that those fluctuations are given (to a first approximation) by an infinite system of Ornstein–Uhlenbeck processes, coupled through an additional drift term. We then provide an expression for its (reversible) stationary distribution, when it exists.

In order to guarantee that the deterministic system has a stable stationary state, and that fluctuations around that steady state can be controlled, we make some additional assumptions on the parameters of our model.

ASSUMPTIONS 1.2.12. For the results in this subsection (and proofs in §1.5), we shall make the following additional assumptions.

1. The function $g(u)$ has a bounded second derivative, and the function $G : \mathbf{R}_+ \to \mathbf{R}_+$, defined by $u \mapsto ug(u)$, has a bounded first and second derivative.
2. The function $u \mapsto \Phi(u)$ has bounded first and second derivatives.
3. There exists $u_* \in (0, \infty)$ with $g(u_*) = 0$.

We further assume that $\delta_n = \eta_n/\tau_n \to 0$ and that

$$\frac{g_n - g}{\delta_n^{1/2}} \to 0, \tag{1.2.6}$$

and

$$\frac{\gamma_n - \gamma}{\delta_n^{1/2}} \to 0$$

as $n \to \infty$, where the convergence is uniform on compact sets and bounded on $\mathbf{R}_+$.

By Theorem 1.2.5, if $\delta_n = \eta_n/\tau_n \to 0$ and $X_0^n \Rightarrow u$ in $\ell_{\Gamma,p}$ for some $p \geq 2$ then

$X^n \Rightarrow X$ in $D([0,T], \mathbf{R}_+^{\mathbf{Z}})$, where X is the solution to the differential equation

$$\begin{aligned} X_0 &= u, \\ X_t(x) &= \int_0^t X_s(x) g(X_s(x))\,\mathrm{d}s + \int_0^t \frac{m}{2} r(X_s(x)) \Delta_1(\Phi \circ X_s)(x)\,\mathrm{d}s. \end{aligned} \tag{1.2.7}$$

Moreover, under Assumptions 1.2.12, $u(x) \equiv u_*$ is a stable fixed point of the system (1.2.7). We abuse notation and write u_* for the vector in $\mathbf{R}_+^{\mathbf{Z}}$ equal to u_* at every site.

If $X_0^n \Rightarrow u_*$ sufficiently quickly, then we can expect that $X_t^n \Rightarrow u_*$ for all $0 \le t \le T$. We should like to describe the fluctuations of X^n about the constant vector u_*, which we expect to be of order $\delta_n^{1/2}$. We define

$$Z_t^n(x) := \frac{X_t^n(x) - u_*}{\delta_n^{1/2}}, \qquad x \in \mathbf{Z}, 0 \le t \le T. \tag{1.2.8}$$

Our next result tells us that under the conditions of Assumptions 1.2.12, Z^n is weakly convergent to an infinite-dimensional system of Ornstein–Uhlenbeck processes, coupled through an additional drift term.

NOTATION 1.2.13. For any function of the population density $f : \mathbf{R}_+ \to \mathbf{R}$ we write $f_* := f(u_*)$ for its value at u_*.

THEOREM 1.2.14. *Let $\mathbf{R}_+^{\mathbf{Z}}$ be endowed with the product topology. Consider a sequence X^n of populations evolving according to the model of §1.2.1 with parameters satisfying Assumptions 1.2.3 and Assumptions 1.2.12. Let Z^n be defined as in equation* (1.2.8)*. Moreover, suppose that $\delta_n \to 0$ and that $Z_0^n \Rightarrow \zeta$ in $\ell_{\Gamma,p}$ for some $p \ge 2$. Then as $n \to \infty$, the sequence of processes $Z^n \Rightarrow Z$ in $D([0,T], \mathbf{R}_+^{\mathbf{Z}})$ and Z is the weak solution to the infinite-dimensional system of stochastic differential equations*

$$\begin{aligned} Z_0 &= \zeta, \\ Z_t(x) &= Z_0(x) + \int_0^t \left(u_* g_*' Z_s(x)\,\mathrm{d}s + \frac{m}{2} r_* \Phi_*' \Delta_1 Z_s(x) \right) \mathrm{d}s \\ &\qquad + \sqrt{\mu_*(2-\mu_*)u_*}\, B_t(x), \end{aligned} \tag{1.2.9}$$

where $\{B(x)\}_{x \in \mathbf{Z}}$ denotes a collection of independent Brownian motions.

As is the case in Theorem 1.2.6, we are able to strengthen the convergence if we have stronger assumptions on the initial condition.

THEOREM 1.2.15. *Under the conditions of Theorem 1.2.14, suppose further that*

$$\sup_{n\in\mathbf{N}} \mathbf{E}\left[\|Z_0^n\|_\infty^p\right] = \sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{x\in\mathbf{Z}^d} |Z_0^n(x)|^p\right] < \infty,$$

and again let Z denote the solution to equation (1.2.9)*. Then $Z^n \Rightarrow Z$ in $D([0,T], \ell_{\Gamma,p})$.*

Equation (1.2.9) may be recast as a stochastic equation in infinite dimensions; written

$$\mathrm{d}Z_t = AZ_t\,\mathrm{d}t + \sqrt{\mu_*(2-\mu_*)u_*}\,\mathrm{d}B_t,$$

where B is a space-time white noise and A is the operator

$$Af := \frac{mr(u_*)\Phi'(u_*)}{2}\Delta_1 f + u_* g'(u_*) f. \tag{1.2.10}$$

As long as A is sufficiently regularising, we expect solutions Z to (1.2.9) to admit a stationary distribution. We will require some semigroup theory. Our first result tells us that A is bounded and thus generates a semigroup.

PROPOSITION 1.2.16. *The operator A defined by* (1.2.10) *is bounded on $\ell_2(\mathbf{Z})$. Moreover, A generates a strongly continuous semigroup $(T_t)_{t\geq 0}$ defined by*

$$T_t = e^{tA} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}.$$

Together with Assumptions 1.2.12, a further assumption on the operator A will guarantee that the first order correction Z about the stable population density u_* has a reversible stationary distribution, from which we will be able to make some preliminary observations about the effect of the parameters of our model on the evolution of ancestral lineages.

ASSUMPTION 1.2.17. The spectrum of A defined in (1.2.10) is contained entirely in $\mathbf{C}_-$ (the open negative complex half plane).

Assumption 1.2.17 is equivalent to the exponential stability of the semigroup $(T_t)_{t\geq 0}$ generated by A (see, for example, Theorem I.3.14 of Engel and Nagel [35]). This, in turn, will be sufficient for the existence of a unique stationary distribution. In the following proposition, we give equivalent conditions in terms of the parameters of our model for Assumption 1.2.17 to be satisfied.

PROPOSITION 1.2.18. *The condition of Assumption 1.2.17 holds if and only if*

$$g_*' < 0 \wedge \frac{2mr_*\Phi_*'}{u_*}. \tag{1.2.11}$$

The discrete Laplacian in (1.2.9) is a (possibly backward) diffusion term and $u_* g'_* Z_s(x)$ the reaction term, or drift. When $\Phi'_* > 0$ so that the coefficient in front of the discrete Laplacian is positive and we really have a smoothing term, the proposition tells us that $g'_* < 0$ is sufficient. By comparing our case to one-dimensional Ornstein–Uhlenbeck processes, this is unsurprising, since $g'_* < 0$ guarantees that the process is mean-reverting. When $\Phi'_* \leq 0$, the transport term in (1.2.9) is backwards heat flow. In general, we would not expect solutions to be well-behaved. However, as long as the drift towards zero is strong enough, as made precise by (1.2.11), then the proposition guarantees sufficient regularity for the first order correction Z to admit a stationary distribution.

PROPOSITION 1.2.19. *Under conditions of Assumptions 1.2.3, 1.2.12 and 1.2.17, the process determined by the system of equations* (1.2.9) *admits a unique reversible stationary distribution $\pi \in \mathbf{M}_1(\ell_{\Gamma,p})$. Moreover, this stationary distribution is Gaussian and is characterised by its mean and autocovariance function*

$$\mathbf{E}_\pi[Z_t(x)] = 0, \tag{1.2.12}$$

$$\begin{aligned}\mathbf{E}_\pi[Z_0(0)Z_t(x)] &:= \rho_t(x)\\ &= \mu_*(2-\mu_*)u_*\left\langle \int_0^\infty T_{t+s}\delta_x \,\mathrm{d}W_s, \int_0^\infty T_s\delta_0 \,\mathrm{d}W_s \right\rangle,\end{aligned} \tag{1.2.13}$$

where W is a vector of independent Brownian motions indexed by $\mathbf{Z}$, $(T_s)_{s\geq 0}$ is the strongly continuous semigroup with infinitesimal generator A, and angle brackets indicate the covariance.

In the case $\Phi'(u_*) > 0$, this can be written

$$\begin{aligned}\mathbf{E}_\pi[Z_t(x)] &= 0,\\ \mathbf{E}_\pi[Z_0(0)Z_t(x)] &= \mu_*(2-\mu_*)u_* \int_0^\infty e^{u_* g'_*(t+2s)} p_{m r_* \Phi'_*(t+2s)}(x)\,\mathrm{d}s,\end{aligned}$$

where, if $(S_t)_{t\in[0,T]}$ is a continuous time nearest neighbour random walk on $\mathbf{Z}$ jumping at unit rate,

$$p_t(x) := \mathbf{P}_0[S_t = x].$$

Effects of fluctuations on ancestral lineages. Taking our expressions for the rate of jumps of ancestral lineages and expanding about u_* reveals two key (compound) parameters determine the first order effect of the fluctuations on the motion of ancestral lineages. The first is $\Phi'(u_*)/\Phi(u_*)$, which determines whether a lineage prefers to jump towards or away from more crowded demes. The second is $(r\gamma)'(u_*)/(r\gamma)(u_*)$ which speeds up or slows down the jump rate out of more crowded demes according to whether it is negative or positive. We close with a brief discussion in §1.5.5.

1.3. Scaling limit for X^n.

In this section we establish our scaling limits for the (forwards in time) population process. Our approach owes a great deal to Perkins [108] and Sturm [120, 121].

The proof of Theorem 1.2.5 follows a standard pattern. In §1.3.1 we establish C-tightness in the product topology. In the process we shall recast the law of X^n as the solution to an approximate martingale problem. In §1.3.2, we characterise the limit points in terms of a martingale problem which we show is equivalent to saying that we have a weak limit to the system of stochastic differential equations (1.2.2).

To prove Theorem 1.2.6, we must establish tightness in $\ell_{\Gamma,p}$. Compact containment is a more exacting requirement here than in the product topology. We must establish a condition that plays the rôle of uniform integrability to guarantee convergence of subsequences in this topology. For this, borrowing ideas from Sturm [120, 121], we exploit the condition (1.2.3) on the initial value of the system. This programme is carried out in §1.3.3.

Proposition 1.2.8, uniqueness of limit points, is established in §1.3.4.

1.3.1. C-tightness of X^n in the product topology. In proving C-tightness, we shall recast the process as the solution to an approximate martingale problem. The first step will be to identify the putative martingales. The integrability that guarantees that they are in fact square integrable martingales follows from the comparison to branching random walk and the calculations in the proof of Lemma 1.3.2 below.

Let us write $A_n = O_{\mathbf{P}}(B_n)$ if and only if for each $\varepsilon > 0$, there exists $K > 0$ such that

$$\sup_n \mathbf{P}\big[|A_n/B_n| > K\big] < \varepsilon.$$

For each $x \in \mathbf{Z}$, define the process $M^n(x)$ by

$$M_t^n(x) := X_t^n(x) - X_0^n(x) - \int_0^{\underline{t}} g_n(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s \\ - \frac{m}{2}\int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x))\Delta_1(\Phi_n \circ X_{\underline{s}}^n)(x)\,\mathrm{d}s.$$

The idea is that this will be an $\mathcal{F}_t^n$-martingale with quadratic variation

$$\delta_n \int_0^{\underline{t}} \mu_n(X_{\underline{s}}^n(x))\big(2 - \mu_n(X_{\underline{s}}^n(x))\big) X_{\underline{s}}^n(x)\,\mathrm{d}s + O_{\mathbf{P}}(\eta_n).$$

Moreover, for all $x, y \in \mathbf{Z}$,

$$\mathbf{E}[M_t^n(x) M_t^n(y)] = 0.$$

To establish the martingale property, let us define $b^n : \mathbf{R}_+^{\mathbf{Z}} \times \mathbf{Z} \to \mathbf{R}$ and $a^n : \mathbf{R}_+^{\mathbf{Z}} \times \mathbf{Z} \times$

$\mathbf{Z} \to \mathbf{R}$ by

$$b^n(u,x) := \frac{1}{\tau_n}\mathbf{E}\left[X^n_{\tau_n}(x) - X^n_0(x) \middle| X^n_0(\cdot) = u(\cdot)\right],$$
$$a^n(u,x,y) := \frac{1}{\tau_n}\mathrm{cov}\left[X^n_{\tau_n}(x) - X^n_0(x), X^n_{\tau_n}(y) - X^n_0(y) \middle| X^n_0(\cdot) = u(\cdot)\right].$$

By construction, for each $x, y \in \mathbf{Z}$,

$$M^n_t(x) := X^n_t(x) - X^n_0(x) - \sum_{0\le \underline{s} < \underline{t}} \tau_n b^n(X^n_{\underline{s}}, x),$$
$$M^n_t(x)M^n_t(y) - \sum_{0\le \underline{s} < \underline{t}} \tau_n a^n(X^n_{\underline{s}}, x, y),$$

both satisfy the martingale property. We are going to need expressions for a^n and b^n in terms of the parameters of our model.

Summing over all surviving offspring of individuals alive at time $\underline{t}$ that establish at x, we find

$$X^n_{\underline{t}+\tau_n}(x) = \eta_n \sum_{\alpha\sim\underline{t}} \left\{ \mathbf{1}_x(Y^\alpha)(N^\alpha_0 + N^\alpha_=) + \mathbf{1}_x(Y^\alpha+1)N^\alpha_+ + \mathbf{1}_x(Y^\alpha-1)N^\alpha_-\right\}.$$

If $\alpha \sim \underline{t}$, then Y^α is $\mathcal{F}^n_{\underline{t}}$-measurable, Given this, and recalling the definition of g_n from equation (1.2.1),

$$\begin{aligned}
&\mathbf{E}[X^n_{\underline{t}+\tau_n}(x)|\mathcal{F}^n_{\underline{t}}]\\
&= \eta_n \sum_{\alpha\sim\underline{t}} \big(\mathbf{1}_x(Y^\alpha)\mathbf{E}[N^\alpha_0 + N^\alpha_=|\mathcal{F}^n_{\underline{t}}]\\
&\qquad + \mathbf{1}_x(Y^\alpha+1)\mathbf{E}[N^\alpha_+|\mathcal{F}^n_{\underline{t}}] + \mathbf{1}_x(Y^\alpha-1)\mathbf{E}[N^\alpha_-|\mathcal{F}^n_{\underline{t}}]\big)\\
&= \eta_n \sum_{\alpha\sim\underline{t}} \big(\mathbf{1}_x(Y^\alpha)(1-\mu_n(X^n_{\underline{t}}(Y^\alpha)))\\
&\qquad + \mathbf{1}_x(Y^\alpha)(1-m\tau_n)r_n(X^n_{\underline{t}}(Y^\alpha))\gamma_n(X^n_{\underline{t}}(Y^\alpha)))\\
&\qquad + \mathbf{1}_x(Y^\alpha+1)\frac{m}{2}\tau_n r_n(X^n_{\underline{t}}(Y^\alpha+1))\gamma_n(X^n_{\underline{t}}(Y^\alpha))\\
&\qquad + \mathbf{1}_x(Y^\alpha-1)\frac{m}{2}\tau_n r_n(X^n_{\underline{t}}(Y^\alpha-1))\gamma_n(X^n_{\underline{t}}(Y^\alpha))\big)\\
&= (1-\mu_n(X^n_{\underline{t}}(x)))X^n_{\underline{t}}(x) + (1-m\tau_n)r_n(X^n_{\underline{t}}(x))\gamma_n(X^n_{\underline{t}}(x))X^n_{\underline{t}}(x)\\
&\qquad + \frac{m}{2}\tau_n r_n(X^n_{\underline{t}}(x))\gamma_n(X^n_{\underline{t}}(x-1))X^n_{\underline{t}}(x-1)\\
&\qquad + \frac{m}{2}\tau_n r_n(X^n_{\underline{t}}(x))\gamma_n(X^n_{\underline{t}}(x+1))X^n_{\underline{t}}(x+1)\\
&= X^n_{\underline{t}}(x) + \tau_n\Big\{ g_n(X^n_{\underline{t}}(x))X^n_{\underline{t}}(x) + \frac{m}{2}r_n(X^n_{\underline{t}}(x))\Delta_1(\gamma_n(X^n_{\underline{t}})X^n_{\underline{t}})(x)\Big\}.
\end{aligned}$$

Thus

$$b^n(u,x) = g_n(u(x))u(x) + \frac{m}{2} r_n(u(x))\Delta_1(\Phi_n \circ u)(x). \tag{1.3.1}$$

For the covariance, writing

$$\begin{aligned} &X^n_{\underline{t}+\tau_n}(x) - X^n_{\underline{t}}(x) \\ &\qquad = \eta_n \sum_{\alpha\sim\underline{t}} \big(\mathbf{1}_x(Y^\alpha)(N_0^\alpha - 1 + N_=^\alpha) + \mathbf{1}_x(Y^\alpha + 1)N_+^\alpha + \mathbf{1}_x(Y^\alpha - 1)N_-^\alpha\big), \end{aligned}$$

and recalling that for $\alpha \sim \underline{t}$, the Poisson variables N_-^α, $N_=^\alpha$, N_+^α and N_0^α, $\alpha \sim \underline{t}$ are conditionally independent given $\mathcal{F}^n_{\underline{t}}$, we find

$$\begin{aligned} &\operatorname{cov}\Big[X^n_{\underline{t}+\tau_n}(x) - X^n_{\underline{t}}(x), X^n_{\underline{t}+\tau_n}(y) - X^n_{\underline{t}}(y) \Big| \mathcal{F}^n_{\underline{t}} \Big] \\ &= \mathbf{1}_x(y)\eta_n^2 \sum_{\alpha\sim\underline{t}} \Bigg[\mathbf{1}_x(Y^\alpha)\Big\{\mu_n(X^n_{\underline{t}}(Y^\alpha))\big(1-\mu_n(X^n_{\underline{t}}(Y^\alpha))\big) \\ &\qquad\qquad + (1-m\tau_n) r_n(X^n_{\underline{t}}(Y^\alpha))\gamma_n(X^n_{\underline{t}}(Y^\alpha))\Big\} \\ &\qquad\qquad + \mathbf{1}_x(Y^\alpha+1)\frac{m}{2}\tau_n r_n(X^n_0(Y^\alpha+1))\gamma_n(X^n_0(Y^\alpha)) \\ &\qquad\qquad + \mathbf{1}_x(Y^\alpha-1)\frac{m}{2}\tau_n r_n(X^n_0(Y^\alpha-1))\gamma_n(X^n_0(Y^\alpha)) \Bigg] \\ &= \mathbf{1}_x(y)\eta_n\mu_n(X^n_{\underline{t}}(x))\big(2-\mu_n(X^n_{\underline{t}}(x))\big)X^n_{\underline{t}}(x) \\ &\qquad\qquad + \mathbf{1}_x(y)\eta_n\tau_n\left(g_n(X^n_{\underline{t}}(x))X^n_{\underline{t}}(x) + \frac{m}{2} r_n(X^n_{\underline{t}}(x))\Delta_1(\gamma_n(X^n_{\underline{t}})X^n_{\underline{t}})(x) \right). \end{aligned} \tag{1.3.2}$$

We note that the conditional independence guarantees that the covariance is zero when $x \neq y$.

In particular, we identify

$$a^n(u,x,y) = \mathbf{1}_x(y)\big\{\delta_n\mu_n(u(x))(2-\mu_n(u(x)))u(x) + \eta_n b^n(u,x)\big\}. \tag{1.3.3}$$

It will be useful to note that

$$|a^n(u,x,y)| \leq \mathbf{1}_x(y)\Big(2\delta_n u(x) + \eta_n |b^n(u,x)|\Big). \tag{1.3.4}$$

We essentially follow Perkins [108]. The first step will be to bound the (putative) mar-

tingale differences $|M_{\underline{t}+\tau_n}(x) - M_{\underline{t}}(x)|$. Then we use Grönwall's Lemma to bound

$$\sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{0\le t\le T} \|X_t^n\|_{\Gamma,p}^p\right] \tag{1.3.5}$$

for $p \ge 2$ (provided that it is bounded at time zero). Moreover, substituting back into our bound on the martingale differences, yields an improved bound.

Our proof of C-tightness in the product topology will require us to prove, in particular, that X_t^n is tight for any fixed t. The bound (1.3.5) doesn't guarantee tightness of X_t^n in $\ell_{\Gamma,p}$, but it does suffice to give tightness in the product topology on $\mathbf{R}_+^{\mathbf{Z}^d}$.

NOTATION 1.3.1. In the calculations that follow, the constant C occurring in our inequalities will be a function of its parameters that may vary from line to line.

LEMMA 1.3.2 (Control of the martingale differences I). *Let $p \ge 2$. There exists a constant $C \equiv C(p, \overline{\gamma}, \overline{g}, \delta) < \infty$ such that for all $0 \le t \le T$,*

$$\begin{aligned}&\mathbf{E}\left[\sup_{0\le \underline{s}\le t} |M_{\underline{s}+\tau_n}^n(x) - M_{\underline{s}}^n(x)|^p\right]\\ &\qquad \le C\eta_n^{p/2-1}\mathbf{E}\left[\int_0^t \left(1 + |X_{\underline{s}}^n(x-1)|^{p/2} + |X_{\underline{s}}^n(x)|^{p/2} + |X_{\underline{s}}^n(x+1)|^{p/2}\right) \mathrm{d}s\right].\end{aligned}$$

PROOF. Observe that

$$\begin{aligned}M_{\underline{s}+\tau_n}^n(x) - M_{\underline{s}}^n(x) &= X_{\underline{s}+\tau_n}^n(x) - \mathbf{E}\left[X_{\underline{s}+\tau_n}^n(x)|\mathcal{F}_{\underline{s}}^n\right]\\ &= \sum_{\alpha\sim\underline{s}:Y^\alpha=x} \eta_n\left(N_0^\alpha - \mathbf{E}[N_0^\alpha|\mathcal{F}_{\underline{s}}^n] + N_=^\alpha - \mathbf{E}[N_=^\alpha|\mathcal{F}_{\underline{s}}^n]\right)\\ &\quad + \sum_{\alpha\sim\underline{s}:Y^\alpha=x-1} \eta_n\left(N_+^\alpha - \mathbf{E}[N_+^\alpha|\mathcal{F}_{\underline{s}}^n]\right) + \sum_{\alpha\sim\underline{s}:Y^\alpha=x+1} \eta_n\left(N_-^\alpha - \mathbf{E}[N_-^\alpha|\mathcal{F}_{\underline{s}}^n]\right).\end{aligned}$$

We borrow a trick from Perkins [108, Chap. II, Sec. 4]. The right hand side is a sum of $\eta_n^{-1}(X_{\underline{s}}^n(x) + X_{\underline{s}}^n(x-1) + X_{\underline{s}}^n(x+1))$ terms, which are conditionally independent and mean zero given $\mathcal{F}_{\underline{s}}^n$. As a result, we may view the sum as the terminal value of a discrete martingale, which we denote by $\mathcal{M}$, and to which we can apply Burkholder's predictable square function inequality, Lemma 1.6.1. Each increment $\mathcal{M}_k - \mathcal{M}_{k-1}$ of this martingale is a term in the sum of the form $\eta_n\left(N_*^\alpha - \mathbf{E}[N_*^\alpha|\mathcal{F}_{\underline{s}}]\right)$, where $* \in \{0, -, +, =\}$.

The terminal value of the martingale $\mathcal{M}$ is $X_{\underline{s}+\tau_n}^n - \mathbf{E}[X_{\underline{s}+\tau_n}^n|\mathcal{F}_{\underline{s}}]$ and so the terminal value of $\langle\mathcal{M}\rangle$ is $\mathrm{var}(X_{\underline{s}+\tau_n}^n(x)|\mathcal{F}_{\underline{s}}^n)$. Bounding the supremum of the moduli of the summands

above by their sum, Burkholder's predictable square function inequality gives

$$\begin{aligned}
&\mathbf{E}\left[\left|M^n_{\underline{s}+\tau_n}(x) - M^n_{\underline{s}}(x)\right|^p \middle| \mathcal{F}^n_{\underline{s}}\right] \\
&\le C(p)\Big(\mathrm{var}(X^n_{\underline{s}+\tau_n}(x)|\mathcal{F}^n_{\underline{s}})^{p/2} \\
&\quad + \sum_{\alpha\sim\underline{s}:Y^\alpha=x} \eta_n^p\Big(\mathbf{E}\left[\left|N^\alpha_0 - \mathbf{E}[N^\alpha_0|\mathcal{F}^n_{\underline{s}}]\right|^p \middle| \mathcal{F}^n_{\underline{s}}\right] + \mathbf{E}\left[\left|N^\alpha_= - \mathbf{E}[N^\alpha_=|\mathcal{F}^n_{\underline{s}}]\right|^p \middle| \mathcal{F}^n_{\underline{s}}\right]\Big) \\
&\quad + \sum_{\alpha\sim\underline{s}:Y^\alpha=x-1} \eta_n^p\Big(\mathbf{E}\left[\left|N^\alpha_+ - \mathbf{E}[N^\alpha_+|\mathcal{F}^n_{\underline{s}}]\right|^p \middle| \mathcal{F}^n_{\underline{s}}\right]\Big) \\
&\quad + \sum_{\alpha\sim\underline{s}:Y^\alpha=x+1} \eta_n^p\Big(\mathbf{E}\left[\left|N^\alpha_- - \mathbf{E}[N^\alpha_-|\mathcal{F}^n_{\underline{s}}]\right|^p \middle| \mathcal{F}^n_{\underline{s}}\right]\Big).
\end{aligned}$$

Note that $N^\alpha_0 \le 1$ for all α, and, since the establishment probability is at most one, the parameters of the conditionally Poisson random variables N^α_-, $N^\alpha_=$, N^α_+ are uniformly bounded by $\overline{\gamma}$. Thus, the absolute central moments in the sum are bounded (uniformly) by a constant $C \equiv C(p,\overline{\gamma})$ and we can bound the three sums by that constant times the total number of terms. Notice that since $p \ge 2$, for any $X \ge 0$ we have $X \le 1 + X^{p/2}$, and for $\eta_n < 1$ we have $\eta_n^{p-1} \le \eta_n^{p/2}$. Combining these observations with the expression for a^n in (1.3.3),

$$\begin{aligned}
&\mathbf{E}\left[\left|M^n_{\underline{s}+\tau_n}(x) - M^n_{\underline{s}}(x)\right|^p \middle| \mathcal{F}^n_{\underline{s}}\right] \\
&\le C(p,\overline{\gamma})\Big\{\mathrm{var}(X^n_{\underline{s}+\tau_n}(x)|\mathcal{F}^n_{\underline{s}})^{p/2} + \eta_n^{p-1}\big(X^n_{\underline{s}}(x) + X^n_{\underline{s}}(x-1) + X^n_{\underline{s}}(x+1)\big)\Big\} \\
&= C(p,\overline{\gamma})\Big(\left|\eta_n\mu_n(X^n_{\underline{s}}(x))\big(2-\mu_n(X^n_{\underline{s}}(x))\big)X^n_{\underline{s}}(x) + \tau_n\eta_n b_n(X^n_{\underline{s}},x)\right|^{p/2} \\
&\qquad + \eta_n^{p-1}(X^n_{\underline{s}}(x) + X^n_{\underline{s}}(x-1) + X^n_{\underline{s}}(x+1))\Big) \\
&\le C(p,\overline{\gamma},\overline{g})\eta_n^{p/2}\Big(1 + |X^n_{\underline{s}}(x)|^{p/2} + |X^n_{\underline{s}}(x-1)|^{p/2} + |X^n_{\underline{s}}(x+1)|^{p/2}\Big).
\end{aligned} \tag{1.3.6}$$

To reach the final line we have used that

$$|b_n(X^n_{\underline{s}},x)| \le C(\overline{\gamma},\overline{g})(X^n_{\underline{s}}(x-1) + X^n_{\underline{s}}(x) + X^n_{\underline{s}}(x+1)),$$

which follows from (1.3.1).

Once again approximating the supremum of a sequence of non-negative terms by its

sum, using the tower property,

$$\begin{aligned}
&\mathbf{E}\left[\sup_{0\le \underline{s}\le t}|M^n_{\underline{s}+\tau_n}(x)-M^n_{\underline{s}}(x)|^p\right]\\
&\le \mathbf{E}\Big[\sum_{0\le \underline{s}\le t}|M^n_{\underline{s}+\tau_n}(x)-M^n_{\underline{s}}(x)|^p\Big]\\
&= \mathbf{E}\Big[\sum_{0\le \underline{s}\le t}\mathbf{E}\Big[\,|M^n_{\underline{s}+\tau_n}(x)-M^n_{\underline{s}}(x)|^p\,\Big|\,\mathcal{F}^n_{\underline{s}}\Big]\Big] \qquad (1.3.7)\\
&\le C(p,\overline{\gamma},\overline{g})\eta_n^{p/2}\tau_n^{-1}\times\\
&\qquad \mathbf{E}\Big[\int_0^t\Big(1+|X^n_{\underline{s}}(x-1)|^{p/2}+|X^n_{\underline{s}}(x)|^{p/2}+|X^n_{\underline{s}}(x+1)|^{p/2}\Big)\,\mathrm{d}s\Big].
\end{aligned}$$

Now $\eta_n=\tau_n\delta_n$, so $\eta_n^{p/2}/\tau_n=\eta_n^{p/2-1}\delta_n$, and, since δ_n was assumed bounded, the proof is complete. □

Observe that, since $M^n_0(x)=0$, we have

$$\sup_{0\le s\le t}|M^n_s(x)|^p\le C(n,p)\sum_{0\le \underline{s}\le t}|M^n_{\underline{s}+\tau_n}(x)-M^n_{\underline{s}}(x)|^p,$$

and (since X^n is bounded above by a branching random walk with Poisson offspring distribution with parameter $\overline{\gamma}$) we can deduce from Lemma 1.3.2 that for each fixed n, $(M^n_t(x))_{0\le t\le T}$ is a square integrable martingale. We shall use this in the proof of the next lemma.

We now bound the expected value of the supremum of the $\ell_{\Gamma,p}$-norm of X^n_t for $t\in[0,T]$, and thus, a fortiori, $\mathbf{E}[\sup_{0\le t\le T}|X^n_t(x)|^p]$ for fixed $x\in\mathbf{Z}$.

LEMMA 1.3.3 ($\ell_{\Gamma,p}$ estimate). *Let $p\ge 2$. If*

$$\sup_{n\in\mathbf{N}}\mathbf{E}[\|X^n_0\|^p_{\Gamma,p}]<\infty \qquad (1.3.8)$$

then there exists a constant $C\equiv C(p,\overline{g},\overline{\gamma},\delta,T,\Gamma)<\infty$ such that

$$\sup_{n\in\mathbf{N}}\mathbf{E}\big[\sup_{0\le t\le T}\|X^n_t\|^p_{\Gamma,p}\big]\le C.$$

PROOF. The proof is an application of Grönwall's Lemma. Define

$$h^n(t):=\mathbf{E}\big[\sup_{0\le s\le t}\|X^n_s\|^p_{\Gamma,p}\big],\qquad 0\le t\le T.$$

We write X^n as a solution to the stochastic equation

$$X^n_s = X^n_0 + \int_0^{\underline{s}} b^n(X^n_{\underline{u}}, \cdot)\,\mathrm{d}u + M^n_{\underline{s}}. \tag{1.3.9}$$

Substituting (1.3.9) into the definition of $\|\cdot\|_{\Gamma,p}$, for some constant $C(p) < \infty$,

$$\begin{aligned} h^n(t) &= \mathbf{E}\big[\sup_{0\le s\le t}\|X^n_s\|^p_{\Gamma,p}\big] \le \mathbf{E}\Big[\textstyle\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}|X^n_s(x)|^p\Big] \\ &\le C(p)\mathbf{E}\Big[\|X^n_0\|^p_{\Gamma,p} + \sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\Big|\int_0^{\underline{s}} b^n(X^n_{\underline{u}},x)\,\mathrm{d}u\Big|^p + \sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}|M^n_{\underline{s}}(x)|^p\Big]. \end{aligned} \tag{1.3.10}$$

Let us begin by bounding the second term on the right hand side of (1.3.10). As before, we use (1.3.1), from which it follows that

$$|b^n(u,x)| \le C(\overline{g},\overline{\gamma})\Big(u(x-1)+u(x)+u(x+1)\Big). \tag{1.3.11}$$

Applying Jensen's inequality,

$$\begin{aligned} &\Big|\int_0^{\underline{s}} b^n(X^n_{\underline{u}},x)\,\mathrm{d}u\Big|^p \\ &\le \underline{s}^{p-1}\int_0^{\underline{s}} C(\overline{g},\overline{\gamma})^p\big|X^n_{\underline{u}}(x-1)+X^n_{\underline{u}}(x)+X^n_{\underline{u}}(x+1))\big|^p\,\mathrm{d}u \\ &\le C(p,\overline{g},\overline{\gamma},T)\int_0^{\underline{s}}\big(|X^n_{\underline{u}}(x-1)|^p+|X^n_{\underline{u}}(x)|^p+|X^n_{\underline{u}}(x+1)|^p\big)\,\mathrm{d}u. \end{aligned}$$

In particular, since, under our assumptions on the weight function in Definition 1.2.1, the ratios $\Gamma(x+1)/\Gamma(x)$ and $\Gamma(x-1)/\Gamma(x)$ are uniformly bounded,

$$\begin{aligned} &\mathbf{E}\Big[\sup_{0\le s\le t}\sum_{x\in\mathbf{Z}}\Gamma(x)\Big|\int_0^{\underline{s}} b^n(X^n_{\underline{u}},x)\,\mathrm{d}u\Big|^p\Big] \\ &\le \mathbf{E}\Big[\sup_{0\le s\le t}\int_0^{\underline{s}} C(p,\overline{g},\overline{\gamma},T)\sum_{x\in\mathbf{Z}}\Gamma(x)\Big(|X^n_{\underline{u}}(x-1)|^p+|X^n_{\underline{u}}(x)|^p+|X^n_{\underline{u}}(x+1)|^p\Big)\,\mathrm{d}u\Big] \\ &\le C(p,\overline{g},\overline{\gamma},T,\Gamma)\mathbf{E}\Big[\int_0^t\|X^n_{\underline{u}}\|^p_{\Gamma,p}\,\mathrm{d}u\Big] \\ &\le C(p,\overline{g},\overline{\gamma},T,\Gamma)\mathbf{E}\Big[\int_0^t\sup_{0\le s\le u}\|X^n_{\underline{s}}\|^p_{\Gamma,p}\,\mathrm{d}u\Big] \\ &= C(p,\overline{g},\overline{\gamma},T,\Gamma)\int_0^t h^n(u)\,\mathrm{d}u. \end{aligned} \tag{1.3.12}$$

Since the process X^n only experiences a finite number of jumps in $[0,t]$, the apparent interchange of integration over time and summation and expectation in the second and last steps is, in reality, interchanging a finite sum with the sum and the expectation. The key point is that the constant is independent of n.

We now turn to the third term of (1.3.10). Applying the Burkholder predictable square function inequality (Lemma 1.6.1 in the appendix) to the martingale $M^n(x)$, there exists a constant $C(p) < \infty$ such that

$$\begin{aligned}\mathbf{E}\Big[\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}|M^n_{\underline{s}}(x)|^p\Big] &= \sum_{x\in\mathbf{Z}}\Gamma(x)\mathbf{E}\big[\sup_{0\le s\le t}|M^n_{\underline{s}}(x)|^p\big]\\ &\le \sum_{x\in\mathbf{Z}}\Gamma(x)C(p)\mathbf{E}\Big[\Big|\int_0^{\underline{t}}a^n(X^n_{\underline{s}},x,x)\,\mathrm{d}s\Big|^{p/2}+\sup_{0\le s\le t}|M^n_{\underline{t}+\tau_n}(x)-M^n_{\underline{t}}(x)|^p\Big],\end{aligned}$$

where a^n is given by (1.3.3).

Thus, using the bound (1.3.4) on a^n and Lemma 1.3.2, we find

$$\begin{aligned}&\mathbf{E}\Big[\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}|M^n_{\underline{s}}(x)|^p\Big]\\ &\le C(p,\overline{g},\overline{\gamma},\delta)\sum_{x\in\mathbf{Z}}\Gamma(x)\mathbf{E}\Big[\int_0^t\Big(|X^n_{\underline{s}}(x-1)|^{p/2}+|X^n_{\underline{s}}(x)|^{p/2}+|X^n_{\underline{s}}(x+1)|^{p/2}\Big)\,\mathrm{d}s\\ &\qquad+\eta_n^{p/2-1}\int_0^t\Big(1+|X^n_{\underline{s}}(x-1)|^{p/2}+|X^n_{\underline{s}}(x)|^{p/2}+|X^n_{\underline{s}}(x+1)|^{p/2}\Big)\,\mathrm{d}s\Big]\\ &\le C(p,\overline{g},\overline{\gamma},\delta)\sum_{x\in\mathbf{Z}}\Gamma(x)\mathbf{E}\Big[\int_0^t\Big(1+|X^n_{\underline{s}}(x-1)|^{p}+|X^n_{\underline{s}}(x)|^{p}+|X^n_{\underline{s}}(x+1)|^{p}\Big)\,\mathrm{d}s\Big]\\ &\le C(p,\overline{g},\overline{\gamma},\delta,T,\Gamma)\Big[1+\int_0^t\mathbf{E}\big[\|X^n_{\underline{s}}\|^p_{\Gamma,p}\big]\,\mathrm{d}s\Big]\\ &\le C(p,\overline{g},\overline{\gamma},\delta,T,\Gamma)\Big[1+\int_0^t h^n(s)\,\mathrm{d}s\Big].\end{aligned}\tag{1.3.13}$$

To obtain the second inequality we have used the trivial bound $X \le 1 + X^2$ for $X \ge 0$. Substituting the bounds (1.3.12) and (1.3.13) in (1.3.10), there is a constant $C \equiv C(p,\overline{g},\overline{\gamma},\delta,T,\Gamma) < \infty$ such that

$$0 \le h^n(t) \le C\Big(1+\int_0^t h^n(s)\,\mathrm{d}s\Big).$$

Since C is independent of n, Grönwall's Lemma yields

$$\sup_{n\in\mathbf{N}}\mathbf{E}\big[\sup_{0\le s\le t}\|X^n_s\|^p_{\Gamma,p}\big] \le C_1\exp(C_2t) < \infty,$$

which is uniformly bounded for $0 \leq t \leq T$. $\square$

We characterise the law of X^n as a solution to an approximate martingale problem.

LEMMA 1.3.4 (Approximate martingale problem). *The law of X^n is a solution to the following approximate martingale problem.*

For each $x \in \mathbf{Z}$,

$$M_t^n(x) := X_t^n(x) - X_0^n(x) - \int_0^{\underline{t}} g_n(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s - \frac{m}{2}\int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x))\Delta_1(\Phi_n \circ X_{\underline{s}}^n)(x)\,\mathrm{d}s$$

is an $\mathcal{F}_t^n$-martingale such that

$$M_t^n(x)^2 - \delta_n \int_0^{\underline{t}} \mu_n(X_{\underline{s}}^n(x))\big(2 - \mu_n(X_{\underline{s}}^n(x))\big) X_{\underline{s}}^n(x)\,\mathrm{d}s + O_{\mathbf{P}}(\eta_n),$$

is also a martingale, and for each pair $x, y \in \mathbf{Z}$,

$$\mathbf{E}[M_t^n(x) M_t^n(y)] = 0.$$

PROOF. The martingale property is satisfied by construction and the required integrability and control of the error follow immediately from our bounds above. $\square$

Combining Lemma 1.3.3 with Lemma 1.3.2 provides further control of the jumps of $M^n(x)$.

LEMMA 1.3.5 (Control of the martingale differences II). *Let $p \geq 2$ and suppose that* (1.3.8) *holds. Then there exists a constant $C \equiv C(p, \overline{\gamma}, \overline{g}, \Gamma, T, \delta)$ such that*

$$\mathbf{E}\left[\sup_{0\leq t\leq T} \sum_{x\in\mathbf{Z}} \Gamma(x) |M_{\underline{t}+\tau_n}^n(x) - M_{\underline{t}}^n(x)|^{2p}\right] \leq C\,\eta_n^{p-1}.$$

Moreover, $\sup_{0\leq t\leq T} |M_{\underline{t}+\tau_n}^n(x) - M_{\underline{t}}^n(x)| = O_{\mathbf{P}}(\eta_n^{1/2-1/2p})$ *as* $n \to \infty$.

PROOF. This is a trivial consequence of Lemmas 1.3.2 and 1.3.3. By Lemma 1.3.2

(with p replaced by $2p$), there is a constant $C(p,\overline{\gamma},\overline{g},\delta)<\infty$ such that

$$\begin{aligned}
&\mathbf{E}\Big[\sup_{0\le \underline{s}\le t}\sum_{x\in\mathbf{Z}}\Gamma(x)|M^n_{\underline{t}+\tau_n}(x)-M^n_{\underline{t}}(x)|^{2p}\Big]\\
&\le C(p,\overline{\gamma},\overline{g},\delta)\eta_n^{p-1}\times\\
&\qquad\mathbf{E}\Big[\sum_{x\in\mathbf{Z}}\Gamma(x)\int_0^t(1+|X^n_{\underline{s}}(x-1)|^p+|X^n_{\underline{s}}(x)|^p+|X^n_{\underline{s}}(x+1)|^p)\,\mathrm{d}s\Big]\\
&\le C(p,\overline{\gamma},\overline{g},\delta,\Gamma,T)\eta_n^{p-1}\mathbf{E}\Big[1+\sup_{0\le t\le T}\|X^n_{\underline{t}}\|^p_{\Gamma,p}\Big],
\end{aligned}$$

and, by Lemma 1.3.3, the last line is bounded by $C(p,\overline{\gamma},\overline{g},\delta,\Gamma,T)\eta_n^{p-1}$. We used again that there are only finitely many jumps of X^n over $[0,t]$ so that we can interchange 'integration' over time (really a finite sum) with summation and expectation.

The final statement of the lemma follows from Markov's inequality and Lemma 1.3.3:

$$\begin{aligned}
&\mathbf{P}\Big[\sup_{0\le t\le T}|M^n_{\underline{t}+\tau_n}(x)-M^n_{\underline{t}}(x)|>K\eta_n^{\frac12-\frac{1}{2p}}\Big]\\
&=\mathbf{P}\Big[\sup_{0\le t\le T}\Gamma(x)|M^n_{\underline{t}+\tau_n}(x)-M^n_{\underline{t}}(x)|^{2p}>\Gamma(x)K^{2p}\eta_n^{p-1}\Big]\\
&\le\frac{\mathbf{E}\Big[\sup_{0\le t\le T}\Gamma(x)|M^n_{\underline{t}+\tau_n}(x)-M^n_{\underline{t}}(x)|^{2p}\Big]}{\Gamma(x)K^{2p}\eta_n^{p-1}}\\
&\le\frac{C(p,\overline{\gamma},\overline{g},\delta,T)}{\Gamma(x)}\frac{1}{K^{2p}}.
\end{aligned}$$

□

We now prove C-tightness of X^n in the product topology. For ease of reference, we record the criteria that we are going to use in order to verify C-tightness. They are a combination of Theorem 3.8.6 of Ethier and Kurtz [40] and Proposition VI.3.26 of Jacod and Shiryaev [67].

THEOREM 1.3.6. *Let (E,d) be a complete and separable metric space and let X^n be a sequence of processes with paths in $D([0,\infty),E)$. The sequence is tight in $D([0,\infty),E)$ if the following conditions hold:*

1. *X^n_t is tight for every rational $t\ge 0$.*

2. *For every $T > 0$, there exists $p > 0$ such that*

$$\lim_{\theta\to 0}\sup_n \mathbf{E}\left[\sup_{0\le t\le T}\sup_{\substack{0\le u\le\theta\\ 0\le v\le\theta\wedge t}}\mathbf{E}\left[\left(d(X^n_{t+u},X^n_t)\wedge 1\right)^p\left(d(X^n_{t-v},X^n_t)\wedge 1\right)^p|\mathcal{F}^n_t\right]\right]=0. \tag{1.3.14}$$

The sequence of processes is C-tight in $D([0,\infty),E)$ if it is tight and satisfies

$$\lim_{n\to\infty}\mathbf{P}\left[\sup_{0\le t\le T} d\left(X^n_{t-},X^n_t\right)>\varepsilon\right]=0 \tag{1.3.15}$$

for all $T\in\mathbf{N}$ and $\varepsilon>0$.

PROPOSITION 1.3.7 (Tightness of X^n). *Let $\mathbf{R}_+^{\mathbf{Z}}$ be equipped with the product topology. If for some $p\ge 2$ we have* (1.3.8)*, then the sequence of processes $(X^n)_{n\in\mathbf{N}}$ is C-tight in $D([0,T],\mathbf{R}_+^{\mathbf{Z}})$.*

PROOF. It will suffice to show that for each $x\in\mathbf{Z}$, $\{(X^n_t(x)_{0\le t\le T}\}_{n\ge 1}$ is tight in $D([0,T],\mathbf{R}_+)$ and the limit points are continuous, since then if we consider the pair of coordinates $\{X^n(x),X^n(x')\}$, there is a convergent subsequence $X^{n_k}(x)$, say, at x and by taking a further subsequence if necessary we can find a subsequence which converges both at x and x'. Evidently this diagonalisation argument carries over to any finite collection of points and hence yields C-tightness in the product topology.

We fix $x\in\mathbf{Z}$ and consider the sequence $X^n(x)$. Tightness of $X^n_t(x)$ for fixed t and x is trivially satisfied by an application of Markov's inequality and Lemma 1.3.3.

Now consider the increments of $X^n(x)$. In order to check the second condition of Theorem 1.3.6, it suffices to show that for $\theta<1$,

$$\sup_n\mathbf{E}\left[\sup_{0\le t\le T}\sup_{\substack{0\le u\le\theta\\ 0\le v\le\theta\wedge t}}\mathbf{E}\left[\,|X^n_{t+u}(x)-X^n_t(x)|^p\,\middle|\,\mathcal{F}^n_t\right]\right]\le C(p,\overline{g},\overline{\gamma},\Gamma,T)\theta, \tag{1.3.16}$$

where $C(p,\overline{g},\overline{\gamma},\Gamma,T)$ is independent of n, and indeed this also guarantees that condition (1.3.15) is satisfied.

Noting that X^n is a Markov process, we first consider

$$\mathbf{E}\left[\,|X^n_{t+u}(x)-X^n_t(x)|^p\,\middle|\,\mathcal{F}^n_t\right]. \tag{1.3.17}$$

Using the stochastic equation (1.3.9), write

$$X^n_{t+u}(x)-X^n_t(x)=\int_{\underline{t}}^{\underline{t+u}} b^n(X^n_s,x)\,\mathrm{d}s+M^n_{t+u}(x)-M^n_t(x).$$

We shall bound the contribution to (1.3.17) of the two terms on the right hand side separately. For the first, we can use the bound (1.3.11) on b^n and Jensen's inequality to obtain that; for the second we apply Burkholder's predictable square function inequality to the martingale term $M^n_{t+u}(x) - M^n_t(x)$ to obtain

$$\mathbf{E}\left[\sup_{t\le u\le t+\theta}|M^n_u(x)-M^n_t(x)|^p\,\middle|\,\mathcal{F}^n_t\right] \le C(p)\left\{\mathbf{E}\left[\left|\int_t^{t+\theta} a^n(X^n_u,x)\,\mathrm{d}u\right|^{p/2}\,\middle|\,\mathcal{F}^n_t\right] + \mathbf{E}\left[\sup_{t\le \underline{s}\le t+\theta}\left|M^n_{\underline{s}+\tau_n}(x)-M^n_{\underline{s}}(x)\right|^p\,\middle|\,\mathcal{F}^n_t\right]\right\}. \quad (1.3.18)$$

After a simple modification to the proof of Lemma 1.3.2 (to allow us to replace 0 and t in the statement of the result by t and $t+\theta$ under the conditional expectation), we obtain that for some constant $C \equiv C(p,\overline{\gamma},\overline{g},\delta) < \infty$,

$$\mathbf{E}\left[\sup_{t\le \underline{s}\le t+\theta}|M^n_{\underline{s}+\tau_n}(x)-M^n_{\underline{s}}(x)|^p\,\middle|\,\mathcal{F}^n_t\right] \le C\eta_n^{p/2-1}\mathbf{E}\left[\int_{\underline{t}}^{t+\theta}\left(1+|X^n_{\underline{s}}(x-1)|^{p/2}+|X^n_{\underline{s}}(x)|^{p/2}+|X^n_{\underline{s}}(x+1)|^{p/2}\right)\mathrm{d}s\,\middle|\,\mathcal{F}^n_t\right].$$

Combining these estimates with the bound (1.3.4) on a^n and applying Jensen's inequality to the integrals over time we find that (1.3.17) is bounded above by

$$C(p,\overline{g},\overline{\gamma},\delta,\Gamma,T)\left(\theta^p\mathbf{E}\left[\sup_{0\le u\le T}|X^n_u(x)|^p\right]+\theta^{p/2}\mathbf{E}\left[\sup_{0\le u\le T}|X^n_u(x)|^{p/2}\right] + \eta_n^{p/2-1}\theta\mathbf{E}\left[\sup_{0\le u\le T}|X^n_u(x)|^{p/2}\right]\right),$$

where the first term corresponds to controlling b^n, the second to a^n and the third to the martingale increments. Since $p \ge 2$, we obtain (1.3.16).

Condition (1.3.15) follows by an application of Markov's inequality. □

REMARK 1.3.8. We emphasize that tightness here is with respect to the product topology on $\mathbf{R}_+^{\mathbf{Z}}$. The stronger property of tightness with respect to the $\ell_{\Gamma,p}$-topology, which requires some 'uniform integrability' of the $X^n(x)$, will be proved under stronger (but biologically natural) restrictions on the initial condition in Proposition 1.3.15 below.

An almost identical argument yields the C-tightness of $M^n(x)$.

LEMMA 1.3.9 (Tightness of $M^n(x)$). *Under the conditions of Proposition 1.3.7, the sequence $\{M^n\}_{n\in\mathbf{N}}$ is C-tight in $D([0,T],\mathbf{R}^{\mathbf{Z}})$ with the product topology.*

PROOF. By, for example, Theorem VI.4.13 of Jacod and Shiryaev [67], in order to verify that $\{M^n\}_{n\in\mathbf{N}}$ is tight, it is sufficient to show that for each $x \in \mathbf{Z}$, the sequence of processes

$$\begin{aligned}\langle M^n(x)\rangle_t &= \int_0^{\underline{t}} a^n(X^n_{\underline{s}}, x, x)\,\mathrm{d}s \\ &= \delta_n \int_0^{\underline{t}} \mu_n(X^n_{\underline{s}}(x))\big(2-\mu_n(X^n_{\underline{s}}(x))\big)X^n_{\underline{s}}(x)\,\mathrm{d}s + \eta_n \int_0^{\underline{t}} b^n(X^n_{\underline{s}}, x)\,\mathrm{d}s\end{aligned}$$

is C-tight.

The calculations in the proof of Lemma 1.3.3 yield

$$\mathbf{E}\big[|\langle M^n(x)\rangle_t|^p\big] \le C(p,\overline{g},\overline{\gamma},\delta,T,\Gamma,x),$$

and, just as in the derivation of (1.3.16),

$$\begin{aligned}&\mathbf{E}\Big[\sup_{0\le t\le T}|\langle M^n(x)\rangle_t - \langle M^n(x)\rangle_{t-}|^p\Big] \\ &\le C(p)\mathbf{E}\Bigg[\sup_{0\le t\le T}\delta_n^p\tau_n^p\big|\mu_n(X^n_{\underline{t}-\tau_n}(x))\big(2-\mu_n(X^n_{\underline{t}-\tau_n}(x))\big)X^n_{\underline{t}-\tau_n}(x)\big|^p \\ &\qquad\qquad + \sup_{0\le t\le T}\eta_n^p\tau_n^p\big|b^n(X^n_{\underline{t}-\tau_n},x)\big|^p\Bigg] \\ &\le C(p,\overline{g},\overline{\gamma},\delta,T,\Gamma,x)\eta_n^p.\end{aligned}$$

Thus $\langle M^n(x)\rangle$ is C-tight.

To verify C-tightness of $M^n(x)$, it is immediate from Lemma 1.3.5 that for all $\varepsilon > 0$,

$$\lim_{n\to\infty}\mathbf{P}\Bigg[\sup_{0\le t\le T}\big|M^n_t(x) - M^n_{t-}(x)\big| > \varepsilon\Bigg] = 0,$$

and the proof is complete. □

1.3.2. Proof of Theorem 1.2.5. We are now in a position to prove Theorem 1.2.5.

From the proof of Lemma II.4.5 of Perkins [108], the following lemma is enough to guarantee that any weak limit point of the sequence of martingales $M^n(x)$ is an L^2-martingale.

LEMMA 1.3.10. *Suppose that* (1.3.8) *holds for some $p \ge 2$. Then for each $x \in \mathbf{Z}$ and $0 \le t \le T$, the family of random variables $(|M^n_{\underline{t}}(x)|^2)_{n\in\mathbf{N}}$ is uniformly integrable.*

PROOF. We have already done all the work. It is enough to show that

$$\sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{0\le t\le T} |M^n_{\underline{t}}(x)|^{2p}\right] < \infty.$$

Setting $t = 0$ in (1.3.18), and replacing p by $2p$, the result follows from our bound (1.3.4), Lemma 1.3.2, and Lemma 1.3.3, exactly as in the proof of Proposition 1.3.7. □

The next step is to characterise any limit point of $(X^n)_{n\in\mathbf{N}}$ as a solution to a martingale problem.

PROPOSITION 1.3.11 (Limiting martingale problem). *Suppose that $X^n_0 \Rightarrow u$ in $\ell_{\Gamma,p}$ for some $p \ge 2$. Then $(X^n)_{n\in\mathbf{N}}$ is C-tight in $D([0,T],\mathbf{R}^{\mathbf{Z}}_+)$ and any weak limit point X solves the following martingale problem:*

$$\begin{cases} \text{For each } x \in \mathbf{Z}, \\ M_t(x) := X_t(x) - X_0(x) - \int_0^t g(X_s(x))X_s(x)\,\mathrm{d}s \\ +\frac{m}{2}\int_0^t r(X_s(x))\Delta_1(\Phi\circ X_s)(x)\,\mathrm{d}s \\ \text{is an } \mathcal{F}_t\text{-martingale such that} \\ M_t(x)^2 - \int_0^t \delta\mu(X_s(x))\big(2-\mu(X_s(x))\big)X_s(x)\,\mathrm{d}s, \\ \text{is also a martingale and for all } x, y \in \mathbf{Z}, \\ \mathbf{E}[M_t(x)M_t(y)] = 0. \end{cases} \tag{MP}$$

PROOF. The C-tightness is the content of Proposition 1.3.7.

After passing to a subsequence if necessary, assume that X^n converges weakly to X in $D([0,T],\mathbf{R}^{\mathbf{Z}}_+)$, where $\mathbf{R}^{\mathbf{Z}}_+$ is equipped with the product topology. By Lemma 1.3.4, for each $x \in \mathbf{Z}$,

$$M^n_{\underline{t}}(x) = X^n_t(x) - X^n_0(x) - \int_0^{\underline{t}} X^n_{\underline{s}}(x) g_n(X^n_{\underline{s}}(x))\,\mathrm{d}s \\ - \int_0^{\underline{t}} \frac{m}{2} r_n(X^n_{\underline{s}}(x))\Delta_1(\Phi_n\circ X^n_{\underline{s}})(x)\,\mathrm{d}s \tag{1.3.19}$$

is a martingale. Fix $x \in \mathbf{Z}$. Since $X^n \Rightarrow X$ in $D([0,T],\mathbf{R}^{\mathbf{Z}}_+)$, $(X^n(x-1), X^n(x), X^n(x+1)) \Rightarrow (X(x-1), X(x), X(x+1)$ in $D([0,T],\mathbf{R}^3_+)$ and, by the Skorokhod Representation Theorem, we can find a probability space, and on it a sequence $\widetilde{X}^n$, such that $\widetilde{X}^n$ has the same law as X^n and $\widetilde{X}^n$ converges to $\widetilde{X}$ almost surely in $D([0,T],\mathbf{R}^{\mathbf{Z}}_+)$. We suppose that we have done this, and drop the 'tilde' from our notation. By, for example, Lemma 3.10.1 of Ethier and Kurtz [40], if a sequence $(f_n)_{n\in\mathbf{N}} \subset D([0,T],\mathbf{R}_+)$ converges in the Skorokhod topology and the limit is continuous then $(f_n)_{n\in\mathbf{N}}$ also converges in the topology

of uniform convergence on compact sets. Since g_n, r_n and γ_n also converge uniformly on compacts, the right hand side of (1.3.19) converges almost surely to

$$M_t(x) := X_t(x) - X_0(x) - \int_0^t X_s(x) g(X_s(x))\,\mathrm{d}s - \int_0^t \frac{m}{2} r(X_s(x)) \Delta_1(\Phi \circ X_s)(x)\,\mathrm{d}s.$$

By Lemma 1.3.10, $M(x)$ is an L^2-martingale, and since

$$M_t^n(x)^2 - \delta_n \int_0^{\underline{t}} \mu_n(X_{\underline{s}}^n(x))\big(2 - \mu_n(X_{\underline{s}}^n(x))\big) X_{\underline{s}}^n(x)\,\mathrm{d}s - \eta_n \int_0^{\underline{t}} b^n(X_{\underline{s}}^n, x)\,\mathrm{d}s$$

is a martingale, and

$$\mathbf{E}\left[\sup_{0\le t\le T}\left|\eta_n \int_0^{\underline{t}} b^n(X_{\underline{s}}^n, x)\,\mathrm{d}s\right|^p\right] \le C(p, \overline{g}, \overline{\gamma}, T, \Gamma, x)\eta_n^p,$$

we identify

$$\langle M(x)\rangle_t = \delta \int_0^t \mu(X_s(x))\big(2 - \mu(X_s(x))\big) X_s(x)\,\mathrm{d}s.$$

Apart from minor modifications, the same argument gives convergence of the square integrable martingale $M^n(x) + M^n(y)$ to a square integrable martingale. Furthermore, since $\mathbf{E}[M_t^n(x) M_t^n(y)] = 0$ for $x \neq y$, a polarisation argument yields $\mathbf{E}[M_t(x) M_t(y)] = 0$. □

Proposition 1.3.11 characterised the weak limit points of the scaled individual based population models as solutions to a martingale problem. In fact, as we now show, all such solutions are also (weak) solutions to the system of SDEs (1.2.2).

LEMMA 1.3.12 (Equivalence of (MP) and (1.2.2)). *Any solution to the martingale problem (MP) is a $\mathbf{R}_+^{\mathbf{Z}}$-valued weak solution to the infinite system of stochastic differential equations* (1.2.2).

PROOF. The proof is standard. We have that

$$M_t(x) = X_t(x) - X_0(x) - \int_0^t X_s(x) g(X_s(x))\,\mathrm{d}s - \int_0^t \frac{m}{2} r(X_s(x)) \Delta_1(\Phi \circ X_s)(x)\,\mathrm{d}s,$$

$$M_t(x)^2 - \delta \int_0^t \mu(X_s(x))\big(2 - \mu(X_s(x))\big) X_s(x)\,\mathrm{d}s,$$

are both martingales. Let $\{\widetilde{B}(x)\}_{x\in\mathbf{Z}}$, independent of X, be a collection of independent Brownian motions. For each $x \in \mathbf{Z}$ define the process

$$B_t(x) := \int_{\substack{0\le s\le t\\ X_s(x)>0}} \big(\delta\mu(X_s(x))\big(2 - \mu(X_s(x))\big) X_s(x)\big)^{-1/2}\,\mathrm{d}M_s(x) + \int_{\substack{0\le s\le t\\ X_s(x)=0}} \mathrm{d}\widetilde{B}_s(x).$$

Then $\langle B(x)\rangle_t = t$, and $\langle B(x), B(y)\rangle_t = 0$ for $x \neq y$, so, by Lévy's characterisation, $\{B(x)\}_{x\in\mathbf{Z}}$ are a collection of independent Brownian motions. A simple calculation shows that

$$M_t(x) = \int_0^t \sqrt{\delta\mu(X_s(x))\big(2-\mu(X_s(x))\big)X_s(x)}\,\mathrm{d}B_s(x)$$

and thus (X, B) solves (1.2.2). □

Since we only have tightness in the product topology, to complete the proof of Theorem 1.2.5 we must check that if $X_0 \in \ell_{\Gamma,p}$ for some $p \geq 2$, then so is the solution X to the system (1.2.2).

LEMMA 1.3.13. *If $u \in \ell_{\Gamma,p}$, then any solution to* (1.2.2) *is $\ell_{\Gamma,p}$-valued.*

PROOF. The calculations that we need are analogous to those of Lemma 1.3.3, but here we have to work harder to justify interchange of orders of integration. First we consider X restricted to the interval $[-J, J] \cap \mathbf{Z}$.

Define the stopping time $\tau^{J,K} = \inf\{t \geq 0 : \|\mathbf{1}_{|\cdot|\leq J} X_t(\cdot)\|^p_{\Gamma,p} > K\}$ and set

$$h_{J,K}(t) := \mathbf{E}\Big[\sup_{0\leq s\leq t\wedge\tau^{J,K}} \|\mathbf{1}_{|\cdot|\leq J} X_s(\cdot)\|^p_{\Gamma,p}\Big], \qquad 0 \leq t \leq T.$$

By (1.2.2),

$$\begin{aligned}
&h_{J,K}(t)\\
&\leq \mathbf{E}\Big[\sum_{\substack{x\in\mathbf{Z}\\|x|\leq J}} \Gamma(x) \sup_{0\leq s\leq t\wedge\tau^{J,K}} |X_s(x)|^p\Big]\\
&\leq \mathbf{E}\Big[\|X_0\|^p_{\Gamma,p} + \sum_{\substack{x\in\mathbf{Z}\\|x|\leq J}} \Gamma(x) \sup_{0\leq s\leq t\wedge\tau^{J,K}} \Big|\int_0^s X_u(x) g(X_u(x))\,\mathrm{d}u\Big|^p\\
&\quad + \sum_{\substack{x\in\mathbf{Z}\\|x|\leq J}} \Gamma(x) \sup_{0\leq s\leq t\wedge\tau^{J,K}} \Big|\int_0^s \frac{1}{2} r(X_u(x))\Delta_1(\Phi\circ X_u)(x)\,\mathrm{d}u\Big|^p\\
&\quad + \sum_{\substack{x\in\mathbf{Z}\\|x|\leq J}} \Gamma(x) \sup_{0\leq s\leq t\wedge\tau^{J,K}} \Big|\int_0^s \sqrt{\delta\mu(X_u(x))\big(2-\mu(X_u(x))\big)X_u(x)}\,\mathrm{d}B_u(x)\Big|^p\Big].
\end{aligned}$$

Proceeding exactly as in the proof of Lemma 1.3.3,

$$h_{J,K}(t) \leq C(p, \overline{\gamma}, T, \Gamma)\Big(1 + \int_0^t h_{J,K}(s)\,\mathrm{d}s\Big),$$

and since the constant is independent of J, K, Grönwall's Lemma provides a uniform bound of $h_{J,K}(t)$ for $t \in [0, T]$. Since the sample paths of the solutions to (1.2.2) are

continuous, $X_t(x)$ is bounded on $[0, T]$ for each x and so for J fixed, $\mathbf{P}[\tau^{J,K} \leq T] \to 0$ as $K \to \infty$. As a result,

$$\lim_{K\to\infty} \sup_{0\leq t\leq T\wedge\tau^{J,K}} \|\mathbf{1}_{|\cdot|\leq J} X_t(\cdot)\|_{\Gamma,p}^p = \sup_{0\leq t\leq T} \|\mathbf{1}_{|\cdot|\leq J} X_t(\cdot)\|_{\Gamma,p}^p$$

almost surely, and so by Fatou's Lemma,

$$\sup_J \mathbf{E}\left[\sup_{0\leq t\leq T} \|\mathbf{1}_{|\cdot|\leq J} X_t(\cdot)\|_{\Gamma,p}^p\right] \leq \sup_J \liminf_{K\to\infty} h_{J,K}(T) \leq C(p, \overline{\gamma}, T, \Gamma).$$

Letting $J \to \infty$, monotone convergence yields

$$\mathbf{E}\left[\sup_{0\leq t\leq T} \|X_t\|_{\Gamma,p}^p\right] < \infty,$$

completing the proof. □

1.3.3. Proof of Theorem 1.2.6. So far we have taken arbitrary initial conditions in $\ell_{\Gamma,p}$ and, provided those initial conditions converge, shown tightness in the product topology. The stronger conclusion of tightness in the topology of $\ell_{\Gamma,p}$ requires us to prove a uniform integrability condition.

LEMMA 1.3.14. *A set $K \subseteq \ell_{\Gamma,p}$ is relatively compact if*

1.

$$\sup_{X\in K} \|X\|_{\Gamma,p} < \infty;$$

2.

$$\lim_{J\to\infty} \sup_{X\in K} \sum_{|x|\geq J} \Gamma(x)|X(x)|^p = 0.$$

PROOF. When $\Gamma(x) \equiv 1$ this is the Fréchet–Kolmogorov Compactness Theorem. See Hanche-Olsen and Holden [57] for a historical discussion.

