STUDIES OF EVOLUTION IN POPULATIONS WITH LONG-RANGE DISPERSAL

by

NATHAN VILLIGER

A DISSERTATION

Presented to the Department of Physics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2023

DISSERTATION APPROVAL PAGE

Student: Nathan Villiger

Title: Studies of Evolution in Populations with Long-range Dispersal

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Raghuveer Parthasarathy	Chair
Jayson Paulose	Advisor
Ben Farr	Core Member
Peter Ralph	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2023

DISSERTATION ABSTRACT

Nathan Villiger

Doctor of Philosophy

Department of Physics

June 2023

Title: Studies of Evolution in Populations with Long-range Dispersal

Long-range dispersal of offspring is ubiquitous in nature, from seeds that disperse random distances thanks to being carried by animals, to pollen that gets carried long distances by the wind, and even viruses that spread around the world with the help of infected travelers on intercontinental airplane journeys. Long-range dispersal can lead to founder events throughout a landscape, as the first individual to colonize a new region benefits from abundant resources and a lack of competition, which can result in that individual's genes making a disproportionately large contribution to future generations near the territory it colonized. Long-range dispersal can drive range expansions when individuals disperse beyond the bounds of the population's current range. Range expansions driven by long-range dispersal can have dramatic consequences, for example as invasive species take over habitats with no ecological architecture to keep them in check or pandemics rapidly spread around the world.

Range expansions driven by long-range dispersal accelerate as they progress and have remarkably different dynamics than the constant-speed expansions carried out by populations with exclusively short-range dispersal. These jump-driven expansions can be challenging to model in part because the dynamics are dominated by the rare longest dispersal events. Recent theoretical advances have enabled predictions about such quantities as population growth rates and the evolution of neutral diversity during range expansions driven by power law dispersal kernels. However, these theories rely on various simplifying assumptions which are not always met by natural populations, and

their applicability to more complex but realistic population dynamics remains an open question. Another open question is how to connect theoretical results with real-world biological populations. This dissertation addresses these open questions by developing methods of simulating range expansions with more realistic population dynamics and extracting dispersal parameters from genomic data.

In Chapter II, we use simulations to explore the consequences of departing from assumptions of the simplified models that led to the aforementioned predictions about population growth and the evolution of neutral diversity. We show that qualitative trends are preserved but reveal quantitative signals of the more realistic local dynamics. In Chapter III, we use simulations to investigate what determines the fate of fitness-affecting mutations that appear during range expansions driven by long-range dispersal, a situation for which there is no existing theory. We find that mutation outcomes are independent of the fitness effect they confer across a wide range of effect sizes. In Chapter IV, we show that convolutional neural networks can learn dispersal parameters from genomic samples taken from individuals in populations with long-range dispersal, bringing the growing body of theoretical work in this field closer to samples that could be taken from actual biological populations.

This dissertation contains previously published and unpublished coauthored material.

CURRICULUM VITAE

NAME OF AUTHOR: Nathan Villiger

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA
Butler University, Indianapolis, IN, USA

DEGREES AWARDED:

Doctor of Philosophy, Physics, 2023, University of Oregon
Bachelor of Science, Physics and Astronomy, 2017, Butler University

AREAS OF SPECIAL INTEREST:

Evolutionary dynamics in spatially structured populations
Machine learning

PROFESSIONAL EXPERIENCE:

Research Assistant, University of Oregon, 2019-2023
Teaching Assistant, University of Oregon, 2018-2023
Undergraduate Research Assistant, Butler University, 2015-2017
Tour Guide, Holcomb Observatory and Planetarium, 2014-2017

PUBLICATIONS:

N. Villiger & J. Paulose (2023). The influence of explicit local dynamics on range expansions driven by long-range dispersal. *G3 Genes |Genomes|Genetics*.

ACKNOWLEDGEMENTS

Thanks to my wife, Emily, for agreeing to move across the country and go on this adventure together, and for later doubling down and agreeing to get married and do life together. It's been so fun moving out west, starting our new lives, and taking in the sweetest cat, and I'm excited to see what else is in store for us. Thank you to the entire Trygstad family for being such a fun, loving, and welcoming second family.

Finally, I'd like to thank all the friends who helped make this happen. I'm grateful to more people than I could possibly list individually, but I'd especially like to thank a few people who have made my time in Eugene so special: Becca and David, for always being great friends and adventure buddies; JD, for the bike rides and ping pong games during our first year and the fun outings throughout our time here, your willingness to teach and share experiences has changed my life; and Merilea, for being a great friend who made a challenging time significantly better, I'm so glad we happened to meet again.

TABLE OF CONTENTS

LIST OF FIGURES

xiii

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 Introduction

Long-range dispersal of offspring is ubiquitous in nature and can take many
forms [1]. Plants often release their seeds and pollen to be carried random distances by
the wind [2], such as maple trees which encapsulate their seeds in helicopters that flutter
away to hopefully land in a fertile sunny place. Animals mediate long-range dispersal
of other organisms often by feeding, migration, or some combination of the two [3].
Examples include blackberries, an invasive plant around Eugene whose seeds are spread
by animals eating the berries and excreting the seeds elsewhere [4, 5], and glacier ice
worms, which have been shown to disperse hundreds of miles between glaciated peaks
of the Pacific Northwest likely as the result of hitchhiking with migratory birds [6]. Of
course, humans play a role in long-range dispersal of other organisms, too. A relatively
benign example is human-mediated dispersal of plant seeds e.g. by seeds sticking to
shoes or clothes [7]; a deadly example is the spread of the Covid-19 pandemic, which
was enabled and accelerated by humans carrying the virus on our airplane journeys all
over the world [8].

Long-range dispersal can drive range expansions when individuals are transported
to new territories beyond the bounds of the population's current range. Range expansions
can proceed at a literally glacial pace as glaciers retreat and open up new habitats (e.g.
for Pacific salmon [9]), or they can be very fast as invasive species rapidly colonize
habitats without any ecological architecture to keep them in check (e.g. cane toads in
Australia [10]). Range expansions are increasingly forced by global warming [11] as the
changing climate makes species' traditional homelands inhospitable while opening up new
hospitable places, such as in the ecological void left behind by retreating glaciers [9].

Range expansions leave behind genetic signatures in the population that look like
local adaptation but actually arise solely due to the spatial/expansion dynamics [12].
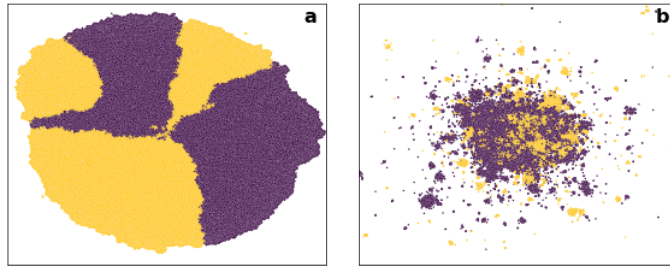
Figure 1. Spatial genetic patterns left behind by simulated neutral expansions in a 2D habitat. Simulations began with 100 individuals tightly packed around the origin and grew to population sizes of one million individuals. Simulations began with a 50/50 mix of two neutral alleles (colors). **a.)** Diffusive expansion. Gene surfing around the perimeter leads to large monoallelic sectors. **b.)** Expansion with long-range dispersal. Sectors are broken up into small speckles of one allele or the other by long-range dispersal across the landscape.

In the absence of long-range dispersal, random genetic drift at the expansion frontier can cause an allele to "surf" the expansion wave [13, 14] and even reach fixation at the front of 1D expansions, guaranteeing that only that allele will be present in individuals who continue the expansion into new territory. In 2D expansions, such as those which start from a droplet at the center of a Petri dish, surfing events around the perimeter of the growing colony lead to the formation of large monoallelic sectors that radiate outwards from the center and persist indefinitely [15, 16, 17, 18], as shown in Fig. 1**a**. The permanent nature of fixation or sectors for 1D or 2D expansions, respectively, is the result of offspring only being able to disperse short distances from their parents; after the initial homeland has saturated, the only place with available resources is at the edge of the colony, so offspring born to parents at the edge of the colony are the only ones fortunate enough to survive and contribute future offspring to future generations. The large regions of decreased genetic diversity after expansions give the impression of local adaptation, as if the allele that formed a sector was uniquely suited to take advantage of unique conditions within that portion of the Petri dish, but they form simply as a result of drift at the edge of the expanding population.

Long-range dispersal provides a mechanism by which individuals deep within the core of expanding populations can still transmit offspring to vacant parts of the landscape and thus play a role in contributing their genes to future generations. Such long dispersal events break up the large sectors typical of diffusive expansions into much smaller regions of space dominated by one allele that are typically seeded by a long-range dispersal event; see the speckles of one color or the other in Fig. 1**b**. Range expansions driven by long-range dispersal accelerate as they progress, leading to much faster growth than the constant-speed expansions driven by short-range dispersal [19]. These "jump-driven" expansions can be challenging to model because it's often the rare longest dispersal events that drive the dynamics. A key ingredient of any model of a population in a spatial landscape is the jump kernel, the distribution of distances from which offspring dispersal distances are randomly drawn. Long-range dispersal is the result of populations dispersing according to a fat-tailed jump kernel that decays slower than exponentially with increasing distance. Many species disperse according to fat-tailed jump kernels [20].

There is a growing body of research into populations with power law jump kernels that decay as $J(r) \sim 1/r^{\mu+1}$ at long distances. Power law kernels are a good description of the dispersal of many plant species [20] and are convenient theoretically because one can continuously traverse from populations that grow diffusively all the way to the well-mixed limit by adjusting one parameter, the kernel exponent μ. Narrow jump kernels with $\mu > d + 1$, where d is the dimensionality of space, lead to diffusive growth as if there were only short-range dispersal, while $\mu \to 0$ creates jump kernels so broad that there is boundless dispersal across the landscape; population growth approaches the exponential growth seen in well-mixed environments [21] (μ must be greater than zero to ensure a normalizable jump kernel). We next briefly describe how we build on previous work to learn new information about evolution in populations with power law dispersal.

1.2 Simulating expansions with explicit local dynamics

Ref. [21] used a lattice-based model to derive the analytical forms of how colony growth depends on the kernel exponent. Refs. [22] and [23] built on that work by studying how soft sweep patterns and the evolution of neutral diversity, respectively, depend on the kernel exponent. All three of these studies used lattice-based models which divided space into a grid of habitats called "demes" and made simplifying assumptions about spatial interactions like local saturation of demes and migration between demes.

There is a long history of researchers using lattice-based models to study evolution in spatial landscapes, going all the way back to the early years of the field (e.g. Ref. [24]). Lattice models are a useful tool because they allow researchers to write down relationships for quantities like migration between demes or deme-wide mutation rates while also making simplifying assumptions (e.g. random mating within demes or a local founder takes all effect) that allow them to focus on macroscopic population dynamics. Lattice-based models are also common in computational studies of evolution; see e.g. Refs. [21, 22, 23, 25, 26, 27, 28, 29, 30, 31]. Obviously, nature is not subdivided into neat grids of habitats, so one may be tempted to increase the spatial resolution of their lattice to make their model better approximate a continuum; however, doing so can sometimes make models *less* able to capture the true population dynamics [32]. In Chapter II, we leverage recent computational advances [33] to simulate range expansions driven by power law dispersal on a continuous landscape and explore how the predicted colony growth and evolution of neutral diversity stand up to realistic simulations that break assumptions of the simplified models. That work was co-authored by Jayson Paulose and has been published in the journal *G3* [34].

1.3 The fate of adaptive mutations during range expansions

The Covid pandemic was a shocking example of long-range dispersal accelerating a range expansion as infected passengers carried the virus on airplane journeys all over the world [8]. Early on, most people who became infected contracted the original

strain of the virus; however, multiple variants appeared later and sequentially overtook their predecessors to become the prominent strain of the virus worldwide. The virus accumulated adaptive mutations that enabled it to infect and/or spread more easily as it was undergoing a jump-driven range expansion.

While the outcomes for adaptive mutations have been studied in diffusive [35] and well-mixed [36] expansions, little is known about the fate of such mutations that appear during range expansions driven by fat-tailed jump kernels [37]. In Chapter III, we use extensive simulations to study what determines the outcomes of fitness-affecting mutations that appear during range expansions driven by power law jump kernels; that work was co-authored by Jayson Paulose.

1.4   Estimating dispersal parameters using neural networks

Machine learning touches many aspects of our everyday lives, performing varied tasks like automated resume screening of job applicants or suggesting which shows we should watch next on Netflix. Neural networks are a type of machine learning model and are useful in part because they can learn from large numbers input features to predict quantities of interest, such as using the pixel values from an image to classify what is shown in the photo. They are increasingly used in population genetics [38, 39, 40, 41, 42, 43, 44] because they can make inferences from raw genomic data, which can contain thousands of measurements per individual, thus harnessing all available information without requiring that the data be compressed into summary statistics before being fed into a model. In Chapter IV, we present preliminary results using convolutional neural networks to estimate dispersal parameters from genomic samples of populations with long-range dispersal; this work was done in collaboration with Chris Smith.

1.5   Continuous-space simulations

A key method throughout this dissertation is continuous-space simulations using the forwards-time, agent-based simulation software SLiM [33]. It is forwards-time in the

sense that simulation time moves forwards and the final state is unknown at runtime, contrasting with the also common backwards in time simulations that begin with the evolved population and simulate an ancestry that could have generated the population in its current state (see e.g. the simulation software msprime [45]). SLiM is agent-based because it records and manages everything at the level of individuals, without resorting to any deme-based simplifications. It efficiently handles all biology (reproduction, mutation, recombination, etc.) and spatial interactions (mate choice, competition, dispersal, etc.) for all individuals on a continuous landscape, both of which make it an invaluable tool in the coming chapters. Detailed simulation protocols will be discussed in the following sections: Section 2.3, Section 2.6.1, Section 3.2.1, Section 3.5.1, and Section 4.3.

CHAPTER II

THE INFLUENCE OF EXPLICIT LOCAL DYNAMICS ON RANGE EXPANSIONS

DRIVEN BY LONG-RANGE DISPERSAL

This work was published as Villiger and Paulose (2023) [34].

2.1 Introduction

Range expansion—the act of a population expanding into new territory—is common in biological populations. Range expansions occur naturally and randomly all the time, often as the result of a species' natural movement, such as by animals moving into new territory or maple helicopters carrying seeds away from their parent tree. Researchers have documented range expansions in a wide variety of organisms, such as plants [46], birds [47], sea creatures [48, 49], and terrestrial animals [10, 50], even humans [51]. Range expansions are increasingly forced by global warming as the changing climate makes traditional habitats inhospitable, while potentially opening up new hospitable regions [11].

Range expansions leave distinctive signatures in the patterns of genetic diversity of a population that can mimic the effects of natural selection [12]. Individuals at the frontier of an expanding population make a large contribution to the subsequent expansion wave, even if their frontier position was solely due to chance; as a result, genetic variants they carry can acquire high frequencies in the population in a phenomenon termed gene surfing [13, 14]. Independent surfing events in separate sections of the expansion front cause the population to segregate into genetically distinct sectors, promoting an illusion of local adaptation from purely neutral mutations [15, 17, 16, 18]. Modeling the combined effect of spatial structure and stochasticity on neutral genetic diversity is key to understanding the biological origins of established genetic patterns, and to the successful prediction of future genetic diversity in pandemics and ecological expansions.

The influence of random chance on genetic diversity during range expansions can be amplified by long-range dispersal [21]. Many species have evolved ingenious ways of dispersing offspring over long distances with help from natural forces and from other organisms [1]. Plants rely on the dispersal of seeds and pollen by wind, waves, and animals [2]. Glacier ice worms can travel hundreds of miles, likely carried by migratory birds [6]. Modern pandemics are driven by microorganisms hitchhiking on air travelers to find new uninfected populations [8]. Even if long-range dispersal events are rare, they have an outsized influence on the expansion because they enable pioneers to seed satellite colonies in uninhabited areas. If a pioneer happens to land in a place with abundant resources and little to no competition, its descendants may flourish. The pioneer's genes will then propagate and any genetic variants they carry will reach high frequencies in the vicinity of the satellite [25, 26, 27, 28, 29] even in the absence of a selective advantage; random chance alone has caused the pioneer's genes to become prominent by means of a founder effect, leading to a *suppression* of local diversity within satellites. However, long-range dispersal also *favors* neutral diversity at larger scales, by ensuring that individuals well within the expanding population have a chance of contributing to growth. The evolution of overall diversity during the range expansion is governed by the trade-off between the two effects, and can depend sensitively on the degree of long-range dispersal experienced by the population [52, 23].

Modeling the general characteristics of range expansions requires two minimal ingredients: a probability distribution of dispersal distances $J(r)$, also called the *jump kernel*, from which dispersal events are randomly drawn; and a method of local density regulation to model the existence of a finite carrying capacity. When long-range jumps are present, the tail of the jump kernel, i.e. its behavior at long distances, critically influences the fate of the population at long times. Fundamental differences from short-range dispersal are observed when the jump kernel is "fat-tailed"; i.e. it decays slower than exponentially with increasing distance. Fat-tailed jump kernels lead to expansions

8

that accelerate as they progress, unlike the constant-speed expansions that occur when dispersal is exclusively short-range [19].

A commonly used fat-tailed kernel is the power-law jump kernel $J(r) \sim 1/r^{\mu+1}$. Besides providing a good description of the dispersal behavior of many species [20], power-law kernels are a useful tool for analyzing and classifying the breadth of potential population outcomes due to long-range dispersal [21]. The exponent μ is a key factor governing the long-time characteristics of the growth and the dispersal patterns, whereas other details of $J(r)$ such as its short-distance functional behavior are less relevant [21]. A broad range of expansion behaviors is encompassed by varying the kernel exponent (limited to $\mu > 0$ to ensure a normalizable distance distribution). At high μ, the jump kernel decays quickly with increasing distance, and a colony expands at a constant rate as if there were exclusively short-range dispersal. As $\mu \to 0$, spatial structure becomes irrelevant and a colony grows as if it were in a well-mixed liquid environment. The intermediate range of kernel values connects these two extremes in a tunable manner.

Recent work has catalogued the distinctive features of population growth dynamics [21] and spatial genomic patterns [23] that can be achieved upon varying the kernel exponent in range expansions driven by power-law growth kernels (a detailed summary is provided in Section 2.2). These studies have identified a critical value of the kernel exponent μ below which the population grows nearly as fast as a well-mixed population, and a significant fraction of the neutral variation in the originating population is preserved for arbitrary long times due to serial reintroduction of variants from the core of the expanding population. For kernel exponents close to but above the critical threshold, population growth is slowed down dramatically and neutral diversity is steadily eroded. However, at even higher values of μ, the behavior approaches that of short-ranged jump kernels, where the population advances as a front moving outward at constant speed. In this situation, a small fraction of the diversity in the originating population persists due to the formation of sectors [15, 53].

Less well understood is the influence of the second key feature of spatial population models: the density regulation mechanism. Modeling growing populations in a spatial continuum presents challenges to both the forward-in-time [54] and backward-in-time [55, 56] approaches, due to the necessity of systematically imposing a local region of influence within which each individual can impact the growth of its neighbors. Local density regulation is commonly implemented by dividing up space into a regular grid of well-mixed subpopulations called demes, each of which has a fixed carrying capacity. Migration events, drawn from the jump kernel, transport individuals across demes. Deme-based models and their variants are widely used in population genetics [57], including for the study of range expansions [12, 53]. However, models that rely on a lattice of demes have their limitations. By design, they do not capture spatial structure and stochasticity at scales smaller than the effective deme size. Imposing an artificial grid of demes also introduces artifacts to the population structure, which can in some instances get *worse* upon increasing the grid resolution to better approximate a continuum [32].