To extend to the weighted spaces considered here, let $K \subseteq \ell_{\Gamma,p}$ satisfy Conditions 1 and 2. Define

$$K_{\Gamma,p} = \{(\Gamma(x)^{1/p} X(x))_{x\in\mathbf{Z}} : X \in K\}.$$

Writing

$$Y(x) = (\Gamma(x))^{1/p} X(x), \tag{1.3.20}$$

we see that $\Gamma(x)|X(x)|^p = |Y(x)|^p$, and so $X \in K$ if and only if $Y \in K_{\Gamma,p}$ and X satisfies Conditions 1 and 2 if and only if Y satisfies the corresponding conditions with $\Gamma \equiv 1$.

These guarantee that $K_{\Gamma,p}$ is relatively compact in $\ell_{1,p}$; that is every sequence in $K_{\Gamma,p}$ has an $\ell_{1,p}$-convergent subsequence. Taking a sequence $(X^n)_{n\in\mathbf{N}}$ in K, the corresponding sequence $(Y^n)_{n\in\mathbf{N}}$ in $K_{\Gamma,p}$ has an $\ell_{1,p}$-convergent subsequence $(Y^{n_k})_{k\in\mathbf{N}}$. Evidently this corresponds through (1.3.20) to an $\ell_{\Gamma,p}$-convergent subsequence $(X^{n_k})_{k\in\mathbf{N}}$ of $(X^n)_{n\in\mathbf{N}}$ as required. □

To prove tightness in $\ell_{\Gamma,p}$, we impose stronger, but biologically natural, conditions on the population at time zero.

PROPOSITION 1.3.15 (Tightness in $\ell_{\Gamma,p}$). *If for some $p \geq 2$ we have*

$$\sup_{n\in\mathbf{N}} \mathbf{E}\left[\|X_0^n\|_\infty^p\right] := \sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{x\in\mathbf{Z}} |X_0^n(x)|^p\right] < \infty, \tag{1.3.21}$$

then the sequence of processes $(X^n)_{n\in\mathbf{N}}$ is C-tight in $D([0,T],\mathbf{R}_+^{\mathbf{Z}})$ where $\mathbf{R}_+^{\mathbf{Z}}$ is equipped with the $\ell_{\Gamma,p}$-topology.

PROOF. Once again we use Theorem 1.3.6 which stipulates that the sequence of processes is tight if $(X_t^n)_{n\in\mathbf{N}}$ is tight for each $t \in [0,T]$ and the continuity estimate (1.3.14) holds. The latter follows from the same argument as (1.3.16). In verifying the first condition, Lemma 1.3.14 and Lemma 1.3.3 reduce the proof to showing that at each fixed time

$$\lim_{N\to\infty} \sup_{n\in\mathbf{N}} \mathbf{P}\left[\sum_{|x|>N} \Gamma(x)\left|X_t^n(x)\right|^p > \varepsilon\right] = 0.$$

We mimic the proof of Proposition 4.6 in Sturm [120].

Define

$$\Gamma^{(r)}(x) = \begin{cases} \Gamma(x) & |x| > r \\ \Gamma(r) & |x| \leq r \end{cases}$$

Then

$$\frac{\Gamma^{(r)}(x)}{\Gamma^{(r)}(y)} \leq \begin{cases} 1 & |x|,|y| \leq r \\ f_\Gamma(|x|-r) & |x| > r, |y| \leq r \\ f_\Gamma(|x-y|) & |x| \geq r, |y| \geq r. \end{cases}$$

Thus, in particular, $\Gamma^{(r)}$ satisfies the conditions that we required of a weight function in Definition 1.2.1.

Following the proof of Lemma 1.3.3 with Γ replaced by $\Gamma^{(r)}$, we see that we can take the constant $C(p,\overline{g},\overline{\gamma},\delta,T,\Gamma^{(r)})$ to be independent of r, since the dependence on $\Gamma^{(r)}$ is through

$$\frac{\Gamma^{(r)}(x+1)}{\Gamma^{(r)}(x)} \leq f_\Gamma(1) \vee 1.$$

Thus, substituting for $\Gamma^{(r)}$, there exists $C(p, \overline{g}, \overline{\gamma}, \delta, T, \Gamma)$ such that

$$\sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{0\le s\le t} \sum_{|x|>r} \Gamma(x)\left|X_s^n(x)\right|^p\right] \le C\left\{(r+1)\Gamma(r) + \sum_{|x|>r}\Gamma(x)\right\} \sup_{n\in\mathbf{N}} \mathbf{E}\left[\|X_0^n\|_\infty^p\right]. \tag{1.3.22}$$

The assumption (1.3.21) and our conditions on Γ ensure that the quantity on the right of (1.3.22) tends to zero as $r \to \infty$, and so Markov's inequality yields

$$\lim_{r\to\infty} \sup_{n\in\mathbf{N}} \mathbf{P}\left[\sup_{0\le s\le t} \sum_{|x|>r} \Gamma(x)\left|X_s^n(x)\right|^p > \varepsilon\right] = 0.$$

Thus, for each $k \in \mathbf{N}$, there exists r_k such that

$$\sup_{n\in\mathbf{N}} \mathbf{P}\left[\sup_{0\le s\le t} \sum_{|x|>r_k} \Gamma(x)\left|X_s^n(x)\right|^p > \frac{1}{k}\right] \le \frac{\varepsilon}{2}\frac{1}{2^k}.$$

Now

$$C_K := \left\{X \in \ell_{\Gamma,p} : \|X\|_{\Gamma,p}^p \le K\right\} \cap \bigcap_{k=1}^{\infty} \left\{X \in \ell_{\Gamma,p} : \sum_{|x|>r_k} \Gamma(x)\left|X(x)\right|^p \le \frac{1}{k}\right\}$$

is relatively compact in $\ell_{\Gamma,p}$ and

$$\begin{aligned}\mathbf{P}\left[X_t^n \notin C_K\right] &\le \mathbf{P}\left[\|X_t^n\|_{\Gamma,p}^p > K\right] + \sum_{k=1}^{\infty} \mathbf{P}\left[\sum_{|x|>r_k} \Gamma(x)\left|X_t(x)\right|^p > \frac{1}{k}\right] \\ &\le \frac{1}{K}\sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{0\le s\le t} \|X_s^n\|_{\Gamma,p}^p\right] + \frac{\varepsilon}{2}.\end{aligned}$$

By Lemma 1.3.3 we can choose K, independent of n, large enough that the first term is less than $\varepsilon/2$ and so $\mathbf{P}[X_t \notin C_K] < \varepsilon$ uniformly in n, and the proof is complete. □

1.3.4. Uniqueness of solutions. In order to deduce convergence of our sequence of rescaled processes, we need to establish uniqueness of the limit point. For this, we invoke a result of Sturm [121].

THEOREM 1.3.16 ([121], Theorem 5.2.1). *Let $p \ge 1$. Consider the infinite system of stochastic differential equations*

$$X_t(x) = X_0(x) + \int_0^t b(X_s, x)\,\mathrm{d}s + \int_0^t \sigma(X_s(x), x)\,\mathrm{d}B_s(x), \qquad x \in \mathbf{Z}, 0 \le t \le T,$$

where $\{B(x)\}_{x\in\mathbf{Z}}$ denotes a collection of independent Brownian motions and b and σ are continuous in their components, with b continuous in X with respect to the topology generated by $\|\cdot\|_{\Gamma,p}$. If b and σ satisfy the growth conditions

$$\|b(X,\cdot)\|_{\Gamma,p} \leq C(1+\|X\|_{\Gamma,p})$$

$$|\sigma(y,x)| \leq C(1+|y|)$$

then for each initial condition $X_0 \in \ell_{\Gamma,2p}$ there exists a solution X to the system of stochastic differential equations with paths in $C([0,T],\ell_{\Gamma,2p})$ such that

$$\mathbf{E}\big[\sup_{0\leq t\leq T}\|X_t\|^{2p}_{\Gamma,2p}\big] < C(T).$$

If, moreover, b satisfies the Lipschitz condition

$$\|b(X,\cdot)-b(Y,\cdot)\|_{\Gamma,1} \leq C\|X-Y\|_{\Gamma,1}$$

and σ the Yamada–Watanabe condition

$$|\sigma(y,x)-\sigma(z,x)| \leq \rho(|y-z|), \tag{1.3.23}$$

where ρ is some strictly increasing function such that $\int_0^\infty \rho^{-2}(s)\,\mathrm{d}s = \infty$, then the solutions to the system are pathwise unique.

Observe that whilst b may depend on the whole system X_s, σ may not, and this is true of the coefficients in the system (1.2.2).

PROOF OF PROPOSITION 1.2.8. Under the assumptions of the proposition, the linear growth and Lipschitz continuity conditions are satisfied. We must just verify the Yamada–Watanabe condition (1.3.23). For this observe that

$$\begin{aligned}
&\big|\sqrt{\delta\mu(X(x))(2-\mu(X(x)))X(x)} - \sqrt{\delta mu(Y(x))(2-\mu(Y(x)))Y(x)}\big| \\
&\leq \sqrt{\big|\delta(\mu(X(x))(2-\mu(X(x)))X(x) - \mu(Y(x))(2-\mu(Y(x)))Y(x))\big|} \\
&\leq C(\delta,\sigma)|X(x)-Y(x)|^{1/2},
\end{aligned}$$

as required. □

1.4. Ancestral lineages.

In this section we wish to establish the behaviour of the ancestral lineage of an individual sampled from a population that is evolving according to our model. In particular, we should

like to understand the effect of the parameters r_n and γ_n on the distribution of the lineage. Our approach will be to condition on the whole history of X^n over the time interval $[0, T]$, and to formulate the behaviour of the lineage ancestral to an individual sampled from the population at time T as a Markov process on $\mathbf{Z}$, described by a random walk in a varying environment. We then estimate the (space-time inhomogeneous) transition probabilities of this process and derive a limit theorem for the lineages. Both in the prelimit and the limit, the ancestral lineage will, with probability one, only make a finite number of jumps over the time interval $[0, T]$. In what follows, we have chosen to take the trivial modification that makes the motion of the lineage càdlàg.

1.4.1. Construction. Consider a realisation of the population size process X^n over the time interval $[0, T]$, corresponding to $\lfloor T\tau_n^{-1} \rfloor$ generations. Select an individual alive at time T (if such an individual exists). The quenched law P^{X^n} on $D([0, T], \mathbf{Z})$ of the position of the ancestors of that individual, as we trace backwards in time, will be that of a nearest neighbour random walk with (space-time inhomogeneous) transition probabilities which depend on X^n. Our first aim is to establish those transition probabilities.

Suppose that $X^n_{\underline{t}+\tau_n}(x) > 0$ for some $0 \le \underline{t} < T$, $x \in \mathbf{Z}$, and suppose that we have sampled an individual from x at time $\underline{t} + \tau_n$. By exchangeability, the position of its parent in generation $\underline{t}$ will be determined by taking a uniform pick from all adults at x at time $\underline{t}$ which survive the generation $\underline{t}$ to $\underline{t} + \tau_n$, and all offspring born to parents alive at x, $x + 1$ or $x - 1$ at time $\underline{t}$ which successfully establish at x at time $\underline{t} + \tau_n$.

Denote by $S^n_{\underline{t}}(x)$ the total number of adults at x at time $\underline{t}$ which survive to time $\underline{t} + \tau_n$. Then $S^n_{\underline{t}}(x)$ is a $\mathrm{Bin}(\eta_n^{-1} X^n_{\underline{t}}(x), 1 - \mu_n(X^n_{\underline{t}}(x)))$ random variable. Similarly, let $B^{n,0}_{\underline{t}}(x)$ (respectively $B^{n,-}_{\underline{t}}(x)$, $B^{n,+}_{\underline{t}}(x)$) be the number of offspring born to parents at x (respectively $x - 1$, $x + 1$) that establish at x at time $(\underline{t} + \tau_n)-$. Finally, let $B^n_{\underline{t}}(x)$ be the total number of offspring establishing at x, ie $B^n_{\underline{t}}(x) = B^{n,0}_{\underline{t}}(x) + B^{n,-}_{\underline{t}}(x) + B^{n,+}_{\underline{t}}(x)$. Note that

$$\mathcal{L}(B^{n,0}_{\underline{t}}(x)) = \mathrm{Poisson}\Big(\eta_n^{-1} X^n_{\underline{t}}(x)(1 - m\tau_n) r_n(X^n_{\underline{t}}(x))\gamma_n(X^n_{\underline{t}}(x))\Big),$$

$$\mathcal{L}(B^{n,\pm}_{\underline{t}}(x)) = \mathrm{Poisson}\Big(\eta_n^{-1} X^n_{\underline{t}}(x \pm 1)\frac{m\tau_n}{2} r_n(X^n_{\underline{t}}(x))\gamma_n(X^n_{\underline{t}}(x \pm 1))\Big),$$

$$\mathcal{L}(B^n_{\underline{t}}(x)) = \mathrm{Poisson}\Big(\eta_n^{-1}\mu_n(X^n_{\underline{t}}(x)) X^n_{\underline{t}}(x) + \tau_n \eta_n^{-1} b_n(X^n_{\underline{t}}, x)\Big),$$

where b_n was defined in (1.3.1), and that conditional on $\mathcal{F}^{X^n}_{\underline{t}} = \sigma(\{X^n_{\underline{s}} : \underline{s} \le \underline{t}\})$, the family of random variables $\{S^n_{\underline{t}}(x), B^{n,0}_{\underline{t}}(x), B^{n,-}_{\underline{t}}(x), B^{n,+}_{\underline{t}}(x) : x \in \mathbf{Z}\}$ are independent.

Denote by $s^n_{\underline{t}+\tau_n}(x)$ the probability that an individual sampled at x at time $\underline{t} + \tau_n$ is a surviving adult from the previous generation; that is

$$s^n_{\underline{t}+\tau_n}(x) = \mathbf{E}\left[\frac{S^n_{\underline{t}}(x)}{S^n_{\underline{t}}(x) + B^n_{\underline{t}}(x)}\middle| X^n\right].$$

Since $B^n_{\underline{t}}(x)$ is dominated by $B^{n,0}_{\underline{t}}(x)$ (the mean of $B^{\pm}_{\underline{t}}$ is order τ_n/η_n, whereas that of $B^{n,0}_{\underline{t}}(x)$ is order $1/\eta_n$), and since both $S^n_{\underline{t}}(x)$ and $B^{n,0}_{\underline{t}}(x)$ can be viewed as the sum of order $1/\eta_n$ independent random variables, one can guess that the law of large numbers kicks in, and the quantity $s^n_{\underline{t}+\tau_n}(x)$ is determined by the ratio of the means, that is $1-\mu_n(X^n_t(x))$ (plus a small correction). In the next section we shall use a local limit theorem to verify that this intuition is correct, at least if $\tau_n^2/\eta_n \to 0$ as $n \to \infty$. Since we are primarily interested in the case in which the limiting population model is stochastic, which requires $\eta_n/\tau_n = \delta_n \to \delta > 0$ and so the condition is automatically satisfied, we won't allow ourselves to be detained by the case in which $\tau_n^2/\eta_n \not\to 0$.

Conditional on the event that the chosen individual at x at time $\underline{t}+\tau_n$ is not one of the surviving adults, it is an offspring of the reproduction in the previous generation and, since such offspring are distributed as the sum of three independent Poisson random variables, the distribution of the location of its parent is determined by their means; ie the probability that the parent was at site x is

$$\frac{(1-m\tau_n)\Phi_n(X^n_{\underline{t}})(x)}{\Phi_n(X^n_{\underline{t}})(x)+\frac{m}{2}\tau_n\Delta_1\Phi_n(X^n_{\underline{t}})(x)},$$

and the probability that the parent was at site $x\pm 1$ is

$$\frac{\frac{m}{2}\tau_n\Phi(X^n_{\underline{t}})(x\pm 1)}{\Phi_n(X^n_{\underline{t}})(x)+\frac{m}{2}\tau_n\Delta_1\Phi_n(X^n_{\underline{t}})(x)},$$

where, as before, $\Phi_n(u)=u\gamma(u)$ denotes the total fecundity when the population size is u. From this description, it is evident that the position of the ancestral lineage is determined by a space-time inhomogeneous Markov chain, jumping at times $\lfloor T/\tau_n\rfloor\tau_n - k\tau_n$, $k = 0,1,\ldots,\lfloor T/\tau_n\rfloor - 1$.

Let us now describe the lineage process formally. We will write

$$\vec{t} := \lceil T/\tau_n\rceil\tau_n - t,$$

and use $\{L^n_t : 0\le t\le T\}$ to denote the process which records the position of the lineage at time t before T. If there are no individuals at the origin at time T, ie if $X^n_T(0)=0$, we will set $L^n = \varnothing$, where $\varnothing$ is an arbitrary symbol denoting the fact that a lineage started at the origin is undefined. If $X^n_T(0)>0$ we will set $L^n_0 = 0$ and let L^n_t denote the position of the ancestral lineage at time $\tau_n\lceil T/\tau_n\rceil - t$. In order that L^n jumps at the times of multiples of whole generations of τ_n, we will formally trace ancestors of individuals alive at time $\tau_n\lceil T/\tau_n\rceil -$. Since $X^n_{\lceil T/\tau_n\rceil -} = X^n_T$ this is done without loss of generality. Denote by $\mathbf{P}[\cdot|X^n]$ the law on $D([0,T],\mathbf{Z})$ of the process L^n conditional on X^n, then we may define

the transition probability at each jump time $\underline{t}$ by

$$p^n_{\underline{t}}(X^n, x, y) := \mathbf{P}[L^n_{\underline{t}+\tau_n} = y | L^n_{\underline{t}} = x, X^n], \qquad x, y \in \mathbf{Z}.$$

In words $p^n_{\underline{t}}(X^n, x, y)$ is the probability, conditional on X^n, that an individual alive and at x at time $\vec{t} = \tau_n\lceil T/\tau_n\rceil - \underline{t}$ is descended from an individual at y at time $\overrightarrow{\underline{t} + \tau_n} = \tau_n\lceil T/\tau_n\rceil - \underline{t} - \tau_n$. Given the above description of these probabilities, we see that

$$p^n_{\underline{t}}(X^n, x, x \pm 1) = \frac{m\tau_n}{2} \frac{(1 - s^n_{\vec{\underline{t}}}(x))\Phi_n(X^n_{\underline{t}})(x \pm 1)}{\Phi_n(X^n_{\vec{\underline{t}}})(x) + \frac{1}{2}m\tau_n\Delta_1\Phi_n(X^n_{\vec{\underline{t}}})(x)},$$

$$p^n_{\underline{t}}(X^n, x, x) = s^n_{\vec{\underline{t}}}(x) + (1 - s^n_{\vec{\underline{t}}}(x)) \frac{(1 - m\tau_n)\Phi_n(X^n_{\underline{t}})(x)}{\Phi_n(X^n_{\vec{\underline{t}}})(x) + \frac{1}{2}m\tau_n\Delta_1\Phi_n(X^n_{\vec{\underline{t}}})(x)},$$

$$p^n_{\underline{t}}(X^n, x, y) = 0, y \notin \{x, x - 1, x + 1\}.$$

1.4.2. Approximation of the probabilities s^n. Observe that

$$s^n_{\underline{t}+\tau_n}(x) = \mathbf{E}\left[\frac{S^n_{\underline{t}}(x)}{S^n_{\underline{t}}(x) + B^n_{\underline{t}}(x)} \middle| X^n \right].$$

LEMMA 1.4.1. *Under the assumptions of Theorem 1.2.10, suppose further that*

$$|X_{\underline{t}+\tau_n} - X_{\underline{t}}| \le J\sqrt{\eta_n}.$$

Then as $n \to \infty$,

$$s^n_{\underline{t}+\tau_n}(x) = 1 - \mu_n(X^n_{\underline{t}}(x)) + o(1)\frac{(1 + J\sqrt{\eta_n})}{\exp(-J^2/\sigma_n^2 c)}, \tag{1.4.1}$$

where the $o(1)$ in the error term is bounded independently of X^n and J.

PROOF. We (arbitrarily) label the individuals in $X^n_{\underline{t}}(x)$ by $1, 2, \ldots X^n_{\underline{t}}(x)/\eta_n$. We can then write

$$S^n_{\underline{t}}(x) = \sum_{i=1}^{X^n_{\underline{t}}(x)/\eta_n} \zeta_i,$$

where the ζ_i are i.i.d. Bernoulli random variables with parameter $1 - \mu(X^n_{\underline{t}}(x))$. We wish to calculate

$$\mathbf{E}\Big[S^n_{\underline{t}}(x) \Big| X^n = \chi^n \Big] = \sum_{i=1}^{\chi^n_{\underline{t}}(x)/\eta_n} \mathbf{E}\Big[\zeta_i | X^n = \chi^n \Big],$$

for which (since individuals in our population are indistinguishable and ζ_1 is a Bernoulli random variable) it suffices to calculate $\mathbf{P}[\zeta_1 = 1 | X^n = \chi^n]$.

Let us write $(B^{n,0}_{\underline{t}}(x))_i$ for the number of offspring of the individual labelled i that contribute to $B^{n,0}_{\underline{t}}(x)$ and write

$$B = B^{n,+}_{\underline{t}}(x) + B^{n,-}_{\underline{t}}(x) + (B^{n,0}_{\underline{t}}(x))_1.$$

Note that B is a Poisson random variable and, since Φ is uniformly bounded, its mean is bounded by $C(1 + \tau_n/\eta_n)$, with a constant which is independent of X^n.

In this notation we can write

$$\frac{X^n_{\underline{t}+\tau_n}(x)}{\eta_n} = B + \sum_{i=2}^{X^n_{\underline{t}}(x)/\eta_n} \xi_i + \zeta_1,$$

where the random variables $\xi_i := \zeta_i + (B^{n,0}_{\underline{t}}(x))_i$ are i.i.d. with mean and variance bounded independently of X^n. Conditioning as above on $X^n = \chi^n$, we shall write

$$\Theta^n = \sum_{i=2}^{\chi^n_{\underline{t}}(x)/\eta_n} \xi_i.$$

It is to this sum that we shall be applying the local limit theorem.

To calculate $\mathbf{P}[\zeta_1 = 1 | X^n = \chi^n]$ we shall partition over possible values of B and use Bayes' rule. We find that

$$\begin{aligned}
\mathbf{P}[\zeta_1 = 1 | X^n = \chi^n] &= \sum_k \mathbf{P}[\zeta_1 = 1, B = k | X^n = \chi^n] \\
&= \sum_k \frac{\mathbf{P}[\zeta_1 = 1, B = k]\mathbf{P}[X^n = \chi^n | \zeta_1 = 1, B = k]}{\mathbf{P}[X^n = \chi^n]} \\
&= \frac{\mathbf{P}[\zeta_1 = 1] \sum_k \mathbf{P}[B = k]\mathbf{P}[\Theta^n = \chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k - 1]}{\sum_k \mathbf{P}[B = k]\Big\{\mathbf{P}[\Theta^n = \chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k - 1]\mathbf{P}[\zeta_1 = 1] \\
&\qquad + \mathbf{P}[\Theta^n = \chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k]\mathbf{P}[\zeta_1 = 0]\Big\}} \\
&= \mathbf{P}[\zeta_1 = 1] \frac{1}{1 + \mathcal{E}^n \mathbf{P}[\zeta_1 = 0]},
\end{aligned}$$

where

$$\mathcal{E}^n = \frac{\sum_k \mathbf{P}[B = k]\Big\{\mathbf{P}[\Theta^n = \chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k - 1] - \mathbf{P}[\Theta^n = \chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k]\Big\}}{\sum_k \mathbf{P}[B = k]\mathbf{P}[\Theta^n = \chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k - 1]}.$$

We must show that $\mathcal{E}^n \to 0$ as $n \to \infty$. We now apply a local limit theorem to Θ^n, which is a sum of at least c/η_n i.i.d. terms.

Setting $e_n = 1 + \tau_n(g_n - \mu_n)$ and $\sigma_n^2 = \mu_n(1 - \mu_n) + (1 - \tau_n)r_n\gamma_n$, where each of the

functions is evaluated at $\chi^n_{\underline{t}}(x)$, Theorem 1.6.2 gives

$$\sup_{m\in\mathbf{Z}} \frac{1}{\sqrt{\eta_n}} \left| \mathbf{P}[\Theta^n = m] - \sqrt{\frac{\eta_n}{2\pi\sigma_n^2 \chi^n_{\underline{t}}(x)}} \exp\left(- \frac{\big(m - e_n \chi^n_{\underline{t}}(x)/\eta_n\big)^2}{2\sigma_n^2 \chi^n_{\underline{t}}(x)/\eta_n} \right) \right| \to 0,$$

as $n \to \infty$.

To ease notation, let us write $P(m)$ for the Gaussian density in this expression. Then substituting in our expression for $\mathcal{E}^n$ and performing a Taylor expansion,

$$\mathcal{E}^n = \frac{\sum_k \mathbf{P}[B = k] P(\chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k) \Big\{ - \frac{\chi^n_{\underline{t}+\tau_n}(x) - \chi^n_{\underline{t}}(x) - \eta_n k - O(\tau_n)\chi^n_{\underline{t}}(x)}{\sigma_n^2 \chi^n_{\underline{t}}(x)} \Big\} + o(\sqrt{\eta_n})}{\sum_k \mathbf{P}[B = k] P(\chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k)}.$$

$$= \frac{-\frac{\chi^n_{\underline{t}+\tau_n}(x) - \chi^n_{\underline{t}}(x)}{\sigma_n^2 \chi^n_{\underline{t}}(x)} \Sigma + O\big(\tau_n\big) + o(\sqrt{\eta_n})}{\Sigma + o(\sqrt{\eta_n})},$$

where

$$\Sigma = \sum_k \mathbf{P}[B = k] P(\chi^n_{\underline{t}+\tau_n}(x)/\eta_n - k).$$

Our result will follow if Σ is at least a constant times $\sqrt{\eta_n}$. The idea is that this will be true if

$$\frac{\big(\chi^n_{\underline{t}+\tau_n}(x)/\eta_n - \chi^n_{\underline{t}}(x)/\eta_n - k\big)^2}{\sigma_n^2 \chi^n_{\underline{t}}(x)/\eta_n}$$

is bounded by a constant for $k \in K$ where $\mathbf{P}[B \in K]$ is order 1. Let us denote by $K := \{0, \dots \lfloor 1/\sqrt{\eta_n} \rfloor\}$. Recalling that for a Poisson random variable D with parameter λ and $x > \lambda$,

$$\mathbf{P}[D > x] \le \frac{(e\lambda)^x e^{-\lambda}}{x^x} \le \exp\Big(- x\big(\log(x/\lambda) - 1\big)\Big),$$

we see that since B has parameter bounded by $C(1 + \tau_n/\eta_n)$,

$$\mathbf{P}[B \in K] \ge 1 - \exp\Big(- \big(C \log(\eta_n/\tau_n^2) - 1\big)/\sqrt{\eta_n}\Big),$$

which tends to 1 as $n \to \infty$.

Under the assumption that $|\chi^n_{\underline{t}+\tau_n} - \chi^n_{\underline{t}}| \le J\sqrt{\eta_n}$, we have

$$\Sigma \ge C \exp\Big(- \frac{J^2}{c\sigma_n^2}\Big)\sqrt{\eta_n},$$

with C independent of χ^n (and c the lower bound on population density of our assumptions).

Our result now follows since

$$\mathcal{E}^n \leq \frac{J\sqrt{\eta_n}}{\sigma_n^2 c} + \frac{C\tau_n/\sqrt{\eta_n} + o(1)}{\exp\left(-\frac{J^2}{c\sigma_n^2}\right)}$$

(where the $o(1)$ term came from the local limit theorem). □

1.4.3. Skorokhod representation of (X^n, L^n). Before going further, let us explain the idea of the rest of the proof. We are going to use the Skorokhod representation theorem to work on a probability space on which the sequence X^n converges almost surely. We allow the lineage process to evolve only until it reaches a deme in which the population size is less than or equal to $c > 0$, after which it is frozen (no longer jumps). As a result, over any given generation, the probability that the lineage jumps to one of the neighbouring demes is uniformly bounded by $C\tau_n$ for some constant C which is independent of X^n. In each deme, in each generation, we use a Bernoulli random variable with parameter $C\tau_n$ to determine *potential* jump times of an ancestral lineage. We then construct a directed branching structure, denoted by $\mathcal{T}^n$, which branches into two at each 'potential' jump time of the lineage, in order to trace all possible paths that it takes. Over the $\lfloor T/\tau_n \rfloor$ generations that we follow, the number of jumps that the lineage makes is almost surely bounded and we can extend our probability space so that this sequence of branching structures also converges almost surely to a limit $\mathcal{T}$ which has at most a finite number of branch points and consequently straddles a bounded set of demes. We shall see that we can take the same sequence $\mathcal{T}^n$ for every X^n.

We construct a perturbation of our ancestral lineage, which we shall denote by $\widetilde{L}^n$, by tracing a path through $\mathcal{T}^n$. The transition probabilities of $\widetilde{L}^n$ are those of L^n, except that we replace $s^n_{\underline{t}}$ by $1 - \mu_n(X^n_{\underline{t}})$. We first check the almost sure convergence of the perturbed lineage process, conditional on the environment (that is the quenched law). To complete the proof of Theorem 1.2.10, we choose a sequence J_n such that the error in Lemma 1.4.1 tends to zero as $n \to \infty$ and use (1.3.6) and Markov's inequality to verify that we can couple L^n and $\widetilde{L}^n$ so that with probability tending to one as $n \to \infty$ they coincide over the time interval $[0, T]$.

We begin with the Skorokhod representation. The pair (X^n, L^n) exist on some (as yet) intangible probability space and their laws are defined with respect to the probability measure $\mathbf{P}$. Their joint law is specified by the marginal of X^n and the conditional law $\mathbf{P}[L^n \in \cdot | X^n]$ whose description is given above. It will be useful to construct random variables $\widehat{X}^n$ and $\widehat{L}^n$ on some explicit probability space with the same joint law as (X^n, L^n) respectively, but such that $\widehat{X}^n \to \widehat{X}$ almost surely, and $\widehat{X}$ has the same distribution as X. Using this representation of (X^n, L^n) will be a key technique in our proofs below.

By the Skorokhod representation theorem, since $X^n \Rightarrow X$ in $D([0, T], \mathbf{R}_+^{\mathbf{Z}})$ which is

complete and separable, there exists a probability space $(\Omega, \mathcal{F}, \widehat{\mu})$ and random variables $\widehat{X}^n$ and $\widehat{X}$ on this space such that $\mathcal{L}(\widehat{X}^n) = \mathcal{L}(X^n)$ and $\mathcal{L}(\widehat{X}) = \mathcal{L}(X)$ with $\widehat{X}^n \to \widehat{X}$ $\widehat{\mu}$-almost surely. In addition, let $([0,1]^\infty, \mathcal{B}([0,1])^\infty, \nu)$ be a probability space with ν the law of a sequence of independent uniform random variables, $(\xi_k)_{k=0}^\infty$, on the interval $[0,1]$. Each pair $(x, \underline{t}) \in \mathbf{Z} \times \{\tau_n, 2\tau_n, \ldots, \lfloor T/\tau_n \rfloor \tau_n\}$ corresponds to a distinct ξ_k. The random variable $(\widehat{X}^n, \widehat{L}^n)$ will be defined on the product space

$$\mathcal{P} = (\Omega \times [0,1]^\infty, \sigma(\mathcal{F} \times \mathcal{B}([0,1]^\infty)), \widehat{\mathbf{P}} := \mu \otimes \nu),$$

so that $(\xi_k)_{k=0}^\infty$ is independent of $\mathcal{F}$. For $(\omega, \xi) \in \Omega \times [0,1]^\infty$ we set $\widehat{X}^n(\omega, \xi) \equiv \widehat{X}^n(\omega)$, and if $\widehat{X}^n_T(\omega, 0) = 0$ then $\widehat{L}^n_t(\omega, \xi) = \varnothing$ for all $t \in [0, T]$; otherwise, if $k(\widehat{L})$ is the index of the uniform random variable corresponding to the current location of $\widehat{L}^n$ after k generations,

$$\begin{aligned}
&\widehat{L}^n_0(\omega, \xi) = 0,\\
&\widehat{L}^n_{(k+1)\tau_n}(\omega, \xi) - \widehat{L}^n_{k\tau_n}(\omega, \xi)\\
&\quad = \begin{cases} -1, & \xi_{k(\widehat{L})} \le p^n_{\underline{t}}\big(\hat{X}^n(\omega), \widehat{L}^n_{k\tau_n}(\omega, \xi), \widehat{L}^n_{k\tau_n}(\omega, \xi) - 1\big),\\ +1, & \xi_{k(\widehat{L})} \ge 1 - p^n_{\underline{t}}\big(\widehat{X}^n(\omega), \widehat{L}^n_{k\tau_n}(\omega, \xi), \widehat{L}^n_{k\tau_n}(\omega, \xi) + 1\big),\\ 0, & \text{otherwise.}\end{cases}
\end{aligned}$$

Notice that we can take the same sequence of uniform random variables $(\xi_k)_{k=0}^\infty$ for every realisation of the population process $\widehat{X}^n$.

For a fixed $\omega \in \Omega$ we will write P^ω for the pushforward measure

$$P^\omega(\widehat{L}^n(\omega, \cdot) \in \cdot) = \nu(\{\xi \in [0,1]^\infty : L^n(\omega, \xi) \in \cdot\})$$

on $D([0,T], \mathbf{Z})$, the law of $\widehat{L}^n$ for a fixed $\omega \in \Omega$. It is this measure which makes our construction useful: instead of conditioning on a sequence of X^n, necessary under $\mathbf{P}$, our measure space allows us instead to average out over sequences $\{\hat{X}^n\}$ which converge almost surely (under $\widehat{\mu}$). Crucially, any statements about the law of L^n under $\mathbf{P}$ hold equivalently for $\widehat{L}$ under $\hat{\mathbf{P}}$ and vice versa.

1.4.4. Tightness of $\hat{L}^n(\omega, \cdot)$. Fix $c > 0$ and for each $n \in \mathbf{N}$, $\omega \in \Omega$ define

$$T_c^{\omega,n} := \inf\{0 \le t \le T : \widehat{X}^n_{\underline{t}}(\omega, \widehat{L}^n_{\underline{t}}) < c\}$$

to be the first time that the lineage $\widehat{L}^n$ arrives at a deme with a population density smaller than c. We adopt the convention that $\inf \emptyset = \infty$.

PROPOSITION 1.4.2. *For $\widehat{\mu}$-almost every $\omega \in \Omega$, the sequence of stopped processes $\widehat{L}^n_{\cdot\wedge T_c^{\omega,n}}(\omega,\cdot)$ is tight in $D([0,T],\mathbf{Z})$ with respect to the measure P^ω.*

To ease notation, we will henceforth assume that the process is stopped when it hits a deme with population density small than c and omit the subscript $T_c^{\omega,n}$; that is we write $\widehat{L}^n_{\cdot\wedge T_c^{\omega,n}}(\omega,\cdot) \equiv \widehat{L}^n_\cdot(\omega,\cdot)$.

PROOF. By Aldous' criteria (see [2]) it is sufficient to show that for $\widehat{\mu}$-almost every ω, the sequence $\{\widehat{L}^n_t(\omega,\cdot)\}$ is tight in $\mathbf{R}_+$ (with respect to P^ω) and that for any sequence of bounded stopping times (ρ_n), for each $\varepsilon > 0$, there exists a $\delta > 0$ and an N_0 such that

$$\sup_{n\geq N_0}\sup_{0\leq\theta\leq\delta} P^\omega[|\widehat{L}^n_{\rho_n+\theta}(\omega,\cdot) - \widehat{L}^n_{\rho_n}(\omega,\cdot)| > \varepsilon] \leq \varepsilon.$$

Fix $\omega \in \Omega$ such that $\widehat{X}^n \to \widehat{X}$. Recall that

$$p^n_{\underline{t}}(\widehat{X}^n(\omega),x,x\pm1) = \frac{\tau_n}{2}\frac{m(1-s^n_{\underline{t}}(x))\Phi_n(\widehat{X}^n_{\underline{t}}(\omega,x\pm1))}{\Phi_n(\widehat{X}^n_{\underline{t}}(\omega,x)) + \frac{1}{2}m\tau_n\Delta_1(\Phi_n\circ\widehat{X}^n_{\underline{t}}(\omega,\cdot))(x)}.$$

Using our assumption that there exists $\overline{\gamma}$ such that $\overline{\gamma}^{-1} \leq \gamma_n(x) \leq \overline{\gamma}$ for all n and x, for all $n \geq n_0(\tau)$ and $\underline{t} \leq T_c^{\omega,n}$,

$$p^n_{\underline{t}}(\widehat{X}^n(\omega),x,x\pm1) \leq m\frac{\tau_n}{2}\frac{\Phi_n(\widehat{X}^n_{\underline{t}}(\omega,x\pm1))}{(1-m\tau_n)\Phi_n(\widehat{X}^n_{\underline{t}}(\omega,x))} \leq m\overline{\gamma}c^{-1}\overline{\Phi}\,\tau_n, \tag{1.4.2}$$

where $\overline{\Phi}$ is the bound on the total fecundity. (The inequality on the right uses that $1/2(1-m\tau_n) \leq 1$ and $\Phi(u) = u\gamma(u)$ where we are only considering $u > c$.)

We thus deduce that there exists a constant $C \equiv C(\overline{\gamma},\overline{\Phi},c) < \infty$ such that we may uniformly bound the probability of jumping out of a site:

$$\sup_{n\in\mathbf{N}}\sup_{\underline{t}\leq T_c^{\omega,n}}\frac{1}{\tau_n}|p^n_{\underline{t}}(\widehat{X}^n(\omega),x,x+1) + p^n_{\underline{t}}(\widehat{X}^n(\omega),x,x-1)| \leq C.$$

Observe that this constant is independent of $\omega \in \Omega$. Before time t, the process $\widehat{L}^n(\omega,\cdot)$ may jump at at most $\underline{t}/\tau_n$ fixed times; at each of these it jumps with probability at most $C\tau_n$. Thus by Markov's inequality,

$$P^\omega[|\widehat{L}^n_t(\omega,\cdot)| > k] \leq \frac{\underline{t}\tau_n^{-1}C\tau_n}{k} = \frac{C}{k},$$

and thus $\{\widehat{L}^n_t(\omega,\cdot)\}$ is tight in $\mathbf{R}_+$ (with respect to P^ω).

Now let us prove the continuity estimate. The key is that the probability of jumping is on the same order as the number times at which the process may jump and thus there is no accumulation of jumps as we pass to the limit. For $\varepsilon \in (0,1)$ observe that the event that $|\widehat{L}^n_{\rho_n+\theta}(\omega,\cdot) - \widehat{L}^n_{\rho_n}(\omega,\cdot)| > \varepsilon$ is equivalent to there being at least one jump on $L^n(\omega,\cdot)$ in $(\rho_n, \rho_n+\theta)$. There are at most θ/τ_n possible jumps; again each jump happens with probability bounded by $C\tau_n$ and thus

$$P^\omega[|\widehat{L}^n_{\rho_n+\theta}(\omega,\cdot) - \widehat{L}^n_{\rho_n}(\omega,\cdot)| > \varepsilon] \le 1 - (1 - C\tau_n)^{\theta/\tau_n}.$$

So for any $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{0\le\theta\le\delta} P^\omega[|\widehat{L}^n_{\rho_n+\theta}(\omega,\cdot) - \widehat{L}^n_{\rho_n}(\omega,\cdot)| > \varepsilon] \le 1 - e^{-2C\delta}.$$

Taking δ sufficiently small, the result follows. □

1.4.5. Proof of Theorem 1.2.10. Define the family of operators $A^\omega = (A^\omega_t)_{0\le t\le T}$ on $B(\mathbf{Z})$ by

$$A^\omega_t f(x) := \frac{1}{2} m \frac{r(\widehat{X}_{T-t}(\omega,x))}{\widehat{X}_{T-t}(\omega,x)} \sum_{y=x\pm1} \Phi(\widehat{X}_{T-t}(\omega,y))\big(f(y) - f(x)\big).$$

Let us first argue that there is a unique solution to the martingale problem for A^ω when it is stopped before arriving at a deme with population size smaller than $c > 0$.

LEMMA 1.4.3. *For each $\omega \in \Omega$, there exists a unique (in law) process $\widehat{L}(\omega)$ such that for $T^{\omega,c} := \inf\{t \ge 0 : \widehat{X}_{T-t}(\omega, \widehat{L}_t(\omega)) < c\}$, $\widehat{L}_t(\omega) = \widehat{L}_{t\wedge T^{\omega,c}}(\omega)$ and*

$$f(\widehat{L}_t(\omega)) - \int_0^{t\wedge T^{\omega,c}} A^\omega_s f(\widehat{L}_s(\omega))\,\mathrm{d}s$$

is a martingale for each $f \in B(\mathbf{Z})$.

PROOF. This follows directly from Theorem 4.7.3 of Ethier and Kurtz [40], noting that the lineage process jumps out of site x at time t at total instantaneous rate

$$\frac{1}{2} m \frac{r(\widehat{X}_{T-t}(\omega,x))}{\widehat{X}_{T-t}(\omega,x)} \sum_{y=x\pm1} \Phi(\widehat{X}_{T-t}(\omega,y)) \le mc^{-1}\overline{\Phi},$$

which is finite. □

We should like to have almost sure convergence of $\widehat{L}^n$ to $\widehat{L}$. The difficulty is that our estimate of $s^n_{\underline{t}}$ requires that $X_{\underline{t}+\tau_n} - X_{\underline{t}}$ is at most $J\sqrt{\eta_n}$, and this will not be true for all

realisations of the process $\widehat{X}^n$. The resolution is to define a perturbation $\widetilde{L}^n$ of $\widehat{L}^n$, on the same probability space as $\widehat{L}^n$, in such a way that the two processes agree for a set of $\widehat{X}^n$ whose probability tends to one as n tends to infinity. We assume that n is large enough that $C\tau_n < 1/3$.

Using the same sequence $(\xi_k)_{k=0}^{\infty}$ of uniform random variables that we used to define $\widehat{L}^n$, define $\mathcal{T}^n$ as follows. At time zero we start a lineage from the origin. In the first generation, if $\xi_0 \le C\tau_n$ (resp. if $\xi_0 \ge 1 - C\tau_n$), where $C = m\overline{\gamma}c^{-1}\overline{\Phi}$ (c.f. (1.4.2)), then we start a new path from the point $x-1$ (resp. $x+1$) at time τ_n. In each generation, for each path alive, we consider the corresponding uniform random variable (an independent variable for each distinct path) and if $\xi_k \le C\tau_n$ (resp. $\xi_k \ge 1 - C\tau_n$) we start a new path from the nearest neighbour to the left (resp. right) of that path, with the understanding that if there is already a path at that location then the creation is suppressed. We denote the resulting branching set of paths by $\mathcal{T}^n$. Note that, by construction, $\widehat{L}^n$ is contained in $\mathcal{T}^n$. Moreover, since $\mathrm{Bin}(\underline{T}/\tau_n, C\tau_n)$ converges to a Poisson random variable with mean CT, $\mathcal{T}^n$ is almost surely finite and the expected position of the leftmost (resp. rightmost) path in $\mathcal{T}^n$ at time T can be bounded uniformly in n.

Now, again using the same uniform random variables $(\xi_k)_{k=0}^{\infty}$, we define the process $\widetilde{L}^n$ in a way which parallels the definition of $\widehat{L}_n$: if $\widehat{X}^n_T(\omega,0) = 0$ then $\widetilde{L}^n_t(\omega,\xi) = \varnothing$ for all $t \in [0,T]$; otherwise, if $k(\widetilde{L})$ is the index of the uniform random variable corresponding to the current location of $\widetilde{L}^n$ after k generations,

$$
\begin{aligned}
&\widetilde{L}^n_0(\omega,\xi) = 0,\\
&\widetilde{L}^n_{(k+1)\tau_n}(\omega,\xi) - \widetilde{L}^n_{k\tau_n}(\omega,\xi)\\
&\quad = \begin{cases} -1, & \xi_{k(\widetilde{L})} \le \widetilde{p}^n_{\underline{t}}\big(\widehat{X}^n(\omega), \widetilde{L}^n_{k\tau_n}(\omega,\xi), \widetilde{L}^n_{k\tau_n}(\omega,\xi) - 1\big),\\ +1, & \xi_{k(\widetilde{L})} \ge 1 - \widetilde{p}^n_{\underline{t}}\big(\widehat{X}^n(\omega), \widetilde{L}^n_{k\tau_n}(\omega,\xi), \widetilde{L}^n_{k\tau_n}(\omega,\xi) + 1\big),\\ 0, & \text{otherwise,}\end{cases}
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{p}^n_{\underline{t}}(\widehat{X}^n, x, x\pm 1) &= \frac{m\tau_n}{2}\frac{\mu^n(\widehat{X}^n_{\underline{t}})\Phi_n(\widehat{X}^n_{\underline{t}})(x\pm 1)}{\Phi_n(\widehat{X}^n_{\underline{t}})(x) + \frac{1}{2}m\tau_n\Delta_1\Phi_n(\widehat{X}^n_{\underline{t}})(x)},\\
\widetilde{p}^n_{\underline{t}}(\widehat{X}^n, x, x) &= 1 - \mu^n(\widehat{X}^n_{\underline{t}}) + \mu^n(\widehat{X}^n_{\underline{t}})\frac{(1-m\tau_n)\Phi_n(\widehat{X}^n_{\underline{t}})(x)}{\Phi_n(\widehat{X}^n_{\underline{t}})(x) + \frac{1}{2}m\tau_n\Delta_1\Phi_n(\widehat{X}^n_{\underline{t}})(x)},\\
\widetilde{p}^n_{\underline{t}}(\widehat{X}^n, x, y) &= 0,\ y \notin \{x, x-1, x+1\}.
\end{aligned}
$$

LEMMA 1.4.4. *Let $c > 0$ as above and let $\widehat{L}(\omega)$ be the unique solution to the stopped martingale problem of Lemma 1.4.3. For $\widehat{\mu}$-almost every $\omega \in \Omega$, $\widetilde{L}^n_{\cdot\wedge T_c^{\omega,n}}(\omega,\cdot) \Rightarrow \widehat{L}(\omega)$*

in $D([0,T],\mathbf{Z})$ *as* $n \to \infty$.

PROOF. Let $\omega \in \Omega$ be such that $\widehat{X}^n(\omega) \to \widehat{X}(\omega)$ in $D([0,T],\mathbf{R}_+^{\mathbf{Z}})$. By construction $\widetilde{L}^n \subseteq \mathcal{T}^n$ and, consequently has almost surely finite range, which we shall denote by $\overline{N}$.

The proof of Proposition 1.4.2 guarantees that the sequence of processes $\widetilde{L}^n_{\cdot\wedge T_c^{\omega,n}}(\omega,\cdot)$ is tight in $D([0,T],\mathbf{Z})$.

For $f \in B(\mathbf{Z})$ define the family of operators $(A^{n,\omega}_{\underline{t}})_{\underline{t}\le T}$ by

$$A^{\omega,n}_{\underline{t}} f(x) = \frac{1}{\tau_n}\widetilde{p}^n_{\underline{t}}(\widehat{X}^n(\omega),x,x+1)[f(x+1)-f(x)] + \frac{1}{\tau_n}\widetilde{p}^n_{\underline{t}}(\widehat{X}^n(\omega),x,x-1)[f(x-1)-f(x)].$$

Then

$$f(\widetilde{L}^n_{\underline{t}}(\omega)) - \int_0^{\underline{t}\wedge T_c^{\omega,n}} A^{\omega,n}_{\underline{s}} f(\widetilde{L}^n_{\underline{s}}(\omega))\,\mathrm{d}s \tag{1.4.3}$$

is a P^ω-martingale, noting that the integral on the right hand side is really a sum because the integrand is a step function, constant between intergeneration times. Now $\widehat{X}^n_{\underline{t}}(x) \to \widehat{X}_{T-t}(x)$ uniformly over $0 \le t \le T$ and $|x| \le \overline{N}$; $\mu_n \to \mu$ and $\gamma_n \to \gamma$ uniformly and therefore for $\widehat{\mu}$-almost every $\omega \in \Omega$, recalling that $\mu(u) = r(u)\gamma(u)$ and $\Phi(u) = u\gamma(u)$,

$$\begin{aligned}\frac{1}{\tau_n}\widetilde{p}^n_{\underline{t}}(\widehat{X}^n(\omega),x,x\pm1) &\to \frac{1}{2}m\frac{\mu(\widehat{X}_{T-t}(x))\Phi(\widehat{X}_{T-t})(x\pm1)}{\Phi(\widehat{X}_{T-t})(x)}\\ &= \frac{1}{2}m\frac{r(\widehat{X}_{T-t}(x))\Phi(\widehat{X}_{T-t})(x\pm1)}{\widehat{X}_{T-t}(x)}\end{aligned}$$

as $n \to \infty$, uniformly over $0 \le t \le T$ and $|x| \le \overline{N}$.

The argument is now completely standard. We pass to the limit and characterise the weak subsequential limits as solutions to a martingale problem. By the Skorokhod representation theorem, there exist random variables with the same distribution as $\widetilde{L}^n_{\cdot\wedge T_c^{\omega,n}}(\omega,\cdot)$ and $\widehat{L}(\omega)$ such that the convergence is almost sure.

Taking expectations of (1.4.3) and passing to the limit, we see that the limiting process is a solution to the stopped martingale problem defined in Lemma 1.4.3. Since this martingale problem is well-posed, the proof is complete. □

PROOF OF THEOREM 1.2.10. It is now straightforward to complete the proof of Theorem 1.2.10. We have shown that for $\widehat{\mu}$-almost every $\omega \in \Omega$, $\widetilde{L}_{\cdot\wedge T_c^{n,\omega}} \Rightarrow \widehat{L}_{\cdot\wedge T_c^{\omega}}$ as $n \to \infty$.

Then for every $h : D([0,T], \mathbf{R}_+^{\mathbf{Z}}) \to \mathbf{R}$ and $g : D([0,T], \mathbf{Z}) \to \mathbf{R}$ bounded and continuous

$$\begin{aligned}\lim_{n\to\infty} \mathbf{E}[h(X^n)g(\widetilde{L}^n_{\cdot\wedge T_c^n})] &= \widehat{\mathbf{E}}[h(\widehat{X}^n)g(\widetilde{L}^n_{\cdot\wedge T_c^n})] \\ &= \lim_{n\to\infty}\int_\Omega h(\widehat{X}^n(\omega))E^\omega[g(\widetilde{L}^{\omega,n}_{\cdot\wedge T_c^{\omega,n}}(\omega,\cdot))]\widehat{\mu}(\mathrm{d}\omega) \\ &= \int_\Omega h(\widehat{X}(\omega))E^\omega[g(\widehat{L}_{\cdot\wedge T_c^\omega}(\omega,\cdot))]\widehat{\mu}(\mathrm{d}\omega) \\ &= \widehat{\mathbf{E}}[h(\widehat{X})g(\widehat{L}_{\cdot\wedge T_c})]\end{aligned}$$

(where interchanging the limit and the integral is valid since both h and g are bounded).

To complete the proof, it is enough to show that $\mathbf{E}[P^\omega[\widetilde{L}^n = \widehat{L}^n]] \to 1$ as $n \to \infty$. Here we benefit from using the same sequence of random variables to construct $\widetilde{L}^n$ and $\widehat{L}^n$. These processes can only differ if at a branch point of $\mathcal{T}^n$, ξ_k lies in one of two intervals whose total length is bounded by twice the modulus of the difference between $\widetilde{p}^n$ and p^n, which is in turn bounded by a constant times

$$s^n_{\underline{t}} - (1 - \mu_n(\widehat{X}^n_{\underline{t}})).$$

The branch points of $\mathcal{T}^n$ are independent of $\widehat{X}^n$ and their total number N^n is bounded. Using the inequality (1.3.16) with θ replaced by τ_n, Markov's inequality, and a union bound, conditioning on $\mathcal{T}^n$ (which we recall is independent of $\widehat{X}^n$), we have that the probability that $|\widehat{X}^n_{\underline{t}+\tau_n} - \widehat{X}^n_{\underline{t}}| > J\sqrt{\eta_n}$ at one of the branch times of $\mathcal{T}^n$ is bounded by a constant times N^n/J.

Now in Lemma 1.4.1 choose a sequence $J_n \to \infty$ so that the right hand side of (1.4.1) tends to zero as $n \to \infty$ and recall that $\mathbf{E}[N^n]$ is uniformly bounded in n and the proof is complete. □

1.5. Central limit theorem for X^n.

Although we now have an expression for the scaling limit of our population model, and the behaviour of an ancestral lineage conditional on the history of the population, since we do not know how to reverse the population model with respect to time, or even conditions under which it has a stationary distribution, it is hard to extract quantitative results. Instead, we now consider the case in which the population size is almost constant. More precisely, we consider the case in which $\delta_n \to 0$ as $n \to \infty$, so that the limiting population model is deterministic and, under conditions that guarantee that this deterministic limit has a stable equilibrium, we examine the small stochastic fluctuations about that equilibrium as we pass to the limit. More precisely, in this section we prove Theorem 1.2.14, which establishes the

limiting behaviour of

$$Z_t^n(x) := \frac{X_t^n(x) - u_*}{\sqrt{\delta_n}},$$

where u_* is the equilibrium of the deterministic limiting system.

In §1.5.1 we establish the C-tightness of the sequence of processes $(Z^n)_{n\in\mathbf{N}}$ in the product topology. This follows a similar scheme to, and uses results of, §1.3.1. In §1.5.2, we characterise the weak limit points as a solution to a martingale problem that itself uniquely characterises the Ornstein–Uhlenbeck corrections. In §1.5.3 we establish the convergence of the sequence in $\ell_{\Gamma,p}$.

1.5.1. C-tightness of Z^n in the product topology. The proof that Z^n is C-tight in $D([0,T],\mathbf{R}_+^{\mathbf{Z}})$ follows the same schedule as the proof of the corresponding result for X^n, ie via a series of Grönwall estimates.

We begin by using Lemma 1.3.4 to characterise the law of Z^n as a solution to an approximate martingale problem.

LEMMA 1.5.1 (Rescaled approximate martingale problem). *The law of Z^n is a solution to the following approximate martingale problem.*

For each $x \in \mathbf{Z}$,

$$\begin{aligned}\mathcal{M}_t^n(x) &:= Z_t^n(x) - Z_0^n(x)\\ &\quad - \delta_n^{-1/2}\int_0^{\underline{t}} g_n(X_{\underline{s}}^n(x))X_{\underline{s}}^n(x)\,\mathrm{d}s - \frac{m}{2}\delta_n^{-1/2}\int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x))\Delta\Phi_n(X_{\underline{s}}^n)(x)\,\mathrm{d}s \quad (1.5.1)\end{aligned}$$

is an $\mathcal{F}_t^n$-martingale such that

$$\mathcal{M}_t^n(x)^2 - \int_0^{\underline{t}} \mu_n(X_{\underline{s}}^n(x))(2-\mu_n(X_{\underline{s}}^n(x)))X_{\underline{s}}^n(x)\,\mathrm{d}s + O_{\mathbf{P}}(\tau_n)$$

is also a martingale, and for each pair $x, y \in \mathbf{Z}$ with $x \neq y$,

$$\mathbf{E}[\mathcal{M}_t^n(x)\mathcal{M}_t^n(y)] = 0.$$

PROOF. The result follows immediately from Lemma 1.3.4, defining $\mathcal{M}^n := \delta_n^{-1/2}M^n$ and recalling the definition of the process Z^n. □

As in Lemma 1.3.5, we are able to use the estimates of Lemma 1.3.2 and Lemma 1.3.3 directly to obtain a bound for the martingale differences $\mathcal{M}_{\underline{t}+\tau_n}^n - \mathcal{M}_{\underline{t}}^n$. As with the proofs of Section 1.3.1, this estimate will be necessary for using the Burkholder predictable square function inequality in obtaining the tightness of Z^n in the product topology.

LEMMA 1.5.2. *Let $p \geq 1$ and suppose that* (1.3.8) *holds. Then there exists a constant $C \equiv C(p, \overline{\gamma}, \overline{g}, \Gamma, T, \delta)$ such that*

$$\mathbf{E}\Big[\sup_{0\leq t\leq T} \|\mathcal{M}^n_{\underline{t}+\tau_n} - \mathcal{M}^n_{\underline{t}}\|^{2p}_{\Gamma,2p}\Big] \leq C\tau_n^{p-1}.$$

Moreover, $\sup_{0\leq t\leq T} |\mathcal{M}^n_{\underline{t}+\tau_n}(x) - \mathcal{M}^n_{\underline{t}}(x)| = O_{\mathbf{P}}(\tau_n^{1/2}\eta_n^{-1/2p})$ *as* $n \to \infty$.

PROOF. We have done most of the work. By the inequality (1.3.7) in the proof of Lemma 1.3.2 with p replaced by $2p$, there is a constant $C(p, \overline{\gamma}, \overline{g}) < \infty$ such that

$$\begin{aligned}&\mathbf{E}\Big[\sup_{0\leq \underline{s}\leq t}|M^n_{\underline{s}+\tau_n}(x) - M^n_{\underline{s}}(x)|^{2p}\Big]\\ &\quad\leq C(p, \overline{\gamma}, \overline{g})\eta_n^p\tau_n^{-1}\mathbf{E}\Big[\int_0^t \Big(1 + |X^n_{\underline{s}}(x-1)|^p + |X^n_{\underline{s}}(x)|^p + |X^n_{\underline{s}}(x+1)|^p\Big)\,\mathrm{d}s\Big].\end{aligned}$$

Multiplying by δ_n^{-p}, the result follows after invoking Lemma 1.3.3. □

Following a similar scheme to Lemma 1.3.3, we now bound the expected value of the supremum of the $\ell_{\Gamma,p}$-norm of Z^n_t for $t \in [0, T]$. Here, slightly more work is required as we wish to approximate the values of the parameters at the stationary population density u_*; this requires controlling the relevant remainder terms.

LEMMA 1.5.3 ($\ell_{\Gamma,p}$ estimate for Z^n). *Under Assumptions 1.2.12 suppose further that $p \geq 2$, and*

$$\sup_{n\in\mathbf{N}} \mathbf{E}[\|Z^n_0\|^p_{\Gamma,p}] < \infty \tag{1.5.2}$$

then there exists a constant $C \equiv C(p, T, \Gamma, \overline{g}, \overline{\gamma}) < \infty$ such that

$$\sup_{n\in\mathbf{N}} \mathbf{E}\big[\sup_{0\leq t\leq T}\|Z^n_t\|^p_{\Gamma,p}\big] \leq C.$$

REMARK 1.5.4. Observe that under Assumption (1.5.2),

$$\sup_{n\in\mathbf{N}} \mathbf{E}[\|X^n_0\|^p_{\Gamma,p}] = \sup_{n\in\mathbf{N}} \mathbf{E}[\|u_* + \delta_n^{1/2}Z^n_0\|^p_{\Gamma,p}] < \infty.$$

In particular, the estimates of Lemma 1.3.3 hold.

PROOF OF LEMMA 1.5.3. We shall use repeatedly—and so note for the record—that for $f : \mathbf{R} \to \mathbf{R}$ k times differentiable at x, the Taylor formula may be expressed as

$$f(x+h) - \sum_{j=0}^{k} \frac{f^{(j)}(x)}{j!}h^j = h^{k+1}\int_0^1 \frac{u^k}{k!} f^{(k+1)}(x + (1-u)h)\,\mathrm{d}u. \tag{1.5.3}$$

As in the proof of the corresponding result for X^n, we define

$$h^n(t) := \mathbf{E}\big[\sup_{0\le s\le t}\|Z^n_s\|^p_{\Gamma,p}\big], \qquad 0 \le t \le T,$$

for which we derive growth estimates and apply Grönwall's lemma. We write Z^n as a solution to the stochastic equation

$$Z^n_s(x) = Z^n_0(x) + \delta_n^{-1/2}\int_0^{\underline{s}} g_n(X^n_{\underline{u}}(x))X^n_{\underline{u}}(x)\,\mathrm{d}u + \frac{m}{2}\delta_n^{-1/2}\int_0^{\underline{s}} r_n(X^n_{\underline{u}}(x))\Delta_1\Phi_n(X^n_{\underline{u}})(x)\,\mathrm{d}u + \mathcal{M}^n_{\underline{s}}(x).$$

Taking the $\ell_{\Gamma,p}$-norm of this expression, there is a constant $C(p) < \infty$ such that

$$\begin{aligned}
&h^n(t)\\
&= \mathbf{E}\big[\sup_{0\le s\le t}\|Z^n_s\|^p_{\Gamma,p}\big]\\
&\le \mathbf{E}\Big[\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}|Z^n_s(x)|^p\Big]\\
&\le C(p)\mathbf{E}\Big[\|Z^n_0\|^p_{\Gamma,p} + \delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\Big|\int_0^{\underline{s}} g_n(X^n_{\underline{u}}(x))X^n_{\underline{u}}(x)\,\mathrm{d}u\Big|^p && (1.5.4)\\
&\qquad + \delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\Big|\int_0^{\underline{s}} r_n(X^n_{\underline{u}}(x))\Delta_1\Phi_n(X^n_{\underline{u}})(x)\,\mathrm{d}u\Big|^p && (1.5.5)\\
&\qquad + \sum_{x\in\mathbf{Z}}\Gamma_x\sup_{0\le s\le t}|\mathcal{M}^n_{\underline{s}}(x)|^p\Big]. && (1.5.6)
\end{aligned}$$

We proceed by estimating the second term within the expectation in (1.5.4). Recall that we have defined $G : u \mapsto ug(u)$ which has bounded first derivative. By Taylor's formula

applied to the map G, and using that $(a+b)^p \le C(p)(a^p+b^p)$ for $a,b>0$ since $p \ge 1$,

$$
\begin{aligned}
&\delta_n^{-p/2} \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \left| \int_0^{\underline{s}} g_n(X^n_{\underline{u}}(x)) X^n_{\underline{u}}(x)\,\mathrm{d}u \right|^p \\
&\le \delta_n^{-p/2} \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \left| \int_0^{\underline{s}} (g_n-g)(X^n_{\underline{u}}(x)) X^n_{\underline{u}}(x)\,\mathrm{d}u \right|^p \\
&\qquad + \delta_n^{-p/2} \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \left| \int_0^{\underline{s}} g(X^n_{\underline{u}}(x)) X^n_{\underline{u}}(x)\,\mathrm{d}u \right|^p \\
&\le C(p)\delta_n^{-p/2} \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \left| \int_0^{\underline{s}} (g_n-g)(X^n_{\underline{u}}(x)) X^n_{\underline{u}}(x)\,\mathrm{d}u \right|^p \\
&\qquad + C(p)\delta_n^{-p/2} \times \\
&\qquad\qquad \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \left| \int_0^{\underline{s}} \delta_n^{1/2} Z^n_{\underline{s}}(x) \int_0^1 G'(u_* + (1-\theta)\delta_n^{1/2} Z^n_{\underline{s}}(x))\,\mathrm{d}\theta\,\mathrm{d}u \right|^p .
\end{aligned}
\tag{1.5.7}
$$

We have used that $G(u_*) = u_* g(u_*) = 0$. Invoking Lemma 1.3.3, we have the uniform bound that $\sup_{n\in\mathbf{N}} \mathbf{E}[\sup_{0\le t\le T} \|X^n_t\|^p_{\Gamma,p}] < \infty$. In particular, applying Jensen's inequality, and using (1.2.6),

$$
\begin{aligned}
&\mathbf{E}\left[\delta_n^{-p/2} \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \left| \int_0^{\underline{s}} (g_n-g)(X^n_{\underline{u}}(x)) X^n_{\underline{u}}(x)\,\mathrm{d}u \right|^p \right] \\
&\le \mathbf{E}\left[\delta_n^{-p/2} \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} T^{p-1} \int_0^{\underline{s}} |(g_n-g)(X^n_{\underline{u}}(x))|^p |X^n_{\underline{u}}(x)|^p\,\mathrm{d}u \right] \\
&\le \sup_{n\in\mathbf{N},u\in\mathbf{R}_+} \frac{|g_n(u)-g(u)|^p}{\delta_n^{p/2}} \sum_{x\in\mathbf{Z}} \Gamma(x) \mathbf{E}\left[\sup_{0\le s\le t} T^{p-1} \int_0^{\underline{s}} |X^n_{\underline{u}}(x)|^p\,\mathrm{d}u \right] \\
&\le \sup_{n\in\mathbf{N},u\in\mathbf{R}_+} \frac{|g_n(u)-g(u)|^p}{\delta_n^{p/2}} \sum_{x\in\mathbf{Z}} \Gamma(x) \mathbf{E}\left[T^{p-1} \int_0^{\underline{t}} |X^n_{\underline{u}}(x)|^p\,\mathrm{d}u \right] \\
&\le \sup_{n\in\mathbf{N},u\in\mathbf{R}_+} \frac{|g_n(u)-g(u)|^p}{\delta_n^{p/2}} T^p \sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{0\le t\le T} \|X^n_t\|^p_{\Gamma,p} \right] \\
&\le C(p,g,\overline{\gamma},\Gamma,T).
\end{aligned}
$$

As in Section 1.3.1, we are able to freely interchange the above 'integrals' over time with summation and expectation as we note that they are really just finite sums over a finite number of generation times.

Turning to the second term in the sum on the right hand side of the last inequality

in (1.5.7), by Jensen's inequality,

$$\begin{aligned}
&\mathbf{E}\Big[\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\Big|\int_0^{\underline{s}}\delta_n^{1/2}Z^n_{\underline{s}}(x)\int_0^1 G'(u_*+(1-\theta)\delta_n^{1/2}Z^n_{\underline{s}}(x))\,\mathrm{d}\theta\,\mathrm{d}u\Big|^p\Big]\\
&\le \mathbf{E}\Big[T^{p-1}\|G'\|_\infty^p\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\int_0^{\underline{s}}|Z^n_{\underline{s}}(x)|^p\,\mathrm{d}u\Big]\\
&\le T^{p-1}\|G'\|_\infty^p\int_0^{\underline{t}}h^n(s)\,\mathrm{d}s.
\end{aligned}$$

We now we bound (1.5.5). Using Jensen's inequality again (and that r_n, being a probability, is at most 1),

$$\begin{aligned}
&\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\Big|\int_0^{\underline{s}}r_n(X^n_{\underline{u}}(x))\Delta_1\Phi_n(X^n_{\underline{u}})(x)\,\mathrm{d}u\Big|^p\\
&\le T^{p-1}\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\int_0^{\underline{s}}|\Delta_1\Phi_n(X^n_{\underline{u}})(x)|^p\,\mathrm{d}u\\
&\le C(p,T)\Big[\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\int_0^{\underline{s}}|\Delta_1(\Phi_n-\Phi)(X^n_{\underline{u}})(x)|^p\,\mathrm{d}u\\
&\qquad+\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\int_0^{\underline{s}}|\Delta_1\Phi(X^n_{\underline{u}})(x)|^p\,\mathrm{d}u\Big].
\end{aligned}$$

These terms are handled in much the same way as those above for G.

Recalling Assumptions 1.2.12,

$$\begin{aligned}
&\mathbf{E}\Big[\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\int_0^{\underline{s}}|\Delta_1(\Phi_n-\Phi)(X^n_{\underline{u}})(x)|^p\,\mathrm{d}u\Big]\\
&\le C(p)\sup_{n\in\mathbf{N}}\frac{\|\gamma_n-\gamma\|_\infty^p}{\delta_n^{p/2}}\times\\
&\qquad\mathbf{E}\Big[\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\int_0^{\underline{s}}|X^n_{\underline{u}}(x+1)|^p+2|X^n_{\underline{u}}(x)|^p+|X^n_{\underline{u}}(x-1)|^p\,\mathrm{d}u\Big]\\
&\le C(p,T,\Gamma)\sup_{n\in\mathbf{N}}\frac{\|\gamma_n-\gamma\|_\infty^p}{\delta_n^{p/2}}\sup_{n\in\mathbf{N}}\mathbf{E}\Big[\sup_{0\le t\le T}\|X^n_t\|^p_{\Gamma,p}\Big]\\
&\le C(p,g,\overline{\gamma},\Gamma,T).
\end{aligned}$$

In order to control

$$\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\sup_{0\le s\le t}\int_0^{\underline{s}}|\Delta_1\Phi(X^n_{\underline{u}})(x)|^p\,\mathrm{d}u$$

we replace each of the terms within the discrete Laplacian, $\Phi(X^n_{\underline{u}}(y))$, $y\in\{x-1,x,x+1\}$

by

$$\Phi(X^n_{\underline{u}}(y)) = \Phi(u_*) + \delta_n^{1/2} Z^n_{\underline{u}}(y) \int_0^1 \Phi'(u_* + (1-\theta)\delta_n^{1/2} Z^n_{\underline{u}}(y))\,\mathrm{d}\theta.$$

Recall that we have assumed that Φ has a bounded first derivative. Then we see that

$$\begin{aligned}
&\mathbf{E}\Big[\delta_n^{-p/2} \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \int_0^{\underline{s}} |\Delta_1 \Phi(X^n_{\underline{u}})(x)|^p \,\mathrm{d}u\Big] \\
&\le C(p)\mathbf{E}\Big[\|\Phi'\|_\infty^p \sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} \int_0^{\underline{s}} \big[|Z^n_{\underline{u}}(x+1)|^p + |Z^n_{\underline{u}}(x)|^p + |Z^n_{\underline{u}}(x-1)|^p\big]\,\mathrm{d}u\Big] \\
&\le \mathbf{E}\Big[C(\Gamma,\Phi) \int_0^{\underline{t}} \|Z^n_{\underline{u}}\|^p_{\Gamma,p}\,\mathrm{d}u\Big] \\
&\le C(\Gamma,\Phi) \int_0^{\underline{t}} h^n(s)\,\mathrm{d}s.
\end{aligned}$$

Finally, we must control (1.5.6). By the Burkholder predictable square function inequality, Lemma 1.6.1, there is a constant $C(p) < \infty$ such that

$$\begin{aligned}
&\mathbf{E}\Big[\sum_{x\in\mathbf{Z}} \Gamma(x) \sup_{0\le s\le t} |\mathcal{M}^n_{\underline{s}}(x)|^p\Big] \\
&\le C(p)\Big\{\sum_{x\in\mathbf{Z}} \Gamma(x)\mathbf{E}\Big[\Big|\int_0^{\underline{t}} \mu_n(X^n_{\underline{s}}(x))(2-\mu_n(X^n_{\underline{s}}(x)))X^n_{\underline{s}}(x)\,\mathrm{d}s \\
&\qquad + \tau_n \int_0^{\underline{t}} b^n(X^n_{\underline{s}},x)\,\mathrm{d}s\Big|^{p/2}\Big] \\
&\qquad + \mathbf{E}\Big[\sup_{0\le s\le t} \|\mathcal{M}^n_{\underline{s}+\tau_n} - \mathcal{M}^n_{\underline{s}}\|^p_{\Gamma,p}\Big]\Big\} \\
&\le C(p,T)\sum_{x\in\mathbf{Z}} \Gamma(x)\mathbf{E}\Big[\int_0^{\underline{t}} |X^n_{\underline{s}}(x)|^{p/2}\,\mathrm{d}s \\
&\quad + \tau_n^{p/2}\int_0^{\underline{t}} \big\{|X^n_{\underline{s}}(x+1)|^{p/2} + |X^n_{\underline{s}}(x)|^{p/2} + |X^n_{\underline{s}}(x-1)|^{p/2}\big\}\,\mathrm{d}s\Big] \\
&\qquad + C(p,\overline{\gamma},\overline{g},\Gamma,T,\delta)\tau_n^{p/2-1} \\
&\le C(p,\overline{\gamma},\overline{g},\Gamma,T,\delta)\mathbf{E}\Big[1 + \int_0^t \|Z^n_{\underline{s}}\|^p_{\Gamma,p}\,\mathrm{d}s\Big] \\
&\le C(p,\overline{\gamma},\overline{g},\Gamma,T,\delta)\Big\{1 + \int_0^t h^n(s)\,\mathrm{d}s\Big\},
\end{aligned}$$

where to arrive at the second inequality we have used that $\|\mu_n(2-\mu_n)\|_\infty \le 1$ (since μ_n is a probability) and the estimate (1.3.11) on b^n; the third follows by substituting the definition of the process Z^n and the bound $X \le 1 + X^2$ for $X \ge 0$.

Returning to (1.5.4)–(1.5.6), combining the estimates above, we see that there exists a

constant C, independent of n such that

$$h^n(t) \leq C\left(1 + \int_0^t h^n(s)\,\mathrm{d}s\right)$$

and invoking Grönwall's lemma, we see that h^n is bounded uniformly in $0 \leq t \leq T$ and $n \in \mathbf{N}$, completing the proof. □

We now prove C-tightness of Z^n in the product topology. As in Section 1.3.1 we shall use Theorem 1.3.6.

PROPOSITION 1.5.5 (Tightness of Z^n). *Let $\mathbf{R}_+^{\mathbf{Z}}$ be equipped with the product topology. If for some $p \geq 2$ we have* (1.5.2)*, then the sequence of processes $(Z^n)_{n\in\mathbf{N}}$ is C-tight in $D([0,T], \mathbf{R}_+^{\mathbf{Z}})$.*

PROOF. Following the scheme set out in the proof of Proposition 1.3.7, it suffices to show that each individual coordinate $\{(Z_t^n(x))_{0\leq t\leq T}\}_{n\in\mathbf{N}}$ is tight, for then a diagonalisation argument yields tightness of the whole process in the product topology.

Fix $x \in \mathbf{Z}$ and consider the sequence $Z^n(x)$. We wish to apply Theorem 1.3.6. Using Lemma 1.5.3 and Markov's inequality we may bound the coordinates of Z^n in probability and so infer the tightness of $Z_t^n(x)$ for each fixed t.

Since the process is Markov, to verify Condition (1.3.14) it suffices to show that for some $p \geq 2$,

$$\sup_{n\in\mathbf{N}} \sup_{\substack{0\leq s\leq t\leq T \\ |t-s|\leq\theta<1}} \mathbf{E}\left[\|Z_t^n - Z_s^n\|_{\Gamma,p}^p\right] \leq C\theta, \tag{1.5.8}$$

for some constant $C < \infty$. This will be our aim.

First we use the stochastic equation (1.5.1) to write

$$\begin{aligned} Z_t^n(x) - Z_s^n(x) = \delta_n^{-1/2} \int_{\underline{s}}^{\underline{t}} g_n(X_{\underline{u}}^n(x)) X_{\underline{u}}^n(x)\,\mathrm{d}u \\ + \delta_n^{-1/2}\frac{m}{2}\int_{\underline{s}}^{\underline{t}} r_n(X_{\underline{u}}^n(x))\Delta_1\Phi_n(X_{\underline{u}}^n)(x)\,\mathrm{d}u + \mathcal{M}_{\underline{t}}^n(x) - \mathcal{M}_{\underline{s}}^n(x). \end{aligned}$$

Taking the $\ell_{\Gamma,p}$-norm,

$$\begin{aligned}
&\mathbf{E}\big[\|Z_t^n - Z_s^n\|_{\Gamma,p}^p\big] \\
&= \mathbf{E}\bigg[\sum_{x\in\mathbf{Z}} \Gamma(x)|Z_t^n(x) - Z_s^n(x)|^p\bigg] \\
&\le C(p)\mathbf{E}\bigg[\delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\bigg|\int_{\underline{s}}^{\underline{t}} g_n(X_{\underline{u}}^n(x))X_{\underline{u}}^n(x)\,\mathrm{d}u\bigg|^p \\
&\qquad + \delta_n^{-p/2}\sum_{x\in\mathbf{Z}}\Gamma(x)\bigg|\int_{\underline{s}}^{\underline{t}} r_n(X_{\underline{u}}^n(x))m\Delta_1\Phi_n(X_{\underline{u}}^n)(x)\,\mathrm{d}u\bigg|^p \\
&\qquad + \sum_{x\in\mathbf{Z}}\Gamma(x)|\mathcal{M}_{\underline{t}}^n(x) - \mathcal{M}_{\underline{s}}^n(x)|^p\bigg] \\
&\le C\theta^{p/2}.
\end{aligned}$$

The remaining details are omitted; the proof mimics that of Proposition 1.3.7, using calculations entirely similar to those in the proof of the previous lemma (in particular, we use the uniform estimate on Z^n derived there). That the error behaves as $\theta^{p/2}$ is a consequence of the Burkholder predictable square function inequality used to estimate the expectation of the terms involving the martingale. Since $p \ge 2$ we obtain (1.5.8).

Condition (1.3.15) follows by another application of Markov's inequality. □

1.5.2. Proof of Theorem 1.2.14. We are now able to prove Theorem 1.2.14.

As in the proof of Theorem 1.2.5, we use the fact, taken from the proof of Lemma II.4.5 of [108], that the following uniform integrability condition guarantees that any weak limit point of $\mathcal{M}^n(x)$ is an L^2-martingale.

LEMMA 1.5.6 (Uniform integrability). *Suppose that* (1.5.2) *holds for some* $p \ge 2$. *Then for each* $x \in \mathbf{Z}$ *and* $0 \le t \le T$, *the family of random variables* $\{|\mathcal{M}_{\underline{t}}^n(x)|^2\}_{n\in\mathbf{N}}$ *is uniformly integrable.*

PROOF. It is sufficient to show that

$$\sup_{n\in\mathbf{N}} \mathbf{E}\big[\sup_{t\le T}|\mathcal{M}_{\underline{t}}^n(x)|^{2p}\big] < \infty,$$

By the Burkholder predictable square function inequality (Lemma 1.6.1) and Lemma 1.5.2

$$\begin{aligned}
\mathbf{E}\big[\sup_{t\le T}|\mathcal{M}_{\underline{t}}^n(x)|^{2p}\big] \le C(p)\mathbf{E}\bigg[&\bigg|\int_0^{\underline{t}} \mu_n(X_{\underline{s}}^n(x))(2-\mu_n(X_{\underline{s}}^n(x)))X_{\underline{s}}^n(x)\,\mathrm{d}s\bigg|^p \\
&+ \bigg|\tau_n\int_0^t b_n(X_{\underline{s}}^n, x)\,\mathrm{d}s\bigg|^p + C\tau_n^{p-1}\bigg].
\end{aligned}$$

Bounding this quantity (uniformly in n) follows exactly the same pattern as above, exploiting Jensen's inequality and Lemma 1.3.3, and is omitted. □

Given the tightness of Z^n (in the product topology) established in Proposition 1.5.5, the next step is to characterise its weak limit points as a solution to a martingale problem corresponding to solutions of the system of equations (1.2.9). Referring to Theorem 1.3.16, it is clear that such a solution is unique. By Lévy's characterisation, the proof of Theorem 1.2.14 is complete once we have established the following.

PROPOSITION 1.5.7 (Limiting martingale problem). *Suppose $Z_0^n \Rightarrow \zeta$ in $\ell_{\Gamma,p}$ for some $p \geq 2$. Then $(Z^n)_{n\in\mathbf{N}}$ is C-tight in $D([0,T], \mathbf{R}_+^{\mathbf{Z}})$ with the product topology and any weak limit point Z solves the following martingale problem:*

$$\begin{cases} \text{For all } x \in \mathbf{Z}, \\ \mathcal{M}_t(x) := Z_t(x) - Z_0(x) - \int_0^t \left(u_* g'_* Z_s(x) + \frac{1}{2} m r_* \Phi'_* \Delta_1 Z_s(x)\right) \mathrm{d}s \\ \text{is an } \mathcal{F}_t\text{-martingale such that} \\ \mathcal{M}_t(x)^2 - q_*(2-q_*)u_* t, \\ \text{is also a martingale and for all } x, y \in \mathbf{Z} \text{ with } x \neq y, \\ \mathbf{E}[\mathcal{M}_t(x)\mathcal{M}_t(y)] = 0. \end{cases} \tag{MP-Z}$$

PROOF OF PROPOSITION 1.5.7. The C-tightness is the content of Proposition 1.5.5.

Since $Z_0^n \Rightarrow \zeta$ in $\ell_{\Gamma,p}$ we have $X_0^n \Rightarrow u_*$ in $\ell_{\Gamma,p}$. In particular, condition (1.5.2), and thus condition (1.3.8), holds. First we use the stochastic equation (1.5.1) to write

$$\begin{aligned} \mathcal{M}_t^n(x) &= Z_t^n(x) - Z_0^n(x) \\ &\quad - \delta_n^{-1/2} \int_0^{\underline{t}} g_n(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x) \,\mathrm{d}s - \frac{1}{2} m \delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta\Phi_n(X_{\underline{s}}^n)(x) \,\mathrm{d}s. \end{aligned} \tag{1.5.9}$$

Let us begin with the first integral. Using that $G(u_*) = g(u_*)u_* = 0 = g(u_*)$,

$$\begin{aligned}
&\delta_n^{-1/2} \int_0^{\underline{t}} g_n(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s \\
&= \delta_n^{-1/2} \int_0^{\underline{t}} G(X_{\underline{s}}^n(x))\,\mathrm{d}s + \delta_n^{-1/2} \int_0^{\underline{t}} (g_n - g)(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s \\
&= \delta_n^{-1/2} \int_0^{\underline{t}} G(u_* + \delta_n^{1/2} Z_{\underline{s}}^n(x))\,\mathrm{d}s + \delta_n^{-1/2} \int_0^{\underline{t}} (g_n - g)(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s \\
&= \int_0^{\underline{t}} u_* g'(u_*) Z_{\underline{s}}^n(x)\,\mathrm{d}s + \delta_n^{-1/2} \int_0^{\underline{t}} R_1(G, u_*, \delta_n^{1/2} Z_{\underline{s}}^n(x))\,\mathrm{d}s \\
&\qquad + \delta_n^{-1/2} \int_0^{\underline{t}} (g_n - g)(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s \\
&= \int_0^{\underline{t}} u_* g'(u_*) Z_{\underline{s}}^n(x)\,\mathrm{d}s \\
&\qquad + \delta_n^{1/2} \int_0^{\underline{t}} Z_{\underline{s}}^n(x)^2 \int_0^1 \theta G''(u_* + (1-\theta)\delta_n^{1/2} Z_{\underline{s}}^n(x))\,\mathrm{d}\theta\,\mathrm{d}s \\
&\qquad + \delta_n^{-1/2} \int_0^{\underline{t}} (g_n - g)(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s.
\end{aligned}$$

(We have used R_1 to denote the remainder in the Taylor expansion.) Let us begin by estimating the third of these terms. By assumption $\delta_n^{-1/2}(g_n - g) \to_{\mathrm{ucb}} 0$ and thus, by invoking Jensen's inequality,

$$\begin{aligned}
&\mathbf{E}\left[\left|\delta_n^{-1/2} \int_0^{\underline{t}} (g_n - g)(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s\right|^p\right] \\
&\le \sup_{n\in\mathbf{N}, u\in\mathbf{R}_+} \frac{|g_n(u) - g(u)|^p}{\delta_n^{p/2}} \cdot T^p \cdot \sup_{n\in\mathbf{N}} \frac{\mathbf{E}\big[\sup_{0\le t\le T} \|X_t^n\|_{\Gamma,p}^p\big]}{\Gamma(x)}.
\end{aligned}$$

In particular, by Lemma 1.3.3,

$$\delta_n^{-1/2} \int_0^{\underline{t}} (g_n - g)(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)\,\mathrm{d}s \overset{L^p}{\to} 0. \tag{1.5.10}$$

Similarly, since G'' is bounded by assumption, we use that $p \ge 2$ and Jensen's inequality to obtain

$$\begin{aligned}
&\mathbf{E}\left[\left|\delta_n^{1/2} \int_0^{\underline{t}} Z_{\underline{s}}^n(x)^2 \int_0^1 \theta G''(u_* + (1-\theta)\delta_n^{1/2} Z_{\underline{s}}^n(x))\,\mathrm{d}\theta\,\mathrm{d}s\right|^{p/2}\right] \\
&\qquad\qquad \le C\delta_n^{p/4} \sup_{n\in\mathbf{N}} \frac{\mathbf{E}\big[\sup_{0\le t\le T} \|Z_t^n\|_{\Gamma,p}^p\big]}{\Gamma(x)}.
\end{aligned}$$

By Lemma 1.5.3,

$$\delta_n^{1/2} \int_0^{\underline{t}} Z_{\underline{s}}^n(x)^2 X_{\underline{s}}^n(x) \int_0^1 \theta g''(u_* + (1-\theta)\delta_n^{1/2} Z_{\underline{s}}^n(x))\, \mathrm{d}\theta\, \mathrm{d}s \overset{L^{p/2}}{\to} 0. \tag{1.5.11}$$

Now we turn to the second integral in (1.5.9)

$$\begin{aligned}
&\frac{m}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta\Phi_n(X_{\underline{s}}^n)(x)\, \mathrm{d}s \\
&= \frac{m}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta\Phi(X_{\underline{s}}^n)(x)\, \mathrm{d}s + \frac{m}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r(X_{\underline{s}}^n(x)) \Delta(\Phi_n - \Phi)(X_{\underline{s}}^n)(x)\, \mathrm{d}s \\
&= \frac{m}{2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Phi'(u_*) \Delta_1 Z_{\underline{s}}^n(x)\, \mathrm{d}s \\
&\quad + \frac{m}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta_1 R_1(\Phi, u_*, \delta_n^{1/2} Z_{\underline{s}}^n(\cdot))(x)\, \mathrm{d}s \\
&\quad + \frac{m}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta_1 (\Phi_n - \Phi)(X_{\underline{s}}^n)(x)\, \mathrm{d}s.
\end{aligned}$$

To control the last term we apply Jensen's inequality to obtain

$$\begin{aligned}
&\mathbf{E}\left[\left| \frac{1}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta_1 (\Phi_n - \Phi)(X_{\underline{s}}^n)(x)\, \mathrm{d}s \right|^p \right] \\
&\le \frac{1}{2^p}\delta_n^{-p/2} T^{p-1} \mathbf{E}\left[\int_0^{\underline{t}} |\Delta_1 (\Phi_n - \Phi)(X_{\underline{s}}^n)(x)|^p\, \mathrm{d}s \right] \\
&\le C(p,T)\delta_n^{-p/2} \Bigg\{ \mathbf{E}\left[\int_0^{\underline{t}} |(\gamma_n - \gamma)(X_{\underline{s}}^n(x+1)) X_{\underline{s}}^n(x+1)|^p\, \mathrm{d}s \right] \\
&\qquad\qquad + \mathbf{E}\left[\int_0^{\underline{t}} |(\gamma_n - \gamma)(X_{\underline{s}}^n(x-1)) X_{\underline{s}}^n(x-1)|^p\, \mathrm{d}s \right] \\
&\qquad\qquad + \mathbf{E}\left[\int_0^{\underline{t}} |(\gamma_n - \gamma)(X_{\underline{s}}^n(x)) X_{\underline{s}}^n(x)|^p\, \mathrm{d}s \right] \Bigg\} \\
&\le C(p,T,\Gamma,x) \sup_{n\in\mathbf{N}} \frac{\|\gamma_n - \gamma\|_\infty^p}{\delta_n^{p/2}} \sup_{n\in\mathbf{N}} \mathbf{E}\big[\sup_{0\le s\le T} \|X_s^n\|_{\Gamma,p}^p\big].
\end{aligned}$$

By Lemma 1.3.3,

$$\frac{1}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta_1 (\Phi_n - \Phi)(X_{\underline{s}}^n)(x)\, \mathrm{d}s \overset{L^p}{\to} 0. \tag{1.5.12}$$

Using that Φ has bounded second derivative (and (1.5.3))

$$\frac{1}{2}\delta_n^{-1/2} \int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x)) \Delta_1 R_1(\Phi, u_*, \delta_n^{1/2} Z_{\underline{s}}^n(\cdot))(x)\, \mathrm{d}s \overset{L^{p/2}}{\to} 0. \tag{1.5.13}$$

Combining (1.5.10), (1.5.11), (1.5.12) and (1.5.13), we may recast the stochastic equation (1.5.1) as

$$\mathcal{M}_t^n(x) = Z_t^n(x) - Z_0^n(x) - \int_0^{\underline{t}} u_* g'(u_*) Z_{\underline{s}}^n(x)\,\mathrm{d}s - \frac{m}{2}\int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x))\Phi'(u_*)\Delta_1 Z_{\underline{s}}^n(x)\,\mathrm{d}s - E_{\underline{t}}^n(x), \quad (1.5.14)$$

where $E_{\underline{t}}^n(x) \overset{L^{p/2}}{\to} 0$ as $n \to \infty$.

After passing to a subsequence if necessary, assume that Z^n converges weakly to Z in $D([0,T], \mathbf{R}_+^{\mathbf{Z}})$, where $\mathbf{R}_+^{\mathbf{Z}}$ is equipped with the product topology. Fix $x \in \mathbf{Z}$. By the Skorokhod representation theorem, there exists a probability space, and on it a sequence of random variables $\tilde{Z}^n$, such that $\tilde{Z}^n$ has the same law as Z^n and $\tilde{Z}^n$ converges almost surely to Z in $D([0,T], \mathbf{R}_+^{\mathbf{Z}})$. We suppose that we have done this and drop the 'tilde' from our notation.

By, for example, Lemma 3.10.1 of Ethier and Kurtz [40], the convergence of $Z^n(x)$ to $Z(x)$ is uniform on compact sets. Since r_n converges uniformly on compacts, we have that,

$$\int_0^{\underline{t}} u_* g'(u_*) Z_{\underline{s}}^n(x)\,\mathrm{d}s \overset{\text{a.s.}}{\to} \int_0^t u_* g'_* Z_s(x)\,\mathrm{d}s,$$

and

$$\frac{m}{2}\int_0^{\underline{t}} r_n(X_{\underline{s}}^n(x))\Phi'(u_*)\Delta_1 Z_{\underline{s}}^n(x)\,\mathrm{d}s \overset{\text{a.s.}}{\to} \frac{m}{2}\int_0^t r_*\Phi'_*\Delta_1 Z_s(x)\,\mathrm{d}s.$$

Passing to the limit $n \to \infty$ in (1.5.14), there is a random variable $\mathcal{M}_t(x)$ such that $\mathcal{M}^n(x) \overset{P}{\to} \mathcal{M}(x)$ and

$$\mathcal{M}_t(x) = Z_t(x) - Z_0(x) - \int_0^t u_* g'_* Z_s(x)\,\mathrm{d}s - \frac{1}{2}\int_0^t r_*\Phi'_*\Delta_1 Z_s(x)\,\mathrm{d}s.$$

By Lemma 1.5.6, $\mathcal{M}(x)$ is an L^2-martingale, and since

$$\mathcal{M}_t^n(x)^2 - \int_0^{\underline{t}} \mu_n(X_{\underline{s}}^n(x))(2-\mu_n(X_{\underline{s}}^n(x)))X_{\underline{s}}^n(x)\,\mathrm{d}s - \tau_n\int_0^{\underline{t}} b^n(X_{\underline{s}}^n, x)\,\mathrm{d}s$$

is a martingale, and

$$\mathbf{E}\left[\sup_{0\le t\le T}\left|\tau_n\int_0^{\underline{t}} b^n(X_{\underline{s}}^n, x)\,\mathrm{d}s\right|^p\right] \le C(p,\overline{g},\overline{\gamma},T,\Gamma,x)\tau_n^p$$

we identify

$$\langle \mathcal{M}(x)\rangle_t = \mu_*(2-\mu_*)u_* t.$$

After an argument identical to that at the end of the proof of Proposition 1.3.11 we may conclude that $\mathbf{E}[\mathcal{M}_t(x)\mathcal{M}_t(y)] = 0$ for all $x \neq y$. □

1.5.3. Proof of Theorem 1.2.15. As was the case in Section 1.3.3, the stronger conclusion of tightness in the topology of $\ell_{\Gamma,p}$ requires us to prove a uniform integrability condition. Again, we make use of the criteria for relative compactness in $\ell_{\Gamma,p}$ established in Lemma 1.3.14.

It is sufficient to establish the following.

PROPOSITION 1.5.8 (Tightness in $\ell_{\Gamma,p}$). *If for some $p \geq 2$ we have*

$$\sup_{n\in\mathbf{N}} \mathbf{E}\left[\|Z_0^n\|_\infty^p\right] = \sup_{n\in\mathbf{N}} \mathbf{E}\left[\sup_{x\in\mathbf{Z}} |Z_0^n(x)|^p\right] < \infty,$$

then the sequence of processes $(Z^n)_{n\in\mathbf{N}}$ is C-tight in $D([0,T],\ell_{\Gamma,p})$.

PROOF. Once again we use Theorem 1.3.6. The required continuity estimate, the second condition of the Theorem, is established in equation (1.5.8). Given Lemma 1.5.3 it is sufficient to establish that for each fixed time t,

$$\lim_{N\to\infty} \sup_{n\in\mathbf{N}} \mathbf{P}\left[\sum_{|x|>N} \Gamma(x)|Z_t^n(x)|^p > \varepsilon\right] = 0.$$

The proof is identical to that of Proposition 1.3.15 and is thus omitted. □

1.5.4. Proof of Propositions 1.2.16, 1.2.18, and 1.2.19. To pave the way to making some preliminary remarks about the impact of the fluctuations in population density on the motion of ancestral lineages, in this short section we write down an expression for the stationary distribution of the first order fluctuations Z, when it exists.

We begin by giving proofs of Proposition 1.2.16 and Proposition 1.2.18.

PROOF OF PROPOSITION 1.2.16. Setting

$$a_0 := u_* g'_* - m r_* \Phi'_*,$$

and

$$a_{-1} := a_1 := \frac{1}{2} m r_* \Phi'_*,$$

and $a_n = 0$ for all $|n| > 1$, the operator A given in (1.2.10) may be expressed as the bi-infinite matrix $A = (a_{ij})$ with entries $a_{ij} = a_{i-j}$, $i, j \in \mathbf{Z}$. Acting on $\ell_2(\mathbf{Z})$, such operators are called Laurent operators.

Necessary and sufficient conditions for Laurent operators to be bounded are given by a theorem of Toeplitz [125], Theorem 1.6.3 in the appendix. If $\mathbf{T}$ denotes the unit circle in the complex plane, the theorem asserts that the operator A is bounded if and only if there exists a function $f \in L^\infty(\mathbf{T})$ with Fourier coefficients $\{a_n\}_{n\in\mathbf{Z}}$.

For our choice of the collection $\{a_n\}_{n\in\mathbf{Z}}$ the function $f \in L^\infty(\mathbf{T})$ given by

$$f(z) := a_0 + a_1(z^{-1} + z) \tag{1.5.15}$$

has Fourier coefficients $\{a_n\}_{n\in\mathbf{Z}}$ and moreover A is bounded. □

PROOF OF PROPOSITION 1.2.18. As we have seen in the proof of Proposition 1.2.16, the operator A given in (1.2.10) is a Laurent operator. In particular, $A = (a_{ij})$ with entries $A_{ij} = a_{i-j}$, $i, j \in \mathbf{Z}$, and there exists a function $f \in L^\infty(\mathbf{T})$ with Fourier coefficients $\{a_n\}_{n\in\mathbf{Z}}$ defined by (1.5.15). In this case, Toeplitz's Theorem (Theorem 1.6.3 in the appendix) also asserts that $\mathrm{spec}(A)$ is given by the essential range of f.

Since $f(e^{i\theta}) = a_0 + 2a_1\cos\theta$, this function f has essential range given by the interval $[a_0 - 2|a_1|, a_0 + 2|a_1|]$. The spectrum of A is thus wholly contained in the open negative half plane if and only if $a_0 + 2|a_1| < 0$. Equivalently, if

$$u_* g'_* + m r_*(|\Phi'_*| - \Phi'_*) < 0,$$

giving the condition (1.2.11). □

We turn now to the existence of a unique stationary distribution for solutions to the system of equations (1.2.9).

LEMMA 1.5.9. *Under the assumptions of Theorem 1.2.14 and Assumption 1.2.17, the system of equations* (1.2.9) *admits a unique reversible stationary distribution π on $\ell_{\Gamma,p}$. Moreover, the stationary distribution π is Gaussian.*

PROOF. By Corollary 6.2 and Theorem 6.6 of Jnowska-Michalik [69], existence and uniqueness of a stationary distribution for the system of equations (1.2.9) is implied by the exponential stability of the semigroup $(T_t)_{t\geq 0}$ generated by the operator A defined in (1.2.10). Since, by Proposition 1.2.16, A is a bounded operator, exponential stability of $(T_t)_{t\geq 0}$ is equivalent to the condition of Assumption 1.2.17; that is, that the spectrum of A lies in the open negative half plane $\mathbf{C}_-$ (see, for example, Theorem I.3.14 of Engel and Nagel [35]).

That the stationary distribution is Gaussian follows from Proposition 6.1 of Jnowska-Michalik [69]. □

PROPOSITION 1.5.10. *The Gaussian distribution π of Lemma 1.5.9 is characterised by equations* (1.2.12) *and* (1.2.13).

PROOF. Again this is standard. We establish the covariance of $Z_{T+t}(x)$ and $Z_T(x)$ for an arbitrary initial condition, and let $T \to \infty$.

Proceeding formally first, following the notation of [29],

$$Z_{T+t} = T_{T+t} Z_0 + \int_0^{T+t} T_{T+t-s} \sqrt{\mu_*(2-\mu_*)u_*}\, \mathrm{d}W_s,$$
$$Z_T = T_T Z_0 + \int_0^{T} T_{T-s} \sqrt{\mu_*(2-\mu_*)u_*}\, \mathrm{d}W_s;$$

in other words, for a function $f \in \ell_2(\mathbf{Z})$,

$$\sum_{i\in\mathbf{Z}} f(i) Z_{T+t}(i) = \sum_{i\in\mathbf{Z}} \big(T_{T+t} f\big)(i) + \int_0^{T+t} \big(T_{T+t-s} f\big)(i) \sqrt{\mu_*(2-\mu_*)u_*}\, \mathrm{d}W_i(s).$$

To recover our expression for the covariance of the invariant measure, note that $T_T Z_0 \to 0$ as $T \to \infty$ and that

$$\mathbf{E}\Big[\int_0^{T+t} T_{T+t-s} f\, \mathrm{d}W_s \int_0^T T_{T-s} g\, \mathrm{d}W_s\Big] = \mathbf{E}\Big[\int_0^T T_{T+t-s} T_t f\, \mathrm{d}W_s \int_0^T T_{T-s} g\, \mathrm{d}W_s\Big]$$
$$= \mathbf{E}\Big[\int_0^T T_{T+s} f\, \mathrm{d}W_{T-s} \int_0^T T_s g\, \mathrm{d}W_{T-s}\Big] = \mathbf{E}\Big[\int_0^T T_{T+s} f\, \mathrm{d}\widehat{W}_s \int_0^T T_s g\, \mathrm{d}\widehat{W}_s\Big],$$

where $(\widehat{W}_s)_{0\le s\le T}$ is a collection of independent Brownian motions.

We now spell out the proof in detail in the case $\Phi'_* > 0$, when the semigroup T_t takes a more familiar form. First, we follow Walsh [134] in writing down a 'Green function representation' of Z_{t+T} and Z_T.

For $\phi = (\phi(x))_{x\in\mathbf{Z}} \in \ell_1$ denote by

$$\langle \phi, Z\rangle := \sum_{x\in\mathbf{Z}} \phi(x) Z(x).$$

Let $\phi : \mathbf{R}_+ \times \mathbf{Z} \to \mathbf{R}$ given by $(t,x) \mapsto \phi_t(x)$ be differentiable in t. By Itô's formula applied to the map $(t,\zeta) \mapsto \phi_t(x)\zeta$

$$\phi_t(x) Z_t(x) = \phi_0(x) Z_0(x) + \int_0^t \partial_s \phi_s(x) Z_s(x)\, \mathrm{d}s$$
$$+ \int_0^t \phi_s(x)\big(u_* g'_* Z_s(x)\, \mathrm{d}s + \frac{m}{2} r_* \Phi'_* \Delta_1 Z_s(x)\big)\, \mathrm{d}s$$
$$+ \int_0^t \phi_s(x) \sqrt{\mu_*(2-\mu_*)u_*}\, \mathrm{d}B_s(x).$$

Summing over $x \in \mathbf{Z}$ and applying summation by parts,

$$\langle \phi_t, Z_t \rangle = \langle \phi_0, Z_0 \rangle + \int_0^t \langle \partial_s \phi_s + u_* g'_* \phi_s + \frac{m}{2} r_* \Phi'_* \Delta_1 \phi_s, Z_s \rangle + \sqrt{\mu_*(2-\mu_*)u_*} \sum_{x \in \mathbf{Z}} \int_0^t \phi_s(x)\, \mathrm{d}B_s(x). \quad (1.5.16)$$

Let $p_t(x)$ be the transition semigroup of a nearest neighbour random walk on $\mathbf{Z}$ jumping at unit rate, started from the origin. Then $p_t(x)$ is the solution to

$$\begin{cases} p_0(x) = \mathbf{1}\{x = 0\}, \\ \partial_t p_t(x) = \dfrac{1}{2} \Delta_1 p_t(x). \end{cases}$$

Thus, defining

$$\phi_s(y) := e^{u_* g'_* (t-s)} p_{m r_* \Phi'_* (t-s)}(y - x),$$

$\phi_s(x)$ is the solution to

$$\begin{cases} \phi_t(y) = \mathbf{1}\{x = y\}, \\ \partial_s \phi_s + u_* g'_* \phi_s + \dfrac{m}{2} r_* \Phi'_* \Delta_1 \phi_s = 0. \end{cases}$$

Inserting this particular test function into the expression (1.5.16),

$$Z_t(x) = e^{u_* g'_* t} \sum_{y \in \mathbf{Z}} p_{m r_* \Phi'_* t}(y - x) Z_0(y) + \sum_{y \in \mathbf{Z}} \sqrt{\mu_*(2-\mu_*)u_*} \int_0^t e^{u_* g'_* (t-s)} p_{m r_* \Phi'_* (t-s)}(y - x)\, \mathrm{d}B_s(y). \quad (1.5.17)$$

For any $t_0 \in \mathbf{R}$, by stationarity we have that $\rho_t(x) = \mathbf{E}[Z_{t_0+t}(x) Z_{t_0}(0)]$. Then by the

Green function representation (1.5.17) and Itô's isometry,

$$
\begin{aligned}
&\rho_t(x)\\
&= \mathbf{E}[Z_{t_0+t}(x)Z_{t_0}(0)]\\
&= e^{u_*g'_*(2t_0+t)}\sum_{y,y'\in\mathbf{Z}} p_{mr_*\Phi'_*(t_0+t)}(y-x)\,p_{mr_*\Phi'_*t_0}(y')\mathbf{E}[Z_0(y)Z_0(y')]\\
&+ \mu_*(2-\mu_*)u_*\mathbf{E}\Bigg[\sum_{y,y'\in\mathbf{Z}}\int_0^{t_0+t} e^{u_*g'_*(t_0+t-s)}\,p_{mr_*\Phi'_*(t_0+t-s)}(y-x)\,\mathrm{d}B_s(y)\\
&\qquad\qquad\times\int_0^{t_0} e^{u_*g'_*(t_0-s)}\,p_{mr_*\Phi'_*(t_0-s)}(y')\,\mathrm{d}B_s(y')\Bigg]\\
&= e^{u_*g'_*(2t_0+t)}\sum_{y,y'\in\mathbf{Z}} p_{mr_*\Phi'_*(t_0+t)}(y-x)\,p_{mr_*\Phi'_*t_0}(y')\mathbf{E}[Z_0(y)Z_0(y')]\\
&+ \mu_*(2-\mu_*)u_*\sum_{y\in\mathbf{Z}}\int_0^{t_0} e^{u_*g'_*(2t_0+t-2s)}\,p_{mr_*\Phi'_*(t_0+t-s)}(y-x)\,p_{mr_*\Phi'_*(t_0-s)}(y)\,\mathrm{d}s\\
&= e^{u_*g'_*(2t_0+t)}\sum_{y,y'\in\mathbf{Z}} p_{mr_*\Phi'_*(t_0+t)}(y-x)\,p_{mr_*\Phi'_*t_0}(y')\mathbf{E}[Z_0(y)Z_0(y')]\\
&+ \mu_*(2-\mu_*)u_*\sum_{y\in\mathbf{Z}}\int_0^{t_0} e^{u_*g'_*(t+2s)}\,p_{mr_*\Phi'_*(t+s)}(y-x)\,p_{mr_*\Phi'_*s}(y)\,\mathrm{d}s\\
&= e^{u_*g'_*(2t_0+t)}\sum_{y,y'\in\mathbf{Z}} p_{mr_*\Phi'_*(t_0+t)}(y-x)\,p_{mr_*\Phi'_*t_0}(y')\mathbf{E}[Z_0(y)Z_0(y')]\\
&+ \mu_*(2-\mu_*)u_*\int_0^{t_0} e^{u_*g'_*(t+2s)}\,p_{mr_*\Phi'_*(t+2s)}(x)\,\mathrm{d}s,
\end{aligned}
$$

where the last line follows from the Chapman–Kolmogorov equations. Taking $t_0 \to \infty$ the result follows. □

1.5.5. The effect of small fluctuations on ancestral lineages. Theorem 1.2.10 has a specific form in this regime.

THEOREM 1.5.11. *Fix $0 < c < u_*$ and define the stopping time $T^{n,c} := \inf\{\underline{t} \geq 0 : X^n_{\underline{t}}(L^n_{\underline{t}}) < c\}$, the first time at which the population density at the site of the lineage falls below c. The process $(X^n, L^n_{\cdot\wedge T^{n,c}}) \Rightarrow (u_*, L)$ in $D([0,T], \mathbf{R}_+^{\mathbf{Z}} \times \mathbf{Z})$ as $n \to \infty$, where L is the unique process such that $L_0 = 0$ and*

$$
f(L_t) - f(L_0) - \int_0^t \frac{m}{2} r_*\gamma_*\Delta_1 f(L_s)\,\mathrm{d}s
$$

is a martingale for each $f \in B(\mathbf{Z})$.

That is, L is a continuous time simple symmetric random walk jumping at rate $r_*\gamma_*$.

In this short section, we make some preliminary remarks about the effect of the small fluctuations captured by Theorem 1.2.14 have on this motion.

Recall that the jump rate of the ancestral lineage is

$$\frac{m}{2}\frac{r(X_{\tilde{t}}(x))\Phi(X_{\tilde{t}}(x+1)) + r(X_{\tilde{t}}(x))\Phi(X_{\tilde{t}}(x-1))}{X_{\tilde{t}}(x)},$$

and that if the lineage is currently at x, conditional on a jump taking place at time t, it is to $x+1$ with probability

$$\frac{\Phi(X_{\tilde{t}}(\omega, x+1))}{\sum_{y=x\pm1}\Phi(X_{\tilde{t}}(\omega, y))}. \tag{1.5.18}$$

Let us first examine (1.5.18) when $X_{\tilde{t}}$ is close to the deterministic u_*. Ignoring higher order corrections, write $X_{\tilde{t}} = u_* + \varepsilon Z_t$, where to save typing we have written $\varepsilon = \sqrt{\delta_n}$. Performing a Taylor expansion of Φ about u_* we find

$$\frac{\Phi\big(u_* + \varepsilon Z_t(x+1)\big)}{\Phi\big(u_* + \varepsilon Z_t(x+1)\big) + \Phi\big(u_* + \varepsilon Z_t(x-1)\big)} = \frac{1}{2} + \frac{\varepsilon}{4}\big(Z_t(x+1) - Z_t(x-1)\big)\frac{\Phi'_*}{\Phi_*} + O(\varepsilon^2).$$

A similar calculation shows that the overall rate of jumps is

$$\frac{m r_* \Phi_*}{u_*}\Big\{1 + \varepsilon\Big(\frac{(u_* r'_* - r_*)}{u_* r_*}Z(x) + \frac{\Phi'_*}{2\Phi_*}\big(Z(x-1) + Z(x+1)\big)\Big)\Big\} = \frac{m r_* \Phi_*}{u_*}\Big\{1 + \varepsilon\Big(\frac{(r\gamma)'_*}{(r\gamma)_*}Z(x) + \frac{\Phi'_*}{\Phi_*}\Delta_1 Z(x)\Big)\Big\}.$$

We see two key quantities, $(\log\Phi)'_*$ and $(\log(r\gamma)'_*$. If $\Phi'_* > 0$ is positive, ancestral lineages jump to demes of higher population density, and if $(r\gamma)'_* < 0$, once at a local maximum, they tend to stay there, in the sense that they jump more slowly than average. On the other hand if $\Phi'_* < 0$, the lineage tends to jump to regions of lower population density, and will loiter in local minima if $(r\gamma)'_* < 0$.

1.6. Appendix.

1.6.1. Burkholder's predictable square function inequality.

LEMMA 1.6.1 (Burkholder [18],Theorem 21.1). *Let $p > 1$ and suppose $(V_k)_{k\in\mathbf{N}}$ is a discrete parameter martingale with respect to the filtration $\mathcal{F}_k$ and that $V_0 = 0$. Define $\langle V\rangle_n := \sum_{k=0}^{n-1}\mathbf{E}[(V_{k+1} - V_k)^2|\mathcal{F}_k]$. Then there exists a constant $C(p) < \infty$ such that*

$$\mathbf{E}[\sup_{0\le k\le n}|V_k|^p] \le C(p)\big(\mathbf{E}[|\langle V_n\rangle|^{p/2}] + \mathbf{E}[\sup_{0\le k\le n-1}|V_{k+1} - V_k|^p]\big).$$

1.6.2. A discrete local limit theorem.

THEOREM 1.6.2 (Durrett [31], Theorem 5.2). *Let $X_1, X_2, \dots$ be i.i.d. copies of an integer-valued random variable χ of mean μ and variance σ^2. Suppose furthermore that there is no infinite subprogression $a + q\mathbf{Z}$ of $\mathbf{Z}$ with $q > 1$ for which X takes values almost surely in $a + q\mathbf{Z}$. Then one has*

$$\mathbf{P}[S_n = m] = \frac{1}{\sqrt{2\pi n}\sigma} \exp\Big(- (m - n\mu)^2 2n\sigma^2 \Big) + o(1/\sqrt{n})$$

for all $n \geq 1$ and all integers m, where the error term $o(1/\sqrt{n})$ is uniform in m. In other words,we have

$$\sup_{m \in \mathbf{Z}} \sqrt{n} \left| \mathbf{P}[S_n = m] - \frac{1}{\sqrt{2\pi n}\sigma} \exp\Big(- (m - n\mu)^2 2n\sigma^2 \Big) \right| \to 0$$

as $n \to \infty$.

1.6.3. Spectra of Laurent operators. Let $\{a_n\}_{n \in \mathbf{Z}} \subset \mathbf{C}$ be a sequence of complex numbers and denote by A the bi-infinite matrix with values $a_{ij} = a_{i-j}$, $i, j \in \mathbf{Z}$, ie defined such that the matrix A is constant along its diagonals. The matrix A acting on $\ell_2(\mathbf{Z})$ is called a *Laurent operator*. A theorem of Toeplitz [125] tells us when a Laurent operator is bounded and provides its spectrum. The version here is a combination of Theorem 1.1 and Theorem 1.2 of Böttcher and Silbermann [15].

THEOREM 1.6.3. *Let $\mathbf{T} := \{z \in \mathbf{C} : |z| = 1\}$ denote the unit circle in the complex plane. The matrix A is a bounded operator on $\ell_2(\mathbf{Z})$ if and only if there exists a function $f \in L^\infty(\mathbf{T})$ such that $\{a_n\}_{n \in \mathbf{Z}}$ is a sequence of Fourier coefficients for f; that is,*

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{-in\theta} \, \mathrm{d}\theta,$$

for all $n \in \mathbf{Z}$. In this case, the spectrum of A is given by the essential range of f; that is,

$$\mathrm{spec}(A) = \{z \in \mathbf{C} : |\{\theta \in [0, 2\pi) : |f(e^{i\theta}) - z| < \varepsilon\}| > 0 \text{ for all } \varepsilon > 0\}.$$

CHAPTER 2

GENE FLOW IN PATCHY HABITATS

Joint work with
ALISON M. ETHERIDGE AND PETER L. RALPH

The demographic details of an individual-based stochastic population model determine not only the time evolution of population sizes, but also the dynamics of *lineages* – the description of where the ancestors of modern individuals lived at times in the past. Here, we assume that there are only two types of habitat, *bad* and *good*, and that population sizes in good habitats are so large that, given the opportunity, lineages will always move from bad to good habitats. Lineages traced back through time can be explicitly described and we highlight the role of the spatial and temporal autocorrelation structure in determining long-term behaviour of lineage movement. The typical displacement between an individual and their ancestor at some time in the past may differ substantially from that predicted based on the typical displacement between parent and child.

2.1. Introduction.

The dynamics of a population are determined by the demography of the individuals that make up the population: fecundities, mortalities, movement rates, and so forth. Stochastic population models can lead to complex behaviour, especially across continuous geography, and a great deal of work has focused on describing them. Inferring these demographic parameters of populations can be of substantial practical importance, as for instance when working to conserve an endangered species or to control an invasive species or disease vector. Although reproduction is a fundamental way by which populations grow, the genealogical connections between parents and offspring can be ignored if one is only interested in population size changes. However, these relationships clearly contain important information about population demography, and so there is substantial interest in using the patterns of relatedness inferred from genetic data to learn about the demographic dynamics and history of real populations. For instance, genome sequencing has provided surprising

information about the history of our own species both in the recent and distance past (see, eg, [89]). However, the relationship between population demographic dynamics and the resulting patterns of genealogical relationships have only been described in the simplest possible models. In this paper, we study how genealogical inheritance, traced back through time, is determined by the (forwards-time) process of population size changes. The models we study are greatly simplified, but are chosen to highlight the important aspects of this relationship.

If we begin with a single individual from some population alive today and choose a particular bit of her genome, then, looking back through history, we can imagine recording the geographic location of the ancestor from whom that individual inherited that bit of genome. This spatial motion, as a function of amount of time in the past, we call a *lineage*. For concreteness, suppose we follow inheritance of the chloroplast, and thus the maternal lineage, in a population of plants. Suppose we know the birth location of a given ancestor at a particular time. Roughly speaking, the location of the next ancestor might be chosen from among the possible ancestors, weighted by the reproductive output of those ancestors that established at that location at that time. On a broad scale, therefore, lineages when traced back through time should tend to follow regions of high population density, if the regions have correspondingly high total reproductive output.

The *dispersal distance*, σ, characterizes the typical distance between parent and offspring (concretely, it is the standard deviation of the displacement between the two along an arbitrary axis). Imagining a lineage as a random walk, one might guess that in a homogeneous population, the distance from a given individual to their ancestor from T generations ago is proportional to $\sigma\sqrt{T}$. Although this may be close in some cases, this can be dramatically wrong in others. The *effective dispersal distance* is the corresponding quantity seen by averaging back along a lineage – if D_t is the distance from an individual to their ancestor living t generations ago, then $\mathrm{var}[D_t] \approx t\sigma_e^2$. For instance, if there is significant competition between siblings, individuals that disperse further from their parents may be more successful. Since these long dispersers are more successful, they contribute more to the modern-day gene pool, and hence are more likely to be seen when looking back along a lineage – so, in this case, we would expect $\sigma_e > \sigma$. Conversely, if habitat quality is highly heterogeneous, so that longer-distance dispersers are likely to end up in an inhospitable region, then σ_e may be substantially less than σ. More generally, spatial and temporal autocorrelation in habitat quality can lead to correlations between dispersal ability and long-term fitness, and therefore selection on dispersal ability itself [112].

In this paper, we study the movement of linages in populations whose sizes are entirely determined by exogenous factors, eg transient, patchy resources. By varying the spatial and temporal autocorrelation of habitat quality in several models, we can highlight how these determine the long-term behaviour of lineage movement, and hence spatial patterns of re-

latedness and gene flow. For a concrete situation, imagine that we are tracking populations of a soil microbe across a grassland, and the microbe is tightly associated with a particular type of shrub. A low density of microbes are found in the soil away from shrubs, but much higher densities live in the root zones of patches of shrubs that expand, contract, appear, and disappear over time. New patches of shrubs initiate only rarely, but when they do, the lucky microbes living nearby to the location where the patch originated reproduce rapidly, and since the area surrounding the patch produces very few offspring, the microbes found in that patch are descendants of those initially lucky individuals for as long as the patch exists. This means that as one traces back the lineage of a microbe found in a patch today, the lineage must remain in that patch until its origination – and so, the motion of lineages is strongly constrained by the dynamics of the patches.

In the next section we describe a general discrete-time model for the forward-in-time dynamics of a population. In this model, offspring are dispersed from their parents at birth. We then assume that the lineage, as described above, can be described as a Markov process conditional on the history of population sizes and give its transition probabilities.

We then proceed to specialise to our case of interest, when the forward-in-time dynamics of the population are entirely determined by exogenous factors, eg transient, patchy resources. We describe a model where there are just two kinds of habitat, labelled *bad* and *good* and assume that, when given the choice, population sizes in good habitats are so large compared to those in bad habitats that lineages always move to locations of good habitat (chosen from those available) moving into the past. Since the structure of the habitat is random, we are studying a random walk in a random environment, a well-studied object.

In order to assess the impact on spatial and temporal correlations in the habitat quality on the effective dispersal of the lineages we consider a series of models for the dynamics of the patch quality. Central to all of them will be a fixed stationary distribution controlling the degree of spatial correlations. Given this we consider three further scenarios. First, we consider the case that the habitats are resampled in every discrete generation of the population — minimising the temporal correlations in habitat quality. We find that lineages can move faster or slower than one which moves freely in a homogeneous environment. Secondly, as a toy example, we consider a model where the environment is resampled every two generations — slightly increasing the degree of temporal correlation in habitat quality. Again, lineages may move faster or slower than one which moves freely in a homogeneous environment.

We subsequently define a model where the habitat quality is allowed to evolve locally in time, ie in each generation there is a small probability that the quality of the habitat an single location flips from good to bad, or *vice versa*. This probability is allowed to depend on the quality of the habitat within a local neighbourhood, allowing for interesting spatial correlations. Passing to a continuous-time limit, the dynamics of the habitat quality

evolve according to a Markov process, essentially an interacting particle system [90], with a stationary distribution given by a random cluster measure [48].

In order to infer the effective dispersal distance of the lineages in this setting, a thorough understanding of the geometry of patches of good habitat is needed. In Section 2.3 under diffusive rescaling, we describe both how new patches appear in the environment and how these patches subsequently evolve. The proofs of all statements are the content of Sections 2.4-2.8. Finally, in Section 2.9 we conjecture the value of the effective dispersal distance in this setting; unfortunately it has not been possible to find a proof, but the key steps to proving such a result are outlined.

2.2. The model.

Suppose that we study a species of plant that lives in discrete populations, and that every year, the entire population reproduces and then dies. To make things even simpler, we suppose that every individual produces a large number of seeds (the same number regardless of location) and that we only follow ancestry to the seed parent (ignoring pollen). Then, if the proportion of the seeds of a plant at location x that arrive at location y is $m(x, y)$, and the number of individuals at location x, k generations ago, is $N(k, x)$, then the total number of seeds from x that arrived at y in the subsequent generation, ie $k-1$ generations ago, is proportional to $N(k, x)m(x, y)$. If all these seeds are roughly equivalent, so that the subsequent generation was composed of a random choice of all available seeds, then the chance that a given plant living at y some $k-1$ generations ago grew from a seed originating in x in the previous generation, is

$$\frac{N(k, x)m(x, y)}{\sum_z N(k, z)m(z, y)}, \tag{2.2.1}$$

where the sum is over all locations.

In this paper, we then assume that the lineage of an individual can be described as a Markov process given the population sizes. Concretely, if L_k is the location k generations ago of the (maternal) ancestor of a given individual who is today randomly chosen from $L(0)$, then given the population sizes histories $(N(k, z))_{z,k}$, the series of locations $(L_k)_{k\geq 0}$ forms a Markov chain with $\mathbf{P}[L_k = x \mid L_{k-1} = y]$ given by equation (2.2.1). This is a standard assumption [132] working in the regime of 'infinitely' large population sizes, so that conditioning on the ancestral process $(L_k)_{k\geq 0}$ does not change the dynamics of the population, ie a single individual contributes so little to the macroscopic population that the dynamics are unaffected.

2.2.1. Patchy habitats. We will assume that there are only two types of habitat, *bad* and *good*, and that population sizes in good habitats are so large that, given the opportunity,

lineages will always move from bad to good habitats. In the motivating example above, good habitat occurs where the shrub lives.

Our analysis will occur mostly in one spatial dimension, beginning on the discrete lattice, with nearest-neighbour dispersal. We will assume that the habitat evolves according to a number of different Markov processes, each with the same stationary distribution (described below). Taking the environmental dynamics to be reversible and at stationarity, we can then simply track the dynamics of the lineage and the habitat together back into the remote past. In spite of its apparent simplicity, the model allows us to describe a number of important features of the general problem, while retaining some analytical and numerical tractability. Crucially, our model will be defined so that we are able to control the spatial and temporal correlations in the habitat quality. By varying the strength of these spatial and temporal correlations, our aim is to establish how they determine the long-term behaviour of lineage movement, and hence spatial patterns of relatedness and gene flow.

In the setting of equation (2.2.1), we will assume that for each x, k, the population sizes in discrete sites take one of two potential sizes, ie $N(k, x) \in \{B, G\}$, where B and G correspond to the population sizes in bad and good habitats respectively with $B \ll G$. In order to arrive at a model where, given the choice, lineages always move to good habitats, we will pass to the limit $B/G \to 0$ so that, after sensibly rescaling population sizes by a factor of G, we may assume that $B = 0$ and $G = 1$. After passing to this limit, the lineage $(L_k)_{k\geq 0}$ forms a Markov chain with transition probabilities

$$\mathbf{P}[L_k = x \mid L_{k-1} = y] = \begin{cases} \dfrac{m(x,y)}{\sum_z m(z,y)}, & \text{if } \displaystyle\sum_z N(k,z)m(z,y) = 0, \\ \dfrac{N(k,x)m(x,y)}{\sum_z N(k,z)m(z,y)}, & \text{otherwise,} \end{cases} \tag{2.2.2}$$

ie the probability that the location of the maternal ancestor k generations ago conditional on being at y $k-1$ generations ago is x is proportional to $m(x, y)$ when all sites z such that $m(z, y) > 0$ are regions of bad habitat k generations ago, and chosen proportional to $N(k, x)m(x, y)$ otherwise.

Our model will allow for the population sizes at different locations to be correlated, ie $\mathrm{cov}[N(k, x), N(k, y)] > 0$. For the sake of simplicity, we restrict our focus to dimensions $d = 1$, assuming that individuals live in demes indexed by the one-dimensional lattice $\mathbf{Z}$. The spatial distribution of habitat at any particular time can be thought of as a sequence of *patches* of consecutive runs of ones (denoting a region of *good* habitat), separated by runs of at least one zero (denoting a region of *bad* habitat). We will model these by specifying that, passing through space, the corresponding sequence of zeroes and ones forms a stationary and reversible Markov chain whereby each patch has a Geometric(p_1) number of ones, each

space between two patches is composed of a Geometric(p_0) number of zeroes, and that all such are independent. In other words, the probability that the habitat at location $x = 0$ is *good*, ie $N(0, k) = 1$, is $p_0/(p_0 + p_1)$, and the sequence of habitats one encounters as one walks along the lattice (in either direction) is a Markov chain that moves from one to zero with probability p_1 and moves from zero to one with probability p_0. We will also specify the temporal dynamics of the habitat in a way that preserves this distribution.

The habitat is then described by a binary sequence $\omega \in \{0, 1\}^{\mathbf{Z}}$, and we call the probability distribution above π. An equivalent description of π is as a *random cluster measure*, also sometimes known as a Fortuin–Kasteleyn measure. For a comprehensive introduction to random cluster measures, we direct the reader to Grimmett [48]. In order to recast π in this language, we fist work on the bounded subset $[-L, L] \cap \mathbf{Z}$ of the lattice. We define a measure on sequences of possible habitats $(\omega_{-L}, \omega_{-(L-1)}, \ldots, \omega_L)$ subject to the boundary conditions $\omega_{-L} = \omega_L = 0$, and obtain π by passing to the limit $L \to \infty$. Conditioning on the outcome $\omega_{-L} = \omega_L = 0$, a sequence $(\omega_{-(L-1)}, \ldots, \omega_{L-1})$ has probability

$$(1 - p_0)^{N_0}(1 - p_1)^{N_1} \left(\frac{p_0 p_1}{(1 - p_0)(1 - p_1)} \right)^{N_c},$$

where N_0 and N_1 are the number of zeroes and ones, respectively, and N_c is the number of patches of good habitat, ie the number of runs of ones.

Our goal now is to specify temporal evolution of the habitat, introducing varying degrees or temporal correlations. The dynamics of the habitat will evolve according to a Markov process for which π is a reversible stationary distribution; we can then simply track the dynamics of the lineage and the habitat together back into the remote past and establish the effective dispersal distance.

2.2.2. Temporally uncorrelated patchy habitats. The simplest case is when the habitat is simply resampled independently from this distribution every generation, and that offspring are divided evenly between the location of their parent and the locations directly on either side of the parent.

LEMMA 2.2.1. *Suppose that the habitat is resampled independently in each generation, and that* $m(x, x) = m(x, x + 1) = m(x, x - 1) = 1/3$. *The variance of the displacement of an ancestral lineage over a single generation is*

$$\operatorname{var}[L_1] = \frac{2}{3} + \frac{p_0 p_1}{3(p_0 + p_1)} \big(1 - (p_0 + p_1)\big).$$

In particular, lineages move faster than those in a homogeneous environment if and only if $p_0 + p_1 < 1$.

The (elementary) proof is deferred until Section 2.4.

In other words, more positive spatial correlations in the environment cause the walk to move more quickly, and more negative ones (ones more in the negative direction) cause it to move more slowly. The eigenvalues of the transition matrix of the habitats are 1 and $1-(p_0+p_1)$, so the transition corresponds to that from a smooth to an oscillating habitat. At first this result seems a little counterintuitive. However, consider what happens when $p_0 = p_1 = 1$, so that the habitat alternates between good and bad in space; once the habitat at the origin is fixed, everything is determined. Then in each generation, there is probability $1/2$ that the ancestral lineage does not move at all, and probability $1/2$ that it moves to a neighbouring deme. The variance of the movement of a lineage over one generation is then $1/2$, less than the $2/3$ corresponding to the symmetric walk corresponding to m, which chooses each of the three potential ancestral demes with equal probability. Conversely, if p_0 and p_1 are both small, then there is some chance a lineage will find itself near the edge of a large patch of good habitat, thus giving it an extra "push". However, the speed is maximized at $p_0 = p_1 = 1/4$, but is only sped up by a factor of $33/32$: the effect is not large. It is also interesting to note that since the lineage only moves to bad habitat if there is no available good habitat, the proportion of the time a lineage spends in bad habitat is $p_1(1-p_0)^2/(p_0+p_1)$. This is the proportion of future generations who are direct descendants of today's bad habitats, ie proportional to the total long-term fitness of bad habitats. This is despite the fact that in any particular generation, the vast majority of individuals may live in good habitats — but, long-term genetic contributions to the population is mostly determined by who initiates new patches, rather than who lives in them.

2.2.3. A first look at temporally correlated patchy habitats. Of course, there is no reason to suppose that the habitat and the resident population have the same intergeneration time. It is instructive to consider what happens if we suppose that several generations of the population pass during each time step of the habitat. Evidently, if the habitat never changes, then any lineage will become "trapped" in a cluster of good habitat.

It is tedious, but not difficult, to prove the following lemma.

LEMMA 2.2.2. *Suppose that the habitat is resampled every two generations (in between which it is held fixed). The variance of the position of a lineage after two generations is*

$$\operatorname{var}[L_2] = \frac{4}{3} + \frac{p_0 p_1}{18(p_0+p_1)}\big(47 - 77p_0 - 62p_1 + 27p_0p_1 + 28p_0^2 + 7p_1^2\big). \qquad (2.2.3)$$

The term $4/3$ in (2.2.3) is precisely what we would have for the symmetric walk in a homogeneous environment. If p_0 and p_1 are close to one then the second term is always

negative, and the walk slows down in the random environment. If p_0 and p_1 are close to zero, then the walk always goes faster in the random than in the homogeneous environment, but the factor $p_0 p_1$ tells us that the effect is small.

It is not hard to convince oneself that the longer the habitat is held fixed, the smaller p_0 and p_1 must be if ancestral lineages are not to be slowed down compared to simple random walk. We require p_1 to be small in order that a cluster is big enough that the motion of a lineage is not be too constrained before the habitat is resampled; we require p_0 to be small so that we rarely see clusters of favourable environment and so do not become trapped. The longer the habitat is held constant, (where time is measured in numbers of generations of the population), the more the ancestral lineages will be slowed compared to those sampled from a homogeneous population. Of course, holding the habitat constant and then resampling is not particularly natural, and so we now turn to finding a Markovian evolution of the environment.

2.2.4. Continuous-time dynamics. Holding the habitat constant for a fixed number of generations and then (independently) resampling, seems very artificial. Instead we should like it to evolve according to a Markov process. The simplest way to achieve this is to recast our model in the setting of continuous-time.

NOTATION 2.2.3. For $\omega \in \{0,1\}^{\mathbf{Z}}$ and $x \in \mathbf{Z}$ we will denote by $\omega^x \in \{0,1\}^{\mathbf{Z}}$ the configuration with $\omega^x_y = \omega_y$ for all $y \neq x$ and $\omega^x_x = 1 - \omega_x$.

Letting $h > 0$ be a scaling parameter, we will define a sequence of $\{0,1\}^{\mathbf{Z}}$-valued processes $(N^h(k,\cdot))_{k\geq 0}$ indexed by h as follows. We specify that, when conditioning on the current state of the environment, the state of the habitat at location $x \in \mathbf{Z}$ one further generation into the past switches from a bad to a good habitat, or *vice versa*, independently of all other locations. Furthermore, we shall suppose that there exists a $\lambda : \mathbf{Z} \times \{0,1\}^{\mathbf{Z}} \to \mathbf{R}_+$ such that for $h > 0$, the probability that the habitat at location x switches state in passing to the previous generation satisfies

$$\mathbf{P}[N^h(k+1,x) = 1 - \omega_x \mid N^h(k,\cdot) = \omega] = \begin{cases} h\lambda(x,\omega) + o(h), & |x| \leq h^{-1/2}, \\ 0, & |x| > h^{-1/2}, \end{cases} \tag{2.2.4}$$

as $h \to 0$, for all $k \geq 0$, $x \in \mathbf{Z}$ and $\omega \in \{0,1\}^{\mathbf{Z}}$. The term $o(h)$ is assumed uniform in x and, moreover, these transition probabilities are supposed homogeneous in k so that the Markov evolution is time-homogeneous. The condition that the probabilities are set to 0 for all $|x| > h^{-1/2}$ is technical and allows us to control the number of flips in any one discrete generation when passing to the limit. Since any biologically relevant applications occur in a finite environment, we consider this assumption to be mild.

ASSUMPTION 2.2.4. We will assume that the function $\lambda : \mathbf{Z} \times \{0,1\}^{\mathbf{Z}} \to \mathbf{R}_+$ is translationally invariant and finite-range in the following sense. There exists an $N < \infty$ such that for all $x \in \mathbf{Z}$, $\lambda(x,\omega)$ is a function of values in the set $\{\omega_y : |x-y| \leq N\}$ only. In particular, λ may take at most 2^8 values and thus there exists a $\bar{\lambda} < \infty$ such that $\sup_{x\in\mathbf{Z}} \sup_{\omega\in\{0,1\}^{\mathbf{Z}}} \lambda(x,\omega) < \bar{\lambda}$.

Our aim is to rescale time by a factor of h, so that one unit of continuous time corresponds to $1/h$ discrete generations and, taking $h \to 0$, to construct a continuous-time limit.