Additionally, using deme-based models forces researchers to make decisions about the specifics of deme saturation and population management. The following selection of recent work exemplifies various possible strategies. Some may choose to have demes that instantaneously change from being empty to full upon the arrival of the first migrant [21, 23], while others may let the deme population grow logistically at a predetermined rate [30, 31] or let the growth be determined by random migration events that bring in individuals from other demes [58]. Death can occur in various ways, such as by attempting to disperse into an already full deme [21, 23] or by being randomly resampled out of an overfull deme's population [58]. If the density regulation unit is the deme population as a whole rather than the individuals in the deme, death may not explicitly occur to any individuals, but the deme population size changes from one time step to the next [30, 31]. Since most computational studies involving long-range dispersal, including the quoted prior results [25, 26, 27, 28, 29, 21, 22, 23, 30, 31], have relied on

10

deme-based approximations, the applicability of their conclusions to continuum-space population growth remains an open question.

The aforementioned results on the population dynamics and neutral evolution of range expansions driven by power-law kernels [21, 23] were derived using a lattice of demes with an additional simplifying assumption: upon arrival at an empty deme, the pioneer immediately saturates the deme, excluding any other migrants from establishing themselves. Not only does this assumption exclude any effects of local dynamics on population growth, it also enforces a local *founder takes all* effect where only one migrant is allowed to contribute to the genetic makeup of a density regulation region. Instant local saturation is justified when long-range jumps are rare and most offspring land within a short distance of their parents; then, the local logistic growth within a deme occurs extremely fast compared to the typical time to arrival of another migrant from a different deme, and can be treated as instantaneous. However, the instant saturation and founder-takes-all assumptions can be invalid when the time scales of local and long-range dispersal are comparable, in which case a local region might receive and send out several migrants while it is being saturated. The influence of the breakdown of fast local saturation on the population dynamics and the spatial genomic structures left behind by long-range dispersal is unknown.

In this work, we address these gaps in our knowledge of range expansions driven by long-range dispersal by performing and analyzing continuum space, individual-based simulations of range expansions driven by power-law kernels. Our simulations were implemented in the population genetics program SLiM [33], and do not use a grid of demes or assume instant saturation of the local carrying capacity by the first arrival. Instead, individuals occupy positions in continuum space and their survival depends on the number of other individuals present within a defined region of influence at the time of their birth. When possible, we compared the outputs to the predictions from models based on lattices of demes of the population growth rate [21] and the evolution of

11

neutral genetic diversity [23]—we term these prior models "lattice-based" predictions. We found that our results often agreed with the lattice-based predictions, giving conditional support to prior results based on models that only focus on the founders. However, when individuals can share resources with many others, we found that focusing exclusively on the founders misses important dynamics between coexisting or competing alleles. In those cases, it becomes necessary to also consider individuals who arrive after the pioneer. We identify parameter regimes where using the lattice-based models is justified, and show that they depend on the specific kernel exponent.

2.2 Background

We first summarize prior results [21, 23] on range expansion dynamics for populations experiencing long-range dispersal with fat-tailed kernels, which were obtained using lattice models conforming to the founder-takes-all assumption at the deme level. Ref. [21] used a lattice model to quantify how a colony expands into unoccupied space when offspring are dispersed according to a power law jump kernel that decays according to $J(r) \sim 1/r^{\mu+1}$. The authors showed that the power-law tail captures the qualitative features of the long-term population growth, and that the short-range behavior of the jump kernel has a negligible impact on the long term population growth. The model of Ref. [21] (hereafter "the lattice model") divides d-dimensional space into a lattice of habitats or "demes". Occupied demes generate offspring according to a Poisson process; offspring attempt to migrate to a new deme randomly chosen by drawing a dispersal distance from $J(r)$ and a random direction relative to the originating deme. Instant local saturation is assumed and is enforced in the model by allowing only two states to each deme: occupied, or empty. A migration attempt to an empty deme is successful, and immediately turns the state of that deme to occupied. A migration attempt to an occupied deme is unsuccessful, and the offspring dies. These assumptions guarantee a founder-takes-all effect at the local level. Henceforth, when we refer to the lattice model, it is implied that instant local saturation and local founder-takes-all are enforced. Much

of our current understanding of jump-driven range expansions derives from the lattice model, as summarized below.

2.2.1 Population growth and time-doubling hierarchy.

Analysis of the lattice model [21] showed that at all times t, a core region of the colony can be identified that is centered at the originating population of the range expansion and within which most demes are occupied. The size of this core region is proportional to the total population size $M(t)$ of occupied demes in the expansion. The long-time asymptotic behavior of the radius of the core region $\ell(t) \propto [M(t)]^{1/d}$ depends on the "heaviness" of the tail of the jump kernel, which is set by the kernel exponent μ. There are two distinct growth possibilities, separated by the value $\mu = d + 1$: the colony expands at a constant rate for $\mu > d + 1$ and it expands faster than linearly when $\mu < d + 1$ [21]. The faster-than-linear growth regime is driven by long jumps whose characteristic size continues to increase as the core expands: this "jump-driven" growth regime, in which pioneers have a large impact on growth, will be the focus of this paper. Within the jump-driven regime, a second special value $\mu = d$ separates two distinct asymptotic behaviors of the core growth at long times ($t \to \infty$): when $d < \mu < d + 1$, the core grows asymptotically as a power law which is faster-than-linear in time ($\ell(t) \propto t^{\frac{1}{\mu-d}}$), in contrast to stretched-exponential growth when $\mu < d$ ($\ell(t) \propto \exp(B_\mu t^\eta)$, where B_μ and η themselves depend on μ and d).

A key result of Ref. [21] was that expansions in the jump-driven regime are governed by a hierarchical time-doubling structure, as depicted in Fig. 2. The core of the colony expands by "absorbing" satellite colonies that were seeded at an earlier time by a rare but consequential long jump. A typical satellite being absorbed into the core at time t was seeded approximately at time $t/2$ by an offspring who dispersed roughly a distance $\ell(t)$ from its parent in the core of the colony. Mathematically, this self-consistency condition can be expressed as

$$\ell(t)^{d+\mu} \sim t\ell(t/2)^{2d}, \tag{2.1}$$

13
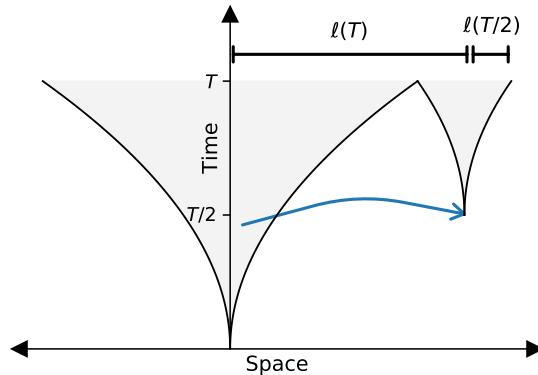
Figure 2. Schematic diagram of the time doubling hierarchy discovered by Ref. [21]. Shaded parts of the plot represent regions of space that are occupied at a given time. The core of the colony (central funnel) grows by absorbing satellites that were seeded at an earlier time by long-range dispersal. A typical satellite being absorbed into the core at time T (smaller funnel at right) was seeded at time of order $T/2$ by an offspring who dispersed a distance of roughly $\ell(T)$ from its parent in the core of the colony; it has grown to a size of order $\ell(T/2)$ when it merges with, and becomes part of, the core.

where the tilde signifies agreement of the leading functional dependence of either side of the relation on the time variable, without including time-independent prefactors or terms whose fractional contributions vanish at long times. The time-doubling hierarchy and Eq. (2.1) form the basis for deriving the asymptotic functional forms of $\ell(t)$ summarized above; unlike those asymptotic forms that are valid only at very long times $t \to \infty$, the self-consistency condition holds as long as the population is large enough that an appreciable number of long-range jumps have occurred [21]. Equation (2.1) forms a basis for more accurate functional forms of the outbreak growth dynamics [21], and also leads to quantitative insights into the evolution of genetic diversity when multiple variants are present in a population experiencing long-range dispersal [22, 23]. Note that the time-doubling hierarchy only relies on the assumption of instant local saturation, and does not require that space be discretized into a regular lattice of demes.

2.2.2 Persistence of initial neutral variation. A striking consequence of range expansions is that the combination of stochasticity and spatial structure can leave behind patterns of neutral genetic variation that are typically associated with selection,

14
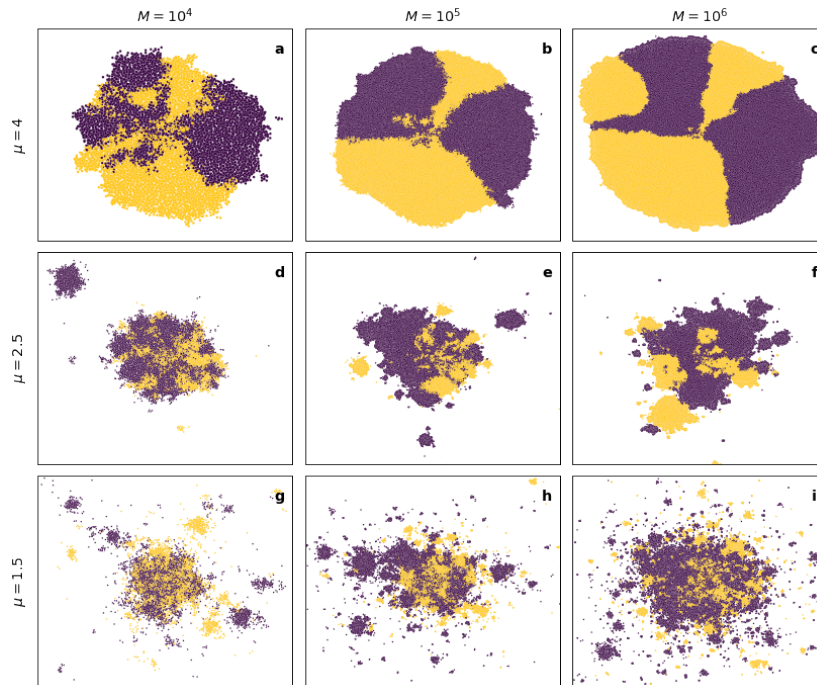
Figure 3. Snapshots of simulated range expansions at different population sizes M. These simulations began with 100 individuals equally split between two neutral alleles, labeled as either purple or yellow in these plots. **a-c.)** Diversity is preserved by the formation of monoallelic sectors for $\mu > d + 1$. **d-f.)** The small number of satellite outbreaks act as bottlenecks, eroding diversity for $d < \mu < d + 1$. **g-i.)** Long-range jumps transport alleles from the core to the exterior of the colony, preserving diversity for $\mu < d$. Additional parameters are $K = 10$ and $p = 0$.

such as sweep-like enrichment of individual alleles [13], diversity gradients [14], and segregation of variants into distinct regions [15]. Simplified models of neutral evolution in spatially structured populations enable us to understand such patterns and to distinguish them from the outcomes of selective events. One aspect of neutral variation that is closely tied to the mode of dispersal is the persistence of initial genetic diversity in the originating population during its expansion into new territory [25, 52]. When dispersal is exclusively short-ranged, only individuals near the edge of the range expansion contribute to future variation; in the absence of new mutations, much of the initial diversity can be lost over time due to successive founder events at the edge. Long-range dispersal enables regions far within the population to contribute to the expansion, which maintains their alleles in the growing population and favors diversity. However, founder effects are not eliminated: each long-range jump seeds a satellite outbreak in which all offspring share the allelic identity of the seeding pioneer, acting as a genetic bottleneck which eliminates diversity locally in the absence of mutations. The fate of the initial neutral variation as the expansion progresses is determined by the balance between these contrasting effects.

The evolution of initial neutral diversity in jump-driven range expansions was analyzed in Ref. [23], using a lattice model in which neutral variation was introduced in the starting population, and no new mutations appeared during the expansions. The existence of the time-doubling hierarchy, Eq. (2.1), was used to identify an effective population of homogeneous satellites whose evolution captures the balance between diversification and coarsening for a given jump kernel exponent. As with the behavior of the core radius growth, the amount of initial diversity preserved after a range expansion was shown to suffer different fates depending on the value of the kernel exponent relative to the spatial dimension. When $\mu < d$, the diversifying influence of long jumps dominates; note the large number of satellites well separated from the core in Fig. 3g–i. The seeding of many satellites by long-range dispersal events from the core enables the population to preserve a finite amount of its initial heterozygosity at long times. By contrast,

16

when $d < \mu < d + 1$, the local coarsening of diversity due to bottlenecks becomes more significant; note the small number of large monoclonal satellites in Fig. **3d–f**. The heterozygosity decays inexorably towards zero as the range expansion progresses, albeit at a slow rate. As μ approaches d, the heterozygosity approaches a finite value but the convergence to this value becomes extremely slow and cannot be observed over practical simulation times. Notably, for $\mu > d+1$, some diversity is also preserved at long times due to the formation of sectors in outward range expansions, as shown in Fig. **3a–c** [15, 53]. Jump kernels of intermediate breadth ($d < \mu < d + 1$) therefore support lower neutral diversity than broader ($\mu \leq d$) and narrower ($\mu \geq d + 1$) kernels.

In summary, Ref. [23] established that long-range dispersal can preserve some of the genetic diversity from the originating population at long times, but only for jump kernels broader than a dimension-dependent threshold. Narrower kernels cause diversity to erode over the course of the expansion due to successive founder events, which can erase even the limited heterozygosity preserved due to the formation of sectors in range expansions with exclusively short-ranged dispersal. However, these features were observed in lattice models which assumed instant local dynamics and founder-takes-all at the deme level; the influence of slow local saturation on the evolution of heterozygosity could not be gauged. In this study, we aim to establish whether insights derived from lattice models of range expansions still apply in a continuous-space model for which the lattice model assumptions can be violated to a controllable degree, and to quantify the effect of explicit local dynamics on neutral genetic variation as the expansion progresses. We next introduce our simulation model which we use to investigate these questions.

2.3   Methods

In order to study jump-driven range expansions which rely on neither a lattice nor the assumption of instant local dynamics, we used the evolutionary simulation software SLiM [33] to simulate range expansions on a 2D continuous landscape without restricting ourselves to a lattice of demes. Individuals produce offspring at a constant
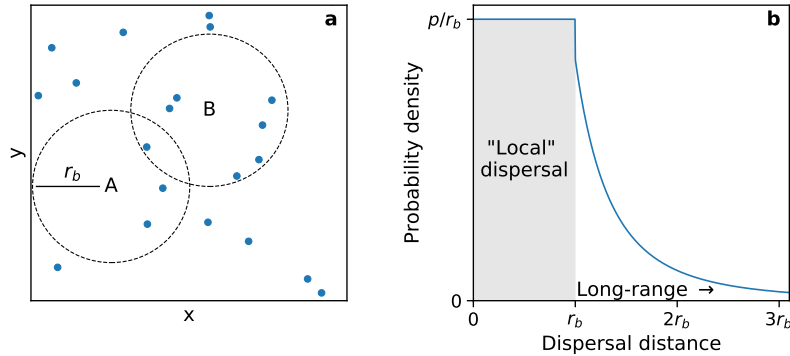
Figure 4. An outline of the simulation procedure. **a.)** A snapshot of a population during a range expansion. The dots represent individuals in the population. Suppose the local carrying capacity is 5. An individual born at position A would only count three others in its local region (dashed circle centered at A), so it would survive. An individual born at position B would count seven others within its local region. That is too many for the individual to successfully compete against, so it would die. **b.)** An example jump kernel. There is a probability p of dispersing within the "local" (shaded) region, that is, within distance $r_{\rm b}$. The jump kernel decays according to the power law $J(r) \sim 1/r^{\mu+1}$ beyond the local region.

rate, and offspring attempt to establish themselves by dispersing in a random direction with dispersal distances drawn from a jump kernel incorporating short-ranged and long-ranged dispersal which we define below (Eq. (2.3)). To focus on the effects of the spreading process and enable direct comparison with previous work (see Section 2.2), our model includes two simplifying assumptions. First, each individual has an allelic identity which is passed on exactly to offspring with no possibility of new mutations; this enables us to evaluate the persistence of initial neutral variation purely due to dispersal and spatial structure during spreading. Second, once offspring are successfully established, they do not move, die, or renew themselves. This assumption allows us to hone in on the dynamics of establishment and expansion, without confounding effects or computational expense from reshuffling and replenishment of regions that have already been saturated. Immortality and immobility post-establishment provide a reasonable approximation for trees that produce massive numbers of seeds over scores of growing seasons, or perennial plants that replenish themselves in place once established. Even in populations for

18

which these assumptions do not hold, the patterns left behind by the initial expansion can still be representative of long-time trends despite the subsequent gene flow due to replenishment and reshuffling of individuals [16, 18].

In the absence of demes with a fixed carrying capacity, a different mechanism to regulate population growth is needed. We assume that the environment has uniformly-distributed resources which can support a uniform carrying capacity per unit area, quantified by a maximum population density ρ. We introduce an interaction distance r_{b} which demarcates a disc-shaped region within which an individual competes with others for resources (Fig. 4**a**). The population density and interaction distance can be combined to define a local carrying capacity K via

$$K = \rho \pi r_{\mathrm{b}}^2. \tag{2.2}$$

When an individual is born, it undertakes a random dispersal event and counts the number of individuals within the interaction region surrounding its new location. If there are at least K other individuals in the interaction region, the duplication event is unsuccessful and the new individual dies. If there are fewer than K other individuals, the new individual establishes successfully in its new location and survives for the remainder of the simulation.

The local interaction region in our continuous-space simulation resembles the geographic subdivision unit (the deme) used in lattice-based models. The concept of instantaneous local saturation, or a local founder-takes-all effect, would therefore correspond to an individual quickly filling its interaction region with its offspring before it (or its descendants) attempted any long-range dispersal events. In order to smoothly depart from the assumptions of the lattice model, it would be useful to control the fraction of dispersal events which are "local", i.e. within the interaction region, as opposed to long-range. To do so, we used a two-part jump kernel that allows us to explicitly specify the probabilities of local versus long-range dispersal, as sketched in

19

Fig. 4**b**. In full, the jump kernel is as follows:

$$J(r) = \begin{cases} p/r_{\text{b}} & r \leq r_{\text{b}} \\ \dfrac{(1-p)}{r_{\text{b}}^{-\mu}}\mu r^{-(\mu+1)} & r > r_{\text{b}} \end{cases} \tag{2.3}$$

where p is the probability of dispersing within the local region. The short-range part of the jump kernel is chosen to be featureless, with the only notable property being that the integrated probability $\int_0^{r_{\text{b}}} J(r)\,dr = p$. The long-range part of the jump kernel matches the power-law kernel used in the prior works discussed [21, 22, 23] and the prefactor ensures the normalization $\int_{r_{\text{b}}}^{\infty} J(r)\,dr = 1 - p$. Jump distances are randomly drawn from this distribution using inverse transform sampling (detailed procedure in Section 2.6.1).

A few comments about our choice of jump kernel, Eq. (2.3), are in order. Our aim is not to exactly reproduce a biologically measured jump distribution at all lengths, but rather to capture the two main features of interest in a simplified kernel—a tunable balance between short- and long-range dispersal determined by the parameter p, and a fat-tailed kernel with a specified power-law falloff controlled by the exponent μ. For simplicity, we chose the short-range part of the jump kernel to be constant with distance r; other forms are expected to lead to similar results provided the integrated probability of jump lengths between 0 and r_{b} evaluates to p. The chosen form also implicitly assumes that the same length scale r_{b} governs the interaction distance for the density regulation and the dispersal behavior. We could have built a model with an additional length parameter dictating the spatial features of the dispersal kernel, but at the cost of added complexity and a larger parameter space. Our simplified choice allows us to dial in a specific balance between local and long-range dispersal by adjusting the parameter p alone, which enables direct comparisons of different simulations where the kernel exponent, local carrying capacity, and size of the density regulation region are kept unchanged. Since the exact shape of the jump kernel at short distances is not biologically realistic (for instance, it has a discontinuity at $r = r_{\text{b}}$), we do not use our model to draw

any conclusions about the spatial distribution of individuals on scales smaller than the interaction distance.

We now specify appropriate units for length and time in our simulations. Since the individuals and the environment are both featureless, and the same length scale r_b governs both the density regulation and the dispersal, the interaction distance is the natural length unit in our model. In our simulations, we set r_b to one, so that all distances reported from simulations are in units of r_b. Time units are chosen such that each individual generates offspring via a Poisson process with a duplication rate of one; i.e. time is reported in units of the average generation time for an individual. Note that not all offspring survive, because of the density regulation mechanism.

Once the length and time units have been fixed, the consequential parameters are the kernel exponent μ, the probability of local dispersal p, and the local carrying capacity K (which determines the local density ρ via Eq. (2.2)). Simulations begin with $10K$ individuals whose x and y positions are random draws from a Gaussian distribution with mean zero and standard deviation $2r_\mathrm{b}$. Everyone in the population gets a chance to produce offspring every time step, which disperse according to the jump kernel with relevant p and μ and then either survive or don't depending on the population density where they happen to land. Simulations end once the population size exceeds a predetermined threshold, usually four orders of magnitude larger than the initial population size. See Section 2.6.1 for more details on the simulation procedure.

We next identify characteristic time scales in the problem which will enable us to choose parameters which violate the instant local dynamics and local founder-takes-all assumptions. (For an expanded discussion with potential improvements, see Section 2.6.2). First let us consider the characteristic saturation time scale for a single interaction region (which takes the place of a deme in our model). While the full saturation dynamics is complicated because of the influence of offspring from nearby interaction regions, we can make a simplified estimate of the saturation time by

21

considering only the descendants of the pioneer individual which undergo local dispersal. Assuming that all these descendants land in the same interaction region, we have an effective division rate of p (in our units) for the local population. In this simplified model, the interaction region fills up according to a logistic function with growth rate p, for which the saturation dynamics are set by the characteristic time scale $\tau_{\rm s} \equiv 1/p$. (The actual saturation time for a deme with a discrete population has an additional logarithmic dependence on the carrying capacity, see Section 2.6.3; we ignore this weaker dependence compared to the dominant $1/p$ dependence in the present discussion of characteristic time scales.) The saturation time scale must be compared to the typical time for the interaction region to send out long-range jumps. The highest possible rate occurs when the region has saturated to population K and sends out long-range jumps at a rate $K(1-p)$. Therefore, we identify $\tau_{\rm j} \equiv 1/(K(1-p))$ as the characteristic time scale separating long-range jumps out of an interaction region. Note that the local saturation time scale is independent of K, whereas the rate of long-range jumps out of an interaction region does depend on K.

The instantaneous local dynamics assumed in the lattice model is approached when the local saturation time is much smaller than the typical time between long-range jumps; i.e. $\tau_{\rm s} \ll \tau_{\rm j}$. Using the above estimates for the characteristic times, we find the criterion

$$\frac{p}{1-p} \gg K \tag{2.4}$$

for fast local dynamics. This criterion is always satisfied as $p \to 1$. When K is large, p must be at least $1 - 1/K$ for Eq. (2.4) to be satisfied: for appreciable local carrying capacities, the fraction of local dispersal events must be very close to one for the criterion to hold. If an individual competes with a large number of other individuals in its neighborhood for resources, Eq. (2.4) is satisfied only if the vast majority of dispersal events are local and long-range jumps are exceedingly rare. Our estimate emphasizes

the need for simulations with explicit local dynamics to investigate the broad range of parameters where the lattice model assumptions do not hold during jump-driven range expansions.

The criterion $\tau_s \ll \tau_j$ ensures that new migrants originate from fully saturated regions. To satisfy the second assumption of the lattice model—the local founder-takes-all effect—we additionally require that the characteristic time between the arrival of the first migrant and a potential second migrant by long-range dispersal, which we call τ_2, is much larger than the local saturation time scale τ_s. Unlike τ_s and τ_j, however, we do not have direct control over τ_2; the latter time scale will depend not only on the model parameters but also on the location of the region being colonized. For example, the expected time to second arrival will be different for a region near the core of a colony compared with a region far from the core that was recently seeded by long-range dispersal. Nevertheless, we expect that τ_2 is closely related to the time scale τ_j associated with sequential long-range jumps out of any given region: if long-range dispersal from all regions is exceedingly rare (τ_j is large), it will take a very long time for a second migrant to arrive into a newly colonized region (τ_2 is large as well). Therefore, we use the same criterion, Eq. (2.4), to gauge whether both assumptions underlying the lattice model are satisfied in our continuum model. In the next section, we directly verify that our simulation results include regimes which violate the assumptions of instantaneous local saturation (Section 2.4.1) and local founder-takes-all (Section 2.4.2), thereby departing strongly from the prior lattice models.

In summary, to violate the lattice model assumptions we require local dispersal probability values comparable to or lower than $1 - 1/K$. If we evenly sample values of p between zero and one, we find that the lattice model assumptions are violated at most parameter values. For instance, if we set the carrying capacity to $K = 10$, the criterion is violated for p values up to around 0.9; when $K = 100$, the criterion is satisfied only for $p > 0.99$. In our simulations, we choose values of carrying capacity K between 10
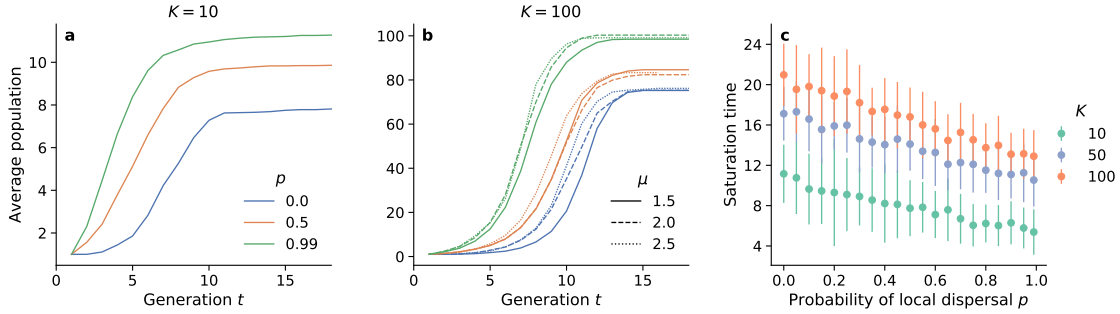
Figure 5. Saturation dynamics of interaction regions around pioneers. **a.)** and **b.)** show the population growth within the interaction region of pioneers (individuals which land in an empty region) as a function of time from establishment of the pioneer, averaged across many pioneers for different values of p (colors). **a.**, $K = 10$ and $\mu = 1.5$; **b.**, $K = 100$ and three different kernel exponents (dashes). Each curve in panels (a.) and (b.) is the average of the local saturation around approximately 60 pioneers gathered across multiple simulations. **c.)** Saturation time of interaction regions, defined as the time taken for the fitted logistic growth function describing the population within an interaction region to reach one less than the saturating population (see Section 2.6.3 for details), for $\mu = 1.5$. We fit the logistic growth function to the local saturation data of approximately 60 interaction regions around pioneers and then computed the saturation time for each region based on the fitted growth rate and carrying capacity. The points in the plot are averages and the error bars are the standard deviations of the computed saturation times of individual interaction regions at each set of parameters.

and 100, and local dispersal probabilities in the range $0 \leq p \leq 0.997$. We expect the lattice model assumptions to be violated over most of these parameter values, except at the upper range of values of p.

2.4 Results

2.4.1 Local dynamics are consistent with logistic growth.
We first analyze the effect of modifying the local dispersal probability p on the population dynamics within interaction regions. Consider the fate of the interaction region surrounding a pioneer that has landed in an empty part of the range. If all local dispersal events experienced by the pioneer and its offspring landed within the pioneer's interaction region, we would expect exponential growth of the local population with rate p until the carrying capacity K is reached. In practice, the interaction regions of the offspring only partially overlap with that of the pioneer, so the population growth levels off smoothly

24

upon approaching the maximum value. When saturation curves across many interaction regions are averaged for a given set of parameters, the average curve takes on the form of a logistic function as shown in Fig. 5**a**–**b**. Upon varying p and μ independently, we find that the saturation proceeds faster as p is increased whereas it is not strongly affected by the kernel exponent (Fig. 5**b**).

We use the logistic growth rate, extracted from a two parameter fit to the average growth curves (see Section 2.6.3 for details), to quantify the local saturation dynamics. As expected, we find that the growth rate is largely independent of carrying capacity and is determined by the local probability p (Fig. 13). The growth rate remains nonzero as $p \to 0$, due to multi-step colonization: although no direct offspring of the pioneer can land in its own interaction region, the descendants of these offspring can land within the interaction region of the pioneer which eventually gets filled. Multi-step effects are also responsible for generating saturation curves whose final population values do not exactly equal the carrying capacity K (plateaus at large t in Fig. 5**a**–**b**), as outlined in Section 2.6.3. The true saturation value of the population within an interaction region can be extracted from the logistic fit and is denoted as K'.

Although the logistic growth rate is set by the local dispersal probability and not the carrying capacity, the typical time taken to fill the interaction region of a pioneer depends on both quantities. Since the logistic growth function is continuous and strictly reaches K' only as $t \to \infty$, we define the time taken to reach a local population of $K' - 1$ as the saturation time for an interaction region. We find that the saturation time falls with increasing local dispersal levels, and rises with increasing local carrying capacity, as shown in Fig. 5**c**. The interaction region around a pioneer that seeds a distant satellite takes longer to fill up at low local dispersal rates and/or at high carrying capacities. Notably, the saturation time falls linearly with p, but has a slow (roughly logarithmic) functional dependence on the carrying capacity.
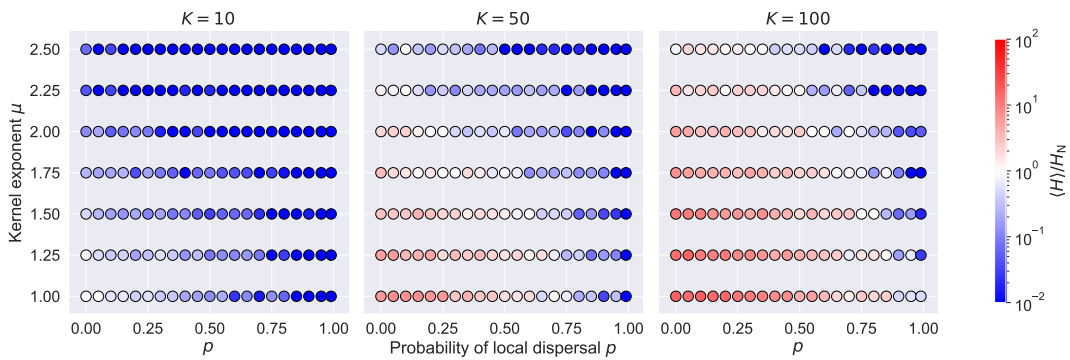
Figure 6. Influence of slow local dynamics on local diversity. Each symbol represents a triplet of parameter values (K, p, μ) and its color shows the average heterozygosity within the interaction regions around several pioneers who seeded distant satellites, normalized against the heterozygosity at which a fully occupied interaction region is expected to have one individual with an allele different from everyone else in the region. Interaction regions are expected to be homogeneous at parameter combinations where the average normalized heterozygosity is less than one (blue points). They are expected to have more than one allele where the average normalized heterozygosity is greater than one (red points), indicating that other individuals typically disperse into and establish themselves within a pioneer's interaction region before it fills up with descendants of the pioneer. The values reported come from the averages across about 50 interaction regions gathered from multiple simulations at each set of parameters.

26

2.4.2 Slow local saturation invalidates founder-takes-all assumption within interaction regions.

Slow saturation of the pioneer's local region increases the chance that other individuals who are not descendants of the original pioneer will disperse into the region and establish themselves before the region is full. If an individual who arrives later has a different allele than the original pioneer, there will be multiple alleles within the region, which introduces genetic diversity within interaction regions in stark contrast to the homogeneous demes imposed by the lattice model. This creates a measurable signal that local saturation times are now comparable to or slower than the typical time gap between the arrivals of first and second migrants by long-range dispersal, τ_2.

To quantify the deviation of local population structure from the local founder-takes-all assumption as the saturation time is increased, we introduced neutral genetic variation in the initial population. Every individual in the initial population was assigned a unique allele, which did not affect the dispersal or reproduction dynamics but was passed on to offspring. The establishment of multiple alleles in the same interaction region was detected by computing the local heterozygosities in the interaction region of isolated pioneers. The heterozygosity, H, is the probability that any two randomly selected individuals will have different alleles. Upon counting the fraction f_i of individuals with each neutral allele i in an interaction region, the heterozygosity of that region is computed as

$$H = 1 - \sum_i f_i^2.$$

A nonzero heterozygosity indicates that more than one allele is present in the region; the larger the heterozygosity, the more evenly distributed the different alleles are in frequency, corresponding to a region in which no single allele dominates.

We averaged the local heterozygosity within the interaction regions of many independent pioneers to obtain a characteristic measurement of the local diversity for

each parameter value. The averaged heterozygosities are normalized against the value at which one expects a fully occupied interaction region to have exactly one individual with a different allele than the pioneer: $H_\mathrm{N} \equiv 2(1/K)(1 - 1/K)$. With this definition, the normalized average heterozygosity $\langle H \rangle / H_\mathrm{N}$ has the following interpretation: normalized average heterozygosities less than one indicate that interaction regions typically have a single allele, whereas values greater than one indicate the expected presence of more than one allele signaling a deviation from the founder "taking all" at the level of the interaction region.

We find that the local heterozygosity is high at low local dispersal rates and high carrying capacities (Fig. 6), consistent with our expectations from the slow saturation dynamics in this part of parameter space. At the smallest carrying capacity ($K = 10$), heterozygosity levels are low across nearly all jump kernels: local saturation occurs fast enough that interaction regions are filled by descendants of the pioneer individual that first arrived in the vicinity. This situation most closely parallels the lattice models. As the carrying capacity is increased, however, we observe appreciable levels of heterozygosity at low levels of local dispersal where the saturation dynamics of regulation regions is slowest (Fig. 5c). As the local dispersal rate increases, a smooth crossover occurs from high to low heterozygosity. The value of p at which this crossover occurs is larger for broader jump kernels (lower μ): longer dispersal events favor mixing of alleles. We expect these trends to continue for carrying capacities on either side of the range we show here. For lower carrying capacities, local diversity would become lower everywhere. For higher carrying capacities, the boundary between pioneer-dominated and not pioneer-dominated (blue points and red points, respectively) would continue to move to the right. The region of parameter space where founders typically "take all" will continue to shrink as carrying capacity increases.

In summary, measurements of local heterozygosity (Fig. 6) indicate a breakdown of founder takes all over wide swaths of parameter space, especially for high carrying

28
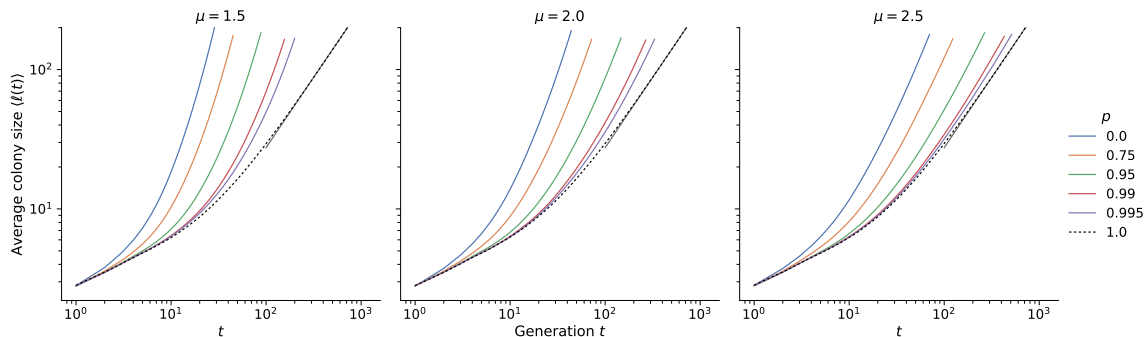
Figure 7. Average growth curves for different μ and p at $K = 10$. These are averages of growth curves from 241 individual simulations at each set of parameters. Average curves at $K = 100$ are shown in Fig. 12. The dashed line corresponds to simulations with only local dispersal; the resulting growth at long times is consistent with a linear relationship (gray line in the upper right of each panel).

capacities and broad jump kernels. While local interaction regions remain largely monoallelic when long-range dispersal is very rare ($p \gtrsim 0.9$), we find evidence that multiple incursions into the same region leave a persistent contribution to the local genetic makeup within interaction regions when long-range and local dispersal rates are of similar order. We next investigate the extent to which these *local* deviations from founder takes all impact *global* features of the population expansion, and in particular whether they lead to departures from the population-level behavior of jump-driven range expansions predicted using lattice-based models in Refs. [21, 23].

2.4.3 Increased long-range dispersal favors faster population growth.

The salient feature of the global population growth under jump-driven expansions is their dramatic speedup compared to expansions that only involve short-range jumps: the typical radial extent of the core region $\ell(t)$ grows faster-than-linearly with time when $\mu < d + 1$. This boost occurs because offspring attempting short-range jumps will land close to their parents and siblings, and are more likely to be unsuccessful due to a lack of local carrying capacity. By contrast, long-range jumps tend to transport offspring to empty areas where they establish and proliferate successfully. Therefore, lower values of

29

the local dispersal probability p are expected to favor faster population growth overall, even though the local saturation is slower.

We measured the population growth with time, $M(t)$, for many independent range expansions at each parameter value. To connect with the results from lattice-based models described in Section 2.2.1, we need an estimate of the core region within which the population has reached saturation. When growth is driven by long-range jumps, there is no sharp boundary between occupied and empty regions even in the lattice model. Rather, the local density is close to ρ out to some distance from the origin, beyond which it crosses over to a power-law decline in density determined by the value of μ [21, 22]. This smoothly varying occupancy profile leaves some ambiguity in precisely defining the core region. We follow Ref. [22] in using the mass-equivalent radius of the population as our best estimate of the core radius from our simulation data:

$$\ell(t) \equiv \sqrt{\frac{M(t)}{\pi\rho}}, \tag{2.5}$$

which provides the required scaling $M(t) \sim [\ell(t)]^d$ [21]. This definition assumes that the bulk of the population is present in regions where the population has reached its maximum density locally. We averaged $\ell(t)$ trajectories across different instances at each set of parameters to get a *growth curve* characterizing the average growth in extent of the population.

We found that the acceleration of range expansion due to long-range dispersal is preserved in the continuum model, as shown by the growth curves in Fig. 7. We focus on the behavior at long times beyond the saturation time scale of a single interaction region (which is of order 10 for $K = 10$, see Fig. 5**c**). When all dispersal is short-range ($p = 1$), the average colony size approaches a linear relationship at long times (dashed curves; linear fit shown with solid curves at upper right), signifying the expected constant-speed outward advance of the population front [53]. Small levels of long-range dispersal (solid curves) are sufficient for the size to grow faster than linearly with time, as evidenced by

a steeper slope on log-log axes compared to the dashed curves. The growth at long times appears to be faster than any power law (i.e. faster than linear on log-log axes) for all values of p at $\mu = 1.5$ and $\mu = 2.0$ (Fig. **7a**–**b**), in line with expectations from the lattice model. Growth approaches a power law in time with exponent greater than one at the the two largest p values for $\mu = 2.5$, but is faster than power-law for the smaller values of p over the population sizes simulated. In all cases, decreasing the probability of short-range dispersal speeds up the colony expansion, as expected: long-range jumps are far more likely to land in empty regions and succeed, compared to local jumps.