REMARK 2.2.5. At first sight, it may appear that it would have been more natural to directly define the continuous-time analogue of our process to be the $\{0,1\}^{\mathbf{Z}}$-valued process whose value at each site flips from a 0 to 1 or *vice versa* at the times of an independent Poisson process with parameter given by some some function of the states of the habitat within some local neighbourhood. Whilst this is the limit we will obtain after a rescaling of time in the sequence of environment process N^h, there is ambiguity as to how the dynamics of the lineages at the boundaries of patches should be defined. Whilst it is natural to set that the lineages will jump at the times of a (further independent) Poisson process of constant rate, with transition probabilities defined by equation (2.2.2), the question of what happens to a lineage when the state of the environment at its current deme flips remains to be addressed. In particular, when the habitat at the current location of the lineage flips from a *good* habitat to a *bad* one, should we *a priori* define that the lineage will instantaneously move to a neighbouring region of *good* habitat, if one is available? By defining our continuous-time model as a limit of our original process in this way, we avoid making this arbitrary decision and allow for direct comparison with the simple examples of Lemma 2.2.1 and Lemma 2.2.2.

We will define a sequence of rescaled processes as follows. For all $t \in [0,T] \subset \mathbf{R}_+$, define the sequence of $\{0,1\}^{\mathbf{Z}}$-valued processes $(\Xi_t^h)_{t\in[0,T]}$ indexed by h by

$$\Xi_t^h(x) := N^h(\lfloor t/h \rfloor, x).$$

Since the transition probabilities only depend on a finite neighbourhood of the site in question, it will be useful to define $V_t^h(x,k) := (\Xi_t^h(x+y) : |y| \leq k) \in \{0,1\}^{2k+1}$, the view of the environment at time t as seen from deme x, extending up to and including its $2k$ nearest neighbours.

We proceed now to define the dynamics of the lineage under the same rescaling of time. Again, for each $h > 0$, we define a Markov chain $(L_k^h)_{k\geq 0}$, corresponding to the passage of the lineage through the population $(N^h(k,\cdot))_{k\geq 0}$, with transition probabilities as defined in equation (2.2.2) and dispersal probabilities $m^h(x,y)$ now indexed by h. For simplicity,

we will assume that we have nearest-neighbour seed dispersal and that for some constant $\sigma > 0$,

$$m^h(x, x+1) = m^h(x, x-1) = \frac{\sigma^2 h}{2} + o(h),$$

and $m^h(x, y) = 0$ for all $y \neq x-1, x, x+1$. The little-o term here, $o(h)$, is assumed, for simplicity, to be uniform in x. In fact, any sequence of dispersal probabilities satisfying $\mathbf{E}[\Delta L_k] = 0$ and $\mathbf{E}[(\Delta L_k)^2] = \sigma^2 h + o(h)$ will suffice. For all $t \in [0, T] \subset \mathbf{R}_+$, we subsequently define a sequence of rescaled lineage processes $(Y_t^h)_{t\in[0,T]}$ indexed by h by

$$Y_t^h := L^h_{\lfloor t/h \rfloor}.$$

Unfortunately, the sequence of processes $\{Y^h\}$ is not tight with respect to the Skorokhod topology on $D([0, T], \{0, 1\}^{\mathbf{Z}} \times \mathbf{Z})$. Taking a closer look at the dynamics this is to be expected since if the lineage approaches the boundary of a region of zeroes, then (with asymptotically high probability) the lineage jumps to the neighbouring region of ones at the time of the next generation into the past. In particular, it is possible that two jumps of macroscopic size occur within an asymptotically small window of time. Such behaviour is forbidden if the sequence of processes is to be relatively compact, at least in the Skorokhod topology where the behaviour at discontinuities must satisfy certain regularity conditions (see eg Ethier and Kurtz [40], Lemma 3.8.1). Instead we take the following approach. We define a process X^h whose evolution is coupled to Y^h. The dynamics of X^h will be judiciously chosen so to avoid the behaviour at the boundary of a patch which lead to non-tightness of the original process.

The process X^h will be defined to have the same transition rates over discrete generations as Y^h, as defined in equation (2.2.2), except until a time when $V^h_{t+h}(X^h_t, 1) = (0, 0, 0)$ but $V^h_{t+h}(X^h_t, 2)$ contains a one, ie when looking one further discrete generation into the past, the habitats of its current location and its nearest neighbours is *bad* but there is at least one *good* habitat looking one further step beyond those. It is at precisely these times where our original process Y^h has the potential to go awry, since if it now jumps to one of its nearest neighbours whose neighbour contains a one, then (with asymptotically high probability, as $h \to 0$) the lineage then jumps to its new neighbour, containing a one, in the next discrete generation. At such a time, X^h will jump left or right with the same probabilities defined in equation (2.2.2), however, if this jump places it adjacent to a *good* site then it must jump one step further, placing it in the *good* deme, in the same discrete generation. For example, for $k \neq 0$,

$$\mathbf{P}[X^h_{t+h} = x + k \mid X^h_t = x, V^h_{t+h}(x, 2) = (0, 0, 0, 0, 0)]$$

$$= \begin{cases} \dfrac{\sigma^2 h}{2} + o(h), & k \in \{-1, 1\}, \\ 0, & k \notin \{-1, 1\}, \end{cases}$$

just as the original process Y^h, but

$$\mathbf{P}[X^h_{t+h} = x + k \mid X^h_t = x, V^h_{t+h}(x, 2) = (0, 0, 0, 0, 1)]$$
$$= \begin{cases} \dfrac{\sigma^2 h}{2} + o(h), & k \in \{-1, 2\}, \\ 0, & k \notin \{-1, 2\}. \end{cases}$$

We will choose a convenient coupling of X^h and Y^h, preserving their marginal distributions. We will first set $X^h_0 = Y^h_0$. The two processes are initially *coalesced* and follow the path of X^h until a time when $V^h_{t+h}(X^h_t, 1) = (0, 0, 0)$ but $V^h_{t+h}(X^h_t, 2)$ contains a one. At such a time, if Y^h jumps to the right (respectively left) then so must X^h. However, if the new right (respectively left) neighbour is a good habitat then the two processes *decouple* and X^h jumps to this deme in the same generation, leaving Y^h behind. For example, for $(k, j) \neq (0, 0)$,

$$\mathbf{P}[X^h_{t+h} = x + k, Y^h_{t+h} = x + j \mid X^h_t = Y^h_t = x, V^h_{t+h}(x, 2) = (0, 0, 0, 0, 1)]$$
$$= \begin{cases} \dfrac{\sigma^2 h}{2} + o(h), & (k, j) \in \{(2, 1), (-1, -1)\}, \\ 0, & (k, j) \notin \{(2, 1), (-1, -1)\}. \end{cases}$$

The two processes then evolve independently of one another, according to their respective transition rates, until such a time that that they meet in the same deme. When they meet the two particles coalesce and follow the same trajectories until a time again when $V^h_{t+h}(X^h_t, 1) = (0, 0, 0)$ but $V^h_{t+h}(X^h_t, 2)$ contains a one and they transition as before.

The validity in working with such a modification X is justified by the following result.

PROPOSITION 2.2.6. *For every $\varepsilon > 0$,*

$$\lim_{h \to 0} \mathbf{P}\left[\int_0^T 1\{X^h_t \neq Y^h_t\} \, \mathrm{d}t > \varepsilon \right] = 0.$$

This result is unsurprising since we expect that the coalesced process should encounter the external boundary of a patch of ones only a finite number of times before time T and, once they decouple, the original process Y^h will jump to the neighbouring good deme (and the location of X^h) in the next discrete generation with probability $1 - o(h)$. The full proof is deferred until Section 2.5.

Furthermore, unlike (Ξ^h, Y^h), the sequence of processes (Ξ^h, X^h) is convergent.

THEOREM 2.2.7. *Suppose that* $\lambda : \mathbf{Z} \times \{0,1\}^{\mathbf{Z}} \to \mathbf{R}_+$ *satisfies Assumption 2.2.4. Let the space* $\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}$ *be endowed with the product topology and suppose that the sequence of initial conditions satisfy* $(\Xi_0^h, X_0^h) \Rightarrow (\zeta, 0)$ *as* $h \to 0$ *where if* $\zeta_0 = 0$ *then* $\zeta_1 = \zeta_{-1} = 0$. *Then* $(\Xi^h, X^h) \Rightarrow (\Xi, X)$ *in* $D([0,T], \{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$ *as* $h \to 0$, *where* (Ξ, X) *is the Markov process satisfying* $\mathbf{P}[(\Xi_0, X_0) = (\zeta, z)] = 1$ *and has generator* $A : C(\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}) \to C(\{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$ *defined by*

$$Af(\omega, x) = \sum_{\xi \in \{0,1\}^{\mathbf{Z}}} \sum_{y \in \mathbf{Z}} c((\omega, x), (\xi, y))\{f(\xi, y) - f(\omega, x)\} \tag{2.2.5}$$

for all $f \in C_0(\{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$, *defined to be the set of continuous functions which depend only on finitely many coordinates of* ω *and constant for all* $|x|$ *sufficiently large. The rates* $c((\omega, x), (\xi, y))$ *are defined as follows.*

If $\omega_x = 1$,

$$c((\omega, x), (\omega^y, x)) = \lambda(y, \omega), \qquad y \neq x,$$

$$c((\omega, x), (\omega^x, x)) = \lambda(x, \omega)(1 - \omega_{x-1})(1 - \omega_{x+1}),$$

$$c((\omega, x), (\omega^x, x+1)) = \lambda(x, \omega)\Big[(1 - \omega_{x-1})\omega_{x+1} + \frac{1}{2}\omega_{x-1}\omega_{x+1}\Big],$$

$$c((\omega, x), (\omega^x, x-1)) = \lambda(x, \omega)\Big[(1 - \omega_{x+1})\omega_{x-1} + \frac{1}{2}\omega_{x-1}\omega_{x+1}\Big],$$

$$c((\omega, x), (\omega, x+1)) = \frac{\sigma^2}{2}\omega_{x+1},$$

$$c((\omega, x), (\omega, x-1)) = \frac{\sigma^2}{2}\omega_{x-1},$$

and $c((\omega, x), (\xi, y)) = 0$ *for all other pairs* (ξ, y).

If $\omega_{x-1} = \omega_x = \omega_{x+1} = 0$, *then*

$$c((\omega, x), (\omega^y, x)) = \lambda(y, \omega), \qquad y \neq x-1, x+1,$$

$$c((\omega, x), (\omega^{x+1}, x+1)) = \lambda(x+1, \omega),$$

$$c((\omega, x), (\omega^{x-1}, x-1)) = \lambda(x-1, \omega),$$

$$c((\omega, x), (\omega, x+1)) = \frac{\sigma^2}{2}(1 - \omega_{x+2}),$$

$$c((\omega, x), (\omega, x+2)) = \frac{\sigma^2}{2}\omega_{x+2},$$

$$c((\omega, x), (\omega, x-1)) = \frac{\sigma^2}{2}(1 - \omega_{x-2}),$$

$$c((\omega, x), (\omega, x-2)) = \frac{\sigma^2}{2}\omega_{x-2}.$$

and $c((\omega, x), (\xi, y)) = 0$ *for all other pairs* (ξ, y). *The rates* $c((\omega, x), (\xi, y))$ *are undefined for any pair* (ω, x) *such that* $\omega_x = 0$ *and either* $\omega_{x+1} = 1$ *or* $\omega_{x-1} = 1$, *but these states are inaccessible.*

Again, the proof is deferred until Section 2.5.

REMARK 2.2.8. The requirement on the limiting initial condition, that if $(\Xi_0, X_0) = (\zeta, z)$ and $\zeta_z = 0$ then $\zeta_{z+1} = \zeta_{z-1} = 0$, is a requirement that the process does not start in a state that is otherwise inaccessible according to the above rates. The condition corresponds to forbidding the lineage to begin in a *bad* habitat when there are *good* habitats within its neighbourhood (which it would necessarily instantaneously jump to).

In other words, the sequence of processes $(\Xi^h)_{h\geq 0}$ approximates a continuous-time Markov chain Ξ taking values in $\{0, 1\}^{\mathbf{Z}}$ such that the state of the habitat at each site $x \in \mathbf{Z}$ flips from *bad* to *good*, or *vice versa*, at instantaneous rate $\lambda(x, \omega)$ when the state of the environment is ω. All other transitions are forbidden.

The dynamics of the limiting lineage process X evolve as follows. If the lineage currently inhabits a *good* habitat x, then it attempts to jump one site to the right, to $x + 1$, at instantaneous rate $\sigma^2/2$; the jump is permitted only if the habitat at $x + 1$ is good. Similarly, the lineage attempts to jump one step to the left, at instantaneous rate $\sigma^2/2$ and the jump is permitted only if the habitat at $x - 1$ is good. The lineage otherwise remains at x until such a time when $\Xi_{t-}(x) = 1$ but $\Xi_t(x) = 0$, ie the habitat at x switches from good to bad. At this time, if such a location exists, the lineage resamples its position by choosing uniformly from all neighbouring sites of good habitat; otherwise it remains at x. For example, if $V_{t-}(x, 1) = (0, 1, 1)$ but $V_t(x) = (0, 0, 1)$, then a lineage at x immediately before time t, will jump to $x + 1$ at time t. On the other hand, if $V_{t-}(x, 1) = (0, 1, 0)$ but $V_t(x) = (0, 0, 0)$, then the lineage remains at x.

If at time t the lineage is at x and $V_t(x, 2) = (0, 0, 0, 0, 0)$, then it jumps one deme to the right and instantaneous rate $\sigma^2/2$ and the left at instantaneous rate $\sigma^2/2$. If there is a nearby patch of ones — specifically if $V_t(x, 2) = (1, 0, 0, 0, 0)$, $V_t(x, 2) = (0, 0, 0, 0, 1)$ or $V_t(x, 2) = (1, 0, 0, 0, 1)$ — then it will continue on to establish there in the same jump, if it jumps in that direction. For example, if $V_t(x, 2) = (1, 0, 0, 0, 0)$, then it jumps one deme to the right at rate $\sigma^2/2$ but two demes to the left, to the patch of ones, at rate $\sigma^2/2$. Finally, if the environment changes ahead of the lineage in either of $x - 1$, x or $x + 1$, then it necessarily jumps to this site.

2.2.5. Temporally correlated patchy habitats. The analysis that follows will be restricted to the continuous-time limit (Ξ, X).

We allow each deme to change state at an instantaneous rate that is chosen in such a way as to ensure that the detailed balance equations are satisfied, so that the dynamics of the habitat are reversible, and the distribution π is stationary for the process. There are many ways to do this, and part of our purpose is to compare the effects on the motion of ancestral lineages of different dynamics of the habitat, for which the underlying (stationary) distribution of the habitat are the same. To maintain detailed balance, we need to set things up so that the ratio of instantaneous transition rates between two configurations is equal to the ratio of their probabilities under π. The simplest, and chosen, way to do this is to suppose that the habitat at location x switches state at instantaneous rate depending only upon its current state and the states of its two nearest neighbours. In the notation above, we specify then that

$$\lambda(x, \omega) \equiv \lambda(\omega_{x-1}, \omega_x, \omega_{x+1}), \tag{2.2.6}$$

for each $\omega \in \{0,1\}^{\mathbf{Z}}$ and $x \in \mathbf{Z}$. It is clear that any such λ satisfies the conditions of Assumption 2.2.4. The evolution of the habitats is then governed by the eight parameters $\{\lambda(\omega_{-1}, \omega_0, \omega_1) : (\omega_{-1}, \omega_0, \omega_1) \in \{0,1\}^3\}$. It is by varying these parameters that we control the degree of temporal autocorrelation in the environment.

In order for detailed balance to be satisfied, it must be the case that

$$\lambda(x, \omega)\pi(\mathrm{d}\omega) = \lambda(x, \omega^x)(\pi \circ \phi_x)(\mathrm{d}\omega),$$

where for every $x \in \mathbf{Z}$, the map $\phi_x : \{0,1\}^{\mathbf{Z}} \to \{0,1\}^{\mathbf{Z}}$ is defined by $\phi_x : \omega \mapsto \omega^x$.

We begin by considering the first pair of rates $\lambda(0,0,0)$ and $\lambda(0,1,0)$ corresponding to what we will call *birth* and *death* events respectively, ie when a new patch of ones appears in a region of zeroes or, *vice versa*, when a patch of ones is removed. In order to preserve detailed balance, the rates must be specified such that

$$\frac{\lambda(0,0,0)}{\lambda(0,1,0)} = \frac{\pi((\ldots,0,1,0,\ldots))}{\pi((\ldots,0,0,0,\ldots))} = \frac{p_0 p_1}{(1-p_0)^2}.$$

We may formalise the notion of of the central fraction here by way of the Radon–Nikodym derivative $\mathrm{d}(\pi \circ \phi_x)/\,\mathrm{d}\pi$ however we omit the arguments. For any $\omega \in \{0,1\}^{\mathbf{Z}}$ and $x \in \mathbf{Z}$, we may sensibly understand this ratio as the limit

$$\lim_{L\to\infty} \frac{\pi(\{\omega' : \omega'_y = \omega_y, |y| \le L\})}{\pi(\{\omega' : \omega'_y = \omega^x_y, |y| \le L\})},$$

which we note is independent of x. In order to preserve detailed balance between birth and

death events we will choose

$$\lambda(0,0,0) := \gamma_b \frac{p_0 p_1}{1-p_0}, \qquad \lambda(0,1,0) := \gamma_b(1-p_0), \tag{2.2.7}$$

for some parameter $\gamma_b \geq 0$ which we allow to vary. Of course, there are many ways in which one may specify these rates preserving detailed balance, but we find that this is most convenient.

Next we consider the rates $\lambda(1,1,1)$ and $\lambda(1,0,1)$ corresponding to what we will call *split* and *merge* events respectively, ie when a patch of ones develops a "hole" or, *vice versa*, when two adjacent patches of ones, separated by a sole zero between them, merge. The rates must be specified such that

$$\frac{\lambda(1,1,1)}{\lambda(1,0,1)} = \frac{\pi((\ldots,1,0,1,\ldots))}{\pi((\ldots,1,1,1,\ldots))} = \frac{p_1 p_0}{(1-p_1)^2}.$$

We will set

$$\lambda(1,1,1) := \gamma_s \frac{p_1 p_0}{1-p_1}, \qquad \lambda(1,0,1) := \gamma_s(1-p_1), \tag{2.2.8}$$

for some parameter $\gamma_s \geq 0$ which we allow to vary.

The parameters $\lambda(0,0,1)$ and $\lambda(1,0,0)$ correspond to events which we will call *left* and *right boundary expansion* respectively, ie when a region of ones expands its border by one position into a region of zeroes. To preserve symmetry, we will set these to be equal. Similarly, the parameters $\lambda(0,1,1)$ and $\lambda(1,1,0)$ correspond to events which we will call *left* and *right boundary contraction* respectively, ie when the boundary of a patch of ones recedes. Again, preserving symmetry, we will set these to be equal. In order to preserve detailed balance during either of these pairs of events, we must specify the rates such that

$$\frac{\lambda(0,0,1)}{\lambda(0,1,1)} = \frac{\pi((\ldots,0,1,1,\ldots))}{\pi((\ldots,0,0,1,\ldots))} = \frac{1-p_1}{1-p_0}.$$

We will set

$$\lambda(0,0,1) = \lambda(1,0,0) := \gamma_m(1-p_1), \qquad \lambda(0,1,1) = \lambda(1,1,0) := \gamma_m(1-p_0), \tag{2.2.9}$$

for some parameter $\gamma_m \geq 0$ which we allow to vary.

We summarise the above discussion in the following lemma.

LEMMA 2.2.9. *Let $(\Xi_t)_{t\in[0,T]}$ be the $\{0,1\}^{\mathbf{Z}}$-valued Markov process defined in Theorem 2.2.7 with rates $\lambda(x,\omega)$ satisfying* (2.2.6) *and $\{\lambda(\omega_{-1},\omega_0,\omega_1) : (\omega_{-1},\omega_0,\omega_1) \in \{0,1\}^3\}$ defined by equations* (2.2.7), (2.2.8) *and* (2.2.9). *Then the process Ξ is reversible and the distribution π is stationary for the process.*

PROOF. See the discussion above. □

2.3. Patch geometry under diffusive rescaling.

Our goal is to understand the effective dispersal speed of the lineage. Lineages remain trapped in patches until their demise; entering a region of zeroes at this time, they subsequently move according to a nearest neighbour random walk with speed σ^2 until the next patch is encountered, where it immediately jumps. It is clear then that the long-term speed of the lineage depends in a fundamental way on, one, the geometry of patches in the space-time plane and, two, on the abundance of patches. By varying the parameters of the previous section it is possible to obtain differing characteristics whilst maintaining the same stationary distribution on configurations of the environment.

Because the system is discrete, it is rather unwieldy to work with. To make headway, we rescale space such that the boundaries of the patches of ones and the lineage behave locally (making exceptions for the times that they are interacting) as real-valued diffusions.

We begin by defining a rescaling of the process (Ξ, X), heuristically motivating a choice of scaling parameters leading to a stationary distribution of patches of exponential width and whose boundaries behave as a system of interacting diffusions. In the following section we prove that after passing to a limit where patches are far apart, a patch's boundaries evolve according to Brownian motion plus drift; and that its boundaries reflect off one another until an exponential amount of local time at width zero, when it dies.

2.3.1. The rescaled parameters. Let $n \in \mathbf{N}$ be a scaling parameter. Our goal is to define a sequence of processes on a finer and finer lattice $\mathbf{Z}/n$. In correspondence with this scaling, we will define a sequence of $\{0,1\}^{\mathbf{Z}/n} \times \mathbf{Z}/n$-valued processes (Ξ^n, X^n) by specifying their distribution relative to the process (Ξ, X) of Theorem 2.2.7. We set, somewhat abusing notation,

$$(\Xi^n, X^n) \stackrel{d}{:=} \big((\Xi(nx) : x \in \mathbf{Z}/n), X/n\big),$$

where, the process indexed by n will depend on parameters p_0^n and p_1^n, as defined in Subsection 2.2.1; and parameters γ_b^n, γ_s^n and γ_m^n, as defined in equations (2.2.7), (2.2.8) and (2.2.9) respectively — all now dependent on the scaling parameter n.

Denote by π^n the measure on $\{0,1\}^{\mathbf{Z}/n}$, the natural rescaling of the stationary measure π which assigns geometric lengths to regions of ones, with parameter p_1^n; and geometric lengths to regions of zeroes, with parameter p_0^n. We wish to choose the parameters p_0^n and p_1^n in such a way that patches retain a macroscopic width after passage to the limit.

PROPOSITION 2.3.1. *Let p_0^n and p_1^n be defined by*

$$p_0^n := \frac{v_0}{n}, \qquad p_1^n := \frac{v_1}{n}.$$

View $\{0,1\}^{\mathbf{Z}/n}$ *as a subset of the* $\{0,1\}$*-valued right-continuous functions* $D(\mathbf{R},\{0,1\})$ *by setting* $\omega(x) := \omega_{\lfloor nx \rfloor/n}$ *for all* $x \in \mathbf{R}$ *and equip* $D(\mathbf{R},\{0,1\})$ *with the Skorokhod topology. Then* $\pi^n \Rightarrow \pi$ *as* $n \to \infty$*, where* π *is the probability measure on* $D(\mathbf{R},\{0,1\})$ *describing the law of the stationary and reversible* $\{0,1\}$*-valued Markov chain, where*

$$\pi(\omega(\mathrm{d}x) = 1 \,|\, \omega(0) = 0) = v_0\,\mathrm{d}x + o(\mathrm{d}x),$$

and, similarly,

$$\pi(\omega(\mathrm{d}x) = 0 \,|\, \omega(0) = 1) = v_1\,\mathrm{d}x + o(\mathrm{d}x).$$

The proof is deferred until Section 2.6.

Importantly, from this proposition we may deduce that if Ξ^n is weakly convergent, then its limit must necessarily admit π as a stationary distribution.

COROLLARY 2.3.2. *Suppose that* $\Xi^n \Rightarrow \Xi$ *in* $D(\mathbf{R},[0,1])$ *as* $n \to \infty$*, with the embedding to this space as in Proposition 2.3.1. Then* π *is a stationary distribution for* Ξ*.*

PROOF. This follows immediately from Theorem 4.9.12 of [40]. □

In other words, after passing to the limit $n \to \infty$, the environment is distributed continuously across the real line. Regions of *good* habitat have width converging to an exponential random variable with rate v_1; and regions of *bad* habitat have width converging to an exponential random variable with rate v_0. Moving from left to right through space, the continuous-time Markov chain describing the sequence of habitats is stationary and the one-dimensional distributions satisfy $\pi(\omega(x) = 1) = v_0/(v_0 + v_1)$.

Consider now the right hand boundary of an arbitrarily chosen patch of ones. We will refer to this as the *focal* patch. According to (2.2.9), under the current rescaling, it jumps a distance $1/n$ to the right at rate $\gamma_m^n(1-p_1^n)$ (as long as there is no patch of ones immediately to the right, ie separated by just one zero) and a distance $1/n$ to the left at rate $\gamma_m^n(1-p_0^n)$ (as long as the patch currently has a width strictly greater than one). In order to obtain a diffusive limit, we take

$$\gamma_m^n := n^2.$$

Denoting the position of the focal right-hand boundary by R_t^n, at least locally,

$$\mathbf{E}[R_t^n - R_0^n \mid R_0^n = r] \approx -(v_1 - v_0)t + o(t),$$

and

$$\mathbf{E}[(R_t^n - R_0^n)^2 \mid R_0^n = r] \approx 2t + o(t),$$

ie R^n approximates a Markov process which behaves locally as a diffusion process with speed 2 and drift $-(v_1 - v_0)$. In particular, if $v_0 < v_1$ — so that regions of zeroes are on average wider than regions of ones — then the right hand boundary drifts to the left. By a similar argument, the left hand boundary approximates a diffusion process with speed 2 but drift $+(v_1 - v_0)$, in the opposite direction.

When the left- and right-hand boundaries coincide, ie when the patch consists of a sole one, either it may expand a distance $1/n$ to the left or right — both with rate $\gamma_m^n(1 - p_1^n)$ in accordance with definitions (2.2.9) — *or*, in accordance with (2.2.7), the patch of ones is removed at rate $\gamma_b^n(1 - p_0^n)$. We will take

$$\gamma_b^n := \gamma_b n.$$

The same parameter controls the rate at which new patches are thrown down in a region of zeroes. At the nth stage of rescaling, a region of zeroes of unit width contains approximately n zeroes; in accordance with the definitions (2.2.7), at rate

$$\gamma_b^n \frac{p_0^n p_1^n}{1 - p_0^n} \simeq \frac{\gamma_b v_0 v_1}{n}$$

those zeroes with no neighbouring demes of good habitat flip to a one. In particular, we can readily believe that in passing to the limit $n \to \infty$, new patches of ones are thrown down in regions of zeroes according to a Poisson point process with rate $\gamma_b v_0 v_1$.

It remains to define γ_s^n under the rescaling. This controls the rate at which patches of good habitat split into two and adjacent patches of good habitat merge, in accordance with the definitions (2.2.8). However, allowing such events will greatly complicate the analysis, so as a final assumption we set these rates equal to zero.

2.3.2. Appearance of new patches. We begin by discussing the mechanism through which new patches of good habitat arrive into the system. As the system evolves, a deme of bad habitat flips to good habitat at rate

$$\gamma_b^n \frac{p_0^n p_1^n}{1 - p_0^n} = \frac{1}{n} \frac{\gamma_b v_0 v_1}{1 - v_0/n} \tag{2.3.1}$$

if and only if both of its neighbouring demes are also bad habitats, ie if locally the view from the deme is $(0, 0, 0)$. We may however view this from another angle. We may describe *potential* arrivals of new clusters according to points in a Poisson random measure P^n on $(\mathbf{R}_+ \times \mathbf{R}, \mathcal{B}(\mathbf{R}_+ \times \mathbf{R}))$ with intensity measure m_n defined by

$$m_n(\mathrm{d}t, \mathrm{d}x) = \frac{1}{n} \frac{\gamma_b v_0 v_1}{1 - v_0/n} \, \mathrm{d}t \otimes \sum_{x \in \mathbf{Z}/n} \delta_x(\mathrm{d}x). \tag{2.3.2}$$

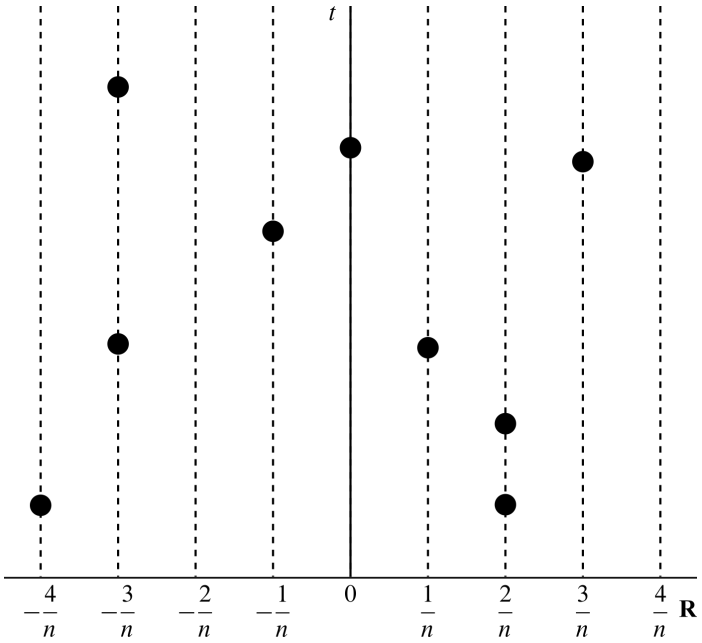


FIGURE 2.1

Here, as usual, $\mathcal{B}(\mathbf{R}_+ \times \mathbf{R})$ denotes the Borel σ-algebra generated by open sets and δ_x denotes the Dirac measure at x. If $(t, x) \in P^n$, then we say that there is a *potential* arrival of a new cluster at the deme x at time t. This is merely ignored if any of the habitats at x or its two nearest neighbours are good, but otherwise accepted. In simpler terms, we associate to each $x \in \mathbf{Z}/n$ an independent Poisson process with rate given by equation (2.3.1); at the time of an arrival of the Poisson process at x, we attempt to throw down a new patch at x but reject it if either its or its neighbours' habitats are good. This Poisson process of points in the space-time place is depicted in Figure 2.1. The description as a Poisson random measure however will be useful for describing the behaviour after passing to the limit because it means that we can sample the potential arrival times at initialisation, independently of future dynamics of the process. Importantly, we need only sample the behaviour of the boundaries of patches of good habitat present at time 0, and those which are allowed to arrive according to this mechanism.

Observing that for any $A \subset \mathbf{R}$, the set $A \cap \mathbf{Z}/n$ contains $O(n)$ demes, it is believable that P^n converges; in particular, we expect that the sequence of Poisson random measures converges to that with mean measure $\gamma_b v_0 v_1 \, \mathrm{d}t \otimes \mathrm{d}x$.

In order to give meaning to such a statement, we first need to establish the space in which our sequence of random measures will converge. We specialise the following notions to the space $\mathbf{R}^d$ but it extends to any complete separable metric space. We denote by $\mathcal{M}(\mathbf{R}^d)$ the space of locally finite measures on $\mathbf{R}^d$. We will say that a sequence of measures

$(\mu_n)_{n\in\mathbf{N}} \subset \mathcal{M}(\mathbf{R}^d)$ converges *vaguely* to μ if and only if

$$\int f(x)\mu_n(\mathrm{d}x) \to \int f(x)\mu(\mathrm{d}x)$$

for all $f \in C_c^+(\mathbf{R}^d)$, the space of all real-valued non-negative continuous functions on $\mathbf{R}^d$ with compact support. The space $\mathcal{M}(\mathbf{R}^d)$ is topologised by vague convergence. A streamlined presentation of the theory weak convergence of random measures is given in Chapter 4 of Kallenberg [70], and, applying their results, it is relatively straightforward to prove the following.

THEOREM 2.3.3. *Let $(P^n)_{n\in\mathbf{N}}$ denote the sequence of Poisson random measures with intensity measure m_n defined by* (2.3.2)*. Then, as $n \to \infty$, $P^n \Rightarrow P$ in $\mathcal{M}(\mathbf{R}^2)$ equipped with the vague topology, where P is the Poisson random measure with intensity measure*

$$m(\mathrm{d}t, \mathrm{d}x) = \gamma_b v_0 v_1 \, \mathrm{d}t \otimes \mathrm{d}x.$$

The proof is deferred until Section 2.7.

2.3.3. The geometry of a single patch. Because of the complicated interactions between patches, it is difficult to specify the full behaviour of the system as we pass to the limit $n \to \infty$. However, patches are on average spaced a distance v_0^{-1} apart and so, if we additionally take $v_0 \to 0$ as we pass to the limit, it is possible to ignore the effect of patch interactions.

We will take a sequence of initial conditions satisfying $\Xi_0^n(0) = 1$. Our interest lies in characterising the geometry (in the space-time plane) of the patch of good habitat initially containing containing the origin. We will refer to this patch as the *focal patch*. In order to do this, we will need some notation. We will denote by R_t^n the position in $\mathbf{Z}/n$ of the right-hand boundary of this patch; by L_t^n the position of the left-hand boundary; by $M_t^n := (R_t^n + L_t^n)/2$ the midpoint of the patch; and by $W_t^n = R_t^n - L_t^n$ its width. These four processes are defined until a time when the patch is removed which we denote by T^n, although it will be useful to denote by (L^n, R^n, M^n, W^n) the processes with the same law which evolve indefinitely. Letting U denote an independent exponentially distributed random variable with mean one, we have that

$$T^n = \inf\left\{t \geq 0 : \int_0^t 1\{W_s^n = 0\}\,\mathrm{d}s \geq \frac{U}{\gamma_b^n(1-p_0^n)}\right\}, \tag{2.3.3}$$

since, by equation (2.2.7), the patch is removed from the system at rate $\gamma_b^n(1-p_0^n)$. In fact, we need only keep track of the midpoint and of the width of the patch, and the law of the

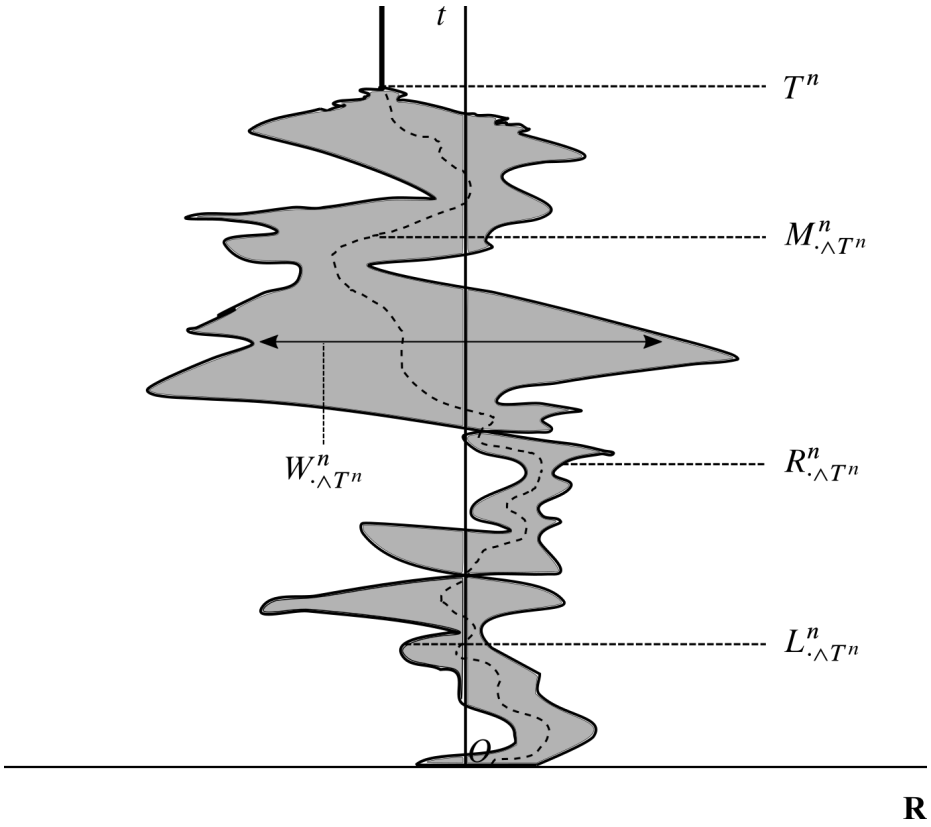


FIGURE 2.2

patch is then described by the stopped process

$$(M^n_{\cdot\wedge T^n}, W^n_{\cdot\wedge T^n}, T^n).$$

The set up is depicted in Figure 2.2.

Our main result is formulated as follows.

THEOREM 2.3.4. *Suppose that* $(\Xi^n_0)_{n\in\mathbf{N}}$ *be a sequence of initial conditions such that* $\Xi^n_0(0) = 1$ *almost surely, distributed according to* $\pi^n(\mathrm{d}\omega \mid \Xi^n_0(0) = 1)$. *Denote by the 4-tuple* (L^n, R^n, M^n, W^n) *the ordered product recording, in order, left-hand boundary, the right-hand boundary, the midpoint and width of the patch initially straddling the origin, evolving indefinitely and never killed, and let* T^n *be the stopping time defined by* (2.3.3).

Let $-L_0$ *and* R_0 *denote independent exponentially distributed random variables with parameter* v_1 *and subsequently define*

$$M_0 = \frac{R_0 - L_0}{2}$$

and

$$W_0 := R_0 - L_0.$$

Denote by B^M and B^W a pair of independent standard Brownian motions on $\mathbf{R}$ *and let M, S, W and ℓ be the processes defined by*

$$M_t := M_0 + B_t^M,$$

$$S_t := W_0 - 2v_1 t + 2B_t^W,$$

$$\ell_t := 0 \vee \big(- \inf_{s \le t} S_s \big),$$

and

$$W_t := S_t + \ell_t.$$

With U an independent exponential random variable with unit mean, let

$$T := \ell^{-1}(2U/\gamma_b) := \inf\{t \ge 0 : \ell_t \ge 2U/\gamma_b\}.$$

Then as $n \to \infty$ and $v_0 \to 0$, the triple $(M^n_{\cdot \wedge T^n}, W^n_{\cdot \wedge T^n}, T^n) \Rightarrow (M_{\cdot \wedge T}, W_{\cdot \wedge T}, T)$ in the space $D([0,\infty), \mathbf{R}^2) \times \mathbf{R}_+$.

The proof is deferred until Section 2.8.

Let us parse the long statement of this result. Proposition 2.3.1 states that the sequence of stationary distributions converges to one that gives patches and the distances between them independent exponential widths, with parameters v_1 and v_0 respectively. It therefore comes as no surprise that our limiting patch will have boundaries with initial values (L_0, R_0) with $-L_0$ and R_0 described by independent exponentials with parameter v_1. The process W defined as in the statement of the theorem would more colloquially be referred to as the *reflected* version of the diffusion process S, reflected at the origin. In fact, the ordered pair (W, ℓ) is the unique solution to the Skorokhod problem for S, which we define precisely when we come to the proof in Section 2.8, and ℓ is the *local time* of W at the origin. The process ℓ, then, gives meaning to a concept of the *amount of time* the patch spends at width zero. The time T, is the first time that the width process W has spent an exponential amount of local time at width zero; this exponential has parameter $\gamma_b/2$. Moreover, in the limit, the midpoint process M is independent of the width of the patch W and the stopping time T.

2.4. Proof of Lemma 2.2.1.

The motion of a lineage over a single generation depends just on the state of the habitat at that location and its two nearest neighbours in the previous generation. Recall that we use zero to denote a region of *bad* habitat and one to denote regions of *good* habitat. If the lineage is currently at the origin, ie $L_0 = 0$, then $\Delta L_0 := L_1 - L_0$, the change of the location of the lineage over a single generation, has a distribution which can be specified

conditional on the population sizes $(N(1,-1), N(1,0), N(1,1))$ one generation ago. For example, if $(N(1,-1), N(1,0), N(1,1)) = (1,0,1)$, which indicates that the population is *bad* at the current location of the lineage but *good* in both neighbouring demes one generation into the past, then according to our rules, the lineage will jump to the left or right with equal probability.

By symmetry, $\mathbf{E}[\Delta L_0] = 0$ and thus it suffices to calculate $\mathbf{E}[(\Delta L_0)^2]$. We simply tabulate:

$(\omega_{-1}, \omega_0, \omega_1)$	$\mathbf{P}[N(1,i) = \omega_i, i = -1, 0, 1]$	$\mathbf{E}[(\Delta L_0)^2 \mid N(1,\cdot) = \omega]$
$(0,0,0)$	$\frac{p_1}{p_0+p_1}(1-p_0)^2$	$\frac{2}{3}$
$(0,0,1)$ or $(1,0,0)$	$\frac{p_1}{p_0+p_1}(1-p_0)p_0$	1
$(0,1,0)$	$\frac{p_1}{p_0+p_1}p_0p_1$	0
$(0,1,1)$ or $(1,1,0)$	$\frac{p_1}{p_0+p_1}p_0(1-p_1)$	$\frac{1}{2}$
$(1,0,1)$	$\frac{p_1}{p_0+p_1}p_0p_1$	1
$(1,1,1)$	$\frac{p_1}{p_0+p_1}(1-p_1)^2$	$\frac{2}{3}$

Using these probabilities we may calculate that

$$\begin{aligned}
&\mathbf{E}[(\Delta L_0)^2] \\
&= \frac{1}{3(p_0+p_1)}\big(2p_1(1-p_0)^2 + 6p_1(1-p_0)p_0 \\
&\qquad\qquad + 3p_1p_0(1-p_1) + 3p_0^2p_1 + 2p_0(1-p_1)^2\big) \\
&= \frac{2}{3} + \frac{p_0p_1}{3(p_0+p_1)}\big(1-(p_0+p_1)\big),
\end{aligned}$$

as per the statement of the lemma.

2.5. Convergence to the continuous-time process.

This section contains the proofs of Proposition 2.2.6 and Theorem 2.2.7, beginning with the latter. Our proof will follow a standard scheme, see eg Ethier and Kurtz [40]. First we will prove that the sequence of processes (Ξ^h, X^h) is tight in the space $D([0,T], \{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$. We will then characterise the (weak) limit points of this sequence by describing them as a Markov process on $\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}$ whose value at each coordinate in $\mathbf{Z}$ changes at a rate depending on a local neighbourhood. We recast the process as an interacting particle system and prove that such a description uniquely characterises the process. We conclude with the proof of Proposition 2.2.6.

It will be useful to note that

$$d((\omega,x),(\xi,y)) := \sum_{z\in\mathbf{Z}} \frac{1}{2^{|z|}}|\omega_z - \xi_z| + |x-y|$$

defines a metric for the product topology on $\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}$. This observation allows us to make use of the following tightness criterion, a special case of Theorem 3.8.6 in Ethier and Kurtz [40].

THEOREM 2.5.1. *Let (E,d) be a complete separable metric space and let $(S^h)_{h\geq 0}$ be a family of processes with sample paths in $D([0,\infty),E)$, and $\mathcal{F}_t^h := \sigma(\{S_s^h : s \leq t\})$ its natural filtration. The sequence $(S^h)_{h\geq 0}$ is tight in $D([0,\infty),E)$ if the following two conditions hold.*

1. *For every $\varepsilon > 0$ and $t \in \mathbf{Q}$, there exists a compact set $K \subseteq E$ such that*

$$\limsup_{h\to 0} \mathbf{P}[S_t^h \in E\backslash K] \leq \varepsilon.$$

2. *For each $T < \infty$, there exists a $\beta > 0$ such that*

$$\lim_{\delta\to 0} \limsup_{h\to 0} \mathbf{E}[\sup_{0\leq t\leq T} \sup_{\substack{0\leq u\leq \delta \\ t+u\leq T}} \mathbf{E}[(d(S_{t+u}^h, S_t^h)\wedge 1)^\beta \mid \mathcal{F}_t^h]] = 0.$$

We are able to use the theorem directly in proving tightness.

LEMMA 2.5.2. *The sequence $\{(\Xi^h, X^h)\}_{h\geq 0}$ is tight in $D([0,T],\{0,1\}^{\mathbf{Z}}\times\mathbf{Z})$.*

PROOF. We begin with the first condition of Theorem 2.5.1. For any $K \in \mathbf{Z}_+$, the set $\{0,1\}^{\mathbf{Z}} \times ([-K,K]\cap\mathbf{Z})$ is compact in $\{0,1\}^{\mathbf{Z}}\times\mathbf{Z}$ in the product topology. It is therefore sufficient to prove that for all $t \in [0,T]$ and $\varepsilon > 0$, there exists a K such that

$$\sup_{h>0} \mathbf{P}[|X_t^h| > K] < \varepsilon. \tag{2.5.1}$$

Defining

$$J_T^h := \#\{t \in [0,T] : X_t^h \neq X_{t-}^h\},$$

ie the number of times that the lineage jumps before time T, (2.5.1) would certainly hold if it can be shown that

$$\sup_{h>0} \mathbf{E}[J_T^h] < \infty. \tag{2.5.2}$$

This follows by a straightforward application of Markov's inequality, noticing that $|X_t^h| > K$ implies that there have been at least K jumps by time T.

When $V_t^h(X_t^h,0) = 1$, the probability that X^h jumps in the next discrete generation into the past is bounded above by $(\bar{\lambda}+\sigma^2)h + o(h)$. Here, $\bar{\lambda}$ is the constant appearing in Assumption 2.2.4. The $\Theta(h)$ term corresponds to the events that the habitat at lineage's current location state (possibly forcing the lineage to jump if either of the nearest neighbours

contains a one), or the event that these three states remain the same but the lineage jumps. If $V_t^h(X_t^h, 1) = (0, 0, 0)$ then the probability that X^h jumps in the next discrete generation into the past is bounded above by $(2\bar{\lambda} + \sigma^2)h + o(h)$. The $\Theta(h)$ term corresponds to the events that either the habitats at lineage's two nearest neighbours flip to a good state, or the event that these three states remain the same but the lineage jumps. (This jump may possibly be two steps however for now we are only interested in the time between jumps rather than their size.) We also observe that because of the way the process X^h was constructed, there can be no time such that $V_t^h(X_t^h, 0) = 0$ but $V_t^h(X_t^h, 1)$ contains a one. In particular, at any time, the probability that X^h jumps in the next discrete generation is bounded above by $(2\bar{\lambda} + \sigma^2)h + o(h)$ and so the expected number of jumps in $\lfloor T/h \rfloor$ is at most $\lfloor T/h \rfloor[(2\bar{\lambda} + \sigma^2)h + o(h)]$. Whence (2.5.2) clearly holds.

We turn now to the second condition of Theorem 2.5.1, controlling the regularity of the paths of the process. We will take $\beta = 1$ and, noticing that

$$\begin{aligned}\mathbf{E}[d((\Xi_{t+u}^h, X_{t+u}^h), (\Xi_t^h, X_t^h)) \wedge 1 \,|\, \mathcal{F}_t^h] \\ \leq \mathbf{E}\left[\sum_{x \in \mathbf{Z}} \frac{1}{2^{|x|}} |\Xi_{t+u}^h(x) - \Xi_t^h(x)| \,\middle|\, \mathcal{F}_t^h \right] + \mathbf{E}[|X_{t+u}^h - X_t^h| \,|\, \mathcal{F}_t^h], \quad (2.5.3)\end{aligned}$$

we will estimate both of the terms on the right hand side in turn.

We first note that we may partition any time window $(t, t+u)$ into non-overlapping discrete generations: for fixed $t, u > 0$ let $s_k = (\lfloor t/h \rfloor + k)h$ for $k = 0, 1, \ldots, \lfloor u/h \rfloor$. This allows us to write, by a simple application of the tower property for conditional expectation,

$$\begin{aligned}\mathbf{E}\left[\sum_{x \in \mathbf{Z}} \frac{1}{2^{|x|}} |\Xi_{t+u}^h(x) - \Xi_t^h(x)| \,\middle|\, \mathcal{F}_t^h \right] \\ \leq \sum_{k=1}^{\lfloor u/h \rfloor} \mathbf{E}\left[\mathbf{E}\left[\sum_{x \in \mathbf{Z}} \frac{1}{2^{|x|}} |\Xi_{s_k}^h(x) - \Xi_{s_{k-1}}^h(x)| \,\middle|\, \mathcal{F}_{s_{k-1}}^h \right] \middle|\, \mathcal{F}_t^h \right].\end{aligned}$$

Recalling that habitats flip states independently of one another and that because of the form of the transition probabilities defined in equation (2.2.4) and Assumption 2.2.4, the probability that the habitat at x flips moving from one generation to the next is zero for $|x| > h^{-1/2}$ and is bounded by ch uniformly in x for some constant $c < \infty$. Below, we will let c be a constant which varies from line to line. In particular, conditioning on n-tuples of sites which flip passing from one generation to the next, for any discrete time s_k

as above,

$$\begin{aligned}
\mathbf{E}\Big[\sum_{x\in\mathbf{Z}}\frac{1}{2^{|x|}}|\Xi^h_{s_k}(x)-\Xi^h_{s_{k-1}}(x)|\,\Big|\,\mathcal{F}^h_{s_{k-1}}\Big] &\le \sum_{n=1}^{2h^{-1/2}+1}(ch)^n \sum_{\substack{A\subseteq\{x\in\mathbf{Z}:|x|\le h^{-1/2}\}\\ |A|=n}}\sum_{x\in A}\frac{1}{2^{|x|}}\\
&\le \sum_{n=1}^{2h^{-1/2}+1}(ch)^n \sum_{|x_1|,\dots,|x_n|\le h^{-1/2}}\sum_{i=1}^{n}\frac{1}{2^{|x_i|}}\\
&= \sum_{n=1}^{2h^{-1/2}+1}(ch)^n \sum_{i=1}^{n}\sum_{|x_1|,\dots,|x_n|\le h^{-1/2}}\frac{1}{2^{|x_i|}}\\
&\le \sum_{n=1}^{2h^{-1/2}+1}(ch)^n 3n(2h^{-1/2}+1)^{n-1}\\
&\le 3ch\sum_{n=1}^{2h^{-1/2}+1} n(3ch^{1/2})^{n-1}\\
&\le 3ch\sum_{n=1}^{\infty} n(3ch^{1/2})^{n-1}\\
&= \frac{3ch}{(1-3ch^{1/2})^2}.
\end{aligned}$$

Moreover, for all $t\in[0,T]$, $u\le\delta$,

$$\mathbf{E}\Big[\sum_{x\in\mathbf{Z}}\frac{1}{2^{|x|}}|\Xi^h_{t+u}(x)-\Xi^h_t(x)|\,\Big|\,\mathcal{F}^h_t\Big]\le\Big\lfloor\frac{\delta}{h}\Big\rfloor\frac{3ch}{(1-3ch^{1/2})^2}. \tag{2.5.4}$$

We will now estimate the second term on the right hand side of equation (2.5.3), controlling the regularity of the paths of the lineage. Again, partitioning the time interval $(t,t+u)$ as above, we may write

$$\mathbf{E}[|X^h_{t+u}-X^h_t|\,|\,\mathcal{F}^h_t]\le\sum_{k=1}^{\lfloor u/h\rfloor}\mathbf{E}[\mathbf{E}[|X^h_{s_k}-X^h_{s_{k-1}}|\,|\,\mathcal{F}^h_{s_{k-1}}]\,|\,\mathcal{F}^h_t].$$

As stated in the proof of the first condition of Theorem 2.5.1, above, the process X^h jumps in any one discrete generation with probability at most $(2\bar{\lambda}+\sigma^2)h+o(h)$. Moreover, it jumps by at most two steps. Thus for all $t\in[0,T]$ and $u\le\delta$,

$$\mathbf{E}[|X^h_{t+u}-X^h_t|\,|\,\mathcal{F}^h_t]\le\Big\lfloor\frac{\delta}{h}\Big\rfloor((4\bar{\lambda}+2\sigma^2)h+o(h)). \tag{2.5.5}$$

Given (2.5.3), (2.5.4) and (2.5.5), the second condition of Theorem 2.5.1 holds and we conclude that the sequence (Ξ^h, X^h) is tight. □

We proceed to characterise the limit points of $\{(\Xi^h, X^h)\}_{h\geq 0}$ by specifying there transition rates.

LEMMA 2.5.3. *Let (Ξ, X) be any weak limit point of the sequence $\{(\Xi^h, X^h)\}_{h\geq 0}$ converging in $D([0,T], \{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$. Then (Ξ, X) has generator $A : C(\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}) \to C(\{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$ satisfying* (2.2.5).

PROOF. Let (Ξ, X) be any weak limit point of the sequence $\{(\Xi^h, X^h)\}_{h\geq 0}$ converging in $D([0,T], \{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$ and let us pass to this convergent subsequence and write $(\Xi^h, X^h) \Rightarrow (\Xi, X)$. Fix $f \in C_0(\{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$. Since $f \in C_0(\{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$, there exists some $N_0 < \infty$ such that, for any $N \geq N_0$, if $\omega, \xi \in \{0,1\}^{\mathbf{Z}}$ satisfy $\omega_x = \xi_x$ for all $|x| \leq N$ then $f(\omega, x) = f(\xi, x)$ for all $x \in \mathbf{Z}$, and $f(\omega, x) = f(\omega, y)$ for all ω whenever $|x|, |y| > N$. For $\eta \in \{0,1\}^N$ and $y \in \mathbf{Z}$, denote by

$$C_{\eta,y} := \{(\xi, y) : \xi \in \{0,1\}^{\mathbf{Z}}, \xi_x = \eta_x, |x| \leq N\}.$$

Observe that f is constant on $C_{\eta,y}$ for fixed η. For $\eta \in \{0,1\}^N$, denote by $\hat{\eta}$ its extension to an element of $\{0,1\}^{\mathbf{Z}}$ which agrees with η for all $|x| \leq N$ and is equal to zero for all $|x| > N$.

Our aim is to compute

$$Af(\omega, x) := \lim_{t \searrow 0} \frac{1}{t} \mathbf{E}[f(\Xi_t, X_t) - f(\omega, x) \,|\, (\Xi_0, X_0) = (\omega, x)]$$

for all (ω, x) satisfying the condition that if $\omega_x = 0$ then $\omega_{x-1} = \omega_{x+1} = 0$ too; to do so, we will make explicit use of the form of f. In particular, it allows us to swap limits and sums over states (ξ, y) in a convenient way: we can compute that

$$\begin{aligned}
Af(\omega, x) &= \lim_{t \searrow 0} \frac{1}{t} \sum_{\xi \in \{0,1\}^{\mathbf{Z}}} \sum_{y \in \mathbf{Z}} \{f(\xi, y) - f(\omega, x)\} \mathbf{P}_{(\omega,x)}[(\Xi_t, X_t) = (\xi, y)] \\
&= \lim_{t \searrow 0} \frac{1}{t} \sum_{\eta \in \{0,1\}^N} \sum_{y \in \mathbf{Z}} \{f(\hat{\eta}, y) - f(\omega, x)\} \mathbf{P}_{(\omega,x)}[(\Xi_t, X_t) \in C_{\eta,y}] \\
&= \sum_{\eta \in \{0,1\}^N} \sum_{|y| \leq N} \{f(\hat{\eta}, y) - f(\omega, x)\} \lim_{t \searrow 0} \frac{1}{t} \mathbf{P}_{(\omega,x)}[(\Xi_t, X_t) \in C_{\eta,y}] \\
&\quad + \sum_{\eta \in \{0,1\}^N} \{f(\hat{\eta}, N+1) - f(\omega, x)\} \lim_{t \searrow 0} \frac{1}{t} \mathbf{P}_{(\omega,x)}\Big[(\Xi_t, X_t) \in \bigcup_{|y| > N} C_{\eta,y}\Big].
\end{aligned}$$

The last line follows from the fact that $f(\hat{\eta}, y) = f(\hat{\eta}, N+1)$ for all $|y| > N$.

Fix $N > 1$. We start by computing

$$\lim_{t \searrow 0} \frac{1}{t} \mathbf{P}_{(\omega,x)}[(\Xi_t, X_t) \in C_{\eta,y}]$$

for possible choices of η and y. Since convergence in distribution on $D([0,T], \{0,1\}^{\mathbf{Z}} \times \mathbf{Z})$ implies convergence of the finite-dimensional distributions (see eg Theorem 3.7.8 of Ethier and Kurtz [40]) and, in particular, convergence of the joint distribution of (Ξ_t, X_t) and (Ξ_0, X_0),

$$\lim_{t \searrow 0} \frac{1}{t} \mathbf{P}_{(\omega,x)}[(\Xi_t, X_t) \in C_{\eta,y}] = \lim_{t \searrow 0} \frac{1}{t} \lim_{h \searrow 0} \mathbf{P}[(\Xi_t^h, X_t^h) \in C_{\eta,y} \mid (\Xi_0^h, X_0^h) = (\omega, x)].$$

Up until the time t, there are $O(h^{-1})$ discrete generations of (Ξ^h, X^h). In each generation, the state of the habitat at each x with $|x| \leq N$ flips independently of one another and with probability $O(h)$. Such a flip may possibly lead to a jump of the lineage. Similarly, the lineage jumps with probability $O(h)$. In particular, in passing to the limit $h \to 0$, events which involve two or more habitat flips, or a habitat flip and a jump of the lineage, are negligible. Denote by $\alpha_N(\omega, x)$

$$\lim_{h \to 0} \frac{1}{h} \mathbf{P}_{(\omega,x)}[(\Xi_h^h(-N), \ldots, \Xi_h^h(N), X_h^h) \neq (\Xi_0^h(-N), \ldots, \Xi_0^h(N), X_0^h)],$$

the $O(h)$ term in the probability that any of the habitats at locations $|y| \leq N$ flip or the lineage jumps in any one discrete generation. This is necessarily finite by the above discussion. Furthermore, events which involve jumps in two or more discrete generations are $O(t^2)$. These events are thus also negligible.

We proceed focussing on one example as the methods are repetitive. For some z with $z \neq x-1, x, x+1$ and $|z| \leq N$ let $\eta_w = \omega_w^z$ and $y = x$. The only non-negligible event when passing to the limit is that when the habitat at z flips for some discrete generation before time t, but before and after that there is no change to the habitats $|w| \leq N$ or to the position of the lineage. Conditioning on the generation in which the switch of state at z

occurs, we compute that

$$\begin{aligned}
&\mathbf{P}[(\Xi_t^h, X_t^h) \in C_{\omega^z,x} \mid (\Xi_0^h, X_0^h) = (\omega, x)] \\
&= \sum_{i=1}^{\lfloor t/h \rfloor} (1-\alpha_N(\omega,x)h)^{i-1}\lambda(z,\omega)h(1-\alpha(\omega^z,x)h)^{\lfloor t/h\rfloor - i} + O(t^2) \\
&= \lambda(z,\omega)(1-\alpha_N(\omega^z,x)h)^{\lfloor t/h\rfloor -1} h \sum_{i=1}^{\lfloor t/h\rfloor} \left(\frac{1-\alpha_N(\omega,x)h}{1-\alpha_N(\omega^z,x)h}\right)^{i-1} + O(t^2) \qquad (2.5.6)\\
&= \lambda(z,\omega)(1-\alpha_N(\omega^z,x)h)^{\lfloor t/h\rfloor -1} \\
&\qquad \times h\left(1-\frac{1-\alpha_N(\omega,x)h}{1-\alpha_N(\omega^z,x)h}\right)^{-1}\left[1-\left(\frac{1-\alpha_N(\omega,x)h}{1-\alpha_N(\omega^z,x)h}\right)^{\lfloor t/h\rfloor}\right] + O(t^2),
\end{aligned}$$

and in particular we obtain that

$$\begin{aligned}
&\lim_{t\searrow 0}\frac{1}{t}\lim_{h\searrow 0}\mathbf{P}[(\Xi_t^h, X_t^h) \in C_{\eta,y} \mid (\Xi_0^h, X_0^h) = (\omega, x)] \\
&= \lim_{t\searrow 0}\frac{1}{t}\lambda(x,\omega)e^{-\alpha_N(\omega^z,x)t}(\alpha_N(\omega,x)-\alpha(\omega^z,x))^{-1}(1-e^{-(\alpha_N(\omega,x)-\alpha_N(\omega^z,x))t}) \\
&= \lambda(x,\omega),
\end{aligned}$$

as required. Proceeding in exactly the same way, one can show that for all $\eta \in \{0,1\}^N$, $|y| \le N$

$$\lim_{t\searrow 0}\frac{1}{t}\mathbf{P}_{(\omega,x)}[(\Xi_t, X_t) \in C_{\eta,y}] = c((\omega,x),(\omega^\eta, y)),$$

where $\omega_z^\eta := \eta_z$ for $|z| \le N$ and $\omega_z^\eta := \omega_z$ otherwise.

Finally we compute

$$\lim_{t\searrow 0}\frac{1}{t}\mathbf{P}_{(\omega,x)}\left[(\Xi_t, X_t) \in \bigcup_{|y|>N} C_{\eta,y}\right] = \lim_{t\searrow 0}\frac{1}{t}\lim_{h\searrow 0}\mathbf{P}_{(\omega,x)}\left[(\Xi_t^h, X_t^h) \in \bigcup_{|y|>N} C_{\eta,y}\right].$$

Recall that $N \ge N_0$ was arbitrary. Choose $N > |x| + 3$. The event on the right hand side implies that X^h has jumped at least twice before time t. The lineage X^h jumps in each discrete generation with probability at most $O(h)$ and thus, whilst it is possible that X^h jumps more than twice before time t, it does so with probability $O(t^2)$. In particular, the limit is equal to 0.

This completes the proof. □

We now prove that the characterisation of the limit points given by the previous result uniquely defines a Markov process on $\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}$.

LEMMA 2.5.4. *There is a unique Markov process* (Ξ, X) *on* $\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}$ *with transition*

rates $c((\omega, x), (\xi, y))$ of Theorem 2.2.7.

PROOF. We may equally view the process (Ξ, X) as a process (Ξ, Γ) on $\{0,1\}^{\mathbf{Z}} \times \{0,1\}^{\mathbf{Z}}$, where $\Gamma_t(x) := 1\{X_t = x\}$. With this notation, the process (Ξ, X) can be recast as an interacting particle system on $(\{0,1\}^2)^{\mathbf{Z}}$, where if $(a, b) \in \{0,1\}^2$ is the value at the xth coordinate, then $a = 1$ if the habitat at x is *good* and $a = 0$ if it is *bad*; and $b = 1$ if the lineage is current at the site x and $b = 0$ if it is elsewhere. Assumption 2.2.4 guarantees that λ depends only on finite number of coordinates and is translationally invariant; furthermore, the motion of the lineage depends on at most its four nearest neighbours. Whence the interacting particle system (Ξ, Γ) is *finite-range* in the sense of Definition I.4.17 of Liggett [90]; in particular, the transition rates satisfy his conditions (3.3) and (3.8) (of Chapter I) and there is thus a unique Markov process with these rates by his Theorem I.3.9. See also Proposition 1.7 in the lecture notes by Grosskinsky [49]. □

PROOF OF THEOREM 2.2.7. This is routine consequence of the above results. Since the characterisation of the limit points defines a unique Markov process on $\{0,1\}^{\mathbf{Z}} \times \mathbf{Z}$, all weakly convergence subsequences converge to the same limit. It follows (see eg Durrett [31], Lemma 8.3.3) that the full sequence converges in distribution to the same limit. □

PROOF OF PROPOSITION 2.2.6. Set $U_0^h := 0$ and define inductively, for $k \geq 1$,

$$T_k^h := \inf\{U_{k-1}^h < t < T : X_t^h \neq Y_t^h\} \wedge T,$$

$$U_k^h := \inf\{T_k^h < t < T : X_t^h = Y_t^h\} \wedge T,$$

the successive decoupling and coalescence times. Here, we adopt the convention that $\inf \emptyset = \infty$. We also define

$$K^h := \sup\{k : T_k^h < T\},$$

the number of decouplings before time T. Observe that $1\{X_t^h \neq Y_t^h\} = 1$ if and only if $T_k^h < t < U_k^h$ for some $k \leq K^h$ and, in particular, for any $m < \infty$,

$$\mathbf{P}\left[\int_0^T 1\{X_t^h \neq Y_t^h\}\,\mathrm{d}t > \varepsilon\right] = \mathbf{P}\left[\sum_{k=1}^{K^h}(U_k^h - T_k^h) > \varepsilon\right].$$

Consider what has happened at the time of a decoupling $t = T_k^h$ for some $k \leq K^h$. Just before t, $X_{t-}^h = Y_{t-}^h$ and $V_t^h(X_{t-}^h, 1) = (0,0,0)$ but $V_t^h(X_{t-}^h, 2)$ contains a one. As there is a decoupling, both processes Y^h and X^h jump in the direction of this good site (possibly one of two on either side of its current position); let us assume without loss of generality that this direction is to the right of its position just before T_k^h. At time T_k^h we are in the

following setting

$$V^h_{T^h_k}(Y^h_{T^h_k}, 1) = (0, 0, 1), \qquad X^h_{T^h_k} = Y^h_{T^h_k} + 1.$$

In the next discrete generation, with probability $1 - O(h)$, the states of the habitats at the location of $Y^h_{T^h_k}$ and its two nearest neighbours remain constant, $X^h_{T^h_k}$ remains stationary, and $Y^h_{T^h_k}$ jumps one step to the right, to the neighbouring good habitat and location of X^h. In this case $U^h_k = T^h_k + h$. All events where $U^h_k - T^h_k > h$ have probability $O(h)$.

Now trivially, since a decoupling can only occur at one of the times that X^h jumps, we have that $K^h \le J^h$ almost surely, where J^h is the number of times that X^h jumps before T defined in the proof of Lemma 2.5.2. Then, for any $m < \infty$ and $h < \varepsilon/m$,

$$\begin{aligned}
\mathbf{P}\Big[\int_0^T 1\{X^h_t \neq Y^h_t\}\,\mathrm{d}t > \varepsilon\Big] &\le \mathbf{P}\Big[\sum_{k=1}^{J^h}(U^h_k - T^h_k) > \varepsilon\Big] \\
&= \mathbf{P}\Big[\sum_{k=1}^{J^h}(U^h_k - T^h_k) > \varepsilon \;\Big|\; J^h \le m\Big]\mathbf{P}[J^h \le m] \\
&\quad + \mathbf{P}\Big[\sum_{k=1}^{J^h}(U^h_k - T^h_k) > \varepsilon \;\Big|\; J^h > m\Big]\mathbf{P}[J^h > m] \\
&\le \mathbf{P}\Big[\sum_{k=1}^{m}(U^h_k - T^h_k) > \varepsilon\Big] + \mathbf{P}[J^h > m] \\
&\le \sum_{k=1}^{m}\mathbf{P}[U^h_k - T^h_k > \varepsilon/m] + \frac{\sup_{h>0}\mathbf{E}[J^h]}{m} \\
&\le Cmh + \frac{\sup_{h>0}\mathbf{E}[J^h]}{m}.
\end{aligned}$$

The penultimate line follows by noticing that a sum of m positive terms is greater than ε only if at least one of them is greater than ε/m and then applying the union bound; the second term in this line follows from Markov's inequality. The last line uses that $\mathbf{P}[U^h_k - T^h_k > \varepsilon/m] \le \mathbf{P}[U^h_k - T^h_k > h] = O(h)$. Taking the limit $h \to 0$ we see that

$$\lim_{h\to 0}\mathbf{P}\Big[\int_0^T 1\{X^h_t \neq Y^h_t\}\,\mathrm{d}t > \varepsilon\Big] \le \frac{\sup_{h>0}\mathbf{E}[J^h]}{m}$$

for every fixed m. Since the denominator is finite, as proved in the proof of Lemma 2.5.2, the statement of the proposition follows trivially. □

2.6. Proof of Proposition 2.3.1.

LEMMA 2.6.1. *The sequence of measures $\{\pi^n\}$ on $D(\mathbf{R}, \{0, 1\})$ is tight.*

PROOF. Kurtz's tightness criterion (Theorem 2.5.1) extends to right-continuous paths on $\mathbf{R}$ in the obvious way. $E = \{0, 1\}$ and so the first condition holds trivially. Denote by $\mathbf{E}_{\pi^n}$ the expectation over $\omega \in \{0,1\}^{\mathbf{Z}/n} \subset D(\mathbf{R}, \{0,1\})$ with respect to the measure π^n. Then

$$\mathbf{E}_{\pi^n}[|\omega(x + 1/n) - \omega(x)| \,|\, \omega(x)] \leq \frac{v_1 + v_2}{n},$$

and so, applying the tower property for conditional expectation at each site $x \in \mathbf{Z}/n$,

$$\mathbf{E}_{\pi^n}[|\omega(x + y) - \omega(x)| \,|\, \mathcal{F}_x] \leq \lfloor ny \rfloor \frac{v_1 + v_2}{n} \leq y(v_1 + v_2).$$

It thus follows that the second condition of Theorem 2.5.1 is satisfied and the sequence π^n is tight. □

PROOF OF PROPOSITION 2.3.1. The sequence π^n is tight and thus there exists a subsequence converging to weakly to a measure π on $D(\mathbf{R}, \{0,1\})$. Let us again denote by n this subsequence so that we may write $\pi^n \Rightarrow \pi$. We proceed to characterise π and prove that such a π is unique.

Since convergence in distribution on $D(\mathbf{R}, \{0,1\})$ implies convergence of the finite-dimensional distributions (see eg Theorem 3.7.8 of Ethier and Kurtz [40]),

$$\pi(\omega(0) = 1) = \lim_{n\to\infty} \pi^n(\omega(0) = 1) = \lim_{n\to\infty} \frac{p_0^n}{p_0^n + p_1^n} = \frac{v_0}{v_0 + v_1}.$$

Furthermore, we also have convergence of the two-dimensional distributions and therefore

$$\begin{aligned}\pi(\omega(x + \mathrm{d}x) = 1 \,|\, \omega(x) = 0) &= \lim_{n\to\infty} \pi^n(\omega(x + \mathrm{d}x) = 1 \,|\, \omega(x) = 0) \\ &= \sum_{k=1}^{\lfloor n\,\mathrm{d}x \rfloor} \left(1 - \frac{v_0}{n}\right)^{k-1} \frac{v_0}{n} \left(1 - \frac{v_1}{n}\right)^{\lfloor n\,\mathrm{d}x \rfloor - k} + O((\mathrm{d}x)^2).\end{aligned}$$

It is straightforward to check via a calculation almost identical to that in (2.5.6) that

$$\lim_{\mathrm{d}x\to 0} \frac{1}{\mathrm{d}x} \pi(\omega(x + \mathrm{d}x) = 1 \,|\, \omega(x) = 0) = v_0.$$

Via the same method, one can show that

$$\lim_{\mathrm{d}x\to 0} \frac{1}{\mathrm{d}x} \pi(\omega(x + \mathrm{d}x) = 0 \,|\, \omega(x) = 1) = v_1.$$

Such a measure π on $D(\mathbf{R}, \{0,1\})$ corresponds to the law of a Markov process initially equal to 1 with probability $v_0/(v_1 + v_0)$ which jumps from 0 to 1 at rate v_0 and from 1 to 0 at rate v_1 and is thus unique. We therefore conclude that $\pi^n \Rightarrow \pi$. □

2.7. Proof of Theorem 2.3.3.

The following is due to Kallenberg [70], Theorem 4.11, specialised to the simpler setting of $\mathbf{R}^d$.

THEOREM 2.7.1. *A sequence of random measures $P^n \Rightarrow P$ in $\mathcal{M}(\mathbf{R}^d)$ equipped with the vague topology if and only if*

$$\mathbf{E}[e^{-\int f \,\mathrm{d}P^n}] \to \mathbf{E}[e^{-\int f \,\mathrm{d}P}]$$

for all $f \in C_c^+(\mathbf{R}^d)$, the space of real-valued non-negative continuous functions with compact support.

As a consequence of this theorem, it is relatively straightforward to prove Theorem 2.3.3.

PROOF OF THEOREM 2.3.3. The distribution of a Poisson random measure P on the measurable space $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}))$ is uniquely determined by its Laplace functional

$$\mathbf{E}[e^{-\int f \,\mathrm{d}P}]$$

for all $f \in C_c^+(\mathbf{R}^d)$. This is because the distribution of $\int f \,\mathrm{d}P$ is thus determined for all such f, which is a separating class of functions, because $\theta f \in C_c^+(\mathbf{R}^d)$ for all $\theta > 0$. Furthermore, if P is a Poisson random measure with mean m then for each $f \in C_c^+(\mathbf{R}^d)$

$$\mathbf{E}[e^{-\int f \,\mathrm{d}P}] = e^{-\int 1-e^{-f} \,\mathrm{d}m}.$$

This result is given, for example, in the appendix of Kurtz and Rodrigues [84] and follows by an approximation of f by simple functions. Since $f \in C_c^+(\mathbf{R}^d)$ implies that $1 - e^{-f} \in C_c^+(\mathbf{R}^d)$, applying the previous theorem we see that it is sufficient to prove that the sequence of intensity measures m_n converges vaguely to m. For $f \in C_c^+(\mathbf{R}^2)$ fixed, denote by $[r, s) \times [a, b) \subset \mathbf{R}_+ \times \mathbf{R}$ a compact set outside which $f = 0$. Then

$$\begin{aligned}\iint_{\mathbf{R}_+\times\mathbf{R}} f(x, y) m_n(\mathrm{d}t, \mathrm{d}x) &= \frac{1}{n} \frac{\gamma_b v_0 v_1}{1 - v_0/n} \iint_{\mathbf{R}_+\times\mathbf{R}} f(t, x) \,\mathrm{d}t \otimes \sum_{x' \in \mathbf{Z}/n} \delta_{x'}(\mathrm{d}x) \\ &\leq \frac{\gamma_b v_0 v_1}{1 - v_0/n} (s - r) \frac{\lfloor n(b - a) \rfloor}{n} \| f \|_\infty,\end{aligned}$$

which is finite. Thus, by the Fubini–Tonelli Theorem,

$$\begin{aligned}\iint_{\mathbf{R}_+\times\mathbf{R}} f(t,x)m_n(\mathrm{d}t,\mathrm{d}x) &= \frac{1}{n}\frac{\gamma_b v_0 v_1}{1-v_0/n}\iint_{\mathbf{R}_+\times\mathbf{R}} f(t,x)\,\mathrm{d}t\otimes\sum_{x'\in\mathbf{Z}/n}\delta_{x'}(\mathrm{d}x)\\ &= \frac{\gamma_b v_0 v_1}{1-v_0/n}\int_{\mathbf{R}_+}\frac{1}{n}\sum_{x\in\mathbf{Z}/n} f(t,x)\,\mathrm{d}t.\end{aligned} \tag{2.7.1}$$

For each t, the function $x\mapsto f(t,x)$ is continuous on a compact subset of $\mathbf{R}$ and is thus Riemann integrable. We identify the sum inside the integral as a Riemann sum and therefore

$$\frac{1}{n}\sum_{x\in\mathbf{Z}/n} f(t,x)\to\int_{\mathbf{R}} f(t,x)\,\mathrm{d}x$$

as $n\to\infty$. Furthermore,

$$\frac{1}{n}\sum_{x\in\mathbf{Z}/n} f(t,x)\le\frac{\lfloor n(b-a)\rfloor}{n}1_{[r,s)}(t)\|f\|_\infty\le(b-a)\|f\|_\infty 1_{[r,s)}(t),$$

which is integrable over $\mathbf{R}_+$. Then, by the dominated convergence theorem, we may pass a limit $n\to\infty$ through the integral in (2.7.1) and obtain

$$\iint_{\mathbf{R}_+\times\mathbf{R}} f(t,x)m_n(\mathrm{d}t,\mathrm{d}x)\to\gamma_b v_0 v_1\int_{\mathbf{R}_+}\int_{\mathbf{R}} f(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

A simple application of Fubini's Theorem yields that

$$\gamma_b v_0 v_1\int_{\mathbf{R}_+}\int_{\mathbf{R}} f(t,x)\,\mathrm{d}x\,\mathrm{d}t=\iint_{\mathbf{R}_+\times\mathbf{R}} f(t,x)m(\mathrm{d}t,\mathrm{d}x),$$

completing the proof. □

2.8. Proof of Theorem 2.3.4.

We prove the theorem in a number of steps. First, we will prove the convergence of the patches ignoring any interactions with neighbouring patches. The law of the complete system can then be constructed by superimposing the behaviour of additional patches: until the first time that these patches meet the boundaries of our focal patch, the focal patch has the same law as the former. We prove that this stopping time converges to infinity, so that we have full convergence.

As a first step, we will consider the behaviour of the focal patch ignoring the interactions with neighbouring patches and also ignoring, for now, the fact that it may be removed with some positive probability when it has width zero. These effects will be reintroduced

later on, in turn. In order to avoid increasingly terse notation, we continue to denote the boundaries of this modified patch by L^n and R^n, its midpoint by M^n and width by W^n; it will be made clear when we return to considering the full dynamics of the process. This modification of the process behaves as follows. The right-hand boundary jumps a distance $1/n$ to the right at rate $n^2(1 - v_1/n)$ and to the left at rate $n^2(1 - v_0/n)$ (except at times when $R^n = L^n$, when left jumps are forbidden); the left-hand boundary jumps a distance $1/n$ to the left at rate $n^2(1 - v_1/n)$ and to the right at rate $n^2(1 - v_0/n)$ (except at times when $R^n = L^n$, when right jumps are forbidden).

2.8.1. The one-dimensional Skorokhod problem. Our first step is to prove that the modified width process W^n converges to a diffusion process reflected at the origin. Our techniques are inspired by those of Iglehart [64], Iglehart and Whitt [65] and Nappo and Torti [99] who, amongst other more technical results, prove convergence of the length of a single-server queue with exponential arrival and service time distribution to reflected Brownian motion. The key to their proofs is to recast the queue length as a solution to a *Skorokhod problem*, which we now define.

The following definition, a refinement but more frequently stated version of the original due to Skorokhod [116, 117], appears in Tanaka [124].

DEFINITION 2.8.1. Let $f \in D([0,\infty), \mathbf{R})$ with $f(0) \geq 0$. We say that a pair $(x, \ell) \in D([0,\infty), \mathbf{R}) \times D([0,\infty), \mathbf{R})$ is a *solution to the one-dimensional Skorokhod problem for* f if for all $t \geq 0$, $x(t) \geq 0$ and

$$x(t) = f(t) + \ell(t),$$

and ℓ is non-decreasing, $l(0) = 0$ and

$$\int_0^\infty 1\{x(s) > 0\}\, \mathrm{d}\ell(s) = 0.$$

The last condition is equivalent to saying that the function ℓ may increase only on the set of times t when $x(t) = 0$. A major result in the theory of diffusion processes with boundary conditions is that the one-dimensional Skorokhod problem has a unique solution. This result is given for completeness.

THEOREM 2.8.2. *For each $f \in D([0,\infty), \mathbf{R})$ with $f(0) \geq 0$, there exists a unique solution (x, ℓ) to the one-dimensional Skorokhod problem for f given by*

$$\ell(t) := 0 \vee \big(- \inf_{s \leq t} f(s) \big)$$

and $x(t) = f(t) + \ell(t)$.

The key observation of [64, 65, 99] is that if $(S^n)_{n\geq 0}$ is a sequence of processes in $D([0,\infty),\mathbf{R})$ and $S^n \Rightarrow S$, then the sequence of solutions to the associated Skorokhod problem for S^n converge. We summarise this result in the following corollary.

COROLLARY 2.8.3. *Suppose that $S^n \Rightarrow S$ in $D([0,\infty),\mathbf{R})$ and (X^n,ℓ^n) is the unique solution to the one-dimensional Skorokhod problem for S^n. Then $(X^n,\ell^n) \Rightarrow (X,\ell)$ in $D([0,\infty),\mathbf{R}^2)$ where (X,ℓ) is the solution to the Skorokhod problem for S.*

PROOF. The key observation is to notice that the map $\Phi : f \mapsto (x,\ell)$ given by mapping $f \in D([0,\infty),\mathbf{R})$ to the unique solution (x,ℓ) to the Skorokhod problem for f is continuous. The result is thus a consequence of the continuous mapping theorem (see eg Theorem 1.2.7 of Billingsley [11]). □

As an illustration of the technique (although we note that the following will not be sufficient for our purposes) we show that the convergence of the width process W^n is almost immediate.

LEMMA 2.8.4. *Suppose that $(L_0^n,R_0^n) \Rightarrow (L_0,R_0)$ as $n\to\infty$ so that $W_0^n = R_0^n - L_0^n \Rightarrow R_0 - L_0$. Let B^W denote a standard Brownian motion on $\mathbf{R}$ and let S be the diffusion process defined by*

$$S_t := R_0 - L_0 - 2v_1 t + 2B_t^W$$

Then as $n\to\infty$ and $v_0 \to 0$, $W^n \Rightarrow W$ where (W,ℓ) is the unique solution to the Skorokhod problem for S. The convergence is in the space $D([0,\infty),\mathbf{R}^2)$ equipped with the topology of uniform convergence on compact sets.

PROOF. The width process W^n behaves as follows. If $W^n \neq 0$ then it jumps a distance $1/n$ to the right at rate $2n^2(1-v_1/n)$ and to the left at rate $2n^2(1-v_0/n)$. If $W^n = 0$ then it jumps a distance $1/n$ to the right at rate $2n^2(1-v_1/n)$. Denote by S^n a process with $S_0^n = W_0^n$ which subsequently jumps a distance $1/n$ to the right at rate $2n^2(1-v_1/n)$ and to the left at rate $2n^2(1-v_0/n)$ and let

$$\ell_t^n := 0 \vee \big(-\inf_{s\leq t} S_s^n\big).$$

Then $W^n = S^n + \ell^n$; in particular, (W^n,ℓ^n) is the solution the Skorokhod problem for S^n. It therefore follows from Corollary 2.8.3 that if $S^n \Rightarrow S$ then $W^n \Rightarrow W$ in $D([0,\infty),\mathbf{R})$ as defined in the statement of the lemma. This follows immediately from the Stroock–Varadhan Theorem for the convergence of Markov chains to diffusions (see the version

stated in Section 8.8.7 of Durrett [31]), noting for the process S^n that

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[S_t^n - S_0^n \mid S_0^n = x] = -2v_1,$$

and

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[(S_t^n - S_0^n)^2 \mid S_0^n = x] = 4.$$

It remains to argue that W^n converges with respect to the topology of uniform convergence on compact sets. Since S is continuous almost surely, so too is W by Theorem 2.8.2. Thus we may apply Lemma 3.10.1 of Ethier and Kurtz [40], which states that weak convergence in $D([0,\infty), \mathbf{R})$ to a continuous limit implies the convergence in $D([0,\infty), \mathbf{R})$ equipped with the topology of uniform convergence on compact sets. □

2.8.2. Convergence of (M^n, W^n). The previous result is insufficient for our proof of Theorem 2.3.4 since it does not prove the joint convergence of the width and midpoint processes. Furthermore, it does not retain any information about the amount of time that the process W^n spends at zero which we will need in order to prove the convergence of the time at which the patch is removed. Note that although the process $\ell_t^n = 0 \wedge (-\inf_{s\le t} S_s^n)$ may refer to the *local time* of W^n at 0 in an abstract sense, we stress that it does not in fact coincide with a more natural definition of the *local time* which we define by

$$C_t^n := \int_0^t 1\{W_s^n = 0\}\,\mathrm{d}s.$$

This distinction was encountered in Nappo and Torti [99], where the authors prove a convergence theorem for the idle time of a Markov single-serve queue to the local time of reflected Brownian motion. However, their methods are purely indicative, referencing lecture notes of Thomas G. Kurtz. In these works, the authors observe that for any sequence $\alpha_n > 0$, the sequence of processes $\alpha_n C^n$ is non-decreasing with $\alpha_n C_0^n = 0$. Moreover, it evidently increases only on the set of times t when $W_t^n = 0$. In particular, the pair $(W^n, \alpha_n C^n)$ is the unique solution to the Skorokhod problem for a process $\hat{S}^n$ defined by $\hat{S}^n := W^n - \alpha_n C^n$. In the setting of the single-server queue with exponential arrival and service time distributions, Nappo and Torti [99] omit the construction of the appropriate choice of the sequence α_n; Thomas G. Kurtz, in the referenced lecture notes, constructs this sequence but does not prove the convergence of the sequence of processes $\hat{S}^n$. In what follows, we construct the corresponding sequence of processes $\hat{S}^n$ in our more general setting and prove that it converges to S defined in Lemma 2.8.4. This enables us to argue the joint convergence of $(M^n, W^n, \alpha_n C^n)$ for an appropriate choice of α_n.

We begin by constructing a process $\hat{S}^n$ for which $(W^n, \alpha_n C^n)$ is the solution to the

associated Skorokhod problem. Denote by Π_+^R, Π_-^R, Π_+^L and Π_-^L four independent Poisson processes with of unit rate. The joint process (M^n, W^n) may be expressed as a solution to the system stochastic equations

$$\begin{aligned} W_t^n = W_0^n &+ \frac{1}{n}\Pi_+^R\big(n^2(1-v_1/n)t\big) \\ &+ \frac{1}{n}\Pi_+^L\big(n^2(1-v_1/n)t\big) \\ &- \frac{1}{n}\Pi_-^R\bigg(n^2(1-v_0/n)\int_0^t 1\{W_s^n > 0\}\,\mathrm{d}s\bigg) \\ &- \frac{1}{n}\Pi_-^L\bigg(n^2(1-v_0/n)\int_0^t 1\{W_s^n > 0\}\,\mathrm{d}s\bigg), \end{aligned}$$

and

$$\begin{aligned} M_t^n = M_0^n &+ \frac{1}{2n}\Pi_+^R\big(n^2(1-v_1/n)t\big) \\ &- \frac{1}{2n}\Pi_+^L\big(n^2(1-v_1/n)t\big) \\ &- \frac{1}{2n}\Pi_-^R\bigg(n^2(1-v_0/n)\int_0^t 1\{W_s^n > 0\}\,\mathrm{d}s\bigg) \\ &+ \frac{1}{2n}\Pi_-^L\bigg(n^2(1-v_0/n)\int_0^t 1\{W_s^n > 0\}\,\mathrm{d}s\bigg). \end{aligned} \tag{2.8.1}$$

In each of these expressions, the first term involving a Poisson process corresponds to the jumps associated with a right boundary expansion; the second to a left boundary expansion; the third to a right boundary contraction; and the fourth to a left boundary contraction. For a Poisson process Π, denote by $\tilde{\Pi}$ the compensated process defined by $\tilde{\Pi}(t) := \Pi(t) - t$. Then we may further express (M^n, W^n) as the solution to

$$\begin{aligned} W_t^n = W_0^n &+ 2(v_0 - v_1)t \\ &+ \frac{1}{n}\tilde{\Pi}_+^R\big(n^2(1-v_1/n)t\big) \\ &+ \frac{1}{n}\tilde{\Pi}_+^L\big(n^2(1-v_1/n)t\big) \\ &- \frac{1}{n}\tilde{\Pi}_-^R\big(n^2(1-v_0/n)(t - C_t^n)\big) \\ &- \frac{1}{n}\tilde{\Pi}_-^L\big(n^2(1-v_0/n)(t - C_t^n)\big) \\ &+ 2n(1-v_0/n)C_t^n, \end{aligned} \tag{2.8.2}$$

and

$$\begin{aligned} M_t^n = M_0^n &+ \frac{1}{2n}\tilde{\Pi}_+^R\big(n^2(1-v_1/n)t\big) \\ &- \frac{1}{2n}\tilde{\Pi}_+^L\big(n^2(1-v_1/n)t\big) \\ &- \frac{1}{2n}\tilde{\Pi}_-^R\big(n^2(1-v_0/n)(t-C_t^n)\big) \\ &+ \frac{1}{2n}\tilde{\Pi}_-^L\big(n^2(1-v_0/n)(t-C_t^n)\big). \end{aligned} \tag{2.8.3}$$

In particular, the pair $(W^n, 2n(1-v_0/n)C^n)$ is the unique solution to the Skorokhod problem for $\hat{S}^n$ defined by $\hat{S}^n := W^n - 2n(1-v_0/n)C^n$. We now prove that $\hat{S}^n$ is weakly convergent to the process S defined in Lemma 2.8.4. In order to do so, we need some auxiliary results on the convergence of the centred and rescaled Poisson processes and the process C^n.

LEMMA 2.8.5. *Let B denote a standard Brownian motion, Π a Poisson process of unit rate and $\tilde{\Pi}$ the compensated process defined by $\tilde{\Pi}(t) = \Pi(t) - t$. Let a_n, b_n be constants such that $a_n \to \infty$ and $a_n/b_n^2 \to 1$. Then as $n \to \infty$,*

$$(\tilde{\Pi}(a_n t)/b_n)_{t\geq 0} \Rightarrow B$$

in $D([0,\infty), \mathbf{R})$ equipped with the topology of uniform convergence on compact sets.

PROOF. The process $Y_t^n := \tilde{\Pi}(a_n t)/b_n$ is a martingale. Its jumps are trivially bounded by b_n^{-1} and so vanish and, furthermore,

$$[Y^n]_t = \Pi(a_n t)/b_n^2 \to t$$

for every $t \geq 0$ by the strong law of large numbers. The convergence in $D([0,\infty), \mathbf{R})$ equipped with the Skorokhod topology then follows from the martingale central limit theorem (see eg Section 7.1 of Ethier and Kurtz [40]). Extending the result to convergence in the topology of uniform convergence on compact sets follows because Brownian motion is continuous and Lemma 3.10.1 of [40]. □

Next we will prove that the local time process $C^n \Rightarrow 0$. The process increases only when $W^n = 0$ and, for each arrival of W^n to zero, we add an exponentially distributed amount of time with parameter $2n^2(1-v_1/n)$ to C^n. We therefore need control on the number of visits to zero and we expect it to be sufficient to prove that this number is $o(n^2)$. Our method involves defining a renewal process counting the number of arrivals of W^n at the origin.

Unfortunately this necessitates the introduction of yet more notation. For each $n \in \mathbf{N}$ let $\tau_0^n := 0$ and $\theta_0^n := 0$ and define inductively, for each $k \geq 1$,

$$\tau_k^n := \inf\{t > \theta_{k-1}^n : W_t^n = 0\},$$
$$\theta_k^n := \inf\{t > \tau_k^n : W_t^n > 0\},$$

the successive arrival and departure times of W^n at the origin. We will adopt the convention that $\inf \emptyset = \infty$. We then define the renewal process counting the number of arrivals at the origin by

$$N_t^n := \sup\{k \geq 0 : \tau_k^n \leq t\}.$$

The successive inter-arrival times of N^n, given by $\tau_{k+1}^n - \tau_k^n$, are independent and may be decomposed as the sum of $\theta_k^n - \tau_k^n$, the exponential waiting time at the origin; and $\tau_{k+1}^n - \theta_k^n$, the time it takes W^n to return to the origin starting from $1/n$.

LEMMA 2.8.6. *For $k \geq 1$, the sequence of random variables $\tau_{k+1}^n - \theta_k^n$ are independent and identically distributed with mean*

$$\mathbf{E}[\tau_2^n - \theta_1^n] = \frac{1}{2n(v_1 - v_0)}$$

and variance

$$\mathrm{var}[\tau_2^n - \theta_1^n] = \frac{1}{2n(v_1 - v_0)^3} + o(1/n^2).$$

PROOF. Recall from the proof of Lemma 2.8.4 the definition of the S^n which jumps a distance $1/n$ to the right at rate $2n^2(1 - v_1/n)$ and to the left at rate $2n^2(1 - v_0/n)$. Conditional on starting at $1/n$, the distributions of W^n and S^n coincide until the time of the first arrival at the origin. In particular, the distribution of $\tau_2^n - \theta_1^n$ coincides with that of the time it takes S^n to pass from $1/n$ to 0.

We may write

$$S_t^n = \frac{1}{n} V_{\Pi_t^n}^n$$

where V^n is a discrete-time simple random walk on $\mathbf{Z}$ jumping to the right with probability $r_n = (1 - v_1/n)/((1 - v_1/n) + (1 - v_0/n))$ and to the left with probability $l_n = 1 - r_n$; and Π^n is a Poisson process with rate parameter $2n^2((1 - v_1/n) + (1 - v_0/n))$. Denoting by $H_0^n = \inf\{k \geq 0 : V_k^n = 0\}$, and by A_i^n a collection of independent and identically distributed exponential random variables with parameter $2n^2((1 - v_1/n) + (1 - v_0/n))$, conditional on $V_0^n = 1$,

$$\tau_2^n - \theta_1^n \stackrel{d}{=} \sum_{k=1}^{H_0^n} A_i^n.$$

In particular,

$$\mathbf{E}[\tau_2^n - \theta_1^n] = \mathbf{E}[H_0^n \mid V_0^n = 1]\mathbf{E}[A_i^n] \tag{2.8.4}$$

and

$$\mathrm{var}[\tau_2^n - \theta_1^n] = \mathrm{var}[H_0^n \mid V_0^n = 1]\mathbf{E}[A_i^n]^2 + \mathbf{E}[H_0^n \mid V_0^n = 1]\mathrm{var}[A_i^n]. \tag{2.8.5}$$

Referring to Proposition 2.10.1 in the appendix to this chapter,

$$\mathbf{E}[H_0^n \mid V_0^n = 1] = \frac{1}{l_n - r_n}$$

and

$$\mathrm{var}[H_0^n \mid V_0^n = 1] = \frac{4l_n r_n}{(l_n - r_n)^3}.$$

Substituting these, along with the expressions for l_n and r_n, into equations (2.8.4) and (2.8.5) gives the result. □

We wish to use Lemma 2.8.6 to bound $\mathbf{E}[N^n(t)]$, ie the expected number of arrivals at the origin by time t. We will make use of the following inequality of Lorden (Theorem 1 in [91]).

LEMMA 2.8.7. *Let $(N_t)_{t\geq 0}$ denote a renewal process with independent inter-renewal times $(Z_k)_{k\geq 1}$ so that $N(t) = \inf\{k \geq 1 : Z_1 + \cdots + Z_k > t\}$. Then*

$$\mathbf{E}[N_t] \leq \frac{t}{\mathbf{E}[Z_1]} + \frac{\mathbf{E}[Z_1^2]}{\mathbf{E}[Z_1]^2}.$$

LEMMA 2.8.8. *There exists a constant $C(t, v_1) < \infty$ such that*

$$\limsup_{\substack{n\to\infty \\ v_0\to 0}} \frac{\mathbf{E}[N_t^n \mid W_0^n = 0]}{n} \leq C(t, v_1).$$

PROOF. Observe that conditional on $W_0^n = 0$, $\tau_1^n = 0$. Applying Lemma 2.8.7 to the renewal process N^n gives that

$$\mathbf{E}[N_t^n] \leq \frac{t}{\mathbf{E}[\tau_2^n - \tau_1^n]} + \frac{\mathrm{var}[\tau_2^n - \tau_1^n]}{\mathbf{E}[\tau_2^n - \tau_1^n]^2} + 1.$$

We may decompose the inter-arrival times as $\tau_2^n - \tau_1^n = (\tau_2^n - \theta_1^n) + (\theta_1^n - \tau_1^n)$ where $\tau_2^n - \theta_1^n$ and $\theta_1^n - \tau_1^n$ are independent. Recall that $\theta_1^n - \tau_1^n$ is the waiting time of W^n at zero and is thus exponentially distributed with parameter $2n^2(1 - v_1/n)$. In particular,

$$\mathbf{E}[N_t] \leq \frac{t}{\mathbf{E}[\tau_2^n - \theta_1^n] + \mathbf{E}[\theta_1^n - \tau_1^n]} + \frac{\mathrm{var}[\tau_2^n - \theta_1^n]}{\mathbf{E}[\tau_2^n - \theta_1^n]^2} + \frac{\mathrm{var}[\theta_1^n - \tau_1^n]}{\mathbf{E}[\theta_1^n - \tau_1^n]^2} + 1.$$

Since $\theta_1^n - \tau_1^n$ is exponentially distributed, the third term on the right hand side is equal to 1. Applying Lemma 2.8.6,

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \frac{1}{n} \frac{t}{\mathbf{E}[\tau_2^n - \theta_1^n] + \mathbf{E}[\theta_1^n - \tau_1^n]} = 2v_1 t,$$

and

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \frac{1}{n} \frac{\mathrm{var}[\tau_2^n - \theta_1^n]}{\mathbf{E}[\tau_2^n - \theta_1^n]^2} = \lim_{\substack{n\to\infty \\ v_0\to 0}} \frac{1}{n} \frac{\frac{1}{2n(v_1-v_0)^3} + o(1/n^2)}{\frac{1}{4n^2(v_1-v_0)^2}} = \frac{2}{v_1}.$$

The result follows. □

Given all of our work, it is now rather straightforward to prove the following.

LEMMA 2.8.9. *As $n \to \infty$ of $C^n \Rightarrow 0$ in $D([0,\infty), \mathbf{R})$.*

PROOF. Observe that $t \mapsto C_t^n$ is continuous, non-decreasing and has gradient bounded above by 1. In particular, the sequence $(C^n)_{n\in\mathbf{N}}$ is tight by Theorem 2.5.1. It will be sufficient to show that for each $t > 0$

$$\limsup_{\substack{n\to\infty \\ v_0\to 0}} \mathbf{E}\left[\sup_{s\le t} |C_s^n| \right] = 0$$

for then the finite-dimensional distributions converge to that of a process which is constant and equal to zero at all times.

The process C^n can be rewritten as

$$C_t^n = \int_0^t 1\{W_s^n = 0\}\, \mathrm{d}s = \sum_{k=1}^{N_t^n} (\theta_k^n - \tau_k^n).$$

Moreover, since C^n is positive and non-decreasing

$$\mathbf{E}\left[\sup_{s\le t} |C_s^n| \right] = \mathbf{E}\left[\sum_{k=1}^{N_t^n} (\theta_k^n - \tau_k^n) \right] \le \mathbf{E}\left[\sum_{k=1}^{N_t^n} (\theta_k^n - \tau_k^n) \,\middle|\, W_0^n = 0 \right].$$

The last inequality is trivial noting that the total amount of time W^n has spent in zero by time t is maximised when $W_0^n = 0$. Observe that N_t^n is a stopping time with respect to the filtration $\mathcal{F}_k^n := \sigma(\{\tau_{j+1}^n - \theta_j^n, \theta_j^n - \tau_j^n : j < k\})$ and $\theta_k^n - \tau_k^n$ is independent of $\mathcal{F}_{k-1}^n$. Then, since $\mathbf{E}[N_t^n \mid W_0^n = 0] < \infty$ by Lemma 2.8.8, we may apply Wald's Lemma to the right-most expectation and obtain that

$$\mathbf{E}\left[\sup_{s\le t} |C_s^n| \right] \le \mathbf{E}[N_t^n \mid W_0^n = 0]\mathbf{E}[\theta_1^n - \tau_1^n] = \frac{\mathbf{E}[N_t^n \mid W_0^n = 0]}{2n^2(1 - v_1/n)}.$$

Applying Lemma 2.8.8, the result follows. □

Given the convergence of the centred and rescaled Poisson processes to Brownian motion and of the local time process C^n to zero, it seems natural to pass to the limit in equations (2.8.2) and (2.8.3) to obtain limiting equations for (M^n, W^n). In order to do so, we require one further lemma reproduced from Billingsley [11], p. 151.

LEMMA 2.8.10. *Suppose that for $n \in \mathbf{N}$, $\Phi^n, \Phi \in D([0,\infty), \mathbf{R})$ are non-decreasing. Suppose further that $X^n, X \in D([0,\infty), \mathbf{R})$. If $(X^n, \Phi^n) \Rightarrow (X, \Phi)$ and X is continuous almost surely, then $X^n \circ \Phi^n \Rightarrow X \circ \Phi$.*

THEOREM 2.8.11. *Suppose that $(L_0^n, R_0^n) \Rightarrow (L_0, R_0)$ as $n \to \infty$. Let B^M and B^W denote a pair of independent standard Brownian motions on $\mathbf{R}$ and let M and S be the diffusion processes defined by*

$$M_t := \frac{1}{2}(L_0 + R_0) + B_t^M,$$

and

$$S_t := R_0 - L_0 - 2v_1 t + 2B_t^W.$$

Then as $n \to \infty$ and $v_0 \to 0$, $(M^n, W^n, 2nC^n) \Rightarrow (M, W, \ell)$ where (W, ℓ) is the unique solution to the Skorokhod problem for S. Moreover, M and (W, ℓ) are independent. The convergence is in the space $D([0,\infty), \mathbf{R}^3)$ equipped with the topology of uniform convergence on compact sets.

PROOF. Recall that $(W^n, 2n(1 - v_0/n)C^n)$ is the solution to the Skorokhod problem for $\hat{S}^n$ defined by

$$\begin{aligned} \hat{S}_t^n = W_0^n &+ 2(v_0 - v_1)t \\ &+ \frac{1}{n}\tilde{\Pi}_+^R\big(n^2(1 - v_1/n)t\big) \\ &+ \frac{1}{n}\tilde{\Pi}_+^L\big(n^2(1 - v_1/n)t\big) \\ &- \frac{1}{n}\tilde{\Pi}_-^R\big(n^2(1 - v_0/n)(t - C_t^n)\big) \\ &- \frac{1}{n}\tilde{\Pi}_-^L\big(n^2(1 - v_0/n)(t - C_t^n)\big). \end{aligned} \tag{2.8.6}$$

Applying Lemma 2.8.5, Lemma 2.8.9 and Lemma 2.8.10, passing to the limit $n \to \infty$ and $v_0 \to 0$ in equations (2.8.3) and (2.8.6) one obtains that $(\hat{S}^n, M^n) \Rightarrow (\hat{S}, M)$ where $(\hat{S}, M)$ is given by the coupled system

$$\hat{S}_t = W_0 - v_1 t + B_+^R(t) + B_+^L(t) - B_-^R(t) - B_-^L(t),$$

and

$$M_t = M_0 + \frac{1}{2}B_+^R(t) - \frac{1}{2}B_+^L(t) - \frac{1}{2}B_-^R(t) + \frac{1}{2}B_-^L(t),$$

where B_+^R, B_-^R, B_+^L and B_-^L are four independent standard Brownian motions. It is straightforward that $B^R := (B_+^R - B_-^R)/\sqrt{2}$ and $B^L := (B_+^L - B_-^L)/\sqrt{2}$ define independent Brownian motions, and, by the rotational symmetry of Brownian motion in $\mathbf{R}^2$,

$$\frac{1}{\sqrt{2}}\begin{pmatrix} B^L + B^R \\ B^R - B^L \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} B^L \\ B^R \end{pmatrix}$$

is equal in distribution to (B^W, B^M), where B^W and B^M are independent Brownian motions. In particular,

$$\hat{S}_t = W_0 - v_1 t + 2B_t^W,$$

and

$$M_t = M_0 + B_t^M,$$

and therefore $\hat{S}$ and M are independent and we identify that $\hat{S} \stackrel{d}{=} S$. Finally, since $(W^n, 2n(1 - v_0/n)C^n)$ is a solution to the Skorokhod problem for $\hat{S}^n$ and $\hat{S}^n \Rightarrow S$, we conclude that $(W^n, 2n(1 - v_0/n)C^n) \Rightarrow (W, \ell)$. Moreover, since S and M are independent, so too are (W, ℓ) and M.

As is now standard, that the convergence holds with respect to the topology of uniform convergence on compact sets follows from the continuity of the limit and Lemma 3.10.1 of Ethier and Kurtz [40]. □

2.8.3. Convergence of T^n. Having shown that the midpoint and width of a patch, ignoring its eventual removal and interactions with its neighbours, converge, we now seek to describe the time at which it is removed. We will often refer to this event as *killing* and the time at which it happens its *death time*. Our method is as follows. We will define a sequence of stopping times T^n for the indefinitely evolving processes (M^n, W^n, C^n) of the previous section. Roughly speaking, T^n will be the first time that W^n has spent an exponential amount of local time at the origin.

In order to make sense of such a definition, we need some further notation. Let us denote by $D_+([0,\infty), \mathbf{R}) \subset D([0,\infty), \mathbf{R})$ the set of non-decreasing paths and, for $x \in D_+([0,\infty), \mathbf{R})$, define its *generalised inverse* ϕ_x by

$$\phi_x(u) := \inf\{t \geq 0 : x(t) \geq u\}.$$

It is well known that ϕ_x is left-continuous and admits right limits (see, for example, Embrechts and Hofert [34]) and thus has only countably many points of discontinuity. This

second statement follows from a minor modification of the arguments in the proof of Lemma 3.5.1 in Ethier and Kurtz [40], where it is proved for right continuous paths admitting left limits.

LEMMA 2.8.12. *For every $u > 0$, the map $\Phi_u : D_+([0,\infty), \mathbf{R}) \to \mathbf{R}$ defined by $\Phi_u(x) = \phi_x(u)$ is continuous at all x continuous and whose generalised inverse ϕ_x is continuous at u.*

PROOF. Let $x \in D_+([0,\infty), \mathbf{R})$ and let $(x_n)_{n\in\mathbf{N}} \subset D([0,\infty), \mathbf{R})$ be such that $x_n \to x$ in $D([0,\infty), \mathbf{R})$. Since x is continuous, the convergence is uniform on compact sets. In particular, for fixed $T > \phi_x(u)$ and $\varepsilon > 0$, there exists an N sufficiently large so that for all $n \geq N$, $\sup_{t\leq T} |x_n(t) - x(t)| < \varepsilon$. Then for all $t \leq \phi_x(u - 2\varepsilon)$,

$$x_n(t) \leq x(t) + \sup_{t\leq T} |x_n(t) - x(t)| \leq u - \varepsilon,$$

and thus $\Phi_u(x_n) > \phi_x(u - 2\varepsilon)$. Similarly, for all $t \geq \phi_x(u + 2\varepsilon)$,

$$x_n(t) \geq x(t) - \sup_{t\leq T} |x_n(t) - x(t)| \geq u + \varepsilon,$$

and thus $\Phi_u(x_n) \leq \phi_x(u + 2\varepsilon)$. In particular, since ϕ_x is continuous at u, passing to the limit $n \to \infty$ and then $\varepsilon \to 0$ completes the proof. □

We now reintroduce the removal of patches of good habitat from the system. Recall that when $W^n = 0$, the patch may expand to non-zero width, and does so at rate $2n^2(1 - v_1/n)$, or it is removed from the system altogether, which occurs with rate $\gamma_b n(1 - v_0/n)$. Consequently, after its first arrival at zero, W^n makes geometric number of excursions with mean

$$\frac{2n}{\gamma_b} \frac{1 - v_1/n}{1 - v_0/n}.$$

before it is removed.

PROPOSITION 2.8.13. *Denote by U an exponential random variable with unit mean and suppose that the conditions and notation of Theorem 2.8.11 are in force. Denote by*

$$T^n := \inf\left\{t \geq 0 : \int_0^t 1\{W^n = 0\}\, \mathrm{d}s \geq \frac{U}{\gamma_b n(1 - v_0/n)}\right\}.$$

Then as $n \to \infty$ and $v_0 \to 0$, $(M^n, W^n, 2nC^n, T^n) \Rightarrow (M, W, \ell, T)$ where M and (W, ℓ) are as defined in Theorem 2.8.11, M and (W, ℓ, T) are independent and

$$T = \ell^{-1}(2U/\gamma_b),$$

with ℓ^{-1}, which is not well-defined, identified with its generalised inverse ϕ_ℓ. The convergence is in the space $D([0,\infty),\mathbf{R}^3)\times\mathbf{R}$ where $D([0,\infty),\mathbf{R}^3)$ is equipped with the topology of uniform convergence on compact sets.

PROOF. The convergence $(M^n, W^n, 2nC^n) \Rightarrow (M, W, \ell)$ and the independence of M and (W, ℓ) is the subject of Theorem 2.8.11.

Observe that by the definition of the local time process C^n we may rewrite

$$T^n = \inf\{t \geq 0 : 2nC^n_t \geq 2U/\gamma_b\}.$$

Let $f : \mathbf{R} \to \mathbf{R}$ be bounded and continuous. Then, conditioning on the value of the exponential random variable U,

$$\mathbf{E}[f(T^n)] = \int_0^\infty e^{-u}\mathbf{E}[f(\phi_{2nC^n}(2u/\gamma_b))]\,\mathrm{d}u,$$

where ϕ_{2nC^n} is the generalised inverse of $2nC^n$. We wish to pass to the limit $n \to \infty$ and $v_0 \to 0$, applying the dominated convergence theorem. We have already shown that $2nC^n \Rightarrow \ell$. Furthermore, ℓ is continuous and its generalised inverse ϕ_ℓ is left-continuous with right limits and thus continuous almost everywhere. In particular, by Lemma 2.8.12, the map $\Phi_u : x \mapsto \phi_x(u)$ is continuous at ℓ for almost every u. Thus by the continuous mapping theorem, for almost all u, $\Phi_{2u/\gamma_b}(2nC^n) \Rightarrow \Phi_{2u/\gamma_b}(\ell)$ and so, for all such u,

$$\mathbf{E}[f(\phi_{2nC^n}(2u/\gamma_b))] \to \mathbf{E}[f(\phi_\ell(2u/\gamma_b))].$$

Since f is bounded, applying the dominated convergence theorem,

$$\mathbf{E}[f(T^n)] \to \int_0^\infty e^{-u}\mathbf{E}[f(\phi_\ell(2u/\gamma_b))]\,\mathrm{d}u = \mathbf{E}[f(\phi_\ell(2U/\gamma_b))],$$

or, in other words, $T^n \Rightarrow \phi_\ell(2U/\gamma_b)$. □

In summary, Proposition 2.8.13 says that the midpoint and width of an patch (without removal) converge to independent diffusions. The midpoint converges to a Brownian motion and it is decorated by the width process which is a diffusion process with drift toward to the origin, where there is a reflecting boundary. Furthermore, the first time at which the process W^n has spent an exponentially distributed amount of time with parameter $\gamma_b n(1 - v_0/n)$ at the origin converges to a random variable T, where T is the first time the local time of W (at zero) exceeds $2U/\gamma_b$, where U is an independent exponentially distributed random variable with unit mean. In other words, T is the first time at which W has accrued an exponentially distributed, with parameter $\gamma_b/2$, amount of local time at the origin.

The sequence of killed patches are only defined up until the times at which they are removed from the system. Furthermore, this time has the same distribution at T^n and, up until this time, the patch dynamics have the same law as (M^n, W^n). In particular, to recover the distribution of the killed patch, we simply need to stop watching (M^n, W^n), which otherwise continue to evolve indefinitely, at time T^n. In particular, if

$$(M^n_{\cdot\wedge T^n}, W^n_{\cdot\wedge T^n}, T^n) \Rightarrow (M_{\cdot\wedge T}, W_{\cdot\wedge T}, T)$$

as $n \to \infty$ and $v_0 \to 0$, then the law of the killed patches converges. This follows almost immediately from the following lemma.

LEMMA 2.8.14. *The map* $\Psi : D([0,\infty), \mathbf{R}^d) \times \mathbf{R} \to D([0,\infty), \mathbf{R}^d)$ *defined by*

$$\Psi(x, \tau) = x(\tau \wedge \cdot)$$

is continuous at all (x, τ) *where* x *is continuous.*

PROOF. Let $(x, \tau) \in D([0,\infty), \mathbf{R}^d) \times \mathbf{R}$ be such that x is continuous and let $(x_n)_{n\in\mathbf{N}} \subset D([0,\infty), \mathbf{R}^d)$ and $(\tau_n)_{n\in\mathbf{N}} \subset \mathbf{R}$ be sequences such that $x_n \to x$ with respect to the Skorokhod topology and $\tau_n \to \tau$. By Proposition 3.5.3 of Ethier and Kurtz [40], it is sufficient to check that for all $T > 0$

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |x_n(t \wedge \tau_n) - x(t \wedge \tau)| = 0.$$

Since x is continuous, it is uniformly continuous on $[0, T]$. In particular, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $|t - s| < \delta$ implies $|x(t) - x(s)| < \varepsilon$. Let N be large enough such that for all $n \geq N$

$$\sup_{t\leq T} |t \wedge \tau_n - t \wedge \tau| \leq |\tau_n - \tau| < \delta.$$

We observe that for all $t \in [0, T]$,

$$\begin{aligned} |x_n(t \wedge \tau_n) - x(t \wedge \tau)| &\leq |x_n(t \wedge \tau_n) - x(t \wedge \tau_n)| + |x(t \wedge \tau_n) - x(t \wedge \tau)| \\ &\leq \sup_{t\in[0,T]} |x_n(t) - x(t)| + \varepsilon. \end{aligned}$$

Since $x_n \to x$ in $D([0,\infty), \mathbf{R}^d)$ and x is continuous, the convergence is uniform on compacts. Moreover, the first term on the last line vanishes in the limit. Since $\varepsilon > 0$ was arbitrary, the proof is complete. □

THEOREM 2.8.15. *Suppose that the conditions and notation of Theorem 2.8.11 and Proposition 2.8.13 are in force. Then as* $n \to \infty$ *and* $v_0 \to 0$,

$$(M^n_{\cdot\wedge T^n}, W^n_{\cdot\wedge T^n}, 2nC^n_{\cdot\wedge T^n}, T^n) \Rightarrow (M_{\cdot\wedge T}, W_{\cdot\wedge T}, \ell_{\cdot\wedge T}, T).$$

The convergence is in the space $D([0,\infty), \mathbf{R}^3) \times \mathbf{R}$ *where* $D([0,\infty), \mathbf{R}^3)$ *is equipped with the topology of uniform convergence on compact sets.*

PROOF. Since M, W and ℓ are all continuous processes, the result follows from Theorem 2.8.11, Proposition 2.8.13, Lemma 2.8.14 and the continuous mapping theorem. □

2.8.4. Interactions with neighbouring patches. Finally, it remains to superimpose the interactions with neighbouring patches, which for now have been neglected. The idea is relatively straightforward: since $v_0 \to 0$, the distance between patches diverges and, consequently, the probability that the focal patches interact with their neighbours before eventually being removed from the system vanishes. Moreover, in the limit, patches have the same law as the quadruple (M, W, ℓ, T) of Theorem 2.8.15. Our strategy is as follows. We first show that we can contain any patch inside a box of area $16v_0^{-1/2}$ in the space-time plane with high probability as $n \to \infty$ and $v_0 \to 0$ (see Figure 2.3). Moreover, since patches are thrown down according to a Poisson random measure with intensity approximately $\gamma_b v_1 v_0 \, \mathrm{d}t \, \mathrm{d}x$ (cf Theorem 2.3.3), as we pass to the limit $v_0 \to \infty$, the probability that two such boxes intersect (and consequently the probability that two patches ever have a chance to meet) vanishes.

In order to prove that a single patch remains within the box described in Figure 2.3, we need to control the maximal width of the patch, its height (in the space-time plane) and the displacement of its midpoint before death. The first of our lemmas controls the maximal width of the patch.

LEMMA 2.8.16. *For all* $0 \le w \le C$,

$$\limsup_{\substack{n\to\infty \\ v_0\to 0}} \mathbf{P}[\sup_{t\le T^n} |W^n_t| \ge C \mid W^n_0 = \lfloor wn \rfloor / n] \le \frac{1}{C}\left(\frac{e^{v_1 w} - 1}{v_1} + \frac{2}{\gamma_b}\right).$$

PROOF. When away from the origin, W^n jumps to the right at rate with probability

$$r_n := \frac{2n^2(1 - v_1/n)}{2n^2(1 - v_1/n) + 2n^2(1 - v_0/n)}$$

and to the left with probability $l_n := 1 - r_n$. In particular, since we are assuming that $v_0 \to 0$ but v_1 remains constant, $l_n > r_n$ and so the width process is recurrent and returns to the origin with probability one whenever it is away from it.

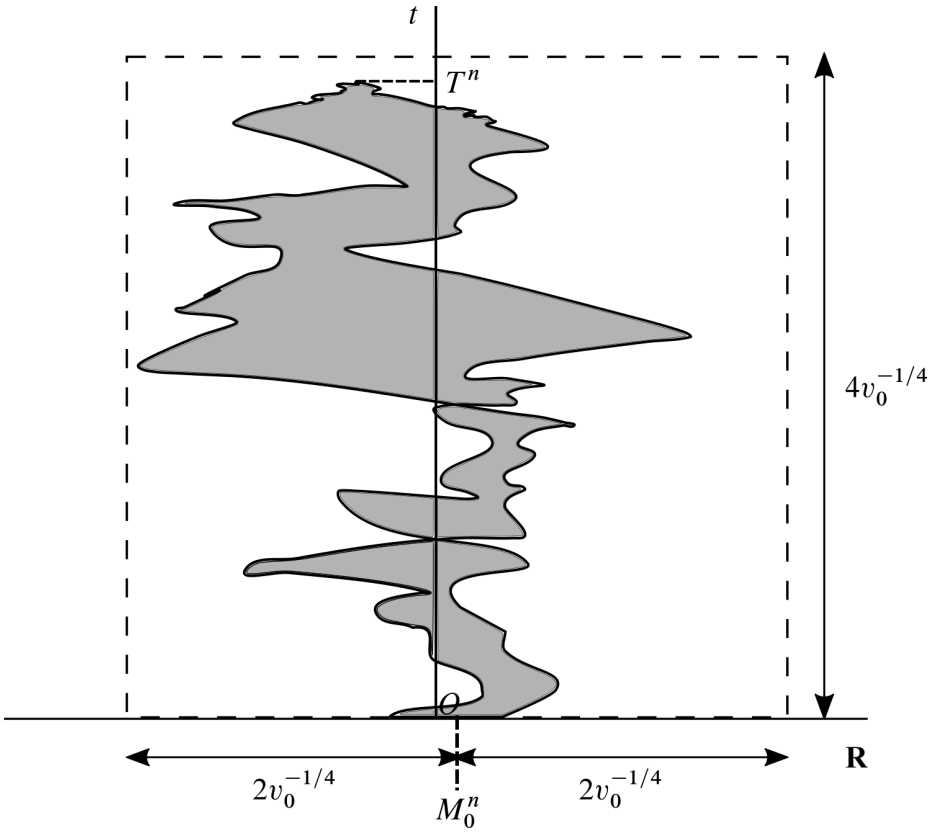


FIGURE 2.3

Initially, $W_0^n = w$. Let T_0^n be the first time W^n hits zero. Before hitting zero for the first time, the jumps of nW^n have the same law as a simple random walk on $\mathbf{Z}$ jumping to the right with probability r_n. By comparison to the gamblers' ruin problem, which for completeness is recalled in Proposition 2.10.2 in the appendix to this chapter, the probability that W^n exceeds C before it first hits zero is given by the probability that a simple random walk on $\mathbf{Z}$, starting from wn and jumping to the right with probability r_n, hits $\lfloor Cn \rfloor$ before it hits 0. Moreover,

$$\begin{aligned}\mathbf{P}[\sup_{t \le T_0^n} W_t^n \ge C \mid W_0^n = w] &= \mathbf{P}[\sup_{t \le T_0^n} nW_t^n > \lfloor Cn \rfloor \mid nW_0^n = nw] \\ &= \frac{(\frac{l_n}{r_n})^{wn} - 1}{(\frac{l_n}{r_n})^{\lfloor Cn \rfloor} - 1} \\ &= \frac{(1 + \frac{v_1 - v_0}{n - v_0})^{wn} - 1}{(1 + \frac{v_1 - v_0}{n - v_0})^{\lfloor Cn \rfloor} - 1}.\end{aligned}$$

Each time the patch width W^n returns to zero, it is killed off with probability

$$\mu_n := \frac{n\gamma_b(1 - v_0/n)}{n\gamma_b(1 - v_0/n) + 2n^2(1 - v_1/n)}.$$

Otherwise, it jumps to $1/n$ and begins an excursion before returning to the origin with probability one. Moreover, the number of excursions made before its eventual removal (not counting the initial displacement from zero) is geometric. Letting G^n denote the number of excursions of W^n before T^n,

$$\mathbf{P}[G^n = k] = (1-\mu_n)^k \mu_n, \qquad \mathbf{E}[G^n] = \frac{1-\mu_n}{\mu_n} = \frac{2n}{\gamma_b}\frac{1-v_1/n}{1-v_0/n}, \tag{2.8.7}$$

for all $k \geq 0$. By a further comparison to gamblers' ruin, the width exceeds C in any one excursion independently and with probability

$$\begin{aligned}\mathbf{P}[\sup_{t\leq T_0^n} W_t^n \geq C \mid W_0^n = 1/n] &= \mathbf{P}[\sup_{t\leq T_0^n} nW_t^n > \lfloor Cn\rfloor \mid nW_0^n = 1] \\ &= \frac{\frac{l_n}{r_n} - 1}{(\frac{l_n}{r_n})^{\lfloor Cn\rfloor} - 1} \\ &= \frac{1}{n-v_0}\frac{v_1 - v_0}{(1+\frac{v_1-v_0}{n-v_0})^{\lfloor Cn\rfloor} - 1}\end{aligned}$$

Moreover, applying Markov's inequality, the probability there is at least one excursion which exceeds a width C is bounded above by

$$\mathbf{E}[G^n]\frac{1}{n-v_0}\frac{v_1-v_0}{(1+\frac{v_1-v_0}{n-v_0})^{\lfloor Cn\rfloor}-1} = \frac{2}{\gamma_b}\frac{1-v_1/n}{1-v_0/n}\frac{n}{n-v_0}\frac{v_1-v_0}{(1+\frac{v_1-v_0}{n-v_0})^{\lfloor Cn\rfloor}-1}.$$

The process W^n exceeds C before being T^n only if it either exceeds C during its first passage to zero, or if it at least one the subsequent independent excursions exceeds C. Moreover, applying the union bound and the previous two calculations,

$$\limsup_{\substack{n\to\infty\\ v_0\to 0}} \mathbf{P}[\sup_{t\leq T^n} |W_t^n| \geq C \mid W_0^n = w] \leq \frac{e^{v_1 w}-1}{e^{v_1 C}-1} + \frac{2}{\gamma_b}\frac{v_1}{e^{v_1 C}-1}.$$

Applying the trivial inequality $1 + x \leq e^x$, the result follows. □

Given the estimate for the maximal width of the patch, the total region spanned during the patch's lifetime can be controlled if one can control the displacement of the midpoint process. This is the substance of the next lemma.

LEMMA 2.8.17. *For all $t_0 > 0$, $C > 0$,*

$$\limsup_{\substack{n\to\infty\\ v_0\to 0}} \mathbf{P}[\sup_{t\leq t_0} |M_t^n - M_0^n| \geq C] \leq \frac{t_0}{C^2}.$$

PROOF. Recalling (2.8.1), $M^n - M_0^n$ is a mean zero martingale. Moreover, since any

convex function of a martingale is a submartingale, $|M^n - M_0^n|^2$ is a submartingale. Thus by Doob's submartingale inequality,

$$\mathbf{P}[\sup_{t \le t_0} |M_t^n - M_0^n| \ge C] \le \frac{\mathbf{E}[|M_{t_0}^n - M_0^n|^2]}{C^2} = \frac{\mathrm{var}[M_{t_0}^n - M_0^n]}{C^2}.$$

Applying the law of total variance, ie first conditioning on C_t^n in equation (2.8.1) and noting that $\mathbf{E}[M_{t_0}^n - M_0^n \mid C_{t_0}^n] = 0$,

$$\mathrm{var}[M_{t_0}^n - M_0^n] = \mathbf{E}[\mathrm{var}[M_{t_0}^n - M_0^n \mid C_{t_0}^n]] \le \frac{t_0}{2}\{(1 - v_1/n) + (1 - v_0/n)\} \le t_0,$$

so completing the proof. □

Finally, controlling the removal time of the patch gives us control on its height (in the space-time plane).

LEMMA 2.8.18. *For all $w > 0$,*

$$\mathbf{E}[T^n \mid W_0^n = \lfloor wn \rfloor/n] = \frac{\lfloor wn \rfloor}{2n(v_1 - v_0)} + \frac{1}{2n^2(1 - v_1/n) + \gamma_b n(1 - v_0/n)} + \frac{2n}{\gamma_b}\frac{1 - v_1/n}{1 - v_0/n}\left(\frac{1}{2n^2(1 - v_1/n) + \gamma_b n(1 - v_0/n)} + \frac{1}{2n(v_1 - v_0)}\right),$$

and furthermore, for all $C > 0$,

$$\limsup_{\substack{n\to\infty \\ v_0 \to 0}} \mathbf{P}[T^n \ge C \mid W_0^n = \lfloor wn \rfloor/n] \le \frac{1}{v_1 C}\left(\frac{w}{2} + \frac{1}{\gamma_b}\right).$$

PROOF. It is straightforward to see that the second statement follows directly from the first by Markov's inequality.

We break the total time until the death of the patch, T^n, into the sum of independent parts. First, there is the time it takes for W^n to hit zero for the first time, starting from $\lfloor wn \rfloor$; then, W^n waits an exponentially distributed amount of time, at the end of which, it may die, or it may begin an excursion away from zero. For each excursion away from zero we have to add the total time to return to zero, plus a further hold time.

Recall the notation for the successive arrival and departure times of W^n at the origin. We set $\theta_0^n := 0$ and defined inductively, for each $k \ge 1$,

$$\tau_k^n := \inf\{t > \theta_{k-1}^n : W_t^n = 0\},$$
$$\theta_k^n := \inf\{t > \tau_k^n : W_t^n > 0\}.$$

Letting G^n denote the number of excursions W^n makes away from zero after its first arrival

and before T^n,

$$T^n = \tau_1^n + (\theta_1^n - \tau_1^n) + \sum_{k=1}^{G^n} \{(\theta_{k+1}^n - \tau_{k+1}^n) + (\tau_{k+1}^n - \theta_k^n)\}. \tag{2.8.8}$$

The sequence of waiting times $\theta_k^n - \tau_k^n$ are independent and exponentially distributed with parameter $2n^2(1-v_1/n)+\gamma_b n(1-v_0/n)$, ie the sum of the rates at which the patch expands and is killed when at width zero. Similarly, the return times $\tau_{k+1}^n - \theta_k^n$ are independent and, by Lemma 2.8.6, for $k \geq 1$ have expectation

$$\mathbf{E}[\tau_{k+1}^n - \theta_k^n] = \frac{1}{2n(v_1 - v_0)}$$

Starting from $\lfloor wn \rfloor / n$, W^n must make $\lfloor wn \rfloor$ independent copies of passages from $1/n$ to the origin. In particular, by Lemma 2.8.6,

$$\mathbf{E}[\tau_1^n \mid W_0^n = \lfloor wn \rfloor / n] = \lfloor wn \rfloor \mathbf{E}[T_0^n \mid W_0^n = 1/n] = \frac{\lfloor wn \rfloor}{2n(v_1 - v_0)}.$$

We wish to take expectations in (2.8.8). To do, we must establish how to compute the expectation of the sum. Possibly surprisingly, the number of excursions G^n is in fact independent of the random variables $\theta_{k+1}^n - \tau_{k+1}^n$. Although not immediately clear, this is standard result from the theory of competing exponential clocks: if A and B are two independent exponential random variables (possibly with different parameters), then $A \wedge B$ is independent of $1\{A \leq B\}$; moreover, how long W^n spends at zero before either being removed or beginning a further excursion is independent of which of these happens. We may thus apply the Wald equation and obtain that

$$\begin{aligned}\mathbf{E}[T^n \mid W_0^n = \lfloor wn \rfloor / n] = \frac{\lfloor wn \rfloor}{2n(v_1 - v_0)} + \frac{1}{2n^2(1 - v_1/n) + \gamma_b n(1 - v_0/n)} \\ + \frac{2n}{\gamma_b} \frac{1 - v_1/n}{1 - v_0/n} \left(\frac{1}{2n^2(1 - v_1/n) + \gamma_b n(1 - v_0/n)} + \frac{1}{2n(v_1 - v_0)} \right),\end{aligned}$$

as required. □

We are now able to prove that the patch does not leave the box in Figure 2.3. Denote by

$$A^n := \{ \sup_{t \leq T^n} M_0^n - L_t^n > 2v_0^{-1/4} \} \cup \{ \sup_{t \leq T^n} R_t^n - M_0^n > 2v_0^{-1/4} \} \cup \{ T^n > 4v_0^{-1/4} \},$$

this event.

PROPOSITION 2.8.19. *For all $w > 0$,*

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \mathbf{P}[A^n \mid W_0^n = \lfloor wn \rfloor / n] = 0.$$

PROOF. The key is to notice that

$$A^n \subseteq \{\sup_{t\le T^n} |W_t^n| > v_0^{-1/4}\} \cup \{\sup_{t\le T^n} |M_t^n - M_0^n| > v_0^{-1/4}\} \cup \{T^n > 4v_0^{-1/4}\},$$

so that by the union bound,

$$\begin{aligned}\mathbf{P}[A^n \mid W_0^n = \lfloor wn \rfloor / n] \le{}& \mathbf{P}[\sup_{t\le T^n} |W_t^n| > v_0^{-1/4} \mid W_0^n = \lfloor wn \rfloor / n] \\ &+ \mathbf{P}[\sup_{t\le T^n} |M_t^n - M_0^n| > v_0^{-1/4} \mid W_0^n = \lfloor wn \rfloor / n] \\ &+ \mathbf{P}[T^n > 4v_0^{-1/4} \mid W_0^n = \lfloor wn \rfloor / n].\end{aligned}$$

Applying the previous three lemmas, the result follows. □

Notice that in the proof of Proposition 2.8.19, we did not use the form of the dimensions of the box, ie that its edge lengths scale like $v_0^{-1/4}$. In fact, any increasing large box would have sufficed since the patch has $O(1)$ size. The dimensions will be useful in completing the proof of Theorem 2.3.4 however: since the box has area scaling like $v_0^{-1/2}$, centring a box on each potential arrival of a new patch, according to the Poisson random measure of Section 2.3.2 with intensity measure (2.3.2), proportional to v_0, we expect no boxes will intersect. In particular, all new arrivals are admitted, and no clusters interact. We are finally able to prove Theorem 2.3.4.

PROOF OF THEOREM 2.3.4. Consider the focal patch, containing the origin at $t = 0$ and its two nearest neighbours. Ignoring interactions, each of them have the same law, given by that of (M^n, W^n, T^n) until the first time any of them meet. Moreover, none of them leave a box described in Figure 2.3 centred on their initial midpoint with probability $\mathbf{P}[A^n]^3 \to 0$ as $n \to \infty$ and $v_0 \to 0$. Thus, as long as the focal patch does not meet any new patches thrown down in its vicinity before its is removed, its law is described by (M, W, ℓ, T) after passing to the limit.

Recall that *potential* new patches are thrown down at initialisation, according to the Poisson random measure P^n with intensity measure

$$m_n(\mathrm{d}t, \mathrm{d}x) = \frac{1}{n} \frac{\gamma_b v_0 v_1}{1 - v_0/n} \mathrm{d}t \otimes \sum_{x' \in \mathbf{Z}/n} \delta_{x'}(\mathrm{d}x).$$

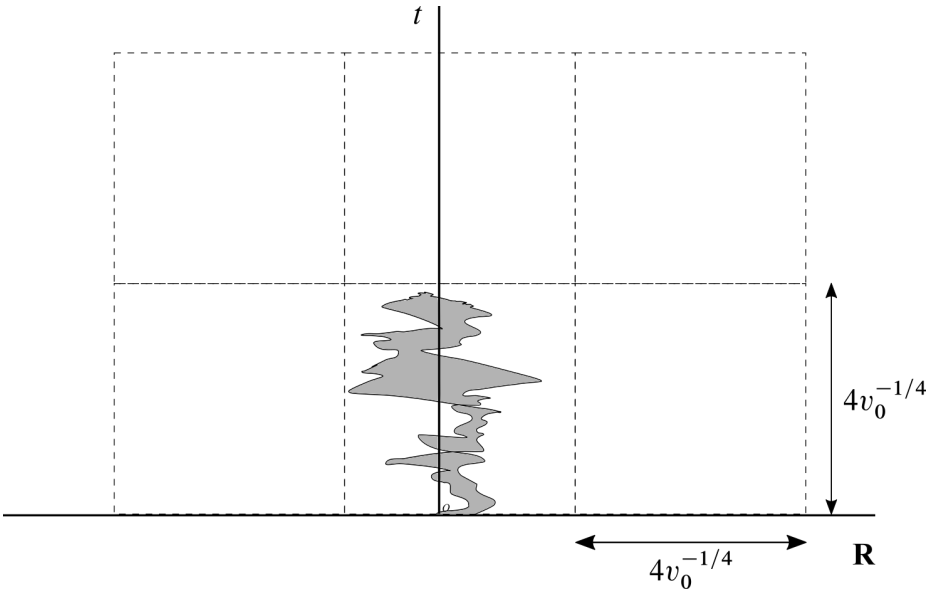


FIGURE 2.4

Therefore it is certainly the case that the focal patch meets no new patch if there are no points of this random measure in the six-box-grid depicted in Figure 2.4. This is because, again by the Proposition 2.8.19, no new patch (initially of width zero) can exit a box centred on its initial midpoint.