Many consequential features of the expansion, however, are determined not by the absolute growth of the population size with time but by the functional form of the growth. For instance, the qualitative differences in global diversity among different kernel ranges (Section 2.2.2) are owed to the different functional forms of $\ell(t)$ observed in the lattice model (see Section 2.2 for a summary). It would be useful to quantify whether and how the local dispersal rate influences the functional form of the population growth curves. A direct comparison of the growth curves to the asymptotic forms derived using the lattice model is not expected to succeed, because the growth curves can take a long time to reach their asymptotic forms, especially for values of μ near the space dimension $d = 2$ [21]. This feature of jump-driven growth is apparent in Fig. **7c**, in which the measured growth curves for $\mu = 2.5$ are nonlinear on logarithmic axes and deviate from the asymptotic power-law form even at long times. Instead, we use the self-consistency condition from Ref. [21], Eq. (2.1), which is expected to hold for times beyond the local saturation time scale but well before the time at which the asymptotic regime is reached in the lattice model. If the entire population in our continuum model were contained in regions that have reached local saturation at all times, then the hierarchy depicted in Fig. 2 would translate to the continuum model as well, and we would expect Eq. (2.1) to be satisfied exactly. This would enable us to predict future population growth given only the current population size and the exponent that characterizes the jump kernel. The

31
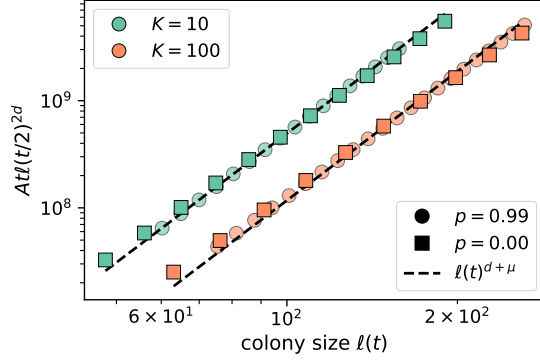
Figure 8. Quantitative test of the hierarchical time-doubling structure. Plots show the RHS of the consistency condition (Eq. (2.1)) versus the colony size $\ell(t)$ for $\mu = 2$. Data are from the average of about 200 growth curves at each set of parameters, and only include the second half of the simulation to exclude expected deviations at short times (see Section 2.6.4 for details). The scaling factor A was adjusted manually to overlay data from different p values for ease of comparison of the apparent power-law exponent (slope of curves on log-log scale). Analogous plots at $\mu = 1.5$ and $\mu = 2.5$ are shown in Fig. 15.

size of deviations from the exact relation could be used to quantify differences in satellite structure between the continuum model and the lattice model.

To test the validity of the consistency condition and its ability to predict population growth, we measured the relationship between the colony size $\ell(t)$ at time t and the quantity $t\ell(t/2)^{2d}$ in our simulations. For t values larger than the local saturation time (order 20 or less for all parameters, Fig. 5**c**), we found that the simulated growth curves are consistent with a power-law relationship between the two quantities across the entire range of local dispersal probability values tested. Data for two representative values of p and two local carrying capacities are shown in Fig. 8; additional curves are shown in Fig. 15. For parameter values which best approximate the assumption of instantaneous filling of density regulation regions (local dispersal probability close to one and low carrying capacity), the power-law exponent quantifying the relationship between $\ell(t)$ and $t\ell(t/2)^{2d}$ also matches the expected exponent of $d + \mu$ (compare green discs to dashed line in Fig. 8). By contrast, the relationship no longer quantitatively matches the consistency condition when local saturation is slowed down by low values of p or high

32
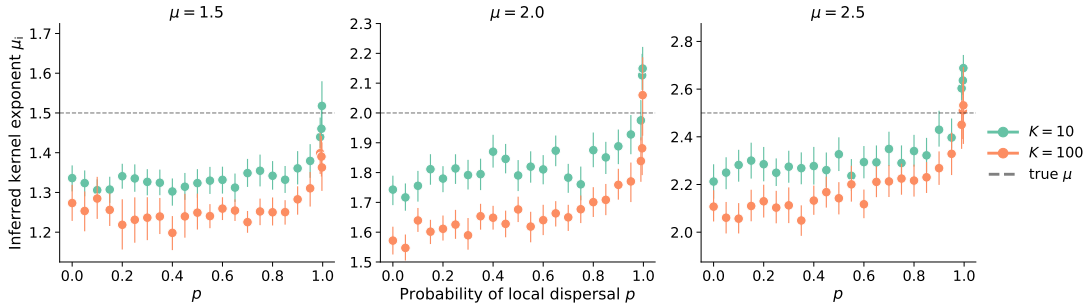
Figure 9. Inferring kernel exponents using the consistency condition. Symbols show the kernel exponent μ_i inferred for the power law relationship between measured $\ell(t)$ and $t\ell(t/2)^{2d}$ (i.e. the slope in Fig. 8) from many individual simulations. Panels are labeled by the true kernel exponent μ used in simulations. The dashed line indicates $\mu_i = \mu$. Each point represents the mean of the individual inferences from roughly 200 independent simulations and the error bars are the 95% confidence interval of the distribution of bootstrapped mean inferred kernel exponents.

values of K. Instead, the population size at time t is larger than that predicted by the population at time $t/2$ according to Eq. (2.1) (square symbols and orange discs in Fig. 8). The functional form of the growth curve appears to be *faster* than would be expected from the time doubling hierarchy, so using Eq. (2.1) leads to an underestimate of the colony size at time t given its size at time $t/2$. Note that Fig. 8 is plotted on logarithmic axes, so the visibly slight difference between the slopes of sets of symbols and dashed lines corresponds to different power law relationships between $\ell(t)$ and $t\ell(t/2)^{2d}$ in our continuum expansions than what is predicted by the time doubling hierarchy derived using lattice models that assume instant saturation of local regions.

To quantify the extent of the deviation from the lattice-model behavior, we fit measurements of the quantity $t\ell(t/2)^{2d}$ to the form $B\ell(t)^\nu$ to extract the power-law exponent ν (see Section 2.6.4 for details). This exponent was used to infer a kernel exponent $\mu_i \equiv \nu - d$ from data such as those shown in Fig. 8, which can be compared to the true kernel exponent μ. To cover the two distinct jump-driven growth regimes and the marginal value $\mu = d$ separating them (as referenced in Section 2.2), we estimate the kernel exponent from growth curves of populations whose jump kernels decay with μ

equal to 1.5, 2.0, and 2.5. We find that the inferred kernel exponent is close to the true exponent when the local dispersal probability approaches one across all jump kernels and carrying capacities tested (Fig. 9). This observation is consistent with our expectation that the limit $p \to 1$ best approximates the lattice model assumptions. However, the inferred kernel exponent is systematically lower than the true value for much of the range $0 < p < 1$, reflecting the shallower-than-expected slopes at low local dispersal in Fig. 8. The inferred exponent grows slowly with the local dispersal probability up to $p \approx 0.9$, and then rises sharply toward the true value as $p \to 1$. This suggests that there could be some functional change to the structure of colony expansion as the parameter changes to nearly all short-range dispersal, while anything less than nearly all short-range dispersal seems to behave similarly regardless of p. The deviations also systematically differ depending on the local carrying capacity, with inferred exponents at $K = 100$ consistently lower than than those at $K = 10$.

We have not isolated the mechanism leading to an inferred kernel exponent μ_i that deviates from the true kernel exponent μ. The fact that $\mu_i < \mu$ implies that the time-doubling hierarchy from the lattice model, quantified in Eq. (2.1), does not hold exactly over much of the range of p values. Furthermore, it shows again that the functional form of the population growth with time is faster in the continuum model than the lattice model. However, this observation by itself does not provide information about how the hierarchy breaks down in the continuum model, or whether a modified version of Eq. (2.1) might be found for continuum space models.

We can nevertheless identify the likely sources of the discrepancy between the continuum and lattice models based on our knowledge of the local and global dynamics. The hierarchy in the lattice model was derived under the assumption that satellites which drive the expansion originate in a core region that has reached its saturation density nearly everywhere, and whose size scales as $[M(t)]^{1/d}$. In our simulations, local regions take some finite amount of time to fill up, but they can begin sending out long

range migrants as soon as they are seeded. An appreciable fraction of satellites may be seeded by individuals dispersing from regions with local densities between zero and ρ; furthermore, the local density could itself vary significantly through the population. These deviations become more prevalent for larger carrying capacities (Fig. 5**c**), which would suggest larger deviations at higher values of K consistent with the behavior of the inferred kernel exponents in Fig. 9.

Altogether, measurements in the continuous-space model reveal small but consistent deviations in the population growth curves from the time-doubling hierarchy predicted in the lattice model. Our simulations indicate that slow local dynamics introduce corrections to the time-doubling hierarchy over a large range of values of the local dispersal probability, consistent with our estimates of the parameter regimes for which the lattice model assumptions break down (Section 2.3). Next, we numerically investigate the impact of these corrections on the dynamics of global diversity, for which the hierarchy of satellite sizes determined the long-time behavior in the lattice model as summarized in Section 2.2.2.

2.4.4   Increased local diversity boosts global heterozygosity but does not overcome long-term trends.   Finally, we investigated the consequences of the enhanced local diversity generated by slow local dynamics (Section 2.4.2 and Fig. 6) on the fate of the initial neutral diversity. Recall the predicted evolution of heterozygosity in prior models assuming fast local saturation and local founder-takes-all effects (summarized in Section 2.2.2): initial variation decayed steadily towards zero for jump kernels with $2 < \mu < 3$ in two dimensions, but some proportion of the initial diversity was preserved for broader ($\mu < 2$) or narrower ($\mu > 3$) kernels. We simulated range expansions where the initial population had equal proportions of two fitness-equivalent alleles (initial global heterozygosity $H_G = 0.5$) and measured the evolution of global heterozygosity. While the outcome of a single simulation is stochastic, we estimated the expected value of the hetorozygosity as a function of population size
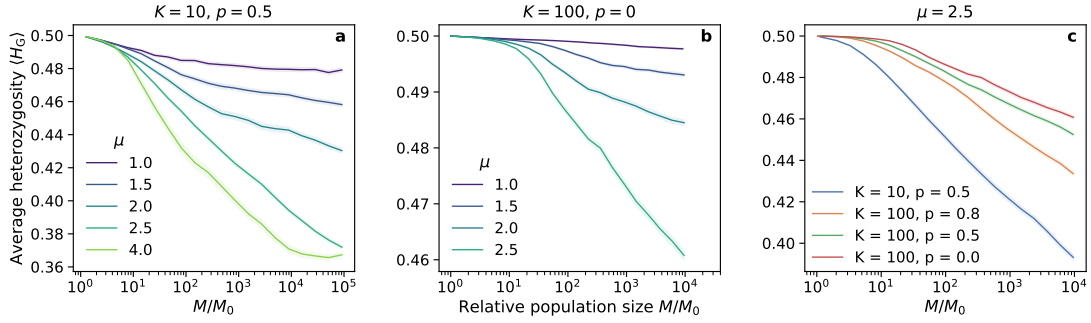
Figure 10. Evolution of global heterozygosity for different kernels and local dispersal rates. **a.)** Average global heterozygosity as a function of the growing population size for different jump kernels, with $K = 10$ and $p = 0.5$. At these parameters, each interaction region is dominated by a single allele (Fig. 6). **b.)** Same as **a** for $K = 100$ and $p = 0$; at these parameters interaction regions tend to harbor multiple alleles. **c.)** Global heterozygosity curves for $\mu = 2.5$ and different local dynamics traversing the spectrum from monoallelic to multiallelic local interaction regions (blue points to red points in Fig. 6). The $K = 10$ data is the same as in panel (a.) but has been truncated for this plot. Data as a function of relative population size were generated by by binning the population sizes from all available simulations and then computing the within-bin $\langle H_G \rangle$; see Section 2.6.5 for details. Shading reports the standard error of the mean within each bin in all panels, as an estimate of the uncertainty in our estimate of the ensemble average. Data come from about 200 independent simulations for $\mu \leq 2$ in panel (a.) and about 400 simulations for $\mu = 2.5$ in panel (a.) and all of panels (b.) and (c.). The curve for $\mu = 4$ in panel (a.) comes from just 24 runs since simulations with $\mu > d + 1$ take much longer and the sectoring mechanism for preserving diversity is well understood (see Fig. 3a–c).

by averaging the outcomes of many independent runs at each set of parameters. Recall that no new mutations appear during the expansions; here we study the long term fate of any pre-existing diversity present in the initial population rather than the emergence of some balance between the loss of diversity (e.g. due to drift) and the promotion of diversity due to new mutations. Although we used a specific initial heterozygosity in our simulations, we expect the observed trends in the *proportion* of initial diversity over time to hold for other values of initial global heterozygosity as well. This proportion is obtained from our simulation data (Fig. 10) by dividing the reported heterozygosity values by 0.5.

We first considered a set of parameters ($K = 10$, $p = 0.5$) for which each interaction region is dominated by the offspring of the seed individual (Fig. 10**a**). This situation approximates the local founder-takes-all mechanism of the lattice models, but does not replicate it exactly as multiple incursions into interaction regions are not strictly excluded. Despite the deviations, we find that the evolution of global diversity in the continuum simulations is consistent with expectations from the lattice model when different kernel exponents are compared. (See Section 2.6.6 for a quantitative comparison.) Average population heterozygosity has settled to a constant proportion of its initial value for $\mu = 1$, and appears to be approaching a constant value as well for $\mu = 1.5$. The slow decay of heterozygosity for $\mu = 2$ is expected; the population may have to grow by several more orders of magnitude before converging to a constant heterozygosity [23]. At $\mu = 2.5$, the heterozygosity decays steadily with no sign of convergence to a finite value, as predicted for lattice models in the range $d < \mu < d + 1$. At $\mu = 4$, a constant heterozygosity is attained at large population sizes due to the formation of persistent sectors with distinct allelic identities (Fig. 3**a**–**c**). In each of the growth regimes separated by the critical kernel exponent values of d and $d + 1$ (2 and 3 respectively in our two-dimensional expansions), the behavior of the global heterozygosity follows the qualitative patterns derived in the lattice model. In spite of the small quantitative differences in the hierarchical structure of satellites merging with the core (Fig. 9), the overall differences in structure which determine the balance between diversification and coarsening in jump-driven expansions are maintained deep within the different growth regimes.

Next, we considered parameters $K = 100$, $p = 0$ for which the local founder-takes-all assumption is violated across all kernels tested according to local heterozygosity measurements. We found that the increased local diversity at these values (as indicated by colors in Fig. 6**c**) contributes to higher global heterozygosities compared to the fast saturation region, as seen in Fig. 10**b** when compared to Fig. 10**a** and Fig. 17. For

instance, at $\mu = 2.5$ the heterozygosity has decayed by around 8% of its initial value when $M/M_0 = 10^4$ in Fig. 10**b**, in contrast to a reduction by over 20% in Fig. 10**a**. The same trend is observed at all kernel exponents: The mix of allelic identities within each interaction region under slower local dynamics provides a reservoir of genetic diversity that allows populations to retain much more diversity than possible under the monoallelic regions imposed by fast local saturation. Nevertheless, the qualitative trends in diversity as the kernel exponent is varied continue to track the expectation for lattice models. In particular, the global heterozygosity steadily decays towards zero for $\mu = 2.5$, albeit at a slower rate compared to the $K = 10$ simulations.

A steady decay in heterozygosity is also observed for other values of the local dispersal probability for the same kernel exponent $\mu = 2.5$, see Fig. 10**c**. Slowing down local dynamics by increasing K and reducing p raises the value of heterozygosity at each population size, but does not prevent the steady decay as a function of M/M_0. These results show that at long times, the diversity-reducing effect of bottlenecks outweighs the local mixing due to slow saturation dynamics for $\mu = 2.5$. We expect that continual heterozygosity loss will be experienced for other kernels in the range $d < \mu < d + 1$ as well, although the rate of decay will be very slow for kernels close to the critical value of $\mu = d$, and for kernels with slower local dynamics (i.e. large carrying capacity and low local dispersal probability). In this regard, the high local heterozygosities observed for kernels with $\mu > 2$ and low p values in Fig. 6 are transients which we expect to decay to lower values if the expansions are allowed to run longer.

2.5   Discussion

Range expansions in populations experiencing long-range dispersal can be dominated by the pioneers who travel long distances and seed satellite colonies. Lattice models that assume that these pioneers quickly saturate the carrying capacity within their local interaction region have provided many insights into the dynamics and population structure of such range expansions [21, 23]. However, real populations operate

in continuous space and with local population dynamics which play out concurrently with the global dynamics driven by long jumps. In particular, the limits on the rates of long-range dispersal for lattice models to be accurate become increasingly strict as the local carrying capacity increases (Eq. (2.4)). We have introduced a continuous-space simulation of range expansions which departs from the gridlike spatial structure and instantaneous local dynamics implied in lattice models, enabling us to quantitatively investigate population growth and neutral diversity in parameter regimes where the lattice models are not expected to be valid.

We found that introducing explicit local dynamics is associated with slow local saturation at low local dispersal rates and especially at high local carrying capacities (Fig. 5). By contrast, the global population growth occurs faster when local dispersal rates are low, because of the increase in long-range jumps that seed satellite populations in unoccupied regions (Fig. 7). The functional forms of the population growth curves show similarities with those from lattice-based models (Fig. 8), but with small yet quantifiable differences (Fig. 9). We suspect that these differences arise due to a violation of a central assumption of the lattice model: that satellites are seeded by long-range migrants who disperse from *fully occupied* source regions. In our continuum model, satellites can begin sending out long-range migrants as soon as they are seeded, which can occur several generations before they saturate at high carrying capacities and low local dispersal rates. In future work, we aim to incorporate this feature into the model of hierarchical population growth sketched in Fig. 2, which would improve the accuracy of theoretical predictions for jump-driven range expansions in situations where local interaction regions are not immediately saturated upon the arrival of a new migrant.

We investigated the effects of departing from instantaneous local saturation on both local and global measurements of neutral diversity. Interrogating the populations within individual interaction regions originally seeded by a long-range dispersal event reveals that multiple lineages, rather than just descendants of the pioneer, become likely

39

as local saturation becomes slower (Fig. 6): our continuum model violates the assumption of a strictly enforced local founder-takes-all effect. Having multiple lineages within interaction regions provides a reservoir of genetic diversity that also enables greater global heterozygosity outside the regime where local founder-takes-all applies: generically, expansions with slower local dynamics exhibit higher global diversity at every stage in the expansion (Fig. 10). Nevertheless, the enhancement in local diversity is not sufficient to overcome long-time trends in global diversity, which continue to be determined by the kernel exponent as was shown in the lattice model [23]. In particular, when $\mu < 2$ the global heterozygosity settles to a stable value after an initial period of decay, whereas for $2 < \mu < 3$ the heterozygosity decays steadily as the range expansion progresses albeit at a slow rate. The decay is a consequence of the repeated coarsening of diversity due to bottlenecks as pioneers expand into their newly occupied surroundings (Fig. **3d–f**). Our results show that this coarsening is slowed down by the increased local diversity when the local founder-takes-all assumption is violated, but it is not completely mitigated and the qualitative long-term trends in global diversity are similar to those predicted using the lattice model. This qualitative agreement with lattice-based predictions is a non-trivial result in light of recent research [32] showing that models based on a discretization of space can leave surprising artifacts in measures of population genetic variation.

Our method of discovering local diversity outside the local founder-takes-all regime was unable to detect if descendants of an individual other than the pioneer were within a local region if they happened to have the same allele as the pioneer by chance. Such information would be useful to investigate genealogical structure beyond the fate of the initial neutral diversity in the population, for example to determine if the pioneer is the most recent common ancestor of everyone else in the interaction region or to study the accumulation of additional neutral mutations during the expansion. A tool like tree sequences [59] could readily be incorporated into our computational model to study such questions, which are a promising target of future work. Understanding the competing

effects of local and long-range dynamics on genealogies in our forward-in-time simulations could also aid the construction of backward-in-time models that incorporate long-range dispersal [60, 61].

Another promising future direction would be to incorporate ongoing local competition among all individuals in the population. In this work, we assumed that established individuals never move or die, modeling populations such as trees which release large numbers of seeds annually and where young saplings stand little chance of outcompeting mature trees around them. However, there are many species of perennial plants, for example, where younger individuals can successfully compete against older individuals in their surroundings. Incorporating population renewal and density-dependent competition in simulations could provide new insights into how these species evolve during range expansions. We suspect that such competition should accelerate the decay of diversity relative to our results for $2 < \mu < 3$ (Fig. 10). Local competition can completely remove alleles from the population, whereas in our model the "losing" allele is surrounded but not lost, and retains a nonzero probability of dispersing an offspring to a faraway vacant habitat.

This work provides a better understanding of the range of validity and the limitations of models of long-range dispersal which rely on instantaneous saturation of local interaction regions and divide continuous space into a lattice. We have confirmed that the conclusions of the lattice model are upheld in populations where pioneers who disperse long distances quickly saturate their immediate surroundings with their descendants; namely in populations with low local carrying capacities and high local dispersal probabilities. Even when the local founder-takes-all condition is violated, we have shown that qualitative trends in population growth and in the evolution of neutral diversity mirror those in the lattice model, albeit with measurable quantitative differences. Heuristics such as the time-doubling hierarchy of Ref. [21] (Fig. 2) and the effective population of satellites identified in Ref. [23] remain useful to understand the

qualitative behavior of expansions under long-range dispersal in non-lattice models. Researchers could employ hybrid discrete/continuous research strategies: identify regimes of interest using the heuristics of the lattice model, and then test and refine these predictions in more realistic continuum simulations.

Our results are relevant to understanding and modeling the dynamics of range expansions in true biological populations, including invasive species, populations fleeing climate catastrophes, and spreading viruses. We now have a better understanding of when the first individual to arrive in a region of space effectively determines the genetic outcome of all others who will later inhabit the same immediate area. Experimenters could estimate the size of the interaction region, the local carrying capacity, and the local dispersal probability in populations of organisms in the lab or in nature. Estimates of those quantities could allow researchers to predict whether or not the founders will "take all" when the population expands its range outwards into new territory, leading to insights about how the population will evolve.

Data availability

Simulation code and the code and data necessary to generate figures are available in the following GitHub repository: https://github.com/paulose-group/explicit-local-dynamics

42

2.6 Appendices

2.6.1 Simulation details. Simulations begin with $10K$ individuals who are given random positions near the origin. Their x and y positions are random draws from a normal distribution with a standard deviation of $2r_{\mathrm{b}}$. Typically about 80% of those individuals survive the density regulation in the first time step. The spatial landscape is large, so the periodic boundary conditions have no effect.

Offspring are produced by cloning without the possibility of mutations, so offspring have the same allele as their parent. Dispersal distances are drawn using inverse transform sampling. Recall, the jump kernel is

$$J(r) = \begin{cases} p/r_{\mathrm{b}} & r \le r_{\mathrm{b}} \\ \dfrac{(1-p)}{r_{\mathrm{b}}^{-\mu}} \mu r^{-(\mu+1)} & r > r_{\mathrm{b}} \end{cases} \tag{2.6}$$

where r_{b} is the boundary between local and long-range and p is the probability of dispersing within the local region.

We begin the sampling procedure by drawing a random number X from the uniform distribution between 0 and 1. That number X is taken to be the the probability of drawing a dispersal distance less than or equal to some distance x (i.e. the integral of the jump kernel from 0 to x). Solving for x gives us our dispersal distance.

If $X \le p$, the offspring disperses locally, so we only need to consider the first term of the jump kernel.

$$X = \int_0^x p/r_{\mathrm{b}}\, dr = \frac{px}{r_{\mathrm{b}}} \rightarrow x = \frac{Xr_{\mathrm{b}}}{p}$$

If $X > p$, the offspring disperses a long distance, so we have

$$X = p + \frac{(1-p)\mu}{r_{\mathrm{b}}^{-\mu}} \int_{r_{\mathrm{b}}}^x r^{-(\mu+1)} dr$$

which leads to

$$x = \left(\frac{(1-p)r_{\mathrm{b}}^{\mu}}{1-X} \right)^{1/\mu}$$
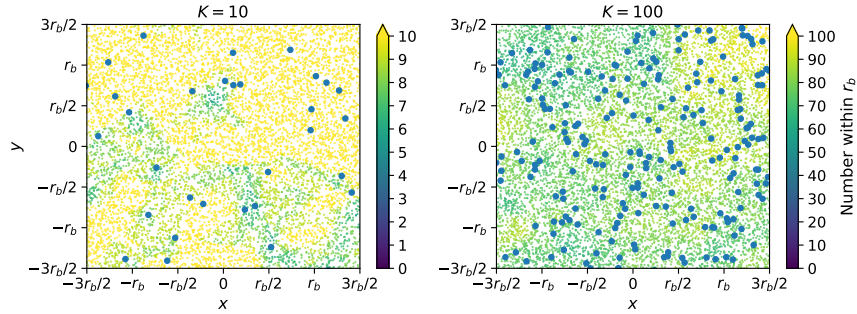
Figure 11. Space doesn't quite fill up to the local maximum population density ρ. The two panels are snapshots from simulations that grew to roughly fifty thousand individuals. The blue points are individuals, and the color of the small yellow and green dots represents how many individuals are within a distance of r_b of that point. Yellow points represent saturated regions; an individual born there would count at least K within its density regulation region. Non-yellow points look hospitable, and one individual born there would count less than K within its density regulation region, but individuals can't fill those spaces since typically everyone produces an offspring every generation and the newborns "destructively interfere".

The dispersal direction is chosen at random from the uniform distribution between 0 and 2π.

All individuals in the population get a chance to produce offspring each time step. Offspring generation is the first thing that happens each time step; the number of offspring for each individual is a random draw from the Poisson distribution with mean 1. Then all newborns *simultaneously* count how many other individuals are within their density regulation regions. This means newborns will count other newborns if they happen to land near each other by chance. It also means that space often doesn't quite fill up to the population density $\rho = K/(\pi r_b^2)$ (as defined in section 2.3). Fig. 11 shows how regions of space may appear habitable, and indeed would be if one single offspring were generated and counted its neighbors at a time, but do not saturate since everyone in the population typically produces an offspring every generation and all newborns count their neighbors simultaneously. All individuals produce one offspring per generation on average, so a region saturated to K individuals is expected to have roughly K newborn individuals attempting to establish within that same region every generation. Our

44
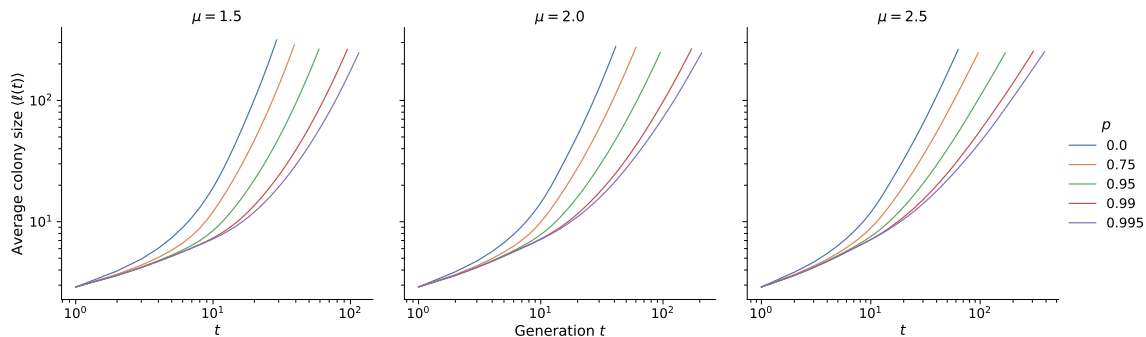
Figure 12. Average growth curves for different μ and p at $K = 100$. Average curves for $K = 10$ are shown in Fig. 7. These are averages of about 140 growth curves at each set of parameters.

density regulation mechanism mimics the biological scenario where none of those newborn individuals are able to get enough resources to survive since there are so many competing for what little is left, a situation termed "scramble competition" in ecology [62]. The typical population that is actually attained in a local density region, which we term K', is estimated using a fit to a logistic growth curve (see Section 2.6.3 below), and deviates by at most 20% from K (Fig. 13). Alternative choices for the density regulation step, such as randomly choosing a subset of newborns to survive so that local density regions can saturate up to the target value K (the "contest competition" scenario), could also be implemented, but at the cost of additional computational resources which would affect the maximum population sizes and growth times that could be simulated.

We typically let the populations grow by about four orders of magnitude, so simulations were ended once the populations exceeded 10^6 or 10^7 individuals for K equal to 10 or 100, respectively. This allowed core radii to grow by about two orders of magnitude, as shown in the average growth curves at $K = 10$ in Fig. 7 and at $K = 100$ in Fig. 12. The solid line indicating the linear relationship between $\langle \ell(t) \rangle$ and t at $p = 1$ in Fig. 7 was generated by fitting a line to the average growth curve from generations 100 to 1000 using NumPy's polyfit() function.

45

2.6.2 Time scales. The assumption of instant local saturation in the lattice models relied on a separation of time scales between local and global dynamics: it is valid provided the time scale for saturation of local regions $\tau_\mathrm{s} = 1/\alpha$ is small compared to the typical time between long-range dispersal attempts from each "deme" or interaction region, which we call τ_j. In our tunable model, our time units are set such that the characteristic time between reproduction attempts is one. The rate of divisions that land within the interaction region is p, which sets the time scale of the logistic growth. Therefore, the characteristic saturation time of local regions is $\alpha = p$, and as a zeroth-order estimate we have $\tau_\mathrm{s} \approx 1/p$. This form is only useful for p close to one, because it ignores the effect of secondary events which land in the interaction region. As a result, the true dependence of α on p is weaker: α grows from 0.4 to 1 as p varies from zero to one (Fig. 13**a**). Therefore τ_s varies weakly from roughly 2.5 to one over the range of p values. For a more accurate estimate, we can use the phenomenological form $\alpha \approx (1+p)/2 \Rightarrow \tau_\mathrm{s} \approx 2/(1+p)$ which does not diverge as $p \to 0$.

When regions have reached local saturation, the rate at which each interaction region sends out long-range jumps is $K(1-p)$. If we assume that the expansion is driven by jumps out of regions that have reached saturation, we have $\tau_\mathrm{j} = 1/(K(1-p))$. Therefore, the condition $\tau_\mathrm{j} \gg \tau_\mathrm{s}$ reduces to

$$\frac{1}{K(1-p)} \gg \frac{1}{p},$$

or

$$K \ll \frac{p}{1-p}.$$

According to this criterion, most of our simulations explicitly do not satisfy the separation of time scales assumed in the lattice model.

2.6.3 Logistic growth description of population dynamics within interaction regions. We started the logistic growth measurements by searching for sufficiently isolated individuals. To find individuals worthy of tracking, we searched
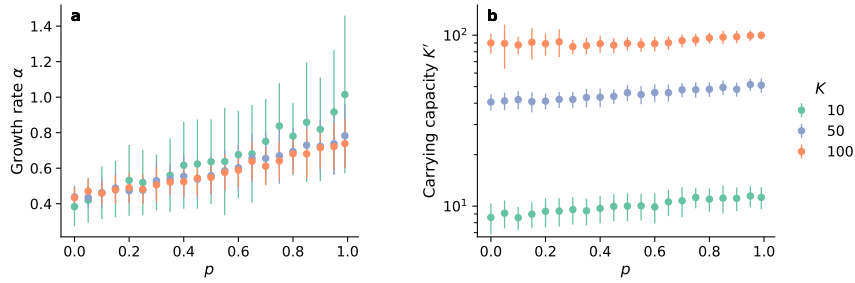
Figure 13. Fitted logistic growth parameters. **a.)** The growth rate increases with increasing short-range dispersal as expected. It does not depend on the carrying capacity because the growth rate is determined by the early growth of the population before the density regulation restricts population growth. **b.)** The fitted local carrying capacity increases slightly with increasing short-range dispersal. Regions often don't saturate all the way to K as discussed above and shown in Fig. 11. We fit the logistic growth function to the saturation data of about 60 interaction regions across multiple simulations at each set of parameters. The points are averages and the error bars are standard deviations of the individual fits. This data comes from expansions with $\mu = 1.5$ and is what formed the saturation times reported in Fig. 5c.

at the end of every generation for individuals who had no one else within a distance

of $10r_{\mathrm{b}}$. Those individuals must have dispersed a long distance. We searched at the

end of every generation until we found at least a predetermined minimum number of

isolated individuals at the same time. We required several at the same time purely for

convenience on the data processing side; these measurements could just as well have been

gathered one at a time as we found the isolated individuals. Nevertheless, once we found

the isolated individuals, we recorded the number within their interaction regions every

generation until the end of the simulation. The local saturation data was then used to fit

for logistic growth parameters. We fit to the logistic function of the form

$$N(t) = \frac{K'}{1 + \left(\frac{K'-N_0}{N_0}\right)e^{-\alpha t}} \tag{2.7}$$

where $N(t)$ is the population at time t, K' is the local carrying capacity, α is the growth

rate, and the initial population is $N_0 = 1$. We used SciPy's curve_fit() function to make

the fits and obtain K' and α. We performed the fits on all individual interaction regions

around the initially isolated individuals that we found that filled up to at least 60% of the local carrying capacity K. Average values and standard deviations are shown in Fig. 13.

We computed the saturation time for an interaction region by setting the population size in Eq. (2.7) equal to $K' - 1$ and then solving for t, which leads to

$$t_{\text{sat}} = \frac{1}{\alpha} \log\left((K' - 1)^2\right) \tag{2.8}$$

In addition to the dominant dependence $\sim 1/\alpha$, where α is itself proportional to p (see Section 2.6.2), we find a logarithmic dependence of the saturation time on the local carrying capacity, which arises from the discrete nature of the local population within a deme. We computed the saturation time for every individual interaction region for which we fit the logistic growth function, using values of α and K' from the fits to the logistic function. We report averages and standard deviations at $\mu = 1.5$ in Fig. 5c.

For the local heterozygosity measurements, every individual in the initial population had a unique allele. We tracked the heterozygosity in the interaction regions of the same individuals for whom we measured logistic growth as described above (i.e. initially isolated individuals). The heterozygosities reported in Fig. 6 are averages of heterozygosities measured across typically about 50 separate interaction regions in the final generation of simulations and gathered from initially isolated individuals in multiple different simulations.

2.6.4 Quantitative assessment of time-doubling hierarchy. We assessed the validity of simulation run times using the consistency condition $\ell(t)^{d+\mu} \sim t\ell(t/2)^{2d}$ (eq. 2.1). The consistency condition is only valid after enough time has elapsed for long dispersal distances to be the driving factor behind a colony's growth [21]. It is necessary to avoid the early times when applying the consistency condition, such as when estimating the kernel exponent as in Fig. 9. Colony growth remains self-consistent once the consistency condition becomes valid. For simulations that ran for T time steps, the values of t we used when applying the consistency condition ran from $T/2$ to T, so the
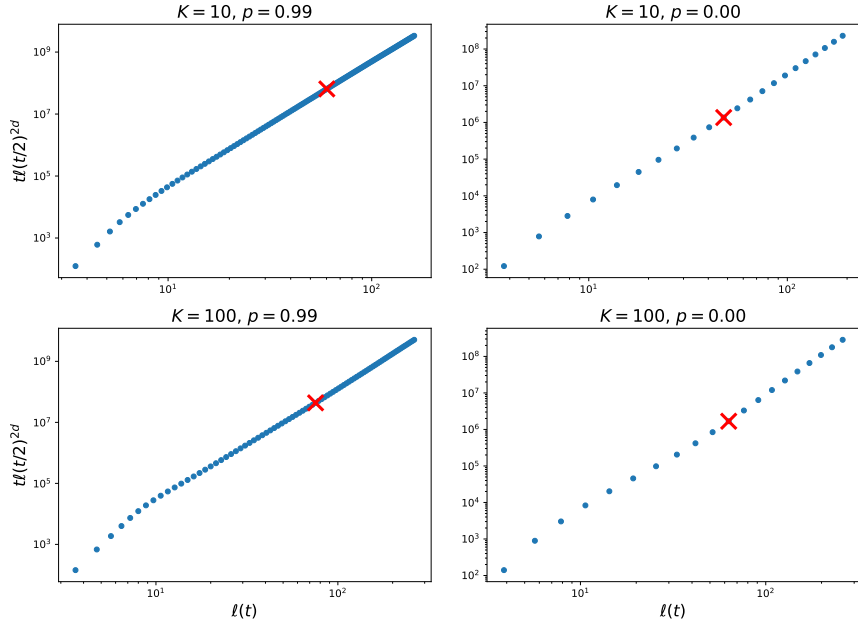
Figure 14. We show $t\ell(t/2)^{2d}$ plotted against $\ell(t)$ from average growth curves at all sets of parameters that went into figure 8. The dots are points from all even-numbered time steps. Simulations with high values of p require many more time steps to reach a given population size than those with low values of p. We use only the points after the red $\times$ in figure 8 and for inferring the kernel exponent as in figure 9. For simulations to be "long enough," we needed at least a handful of points once the growth became self-consistent (i.e. linear on these plots). Using average growth curves from expansions at intermediate probabilities of local dispersal result in plots somewhere between these two extremes: more data points in the linear sections than the $p = 0$ case but not as many as in the $p = 0.99$ case.

values of $t/2$ ran from $T/4$ to $T/2$. The first data point we used with the consistency

condition is marked with a red $\times$ in Fig. 14. We need at least a handful of data points

after the first one to check for agreement with the consistency condition and to estimate

the kernel exponent. Our run times were just enough at the lowest probabilities of local

dispersal and gave us many useful data points at high local dispersal. We compared our

data at $\mu = 2$ against Eq. (2.1) in Fig. 8; analogous plots at $\mu = 1.5$ and 2.5 are shown in

Fig. 15.

    The expansions at high local dispersal require many more time steps to reach

the predetermined population threshold necessary to end simulations than those at low
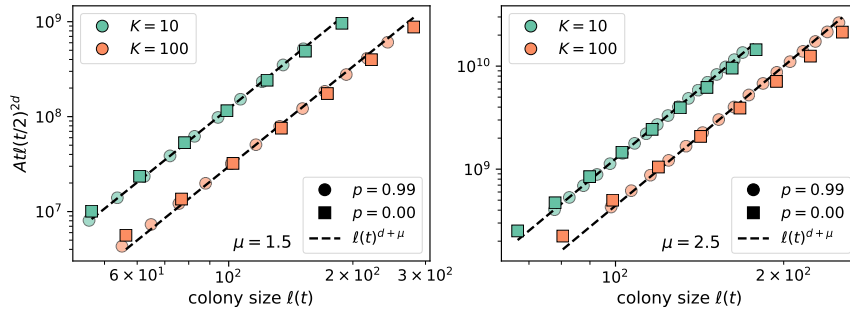
Figure 15. Analogous to Fig. 8.

local dispersal. Expansions at high local dispersal take longer and grow slower than those at low local dispersal since offspring are much more likely to land near their parent in regions that may already be saturated, which means the times t and sizes $\ell(t/2)$ used here are higher at high local dispersal. A scaling factor of $A \approx 20$ was needed to raise the points at low local dispersal to the level of those at high local dispersal. Bringing them together highlights the difference in power laws (slopes) between the sets of points at each value of K.