Integrating out over the intensity measure, the expected number of points of the random measure P^n falling into this region is given by

$$\frac{\gamma_b v_0 v_1}{1 - v_0/n} 8 v_0^{-1/4} \frac{\lfloor 12 v_0^{-1/4} n \rfloor}{n}.$$

Thus by Markov's inequality, the probability that there is at least one point of P^n in this region is bounded above by

$$\frac{96 \gamma_b v_0^{1/2} v_1}{1 - v_0/n} \to 0$$

as $n \to \infty$ and $v_0 \to 0$.

Moreover, in the limit $n \to \infty$ and $v_0 \to 0$, the probability that the focal patch meets any patch existing at $t = 0$ or any new patch vanishes. Thus, its law can be described by the limiting processes (M, W, ℓ, T) which ignore interactions with neighbouring patches, as hoped. This completes the proof. □

2.9. Effective dispersal distance.

In Section 2.3.2 we sought to understand how potential new patches appear after passing to the limit $n \to \infty$; we found that these potential new patches were thrown down according to a Poisson random measure, with on average $\gamma_b v_1 v_0$ arrivals per each unit of area

in the space-time plane. These patches are admitted if the habitat at the location of this potential start point is *bad* at this time. Further, in Section 2.3.3 we sought to understand the geometry of these patches. Passing simultaneously to the limit $n \to \infty$ and $v_0 \to 0$ — the latter so that the distance between patches diverges and interactions between them may be ignored — we found that an individual patch has boundaries which behave as reflecting diffusions, and the patch dies out after an exponential amount of local time at width zero.

Given these results we, in this section, conjecture the speed of the lineages. Unfortunately, time constraints have not permitted a full proof, but we will indicate the key steps to proving such a result and leave it as an open problem.

Recall that the lineage process X^n behaves as follows. In the interior of a patch or in a region of zeroes it performs a nearest neighbour random walk, jumping a distance $1/n$ at rate σ_n^2, choosing a direction with equal probability. A lineage inside a patch will reflect off the patch boundaries, remaining inside the patch until the time at which it is removed from the system; it then wanders around the empty region of space until it approaches the boundary of a new patch, where it is immediate drawn in and behaves as before. This is the substance of Theorem 2.2.7.

Our primary goal is to infer the effective dispersal of the lineages, as measured by the time-average of the mean squared displacement:

$$\lim_{t\to\infty} \frac{1}{t}\mathbf{E}[(X_t^n)^2]. \tag{2.9.1}$$

When a lineage enters a patch, it is constrained to pass through the points where it has width zero and, moreover, remains within the patch until it is eventually removed from the system. Its mean squared displacement within patches, then, is controlled by the mean squared displacement of the midpoint of the patch.

Let ρ_n denote the long-run proportion of time that the lineage spends within patches of good habitat, ie

$$\rho_n := \lim_{t\to\infty} \frac{1}{t}\int_0^t \Xi_s^n(X_s^n)\,\mathrm{d}s,$$

and further define the 'speed' of the midpoint of patches

$$\varsigma_n^2 := \mathbf{E}\left[\frac{(M_{T^n}^n - M_0^n)^2}{T^n}\,\middle|\, W_0^n = 0\right],$$

where (M^n, W^n, T^n) are as in the statement of Theorem 2.3.4. We may then sensibly suggest that the time-averaged mean squared displacement of X^n defined in (2.9.1) is approximated in the limit as follows.

CONJECTURE 2.9.1. *The effective dispersal of the lineages may be approximated by*

$$\lim_{\substack{n\to\infty\\ v_0\to 0}} \lim_{t\to\infty} \frac{1}{t}\mathbf{E}[(X_t^n)^2] = \lim_{\substack{n\to\infty\\ v_0\to 0}} \{\rho_n \varsigma_n^2 + (1-\rho_n)\frac{\sigma_n^2}{n^2}\}.$$

The term on the right hand side is the weighted average of the speed within patches of good habitat and regions of bad habitat — weighted by the proportion of time the lineage spends within each, respectively.

Patches of good habitat are on average a distance $O(v_0^{-1})$ apart. Moreover, in order to achieve a non-trivial limit where X^n spends a positive proportion of time in patches, X^n will need to jump fast enough such that it traverses these distances in an $O(1)$ amount of time. We achieve this by setting

$$\sigma_n^2 := \left(\frac{\sigma n}{v_0}\right)^2.$$

Under this scaling, locally, ie away from patch boundaries,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}[(X_t^n - X_0^n)^2 \mid X_0^n = x] \approx \frac{\sigma^2}{v_0^2},$$

as $n \to \infty$. In particular, taking a further rescaling of space so that one unit of space corresponds to n/v_0 points of the one-dimensional lattice $\mathbf{Z}$, it is not too difficult to convince oneself that $v_0 X^n \Rightarrow Y$, where Y is a diffusion process reflecting off the internal boundaries of patches but otherwise behaving locally as a speed-σ^2 Brownian motion; that is, $\mathrm{d}\mathbf{E}[Y_t^2]/\,\mathrm{d}t = \sigma^2$ away from patch boundaries.

Under the same rescaling of space, potential new patches are thrown down according to a Poisson random measure as follows.

LEMMA 2.9.2. *Let P^n be the Poisson random measure of Proposition 2.3.3 which throws potential new patches onto $\mathbf{R}_+ \times \mathbf{Z}/n$ with intensity measure m_n given by* (2.3.2) *on $\mathbf{R}_+ \times \mathbf{R}$. Denote by Q^n the rescaled measure defined by*

$$Q^n(\mathrm{d}t, \mathrm{d}x) := P^n(\mathrm{d}t, v_0^{-1}\,\mathrm{d}x).$$

Then as $n \to \infty$ and $v_0 \to 0$, the sequence of Poisson random measures $Q^n \Rightarrow Q$ in $\mathcal{M}(\mathbf{R}_+ \times \mathbf{R})$ equipped with the topology of vague convergence, where Q is the Poisson random measure with mean measure

$$\gamma_b v_1\,\mathrm{d}t\,\mathrm{d}x.$$

PROOF. Referring to the mean measure of the random measure P^n as given in equa-

tion (2.3.2), observe that on rectangles of the form $[s,t)\times[a,b)$,

$$\begin{aligned}\mathbf{E}[Q^n([s,t)\times[a,b))] &= \mathbf{E}[P^n([s,t)\times[v_0^{-1}a, v_0^{-1}b))]\\ &= \frac{1}{n}\frac{\gamma_b v_0 v_1}{1-v_0/n}(t-s)\sum_{x\in\mathbf{Z}/n}\delta_x([v_0^{-1}a, v_0^{-1}b))\\ &= \frac{1}{n}\frac{\gamma_b v_0 v_1}{1-v_0/n}(t-s)\sum_{x\in v_0\mathbf{Z}/n}\delta_x([a,b)).\end{aligned}$$

Moreover, since measures on $\mathbf{R}^2$ are uniquely determined by their values on rectangles of this form, Q^n has mean measure defined by

$$\frac{1}{n}\frac{\gamma_b v_0 v_1}{1-v_0/n}\,\mathrm{d}t\sum_{x'\in v_0\mathbf{Z}/n}\delta_{x'}(\mathrm{d}x).$$

Taking $n/v_0 =: n'$, we observe that the statement of the lemma is in fact the same as Theorem 2.3.3 with a change of parameters. □

From the perspective of the increasingly fast moving process X^n, these patches, whose geometries converge in the limit (cf Theorem 2.3.4), will have effective zero width. More accurately, the process v_0X^n will have approximately zero speed inside patches. In an attempt to make this rigorous we make the following conjecture.

CONJECTURE 2.9.3. *In the setting of Theorem 2.3.4,*

$$\lim_{\substack{n\to\infty\\ v_0\to 0}}\varsigma_n^2 = 1.$$

Within a patch, the process v_0X^n travels at average speed $v_0^2\varsigma_n^2\to 0$ as a consequence.

The idea of a proof for this result is as follows. Since by Theorem 2.3.4 the triple $(M^n,W^n,T^n)\Rightarrow(M,W,T)$ and M is a Brownian motion independent of (W,T), we may readily believe that it is possible to show that

$$\begin{aligned}&\lim_{\substack{n\to\infty\\ v_0\to 0}}\mathbf{E}\left[\frac{(M^n_{T^n}-M^n_0)^2}{T^n}\,\middle|\,W^n_0=0\right]\\ &=\mathbf{E}\left[\frac{(M_T-M_0)^2}{T}\,\middle|\,W_0=0\right]\\ &=\mathbf{E}\left[\frac{1}{T}\mathbf{E}\left[(M_T-M_0)^2\,\middle|\,T,W_0=0\right]\,\middle|\,W_0=0\right]\\ &=\mathbf{E}\left[\frac{1}{T}T\,\middle|\,W_0=0\right]\\ &=1,\end{aligned}$$

where the tower rule for conditional expectation has been used to arrive at the second equality. Proving the first equality is the substance of establishing the conjecture.

We now turn our attention to the long-run proportion of time that X^n spends in patches of good habitat. This is of course equal to the long-run proportion of time that $v_0 X^n$ spends in (rescaled) patches of good habitat. Furthermore, we conjecture that this long-run proportion of time is proportional to the ratio of time spent within patches of good habitat and time spent looking for patches of good habitat. This is formulated as follows.

CONJECTURE 2.9.4. *Let $\mathbf{E}[\tau_{10}^n]$ be the mean time it takes for $v_0 X^n$ to exit a patch of good habitat after first entering it, at the time of the patch's death; and let $\mathbf{E}[\tau_{01}^n]$ be the expected time it takes to enter a new patch after exiting the patch at this time. Then*

$$\rho_n = \frac{\mathbf{E}[\tau_{10}^n]}{\mathbf{E}[\tau_{10}^n] + \mathbf{E}[\tau_{01}^n]}.$$

Let us begin by considering $\mathbf{E}[\tau_{10}^n]$, the expected amount of time the lineage spends in a patch after hitting it. Since we have made no further rescaling of time, the height of the patches (in the space-time plane) is given by T^n, and $T^n \Rightarrow T$ where T is also defined as in Theorem 2.3.4.

LEMMA 2.9.5. *In the setting of Theorem 2.3.4,*

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \mathbf{E}[T^n \mid W_0^n = 0] = \frac{1}{\gamma_b v_1},$$

and

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \mathbf{E}[(T^n)^2 \mid W_0^n = 0] = \frac{2}{\gamma_b^2 v_1^2} + \frac{1}{\gamma_b v_1^3}.$$

PROOF. The first limit follows immediately from Lemma 2.8.18.

Recall from the proof of Lemma 2.8.18 that we may write

$$T^n = \tau_1^n + (\theta_1^n - \tau_1^n) + \sum_{k=1}^{G^n} \{(\theta_{k+1}^n - \tau_{k+1}^n) - (\tau_{k+1}^n - \theta_k^n)\},$$

where τ_1^n is the first hitting time of W^n at zero; $\theta_1^n - \tau_1^n$ is the first departure time (either beginning an excursion or killing the patch); and $\tau_{k+1}^n - \theta_k^n$ and $\theta_{k+1}^n - \tau_{k+1}^n$ are the subsequent (independent) excursions and hold times at zero respectively. The term G^n is the geometric number of excursions away from width zero. As noted in the proof of Lemma 2.8.18, all of these individual random variables are independent and G^n is independent of $\tau_{k+1}^n - \theta_k^n$ and $\theta_{k+1}^n - \tau_{k+1}^n$.

Conditional on $W^n = 0$, $\tau_1^n = 0$ almost surely. This term is thus ignored. Denoting by

$$\Sigma_n := \sum_{k=1}^{G^n} \{(\theta_{k+1}^n - \tau_{k+1}^n) + (\tau_{k+1}^n - \theta_k^n)\},$$

we may simply compute that

$$\mathbf{E}[(T^n)^2 \mid W_0^n = 0] = \mathbf{E}[\Sigma_n^2 \mid W^n = 0] + o(n^{-1}),$$

where we have used that as $\theta_1^n - \tau_1^n$ is exponentially distributed with parameter $2n^2(1 - v_1/n) + n\gamma_b(1 - v_0/n)$,

$$\mathbf{E}[\theta_1^n - \tau_1^n \mid W^n = 0] = O(n^{-2}),$$

$$\mathbf{E}[(\theta_1^n - \tau_1^n)^2 \mid W^n = 0] = O(n^{-4}),$$

and

$$\mathbf{E}[\Sigma_n \mid W^n = 0] = \mathbf{E}[G^n]\{\mathbf{E}[\theta_2^n - \tau_2^n] + \mathbf{E}[\tau_2^n - \theta_1^n]\} = O(1).$$

In the last equality, the value of $\mathbf{E}[G^n]$ is given by (2.8.7) in the proof of Lemma 2.8.16 and $\mathbf{E}[\tau_2^n - \theta_1^n]$ from Lemma 2.8.6.

The random variable Σ_n is independent of W_0^n and we subsequently drop the conditioning from the notation. Referring to Lemma 2.10.3 in the appendix to this chapter, which gives an expression for the second moment of a random sum of this type,

$$\mathbf{E}[\Sigma_n^2] = \mathbf{E}[G_n(G_n - 1)]\mathbf{E}[(\theta_{k+1}^n - \tau_{k+1}^n) + (\tau_{k+1}^n - \theta_k^n)]^2 + \mathbf{E}[G_n]\mathbf{E}[\{(\theta_{k+1}^n - \tau_{k+1}^n) + (\tau_{k+1}^n - \theta_k^n)\}^2].$$

Since G_n is geometrically distributed with parameter μ_n, as per (2.8.7),

$$\mathbf{E}[G_n] = \frac{1 - \mu_n}{\mu_n} = \frac{2n}{\gamma_b} \frac{1 - v_1/n}{1 - v_0/n},$$

and

$$\mathbf{E}[G_n(G_n - 1)] = \frac{2(1 - \mu_n)^2}{\mu_n^2} = \frac{8n^2}{\gamma_b^2} \left(\frac{1 - v_1/n}{1 - v_0/n} \right)^2.$$

By Lemma 2.8.6 and since $\theta_{k+1}^n - \tau_{k+1}^n$ is independent and exponentially distributed with parameter $2n^2(1 - v_1/n) + n\gamma_b(1 - v_0/n)$,

$$\mathbf{E}[(\theta_{k+1}^n - \tau_{k+1}^n) + (\tau_{k+1}^n - \theta_k^n)]^2 = \frac{1}{4n^2(v_1 - v_0)^2} + O(n^{-3}),$$

and

$$\begin{aligned}&\mathbf{E}[\{(\theta^n_{k+1}-\tau^n_{k+1})+(\tau^n_{k+1}-\theta^n_k)\}^2]\\&=\operatorname{var}[\theta^n_{k+1}-\tau^n_{k+1}]+\operatorname{var}[\tau^n_{k+1}-\theta^n_k]+\mathbf{E}[(\theta^n_{k+1}-\tau^n_{k+1})+(\tau^n_{k+1}-\theta^n_k)]^2\\&=\frac{1}{2n(v_1-v_0)^3}+O(n^{-2}).\end{aligned}$$

Substituting these expressions, the result follows. □

Since we are arguing merely heuristically, we leave it as a conjecture that

$$\mathbf{E}[T \mid W_0 = 0] = \lim_{\substack{n\to\infty\\ v_0\to 0}} \mathbf{E}[T^n \mid W^n_0 = 0],$$

and similarly for the second moment.

Lineages wander the regions of *bad* habitat until they encounter a patch of *good* habitat, where they become trapped. They are more likely to encounter patches which have long lifetimes than short ones. Furthermore, we expect that the height of a patch hit by a lineage will be distributed according to the size-biased distribution of T^n, ie denoting by H^n the height of the encountered patch,

$$\mathbf{P}[H^n \le h] = \frac{1}{\mathbf{E}[T^n \mid W^n_0 = 0]} \int_0^h t\, \mathbf{P}[T^n \in \mathrm{d}t \mid W^n_0 = 0].$$

Furthermore, the distribution of the remaining height of the patch at the time when the lineage hits it will be a uniform pick from this sized bias distribution. Moreover, one may readily check that the residual height has expectation given by

$$\frac{1}{2}\frac{\mathbf{E}[(T^n)^2 \mid W^n_0 = 0]}{\mathbf{E}[T^n \mid W^n_0 = 0]},$$

where the factor of the half comes from averaging out over the uniform location, and the ratio of the second and first moment of T is the expectation of the size-biased random variable. Assuming the above exchange of limits and expectations is valid, we moreover expect the following.

CONJECTURE 2.9.6.

$$\lim_{\substack{n\to\infty\\ v_0\to 0}} \mathbf{E}[\tau^n_{10}] = \lim_{\substack{n\to\infty\\ v_0\to 0}} \frac{1}{2}\frac{\mathbf{E}[T_n^2 \mid W^n_0 = 0]}{\mathbf{E}[T_n \mid W^n_0 = 0]} = \frac{1}{\gamma_b v_1} + \frac{1}{2v_1^2}.$$

Now let us turn to $\mathbf{E}[\tau^n_{01}]$, the time to re-enter a patch after exiting a patch of good habitat at the time of its death. Outside of patches, $v_0 X^n$ approximates a Brownian motion with

speed σ^2. Further, recall that under the stationary distribution for Ξ^n, the distance to the right to the nearest patch after a rescaling of space by a factor of v_0 is distributed according to

$$\frac{v_0}{n} \times \text{Geometric}(v_0/n)$$

which converges in distribution to an exponential random variable with unit mean in the limit $n \to \infty$, $v_0 \to 0$. Similarly, the distance to the left to the nearest patch of good habitat has the same distribution, is independent and likewise converges to an exponential random variable with unit mean. On average then, the distances to the left and to the right to the nearest patches from position of the lineage at the time it leaves the patch which just died out are both 1. It is well known that the mean first time a speed σ^2 Brownian motion starting at the origin hits either -1 or 1 is given by $1/\sigma^2$. This can, for example, be proved by Dynkin's formula: the expected time is given by $u(0)$ where u is the solution to $\sigma^2 u''(x)/2 = -1$ for all $-1 \leq x \leq 1$ and $u(-1) = u(1) = 0$. We propose then that the following is true.

CONJECTURE 2.9.7. *The mean time it takes for $v_0 X^n$ to exit a patch of good habitat after first entering it satisfies*

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \mathbf{E}[\tau_{01}^n] = \frac{1}{\sigma^2}.$$

Proving such a result seems incredibly difficult and will make substantial use of the methods and results of the literature on random walks or diffusions in random environments. There is a great body of work which considers Brownian motions amongst a Poisson random field of static traps (see, for example, Sznitman [123] for a thorough review). This contrasts with our setting, where the *traps*, in this case patches of good habitat, are dynamic: a trap present at time t is not necessarily present at time $t + s$. Even so, much of this work on Poisson fields of traps is centred on computing the distribution of the Brownian motion conditional on not hitting a trap by time t. Here, we are interested in the mean hitting time of the first trap.

Combining Conjecture 2.9.4, Conjecture 2.9.6 and Conjecture 2.9.7, we expect that it is possible to prove the following.

CONJECTURE 2.9.8. *The long-run proportion of time that the lineage spends within patches of good habitat satisfies*

$$\lim_{\substack{n\to\infty \\ v_0\to 0}} \rho_n = \frac{\frac{1}{\gamma_b v_1} + \frac{1}{2v_1^2}}{\frac{1}{\gamma_b v_1} + \frac{1}{2v_1^2} + \frac{1}{\sigma^2}}.$$

Finally, combining the conjectured results Conjecture 2.9.1, Conjecture 2.9.3 and Conjecture 2.9.8 we come to the main conjecture of the section.

CONJECTURE 2.9.9. *The effective dispersal satisfies*

$$\lim_{v_0\to 0}\lim_{n\to\infty}\lim_{t\to\infty}\frac{v_0^2}{t}\mathbf{E}[(X_t^n)^2] = \frac{1}{\frac{1}{\gamma_b v_1}+\frac{1}{2v_1^2}+\frac{1}{\sigma^2}}.$$

We conclude by partially verifying this result against our intuition. The first thing to notice is that the limit is bounded above by σ^2. This makes perfect sense since patches are only ever an impediment to the lineages, and always slow them down.

As $\gamma_b \to 0$, so that patches never die, lineages become trapped within a patch and are constrained to remain there forever. In this limit then, they move with speed zero. This is confirmed by the result. Similarly, as $v_1 \to 0$, patches, which have expected width v_1^{-1}, become very wide. In particular, they never 'come down' to zero width and are thus given no opportunity to die out. The lineage, then, is constrained to be trapped in the patch forever and has limiting zero speed, as predicted.

As $v_1 \to \infty$, patches are very thin and are killed off quickly. In fact, if we believe Conjecture 2.9.8, then lineages spend asymptotically zero time within patches, because they are killed off so fast. We are there for unsurprised that the conjecture predicts that in this limit, the lineage has limiting speed σ^2, as a lineage which never encounters a patch would.

The final limit, when $\gamma_b \to \infty$ is more complicated. In this regime, patches which are thrown down according to a Poisson random measure with mean measure proportional to γ_b and killed after a exponential, with mean proportional to γ_b amount of local time at width zero. So in this limit, patches are extremely abundant, but are killed very quickly. The limiting speed then depends on the parameters in a highly complicated way.

2.10. Appendix.

2.10.1. Hitting times of the simple random walk on Z. The following uses standard methods (see, for example, Feller [42]) but is stated and proved for completeness.

PROPOSITION 2.10.1. *Let $(V_k)_{k\geq 0}$ denote a simple random walk on $\mathbf{Z}$ jumping to the right with probability $r > 0$ and to the left with probability $l = 1 - r$. Suppose that $l > r$ so that the walk is transient and drifts to the left. Define*

$$H_0 := \inf\{k \geq 0 : V_k = 0\}$$

to be the first hitting time at zero. Then

$$\mathbf{E}[H_0 \mid V_0 = 1] = \frac{1}{l-r}$$

and

$$\mathrm{var}[H_0 \mid V_0 = 1] = \frac{4lr}{(l-r)^3}.$$

PROOF. Let $P(\theta) = \mathbf{E}[\theta^{H_0} \mid V_0 = 1]$ denote the probability generating function of H_0. Since $l > r$, $H_0 < \infty$ almost surely and thus $P(\theta)$ is well-defined for all $\theta \in [0,1]$ and $P(1) = 1$. Furthermore, by conditioning on the first step of V, it is a standard technique to exploit that P can be written as a solution to

$$P(\theta) = l\theta + r\theta P(\theta)^2. \tag{2.10.1}$$

Applying standard facts about generating functions

$$\mathbf{E}[H_0 \mid V_0 = 1] = P'(1)$$

and

$$\mathrm{var}[H_0 \mid V_0 = 1] = P''(1) + P'(1) - P'(1)^2.$$

Differentiating through the recursion (2.10.1) and solving

$$P'(1) = \frac{1}{l-r}$$

and

$$P''(1) = \frac{4r}{(l-r)^2} + \frac{2r}{(l-r)^3}$$

from which one can readily verify the expressions for the expectation and variance of H_0 above. □

PROPOSITION 2.10.2. *Let $(V_k)_{k\geq 0}$ denote a simple random walk on $\mathbf{Z}$ jumping to the right with probability $r > 0$ and to the left with probability $l = 1 - r$. Define for each $a \in \mathbf{Z}$*

$$H_a := \inf\{k \geq 0 : V_k = a\}.$$

Then for $0 \leq a \leq b$,

$$\mathbf{P}[H_b \leq H_0 \mid V_0 = a] = \frac{(\frac{l}{r})^a - 1}{(\frac{l}{r})^b - 1}.$$

2.10.2. Random sums.

LEMMA 2.10.3. *Suppose that $(Y_k)_{k\in\mathbf{N}}$ is a sequence of independent and identically distributed random variables with finite mean and variance, and G is an independent random*

variable, also with finite mean and variance. Define

$$\Sigma := \sum_{k=1}^{G} Y_k,$$

then

$$\mathbf{E}[\Sigma] = \mathbf{E}[G]\mathbf{E}[Y_1]$$

and

$$\mathbf{E}[\Sigma^2] = \mathbf{E}[G(G-1)]\mathbf{E}[Y_1]^2 + \mathbf{E}[G]\mathbf{E}[Y_1^2].$$

PROOF. Let P_G denote the probability generating function of G; and by M_Σ and M_{Y_1} the moment generating functions of Σ and Y_1 respectively. Then it is a standard computation that

$$M_\Sigma(t) = P_G(M_{Y_1}(t)).$$

Differentiating,

$$M'_\Sigma(t) = P'_G(M_{Y_1}(t))M'_{Y_1}(t),$$

and

$$M''_\Sigma(t) = P''_G(M_{Y_1}(t))(M'_{Y_1}(t))^2 + P'_G(M_{Y_1}(t))M''_{Y_1}(t).$$

Taking $t = 0$, it is clear that $\mathbf{E}[\Sigma] = \mathbf{E}[G]\mathbf{E}[Y_1]$ and

$$\mathbf{E}[\Sigma^2] = P''_G(1)M'_{Y_1}(0)^2 + P'_G(1)M''_{Y_1}(0) = \mathbf{E}[G(G-1)]\mathbf{E}[Y_1]^2 + \mathbf{E}[G]\mathbf{E}[Y_1^2],$$

and the result follows. □

CHAPTER 3

GENETIC DIVERSITY IN POPULATIONS SUBDIVIDED INTO SELF-INCOMPATIBLE MATING TYPE CLASSES

Joint work with
GEORGE W. A. CONSTABLE, SARAH PENINGTON,
IAN LETTER AND DANIEL WILSON

The evolutionary mechanism that drove the establishment of self-incompatibility in early sexual eukaryotes is still a debated topic. While a number of competing hypotheses have been proposed, each has received mixed theoretical attention. In particular, the hypothesis that self-incompatibility increases the benefits of genetic recombination in sexual haploids has been comparatively understudied. In this paper we address this topic by mathematically deriving how the probability of mating with a genetically distinct individual changes as a function of the presence or absence of self-incompatible mating type classes. We find that although populations with mating types successfully engage in sexual reproduction less frequently than their self-compatible competitors, they can nevertheless mate with genetically distinct partners more frequently. This occurs when the number of sexual reproductive events per generation is low (ie in small populations with low rates of facultative sexual reproduction). Finally we demonstrate the potential for frequency-dependent selection in competitive dynamics between self-compatible and self-incompatible types. These analytic results provide a baseline for studying the sex advantage enhancer model for the evolutionary origin of mating types within each specific hypothesis for the evolution of recombination.

3.1. Introduction.

A primary challenge encountered by sexually reproducing organisms is that of locating a suitable mate [87]. We might therefore expect that any mutation that increases an organism's ability to find a mate should be selected for, while any mutation that decreases an organism's probability of encountering a mate should be selected against. Framed in this

way, it is perhaps surprising that the vast majority of sexual organisms are not free to mate indiscriminately, but rather must find a complementary partner [88].

Amongst most fungi, ciliates and algae, mating is restricted to occurring between individuals from distinct self-incompatible classes, termed mating types [51]. These mating types are genetically determined by mating type loci that can be crudely understood as ancestral forms of the more familiar sex chromosomes present in many animals [17]. Indeed, both empirical and theoretical studies support the view that mating types form an evolutionary precursor to the true sexes [86, 126]. Unlike the true sexes however, which are defined by the morphological dissimilarity of their sperm-egg gametes (anisogamy), the gametes of distinct mating types are morphologically similar (isogamy).

The lack of size dimorphism between the sex cells of isogamous species raises a number of interesting evolutionary possibilities. For instance, in the absence of the disruptive selection that drives the evolution of the two sexes, the number of mating types is not restricted to two, with up to 23,328 in fungi [82] and 100 in ciliates [109]. However this morphological similarity also raises evolutionary questions. In anisogamous species, the evolutionary stability of self-incompatibility is intuitively clear; selection against sperm-sperm syngamy[1] is driven by low viability of the resultant ill-provisioned zygote while selection for egg-egg syngamy is weakened by their low encounter rate. With such factors not present in isogamous species, we can then ask if self-incompatibility can remain evolutionarily stable in these systems and, more fundamentally, how it even first arose.

Most models of the early evolution of sexual reproduction suppose the existence of a "unisexual" early ancestor that mated indiscriminately and from which mating type self-incompatibility systems evolved [60]. However there is as yet no empirical evidence indicating that such a unisexual system preceded the mating type self-incompatibility systems in isogamous species. Therefore if this supposition is true (and this seems likely given the difficulty of evolving mechanisms for sexual reproduction and two self-incompatibility systems simultaneously) the rapid loss of this unisexual ancestor indicates a nascent advantage to self-compatibility in the evolution of early sex. We next turn our attention to the stability of isogamous self-incompatibility systems.

The loss of self-incompatibility has occurred in multiple isogamous species, with phylogenetic analyses showing that self-compatibility is a derived trait[2] evolving in the presence of pre-existing mating type self-incompatibility systems [111]. This is termed *homothalism*. Homothalism comes in many distinct forms [138] including: primary homothalism (individuals possess multiple ancestral mating type alleles on the same chromosome),

[1] *Syngamy* refers to the fusion of two cells during reproduction. The two fusing cells are called *gametes*. The resultant cell, termed the *zygote*, contains the genetic material from both gametes.

[2] In phylogenetics, an ancestral trait or characteristic is one that has been inherited from the last common ancestor of the lineage. Other traits or characteristics, which have appeared since this time, are *derived traits*.

same-sex mixing (individuals of the same mating type can mate) and mating type switching (individuals carry multiple mating type alleles with alternate mating types expressed in alternate generations of asexual progeny). However the effect of each of these mechanisms is to remove the barriers preventing the progeny of a parent (or recent ancestor in the case of mating type switching) from mating with each other. It is clear that the evolutionary advantage of homothalism is the improved chances of finding a mate for sexual reproduction, which can provide reproductive assurance for obligately sexual species or facultatively[3] sexual species in which sexual reproduction is associated with the formation of a resistant state. However in order to understand when self-incompatibility is selected against we must not just identify what the costs of self-incompatibility are, but also its benefits.

Multiple theories have been proposed as to the evolutionary pressures that may have driven the evolution of self-incompatibility in early sexual eukaryotic[4] ancestors (reviewed in [9]). Some of these assume a physiological advantage to self-incompatible mating types. These include the "by-product model" (that a bipolar pheromone-receptor recognition system[5] improves the efficiency of syngamy between gametes [52, 54, 61]) and the "ploidy model" (that self-incompatibility brings together distinct mating type alleles that in concert may signal a diploid form, with which syngamy should not be instigated [51]). Other models suggest that genetic conflict[6] may lie at the heart of the evolution of self-incompatibility. These include the "organelle inheritance model" (that self-incompatible alleles in fact encode for donor-receiver roles in the uniparental inheritance of cytoplasm[7] [53, 63]) and the "selfish genetic element model" (that self-incompatible alleles emerged as selfish genetic elements that evolved to target "uninfected" non-self gametes for mating [62]). A

[3]Reproduction refers to the general process whereby organisms replicate themselves. In *asexual reproduction*, individuals are able to reproduce without another organism; it does not involve the exchange of genetic material between members of the population and offspring are clones or near-clones of their parents. In *sexual reproduction*, new offspring are formed by the fusion of genetic material from two parents. A species is said to be *facultatively sexual* if it is able to reproduce both asexually and sexually.

[4]All organisms may be divided into two classes: eukaryotes and prokaryotes. *Eukaryotes* are organisms whose cells contain a well-defined nucleus carrying its genetic material enclosed in a membrane. All animals, plants and fungi are eukaryotes. Organisms which lack a well-defined nucleus are *prokaryotes*. Prokaryotes are single-celled organisms; for example all bacteria are prokaryotic.

[5]A *bipolar pheromone-receptor recognition mating system* is one where two mating type classes coexist within the population, defined by the exclusive presence of either a pheromone or pheromone receptor molecule. This bipolar system is used to identify and locate a mate.

[6]A genetic element is said to be *selfish* if, in spite of having no net benefit (or in some cases possessing a net-negative effect) to an organism's fitness, is nevertheless able to enhance its own transmission, for example through an increased productive rate. The resulting interplay between the positive selection for non-selfish genetic elements and competing self-promoting selfish elements is referred to as *genetic conflict*.

[7]*Uniparental inheritance* is a form of genetic inheritance whereby offspring inherit their genotype from a single parent. In some mammals — which reproduce sexually by fusing together the genetic material contained within an egg and sperm cell of the mother and father respectively — the inheritance of genetic material contained within the cell's cytoplasm (all material within the cell's membrane other than the nucleus) is, unlike the nucleic genetic material, uniparental. This is because virtually all of the cytoplasm of the resultant zygote is inherited from the egg cell, which is significantly larger than sperm.

final class of models, and the focus of interest for this paper, investigates how the costs and benefits to sex and genetic recombination may be respectively suppressed or amplified by self-incompatibility.

The "inbreeding depression avoidance model" suggests that as self-incompatible promotes disassortative mating[8], it reduces the rate at which rare deleterious mutations become homozygous in inbred offspring and may therefore be selected for [20, 128, 129]. However, because this hypothesis relies on species being diploid with sporophytic[9] self-incompatibility determined in this diploid stage [9], it is not appropriate for isogamous species which are mostly haploid with gametophytic self-incompatibility determined at the haploid stage. While the model is appropriate for self-incompatible alleles in plants, in this scenario self-incompatibility has evolved on top of a pre-existing anisogamous mating system and therefore lays outside the scope of this paper.

In diploid populations, there has been a considerable amount of attention devoted to considering the number of self-incompatible mating type classes when this type is determined by two alleles at a single locus (see, eg Vekemans and Slatkin [130], Clark [21], Karlin and Lessard [71, 72], and references therein). In these models, the (plant) populations are obligately sexual. Whilst the formulation of the models in these works permits a thorough mathematical treatment, the models do not extend to the haploid populations which are the subject of this paper. In contrast, the "sex advantage enhancer model", which suggests that self-incompatibility amplifies the benefits of genetic recombination[10], is suitable for haploid isogamous species. However, with rare exceptions [26, 101], it has received little theoretical attention.

From a modelling perspective, the reason that the sex advantage enhancer model has been tackled infrequently is clear. The hypothesis is straightforward to to formulate verbally: the benefits of recombination in haploids can only be realized if syngamy occurs between non-identical haploid clones, and the probability this occurs should be increased by the presence of self-incompatible alleles in the population. However, it is significantly harder to formulate the hypothesis mathematically; the precise mechanisms that drive selection for genetic recombination are themselves still an intensely debated subject [1]. A complete investigation of the feasibility of this hypothesis therefore requires that the ef-

[8] *Disassortative mating* is when organisms with more dissimilar genotypes are more likely to mate with one another than would be expected if mates were chosen merely at random.

[9] The life cycle of all land plants and many algae consists of two distinct alternating haploid sexual and diploid asexual phases. A *sporophyte* is the diploid multi-cellular stage. Mature sporophytes reproduce asexually; their offspring are haploid and contain half the genetic material of the parent. These haploid offspring mature into *gametophytes* which reproduce sexually: they form gametes which fuse to form diploid sporophytes containing the genetic material from both parents.

[10] *Genetic recombination* occurs during meiosis, the reproductive process whereby diploid cells divide to produce two offspring each containing half the genetic material of its parent. The genetic material of a single offspring is not an identical copy of either of the two copies held by its diploid parent but, rather, a mixture of the two. The process through which this mixture is formed is genetic recombination.

fects of self-incompatibility are tested within each recombination hypothesis framework (eg Fisher–Muller hypothesis[11], the Hill–Robertson effect[12], and the Red Queen hypothesis[13] [58]), which themselves only yield selection for recombination under a limited range of scenarios.

An alternative approach is to abandon any attempt to model selection for recombination and instead take a courser grained perspective. For instance one could impose an advantage to non-self mating 'by hand', such that the progeny of self-incompatible parents have a enhanced reproductive rate [26, 101]. However this leads to a rather circular argument (self-incompatible is selected for if the reproductive rate enhancement exceeds the cost of mate finding) and does not capture the feedback between sexual reproduction, self-incompatibility and population diversity.

In this paper we take a simplified view that neglects specific hypotheses for the benefits of recombination, but nevertheless cuts to the core of the verbal formulation of the sex advantage enhancer model. In particular, by investigating the relationship between population-level genetic diversity and self-incompatible systems, we calculate how the probability of mating with a genetically distinct individual varies as a function of the presence or absence of mating type alleles and with their varied number. We also investigate how these factors interplay with the rate of facultative sex in the population, which recent theoretical studies have suggested may play a role in the evolution of mating type number [23, 27, 83]. In the context of the models of the evolutionary origins of self-incompatibility, considering the role of facultative sex is important as it is highly likely that sexual reproduction was rare in its early evolutionary history.

3.2. The model.

We consider a haploid population in which the mating type of an individual is determined by one of an infinite set of potential alleles at a single locus. We shall assume that the genome is long and this locus is small so that we may approximate its position by a single point m within the normalized genome length $[0, 1]$ (see Figure 3.1). Using a Moran

[11]Under the Fisher–Muller hypothesis, sex benefits a population as a whole (but not single individuals within it) by increasing the genetic diversity of offspring (through genetic recombination and merging of genetic material from their two parents) and thus making less susceptible to environmental changes than their asexually reproducing competitors.

[12]The Hill–Robertson effect describes a phenomenon whereby the increased likelihood of inheriting genetic material close to one another in an gene sequence from the same parent hampers the efficiency of selection. The effect leads one to conclude that genetic recombination is selected for, so that the likelihood of inheriting nearby genes from the same parent is reduced.

[13]The Red Queen hypothesis states that organisms must continuously evolve, not just in order to adapt to changing environments, but also because their competitors are. Like the Fisher–Muller hypothesis, it is therefore proposed that genetic recombination is selected for, so conferring an increased genetic diversity of offspring and increasing the speed of evolution.

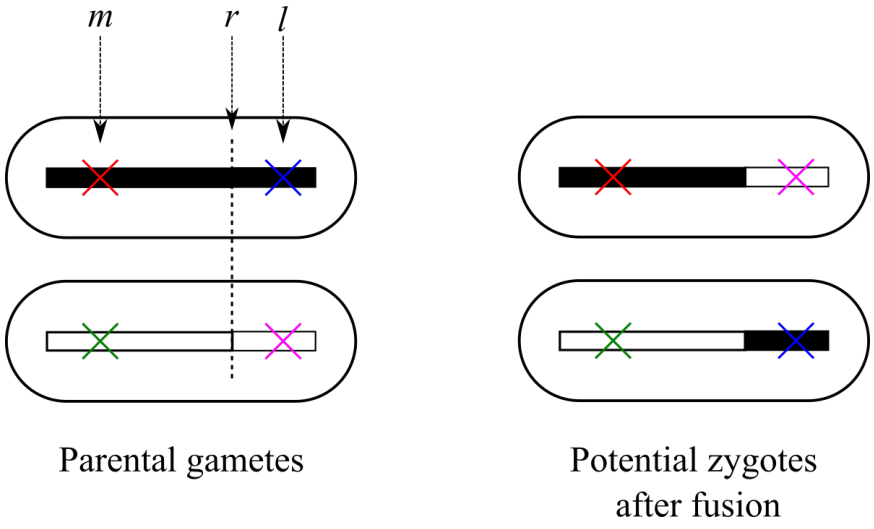


FIGURE 3.1. Each individual haploid cell in a finite population of size N begins a reproductive event at rate N. With probability $1-\bar{\sigma}$, the event is asexual, and the parent cell produces a clone of itself. With probability $\bar{\sigma}$ the event is sexual. The focal parents seeks a compatible mate and succeeds in doing so with probability $\theta_i^N(c)$. If successful, a recombination site r is selected uniformly on the genome, modelled by the unit interval $[0,1]$. The zygote inherits all genetic material to the left of r from the focal parent and from the right of r from the chosen mate, or *vice versa* with equal probability. In either the case of asexual or sexual reproduction and independently of everything else, the allele at the neutral locus l is incorrectly copied from parent to daughter cell with probability $\bar{\mu}$. The daughter cell replaces an individual of its mating type chosen uniformly at random.

modelling framework in which birth and death events are coupled, the population is fixed to a constant size N.

For some fixed parameter $\bar{\sigma} \in [0,1]$, we suppose that each individual reproduces asexually at rate $N(1-\bar{\sigma})$. Upon being selected for an asexual reproduction event, an individual is created that inherits its parental mating type. Meanwhile, at rate $N\bar{\sigma}$, each individual attempts to instigate a sexual reproduction event. Observe that $\bar{\sigma}$ approximately captures the rate of sexual to asexual reproduction. Furthermore, we see on average N^2 reproductive events in each unit of time, corresponding roughly to N generations which, more accurately, are non-overlapping. For facultatively sexual populations ($\bar{\sigma} > 0$), we will consider three scenarios for the mating dynamics.

Firstly, we will consider a 'unisexual' population in which the chosen individual can mate with any other individual in the population irrespective of mating type. The probability that this individual can successfully mate with another chosen randomly from the population is thus one. (We choose the terminology unisexual to emphasise that we here may be capturing a sexual form that preceded mating type self-incompatibility [111]). So to preserve the total population size N, birth and death events will be coupled and the newborn individual replaces an individual chosen uniformly at random from the existing

population.

Secondly we consider a population with k self-incompatible mating types. We denote by N_i the number of individuals of mating type i, such that $\sum_{i=1}^{k} N_i = N$. Having been chosen to attempt a sexual reproduction event, an individual begins what we term a *mating round* (described below) during which they are given a number of attempts to randomly select a partner carrying a distinct mating type allele from the population. The resultant zygote will carry genetic material which is a combination of that of its parents. How this is formed is specified below. Mating type subpopulation sizes will be fixed and, to achieve this, we specify that a newborn individual replaces an individual of the inherited mating type from the existing population.

Finally we consider a case with $k-1$ self-incompatible mating types and a single unisexual (self-compatible) type. If one of the first $k-1$ self-incompatible types is selected for sexual reproduction it instigates a mating round during which it attempts to find a suitable mate. Meanwhile if the unisexual type is selected for sexual reproduction, it finds a mate with probability one. As in the previous instance, zygotes will carry genetic material which is a combination of that of its parent and replace an individual of the same mating type chosen from the existing population (again preserving subpopulation sizes.)

3.2.1. Mating rounds. Self-incompatible mating types are, by definition, unable to mate with those of the same mating type. They are however permitted a random number of attempts to find a compatible mate. During each attempt, a potential mate is selected uniformly at random from the $N-1$ remaining individuals in the population. The attempt is rejected if the selected mate is of the same mating type and accepted otherwise. The mating round ends if either all allowed attempts are rejected or after the first reproduction with a compatible partner. (Individuals thus sexually reproduce at most once per mating round.)

The number of attempts an individual makes will be random and we will assume a Gibbs-type distribution for the number of attempts: for a fixed parameter $c > 0$, if A is the random number of attempts made, we will take $\mathbf{P}[A = n] \propto e^{-cn}$ for all $n \geq 1$ and $\mathbf{P}[A = n] = 0$ otherwise. In particular, this implies that N is geometrically distributed with parameter $1 - e^{-c}$ and

$$\mathbf{P}[A = n] = e^{-c(n-1)}(1 - e^{-c}), \qquad n \geq 1.$$

We note that the limit $c \to \infty$ corresponds to each individual having exactly one chance to find a compatible mate (ie a large cost to mate searching) and the limit $c \to 0$ corresponds to case where individuals may make as many attempts as necessary to find a compatible mate (ie no cost to mate searching). The parameter c can thus be understood as interpolating between mass-action encounter rates at one extreme ($c \to \infty$) and active mate searching on the other.

For fixed $c > 0$, an individual of mating type i mates with an individual of mating type $j \neq i$ with probability

$$\theta_{ij}^N(c) := \sum_{n=1}^{\infty} \mathbf{P}[A = n] \sum_{m=1}^{n} \Big(\frac{N_i - 1}{N - 1}\Big)^{m-1} \frac{N_j}{N - 1} = \frac{1}{1 - e^{-c\frac{N_i - 1}{N-1}}} \frac{N_j}{N - 1}, \qquad (3.2.1)$$

and thus successfully finds a compatible mate during a mating round with probability

$$\theta_i^N(c) := \sum_{j \neq i} \theta_{ij}^N(c) = \frac{1}{1 - \frac{N_i - 1}{N-1} e^{-c}} \Big(1 - \frac{N_i - 1}{N - 1}\Big).$$

3.2.2. Recombination. If the individual is successful in finding a compatible mate during the mating round the juvenile inherits genetic material from both parents as follows. A recombination site r is selected uniformly at random on the unit interval $[0, 1]$ and the offspring inherits all genetic material to the left of r from one parent and the genetic material to the right of r from the other. The juvenile replaces an individual chosen uniformly at random from the subpopulation of its mating type so that birth and death events are coupled and subpopulation sizes remain fixed. Recombination events are depicted in Figure 3.1.

We also consider a highly polymorphic neutral genetic marker at position $l \in [0, 1]$ on the normalised genome. It is one of an infinite set of potential alleles denoted by elements in the set $\{a_1, a_2, \ldots\}$. The allelic type of an individual at l is assumed to have no influence on its mating type. We will denote by $d = |m - l|$ the distance on the genome between the locus determining the individual's mating type and the locus determining its allelic class. In a sexual reproduction event, both mating and allelic types may be inherited from either parent or they may individually be inherited from distinct parents. As the recombination site is chosen uniformly on the interval $[0, 1]$ and the loci carrying the mating type and allelic type are at distance d apart, the probability that these two traits are inherited from distinct parents is d. Calling the individual which attempts sexual reproduction and selects a mate the *focal parent*, the probability of each possible outcome can be tabulated as in Table 3.1.

We note that the assumption that there is only one recombination site significantly simplifies our analysis. In the case where we consider genetic diversity on a separate chromosome, we set $d = 1/2$ (ie there is no linkage between the mating type allele and the neutral marker).

3.2.3. Mutation. In addition we will suppose that in either the case of sexual or asexual reproduction, the genetic material at the neutral locus is incorrectly transmitted to the progeny with probability $\bar{\mu} \in [0, 1]$. It is assumed that the mutation is to a selectively neutral allelic type not yet seen in the population.

Event type	Successful	Recombination	Mutation at l	Probability
Asexual	-	-	✓	$(1-\bar{\sigma})\bar{\mu}$
	-	-	x	$(1-\bar{\sigma})(1-\bar{\mu})$
Sexual	x	-	-	$\bar{\sigma}(1-\theta_i^N(c))$
	✓	✓	✓	$\bar{\sigma}\theta_i^N(c)d\bar{\mu}$
	✓	✓	x	$\bar{\sigma}\theta_i^N(c)d(1-\bar{\mu})$
	✓	x	✓	$\bar{\sigma}\theta_i^N(c)(1-d)\bar{\mu}$
	✓	x	x	$\bar{\sigma}\theta_i^N(c)(1-d)(1-\bar{\mu})$

TABLE 3.1. Inheritance table showing possible outcomes and their probabilities when an individual of the ith mating type begins a reproductive event.

3.3. Mathematical analysis.

In what follows we outline the results of our mathematical analysis. Proofs and mathematical background may be found in the appendix to this chapter.

3.3.1. Allele frequencies at the neutral marker. In this section we analyse the dynamics of the allele frequencies at the neutral marker l under the different mating systems specified in the previous section. As a proxy for a measure of 'genetic diversity', we calculate the probability that two individuals sampled randomly from the population are homozygous at the neutral marker l in each of the cases of an asexual population, a unisexual population and a population with k self-incompatible mating types respectively. We then proceed to approximate this same quantity for a mixed population with a single unisexual type and $k-1$ self-incompatible types.

Readers less interested in the specific mathematical details may proceed to the Results section.

3.3.1.1. The rare sex, rare mutation regime. Our interest lies in the case that both sex, and mutations at the locus l are rare; in particular, it will be assumed that for some parameters $\sigma > 0$ and $\mu > 0$,

$$\bar{\sigma} = \frac{\sigma}{N},$$

and

$$\bar{\mu} = \frac{\mu}{N}.$$

The parameter σ may be interpreted as the average number of initiated sexual mating rounds per N reproductive events. The parameter μ may be interpreted as the average number of successful reproductive events which terminate with a mutation per N successful reproductive events. We stress that this is per *successful* reproductive event noting that when an individual begins a reproductive event it may not result in the successful production of offspring if the event is sexual and a compatible mate is not found. Moreover, as the

average number of reproductive events per unit time varies with sexual reproduction (from N in the asexual $\bar{\sigma} = 0$ case to $\sum_{i=1}^{k} N_i \theta_i^N(c)$ in the obligately sexual $\bar{\sigma} = 1$ case) so too does the number of mutations per N reproductive events. This leads to a natural definition of the effective rate of mutation given by

$$\mu_{\text{eff}}^N := \Big((1-\bar{\sigma}) + \sum_{i=1}^{k} \frac{N_i}{N} \bar{\sigma} \theta_i^N(c)\Big)\mu,$$

which is per individual rate of introducing a new mutation to the population. Equally, μ_{eff}^N is the average number of mutations per N reproductive events.

In the same vein, we may define an effective recombination rate

$$\rho_{\text{eff}}^N := \sum_{i=1}^{k} \frac{N_i}{N} \sigma \theta_i^N(c) d,$$

that is, the average per individual rate of beginning a sexual reproductive event with a crossover between the mating type allele and allele at the neutral marker. Equally, ρ_{eff}^N is the average number of crossovers between these same alleles per N reproductive events. The metric ρ_{eff}^N measures the degree of gene flow between self-incompatible subpopulations.

3.3.1.2. Diffusion approximation of allele frequencies at the neutral marker. In the rare sex, rare mutation regime (as described in the previous section), as $N \to \infty$ the allele frequencies within mating type subpopulations are well approximated by a k-dimensional diffusion process given by the solution to a system of stochastic differential equation (SDE). A brief account of diffusion processes and how they may be used to approximate continuous-time Markov chains is given in the appendix.

Recall that the allelic type at the neutral locus l is one of an infinite set of potential alleles denoted by elements in the set $\{a_1, a_2, \ldots\}$. For the reference type, a_1, denote for each $i = 1, \ldots, k$,

$$X_i^N(t) := \frac{\#\{\text{Type } i \text{ individuals carrying } a_1 \text{ at } l \text{ at time } t\}}{N_i}.$$

That is, $X_i^N(t)$ denotes the proportion of individuals of the ith mating type which carry the allele a_1 at the neutral locus l at time t. Passing to the large population limit ($N \to \infty$), we further define p_i as the fraction of individuals of mating type i ($N_i/N \to p_i$) and x_i as the frequency of a_1 in the ith mating type class ($X_i^N(0) \to x_i$). In the appendix section titled 'Proof of the diffusion approximation' we then show that X^N converges to the solution of the system of stochastic differential equations given by

$$X_i(t) = x_i - \int_0^t \mu X_i(s)\,\mathrm{d}s + \int_0^t \sum_{j\neq i} m_{ij}(p,c)(X_j(s) - X_i(s))\,\mathrm{d}s + \int_0^t \sqrt{2p_i^{-1}X_i(s)(1-X_i(s))}\,\mathrm{d}B_i(s), \quad (3.3.1)$$

where, for each $i, j = 1, \ldots, k$, B_i is an independent Brownian motion and

$$m_{ij}(p,c) := \begin{cases} \dfrac{\sigma d}{2}\dfrac{p_i\theta_{ij}(c) + p_j\theta_{ji}(c)}{p_i}, & i \neq j, \\ 0, & i = j. \end{cases} \quad (3.3.2)$$

3.3.2. Genealogy. By recording only the proportion of each subpopulation carrying the ancestral type a_1 at the neutral marker, $X_i(t)$, we do not retain information about how individuals are related to one another. However, looking backwards-in-time, it is possible to describe the genealogical trees relating individuals in a random sample from the population.

We wish to establish the genealogy at the neutral marker l of a finite sample of individuals taken from the population X^N. Denote by $Y_i^N(0)$ the number from this sample of mating type i. We trace the ancestry of the neutral marker at the neutral locus, denoting by $Y_i^N(t)$ the number of individuals of the ith mating class that are ancestors to the original sample t units of time into the past.

Passing to the large population limit $N \to \infty$, it is shown in the appendix that the process $Y^N = (Y_1^N, \ldots, Y_k^N)$ approximates a structured coalescent process Y which evolves as follows. If $Y(t) = (y_1, \ldots, y_k)$ then

1. each pair of type i ancestral lines *coalesces* at rate $2/p_i$, independently of one another; ie, $y_i \mapsto y_i - 1$ at rate
$$\frac{2}{p_i}\binom{y_i}{2};$$
2. each ancestral line of type i becomes an ancestral line of type j at rate $m_{ij}(p,c)$, independently of one another; ie, $(y_i, y_j) \mapsto (y_i - 1, y_j + 1)$ at rate
$$y_i m_{ij}(p,c);$$
and
3. each ancestral line experiences a mutation at total rate μ, independently of one another. (There is no change to the number of individuals that are ancestors to the original sample during this event.)

The result is proved in the appendix. An illustration of this process is depicted in Figure 3.2.

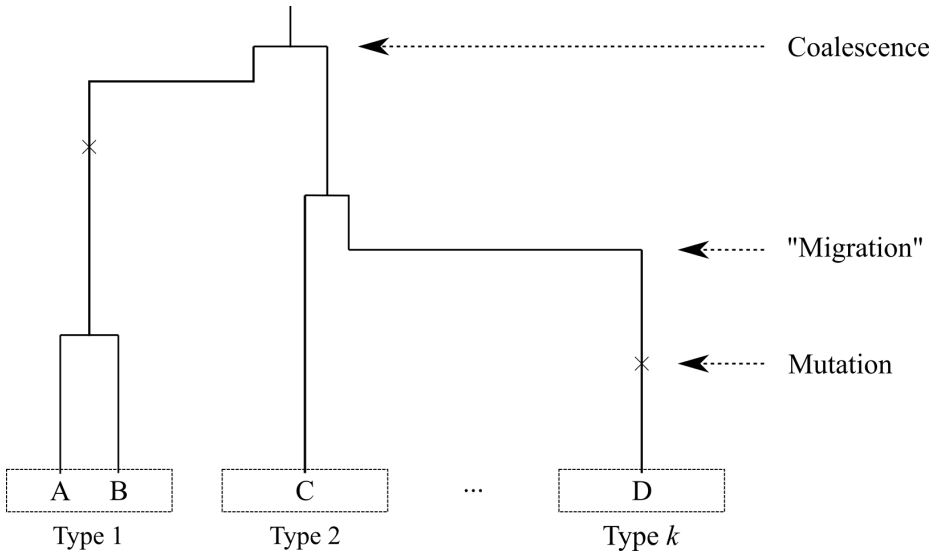


FIGURE 3.2. Figure illustrating the dynamics of the genealogical process $Y(t)$. In this realisation, we see the genealogical history of a sample of four individuals A, B, C and D. The individual C is of the ancestral type, having experienced no mutations in its genealogical history before the time of the last common ancestor of the sample. Individuals A and B are of the same type at the neutral locus. All individuals in the sample have an allelic type ancestral to an individual originally of the second mating type.

The parameter $m_{ij}(p,c)$ may thus be thought of as an effective per individual rate of a crossover between a type i and type j individual. Models for genealogies of this type are often referred to as a structured coalescent, or island model (cf Durrett [32]) and, since the recombination events (item 2 above) manifest as an effective migration of ancestors between mating type subpopulations, we will usually refer to such events as a "migration". This terminology is also consistent with the usual setting of the structured coalescent and the island model.

In the appendix, we demonstrate that the processes X and Y are dual in the sense that given $X(0) = x$ and $Y(0) = y$,

$$\mathbf{E}\Big[\prod_{i=1}^{k} X_i(t)^{y_i}\Big] = \mathbf{E}\Big[\prod_{i=1}^{k} x_i^{Y_i(t)} \exp\Big\{-\mu \int_0^t \sum_{i=1}^{k} Y_i(s)\,\mathrm{d}s\Big\}\Big].$$

Relations of this kind are frequently of interest. In the appendix, we make explicit use of this duality in proving the uniqueness of solutions to the system of equations (3.3.1).

3.3.3. Identity by descent. It is clear from our description of the genealogical process that the two individuals sampled from the population will have a common ancestor if we look far enough back into the past. We will say that the two individuals are *identical by descent* if they are of the same allelic type at the locus l; this will be the case if neither of

the sampled individuals experience a mutation before their eventual coalescence.

We calculate the probability that two individuals sampled from the population are identical by descent, first in the case of *monomorphic populations* (ie when the mating dynamics are uniform across all individuals) and secondly in the case of mixed populations (ie when some individuals are self-incompatible, but others are unisexual).

3.3.3.1. Monomorphic populations. Let us define π_{ij} to be the probability that an individual sampled from the type i mating type subpopulation and an individual sample from the type j subpopulation (distinct from the first if $i = j$) are identical by descent, ie the probability that they carry the same allelic type at the neutral marker l.

We begin by considering a population of k self-incompatible mating types. (Monomorphic asexual populations and unisexual populations can be obtained as limiting forms of this scenario.) In the appendix, we derive recursions for the probabilities π_{ij}, $i, j = 1, 2, \ldots, k$, however, we are not able to find tractable solutions in this generality. To make progress, we set $p_i = 1/k$ for all i. This approach has previously been taken by Ennos and Hu [36], and in the case of k self-incompatible types is equivalent to working in the regime $k/N \ll 1$ and $c > 0$, which is a biologically reasonable regime for scenarios in which mating type extinctions are rare [28].

The above simplification induces some additional symmetry to the system and increases the analytic tractability of the problem. In this setting, only two quantities are of interest; they are π_d, the probability that two individuals sampled from distinct mating type subpopulations are identical by descent, and π_s, the probability that two individuals sampled from the same subpopulation are identical by descent. Observe that $\pi_{ij} \equiv \pi_{12} =: \pi_d$ if $i \neq j$ and $\pi_{ij} \equiv \pi_{11} =: \pi_s$ if $i = j$. Similarly, we can obtain an expression for $\theta_i^N(c)$ (see equation (3.2.1)), where, in the large population limit,

$$\theta_i(c) \equiv \theta_1(c) := \lim_{N\to\infty} \theta_i^N(c) = \frac{k-1}{k-e^{-c}},$$

and

$$\theta_{ij}(c) \equiv \theta_{12}(c) := \lim_{N\to\infty} \theta_{ij}^N(c) = \frac{1}{k-e^{-c}}, \tag{3.3.3}$$

We can now define the "migration" rate, m, (see equation (3.3.2)), which is now equal across types, by

$$m_{ij} \equiv m_{12} =: m = \frac{\sigma d}{k-e^{-c}}. \tag{3.3.4}$$

Recall that by "migration" we are referring to the effective migration of lineages between subpopulations corresponding to gene flow between types resulting from a recombination.

In the appendix it is shown that

$$\pi_s = \frac{k[1+(k-1)\frac{m}{\mu+m}]^{-1}}{k[1+(k-1)\frac{m}{\mu+m}]^{-1}+\mu}, \tag{3.3.5}$$

and

$$\pi_d = \frac{m}{\mu+m}\frac{k[1+(k-1)\frac{m}{\mu+m}]^{-1}}{k[1+(k-1)\frac{m}{\mu+m}]^{-1}+\mu}. \tag{3.3.6}$$

We similarly define π_{tot} to be the probability that two distinct individuals chosen uniformly at random from the *whole* population, are identical by descent. Since the pair are of the same mating type with probability $1/k$ and of distinct types with probability $(k-1)/k$, the quantity π_{tot} is given by the weighted average of π_s and π_d and, specifically

$$\begin{aligned}\pi_{\text{tot}} &= \frac{1}{k}\pi_s + \frac{k-1}{k}\pi_d \\ &= \frac{1}{k[1+(k-1)\frac{m}{\mu+m}]^{-1}} \cdot \frac{k[1+(k-1)\frac{m}{\mu+m}]^{-1}}{k[1+(k-1)\frac{m}{\mu+m}]^{-1}+\mu}.\end{aligned} \tag{3.3.7}$$

In the case of (monomorphic) asexual populations we may obtain the equivalent quantities by taking $\sigma = 0$. The biologically relevant case is $k = 1$, when the population consists of a single well-mixed asexually reproducing population, however, it will be convenient to consider the case where k separated subpopulations reproduce asexually. In this setting, the relevant quantities are

$$\pi_s|_{\sigma=0} = \frac{k}{k+\mu}, \tag{3.3.8}$$

the probability that two distinct individuals sampled from the same 'isolated' subpopulation are identical by descent, and

$$\pi_{\text{tot}}|_{\sigma=0} = \frac{1}{k}\cdot\frac{k}{k+\mu}. \tag{3.3.9}$$

Finally we consider a unisexual population. For a unisexual population, there is only a single well-mixed population ($k = 1$). Importantly, while sexual reproduction is now possible between all individuals, this in fact has no impact on the probability of identity by descent at the neutral locus. Without the population structure imposed by self-incompatibility, the genealogies are the same as that of an asexual population. We thus find that the probability of identity by descent in the unisexual population is given by

$$\pi_{\text{tot}}|_{k=1} = \frac{1}{1+\mu}. \tag{3.3.10}$$

3.3.3.2. Mixed populations. We next consider a mixed population consisting of $k-1$ self-incompatible mating types and a single unisexual type.

Since a unisexual individual is always able to find a mate, it has a higher per capita fecundity than its self-incompatible competitors. Moreover, in the absence of any costs to unisexuality, the unisexual type always sweeps to fixation. We therefore consider three scenarios in which the probabilities π_{ij} are calculated. In the first scenario the frequency of the unisexual type is held at a constant frequency with respect to the $k-1$ self-incompatible types (each at equal frequency). In the second, a rare unisexual type arises in a population of $k-1$ self-incompatible types at equilibrium. In the third, $k-1$ rare self-incompatible types arise in a population of the unisexual type at equilibrium. Of course, in this last case, the biologically relevant case is $k=2$, when a single self-compatible type emerges. Since the machinery developed allows us to consider general k, we do so for reasons of symmetry.

We first consider the setting of a population subdivided into k allelic classes, one of which is unisexual. We will, without loss of generality, denote by $u \in \{1,2,\ldots,k\}$ the index of the unisexual subpopulation. For the sake of clarity, in this revised model individuals of type u are able to mate indiscriminately, and so always find a compatible mate, and individuals of type $i \neq u$ may mate only with those belonging to a class distinct from that of themselves.

In this setting, the probabilities θ_{ij}^N, as computed in equation (3.2.1), are now given by

$$\theta_{ij}^N(c) = \begin{cases} \dfrac{1}{1-e^{-c\frac{N_i-1}{N-1}}}\dfrac{N_j}{N-1}, & i \neq u, \\ \dfrac{N_j}{N-1}, & i = u, \end{cases}$$

converging, as $N \to \infty$, to

$$\theta_{ij}(c) = \begin{cases} \dfrac{1}{1-e^{-c}p_i}p_j, & i \neq u, \\ p_j, & i = u. \end{cases} \tag{3.3.11}$$

Following the mathematical reasoning and methodology above, including the proofs in the appendix, it is established that the proportion of type i individuals carrying the ancestral type at the neutral marker l converges to the system of stochastic differential equations defined in equation (3.3.1), with the "migration" rates m_{ij} defined in equation (3.3.2) now

depending instead on the new calculation of θ_{ij}. To be explicit, they become

$$m_{ij}(\theta, c) := \begin{cases} \dfrac{\sigma d}{2}\left(\dfrac{1}{1-e^{-c}p_i} + \dfrac{1}{1-e^{-c}p_j}\right)p_j, & i, j \neq u, \\ \dfrac{\sigma d}{2}\left(\dfrac{1}{1-e^{-c}p_i} + 1\right)p_j, & i \neq u, j = u, \\ \dfrac{\sigma d}{2}\left(1 + \dfrac{1}{1-e^{-c}p_j}\right)p_j, & i = u, j \neq u, \\ 0, & i = j. \end{cases}$$

Similarly, the backwards-in-time genealogical process converges to the same structured coalescent as above but with these migration rates. From this limiting genealogical process we wish to compute the probabilities of identity by descent π_{ij} but again find that whilst theoretically possible, it is mathematically impractical to do so without imposing some further symmetry. We will assume that a proportion $\alpha \in (0, 1)$ belong to the unisexual class, u, and so set $p_u = \alpha$, and that the remaining classes are held in constant frequency, setting $p_i = (1-\alpha)/(k-1)$ for all $i \neq u$. In this case we may write

$$m_{ij}(\theta, c) := \begin{cases} m, & i, j \neq u, \\ m_{0u}, & i \neq u, j = u, \\ m_{u0}, & i = u, j \neq u, \\ 0, & i = j, \end{cases}$$

where

$$m = \sigma d \frac{1}{1 - e^{-c\frac{1-\alpha}{k-1}}} \frac{1-\alpha}{k-1},$$

$$m_{0u} = \frac{\sigma d}{2}\left(\frac{1}{1 - e^{-c\frac{1-\alpha}{k-1}}} + 1\right)\alpha,$$

and

$$m_{u0} = \frac{\sigma d}{2}\left(1 + \frac{1}{1 - e^{-c\frac{1-\alpha}{k-1}}}\right)\frac{1-\alpha}{k-1}.$$

These migration rates are depicted in Figure 3.3.