We found the inferred kernel exponent μ_i from the growth curves by fitting $B\ell(t)^\nu$ to the quantity $t\ell(t/2)^{2d}$ using SciPy's curve_fit() function. We obtained values for both the prefactor B and the exponent ν, but only the exponent was of any interest for estimating the kernel exponent from the growth curves. We estimated the kernel exponent by performing this fit using data from only a later subset of the time steps as discussed in the previous paragraph. We then compute the inferred exponent as $\mu_i \equiv \nu - d$. We computed μ_i using all available growth curves (typically about 200 at any given set of parameters) to get the averages and confidence intervals reported in Fig. 9.

The exact value of μ_i somewhat depends on the fit method. For comparison, we repeated the process of extracting μ_i by finding the best fit line to the relevant data in log-log space, where the exponent could be found from the inferred slope. These two values would exactly match if we had infinitely long simulations that had perfectly converged to constant power laws, but in practice that is not the case. There is often a

50
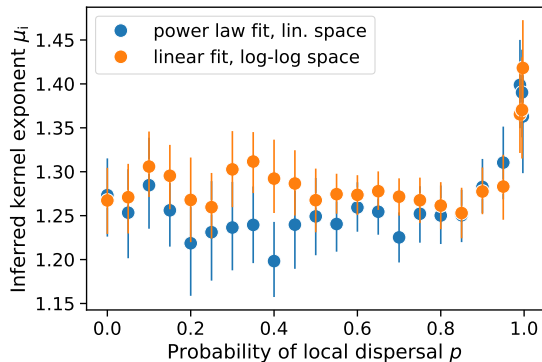
Figure 16. Example comparison between inferring kernel exponents by directly fitting power laws in linear space and fitting for the slope in log-log space. This example data comes from growth curves with $K = 100$ and $\mu = 1.5$ (same data as the orange points in the left panel of Fig. 9); analogous plots look similar at other pairs of parameters. Each point represents the mean of the individual inferences from roughly 50 independent simulations below $p = 0.5$ and roughly 150 simulations above $p = 0.5$. The error bars are the 95% confidence interval of the distribution of bootstrapped mean inferred kernel exponents.

slight difference between the average values of μ_i from the two procedures, as shown at the example parameters in Fig. 16. However, we take the generally overlapping error bars as a signal that it's safe to proceed with our inferred values. This sort of comparison could be used as a test of whether or not population growth has converged to the expected time-doubling hierarchy: consistent gaps between error bars are a warning that simulations may not be long enough. This test led us to run longer simulations to generate the data shown at $p \leq 0.5$ in Fig. 16 and the corresponding data points in Fig. 9. The longer simulations ran until they reached population sizes of 30 million individuals, triple the size of our usual cutoff for simulations with $K = 100$.

2.6.5 Reporting the evolution of global heterozygosity.
There are multiple reasonable ways to compute and display the global heterozygosity as a function of the growing population size as in Fig. 10 and Fig. 17; here we discuss some options and justify our choice. The true independent variable in our simulations is time. Every time step consists of offspring generation and dispersal followed by density regulation as discussed in Section 2.3 and Section 2.6.1. Population size and heterozygosity are

51

recorded at the end of each time step, after individuals have been removed from the population if their birthplaces are too densely occupied. This suggests that the "ground truth" for reporting the evolution of global heterozygosity might be plots of $\langle H_{\mathrm{G}} \rangle$ versus time, where averages and standard errors are computed with all available data at a given time step.

However, for generalizing results or comparing with the results of Ref. [23], it would be useful to compute $\langle H_{\mathrm{G}} \rangle$ as a function of the population size. One method of doing this would be to compute the averages $\langle H_{\mathrm{G}} \rangle$ and $\langle M/M_0 \rangle$ each time step. This method ignores what can be significant variation in population growth rates between individual simulations and generates points whose horizontal and vertical coordinates in the plots of Fig. 10 are both functions of time.

We sought to compute $\langle H_{\mathrm{G}} \rangle$ directly as a function of population size by generating binned population sizes and computing the average heterozygosity from all available simulation time steps where the population size was within a given bin. This means that a single simulation can contribute to a given data point on the $\langle H_{\mathrm{G}} \rangle$ versus M/M_0 curve multiple times or not at all depending on how many time steps the population size was within that bin in that simulation. We used the R function `cut()` to place population sizes within 20 bins of equivalent width in logarithmic space, thus generating equally spaced data points for Fig. 10 and Fig. 17.

We use the standard error of the mean to estimate our uncertainty in $\langle H_{\mathrm{G}} \rangle$. A consequence of the binning procedure is that standard errors of the mean heterozygosity get vanishingly small in Fig. 10 and Fig. 17, despite the fact that the heterozygosity trajectory can vary quite a bit between individual simulations. The bins in those figures often consist of multiple data points from each simulation, especially for the bins at larger population sizes. Even though we generally have 200-400 independent simulations at each parameter combination shown in those figures, the points in the figures are often averages of thousands of data points that fall within each bin, resulting in nearly invisible
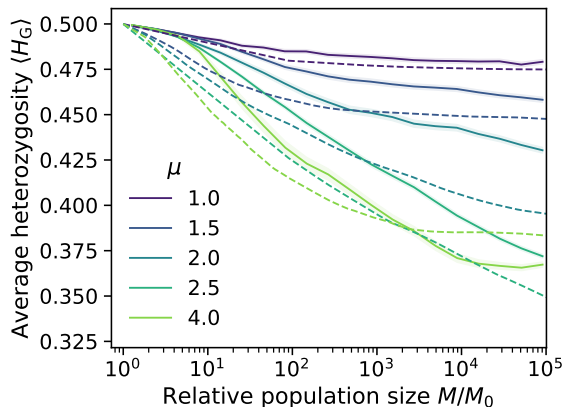
Figure 17. Direct comparison of heterozygosity evolution in the continuum simulations with fast local dynamics ($K = 10$, $p = 0.5$; solid lines and shading are same as in Fig. 10**a**) to that of the lattice-based simulations reported in Ref. [23] (dashed lines). In both cases, the initial population had a 50/50 mix of two alleles (initial heterozygosity of 0.5). Kernel exponents match at all values except $\mu = 2.5$ (continuum), for which the same color refers to $\mu = 2.4$ (lattice).

uncertainty bands since the standard error of the mean is $s/\sqrt{N}$ where s is the sample standard deviation and N is the number of samples. Such a large number of samples within each bin gives us a small uncertainty in our estimate of the average $\langle H_G \rangle$.

2.6.6   Direct comparison to lattice model. We used data for lattice-based simulations from Ref. [23] to compare results between continuum simulations at parameter values $K = 10$, $p = 0.5$ where the averaged normal heterozygosity is less than one (see Fig. 6), approximating local founder-takes-all, and lattice-based simulations where founder-takes-all is imposed at the deme level. The initial conditions in the two types of simulations were not exactly matched: both began with a 50/50 mix of two alleles, but the continuum simulations began with typically about 80 individuals near the origin (Section 2.6.1) while the lattice-based simulations began with 111 occupied demes packed in a disc around the origin. Note that a deme is roughly a discrete analogue of an interaction region, so the continuum simulations' ~80 individuals correspond to roughly $80/K = 8$ occupied demes. Another discrepancy is that Ref. [23] did not generate data at $\mu = 2.5$, so we include their data from $\mu = 2.4$ as a comparison with our $\mu = 2.5$ data.

We observe that the difference in initial conditions leads to different dynamics at early times/small population sizes. In the continuum simulations, most of the early dynamics involves local events which mix and even out the starting population near the origin, and significant changes in heterozygosity only kick in when the population has reached ten times its initial size. By contrast, the lattice simulations only included long-range jumps, and the heterozygosity begins to fall earlier. This discrepancy leads to early differences in the observed heterozygosities between the two sets of models. However, the later trends, especially the contrast between a quick saturation of heterozygosity to a constant value at $\mu = 1$ as opposed to a persistent decay for $\mu = 2.5$ and a decay followed by a delayed saturation for $\mu = 4.0$, are successfully captured by the lattice model. The quantitative discrepancy between the lattice and continuum values of $\langle H_G \rangle$ is largest at $\mu = d = 2$, which is a special point for the underlying dynamics that leads to extremely slow changes in the heterozygosity [21, 23]; we hypothesize that the small discrepancy in the initial conditions persists the longest at this special kernel exponent.

We also observe that the continuum $\langle H_G \rangle$ at large population sizes is higher than that from lattice-based simulations for all jump-driven kernels ($\mu < 3$). This is consistent with the observation that while local heterozygosity is small in the continuum simulations, it is not zero for the chosen parameter values of $K = 10$, $p = 0.5$ (left panel in Fig. 6) and the slight deviations from local founder takes all promote higher heterozygosity compared to the strict founder-takes-all assumption of the lattice model.

CHAPTER III

THE FATE OF ADAPTIVE MUTATIONS IN RANGE EXPANSIONS DRIVEN BY

LONG-RANGE DISPERSAL

This chapter was co-authored with Jayson Paulose.

3.1 Introduction

Range expansions are ubiquitous in biological populations [1]. They occur naturally in various ecosystems, such as when plant seeds are transported random distances by wind or animals [2]. They may be enabled by climate shifts, such as glacial retreat opening up new habitat [9], or forced by global warming as previously habitable places become inhospitable [11]. Humans mediate range expansions in many ways, such as by introducing what will become invasive plants by seeds sticking to the soles of our boots when we go places [7], inadvertently transmitting white-nose syndrome between bats in different caves [63], or carrying the SARS-CoV-2 virus with us when we travel on airplane journeys all over the world [8].

Range expansions can leave behind genomic signatures that may look like local adaptation but actually arise due to recurrent founder events at the expansion frontier [12]. Variants that appear near the edge of a growing colony can "surf" along the outward growth and reach high frequencies in the newly colonized area [13, 14]. Colonies growing in 2D as on a Petri dish stochastically segregate into monoallelic sectors due to drift and founder events leading to fixation of different alleles in different sections of the colony's perimeter [15, 16, 17, 18], see Fig. 18**a**. Genomic regions linked to neutral surfing variants may exhibit decreased diversity solely due to the expansion dynamics [64], although the decreased diversity mimics what can be observed following positive selection [12] and population contractions [65].

Mutations that affect an individual's fitness may appear during a range expansion. The effect conferred by a mutation can be anywhere along the continuum from lethal or strongly deleterious all the way to strongly beneficial. Much work has gone into studying
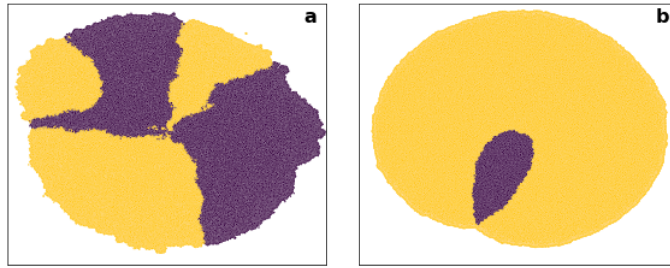
Figure 18. Spatial genetic patterns left behind by simulated diffusive expansions in a 2D habitat. **a.)** Neutral variation. The population started with 100 individuals who were a 50/50 mix between two neutral alleles (colors) and grew to just over one million individuals. **b.)** Selective variation. The population again began with 100 individuals, but this time one individual at the top of the initial colony had an advantageous mutation. The mutation allows the mutant sector (yellow) to grow faster than the wildtype sector (purple). This population grew to about 6.3 million individuals, just enough to capture the mutant sector enclosing the wildtypes.

and quantifying the distribution of fitness effects of new mutations; see, for example,

Ref. [66] and references therein. Mutations that appear in expanding populations may

thrive if they are fortunate enough to appear in an individual near the expansion frontier.

Even deleterious alleles are capable of surfing and taking over the front in 1D expansions,

particularly in populations emerging from a bottleneck [67]. They can also establish

sectors in 2D expansions, although sectors established by less fit alleles are transients

and will eventually be closed/terminated by the invasion of their fitter neighbors as the

expansion proceeds [35]. Sectors are often readily visible when there are a small number

of unique alleles, but even genomically realistic simulation studies show that recurrent

founder events at the expansion frontier leads to gradually decreasing fitness due to an

accumulation of deleterious mutations that are able to surf the wave of expansion [68].

"Fitness" as used here refers to monoallelic sector or colony expansion velocities; a fitter

allele will expand faster and win the race outwards in search of habitat or resources.

The faster growth of fitter alleles enables them to take over the expansion frontier in

1D expansions that begin from a linear inoculant and in 2D expansions that begin from

a droplet, where in both cases the initial population consists of a mixture of two alleles

with a fitness difference between them [35]; see the mutant sector enclosing the wildtype sector in Fig. 18**b**.

Fitness differences between alleles are often measured using their growth rates in well-mixed liquid environments. Strains are allowed to grow for several generations, either in separate environments or in a single test tube together if it's a competition experiment, and then the relative fitness is estimated as

$$s_{\mathrm{wm}} = g_1/g_2 - 1 \qquad\qquad (3.1)$$

where g_1 and g_2 are the growth rates of the two alleles. Well-mixed environments mean spatial structure is nonexistent and the population grows exponentially until it approaches the global carrying capacity; exponential growth has been shown to make selection more effective in terms of fixation probabilities of beneficial alleles [36]. In this case, an allele being "beneficial" simply means that it confers the ability to produce offspring more quickly.

However, the measurement and indeed definition of "fitness" can be context-specific. In spatially structured environments such as Petri dishes, where the population grows outwards diffusively in pursuit of nutrients, an allele's ability to spread across the landscape is of the utmost importance. It has recently been shown that spatially structured landscapes accelerate selection by founder events causing boundless outward propagation of one allele while cutting off the possibility of outward growth for the other for diffusively-growing populations such as yeast on a Petri dish [69]. Spatial assays of relative fitness [35] take advantage of the more rapid expansion of beneficial alleles to quantify the advantage enjoyed by those which are able to win the outward race for nutrients; one such measurement considers the opening angle of mutant sectors such as the faster-growing (fitter) yellow sector in Fig. 18**b**. See Ref. [69] for a detailed overview of different dynamics at the extremes of spatial structuredness.

Between the extremes of diffusively spreading populations (as on a Petri dish) and exponentially growing populations (as in a well-mixed liquid environment) lie populations that are capable of long-range dispersal across their environment. Offspring dispersal is a stochastic event; a probability distribution called the *jump kernel* quantifies how far offspring typically disperse from their parent's position, so any individual offspring's dispersal distance can be thought of as a random draw from the jump kernel. Long-range dispersal is common in nature; in particular, Ref. [20] showed that fat-tailed jump kernels that decay slower than exponentially with increasing distance are a good description for the dispersal of many plant species. Long-range dispersal can lead to "patchiness" in the spatial distribution of genotypes that mimics local adaptation but is really just the result of random dispersal events across the landscape, such as in the case of soft sweeps in populations with long-range dispersal [22].

The influence of random chance is increased in populations capable of long-range dispersal since founder events can occur throughout the landscape even at long distances from the densely occupied core, whereas the aforementioned gene surfing and genetic sectors are limited to form exclusively at the edge of colonies that expand diffusively. This can give rise to "survival of the luckiest" evolutionary dynamics instead of the naive picture of "survival of the fittest": an allele that disperses an offspring into vacant territory far beyond the core may be able to carve out a niche for itself and increase its frequency, even if it is deleterious.

The survival of mutations in expanding populations has been studied for populations that grow diffusively [35] and exponentially [36], but there is an outstanding need to study the outcomes of mutations that appear during invasions driven by long-range dispersal [37]. Range expansions driven by fat-tailed jump kernels are challenging to study because the growth is driven by rare long dispersal events; however, recent theoretical work has enabled quantitative predictions about the evolution of neutral diversity [23] and soft sweep patterns [22] in populations with long-range dispersal.

That work builds on Ref. [21], who determined the analytical forms of colony growth when dispersal distances are drawn from a power law jump kernel that decays as $J(r) \sim 1/r^{\mu+1}$, where the kernel exponent μ sets the "heaviness" of the tail of the jump kernel. Low values of μ lead to broad jump kernels and frequent long-range dispersal events, while higher values of μ create narrower jump kernels and shorter dispersal on average. Power law kernels are common in nature [20], and this form is useful because one can continuously traverse from diffusive colony growth when $\mu > d + 1$ (where d is the spatial dimension) all the way to the well-mixed extreme as $\mu \to 0$ by only adjusting the kernel exponent μ. Colony growth is fastest at low kernel exponents because offspring are more likely to disperse a long distance and land in unoccupied territory away from the core of the growing population. $\mu < d + 1$ is the jump-driven growth regime, where long distance dispersal events drive colony growth. A second division separates different asymptotic functional forms of colony growth within the jump-driven growth regime: colony growth approaches stretched exponentials at long times for $\mu < d$, while it approaches faster than linear power laws for $d < \mu < d + 1$.

There is also an outstanding need to study the outcomes of mutations that appear in quickly-growing populations where individuals can live for multiple generations. Ref. [36] studied the fixation probability of new mutations in exponentially growing populations (among other demographic scenarios), but their non-overlapping generations limit the speed at which low-frequency mutations can accelerate away from loss. Many types of organisms can survive for multiple years, such as perennial plants. If individual survival is determined by some type of density-dependent competition, long-range dispersal could transport a new or low-frequency mutation to vacant territory that virtually eliminates the possibility of loss at a time when the mutation would otherwise be at greatest risk of being lost.

We seek to address these knowledge gaps by studying the outcomes of fitness-affecting mutations that appear during range expansions driven by long-range dispersal.

We measure mutation outcomes in terms of survival rates, i.e. the probability that a mutation that appears during an expansion will still be present in the population much later. We also investigate how quickly new mutations are lost and explore the final frequencies of mutations that remain to form a lasting part of the population.

3.2 Methods

3.2.1 Simulation methods.

We simulated range expansions in an isotropic 2D landscape using SLiM [33]. The spatial landscape is continuous and does not impose any local deme structure. Simulations begin with 100 individuals randomly placed within a small circle around the origin. Density-dependent competition models the struggle for resources experienced by natural populations. We define the carrying capacity density ρ_K and the competitive interaction length scale r_i, which combine to form the local carrying capacity as $K = \rho_K \pi r_i^2$. We then define how individuals compete with one another: spatial competition is quantified using a Gaussian kernel with standard deviation r_i that extends out to distance $3r_i$. The competition strength exerted between a nearby pair of individuals is

$$c(d_{12}) = \exp\left[-d_{12}^2 / \left(2r_i^2\right)\right] \tag{3.2}$$

where d_{12} is the distance between them. We ignore competition between individuals separated by more than $3r_i$ because the Gaussian function values become negligible. The competition is reciprocal, so all pairs of individuals exert equivalent competition strengths on one another.

While K is the local carrying capacity parameter we feed into simulations, that number only captures the saturated population of a region with radius equal to the competition length scale r_i; in that sense it can be seen as a carrying capacity population scale. Individuals are aware of each other and compete with everyone within three length units r_i of their position; therefore the true number of competitors with which an individual can share a saturated region is $\rho_K \pi \left(3r_i\right)^2 = 9K \equiv K'$. The plots and

60

relevant interpretations in Section 3.3 will be in terms of K since that is the parameter we set and the joint parameter N_*/K (defined below) takes nice values, but we focus on K' as the true local carrying capacity in analyses which depend on the actual number of competitors one may face.