Let us denote by π_s the probability that two individuals chosen from the same self-incompatible class are identical by descent; by π_d the same probability, but for two individuals chosen from distinct self-incompatible classes; by π_s^u again the same probability, but for two individuals both of the unisexual class; and by π_d^u the same probability, but for two individuals, one of the unisexual class and one of a self-incompatible class. A system of equations given by (3.6.15), (3.6.16), (3.6.17) and (3.6.18) for the probabilities π_s, π_d, π_s^u and π_d^u is derived in the appendix.

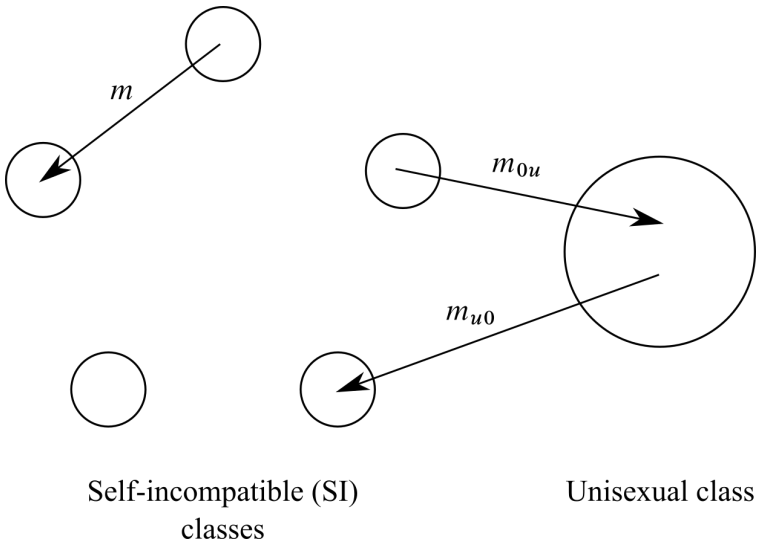


FIGURE 3.3. Migration rates between mating type classes in a population composed of both self-incompatible and unisexual types.

Though straightforward to evaluate numerically, the resulting equations are too algebraically complex to be worthwhile reproducing here. Tractable analytic solutions are available in case that $\alpha = 1/k$ (a case of interest in our simulations) and in the limit $c \to \infty$. In this setting $m = m_{0u} = m_{u0} = \sigma d/k$ and two lineages within the same class coalesce at rate $2k$, independent of whether this class is self-incompatible or unisexual. Because of this symmetry, $\pi_s = \pi_s^u$ and $\pi_d = \pi_d^u$ and the system of equations (3.6.15), (3.6.16), (3.6.17) and (3.6.18) reduces to that of a population consisting of k self-incompatible types held in constant frequency (with $c \to \infty$). In particular, in this special case,

$$\pi_s = \pi_s^u = \frac{k\mu + \sigma d}{k\mu + \sigma d + \mu(\mu + \sigma d)},$$

$$\pi_d = \pi_d^u = \frac{\sigma d}{k\mu + \sigma d + \mu(\mu + \sigma d)}.$$

Suppose next that we have a single unisexual mutant emerge in a self-incompatible population with $k-1$ mating types, leading to k types (including the unisexual type) in the population in total. Initially the population is in the equilibrium described by equations (3.3.5), (3.3.6), (3.3.7) with k replaced by $k-1$, but otherwise unaffected by the presence of the rare mutant. Initially the mutant subpopulation will be fixed for single neutral marker, and so the probability that two individuals both chosen from the new unisexual type are identical by descent is (initially) one.

Next we consider the probability that the new unisexual mutant and an individual sampled from the remaining (self-incompatible population) are identical by descent. Immediately after the appearance of this mutant, these probabilities are identical to that of its

ancestral subpopulation. Moreover, if the second individual belongs to the mutant's ancestral population (with probability $1/(k-1)$, then they are identical by descent with probability given by (3.3.5)) and otherwise with probability given by (3.3.6). This leads to the following expressions for the probability of identity by descent for individuals sampled from the same (π_s and π_s^u) or distinct (π_d and π_d^u) subpopulations, for self-incompatible and unisexual types respectively:

$$\pi_s = \frac{(k-1)[1+(k-2)\frac{m}{\mu+m}]^{-1}}{(k-1)[1+(k-2)\frac{m}{\mu+m}]^{-1}+\mu}, \tag{3.3.12}$$

$$\pi_d = \frac{m}{\mu+m}\frac{(k-1)[1+(k-2)\frac{m}{\mu+m}]^{-1}}{(k-1)[1+(k-2)\frac{m}{\mu+m}]^{-1}+\mu}, \tag{3.3.13}$$

$$\pi_s^u = 1, \tag{3.3.14}$$

$$\pi_d^u = \frac{1}{k-1}\pi_s + \frac{k-2}{k-1}\pi_d, \tag{3.3.15}$$

and m taken from equation (3.3.4).

Finally, we envisage a scenario in which $k-1$ mutant self-incompatible types emerge in a unisexual population at equilibrium. We note that this is a rather artificial scenario (except in the case $k = 2$), in the sense that it is unlikely that multiple self-incompatible mating types will evolve simultaneously. Nevertheless we consider this case as the natural symmetric situation to that discussed previously. The probability of identity by descent for individuals sampled from within the unisexual subpopulation will be initially unchanged by the presence of these rare mutants (ie will still be given by equation (3.3.10)), a property that will also be inherited by individuals sampled between the ancestral unisexual population and mutant self-incompatible populations who, immediately after their emergence, have yet to experience mutations at the neutral marker. In the unlikely event that two individuals are sampled from within a given self-incompatible subpopulation, the probability of identity by descent is one. This leads to the following expressions for the probability of identity by descent:

$$\pi_s = 1, \tag{3.3.16}$$

$$\pi_d = \pi_s^u = \pi_d^u = \frac{1}{1+\mu}. \tag{3.3.17}$$

3.4. Results.

The calculations of the probability of identity by descent in the large population limit, accessible because of the tractability of the limiting genealogy, allow us to consider notions of *genetic diversity* and what we term the *rate of effectual sex*. We consider each of these

in turn.

3.4.1. Genetic diversity. Nei [102] defines the *heterozygosity* or *gene diversity* of a population at a fixed locus l to be the probability that two individuals sampled uniformly at random carry distinct alleles at l. Since we are considering a haploid population, we prefer the term gene diversity and will use this exclusively henceforth. Nei's measure of diversity, which we denote by h_{tot} and h_{sub} for the total population and mating type subpopulations respectively, can be expressed as $h_{\text{sub}} = 1 - \pi_s$ and $h_{\text{tot}} = 1 - \pi_{\text{tot}}$. We proceed to consider these calculations in scenarios of particular interest, focusing on how genetic diversity within the monomorphic populations changes as a function of the population parameters.

3.4.1.1. Monomorphic asexual population. The first case of interest, and used as a basis for comparison, is that of a purely asexually reproducing population. This situation is recovered by taking $\sigma = 0$ in equation (3.3.7).

The case $k = 1$ corresponds to the whole population reproducing according to Moran model dynamics and the subsequent genealogical process being given by the unstructured Kingman coalescent. In this case, π_d has no meaning since the single 'subpopulation' and the total population coincide. The probability of identity is given by

$$\pi_{\text{tot}}|_{\sigma=0,k=1} = \pi_s|_{\sigma=0,k=1} = \frac{1}{1+\mu}.$$

Thus Nei's gene diversity, $h_{\text{tot}} = 1 - \pi_{\text{tot}}$ increases as a function of mutation probability, $\bar{\mu}$, and population size, N, where $\mu = \bar{\mu}N$.

While mating type classes are lost when $\sigma = 0$, it is nonetheless useful as a basis of comparison to consider the above quantities when the mating type frequencies are held constant artificially at $1/k$. In this case, the population is reduced to a set of k non-interacting asexually reproducing subpopulations, and

$$\pi_s|_{\sigma=0} = \frac{k}{k+\mu}, \tag{3.4.1}$$

$$\pi_d|_{\sigma=0} = 0,$$

$$\pi_{\text{tot}}|_{\sigma=0} = \frac{1}{k} \cdot \frac{k}{k+\mu}. \tag{3.4.2}$$

These quantities have natural probabilistic interpretations: two lineages in the same subpopulation of density (that is, the limiting proportion of the total population, as $N \to \infty$) p, coalesce at rate $2/p$; each experience mutations at rate μ. They will be identical by descent if and only if there has been no mutation before their eventual coalescence, with probability $2p^{-1}/(2p^{-1} + 2\mu)$. This gives meaning to $\pi_s|_{\sigma=0}$ when taking $p = 1/k$. Individuals from differing subpopulations are necessarily non-identical at the neutral marker

since subpopulations evolve independently of one another. Thus $\pi_d|_{\sigma=0} = 0$. The total probability of identity, $\pi_{\text{tot}}|_{\sigma=0}$, is given by the product of the probability that the two lineages are from the same subpopulation, $1/k$, and the probability that they coalesce before either mutates, π_s.

We thus see that in the extreme (if artificial) limit of non-interacting mating type subpopulations, Nei's gene diversity at the population level ($h_{\text{tot}} = 1 - \pi_{\text{tot}}$) increases as a function of the number of mating types. Meanwhile gene diversity within the mating type subpopulations ($h_{\text{sub}} = 1 - \pi_s$) decreases as the number of mating types increases.

3.4.1.2. Monomorphic unisexual population. As addressed previously, the probability of identity in the unisexual population is the same as in an asexual population (see equation (3.3.10)):

$$\pi_{\text{tot}}|_{k=1} = \frac{1}{1+\mu}. \tag{3.4.3}$$

Thus Nei's gene diversity at the population level is independent of the rate of sexual reproduction in the population. We stress that we do not expect the genetic diversity to be independent on the rate of sexual reproduction in general. Rather, only the diversity at the neutral locus is unchanged, with the issue of genome-wide diversity obscured by Nei's measure, which focuses on a single locus.

3.4.1.3. Monomorphic self-incompatible population. In a facultatively sexual population with k mating types held in constant frequency, the probability of identity within the subpopulations and total population are given by equations (3.3.5) and (3.3.7) respectively. By comparing these quantities to the base case $\sigma = 0$, when k asexually reproducing populations held in constant frequency evolve independently of one another (see equations (3.3.8) and (3.3.9), we are led to a natural definition of an effective subpopulation and total population density, as we discuss below.

Assigning a 'mass' N^{-1} to each individual in the population at the Nth stage of rescaling, the *density* of a subpopulation i is given by N_i/N. The limiting subpopulation density is given by $p_i = \lim_{N\to\infty} N_i/N$. In this case of interest, where we artificially hold subpopulations in constant frequency, each subpopulation has limiting density k^{-1}. The total, or meta-, population density is 1. In the absence of all other interactions, two lineages in an asexually reproducing subpopulation of density p are identical by descent with probability $p^{-1}/(p^{-1}+\mu)$. It is reasonable then, that a population with probability of identity π has effective population density

$$p_{\text{eff}} := \frac{1}{\mu}\Big(\frac{1}{\pi} - 1\Big).$$

A similar notion of effective population size is defined by in Section 4.6 of Durrett [32] which he calls the *heterozygosity effective population size*.

Substituting π_s as given by equation (3.3.5), the effective subpopulation density is

$$p_{\text{sub,eff}} = \frac{1}{k}\Big[1 + (k-1)\frac{m}{\mu + m}\Big],$$

and substituting π_{tot} as given by equation (3.3.7) the effective total population density is

$$p_{\text{tot,eff}} = 1 + (k-1)\frac{1}{\mu + km}.$$

Elegantly, the quantities given by equations (3.3.5) and (3.3.7) correspond to a purely asexually reproducing population divided into independently evolving subpopulations held in constant frequency with density given by $p_{\text{sub,eff}}$.

For large N, we can translate back to a notion of effective population *size*, writing

$$N_{i,\text{eff}} \approx \frac{N}{k}\Big[1 + (k-1)\frac{\sigma d}{\mu(k - e^{-c}) + \sigma d}\Big],$$

and

$$N_{\text{eff}} \approx N\Big[1 + (k-1)\frac{1}{\mu + \sigma d\,\frac{k}{k-e^{-c}}}\Big].$$

We have used that $m = \sigma d/(k - e^{-c})$. The quantity $N_{i,\text{eff}}$ is approximately equal to N/k, what we would expect for k independent asexual populations, multiplied by a factor which is larger than 1, increasing in k, and increasing to $1 + \sigma d/\mu$: subpopulations are less susceptible to genetic drift, ie invading genotypes are less vulnerable to loss or fixation within subpopulations, when the probability of mutation is small relative to the probability of sex. Similarly, the total effective population size is approximately equal to N multiplied by a factor larger than 1; this factor increases approximately linearly in k, scaling like $k/(\mu + \sigma d)$. We observe that if $\sigma = \Theta(k)$, ie if the rate of sex is on the same order as the number of mating types, then there is no linear growth in the effective population size. Conversely, when the rate of sex is small, the total effective population size can be very large indeed.

In a final remark, we observe that taking $\sigma \to \infty$, that is, taking the rate of sex to be very large, equations (3.3.5) and (3.3.7) are all given by the probability of identity for two lineages in a homogeneously mixed asexually reproducing population given by $\pi_{\text{tot}}|_{\sigma=0,k=1}$. This is consistent with a phenomena called 'the collapse of structure' in structured coalescent models; the reader is directed to the informative summary in Chapter 6 of Etheridge [37] and the original paper by Nordborg and Krone [106].

In summary, the Nei's gene diversity at the population level ($h_{\text{tot}} = 1 - \pi_{\text{tot}}$) increases with the number of mating types, while genetic diversity within a subpopulation ($h_{\text{s}} = 1 - \pi_{\text{s}}$) decreases as a function of the number of mating types. However both these responses

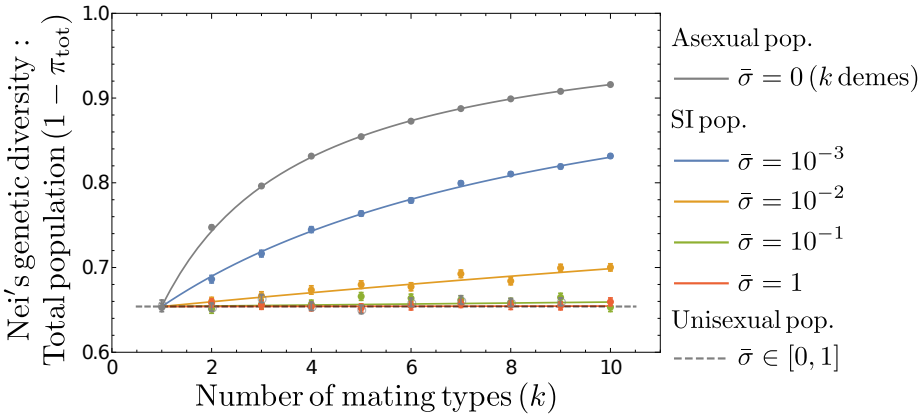


FIGURE 3.4. Total population genetic diversity (as measured by Nei's gene diversity) for asexual, self-incompatible and unisexual populations. Theoretical results (solid lines) well-predict the results of simulations (markers). Theoretical predictions for an asexual population with k non-interacting demes are obtained from equation (3.4.1), for a unisexual population from equation (3.4.3) and for a population with k self-incompatible mating types from equation (3.3.5). Model parameters: $N = 7560$, $\bar{\mu} = 2.5 \times 10^{-4}$, $c = 5$, $d = 0.5$. Simulations are averaged over 10^3 time points sampled regularly over 10^7 generations following an allowed 10^5 generation equilibrating period (initial conditions are all individuals homogeneous at the neutral marker). Note that in the setting of monomorphic asexual or unisexual populations, only the case $k = 1$ is useful for comparison, being the only biologically relevant setting.

are muted by an increasing number of attempted sexual reproductive events per generation, $\sigma = \bar{\sigma} N$.

3.4.1.4. Comparison of monomorphic populations. In Figures 3.4 and 3.5 we compare our analytic results for total population diversity and subpopulation diversity respectively to the results of Gillespie simulations[14]. We see excellent agreement across a range of parameters.

For unisexual populations we find that diversity at the neutral locus is unmodified from the asexual case and thus independent of rate of sexual reproduction.

In self-incompatible populations, we see that the total population level diversity increases monotonically away from the unisexual baseline as a function of the number of mating types (although at a decreasing rate as more types are added). While this effect increases as the rate of sexual reproduction decreases, this is entirely a result of the division of the population into mating type subpopulations. (This can be understood intuitively by con-

[14]A *Gillespie simulation* or *Gillespie algorithm* is a method for simulating a stochastic jump process popularised by Gillespie but initially proposed by Doob. If X is a Markov process taking values on a discrete space S and q_{ij} is the rate at which X jumps from i to j, then its trajectories are simulated as follows. The initial value X_0 is sampled according to some pre-specified distribution. If $X_0 = i$ then the time of the first jump is sampled by sampling from an exponential distribution with parameter $\sum_{j \neq i} q_{ij}$. The value of X after this jump is determined by sampling from the distribution proportional to q_{ij}, $j \neq i$. The algorithm then proceeds in the same fashion.

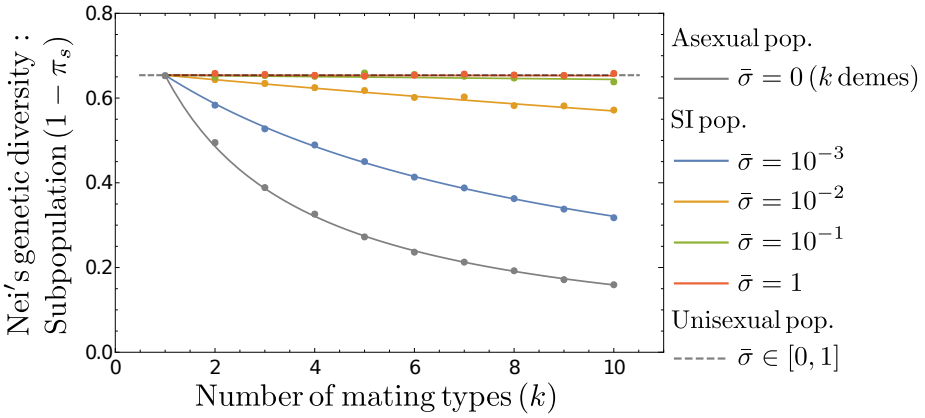


FIGURE 3.5. Subpopulation genetic diversity (as measured by Nei's gene diversity) for asexual, self-incompatible and unisexual populations. Theoretical results (solid lines) well-predict the results of simulations (markers). Theoretical predictions for an asexual population with k non-interacting demes are obtained from equation (3.4.2), for a unisexual population from equation (3.4.3) and for a population with k self-incompatible mating types from equation (3.3.7). Model parameters are the same as those in Figure 3.4. Note that in the setting of monomorphic asexual or unisexual populations, only the case $k = 1$ is useful for comparison, being the only biologically relevant setting.

sidering recombination as providing a mechanism for a "migration" event between mating type backgrounds acting as demes.) As a result the genetic diversity in the total population is in fact maximised when sex is entirely absent. However this is a rather artificial case; in this regime, there is no selection for mating types and so the mating-type population classes would be lost to drift. Similarly when sexual reproduction is rare and selection for equal mating type frequencies weak, mating types become increasingly susceptible to extinctions as their number increases. Meanwhile in the converse limit of obligate sex, the total population diversity relaxes to the values found in the unisexual and asexual cases. Again this can be intuitively understood by considering "obligate migration" between mating type "demes", which erodes the effect of population substructure. This corresponds to the limit $\sigma \to \infty$.

While genetic diversity within the total population increases as a function of the number of mating types, the genetic diversity within a mating type subpopulation decreases. This is a result of smaller subpopulation sizes (N/k), which simultaneously decreases the effective mutation rate and increases genetic drift within each class. The effect is amplified by low rates of sexual reproduction which increases the genetic isolation of the respective subpopulations.

We note that in both the cases of total population genetic diversity and subpopulation genetic diversity the key parameter with regards to facultative sex is the number of sexual events per generation, $\sigma = N\bar{\sigma}$, rather than simply the probability a sexual reproductive

event is attempted, $\bar{\sigma}$. Thus increasing the population size while keeping $\bar{\sigma}$ fixed has a similar qualitative effect to increasing $\bar{\sigma}$ with fixed N; the response of h_{tot} and h_s to increasing k is decreased.

3.4.2. Rate of effectual sex. In this section, we seek to understand how the benefits of sexual reproduction experienced may be affected by the presence or absence of mating types and their number. As the benefits of recombination, what ever they may be, can only be realised when sex occurs between genetically distinct individuals, we calculate the per capita probability that an individual successfully mates with an individual that is genetically distinct at the neutral locus. We term this the probability of effectual sex, which is given by the probability that a compatible mate is successfully located multiplied by the probability that this mate is not identical by descent at the neutral marker. For each $i = 1, \ldots, k$, this can be expressed

$$\phi_i := \sum_{j \neq i} \theta_{ij}(c)(1 - \pi_{ij}). \tag{3.4.4}$$

We note that this term is decreases with the probability of finding a mate (eg in self-incompatible populations) but increases with increasing genetic diversity between selected mates (caused, for example, by mating type population structure). It is the balance between these competing factors that we will explore in the following subsections. We first explore the behaviour of ϕ_i in the cases of monomorphic unisexual and self-incompatible populations, before proceeding to consider how ϕ_i changes in mixed populations of unisexuals and self-incompatible mating types.

3.4.2.1. Monomorphic unisexual population. Because there is no self-incompatibility in the unisexual population, the probability of finding a mate is one. Meanwhile the probability of identity by descent for two randomly selected individuals is $\pi_{\text{tot}}|_{k=1}$, with $\pi_{\text{tot}}|_{k=1}$ given in equation (3.4.3). Thus the probability of effectual sex in a monomorphic unisexual population is given simply by

$$\phi^u = \frac{\mu}{1 + \mu}. \tag{3.4.5}$$

From this we see that ϕ^u is independent of the frequency of sexual reproduction.

3.4.2.2. Monomorphic self-incompatible population. When the population is divided into self-incompatible mating type classes, the probability of finding a compatible mate is now generally less than one (see $\theta_{ij}(c)$ in equation (3.3.3)). Meanwhile the probability of identity by descent for two individuals selected from distinct and compatible mating type subpopulations $i \neq j$ is $\pi_{ij} = \pi_d$, with π_d given in equation (3.3.6). Thus the probability

of effectual sex in a monomorphic self-incompatible population is given by

$$\phi_i \equiv \phi := \theta_1(c)(1-\pi_d) = \frac{k-1}{k-e^{-c}}(1-\pi_d),$$

where we note that ϕ is the same for all mating types as we have assumed each is held at equal frequency $1/k$ (see Section 3.3). Rewriting this in terms of the original model parameters, we find

$$\phi = \frac{k-1}{k-e^{-c}}\left[\frac{\mu}{\mu+m} + \frac{m}{\mu+m}\frac{\mu}{k[1+(k-1)\frac{m}{\mu+m}]^{-1}+\mu}\right]. \tag{3.4.6}$$

Thus we see that the probability of effectual sexual reproduction increases with decreasing c (which increases the probability of locating a compatible mate) and with an increasing number of mating types (which increases the probability of mating with a genetically distinct individual).

3.4.2.3. Mixed self-incompatible and unisexual population. In this section we consider a mixed population consisting of $k-1$ self-incompatible mating types and a single unisexual type. As discussed in Section 3.3, when calculating the probability of identity by descent in the mixed population we are forced to consider simplified cases of the model in order to make analytic progress. This is because the unisexual type will always sweep to fixation in the absence of any reproductive benefits to self-incompatibility. We therefore consider three scenarios under which ϕ_i is calculated in turn: a unisexual type held at frequency α and $k-1$ self-incompatible mating types each at frequency $(1-\alpha)/(k-1)$ in a population at equilibrium; a rare unisexual type arising in a population of $k-1$ self-incompatible types previously at equilibrium; and $k-1$ rare self-incompatible types arising in a population of the unisexual type at equilibrium.

Recall that we denote the probability of identity by descent as π_d for two individuals sampled from distinct mating type classes, π_s for two individuals sampled from the same mating type class, π_d^u for one individual sampled from a mating type class and another sampled from the unisexual class, and π_s^u for two individuals sampled from the unisexual class. We then denote by ϕ and ϕ^u the probability of effectual sexual reproduction for self-incompatible mating types and unisexual types respectively. Using equation (3.4.4), these probabilities can be expressed as

$$\phi = (k-2)\theta_{ij}(c)(1-\pi_d) + \theta_{iu}(c)(1-\pi_d^u), \tag{3.4.7}$$

and

$$\phi^u = (k-1)\theta_{uj}(c)(1-\pi_d^u) + \theta_{uu}(c)(1-\pi_s^u), \tag{3.4.8}$$

where i, j are arbitrary indices representing self-incompatible classes (ie chosen such that $i \neq j$ and $i, j \neq u$) and the forms of $\theta_{ij}(c)$ and the probabilities of identity by descent depend on the scenario under consideration.

When the unisexual type held artificially at a constant frequency and the remaining self-incompatible mating types at equal frequencies, expressions for $\theta_{ij}(c)$ can be taken from equation (3.3.11) with $p_u = \alpha$ and $p_{i \neq u} = (1 - \alpha)/(k - 1)$. Substituting these into equations (3.4.7) and (3.4.8), we obtain

$$\phi = \frac{(k-2)(1-\alpha)(1-\pi_d) + (k-1)\alpha(1-\pi_d^u)}{k - 1 - e^{-c}(1-\alpha)}, \tag{3.4.9}$$

$$\phi^u = (1-\alpha)(1-\pi_d^u) + \alpha(1-\pi_s^u), \tag{3.4.10}$$

with π_d, π_s, π_d^u and π_s^u taken as the solution to the equations (3.6.15), (3.6.16), (3.6.17) and (3.6.18). The full expressions are too algebraically complex to reproduce. However we find the following general patterns. When the frequency of unisexuals is very low ($\alpha \to 0$), the rate of effectual sex is greater for unisexuals than self-incompatible mating types: $\phi^u > \phi$ for all k. Conversely when the unisexual frequency is very high ($\alpha \to 1$), the rate of effectual sex is greater for self-incompatible mating types than unisexuals: $\phi > \phi^u$ for all k. The precise value of α for which this transition occurs is dependent in a complex manner on the model parameters. We shall explore this more thoroughly in the following section.

Next we consider the case that a unisexual mutant emerges in a self-incompatible population with $k - 1$ mating types, previously at equilibrium. Expressions for $\theta_{ij}(c)$ can be taken from equation (3.3.11) with $p_u \approx 0$ and $p_i = 1/(k-1)$ for all $i \neq u$. Substituting these into equations (3.4.7) and (3.4.8), we obtain

$$\phi = \frac{(k-2)(1-\pi_d)}{k - 1 - e^{-c}}, \tag{3.4.11}$$

$$\phi^u = 1 - \pi_d^u, \tag{3.4.12}$$

with π_d, π_s, π_d^u and π_s^u taken from equations (3.3.12), (3.3.13), (3.3.14) and (3.3.15). We find that this qualitatively captures the prediction of equations (3.4.9) and (3.4.10) in the limit $\alpha \to 0$.

Finally we consider the scenario in which $k - 1$ mutant self-incompatible mating types simultaneously emerge in a resident unisexual population, previously at equilibrium. Expressions for $\theta_{ij}(c)$ can be taken from equation (3.3.11) with $p_u \approx 1$ and $p_i = 0$ for all $i \neq u$. Substituting these into equations (3.4.7) and (3.4.8), along with expressions for π_d, π_s, π_d^u and π_s^u taken from equations (3.3.16) and (3.3.17), we obtain

$$\phi = \phi^u = \frac{\mu}{1 + \mu}. \tag{3.4.13}$$

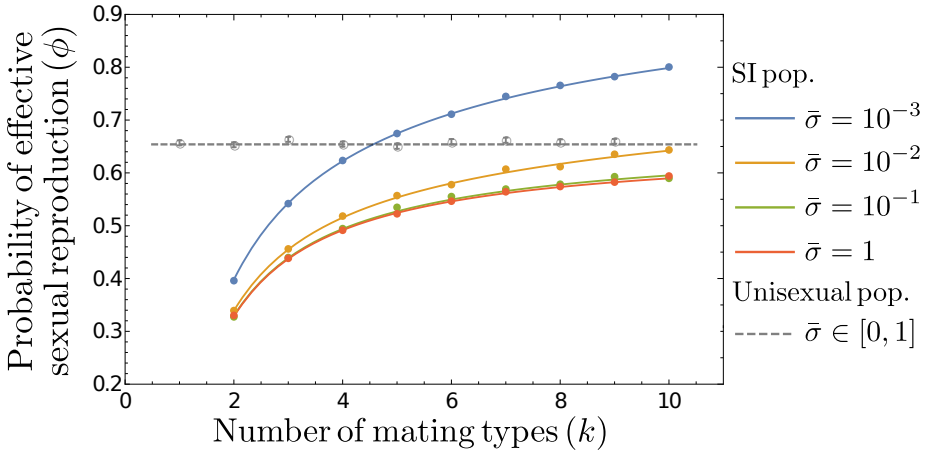


FIGURE 3.6. Probability of a effective sexual reproduction, ϕ, for monomorphic self-incompatible and unisexual populations. Theoretical results (solid lines) well-predict the results of simulations (markers). Theoretical predictions for a unisexual population are taken from equation (3.4.5) and for population with k self-incompatible mating types from equation (3.4.6). Model and simulation parameters are the same as those in Figure 3.4. Note that in the setting of a monomorphic unisexual population, only the case $k = 1$ is useful for comparison, being the only biologically relevant setting.

This indicates that upon emergence, the mutant self-incompatible type has exactly the same probability of effectual sex as its unisexual ancestor.

We note that while this qualitatively disagrees with the prediction obtained in equations (3.4.9) and (3.4.10), (for which $\phi > \phi_u$ at $\alpha = 1$), this is not unexpected. The approximation in equations (3.4.9) and (3.4.10) is obtained assuming that the population is at equilibrium; thus in this limit there is non-zero probability that the self-incompatible types (each at infinitely low frequency) carry a distinct neutral genetic marker to those seen in the unisexual population. The approximation in equations (3.4.13) on the other hand is obtained assuming that the self-incompatible types inherit the neutral marker of their unisexual ancestor; thus in this limit there is zero probability that the self-incompatible types (each at infinitely low frequency) carry a distinct neutral genetic marker to those seen in the unisexual population.

3.4.2.4. Comparison of cases. In Figure 3.6 we compare our analytic results for the rate of effectual sex, ϕ, to the results of Gillespie simulations. Once again, we see excellent agreement across a range of parameters. We see that for self-incompatible populations, the probability of effectual sex increases with an increasing number of mating types, k, and a decreasing probability of instigating a mating round, $\bar{\sigma}$, both of which increase the effect of population structure. When the rate of sexual reproduction is extremely low and the number of mating types moderately high ($\bar{\sigma} = 10^{-3}$, $k = 5$), we see that the probability of effectual sex in the self-incompatible population can in fact exceed that of unisexual

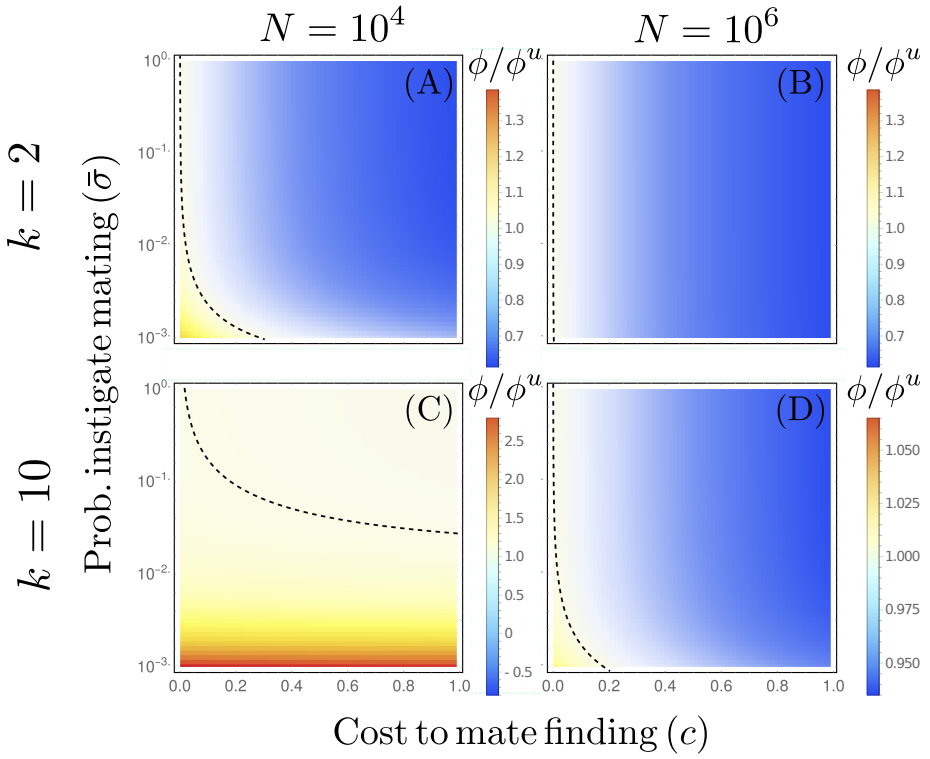


FIGURE 3.7. Theoretical predictions for the ratio of effectual sexual reproduction for self-incompatible mating types, ϕ, over unisexual types ϕ^u, in monomorphic populations (see equations (3.4.5) and (3.4.6). In all panels $\bar{\mu} = 10^{-9}$ and $d = 1/2$. The ratio ϕ/ϕ^u decreases with increasing population size (compare (A) to (B) and (C) to (D)), but increases with a greater number of mating types (compare (A) to (C) and (B) to (D)).

populations. Thus while the unisexual population engages in more sexual reproduction overall, many of these opportunities to realise the benefits of sex are essentially wasted by same-clone mating.

In Figure 3.6, N and c are held fixed. In fact $c = 5$ in these plots, which is approximately equivalent to mass-action encounter rates (a large cost to mate finding). In Figure 3.7 we explore how the relative advantage to the self-incompatible populations over the unisexual population in terms of the rate of effectual sex ϕ varies with these parameters. We note that since the key parameter in equations (3.4.5) and (3.4.6) is in fact the *number* of instigated mating rounds per generation, $\sigma = \bar{\sigma} N$, for the self-incompatible population to have a higher ϕ than the unisexual population for $k = 2$ requires either extremely low $\bar{\sigma}$ or relatively low N (see Figure 3.7(A) and Figure 3.7(B)). However an increased number of mating types can appreciably alleviate this requirement (see Figure 3.7(C) and Figure 3.7(D)).

In the above paragraphs, we have started to identify the potential advantages garnered by self-incompatibility over unisexuality at the population level. However we have done so by comparing populations consisting of purely self-incompatible or unisexual types. We next

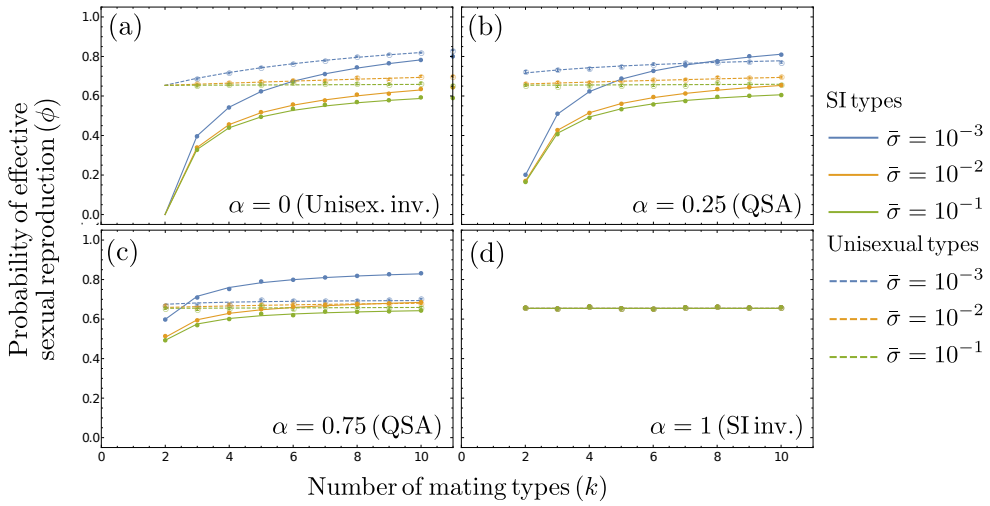


FIGURE 3.8. Probabilities of a effective sexual reproduction, ϕ and ϕ^u, for mixed populations with a unisexual type at frequency α and $k-1$ self-incompatible mating types at frequencies $(1-\alpha)/(k-1)$. Theoretical results (solid lines) well-predict the results of simulations (markers). Panel (A): Scenario of a rare unisexual type invading resident population of mating types, with theoretical results taken from equations (3.4.11) and (3.4.12). Panels (B) and (C): Scenario of each type held at a constant frequency so that the population has time to equilibrate to a quasi-stationary state, with theoretical results taken from equations (3.4.9) and (3.4.10). Panel (D): Scenario of a set of rare self-incompatible types invading a resident unisexual population, with theoretical results taken from equation (3.4.13). Model and simulation parameters are the same as those in Figure 3.4.

discuss how these insights are altered in the mixed population.

As previously described, we began by calculating the probability of effectual sex (see equations (3.4.5) and (3.4.6)) for unisexual and self-incompatible types (ϕ^u and ϕ respectively), under the condition that the frequencies of each of the types was fixed. This is equivalent to making a *quasi-static approximation* which assumes that the population reaches an equilibrium in the distribution of allele frequencies at the neutral marker on a much faster timescale than that on which the frequency of unisexuals changes. In biological terms this can be understood as assuming very weak selection for unisexuals over self-incompatible mating types, or vice versa.

The results of this quasi-static approximation are illustrated in Figure 3.8(B) and Figure 3.8(C). The probability of effectual sexual reproduction for self-incompatible types (solid lines) behaves in a qualitatively similar manner to that in the case of a monomorphic self-incompatible population (see Figure 3.6): ϕ increases with an increasing number of mating types k and a decreasing probability of instigating a mating round $\bar{\sigma}$. However we see qualitatively different behaviour of ϕ^u when comparing the monomorphic and mixed population cases. When rare, they are now able to exploit the diversity generated by the self-incompatible types (see Figure 3.8(B), dashed lines) to increase ϕ^u beyond what is

observed in a monomorphic unisexual population (see Figure 3.6). As, in addition, they are compatible with the entire population, their rate of effectual sex can exceed that of the self-incompatible types. However, as the unisexual frequency increases (see Figure 3.8(C)) their probability of same-clone mating increases, while simultaneously the costs paid by self-incompatible types in terms of restricted mating opportunities decreases. Together these factors can combine to give a higher rate of effectual sex for the self-incompatible population than the unisexual population when mating types are numerous, and sex and the self-incompatible types themselves are rare.

While the quasi-static approximation described above assumes that the population is at equilibrium, we also developed out of equilibrium expressions for the probability of effectual sex, ϕ^u and ϕ, under the scenarios of a unisexual type was invading a resident population of self-incompatible mating types (see equations (3.4.11) and (3.4.12)) and of self-incompatible mating types invading a unisexual population (see equations (3.4.13)). These results are illustrated in Figure 3.8, in panels (A) and (B) respectively. We see that a unisexual invader is able to exploit the genetic diversity generated by resident self-incompatible types (see panel (A)), in qualitative agreement with the quasi-static approximation. However invading self-incompatible mating types initially receive no advantage in terms of an increased ϕ (see panel (D)), in disagreement with the quasi-static approximation. This implies that while self-incompatible types may experience a higher rate of effectual sex than the unisexual population when the self-incompatible types themselves are rare, during their invasion this advantage is contingent on them surviving long enough to genetically differentiate from their unisexual ancestors.

In summary we can paint the following picture. When self-incompatible mating types first emerge in a unisexual population, they will have an equal rate of effectual sex to the resident population (see Figure 3.8(D)). As they are rare they pay no costs in terms of a reduced number of mating opportunities, but simultaneously receive no benefits from self-incompatibility, having not had time to differentiate themselves from their unisexual ancestors. Any initial growth in their frequencies will be driven primarily by drift. However, if these mating types can survive at low frequencies for a sufficient amount of time, they will differentiate genetically from their unisexual ancestors. In this case, when sex is rare ($\bar{\sigma} \ll 1$), self-incompatible types will mate more frequently with genetically distinct individuals than unisexuals (see Figure 3.8(C)). If the benefits of sex lie in recombination, then the mating types may experience a selective advantage over their unisexual ancestors. Although this may temporally drive an increase in the frequency of self-incompatible types, this advantage is short-lived. As the frequency of self-incompatible types increases, their opportunities to find a compatible mate decrease. Simultaneously the unisexual type increasingly exploits the genetic diversity generated by the self-incompatible types. Together these factors can drive the probability of effectual sex in unisexuals above that of

the invading mating types (see Figure 3.8(B)) an effect that becomes ever stronger as the self-incompatible mating types approach fixation.

Conversely, when unisexual types first emerge in a population of self-incompatible mating types, they are able to exploit the genetic diversity of the resident population to obtain all of the benefits of self-incompatibility (decreased probability of mating with an individual of the same allelic type at the neutral marker) while not paying the costs of missed mating opportunities (see Figure 3.8(A)). They will thus receive a selective advantage and increase in frequency. However as they increase, their probability of effectual sex drops below that of their self-incompatible ancestors, an effect that becomes stronger as the unisexual types approach fixation (see Figure 3.8(B) and Figure 3.8(C)). If the benefits of recombination are sufficiently strong, this will generate selection against the unisexual type. Taken together, these results suggest that we may in fact see negative frequency dependent selection for self-incompatibility in haploids.

3.5. Discussion.

The existence of sexual self-incompatibility between gametes has posed an evolutionary puzzle for many years, and a variety of theories to explain the evolution of this discriminatory behaviour have been proposed Billiard et al. [9]. Among these the hypothesis that mating types, by preventing haploid selfing, enhance the advantages of recombination has remained poorly investigated theoretically [10]. With the precise benefits of recombination still a widely debated topic, the reason for this deficit is clear; a full investigation of the theory in fact requires distinct studies testing it within the context of each separate hypothesis for the evolution of recombination. In this paper we have circumvented this issue by first studying a more general problem; as stated succinctly in Billiard et al. [10]:

> If recombination has an advantage, whatever it is, individuals gain this advantage only when the haploid genomes that are recombining are not strictly identical.

We have shown that although self-incompatibility decreases the probability of successful sexual reproduction, it can in fact increase the probability of effectual sexual reproduction between non-clone mates, even exceeding the rates of effectual sexual reproduction in unisexuals. However the strength of this effect is highly dependent on population parameters including its effective size, the rate of sexual reproduction and the population composition.

We have demonstrated that monomorphic unisexual populations are less diverse than populations composed of self-incompatible mating types and further, that the probability of effectual sexual reproduction with an individual genetically distinct at a neutral marker can be lower in a unisexual population than in self-incompatible population. Both of these

population-level benefits to self-incompatibility vary in a qualitatively similar way as the remaining parameters in the model are varied. Increasing the number of mating types and decreasing the number of sexual events per generation ($\sigma = N\bar{\sigma}$) increases the relative benefits to self-incompatibility over unisexuality. Meanwhile decreasing the cost to mate finding (ie allowing more opportunities for the self-incompatible types to find a compatible mate) has a particularly prominent effect on the probability of effective sex in the self-incompatible population, allowing it to more readily exceed that in a unisexual population.

The above discussion is limited to considering the population level benefits of self-incompatible over unisexuality. However for an evolutionary perspective, we must instead understand the costs and benefits of self-incompatible at an individual level. Here our simplified "recombination hypothesis free" approach (where the benefits of recombination have not been explicitly defined) runs into some trouble; as we have shown, without any emergent benefits to mating with genetically distinct individuals, the unisexual type is positively selected for. Although the strength of selection against self-incompatibility is decreased with lowered costs to mate finding, the unisexual type will always sweep to fixation in finite time.

To make progress we investigated how the rate of effectual sex (an individual benefit) varied between unisexual and self-incompatible types in a population held artificially at a polymorphic equilibrium. This approximation to the dynamics of the full model (in which the frequencies of unisexuals and self-incompatible mating types are allowed to vary) is likely to hold when selection for either of the types is relatively weak. We found that in this mixed population, the unisexual individuals can exploit the diversity generated by mating types, obtaining a higher rate of effectual sex than in their monomorphic populations. This scenario is reminiscent of the tragedy of the commons, a comparison previously utilised to explain selection for low-recombination rate modifiers in general models for the evolution of recombination [81]. While this leads to unisexuals having a higher probability of effectual sex than their self-incompatible mating type competitors when rare, this advantage is frequency dependent; as the frequency of unisexuals increases, their rate of effectual sex dips back below that of their self-incompatible competitors. While the precise frequency at which this transition occurs depends in a complex way on the model parameters, our intuition from the case of monomorphic populations carries over: the advantage of self-incompatibility over unisexuality diminishes as the number of mating types decreases, and the number of sexual events per generation ($\sigma = N\bar{\sigma}$) and costs to mating finding increase.

We next turn our attention to the key biological motivation for this paper. In the context of the sex-advantage enhancer hypothesis: (i) how did self-incompatibility first evolve and (ii) how does self-incompatibility remain robust again invasion?

Regarding the first question, we note that initially (when rare) a mutant self-incompatible type would only be weakly selected against. However, in order to drift to more substantial

frequencies, the costs of mate finding would have to be sufficiently low. This requires the possibility of many mating rounds and thus suggests that self-incompatibility first evolved in an aquatic environment, in which individuals have more opportunities to actively find a compatible mate. Such an environment is the natural case for algae. While fungi are more associated with terrestrial environments, empirical evidence suggests that early fungal ancestors lived in water [100]. With the MAT locus that determines mating types in fungi appearing ancient and displaying a remarkable conservation of core homeodomain proteins over trans-specific evolutionary periods [46], it is conceivable that self-incompatibility in fungi arose in just such an environment. However in order to see selection for self-incompatibility, it is not sufficient to simply have low costs; self-incompatibility must also have benefits.

As we have addressed, if the self-incompatible mating types can be maintained in the population for a sufficient amount of time at low-frequencies, they can obtain a higher effective rate of sex to their unisexual ancestors. If genetic recombination with non-clone mates is sufficiently advantageous, this could generate selection for mating types. However the magnitude of this effect is strongly dependent on the remaining population parameters. Crucially, this benefit to self-incompatibility requires that the number of sexual events per generation, $\sigma = N\bar{\sigma}$, is low. While it seems natural to assume that when self-incompatibility first evolved, the rate at which sexual reproduction and mate finding was triggered was very low ($\bar{\sigma} \ll 1$), low effective population sizes (eg $N \ll 10^6$, see Figure 3.7) are harder to justify biologically, especially if we consider these dynamics occurring in a well-mixed aqueous environment. Such low population size values are more consistent with a spatially structured population, within which the costs to mate finding are expected to be larger. We proceed with this caveat in mind.

Assuming that the self-incompatible mating types are selected for at this stage of their invasion trajectory, we emphasise that this advantage over unisexuals is necessarily fleeting. Once the self-incompatible types reach a critical frequency, unisexual types will be able to parasitise their diversity and regain the upper hand. Thus in terms of self-incompatibility having a higher probability of displacing unisexuality than *vice versa*, the very best that can be hoped for is that this transition occurs at a high frequency of self-incompatible types, far into their invasion trajectory. We note that once a resident population of two mating types has established, an evolutionary pressure for the number of mating types to increase arises. New mating types experience a selective advantage both in terms of increased mating opportunities relative to the two resident populations and a (shared) decrease in the probability of mating with a genetically distinct individual. There is then selection for an ever greater number of mating types, which are only kept bounded by genetic drift [23, 28].

We now consider the second question, of how self-incompatibility can remain robust again invasion from unisexual, or homothalic, mutants. Our model suggests that in fact

self-incompatibility may never be truly robust against invasions from unisexuals, as unisexuals can always free-ride on the genetic differentiation of resident self-incompatible mating types. However fixation of unisexuals might be prevented if their probability of effectual sexual reproduction drops below that of self-incompatible residents at sufficiently low frequency. We have shown that in this context, the presence of multiple resident mating types can markedly enhance this possibility, such that self-incompatible types can maintain a higher rate of effectual sex even as the number of instigated sexual events per generation becomes moderately high (see Figure 3.7). It is therefore interesting to note that among the *Basidiomycota*, which are characterised not only by their comparatively high rates of sexual reproduction [85], but also by their large number of mating types [22], homothalism is relatively rare [68]. Homothalism is also far from ubiquitous in isogamous algae, with only a handful of species known to employ this indiscriminate mating strategy [113]. Here our model suggests qualitatively that self-incompatibility may be protected by the rarity of sex in these species [59]. However in a quantitative sense, our model predicts that this rarity should be more offset by the large effective population sizes observed in isogamous algae [103].

Taken together, we see from the above discussion that the sex advantage enhancer hypothesis may agree qualitatively with observed scenarios under which self-incompatibility first evolved (ie an aquatic environment with low costs to mate searching and very rare sexual reproduction), is maintained (ie in population with a low number of sexual reproductive events per generation and a large number of mating types) or is displaced by homothalism (ie populations with more frequent sexual reproduction, a low number of resident mating types and high costs to mate finding). Although in a quantitative sense we see that our model predicts that self-incompatibility may be selected for over a rather restricted range of biologically reasonable parameters, we have made a number of simplifications in order to make the problem tractable, but that may provide an overly stringent estimation of when self-incompatibility may be selected for. Below we detail some of these key simplifications, as well as discussing their potential effects.

First, for simplicity, we have considered only the dynamics of a single neutral genetic marker, and defined sex as 'ineffectual' when when two mating individuals carry the same marker. In reality, particularly in populations undergoing recombination, it is of course possible that while individuals are identical at the neutral marker, they are genetically distinct throughout the rest of their genome. Thus our estimate very much provides an underestimate for the rate of potentially beneficial sex between non-clone mates. Despite this we believe that the qualitative intuition that we have developed will remain valid; frequent sexual reproduction erodes population structure, and is thus likely to diminish the advantage to self-incompatibility in well-mixed populations.

Second, while we have assumed a well-mixed population, it is clear that spatial struc-

ture, by increasing the probability of mating with genetically related individuals, is likely a strong driver of the evolution self-incompatibility. While within our modelling framework we view this effect as manifesting for small population size. While it would be interesting to consider how explicitly accounting for spatial structure would affect the results we have described here, we note any additional insights would likely come at the cost of generality, with any model necessarily having to account for species-specific details of asexual growth environment, mate searching and non-local dispersal. Such models also prohibit tractable analytic results.

Third, we note that our modelling approach is based on the assumption that mating with a genetically distinct individual increases the advantages of sexual reproduction. However, the benefit of sexual reproduction is not restricted to such recombination advantages. In many species, stress-induced sexual reproduction has evolved to be coincident with an environmentally resistant dormant form. In cases such as these, the primary advantages to recombination (survival) are achieved regardless of the genetic identity of the partner. For instance, in the yeasts *Saccharomyces cerevisiae* and *S. pombe*, sexual reproduction is followed by the formation of environmentally resistant spores [47]. Thus while these species fall well within the range where we would expect self-incompatibility to be favourable (with rare rates of sexual reproduction [127] and spatially structured populations with small local effective population sizes [105]) mating type switching has still evolved in these species [104]. This homothalism, while metabolically costly during asexual reproduction, provides reproductive assurance and survival over the longer term can can thus still be selected for [105].

Finally, and perhaps most importantly, our model does not include any explicit benefit to genetic recombination. This was very much a conscious decision, as it allows us to avoid 'baking in' an advantage to non-self mating, in contrast to earlier studies [26, 101]. While our approach falls short of what must be the ultimate goal of investigating how self-incompatibility can be selected for in models in which the benefits of recombination are emergent (ie arise under a recombination hypothesis framework), our results nevertheless provide intuitive insight as to how these models might behave.

For instance, consider a prototypical model of the Red Queen Hypothesis in a haploid population, with two biallelic loci and with selection for recombination generated by gene-for-gene interactions. If sex is regular, we might expect the frequencies of the genes in each mating type subpopulation to be roughly equal. In this scenario we might expect there to be no advantage to self-incompatibility, only a cost in terms of mating opportunities. However, if sex were rare, drift would drive deviations in genotype frequencies between mating type subpopulations, increasing the probability of mating with a genetically distinct individual under self-incompatibility. Our hope is that this problem will be tackled in the near future, and that the results of this paper will help guide future research that specifically look at the

evolution of self-incompatible within each evolution of recombination framework.

3.6. Appendix.

3.6.1. Diffusion processes. A k-dimensional diffusion process denoted

$$\{X(t) = (X_1(t), \ldots, X_k(t))\}_{t \geq 0}$$

is a continuous time Markov process with continuous sample paths. It is described by two groups of quantities, called the drift and diffusion coefficients, which describe the mean and covariance in the change in each coordinate $X_i(t)$ over an infinitesimal time period: if for each $i, j = 1, \ldots, k$ and $h > 0$ one sets $\Delta_h X_i(t) := X_i(t+h) - X_i(t)$,

$$b_i(x) := \lim_{h \searrow 0} \frac{1}{h} \mathbf{E}[\Delta_h X_i(0) | X(0) = x],$$
$$a_{ij}(x) := \lim_{h \searrow 0} \frac{1}{h} \mathbf{E}[\Delta_h X_i(0) \cdot \Delta_h X_j(0) | X(0) = x],$$

then $\{X(t)\}_{t \geq 0}$ is described as a solution to the system of stochastic differential equations

$$X_i(t) = X_i(0) + \int_0^t b_i(X(s))\, \mathrm{d}s + \sum_{j=1}^{k} \int_0^t \sigma_{ij}(X(s))\, \mathrm{d}B_j(s), \tag{3.6.1}$$

often denoted

$$\mathrm{d}X_i(t) = b_i(X(t))\, \mathrm{d}t + \sum_{j=1}^{k} \sigma_{ij}(X(t))\, \mathrm{d}B_j(t),$$

where $B_1, \ldots, B_k$ are independent Brownian motions and the coefficients σ_{ij}, $i, j = 1, \ldots, k$, are defined by the relation $a(x) = \sigma(x)\sigma(x)^T$. Commonly, the coefficients $b(x) = (b_i(x))_{i=1,\ldots,k}$ and $a(x) = (a_{ij}(x))_{i,j=1,\ldots,k}$ are referred to as the drift and diffusion coefficients respectively. Approximately speaking, we may interpret this as saying that

$$X_i(t+h) \approx X_i(t) + b_i(X(t))h + \sum_{j=1}^{k} \sigma_{ij}(X(t))\sqrt{h}\xi_j(t),$$

where for each $j = 1, \ldots, k$ and $t > 0$, $\xi_j(t)$ is a Gaussian random variable with mean zero and unit variance such that $\xi_i(t)$ and $\xi_j(s)$ are independent if either $i \neq j$ or $t \neq s$. The diffusion may equally be characterised by its infinitesimal generator L defined by

$$Lf(x) := \lim_{h \searrow 0} \frac{1}{h} \mathbf{E}[\Delta_h f(X(t)) | X(t) = x]$$

$$= \sum_{i=1}^{k} b_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x).$$

The reader is referred to Karlin and Taylor [73] and Durrett [31] for a further introduction to diffusion processes.

Diffusion processes often arise as the limit of a sequence of Markov processes with increasingly frequent but increasingly small jumps; the canonical example is Brownian motion, which arises as the limit of a suitably rescaled simple random walk. Suppose that we have a sequence of continuous time Markov processes $\{X^N\}_{N=1}^{\infty}$, where, for each N, the process X^N takes values in $S_N \subseteq \mathbb{R}^k$. If we define for $x \in S_N, i, j = 1, \ldots, k$,

$$b_i^N(x) := \lim_{t \searrow 0} \frac{1}{t} \mathbf{E}[\Delta_t X_i^N(0)) | X^N(0) = x],$$
$$a_{ij}^N(x) := \lim_{t \searrow 0} \frac{1}{t} \mathbf{E}[\Delta_t X_i^N(0) \Delta_t X_j^N(0) | X^N(0) = x],$$

and suppose that

$$\lim_{N \to \infty} \sup_{x \in \mathbf{R}^k} |b_i^N(x) - b_i(x)| = 0,$$

$$\lim_{N \to \infty} \sup_{x \in \mathbf{R}^k} |a_{ij}^N(x) - b_{ij}(x)| = 0,$$

and

$$\lim_{N \to \infty} \sup_{x \in \mathbf{R}^k} \lim_{t \searrow 0} \frac{1}{t} \mathbf{E}[|X^N(t) - X^N(0)|^4 | X^N(0) = x] = 0,$$

then the sequence of processes X^N approximates the solution X to the system of equations (3.6.1). This is a version of the seminal result of Stroock and Varadhan, see eg Durrett [31]. A full statement of the result is given in the next section. The fourth moment condition can be relaxed to a condition which controls the size of the jumps of the Markov chain, although, as in our case, it is often easy to check.

3.6.2. The Stroock–Varadhan Theorem. In this appendix, we state a version of the Stroock–Varadhan Theorem for continuous-time Markov chains; it follows after combining the results of Theorem 7.1 and Lemma 8.1 in Durrett [31], Chapter 8. The Stroock–Varadhan Theorem provides the conditions under which a sequence of Markov chains, in either discrete or continuous time, converge to a solution to a martingale problem; a further result states when such a solution may be expressed as a solution to a system stochastic differential equations. An analogous version of the result we state exists for discrete time processes; the reader is directed to Chapter 8 of Durrett [31] for further details.

Let $\{X^N\}_{N=1}^{\infty}$ be a sequence of continuous-time Markov chains, where for each N the process X^N takes values in $S_N \subseteq \mathbb{R}^k$. Define for each $N = 1, 2, \ldots, x \in S_N, A \subseteq$

$\mathbb{R}^k, x \notin A$ a sequence of measures

$$Q^N(x, A) := \lim_{t \searrow 0} \frac{1}{t} \mathbf{P}[X^N(t) \in A \mid X^N(0) = x]. \tag{3.6.2}$$

The measure $Q^N(x, A)$ records the instantaneous rate at which the process X^N jumps from a state x into the subset A. Let us further define the sequence of moments

$$a_{ij}^N(x) := \int_{\mathbb{R}^k} (y_i - x_i)(y_j - x_j) Q^N(x, \mathrm{d}y), \tag{3.6.3}$$

$$b_i^N(x) := \int_{\mathbb{R}^k} (y_i - x_i) Q^N(x, \mathrm{d}y), \tag{3.6.4}$$

$$\gamma^N(x) := \int_{\mathbb{R}^k} |y - x|^4 Q^N(x, \mathrm{d}y). \tag{3.6.5}$$

Let us suppose that

(A) for any closed bounded set $K \subseteq \mathbb{R}^k$,

$$\sup_{x \in K} Q^N(x, \mathbb{R}^k \backslash \{x\}) < \infty,$$

(B) b_i and a_{ij} are continuous coefficients for which, for every initial condition $x \in \mathbb{R}^k$, there exists a unique process X such that $X(0) = x$ and

$$X_i(t) - \int_0^t b_i(X(s))\, \mathrm{d}s, \tag{3.6.6}$$

$$X_i(t) X_j(t) - \int_0^t a_{ij}(X(s))\, \mathrm{d}s \tag{3.6.7}$$

are both local martingales,

and for each $i, j = 1, \ldots, k$, it holds that

(C) $\lim_{N \to \infty} \sup_{x \in \mathbb{R}^k} |a_{ij}^N(x) - a_{ij}(x)| = 0$,

(D) $\lim_{N \to \infty} \sup_{x \in \mathbb{R}^k} |b_i^N(x) - b_i(x)| = 0$, and

(E) $\lim_{N \to \infty} \sup_{x \in \mathbb{R}^k} \gamma^N(x) = 0$.

If, in addition to conditions (A-E), it holds that $\mathbf{P}[X^N(0) = x^N] = 1$ and $x^N \to x$ in $\mathbb{R}^k$ as $N \to \infty$ then, in the same limit, $X^N \Rightarrow X$, where X the unique process such that $X(0) = x$, and (3.6.6) and (3.6.7) are local martingales. Here, '$\Rightarrow$' denotes weak convergence in the space of right-continuous paths $\omega : [0, T] \to \mathbb{R}^k$ with left limits (so-called càdlàg paths) equipped with the Skorokhod topology. The space is frequently denoted $D([0, T], \mathbb{R}^k)$ and

the reader is directed to Section 8.6 of Durrett [31], Chapter 3, for further details of this notion of convergence.

Let us now address the circumstances under which there exists a process X satisfying condition (B). Defining σ such that $a = \sigma\sigma^T$, Theorem 4.5 of Durrett [31], Chapter 5, asserts that there is a one-to-one correspondence between processes X with $X(0) = x$ satisfying the condition that (3.6.6) and (3.6.7) are local martingales, and weak solutions to the system of stochastic differential equations defined in equation (3.6.1). Uniqueness in distribution of solutions to equation (3.6.1), then, guarantees uniqueness of such a process X. Verifying (B) is thus equivalent to verifying that

(B') b_i and a_{ij} are continuous coefficients for which, for every initial condition $x \in \mathbb{R}^k$, solutions to the system of equations (3.6.1) with $a = \sigma\sigma^T$ and $X(0) = x$ are unique in distribution.

One approach to proving such uniqueness is the method of duality, outlined in the following appendix.

3.6.3. The method of duality. In this appendix we outline a working version of the method of duality. This technique, although seemingly contrived, has found widespread application proving distributional uniqueness of solutions to stochastic differential equations and, more generally, martingale problems. Here, we state a special case of a more general version appearing in Section 4.4 of [40].

Distributional uniqueness of solutions to this system of equations is implied by uniqueness of one-dimensional distributions (see, for example, Theorem 4.2 of Ethier and Kurtz [40], Chapter 4); that is, if X and $\tilde{X}$ are two solutions to the system of equations (3.6.1), it is sufficient to show that, for each $t \geq 0$,

$$\mathbf{E}[F(X(t))] = \mathbf{E}[F(\tilde{X}(t))] \tag{3.6.8}$$

for a sufficiently large class of functions F. For example, the set of all smooth and compactly supported $F : \mathbb{R}^k \to \mathbb{R}$ would be one such class (although much smaller classes, as we will make use of later, do exist).

Fix $x \in \mathbb{R}^k$. We outline a method for proving distributional uniqueness of solutions to the system of equations (3.6.1) with initial condition $X(0) = x$. Let $f : \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}$ and $\beta : \mathbb{R}^l \to \mathbb{R}$ be functions. Take $y \in \mathbb{R}^l$ and suppose that

(F) there exists a $\mathbb{R}^l$-valued Markov process Y such that $Y(0) = y$ and, for every solution X to the system of equations (3.6.1) with initial condition $X(0) = x$, X is

independent of Y and

$$\mathbf{E}[f(X(t), y)] = \mathbf{E}\big[f(x, Y(t)) \exp\big\{\int_0^t \beta(Y(s))\,\mathrm{d}s\big\}\big].$$

If, furthermore, for each $t \geq 0$,

(G) $\int_0^t |\beta(Y(s))|\,\mathrm{d}s < \infty$ almost surely,

(H) $\mathbf{E}[|f(X(t), y)|] < \infty$ almost surely, and

(I) $\mathbf{E}[|f(x, Y(t)) \exp\{\int_0^t |\beta(Y(s))|\,\mathrm{d}s|\}] < \infty$ almost surely,

then we say that solutions to (3.6.1) with initial condition $X(0) = x$ are dual to the process Y with respect to the functions f and β. Thus the problem of proving uniqueness of solutions to (3.6.1) with initial condition $X(0) = x$ is transformed into one of establishing the existence of Y for a large enough choice of initial conditions y: if such a Y exists for all initial conditions $y \in S \subseteq \mathbb{R}^l$, where S is a set large enough so that the collection of functions $\{F = f(\cdot, y) : y \in S\}$ is large enough for determining the one-dimensional distributions of solutions to equations (3.6.1) in the sense of equation (3.6.8), then solutions to equations (3.6.1) are unique in distribution.

3.6.4. Uniqueness of solutions to the system of SDEs. In this appendix, we establish a duality relation between solutions X to the system of equations (3.3.1) and the genealogical process Y. Applying the result of the previous appendix, we determine the distributional uniqueness of solutions to this system.

Let X be a solution to the system of equations (3.3.1) with $X(0) = x$ and Y the genealogical process with $Y(0) = y$. The process X has generator

$$L_X g(x) = -\sum_{i=1}^{k} \mu x_i \frac{\partial g}{\partial x_i}(x) + \sum_{i=1}^{k} \sum_{j \neq i} m_{ij}(p, c)(x_j - x_i) \frac{\partial g}{\partial x_i}(x) + \sum_{i=1}^{k} \frac{1}{p_i} x_i (1 - x_i) \frac{\partial^2 g}{\partial x_i^2}(x),$$

and Y has generator

$$L_Y g(y) = \sum_{i=1}^{k} \sum_{j \neq i} m_{ij}(p, c) y_i (f(\ldots, y_i - 1, \ldots, y_j + 1, \ldots) - f(y)) + \sum_{i=1}^{k} \frac{2}{\theta_i} \binom{y_i}{2} (f(\ldots, y_i - 1, \ldots) - f(y)).$$

Proceeding informally, for a function $f : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ and $\beta : \mathbb{R}^k \to \mathbb{R}$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{E}\big[f(X(s),Y(t-s))\exp\Big\{\int_0^{t-s}\beta(Y(u))\,\mathrm{d}u\Big\}\big] \\ = \mathbf{E}\big[\big(L_X f(\cdot,Y(t-s))(X(s)) - L_Y f(X(s),\cdot)(Y(t-s)) \\ - \beta(Y(t-s))f(X(s),Y(t-s))\big) \\ \times\exp\Big\{\int_0^{t-s}\beta(Y(u))\,\mathrm{d}u\Big\}\big].\end{aligned} \tag{3.6.9}$$

If, then, we can identify judicious choices of f and β such that

$$L_X f(\cdot,y)(x) = L_Y f(x,\cdot)(y) + \beta(y)f(x,y),$$

integrating equation (3.6.9) over $0 \le s \le t$ establishes (F). It is easy to verify that taking $f(x,y) := \prod_{i=1}^k x_i^{y_i}$ and $\beta(y) = -\mu\sum_{i=1}^k y_i$ is sufficient, and, given the informal calculation (3.6.9) is valid,

$$\mathbf{E}\big[\prod_{i=1}^k X_i(t)^{y_i}\big] = \mathbf{E}\big[\prod_{i=1}^k x_i^{Y_i(t)}\exp\Big\{-\mu\int_0^t\sum_{i=1}^k Y_i(s)\,\mathrm{d}s\Big\}\big].$$

Moreover, since $X_i(t) \le 1$ and, because Y_i is a decreasing process, $Y_i(t) \le \max_{i=1,\dots,k} y_i$ for all $i = 1,\dots,k$, $t \ge 0$, the estimates (G), (H) and (I) hold trivially. In particular, since the genealogical process Y can be constructed for any initial condition $y \in \mathbb{N}^k$, we have derived an explicit formula for all mixed moments of $X(t)$, $t \ge 0$, and, thus, the one-dimensional distributions of all solutions X to equations (3.3.1) with initial condition $X(0) = x$ coincide. In turn, solutions are unique in distribution.

It remains to justify the informal differentiation of the expectation. We loosely follow the more general arguments of Ethier and Kurtz [40], pp. 192-195. Restricting the analysis to the special case of our setting, the simplicity of the arguments is not lost to abstraction. Setting

$$F(s,t) := \mathbf{E}\big[f(X(s),Y(t))\exp\Big\{\int_0^t\beta(Y(u))\,\mathrm{d}u\Big\}\big],$$

it is clear that

$$\begin{aligned}\frac{F(s+h,t)-F(s,t)}{h} = \mathbf{E}\big[\frac{f(X(s+h),Y(t))-f(X(s),Y(t))}{h} \\ \times\exp\Big\{\int_0^t\beta(Y(u))\,\mathrm{d}u\Big\}\big].\end{aligned}$$

For all test functions g, $M_X(s) := g(X(s)) - \int_0^s L_X g(X(r))\,\mathrm{d}r$ is a martingale and, in

particular, we may therefore express the right hand side of this equation as

$$\frac{F(s+h,t)-F(s,t)}{h} = \mathbf{E}\Big[\frac{1}{h}\int_s^{s+h} L_X f(\cdot, Y(t))(X(r))\,\mathrm{d}r \times \exp\Big\{\int_0^t \beta(Y(u))\,\mathrm{d}u\Big\}\Big], \quad (3.6.10)$$

using that X, and hence M_X, is a process independent of Y and that $E[M_X(s+h) - M_X(s)] = 0$. We proceed to pass to the limit $h \to 0$. With our choice of β, the exponential term inside the expectation on the right hand side is clearly bounded by one; furthermore, by the mean value theorem,

$$\frac{1}{h}\int_s^{s+h} L_X f(\cdot, Y(t))(X(r))\,\mathrm{d}r = L_X f(\cdot, Y(t))(X(r)), \quad \text{a.s.},$$

for some $s < r < s+h$, and so, if there exists some constant $\Gamma < \infty$ such that

$$\sup_{r,t\geq 0} |L_X f(\cdot, Y(t))(X(r))| \leq \Gamma, \quad \text{a.s.}, \quad (3.6.11)$$

then, by Lebesgue's dominated convergence theorem, passing the limit $h \to 0$ under the the expectation on the right hand side of equation (3.6.10) is valid, and

$$\frac{\partial F}{\partial s}(s,t) := \mathbf{E}\Big[L_X f(\cdot, Y(t))(X(s)) \exp\Big\{\int_0^t \beta(Y(u))\,\mathrm{d}u\Big\}\Big].$$

Since, as above, $X_i(t) \leq 1$ and $Y_i(t) \leq \max_{i=1,\ldots,k} y_i$ for all $i = 1,\ldots,k$, $t \geq 0$, it is straightforward to verify the estimate in equation (3.6.11).

Proceeding in the same vein,

$$\begin{aligned}
&\frac{F(s,t+h)-F(s,t)}{h} \\
&= \mathbf{E}\Big[\frac{f(X(s),Y(t+h)) - f(X(s),Y(t))}{h} \\
&\qquad \times \exp\Big\{\int_0^t \beta(Y(u))\,\mathrm{d}u\Big\}\Big] \\
&\quad + \mathbf{E}\Big[f(X(s),Y(t+h)) \exp\Big\{\int_0^t \beta(Y(u))\,\mathrm{d}u\Big\} \\
&\qquad \times \frac{\exp\{\int_t^{t+h} \beta(Y(u))\,\mathrm{d}u\} - 1}{h}\Big].
\end{aligned} \quad (3.6.12)$$

For all test functions g, $M_Y(s) := g(Y(s)) - \int_0^s L_Y g(X(r))\,\mathrm{d}r$ is a martingale. Moreover, $M_Y(t+s) - M_Y(t)$ is independent of $Y(u)$, $u \leq t$, and so the first expectation on the right

hand side may be written as

$$\mathbf{E}\Big[\frac{1}{h}\int_t^{t+h} L_X f(X(s),\cdot)(Y(r))\,\mathrm{d}r \exp\Big\{\int_0^t \beta(Y(u))\,\mathrm{d}u\Big\}\Big].$$

Again applying the mean value theorem, one can show that

$$\frac{1}{h}\int_t^{t+h} L_X f(X(s),\cdot)(Y(r))\,\mathrm{d}r \exp\Big\{\int_0^t \beta(Y(u))\,\mathrm{d}u\Big\} \le L_Y f(X(s),\cdot)(Y(r)), \quad \text{a.s.},$$

for some $t \le r \le t+h$. Since this right hand side is bounded uniformly in $r, s \ge 0$, we may apply the dominated convergence theorem to pass the limit $h \to 0$ under the expectation, exactly as above.

Turning now to the terms in the second expectation on the right hand side of equation (3.6.12), with our choice of the function β,

$$\begin{aligned}\Big|\exp\Big\{\int_t^{t+h}\beta(Y(u))\,\mathrm{d}u\Big\} - 1\Big| &= 1 - \exp\Big\{\int_t^{t+h}\beta(Y(u))\,\mathrm{d}u\Big\}\\ &\le 1 - \exp\{-\mu k \max_{i=1,\dots,k} y_i h\}\\ &\le \mu k \max_{i=1,\dots,k} y_i h.\end{aligned}$$

By a second application of Lebesgue's dominated convergence theorem, passing to the limit $h \to 0$ in equation (3.6.12),

$$\frac{\partial F}{\partial t}(s,t) = E\Big[\big(L_Y f(X(s),\cdot)(Y(t)) + \beta(Y(t)) f(X(s),Y(t))\big)\exp\Big\{\int_0^t \beta(Y(u))\,\mathrm{d}u\Big\}\Big].$$

Finally, we note that

$$\int_0^T\int_0^T \Big(\Big|\frac{\partial F}{\partial s}(s,t)\Big| + \Big|\frac{\partial F}{\partial t}(s,t)\Big|\Big)\,\mathrm{d}s\,\mathrm{d}t < \infty,$$

then, by the Fubini–Tonelli Theorem, exchanging the order of integration yields that

$$\int_0^T\int_0^t (\partial_1 F(s,t-s) - \partial_2 F(s,t-s))\,\mathrm{d}s\,\mathrm{d}t = \int_0^T (F(s,0) - F(0,s))\,\mathrm{d}s.$$

Differentiating with respect to T, we have verified equation (3.6.9) in its integrated form.

3.6.5. Proof of the diffusion approximation. We will verify the conditions of the Stroock–Varadhan Theorem. Condition (B'), and thus (B), was verified establishing a duality relation with the genealogical process in the previous appendix.

Let Q^N be defined as in equation (3.6.2), then, since there are N individuals each inde-

pendently entering a reproductive phase at rate N, it is clear that $Q^N(x, \mathbb{R}^k \backslash \{x\}) < N^2$ and thus condition (A) is trivially satisfied.

Let b_i^N be defined as in equation (3.6.4). By considering each possible outcome of an individual's reproductive phase,

$$
\begin{aligned}
& b_i^N(x) = \\
& + \frac{1}{N_i} \cdot N_i x_i \cdot N(1-\bar{\sigma})(1-\bar{\mu})(1-x_i) \\
& - \frac{1}{N_i} \cdot N_i x_i \cdot N(1-\bar{\sigma})\bar{\mu} x_i \\
& - \frac{1}{N_i} \cdot N_i (1-x_i) \cdot N(1-\bar{\sigma}) x_i \\
& + \frac{1}{N_i} \cdot N_i \cdot N\bar{\sigma} \sum_{j \neq i} \theta_{ij}^N(c) \frac{1}{2}(d x_j + (1-d)x_i)(1-\bar{\mu})(1-x_i) \\
& + \frac{1}{N_i} \cdot \sum_{j \neq i} N_j \cdot N\bar{\sigma} \theta_{ji}^N(c) \frac{1}{2}(d x_j + (1-d)x_i)(1-\bar{\mu})(1-x_i) \\
& - \frac{1}{N_i} \cdot N_i \cdot N\bar{\sigma} \sum_{j \neq i} \theta_{ij}^N(c) \frac{1}{2}(d(1-x_j) + (1-d)(1-x_i))(1-\bar{\mu}) x_i \\
& - \frac{1}{N_i} \cdot \sum_{j \neq i} N_j \cdot N\bar{\sigma} \theta_{ji}^N(c) \frac{1}{2}(d(1-x_j) + (1-d)(1-x_i))(1-\bar{\mu}) x_i \\
& - \frac{1}{N_i} \cdot N_i \cdot N\bar{\sigma} \sum_{j \neq i} \theta_{ij}^N(c) \frac{1}{2} \bar{\mu} x_i \\
& - \frac{1}{N_i} \cdot \sum_{j \neq i} N_j \cdot N\bar{\sigma} \theta_{ji}^N(c) \frac{1}{2} \bar{\mu} x_i .
\end{aligned}
$$

The first term on the right hand side covers the case that the proportion of type i individuals carrying the ancestral type a_1 at l increases by N_i^{-1} during an asexual reproductive phase; this is the case only if one of the $N_i x_i$ type i individuals carrying a_1 enter an asexual reproductive phase (at rate $N(1-\bar{\sigma})$), there is no mutation (with probability $1-\bar{\mu}$), and one of the type i individuals not carrying the ancestral type a_1 at the neutral marker is replaced (with probability $1 - x_i$).

The second and third terms on the right hand side cover the case that the proportion of type i individuals carrying the ancestral type a_1 at l decreases by N_i^{-1} during an asexual reproductive phase; this is the case if, either one of the $N_i x_i$ type i individuals carrying a_0 enters an asexual phase (at rate $N(1-\bar{\sigma})$), there is a mutation (with probability $\bar{\mu}$), and one of the same group is replaced (with probability x_i), or if one of the $N_i(1-x_i)$ type i individuals not carrying a_1 enters an asexual phase (at rate $N(1-\bar{\sigma})$) and one of the a_1-carriers is replaced (with probability x_i). Note that whether a mutation occurs or not

during this latter phase is immaterial since we have assumed there can be no backwards mutation.

The fourth and fifth terms on the right hand side cover the case that the proportion of type i individuals carrying the ancestral type a_1 at l increases by N_i^{-1} during a sexual reproductive phase in which there is no mutation at l. The fourth term corresponds to the first (of two) cases that one of the N_i type i individuals enters a sexual reproductive phase (at rate $N\bar{\sigma}$) and, for any $j \neq i$, successfully finds a mate of type j (with probability $\theta_{ij}^N(c)$); then, with probability $(dx_j + (1-d)x_i)/2$, the offspring inherits the mating type from the focal parent and the ancestral type a_1 at the locus l; there is no mutation (with probability $1-\bar{\mu}$) and the replaced individual does not carry a_1 (with probability $1-x_i$). The fifth term is derived similarly, except the focal parent is of any type $j \neq i$ and the chosen mate is of type i.

The sixth and seventh terms cover the case that a type i individual carrying a_1 is replaced during a sexual reproductive phase and there is no mutation. These may be derived similarly to the fourth and fifth terms.

Lastly, the eighth and ninth terms correspond to the case that, during a sexual reproductive phase, a type i individual carrying the ancestral type a_1 at l is lost due to being replaced by a mutant type. The eighth term is the case that one of the N_i type i individuals enters a sexual reproductive phase (at rate $N\bar{\sigma}$) and for any $j \neq i$ finds a compatible mate (with probability $\theta_{ij}^N(c)$, the progeny inherits the mating type from the focal parent (with probability $1/2$), there is a mutation (with probability $\bar{\mu}$), and a type i individual which carries a_1 is replaced (with probability x_i). The ninth term is derived similarly, except the focal parent is of any type $j \neq i$ and the chosen mate is of type i.

Recall that, in the regime of rare sex and rare mutation, $\bar{\sigma} = \sigma/N$ and $\bar{\mu} = \mu/N$. The first three terms sum to

$$-\big(1-\frac{\sigma}{N}\big)\mu x_i,$$

the subsequent four terms sum to

$$\big(1-\frac{\mu}{N}\big)\frac{\sigma d}{2}\sum_{j\neq i}\frac{N_i\theta_{ij}^N(c)+N_j\theta_{ji}^N(c)}{N_i}(x_j-x_i),$$

and the final two terms sum to

$$-\frac{\mu}{N}\frac{\sigma}{2}x_i\sum_{j\neq i}\frac{N_i\theta_{ij}^N(c)+N_j\theta_{ji}^N(c)}{N_i}.$$

Setting

$$b_i(x) := -\mu x_i + \frac{\sigma d}{2}\sum_{j\neq i}\frac{p_i\theta_{ij}(c)+p_j\theta_{ji}(c)}{p_i}(x_j-x_i),$$

it is straightforward to show that

$$
\begin{aligned}
&|b_i^N(x) - b_i(x)| \leq \\
&+\frac{\sigma}{N}\mu x_i \\
&+\frac{\sigma d}{2}\sum_{j\neq i}\left|\frac{N_i\theta_{ij}^N(c) + N_j\theta_{ji}^N(c)}{N_i} - \frac{p_i\theta_{ij}(c) + p_j\theta_{ji}(c)}{p_i}\right||x_j - x_i| \\
&+\frac{\mu}{N}\frac{\sigma d}{2}\sum_{j\neq i}\frac{N_i\theta_{ij}^N(c) + N_j\theta_{ji}^N(c)}{N_i}(d|x_j - x_i| + 1),
\end{aligned}
$$

and, since we may easily bound $x_i \leq 1$, thus that condition (D) is satisfied.

Let a_{ij}^N be defined as in equation (3.6.3). By considering each possible outcome of an individual's reproductive phase, following the calculation of $b_i^N(x)$ exactly,

$$
\begin{aligned}
&a_{ii}^N(x) = \\
&+\frac{1}{N_i^2}\cdot N_i x_i \cdot N(1-\bar\sigma)(1-\bar\mu)(1-x_i) \\
&+\frac{1}{N_i^2}\cdot N_i x_i \cdot N(1-\bar\sigma)\bar\mu x_i \\
&+\frac{1}{N_i^2}\cdot N_i(1-x_i)\cdot N(1-\bar\sigma)x_i \\
&+\frac{1}{N_i^2}\cdot N_i\cdot N\bar\sigma\sum_{j\neq i}\theta_{ij}^N(c)\frac{1}{2}(dx_j + (1-d)x_i)(1-\bar\mu)(1-x_i) \\
&+\frac{1}{N_i^2}\cdot \sum_{j\neq i} N_j\cdot N\bar\sigma\theta_{ji}^N(c)\frac{1}{2}(dx_j + (1-d)x_i)(1-\bar\mu)(1-x_i) \\
&+\frac{1}{N_i^2}\cdot N_i\cdot N\bar\sigma\sum_{j\neq i}\theta_{ij}^N(c)\frac{1}{2}(d(1-x_j) + (1-d)(1-x_i))(1-\bar\mu)x_i \\
&+\frac{1}{N_i^2}\cdot \sum_{j\neq i} N_j\cdot N\bar\sigma\theta_{ji}^N(c)\frac{1}{2}(d(1-x_j) + (1-d)(1-x_i))(1-\bar\mu)x_i \\
&+\frac{1}{N_i^2}\cdot N_i\cdot N\bar\sigma\sum_{j\neq i}\theta_{ij}^N(c)\frac{1}{2}\bar\mu x_i \\
&+\frac{1}{N_i^2}\cdot \sum_{j\neq i} N_j\cdot N\bar\sigma\theta_{ji}^N(c)\frac{1}{2}\bar\mu x_i.
\end{aligned}
$$

The first three terms sum to

$$
2\big(1-\frac{\sigma}{N}\big)\big(1-\frac{\mu}{N}\big)\frac{N}{N_i}x_i(1-x_i) + \frac{\mu}{N}\big(1-\frac{\sigma}{N}\big)\frac{N}{N_i}x_i(1+2x_i),
$$

the subsequent four terms sum to

$$\frac{1}{N_i}\big(1-\frac{\mu}{N}\big)\frac{\sigma}{2}\sum_{j\neq i}\Big[\frac{N_i\theta_{ij}^N(c)+N_j\theta_{ji}^N(c)}{N_i}\big(2x_i(1-x_i)+d(1-2x_i)(x_j-x_i)\big)\Big],$$

and the final two terms sum to

$$\frac{1}{N_i}\frac{\mu}{N}\frac{\sigma}{2}x_i\sum_{j\neq i}\frac{N_i\theta_{ij}^N(c)+N_j\theta_{ji}^N(c)}{N_i}.$$

Setting

$$a_{ii}(x):=\frac{2}{p_i}x_i(1-x_i),$$

it is straightforward to show that, as with the case of the sequence $b_i^N(x)$ and its putative limit $b_i(x)$, that condition (C) is satisfied for $i=j$. Considering now $i\neq j$, since there may be at most one replacement of an individual during any one reproductive phase, we note that there cannot be simultaneous changes in the proportion of type i and type j individuals carrying a_1 at l. In particular, $a_{ij}^N(x)=0$ for all $i\neq j$ and condition (C) is satisfied with $a_{ij}(x)=0$.

Let γ^N be defined as in equation (3.6.5). It follows from Jensen's inequality that,

$$|x-y|^4\leq k\sum_{i=1}^k|x_i-y_i|^4$$

and thus that

$$\gamma^N(x)\leq k\sum_{i=1}^k\int_{\mathbb{R}^k}|x_i-y_i|^4Q^N(x,\mathrm{d}y)=k\sum_{i=1}^k\frac{1}{N_i^2}a_{ii}^N(x).$$

Condition (E) then follows from (C).

3.6.6. Proof of the approximation of genealogies. The dynamics of the genealogical process form a continuous-time Markov chain. Since this process has a bounded Markov generator, it is sufficient to check the convergence of the individual rates.

Consider first a sample of two type i individuals in a large population of size N. Let us determine how far back in time we have to look until we see an asexual reproductive phase from which our focal individuals were the two outputs (the surviving parent and its progeny). Looking backwards-in-time, either of the ancestral lines we are tracing encounters an asexual phase where it is the parent at total rate $2N(1-\bar{\sigma})$; this event concludes by replacing the second ancestral line with probability N_i^{-1}. Thus, we encounter a common ancestor (for the allelic type at l) after an exponentially distributed amount of time with

rate parameter

$$2N(1-\bar{\sigma})N_i^{-1}.$$

This reproductive phase concludes without mutation with probability $1-\bar{\mu}$, in which case we say that the two individuals coalesce, and with a mutation with probability $\bar{\mu}$. Passing to the limit $N \to \infty$, the two individuals coalesce at total rate

$$\lim_{N\to\infty} 2N(1-\bar{\sigma})N_i^{-1}(1-\bar{\mu}) = \frac{2}{p_i}.$$

Now let us proceed to the more general setting, considering a sample of n individuals from a population of size $N \gg n$. Denote by n_i the number of this sample who are of type i. In particular, $\sum_{i=1}^{k} n_i = n$. We proceed by establishing the genealogy of the sample at the locus l in the large population limit $N \to \infty$.

We begin by thinking about what happens to the genealogy when we encounter an asexual reproductive phase in the total population. Let us first restrict our attention to the case that these conclude without mutation. These events occur at total rate $N^2(1-\bar{\sigma})(1-\bar{\mu})$ so only events of probability $O(1/N^2)$ or greater will be seen after passage to the limit. With probability $n_i/N \cdot (n_i-1)/N_i$ this is an asexual phase involving two ancestral lines in our sample. In this case, the two individuals are said to coalesce. In the limit, this happens then at total rate

$$\lim_{N\to\infty} N^2(1-\bar{\sigma})\frac{n_i}{N}\frac{n_i-1}{N_i}(1-\bar{\mu}) = \frac{2}{p_i}\binom{n_i}{2},$$

and, after this event, we need only continue to trace $n_i - 1$ ancestral lines in order to establish the genealogy of our original sample. We must also consider the case that either, or both, of the ancestral lines involved during an asexual phase are not in our sample of n individuals. Although this happens at total rate

$$N^2(1-\bar{\sigma})\sum_{i=1}^{k}\big(2\frac{n_i}{N}\frac{N_i-n_i}{N_i} + \frac{N_i-n_i}{N}\frac{N_i-n_i}{N_i}\big)(1-\bar{\mu}),$$

which blows up as $N \to \infty$, we still continue to need to trace n_i ancestral lines of type i and thus do not see any change in the observed genealogical history.

We consider the effect on the genealogy now when we encounter sexual reproductive phases; again, we begin by restricting our attention to those which do not conclude with a mutation. These events occur at total rate $N^2\bar{\sigma}(1-\bar{\mu})$. Recalling that $\bar{\sigma} = \sigma/N$, we thus need only consider events of probability $O(1/N)$ or greater, since only these will be retained after passage to the limit. Consider when we will encounter a sexual event where one of the sampled n_i type i individuals inherits its allelic type from a type j individual.

This occurs with probability

$$\frac{N_i - n_i}{N}\theta_{ij}^N(c)\frac{N_j - n_j}{N_j}\frac{d}{2}\frac{n_i}{N_i} + \frac{N_j - n_j}{N}\theta_{ji}^N(c)\frac{N_i - n_i}{N_i}\frac{d}{2}\frac{n_i}{N_i} + O(1/N^2),$$

ie, when the focal parent is of type i but not in our sample of ancestral lines; it successfully finds a type j mate; this type j mate is also not in our sample; the progeny inherits the mating type from the type i individual and the allelic type from the type j; and the replaced type i individual is in our sample. This similarly happens when the focal parent is of type j and the chosen mate is of type i, corresponding to the second term in this sum. It would also be the case if either, or both, of the parents are in our sample, but this event is of probability $O(1/N^2)$ and thus neglected. Thus, after passage to the limit, we need to continue to trace one fewer type i ancestral line and one further type j ancestral line at rate

$$\lim_{N\to\infty} N^2\bar{\sigma} \times \Big(\frac{N_i - n_i}{N}\theta_{ij}^N(c)\frac{N_j - n_j}{N_j} + \frac{N_j - n_j}{N}\theta_{ji}^N(c)\frac{N_i - n_i}{N_i}\Big)\frac{d}{2}\frac{n_i}{N_i}(1-\bar{\mu})$$
$$= \frac{\sigma d}{2}\frac{p_i\theta_{ij}(c) + p_j\theta_{ji}(c)}{p_i}n_i = m_{ij}(p,c)n_i.$$

Sexual phases which conclude with the replacement of an ancestral line not in our sample, although possibly occurring with $O(1)$ probability, are not seen when tracking just the number of ancestral lines needed to establish the genealogy of our original sample of n individuals.

Let us turn now to events which also involve a mutation. In order to observe a change in our genealogical process, the progeny of the event must be in our sample. This occurs with probability n_i/N_i if the inherited mating type is i. Thus, a reproductive phase is sexual, includes a mutation, and actually observed with probability of order $O(\bar{\sigma}\bar{\mu}\sum_{i=1}^k n_i N_i^{-1}) = O(N^{-3})$. Since events occur at total rate N^2, all such events may be neglected. Turning to asexual phases, we observe a mutation when the ancestral line we are tracing is the progeny of an individual not in a sample and a mutation occurs; a type i ancestral line experiences a mutation at l total rate

$$\lim_{N\to\infty} N^2(1-\bar{\sigma})\frac{N_i}{N}\bar{\mu}\frac{n_i}{N} = n_i\mu.$$

3.6.7. Identity by descent in a population with k self-incompatible mating type classes. In this section we derive a system of equations for the probability that two individuals are identical by descent; our system is specialised to and subsequently solved in the case that $p_i = 1/k$ for all $i = 1, 2, \ldots, k$.

To ease notation we will write $m_{ij}(p,c) \equiv m_{ij}$. Conditioning on the first event,

$$\pi_{ii} = \frac{2/p_i}{2\mu + 2/p_i + 2\sum_{j\neq i} m_{ij}} + \sum_{j\neq i} \pi_{ij} \frac{2m_{ij}}{2\mu + 2/p_i + 2\sum_{j\neq i} m_{ij}}.$$

For $i \neq j$,

$$\pi_{ij} = \sum_{l\neq i} \pi_{lj} \frac{m_{il}}{2\mu + \sum_{l\neq i} m_{il} + \sum_{l\neq j} m_{jl}} + \sum_{l\neq j} \pi_{il} \frac{m_{jl}}{2\mu + \sum_{l\neq i} m_{il} + \sum_{l\neq j} m_{jl}}.$$

Tractable solutions to this system of equations do not exist in general. Specialising to the case that $p_i = 1/k$, writing $\pi_{ij} \equiv \pi_{12} =: \pi_d$ if $i \neq j$, $\pi_{ij} \equiv \pi_{11} =: \pi_s$ if $i = j$, and $m_{ij} \equiv m_{12} =: m = \sigma d/(k - e^{-c})$, this system of equations may be rewritten as

$$\pi_s = \frac{k}{\mu + k + (k-1)m} + \pi_d \frac{(k-1)m}{\mu + k + (k-1)m}. \tag{3.6.13}$$

For $i \neq j$,

$$\pi_d = \frac{(k-2)m}{\mu + (k-1)m}\pi_d + \frac{m}{\mu + (k-1)m}\pi_s. \tag{3.6.14}$$

A simple rearrangement of equation (3.6.14) yields that

$$\pi_d = \frac{m}{\mu + m}\pi_s,$$

and subsequently substituting this into equation (3.6.13) gives that

$$\pi_s = \frac{k}{k + \mu(1 + (k-1)\frac{m}{\mu+m})},$$
$$\pi_d = \frac{m}{\mu + m} \frac{k}{k + \mu(1 + (k-1)\frac{m}{\mu+m})}.$$

3.6.8. Identity by descent in mixed self-incompatible and unisexual population. In this section we derive a system of equations for the probability of identity at the neutral marker l for pairs of individuals chosen from a population consisting of $k-1$ self-incompatible mating type classes of equal size, and a unisexual class consisting of a proportion α of the total population. Let the probabilities π_s, π_d, π_s^u and π_d^u be as defined in the section 'Mixed self-incompatible and unisexual populations' above.

We use the limiting genealogical process to compute the probabilities of identity by descent. Two individuals will be identical by descent if they coalesce before either experiences a mutation.

Two lineages of the same self-incompatible type coalesce at rate $2(k-1)/(1-\alpha)$ and

two individuals of the unisexual type coalesce at rate $2/\alpha$. Conditioning on the first event, which may be a coalescence, a mutation or a migration,

$$\pi_s = \frac{2\frac{k-1}{1-\alpha} + 2(k-2)m\pi_d + 2m_{0u}\pi_d^u}{2\mu + 2\frac{k-1}{1-\alpha} + 2(k-2)m + 2m_{0u}}, \tag{3.6.15}$$

$$\pi_d = \frac{2(k-3)m\pi_d + 2m\pi_s + 2m_{0u}\pi_d^u}{2\mu + 2(k-2)m + 2m_{0u}}, \tag{3.6.16}$$

$$\pi_s^u = \frac{2\frac{1}{\alpha} + 2(k-1)m_{u0}\pi_d^u}{2\mu + 2\frac{1}{\alpha} + 2(k-1)m_{u0}}, \tag{3.6.17}$$

$$\pi_d^u = \frac{m(k-2)\pi_d^u + m_{0u}\pi_s^u + (k-2)m_{u0}\pi_d + m_{u0}\pi_s}{2\mu + m(k-2) + m_{0u} + (k-1)m_{u0}}. \tag{3.6.18}$$

We now consider the case that $\alpha = 1/k$ and $c = \infty$; in this case, as discussed in the main body of the paper, symmetry between the self-incompatible and unisexual classes is introduced and

$$\pi_s = \pi_s^u = \frac{k\mu + \sigma d}{k\mu + \sigma d + \mu(\mu + \sigma d)},$$

$$\pi_d = \pi_d^u = \frac{\sigma d}{k\mu + \sigma d + \mu(\mu + \sigma d)}.$$

Let us now turn to computing ϕ and ϕ^u, the probability of mating with an individual distinct at the neutral marker for self-incompatible and unisexual individuals respectively. The symmetry does not carry over to these quantities because the unisexual types retain the ability to mate with themselves. The quantities are straightforward to compute:

$$\begin{aligned}\phi &= \frac{k-2}{k}(1-\pi_d) + \frac{1}{k}(1-\pi_d^u) \\ &= \frac{k-1}{k}(1-\pi_d) \\ &= \frac{k-1}{k}\frac{k\mu + \mu(\mu+\sigma d)}{k\mu + \sigma d + \mu(\mu+\sigma d)},\end{aligned}$$

and

$$\begin{aligned}\phi^u &= \frac{k-1}{k}(1-\pi_d^u) + \frac{1}{k}(1-\pi_s^u) \\ &= \frac{k-1}{k}(1-\pi_d) + \frac{1}{k}(1-\pi_s) \\ &= \frac{(k-1)\mu + \mu(\mu+\sigma d)}{k\mu + \sigma d + \mu(\mu+\sigma d)}.\end{aligned}$$

CONCLUSION AND FURTHER WORK

In this last chapter, we review the contributions of this thesis and comment on possible extensions for further work.

Ancestral lineages.

In spatially structured populations there is a substantial interest in understanding gene flow and the resultant patterns of genetic variation in the population. If we begin with a single individual from some population alive today and choose a particular bit of her genome, then, looking back through history, we can imagine recording the geographic location of the ancestor from whom that individual inherited that bit of genome. This spatial motion, as a function of amount of time in the past, we called a *lineage*. Understanding the spatial motion of these lineages is fundamental to explaining patterns of genetic variation in spatially structure populations.

The demographic details of an individual-based stochastic population model determine not only the time evolution of population sizes, but also the dynamics of lineages. Whilst in the simplest models the motion of the lineage is given by a simple random walk, in more general models they are given by a random walk in a random environment. We sought to understand the motion of these lineages for two contrasting models of spatially distributed populations: in Chapter 1, where the population size dynamics depend on nonlocal interactions between adults and juveniles; and, in Chapter 2, when the population size dynamics are determined entirely by exogenous factors.

Gene flow in a population model with nonlocal dispersal of juveniles.

In Chapter 1 we defined a model where the demographic parameters depended on the local population size in a near neighbourhood. We defined a sequence of individual based models, indexed by $n \in \mathbf{N}$ where, for two sequences τ_n, η_n, of non-negative real numbers, individuals live in *demes* indexed by $\mathbf{Z}^d$ and the population evolves in discrete generations with inter-generation time τ_n. Each individual in the population is assigned 'mass' η_n. We write $X_t^n(x)$ for the size of the population in deme $x \in \mathbf{Z}^d$ at time t, that is

$$X_t^n(x) := \eta_n \times \text{number of particles alive at } x \text{ at time } t.$$

Reproduction is described as follows. For each $t \in \{\tau_n, 2\tau_n, \ldots, \lfloor T/\tau_n \rfloor \tau_n\}$,

1. at each point $x \in \mathbf{Z}^d$, each individual, independently, produces a Poisson number of juvenile offspring with parameter $\gamma_n(X^n_{t-}(x))$;

2. at each point $x \in \mathbf{Z}^d$, after production of juveniles, each adult at x, independently, dies with probability $\mu_n(X^n_{t-}(x))$;

3. with probability $m\tau_n$ each juvenile, independently, migrates to a neighbouring deme (chosen with equal probability from among its nearest neighbours), otherwise it stays in its parental deme;

4. after migration, for each $x \in \mathbf{Z}^d$, each juvenile at x, independently, establishes (and so becomes an adult in the population) with probability $r_n(X_{t-}(x))$. (We treat this maturation as instantaneous.)

Defining g_n by

$$g_n = \frac{r_n\gamma_n - \mu_n}{\tau_n},$$

and assuming $\tau_n \to 0$, $\eta_n \to 0$, $\eta_n/\tau_n := \delta_n \to \delta \in [0, \infty)$, we proved that the sequence of processes $(X^n)_{n\in\mathbf{N}}$ is weakly convergent in the Skorokhod topology to the interacting system of SDEs

$$\begin{aligned} X_0(x) &= u(x), \\ X_t(x) &= X_0(x) + \int_0^t X_s(x) g(X_s(x))\,\mathrm{d}s + \int_0^t \frac{m}{2} r(X_s(x))\Delta_1(\Phi \circ X_s)(x)\,\mathrm{d}s \\ &\qquad + \int_0^t \sqrt{\delta\mu(X_s(x))(2 - \mu(X_s(x)))X_s(x)}\,\mathrm{d}W_s(x), \end{aligned} \tag{3.6.19}$$

under the conditions of Assumptions 1.2.3. This was the statement of Theorem 1.2.5.

Furthermore, denoting by L^n_t the process which records the location at time t before the present of the lineage ancestral to a lineage sampled from the origin (if it exits), under the same rescaling, and after imposing further minor conditions on the parameters, the sequence of processes $(X^n, L^n)_{n\in\mathbf{N}}$ is weakly convergent to the pair (X, L), where the law of L is given conditional on X and satisfies the stopped martingale problem: L is the unique process such that $L_0 = 0$ and

$$f(L_t) - f(L_0) - \int_0^{t\wedge T^c} A^X_s f(L_s)\,\mathrm{d}s$$

is a $\mathbf{P}[\cdot|X]$-martingale for each $f \in B(\mathbf{Z})$, where

$$A_s^X f(x) := \frac{1}{2}\frac{r(X_{T-t}(x))}{X_{T-t}(x)}\Big(\Phi(X_{T-t}(x+1))\big(f(x+1)-f(x)\big) + \Phi(X_{T-t}(x-1))\big(f(x-1)-f(x)\big)\Big),$$

and $T^c := \inf\{t \geq 0 : X_{T-t}(L_t) < c\}$.

Taking $\delta = 0$, under the conditions of Assumptions 1.2.12, the system (3.6.19) has a stable fixed point where the population size is equal to u_* at all demes $x \in \mathbf{Z}$. Defining

$$Z_t^n(x) := \frac{X_t^n(x) - u_*}{\delta_n^{1/2}}$$

to be the fluctuations about this stable fixed point, the sequence of processes $(Z^n)_{n\in\mathbf{N}}$ is weakly convergent to a system of Ornstein–Uhlenbeck processes, coupled through an additional drift term. This is the result of Theorem 1.2.14.

Under further conditions of the parameters which guarantee that the drift term in this system of Ornstein–Uhlenbeck processes is sufficiently regularising (Assumptions 1.2.17), the limiting process Z admits a reversible stationary distribution.

Whilst we are able to make initial progress on the behaviour of the lineages in this regime there is significant scope to push our analysis further. We do however, see two key quantities, $(\log\Phi)'_*$ and $(\log(r\gamma)'_*$ which determine the speed and direction of the ancestral lineages. If $\Phi'_* > 0$ is positive, ancestral lineages jump to demes of higher population density, and if $(r\gamma)'_* < 0$, once at a local maximum, they tend to stay there, in the sense that they jump more slowly than average. On the other hand if $\Phi'_* < 0$, the lineage tends to jump to regions of lower population density, and will loiter in local minima if $(r\gamma)'_* < 0$.

Gene flow in patchy habitats.

In Chapter 2 we sought to quantify the patterns of genetic variation in populations with structure engendered exogenously, by the presence or absence of transient, patchy resources. In our model, the habitat at each location of space may take one of two values, *good* or *bad*, indicating the presence or absence of such a resource. We assumed that the population was distributed across the infinite one-dimensional lattice and the population size at each site was determined by the quality of the habitat there: G in regions of good habitat and B in regions of bad habitat.

We denoted by $N(x,k)$ the population size at location $x \in \mathbf{Z}$ k generations ago and by $m(x,y)$ the probability that offspring of an ancestor at x establish at location y. The lineages $(L_k)_{k\geq 0}$, where L_k is the location k generations ago of the (maternal) ancestor of a given individual who is today randomly chosen from L_0, form a Markov chain with

transition probabilities

$$\mathbf{P}[L_k = x \mid L_{k-1} = y] = \begin{cases} \dfrac{m(x,y)}{\sum_z m(z,y)}, & \text{if } \displaystyle\sum_z N(z,k)m(z,y) = 0, \\ \dfrac{N(x,k)m(x,y)}{\sum_z N(z,k)m(z,y)}, & \text{otherwise.} \end{cases}$$

We passed to a limit where $B \to 0$ so that, when given the opportunity, lineages will always move from bad to good habitats and remain within them until their origination. We set, without loss of generality $G = 1$. In spite of its apparent simplicity, our model allows us to explore a variety of scenarios, while retaining some analytical and numerical tractability.

By varying the spatial and temporal autocorrelation of habitat quality in several models, we can highlight how these determine the long-term behaviour of lineage movement, and hence spatial patterns of relatedness and gene flow. In order to control the spatial autocorrelations we specified that dynamics of the habitat be stationary and reversible and, for $p_0 > 0$ and $p_1 > 0$, with stationary distribution given by the Fortuin–Kastelyn measure

$$(1-p_0)^{N_0}(1-p_1)^{N_1}\left(\frac{p_0 p_1}{(1-p_0)(1-p_1)}\right)^{N_c},$$

where N_0 and N_1 are the number of zeroes and ones, respectively, and N_c is the number of patches of good habitat, ie the number of runs of ones.

In Section 2.2.2 we found that when resampling the environment after each discrete generation we found that lineages could be both faster or slower than one moving in a homogeneous environment. In Section 2.2.3, the same phenomenon was encountered when the environment was held fixed for two generations.

However, this is a rather unnatural thing to do. In Section 2.2.4 we passed to a limit where the patches and the lineages evolved in continuous-time. The system of habitat quality evolved as an interacting particle system; the quality of the habitat at each location $x \in \mathbf{Z}$ flipped from *bad* to *good*, or *vice versa* at a rate dependent on the state of the habitat in some finite neighbourhood. In this same limit, the lineage was given by a continuous-time random walk in a dynamic random environment. Whilst the lineage jumps at a constant rate σ^2 it always chooses the jump to areas of good habitat when they are available.

Our goal was to compute the long-run mean square displacement of the lineages, or their effective dispersal. Since lineage speed is determined by the length of time they are trapped within patches of good habitat and motion of patches during this time, a thorough understanding of patches in the space-time plane was required. In Theorem 2.3.4 it was shown that after making a suitable choice of parameters, after a diffusive scaling of space

and time and passing to a limit where patches do not interact with one another, the midpoint of the patch and its width are well-approximated by independent diffusion processes. The limiting midpoint process M is given by a standard Brownian motion. The width process W is given by the solution to the Skorokhod problem for the diffusion process S given by

$$S_t = W_0 - 2v_1 t + 2B_t^W,$$

where B^W is a Brownian motion independent of M and $v_1^{-1} > 0$ is the average width of patches. Moving backwards through time, the patch reaches its origination at time T. For some constant $\gamma_b > 0$, it was shown that T is given by

$$T = \inf\{t \geq 0 : \ell_t \geq 2U/\gamma_b\},$$

where ℓ is the local time process

$$\ell_t := 0 \vee \big(-\inf_{s \leq t} S_s\big),$$

and U is an exponential random variable with unit mean independent of M and B^W.

The behaviour of the patches of good habitat in this limit allows us to make a number of qualitative inferences about the speed of lineages as the parameters of the model are varied. In a limit where the lineages move fast enough so that they encounter these increasingly far apart patches in finite time, we give an argument to support the conjecture that the speed of the lineage will be proportional to the proportion of time spent in regions of bad habitat. Moreover we conjecture that the effective dispersal rate, after suitable rescaling, should be given by

$$\frac{1}{\frac{1}{\gamma_b v_1} + \frac{1}{2v_1^2} + \frac{1}{\sigma^2}}.$$

Genetic diversity in populations subdivided into self-incompatible mating types.

In Chapter 3 we considered a model for a population structured by its reproductive strategy. Most fungi, ciliates and algae, mating is similarly restricted to occurring between individuals from distinct self-incompatible classes, termed *mating types*. However, the number of mating types is not necessarily restricted to two, with up to 23,328 in fungi and 100 in ciliates. We sought to understand when such population structure would evolve and the reason for such variability in the mating type number.

The novelty in our methods is that we abandon any attempt to model selection for recombination, a major obstacle to previous work, and instead take a courser grained perspective. Our approach is to quantify the patterns of genetic variability by computing the probabil-

ity of identity by descent for individuals sampled from either the same or distinct mating types. For our model, we are able to express this probability as a function of the number of mating types and the frequency of sex. Assuming that k mating types occur in constant frequency in the population, in an infinite population limit, the probability that two individuals sampled from the same mating type class are identical by descent for an allele at a neutral marker on the genome is given by

$$\pi_s = \frac{k[1 + (k-1)\frac{m}{\mu+m}]^{-1}}{k[1 + (k-1)\frac{m}{\mu+m}]^{-1} + \mu},$$

and, for two individuals sampled from different mating type classes, is given by

$$\pi_d = \frac{m}{\mu + m} \frac{k[1 + (k-1)\frac{m}{\mu+m}]^{-1}}{k[1 + (k-1)\frac{m}{\mu+m}]^{-1} + \mu},$$

where σ is the average number of sexually reproductive events per generation, μ is the average number of mutations, and m is the effective gene flow between mating type classes due to recombination. The parameter m is proportional to the probability that an individual is successfully able to find a mate when it attempts sex. Our methods involved using the Stroock–Varadhan theory of approximation of Markov chains by diffusions and constructing the corresponding genealogical process in the limit. This genealogical process is given by the structured coalescent where, unlike in classical models, the structure refers to mating type classes rather than spatial locations.

We have shown that although self-incompatibility decreases the probability of successful sexual reproduction, it can in fact increase the probability of effectual sexual reproduction between non-clone mates, even exceeding the rates of effectual sexual reproduction in unisexuals. However the strength of this effect is highly dependent on population parameters including its effective size, the rate of sexual reproduction and the population composition. Increasing the number of mating types and decreasing the number of sexual events per generation σ increases the relative benefits to self-incompatibility over unisexuality. Meanwhile decreasing the cost to mate finding (ie allowing more opportunities for the self-incompatible types to find a compatible mate) has a particularly prominent effect on the probability of effective sex in the self-incompatible population, allowing it to more readily exceed that in a unisexual population.

We found that in this mixed population consisting of $k - 1$ self-incompatible subpopulations and a single unisexual class, the unisexual individuals can exploit the diversity generated by mating types, obtaining a higher rate of effectual sex than in their monomorphic populations. While this leads to unisexuals having a higher probability of effectual sex than their self-incompatible mating type competitors when rare, this advantage is frequency

dependent; as the frequency of unisexuals increases, their rate of effectual sex dips back below that of their self-incompatible competitors. While the precise frequency at which this transition occurs depends in a complex way on the model parameters, our intuition from the case of monomorphic populations carries over: the advantage of self-incompatibility over unisexuality diminishes as the number of mating types decreases, and the number of sexual events per generation σ and costs to mating finding increase.

In regard to when we would expect self-incompatible classes to be selected for we found that initially (when rare) a mutant self-incompatible type would only be weakly selected against but that in order to drift to more substantial frequencies, the costs of mate finding would have to be sufficiently low. If the self-incompatible mating types can be maintained in the population for a sufficient amount of time at low-frequencies, they can obtain a higher effective rate of sex to their unisexual ancestors. If genetic recombination with non-clone mates is sufficiently advantageous, this could generate selection for mating types.

Answering the question of the number of maintained mating types, we found that new mating types experience a selective advantage both in terms of increased mating opportunities relative to resident population and a (shared) decrease in the probability of mating with a genetically distinct individual. There is then selection for an ever greater number of mating types, which are only kept bounded by genetic drift.

BIBLIOGRAPHY

[1] A. F. Agrawal. Evolution of sex: Why do organisms shuffle their genotypes? *Curr. Biol.*, 16(7):R696–R704, 2006.

[2] D. Aldous. Stopping times and tightness. *Ann. Probability*, 6(2):335–340, 1978.

[3] N. Barton, D. Briggs, J. Eisen, D. Goldstein, and N. Patel. *Evolution*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2007.

[4] N. Barton, A. Etheridge, A. Véber, et al. A new model for evolution in a spatial continuum. *Electronic Journal of Probability*, 15:162–216, 2010.

[5] N. H. Barton and A. M. Etheridge. The relation between reproductive value and genetic contribution. *Genetics*, 188(4):953–973, 2011.

[6] N. H. Barton, A. M. Etheridge, and A. Véber. Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01): P01002, 2013.

[7] R. F. Bass and Z.-Q. Chen. Brownian motion with singular drift. *Ann. Probab.*, 31 (2):791–817, 2003.

[8] A. Benassi and J.-P. Fouque. Hydrodynamical limit for the asymmetric simple exclusion process. *Ann. Probab.*, 15(2):546–560, 1987.

[9] S. Billiard, M. López-Villavicencio, B. Devier, M. E. Hood, C. Fairhead, and T. Giraud. Having sex, yes, but with whom? inferences from fungi on the evolution of anisogamy and mating types. *Biological Reviews*, 86(2):421–442, 2011.

[10] S. Billiard, M. López-Villavicencio, M. Hood, and T. Giraud. Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *J. Evol. Biol.*, 25:1020–1038, 2012.

[11] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[12] N. Biswas, A. Etheridge, and A. Klimek. The spatial lambda-fleming-viot process with fluctuating selection. *arXiv preprint arXiv:1802.08188*, 2018.

[13] J. Blath, A. Etheridge, and M. Meredith. Coexistence in locally regulated competing populations and survival of branching annihilating random walk. *Ann. Appl. Probab.*, 17(5-6):1474–1507, 2007.

[14] W. F. Bodmer and L. L. Cavalli-Sforza. A migration matrix model for the study of random genetic drift. *Genetics*, 59(4):565, 1968.

[15] A. Böttcher and B. Silbermann. *Introduction to large truncated Toeplitz matrices*. Springer Science & Business Media, 2012.

[16] G. S. Bradburd and P. L. Ralph. Spatial population genetics: it's about time. *Annual Review of Ecology, Evolution, and Systematics*, 2019.

[17] S. Branco, H. Badouin, R. C. R. de la Vega, J. Gouzy, F. Carpentier, G. Aguileta, S. Siguenza, J.-T. Brandenburg, M. A. Coelho, M. E. Hood, and T. Giraud. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc. Natl. Acad. Sci.*, 114(27):7067–7072, 2017.

[18] D. L. Burkholder. Distribution function inequalities for martingales. *Ann. Probability*, 1:19–42, 1973.

[19] H. Cayuela, Q. Rougemont, J. G. Prunier, J.-S. Moore, J. Clobert, A. Besnard, and L. Bernatchez. Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Molecular Ecology*, 27(20):3976–4010, 2018.

[20] D. Charlesworth and B. Charlesworth. The evolution and breakdownof s-allele systems. *Heredity*, 43:41–55, 1979.

[21] A. G. Clark. Population genetic aspects of gametophytic self incompatibility. *Plant Species Biology*, 11(1):13–22, 1996.

[22] M. A. Coelho, G. Bakkeren, S. Sun, M. E. Hood, and T. Giraud. Fungal sex: The basidiomycota. *Microbiology spectrum*, 5:0046–2016, 2017.

[23] G. W. A. Constable and H. Kokko. The rate of facultative sex governs the number of expected mating types in isogamous species. *Nat. Ecol. Evol.*, 2(7):1168–1175, 2018.

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

[25] J. F. Crow and M. Kimura. *An introduction to population genetics theory.* New York, Evanston and London: Harper & Row, Publishers, 1970.

[26] T. L. Czaran and R. F. Hoekstra. Evolution of asexual asymmetry. *BMC Evolutionary Biology*, 21:4–34, 2004.

[27] P. Czuppon and G. W. A. Constable. Invasion and extinction dynamics of mating types under facultative sexual reproduction. *Genetics*, 213(2):567–580, 2019.

[28] P. Czuppon and D. Rogers. Evolution of mating types in finite populations: the precarious advantage of being rare. *Journal of Evolutionary Biology*, 32(11):1290–1299, 2019.

[29] G. Da Prato and J. Zabczyck. *Stochastic equations in infinite dimensions*. Encyclopaedia of Mathematics and its Applications. Cambridge University Press, 1992.

[30] C. Darwin. *On the origin of species by means of natural selection*. John Murray, London, 1859.

[31] R. Durrett. *Stochastic Calculus: A Practical Introduction*, volume 6. CRC Press, 1996.

[32] R. Durrett. *Probability models for DNA sequence evolution*. Springer Science & Business Media, 2008.

[33] R. Durrett and W.-T. Fan. Genealogies in expanding populations. *Ann. Appl. Probab.*, 26(6):3456–3490, 2016.

[34] P. Embrechts and M. Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013.

[35] K.-J. Engel and R. Nagel. *One-parameter semigroups for linear evolution equations*. Springer, New York, NY, 2000.

[36] R. A. Ennos and X. S. Hu. Estimating the number of sexual events per generation in a facultatively sexual haploid population. *Heredity*, 2018.

[37] A. Etheridge. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009*, volume 2012. Springer Science & Business Media, 2011.

[38] A. Etheridge, T. Kurtz, P. Ralph, and A. Smith. A class of nonlinear superprocesses. In preperation.

[39] A. Etheridge, N. Freeman, D. Straulino, et al. The brownian net and selection in the spatial λ-fleming-viot process. *Electronic Journal of Probability*, 22, 2017.

[40] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons Inc, 1986.

[41] W. J. Ewens. *Mathematical population genetics 1: theoretical introduction*, volume 27. Springer Science & Business Media, 2012.

[42] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[43] J. Felsenstein. A pain in the torus: some difficulties with models of isolation by distance. *The American Naturalist*, 109(967):359–368, 1975.

[44] R. Forien. Isolation by distance patterns arising from short range and long range dispersal–a forwards in time approach. *arXiv preprint arXiv:1907.07930*, 2019.

[45] R. Forien et al. The stepping stone model in a random environment and the effect of local heterogneities on isolation by distance patterns. *Electronic Journal of Probability*, 24, 2019.

[46] J. A. Fraser, S. Diezmann, R. L. Subaran, A. Allen, K. B. Lengeler, F. S. Dietrich, and J. Heitman. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLOS Biology*, 2(12), 11 2004.

[47] N. Gerber and H. Kokko. Abandoning the ship using sex, dispersal or dormancy: multiple escape routes from challenging conditions. *Phil. Trans. R. Soc. B*, 373: 20170424, 2018.

[48] G. Grimmett. *The random-cluster model*, volume 333 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006.

[49] S. Grosskinsky. Interacting particle systems. *Notes*, 2009.

[50] M. Z. Guo, G. C. Papanicolaou, and S. R. S. Varadhan. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Comm. Math. Phys.*, 118(1):31–59, 1988.

[51] E. Haag. Why two sexes? sex determination in multicellular organisms and protistan mating types. *Seminars in Cell & Developmental Biology*, 18:348–349, 2007.

[52] Z. Hadjivasiliou and A. Pomiankowski. Evolution of the mating type locus with suppressed recombination. *bioRxiv*, page 518654, 2019.

[53] Z. Hadjivasiliou, A. Pomiankowski, R. Seymour, and N. Lane. Selection for mitonuclear co-adaptation could favour the evolution of two sexes. *Proceedings of the Royal Society B: Biological Sciences*, 279:1865–1872, 2012.

[54] Z. Hadjivasiliou, Y. Iwasa, and A. Pomiankowski. Cell-cell signalling in sexual chemotaxis: a basis for gametic differentiation, mating types and sexes. *J. R. Soc. Interface*, 12:20150342, 2015.

[55] B. C. Haller and P. W. Messer. Slim 3: Forward genetic simulations beyond the wright–fisher model. *Molecular biology and evolution*, 36(3):632–637, 2019.

[56] B. C. Haller, J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph. Tree-sequence recording in slim opens new horizons for forward-time simulation of whole genomes. *Molecular ecology resources*, 19(2):552–566, 2019.

[57] H. Hanche-Olsen and H. Holden. The Kolmogorov-Riesz compactness theorem. *Expositiones Mathematicae*, 28(4):385–394, 2010.

[58] M. Hartfield and P. D. Keightley. Current hypotheses for the evolution of sex and recombination. *Integr. Zool.*, 7(2):192–209, 2012.

[59] A. R. Hasan and R. W. Ness. The genomic landscape of recombination rate variation in chlamydomonas reinhardtii reveals effects of linked selection. *bioRxiv*, (https://doi.org/10.1101/340992), 2018.

[60] J. Heitman. Evolution of sexual reproduction: A view from the fungal kingdom supports an evolutionary epoch with sex before sexes. *Fungal Biology Reviews*, 29 (3):108 – 117, 2015. Special Issue: Fungal sex and mushrooms – A credit to Lorna Casselton.

[61] R. F. Hoekstra. On the asymmetry of sex: Evolution of mating types in isogamous populations. *J. of Theor. Biol.*, 98(3):427–451, 1982.

[62] R. F. Hoekstra. The evolution of male-female dimorphism: older than sex? *Journal of Genetics*, 69(1):11–15, 1990.

[63] L. D. Hurst and W. D. Hamilton. Cytoplasmic fusion and the nature of sexes. *Proc. Roy. Soc. B*, 247:189–194, 1992.

[64] D. L. Iglehart. Weak convergence in queueing theory. *Advances in Applied Probability*, pages 570–594, 1973.

[65] D. L. Iglehart and W. Whitt. Multiple channel queues in heavy traffic. i. *Advances in Applied Probability*, 2(1):150–177, 1970.

[66] Y. Iwasa and A. Sasaki. Evolution of the number of sexes. *Evolution*, 41(1):49–65, 1987.

[67] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1987.

[68] T. James. Analysis of mating type locus organization and synteny in mushroom fungi: Beyond model species. In J. Heitman, J. W. Kronstad, J. W. Taylor, and L. A. Casselton, editors, *Sex in Fungi: Molecular Determination and Evolutionary Implications*, pages 317–331. ASM Press, Washington DC, 2007.

[69] A. Jnowska-Michalik. On processes of ornstein-uhlenbeck type in hilbert space. *Stochastics: An International Journal of Probability and Stochastic Processes*, 21 (3):251–286, 1987.

[70] O. Kallenberg. *Random measures, theory and applications*. Springer, 2017.

[71] S. Karlin and S. Lessard. On the optimal sex ratio. *Proceedings of the National Academy of Sciences*, 80(19):5931–5935, 1983.

[72] S. Karlin and S. Lessard. *Theoretical studies on sex ratio evolution*. 1986.

[73] S. Karlin and H. E. Taylor. *A second course in stochastic processes*. Elsevier, 1981.

[74] J. Kelleher, N. H. Barton, and A. M. Etheridge. Coalescent simulation in continuous space. *Bioinformatics*, 29(7):955–956, 2013.

[75] J. Kelleher, A. Etheridge, A. Véber, and N. H. Barton. Spread of pedigree versus genetic ancestry in spatially distributed populations. *Theoretical population biology*, 108:1–12, 2016.

[76] J. Kelleher, K. R. Thornton, J. Ashander, and P. L. Ralph. Efficient pedigree recording for fast population genetics simulation. *PLoS computational biology*, 14(11): e1006581, 2018.

[77] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1985.

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

[79] M. Kimura and G. H. Weiss. A mathematical analysis of the stepping stone model of genetic correlation. *Journal of Applied Probability*, pages 129–149, 1965.

[80] S. Kliem. Convergence of rescaled competing species processes to a class of SPDEs. *Electron. J. Probab.*, 16:no. 22, 618–657, 2011.

[81] H. Kokko. When synchrony makes the best of both worlds even better: How well do we really understand facultative sex? *The American Naturalist*, 195(2):380–392, 2020. PMID: 32017623.

[82] E. Kothe. Mating types and pheromone recognition in the homobasidiomycete schizophyllum commune. *Fungal Genetics and Biology*, 27:146–152, 1999.

[83] Y. Krumbeck, G. W. A. Constable, and T. Rogers. Fitness differences suppress the number of mating types in evolving isogamous species. *Royal Society Open Science*, 7:192126, 2020.

[84] T. G. Kurtz and E. R. Rodrigues. Poisson representations of branching markov and measure-valued branching processes. *The Annals of Probability*, 39(3):939–984, 2011.

[85] S. C. Lee, M. Ni, W. Li, C. Shertz, and J. Heitman. The evolution of sex: a perspective from the fungal kingdom. *Microbiol. Mol. Biol. Rev.*, 74(2):298–340, 2010.

[86] J. Lehtonen and H. Kokko. Two roads to two sexes: unifying gamete competition and gamete limitation in a single model of anisogamy evolution. *Behav. Ecol. Sociobiol.*, 65:445–459, 2011.

[87] J. Lehtonen, M. Jennions, and H. Kokko. The many costs of sex. *Trends Ecol. Evol.*, 27:172–178, 2012.

[88] J. Lehtonen, H. Kokko, and G. A. Parker. What do isogamous organisms teach us about sex and the two sexes? *Phil. Trans. R. Soc. B*, 371:20150532, 2016.

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

[90] T. M. Liggett. *Interacting particle systems*, volume 276. Springer Science & Business Media, 2012.

[91] G. Lorden. On excess over the boundary. *The Annals of Mathematical Statistics*, pages 520–527, 1970.

[92] G. Malécot. Mathématiques de l'hérédité. 1948.

[93] G. Malécot. Heterozygosity and relationship in regularly subdivided populations. *Theoretical population biology*, 8(2):212–241, 1975.

[94] D. R. McDonald. On local limit theorem for integer valued random variables. *Teor. Veroyatnost. i Primenen.*, 24(3):607–614, 1979.

[95] G. Mendel. Versuche über pflanzen-hybriden. *Theor Appl Genet*, 13:221–268, 1941.

[96] M. Meredith. *Asymptotic results for some stochastic population models*. PhD thesis, University of Oxford, 2006.

[97] M. Morgan-Richards, S. S. Langton-Myers, and S. A. Trewick. Loss and gain of sexual reproduction in the same stick insect. *Molecular ecology*, 28(17):3929–3941, 2019.

[98] C. Müller and R. Tribe. Stochastic p.d.e.'s arising from the long range contact and long range voter processes. *Probab. Theory Related Fields*, 102(4):519–545, 1995.

[99] G. Nappo and B. Torti. Continuous time random walks and queues: explicit forms and approximations of the conditional law with respect to local times. *Stochastic processes and their applications*, 116(4):585–610, 2006.

[100] M. A. Naranjo-Ortiz and T. Gabaldón. Fungal evolution: major ecological adaptations and evolutionary transitions. *Biological Reviews*, 94(4):1443–1476, 2019.

[101] M. J. Nauta and R. F. Hoekstra. Evolution of reproductive systems in filamentous ascomycetes. i. evolution of mating types. *Heredity*, 68:405–410, 1992.

[102] M. Nei. Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences*, 70(12):3321–3323, 1973.

[103] R. W. Ness, A. D. Morgan, N. Colegrave, and P. D. Keightley. Estimate of the spontaneous mutation rate in chlamydomonas reinhardtii. 192:1447–1454, 2012.

[104] B. Nieuwenhuis and S. Immler. The evolution of mating-type switching for reproductive assurance. *Bioessays*, 38:1141–1149, 2016.

[105] B. Nieuwenhuis, S. Tusso, P. Bjerling, J. Stå ngberg, J. Wolf, and S. Immler. Repeated evolution of self-compatibility for reproductive assurance. *Nat. Commun.*, 24 (9):1639, 2018.

[106] M. Nordborg and S. M. Krone. Separation of time scales and convergence to the coalescent in structured populations. *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot*, pages 194–232, 2002.

[107] M. F. Norman. Approximation of stochastic processes by Gaussian diffusions, and applications to Wright-Fisher genetic models. *SIAM J. Appl. Math.*, 29(2):225–242, 1975. Special issue on mathematics and the social and biological sciences.

[108] E. Perkins. Dawson-Watanabe superprocesses and measure-valued diffusions. In *Lectures on probability theory and statistics (Saint-Flour, 1999)*, volume 1781 of *Lecture Notes in Math.*, pages 125–324. Springer, Berlin, 2002.

[109] S. S. Phadke and R. A. Zufall. Rapid diversification of mating systems in ciliates. *Biological Journal of the Linnean Society*, 98:187–197, 2009.

[110] A. Pilipenko. *An introduction to stochastic differential equations with reflection*, volume 1. Universitätsverlag Potsdam, 2014.

[111] K. C. Roach, M. Feretzaki, S. Sun, and J. Heitman. Unisexual reproduction. volume 85 of *Advances in Genetics*, pages 255 – 305. Academic Press, 2014.

[112] S. J. Schreiber. Interactive effects of temporal correlations, spatial heterogeneity and dispersal on population persistence. *Proceedings of the Royal Society B: Biological Sciences*, 277(1689):1907–1914, 2010.

[113] H. Sekimoto, J. Abe, and Y. Tsuchikane. New insights into the regulation of sexual reproduction in closterium. volume 297 of *International Review of Cell and Molecular Biology*, pages 309 – 338. Academic Press, 2012.

[114] T. Shiga. Stepping stone models in population genetics and population dynamics. In *Stochastic processes in physics and engineering*, pages 345–355. Springer, 1988.

[115] T. Shiga and A. Shimizu. Infinite-dimensional stochastic differential equations and their applications. *J. Math. Kyoto Univ.*, 20(3):395–416, 1980.

[116] A. V. Skorokhod. Stochastic equations for diffusion processes in a bounded region i. *Theory of Probability & Its Applications*, 6(3):264–274, 1961.

[117] A. V. Skorokhod. Stochastic equations for diffusion processes in a bounded region. ii. *Theory of Probability & Its Applications*, 7(1):3–23, 1962.

[118] M. Slatkin. Gene flow in natural populations. *Annual review of ecology and systematics*, 16(1):393–430, 1985.

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

[120] A. Sturm. On convergence of population processes in random environments to the stochastic heat equation with colored noise. *Electron. J. Probab.*, 8:no. 6, 39, 2003.

[121] A. K. Sturm. *On spatially structured population processes and relations to stochastic partial differential equations*. PhD thesis, University of Oxford, 2002.

[122] S. Sun and J. Heitman. Is sex necessary? *BMC biology*, 9(1):1–4, 2011.

[123] A.-S. Sznitman. *Brownian motion, obstacles and random media*. Springer Science & Business Media, 1998.

[124] H. Tanaka. Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Math. J.*, 9(1):163–177, 1979.

[125] O. Toeplitz. Zur theorie der quadratischen und bilinearen formen von unendlichvielen veränderlichen. *Mathematische Annalen*, 70(3):351–376, 1911.

[126] T. Togashi, J. Bartelt, J. Yoshimura, K. Tainaka, and P. Cox. Evolutionary trajectories explain the diversified evolution of isogamy and anisogamy in marine green algae. *Proc. Natl. Acad. Sci.*, 109:13692–13697, 2012.

[127] I. J. Tsai, D. Bensasson, A. Burt, and V. Koufopanou. Population genomics of the wild yeast saccharomyces paradoxus: Quantifying the life cycle. *Proc. Natl. Acad. Sci.*, 105(12):4957–4962, 2008.

[128] M. K. Uyenoyama. On the evolution of genetic incompatibility systems. ii. initial increase of strong gametophytic self-incompatibility under partial selfing and half-sib mating. *Am. Nat.*, 131:700–722, 1988.

[129] M. K. Uyenoyama. On the evolution of genetic incompatibility systems. iii. introduction of weak gametophytic selfincompatibility under partial inbreeding. *Theor. Popul. Biol.*, 34:47–91, 1988.

[130] X. Vekemans and M. Slatkin. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics*, 137(4):1157–1165, 1994.

[131] J. Wakeley. The Coalescent in an Island Model of Population Subdivision with Variation among Demes. Technical report, 2001.

[132] J. Wakeley. *Coalescent theory: an introduction*. Number 575: 519.2 WAK. 2009.

[133] J. Wakeley and N. Aliacar. Gene genealogies in a metapopulation. *Genetics*, 159(2): 893–905, 2001.

[134] J. B. Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 265–439. Springer, Berlin, 1986.

[135] I. J. Wang and G. S. Bradburd. Isolation by environment. *Molecular ecology*, 23 (23):5649–5662, 2014.

[136] J. F. Wilkins and J. Wakeley. The coalescent in a continuous, finite, linear population. *Genetics*, 161(2):873–888, 2002.

[137] D. Williams. *Probability with martingales*. Cambridge university press, 1991.

[138] A. M. Wilson, P. M. Wilken, M. A. van der Nest, E. T. Steenkamp, M. J. Wingfield, and B. D. Wingfield. Homothallism: an umbrella term for describing diverse sexual behaviours. *IMA Fungus*, 6(1):207–214, 2015.

[139] S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97, 1931.

[140] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the sixth international congress of Genetics*, volume 1, pages 356–366, 1932.

[141] S. Wright. Isolation by distance. *Genetics*, 28(2):114, 1943.

[142] I. Zähle, J. T. Cox, R. Durrett, et al. The stepping stone model. ii: Genealogies and the infinite sites model. *The Annals of Applied Probability*, 15(1B):671–699, 2005.