The Gaussian interaction function causes nearby individuals to compete more fiercely with each other than distant individuals. Each time step, everyone in the population measures the population density ρ in their immediate surroundings by summing up the competition strengths exerted by everyone within distance $3r_i$ and then dividing that total competition strength by the integral of the Gaussian interaction function over the entire interaction region to get a density in units of number per area (see Section 3.5.1 for further discussion). Their probability of surviving to the next time step is ρ_K/ρ or 1, whichever is lower, meaning individuals in not-full regions definitely survive and individuals in full regions may die.

All individuals in the population get a chance to produce offspring by cloning at the start of each time step. The numbers of offspring produced by each individual are random draws from Poisson distributions. Wildtypes produce one offspring per time step on average; mutants produce $1 + s$, where s is the fitness effect of the mutation. Offspring disperse a random distance and direction from their parents. Dispersal distances are drawn from the following jump kernel:

$$
J(r) = \begin{cases} A/r_i & r \leq r_i \\ (A/r_i)(r/r_i)^{-(\mu+1)} & r > r_i \end{cases}
\tag{3.3}
$$

where the normalization constant is $A = \frac{\mu}{\mu+1}$. Note that A equals the total probability of drawing a dispersal distance less than or equal to r_i, so it sets the balance between local and long-range dispersal attempts. Using power law jump kernels means that our colonies will grow according to the forms discovered by Ref. [21] and connects this work to the growing body of knowledge about evolution in populations with long-range

61

dispersal driven by such kernels [22, 23]. The dispersal direction is a random draw from the uniform distribution between zero and 2π. This jump kernel assumes that dispersal distances occur on the same length scale that governs spatial competition, with uniform probability of drawing disprsal distances anywhere between zero and the length scale r_i and decreasing probability of dispersing further away. Combining the two length scales also serves to simplify our parameter space by eliminating the need to introduce a separate length scale for dispersal.

A mutation with fitness effect s is introduced in a newborn individual at the end of the first time step when the population exceeds N_* individuals. Choosing the first individual to have the mutation at the end of this time step guarantees that individual at least one opportunity to produce offspring. The order of operations within any time step is as follows. The first step is offspring generation and dispersal. Then, everyone in the population measures the local population density in their immediate surroundings. After that, individuals survive or die probabilistically depending on the population density they measured in the previous step. All individuals simultaneously carry out the actions of each of the previous substeps. Finally, we check to see if the mutation has been lost or if the population has reached a predetermined cutoff size and terminate the simulation if either condition is true.

Density-dependent competition leads to a significant proportion of the population being replaced every time step, particularly in the densely occupied core of the colony. Since everyone produces an offspring each time step on average, a region that begins a time step saturated to K individuals has roughly K additional new offspring that attempt to establish within the region during the offspring generation process. Everyone will measure densities of approximately $2\rho_K$, leading to roughly a 50/50 chance of surviving to the next time step. The median ages at the end of simulations are always zero or one, but the distribution of ages has a long tail (maximum usually just over 20) since long-range dispersal can transport offspring to faraway vacant habitat where they

62

are *guaranteed* to survive for several time steps ($\sim$10) before their region saturates and they may be squeezed out and unable to compete for resources.

Simulations end when the mutation is lost or when the population reaches a predetermined cutoff size, typically 10^7 individuals. This cutoff size allows us to record five orders of magnitude of population growth, and is far enough above our values of N_* that there was ample time to get statistics on whether or not the mutation will be a lasting contributor to the expanding population. We report the results of 100 simulations at each individual set of parameters discussed here. We focus on $\mu < 3$ in order to explore the dynamics and mutation outcomes in jump-driven expansions, extending beyond what is currently known about beneficial mutations that appear in diffusive expansions [35]. We simulate neutral mutations and mutations with fitness effects at various orders of magnitude spanning from strongly deleterious to strongly beneficial. We sample such a variety of fitness effects in order to gain intuition about mutation outcomes throughout the space of possible advantages, from the common nearly neutral ones all the way out to rarely touched but still relevant extremes. We use $K = 100$ throughout this work, as that allows us considerable options for varying N_*/K without quite being computationally prohibitive. We restrict the parameters shown in the plots of Section 3.3 to focus attention on the salient takeaway from each figure.

3.2.2 Expectations. The small and initially compact population at the start of the range expansion is a bottleneck. It has been shown that proximity to a bottleneck can be a key factor determining the long-term fate of new mutations [67]. For this work, it is most convenient to measure proximity to the bottleneck in terms of N_*, the population size when the mutation appears. However, N_* alone does not take into account if a newborn individual should expect to have to fight for resources. We combine N_* with the local carrying capacity K to form the dimensionless mutation introduction time N_*/K as a unitless measure of how densely or sparsely occupied the landscape is expected to be. "Early-appearing" mutations occur at low values of N_*/K (order one or

63
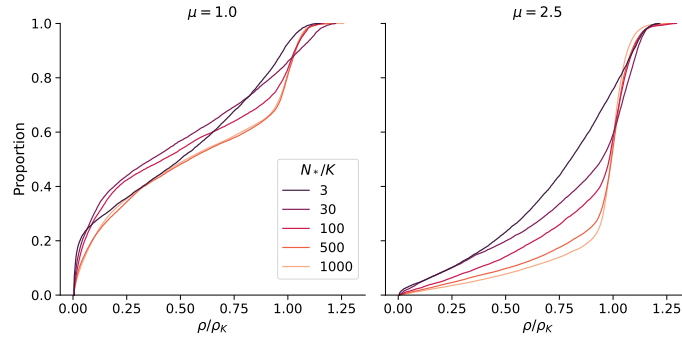
Figure 19. Empirical cumulative distribution functions (ECDFs) of the local population densities surrounding all newborn individuals in the first time step when the population size exceeds N_*. The initial mutant is a random sample from these newborns. We generated this data by running as few simulations as necessary to sample at least 10000 newborns. We got just over 10000 samples for $N_*/K \leq 100$ and roughly 50000 samples for the higher values. This required 1-6 simulations for the highest four values of N_*/K and about 50 simulations for $N_*/K = 3$.

lower), when the homeland has few individuals and individuals are unlikely to be killed by density-dependent competition. The mutations appear in a population that's growing outwards with nothing but vacant habitat beyond. "Late-appearing" mutations occur at high values of N_*/K ($N_*/K \gg 1$), when there are already many individuals competing for resources. The worst case scenario for a new mutation is to appear in the densely occupied core of the colony, where the existing population is largely replaced every time step and the life expectancy is short.

We can interrogate the population densities around newborns at time steps when the mutation is introduced as a crude measure to understand the landscapes that mutations are typically born into. This is the observable outcome that we hope to quantify with the ratio N_*/K. As shown in Fig. 19, initial mutants at low N_*/K in populations with broad jump kernels are likely to be born into sparsely inhabited areas, whereas later initial mutants become more likely to appear in saturated areas, particularly for narrower jump kernels.

We use the worst case scenario to form the lower bound on expected outcomes. Recall that the pre-existing population is largely replaced every time step in saturated

regions. The population size remains fixed at approximately K' individuals within the region, resulting in local dynamics that resemble the Moran model [70]. Unlucky mutations that appear in saturated regions must survive Moran-like local dynamics in their immediate vicinity if they wish to be a lasting part of the population. We estimate the mutation fixation probability using the Moran model following the logic that the mutation must fix locally somewhere in order to remain in the population indefinitely. The eventual fixation probability of a mutation that starts as one copy in a constant-sized population of K' individuals is given by

$$P = \frac{1 - (1+s)^{-1}}{1 - (1+s)^{-K'}} \tag{3.4}$$

where $1 + s$ is the fitness of mutants relative to wildtypes. We expect that the survival probability of a new mutation will be at least this fixation probability. See the fixation probabilities in Table 1.

The Moran model gives us a shorthand for establishing expectations about whether random genetic drift or selection should be the stronger force determining the outcome of mutations that appear in saturated areas. Random genetic drift is expected to dominate when $|K's| \ll 1$, where nonzero values of s are negligible and the fixation probability is approximately $P \approx 1/(K')$. On the other hand, selection should be the more important factor when $|K's| \gg 1$, where the fixation probability for beneficial mutations goes to $P \approx s$. Values of $K's$ are included in Table 1; we expect to see signs of selection continually getting stronger at the increasingly large values of $|K's|$ we've chosen. What remains to be seen is how the interplay of long-range dispersal and spatial structure affects the expected effectiveness of selection. Recent investigations into populations at the extremes of spatial structure [69] suggest that selective mutations will have better outcomes in populations with more spatial structure resulting from narrower jump kernels, but where and how and how quickly that transition occurs is unknown.

s	$K's$	Moran	Moran wt takeback	OW
-0.5	-450.0	1.18e-271	5.00e-01	NA
-0.2	-180.0	1.51e-88	2.00e-01	NA
-0.1	-90.0	7.31e-43	1.00e-01	NA
-0.05	-45.0	4.70e-22	5.00e-02	NA
-0.02	-18.0	2.59e-10	2.00e-02	NA
-0.01	-9.0	1.19e-06	1.00e-02	NA
-0.002	-1.8	3.96e-04	2.40e-03	NA
0.0	0.0	1.11e-03	1.11e-03	0.60
0.002	1.8	2.39e-03	3.97e-04	0.60
0.01	9.0	9.90e-03	1.29e-06	0.61
0.02	18.0	1.96e-02	3.64e-10	0.62
0.05	45.0	4.76e-02	4.25e-21	0.64
0.1	90.0	9.09e-02	5.58e-39	0.68
0.2	180.0	1.67e-01	1.09e-72	0.74
0.5	450.0	3.33e-01	1.65e-159	0.86

Table 1. Fixation probabilities at all values of s and corresponding $K's$ for two different models. Moran gives the Moran model fixation probability for a mutation with a single copy among a population of K' individuals, while Moran wt takeback gives the Moran model probability of a single wildtype eventually fixing among a population of K' individuals that are otherwise all mutants. OW is Otto and Whitlock's survival probability for mutations in exponentially growing populations, computed as the solution to Eq. (3.5).

The Moran model comes in handy again as a way of showing the importance of mutant establishment and local fixation. We can use it to quantify the inverse of the previous paragraphs: how likely are wildtypes to *retake* a region where the mutation has fixed? Now the relative fitness in Eq. (3.4) is the fitness of wildtypes relative to mutants, $(1 + s)^{-1}$. It's unlikely for wildtypes to take back a single region where the mutation has fixed for all but the most deleterious mutations. A mutation with a solid foothold in the population may quickly establish in mutiple areas. A single wildtype takeback is unlikely, but the probability of multiple separate wildtype takebacks is vanishingly small, highlighting the lasting importance of rapid establishment and growth of nascent mutations. The probability of wildtypes taking back a mutant region is included in Table 1.

An upper bound on expected outcomes is harder to establish. The closest theoretical work has been done by Ref. [36], who computed fixation probabilities in exponentially growing populations with non-overlapping generations. Their definition of fixation of an allele was the allele having descendants in the population after a very large amount of time has passed, analogous to our use of survival/survival rate verbiage since any mutants present at the end of our simulations must be descendants of the initial mutant. They extended the branching process approach of Ref. [71] to estimate the survival probability as the solution of

$$1 - P = \exp\left[-(1+s)(1+r)P\right] \tag{3.5}$$

where r is the exponential growth rate and s is assumed to be positive. We will compare our results against their survival probabilities since they offer a benchmark outcome at the well-mixed extreme, but keep in mind that their survival probability does not exactly apply to our simulations due to the different life expectancies/generation structures. Another key difference is that Eq. (3.5) comes from a well-mixed model that does not incorporate space, whereas the spatial landscape is an important player in our range
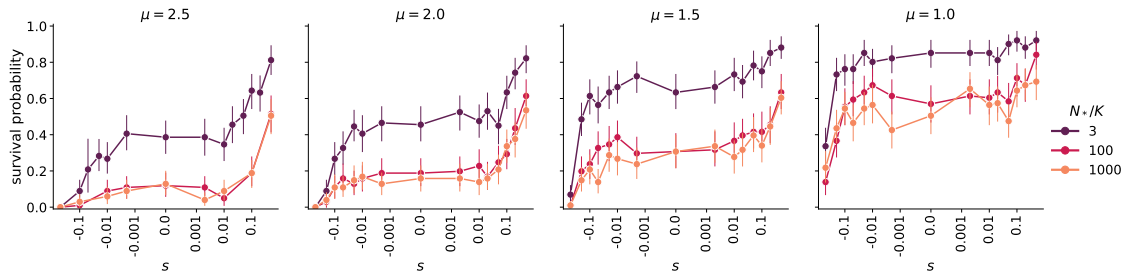
Figure 20. Observed survival rates versus mutation advantage s at three values of N_*/K. Points mark the observed proportion of simulations where the mutation was present in the population at the end of simulations, and error bars are the 95% bootstrap confidence interval over 100 replicate simulations at each parameter set. Note that x axes are scaled logarithmically on either side of $s = 0$. Kernel exponents are arranged in descending order as you move from left to right so that long-range dispersal increases as you move to the right.

expansions. Population growth approaches the well-mixed extreme as $\mu \to 0$, but such broad jump kernels lead to boundless dispersal across the landscape and unrestricted growth as offspring are always able to disperse to locations devoid of competition. The closest comparison we make here is for fast expansions with $\mu = 1$, where we estimate the survival probabilities reported in Table 1 by fitting for an effective growth rate using data from the first few generations of population growth (see Section 3.5.3 for further details).

Non-overlapping generations limit the speed at which low-frequency mutations can accelerate away from loss, whereas individuals can live for many generations in our simulations if they happen to be born in a sparsely inhabited area. Long life expectancies for newborns in relatively vacant habitat can dramatically improve the survival probability of a new mutation if it is fortunate enough to appear there, since it is protected against loss when its frequency is low and it would otherwise be at great risk. We expect to see the greatest survival rates for mutations that appear early in expansions driven by broad jump kernels with low exponents μ. Those survival rates may even surpass what was predicted by the spaceless model of Ref. [36].

3.3 Results

3.3.1 The value of s often does not affect survival rates. The first question of interest is whether or not a mutation that appears during a range expansion will remain to form a lasting part of the population. We measure this in terms of survival rates, the observed proportion of simulations where the mutation is still present in the population at the end of the simulation. We found that a new mutation's fitness effect has no measurable impact on its survival probability for all but the most extreme values of s we explored, as shown in Fig. 20. Extremely beneficial and deleterious mutations have higher and lower survival rates, respectively, but a flat horizontal line can be made to fall within the error bars for all or nearly all values of s within ± 0.05 at any $(\mu, N_*/K)$ pair, which corresponds to $|K's| \leq 45$.

Two spatial factors seem to be the key players determining survival probability: the ability of new offspring to disperse across the landscape, as measured by the kernel exponent μ; and the landscape occupancy when the mutation appears, as measured by N_*/K. The dispersal dependence is visible by looking left to right across the panels of Fig. 20. Survival rates increase as jump kernels get broader (as μ decreases) regardless of the other parameters. Increased ability to disperse offspring long distances across the landscape increases the odds that a nascent mutation will be able to surf the outward expansion and survive to play a longstanding role in the population's evolution.

Early-appearing mutations tend to fare better than those that appear later, as is evident by looking at the different colored survival rates within each panel of Fig. 20. The initial population can be thought of as emerging from a population and spatial bottleneck, so individuals born within the first few generations see nothing but vacant habitat as they look outwards and their descendants may have a large contribution to the final population. While evidently early is best, specifically how late "late-appearing" mutations appear has no impact on their survival probabilities. This is visible based on the indistinguishable survival rates between the two higher values of N_*/K within each

69
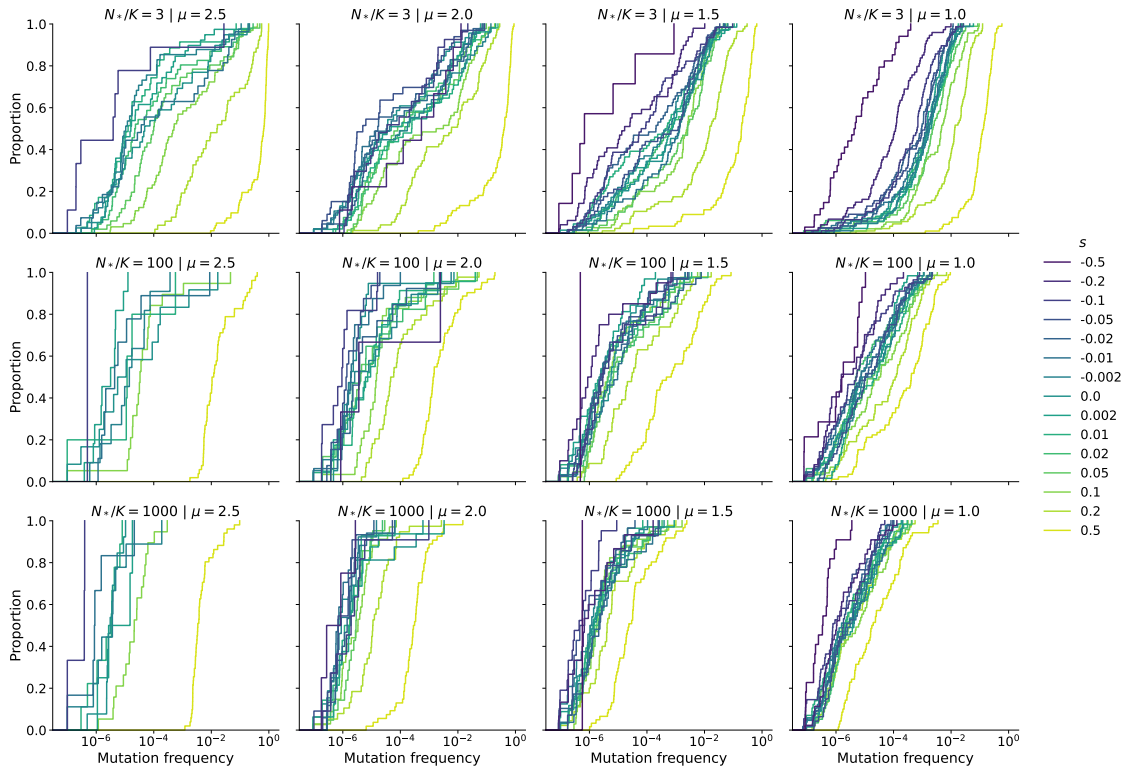
Figure 21. ECDFs of the frequencies of surviving mutations.

panel of Fig. 20: there is no difference in mutation survival rates despite the order of magnitude difference in population size when the mutation appears.

3.3.2 The value of s often does not affect frequency distributions of surviving mutations.

The second question of interest is what frequency may a mutation achieve given that it does survive? Is it present in the population but at very low frequency, teetering at the precipice of loss, or has it been so successful that it may eventually reach fixation? Returning to the Moran model expectations (Section 3.2.2) and local fixation probabilities (Table 1), we expected to see signals of selection continually growing stronger at the increasingly extreme values of $|K's|$ we used in our simulations. The expected signal would be that more beneficial mutations should reach generally higher frequencies, and similarly more deleterious ones should have lower frequencies. We see elements of that signal in the empirical cumulative distribution functions (ECDFs)

70

N_*/K	μ	p value	adjusted p value
3	1.0	0.002576	0.006623
3	1.5	0.008721	0.019027
3	2.0	0.053766	0.107532
3	2.5	0.007886	0.017742
100	1.0	0.516170	0.589909
100	1.5	0.306230	0.424011
100	2.0	0.285060	0.410486
100	2.5	0.200000	0.320000
1000	1.0	0.659620	0.730656
1000	1.5	0.924980	0.946475
1000	2.0	0.298930	0.422019
1000	2.5	0.259420	0.381189

Table 2. p values and adjusted p values from K-sample Anderson-Darling tests on the ECDFs of surviving mutation frequences for the mutations with $|s| \leq 0.05$ shown in Fig. 21. p values were adjusted according to the procedure of Benjamini and Hochberg [72] which controls the false discovery rate.

of frequencies of surviving mutations, shown in Fig. 21. The empirical distributions of mutation frequencies are often shifted higher or lower for the strongest beneficial or deleterious mutations, respectively, where $|K's| \gtrsim 100$.

However, the distributions of frequencies for mutations closer to neutral appear less sensitive to values of s over a wide range, mirroring the insensitivity of survival rates to s shown in Fig. 20. To identify the mutation effect sizes at which differences in frequency distributions are meaningful, we performed K-sample Anderson-Darling tests [73] on the ECDFs of the surviving mutation frequencies for all mutations with various ranges of $|s|$. The null hypothesis being tested is that the observed frequencies at each value of s are samples from some common (unspecified) distribution; rejecting the null hypothesis in this case would mean that the data we measured would be highly unlikely if the samples were really all from some common distribution (where the definition of "highly unlikely" comes from the threshold p value at which we decide to reject the null hypothesis). In Table 2, we show p values from tests that used mutations with $|s| \leq 0.05$. The p values show that we can only reject the null hypothesis for a few

values of μ and the earliest-appearing mutations. Intuitively, this means that nonzero values of s within $|s| \leq 0.05$ don't have any meaningful impact on the frequency a mutation is expected to achieve. The tests show us that fitness effects become significant when mutations get even stronger; an extended table showing results from all tests is presented in Section 3.5.3, Table 3. Incorporating more extreme values of s into the tests leads to the null hypothesis being rejected at $8/12$ $(N_*/K, \mu)$ pairs when mutations with $|s| \leq 0.1$ are used and at all parameter combinations when mutations with $|s| \leq 0.2$ or $|s| \leq 0.5$ are used.

We should note that there is a clear theoretical limit that can be considered here: the expected frequency under a well-mixed model. Studying frequencies with respect to that limit tells us about the influence of space and dispersal across the landscape on mutation outcomes; see Section 3.5.2 for discussion relative to the well-mixed limit. A fact we will mention here in the main text is that the frequency attained by deterministically growing mutants in a well-mixed model with the same final population size does not vary strongly when $|s| < 0.05$, suggesting that the simulations have not run long enough for the ECDFs in Fig. 21 to be well-separated from each other at $\mu = 1$. It is possible that the null hypothesis will be rejected if the range expansion runs for longer, as small differences in growth rates between the wild type and the mutant accumulate over longer times. However, space has been shown to accelerate selection relative to well-mixed environments [69], which might lead one to believe that we should observe greater separation between the ECDFs with $|s| \leq 0.05$ at the higher values of μ. The tightly-packed ECDFs we observe in our data defy that expectation.

3.3.3   Effects of long-range dispersal on survival rates.   We now return to observed mutation survival rates and directly consider how different amounts of long-range dispersal affect mutation outcomes; this perspective also allows us to revisit the theoretical expectations discussed in Section 3.2.2. Recall from Fig. 20 that broad jump kernels and early appearances promote mutation survival, while having late versus
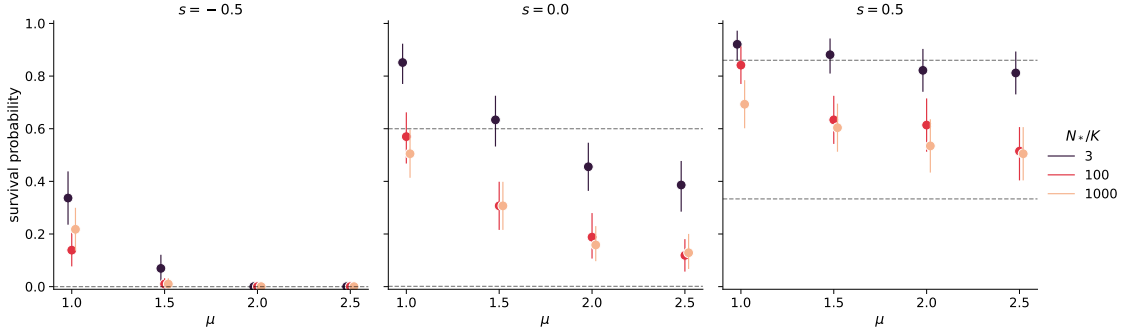
Figure 22. Observed mutation survival rates plotted against kernel exponent μ for neutral and very strong mutations. The upper dashed line marks the survival probability under the exponential growth model of Ref. [36], and the lower dashed line marks the fixation probability under the Moran model as computed in Section 3.2.2. The two lines are indistinguishable at $s = -0.5$ where survival is unlikely. Points mark the observed proportion of simulations where the mutation was present in the population at the end of simulations, and error bars are the 95% bootstrap confidence interval over our 100 replicate simulations.

later values of N_*/K doesn't seem to make a difference. Those points are reiterated in Fig. 22, which reports survival probabilities grouped by values of s, allowing us to compare against the theoretical expectations mentioned above. Survival rates are highest at the broadest jump kernels and decrease as kernels get narrower. The early-appearing mutations with $N_*/K = 3$ survive at much higher rates than late-appearing mutations, and the survival rates of the late-appearing mutations are approximately equal despite the order of magnitude separating their values of N_*/K. The horizontal dashed lines in each panel come from the expectations outlined in Section 3.2.2.

The fixation probability from the Moran model [70] only depends on s and K' (Eq. (3.4)) and forms the lower bound on expected survival rates. That baseline assumes the mutation first appears in a region saturated to K' individuals and considers the outcomes only in terms of local dynamics within that region. Survival rates are always at least as high as the Moran model fixation probability (lower dashed line in each panel of Fig. 22), even for mutations that appear late in populations with the narrowest jump

kernels shown here. Those mutations are likely to appear in fully saturated regions
(Fig. 19), but they persist at higher rates than the Moran model predicts.

The survival probability under the exponential growth model of Ref. [36] depends
only on s and the population growth rate, see Eq. (3.5). The sparsely inhabited landscape
and lack of competitors can allow early-appearing mutations in populations with broad
jump kernels to persist at higher rates than what is predicted by the model of Ref. [36],
whose survival probability is plotted as the upper dashed line in each panel of Fig. 22.
Late-appearing mutations in populations with broad jump kernels survive at rates
approaching the survival probability from Ref. [36], suggesting that it may be a useful
upper bound on expected survival rates for beneficial mutations that appear later into
fast expansions. We also see that even highly deleterious mutations can remain in the
population indefinitely if the mutation happens to appear near the beginning of a range
expansion in a population with a broad jump kernel.

3.3.4 Mutations can persist indefinitely at small population sizes.

The Moran model fixation probabilities of Table 1 show the likelihood of the mutation
reaching fixation among a population of K' individuals and the probability of wildtypes
taking back a region where the mutation has fixed. We interrogated our simulation
results in search of signals of interesting dynamics following local fixation by exploring
the data conditional on whether or not the mutant population was ever greater than or
equal to K' individuals at the end of any single simulation time step. We use whether
or not the mutant population size reached K' individuals as a minimal proxy for
local fixation since our continuous spatial landscape doesn't have any demes or other
boundaries we could use to cleanly impose spatial structure or discriminate between
"local" and non-local events.

In Fig. 23a, we show that most mutations fail to reach population sizes of K'.
A mutation's probability of reaching a population size of K' depends on N_*/K in a
similar way to what was observed in the survival rates of Figs. 20 and 22: early-appearing

74
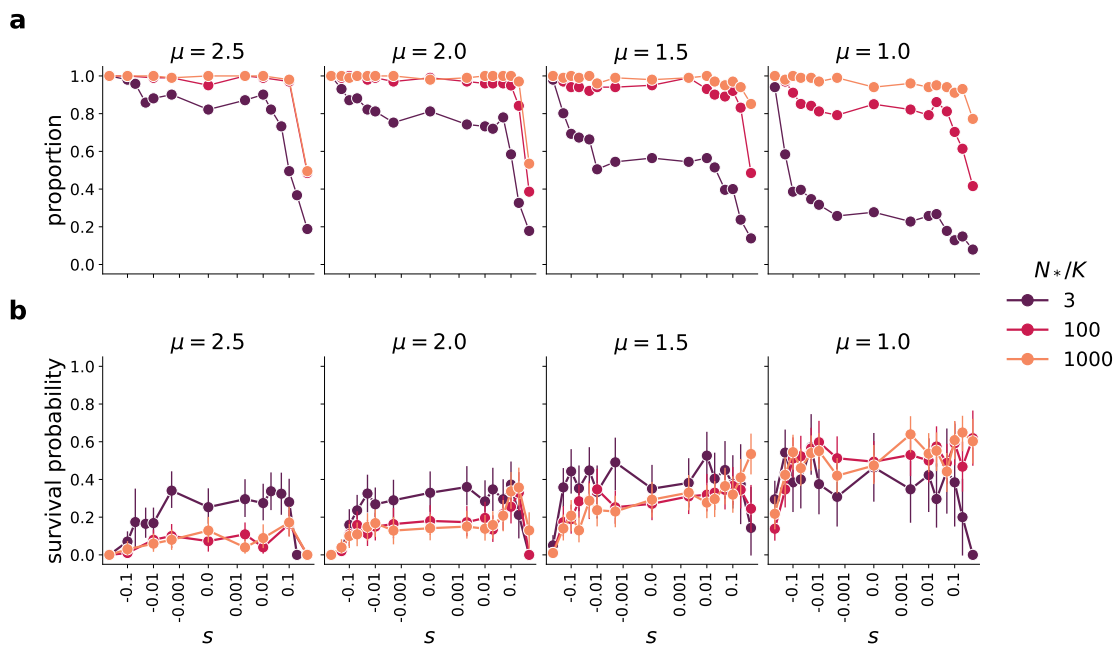
Figure 23. Mutation outcomes in simulations where the mutant population never reached K' individuals. **a.)** The proportion of simulations where the mutant population never reached K' individuals. **b.)** Mutation survival rates in simulations where the mutant population never reached K' individuals. Error bars are the 95% bootstrap confidence interval.

75

mutations are more likely to reach a population size of K', while the probability of doing so doesn't appear to depend on N_*/K at the two higher values despite the order of magnitude between them. There is a sharp increase in the probability of reaching population sizes of K' for the most beneficial mutations, mirroring the improved survival rates for the most beneficial mutations shown in Fig. 20. Reaching K' individuals seems to be a high and unlikely level of success for most mutations, and it seems to guarantee the survival of mutations that are able to do it; we never observed mutations being lost after reaching population sizes of K' individuals.

In Fig. 23b, we show the survival rates of mutations which never reached population sizes of K' individuals. Never reaching K' individuals means that the mutation could not have locally fixed anywhere in the landscape, so it would be destined for loss based on the Moran model expectations that ignore the expanding range (Section 3.2.2). However, we find that it is increasingly possible for the mutation to survive without ever being present in K' individuals as jump kernels become broader, especially for later-appearing mutations. Evidently more long-range dispersal across the landscape supports a mutation's effort to remain in the population, even if it is never able to locally fix anywhere. Perhaps being present at low to intermediate frequencies in multiple regions provides enough defense against loss and sufficient offspring production and dispersal attempts that some mutations can persist indefinitely without ever taking over one single region of space.

3.3.5 Most mutations to be lost are lost quickly. Statistics about mutation survival rates and possible mutation frequencies are influenced by the simulation parameters, in particular the cutoff population size at which simulations are terminated (recall, we use 10^7 individuals as that cutoff throughout this paper). It is possible to learn about how observed survival rates may change in shorter or longer range expansions by interrogating the population sizes at which mutations are lost, as explored in Fig. 24. Most mutations that are lost disappear from the population within the first

76
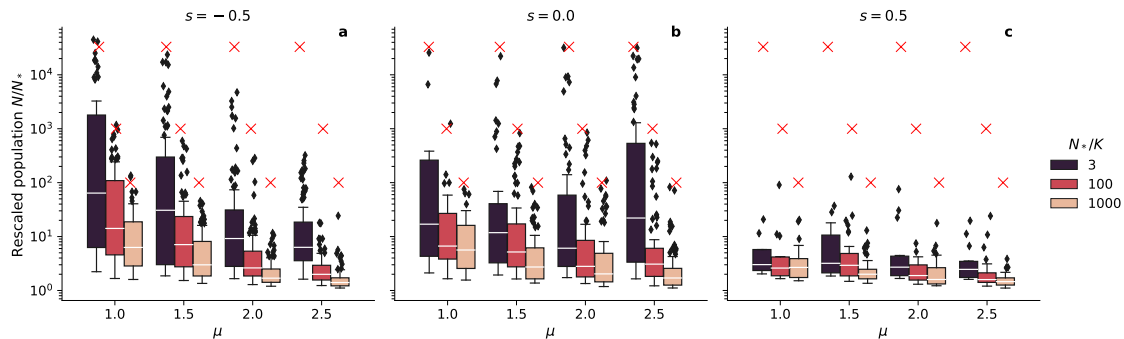
Figure 24. Rescaled population sizes at the time step when mutations are lost from their respective populations. The population size N is rescaled by dividing by the population size N_* when the mutation was introduced; the $\times$ symbols mark the rescaled cutoff population sizes at which simulations are terminated: $10^7/N_*$.

two orders of magnitude of population growth after they appear (all medians fall below 100 in Fig. 24). All extremely beneficial mutations that will be lost are lost more or less immediately after they appear – their fates are decided right away (Fig. 24**c**).

The picture is slightly more nuanced for neutral and strongly deleterious mutations (Fig. 24**a**–**b**). In both cases, there are (μ, N_*/K) pairs where many points are grouped at or just below the cutoff population size. These points suggest that our observed survival rates are transients and still decaying, since there's nothing special about the cutoff population size we chose. Longer simulations could result in notably lower survival rates at those parameter combinations if mutations are still being lost from those populations at a steady rate. Parameter combinations with few or no points near the cutoff population size have likely shown us something close to the "true" survival probability at those parameters; the same mutations would be lost in shorter simulations, and we assume that survival rates would be nearly identical in slightly longer expansions since there's nothing special about our cutoff.

3.4 Discussion

A new mutation's fitness effect has a far weaker impact on survival probabilities (Fig. 20) and mutant frequencies (Fig. 21) than would be expected if fixation within a local density regulation region were necessary for long-term survival. As discussed in

77

Section 3.2.2, we expected to see signs of selection continually becoming more effective as values of $|K's|$ became larger; instead, space seems to be the key factor determining possible mutation outcomes in two ways.

First, mutations that appear early (N_*/K values of order one) appear in individuals who have relatively few competitors and see nothing but vacant habitat beyond them. It pays dividends to appear early, close to the spatial and population bottleneck that is the beginning of the expansion. The specific time at which the mutation appears ceases to matter once $N_*/K \gtrsim 100$. Much of the population growth happens just beyond the core in regions that will soon be absorbed into it, meaning that even mutations that appear to be born away from the core population may quickly be surrounded and thwarted by wildtypes. Regardless of where in relation to the core a new mutation appears, it becomes more likely for juveniles to be born into locally saturated areas as N_*/K increases, particularly for narrower jump kernels as shown in Fig. 19. The primary struggle for a new mutation will be whether or not it can establish *locally*, which has nothing to do with the *global* population size. For later-appearing mutations, the key takeaway seems to be that the local dynamics, such as what is summarized by the Moran model in Eq. (3.4) and Table 1, need to be considered as an important factor even in populations with long-range dispersal.

Second, mutations are more likely to survive in populations with broader jump kernels (Fig. 22), even if they are never able to locally fix anywhere (Fig. 23b). More long-range dispersal helps the population spread faster and enables offspring to disperse long distances in search of vacant habitat. Newborn individuals in populations with broader jump kernels are generally born into less densely occupied areas than those in populations with narrower jump kernels (Fig. 19), increasing the odds that a nascent mutation could establish a solid foothold in the population before experiencing fierce competition from wildtypes. That being said, we do see signs of selective mutations' outcomes improving relative to neutral at lower values of s for narrower jump kernels

(see Fig. 20); we also show in Fig. 27 that the highest frequencies at any given $(N_*/K,$ $s)$ pair are often achieved by populations with narrower jump kernels. These results are in line with the findings of Ref. [69] that spatial structure enhances selection. Given that a mutation isn't lost soon after appearing (as is the most common case; Fig. 24), it may soar to the highest heights in populations with narrower jump kernels if all else is equal.

Our simulation results provide insights into mutation outcomes in populations for which there is no existing theory. We showed that unlucky mutations can be likely to appear in fully saturated regions even in populations with long-range dispersal, necessitating the Moran model fixation probability as a worst case lower bound on expectations. This scenario is most relevant in populations with narrower jump kernels as shown in the right panel of Fig. 19. For fast expansions with broad jump kernels, the closest theoretical work to date was done by Ref. [36], who computed the survival probability for mutations that appear in exponentially growing populations with non-overlapping generations. Our simulations with density-dependent mortality provide a mechanism by which individuals can survive for several time steps if they are fortunate enough to be born in a relatively empty area. This effect launches survival rates above the survival probability from Ref. [36] for mutations that appear early in populations with broad jump kernels. Even if we don't know the exact breadth of a jump kernel's power law tail or exactly how early a mutation appeared during an expansion, Ref. [36]'s survival probability provides a good starting point for estimating mutation survival probabilities. The survival rates we observed were always of the same scale as the probabilities we computed according to the procedure outlined by Ref. [36], which would make that value a worthwhile first estimate for survival probabilities in e.g. experimental populations, where there may be more unknowns and messier conditions than our idealized simulations.

Neither the Moran model nor Ref. [36] consider the influence of space or the expanding range on mutation outcomes. Later-appearing mutations become more likely

to appear in saturated regions, particularly in populations with narrower jump kernels (Fig. 19), meaning a new mutation will have to compete against an otherwise all-wildtype local population filled to the local carrying capacity in order to survive. The Moran model fixation probabilities set our baseline expectations for mutations that appear in saturated areas, and they are often quite low (Table 1). In the Moran model, any mutations that don't reach fixation are lost, so the probabilities in Table 1 suggest that we should see very few surviving mutations at many parameter combinations. However, even though the local fixation probabilities are often low and we rarely see signals of possible local fixation (Fig. 23**a**), we often observe high survival rates that are constant across a wide range of fitness effects s (Fig. 20). These higher than expected survival rates seem to be a result of the population's expansion into unoccupied territory. Mutations that appear in recently colonized areas may experience less competition than those that appear in the densely occupied core and therefore be able to quickly proliferate due to the decreased competition, analogous to the gene surfing that can occur during diffusive expansions. Mutants may take advantage of long-range dispersal to travel outwards from partially inhabited area to partially inhabited area, making it possible for the mutation to survive indefinitely despite never locally fixing anywhere (i.e. the surviving mutations which generated the data plotted in Fig. 23**b**).

We have shown that whether or not a new mutation persists in a population is independent of the effect the mutation confers on its carriers for the most common (non-extreme) range of fitness effects. This may encourage different mutations to become prominent in different habitats that have been invaded by the same species, since different mutations will stochastically appear in individuals in each site and any new mutations are likely to be functionally neutral in terms of their survival probability and expected frequency if they do survive (i.e. within the flat section with $|s| \leq 0.05$ in Fig. 20, the same range of s values that were not statistically significant from neutral as shown in Table 2). This structure could possibly be detected by population genetic measures of

differentiation between populations, such as F_{ST}, although it may be difficult to pin said variation on expansion dynamics rather than other factors (like generic isolation by distance).

In this work, we restricted our focus and studied only mutations that confer a global fitness effect by changing the offspring production rate. A local competitive advantage could provide a different mechanism for a mutation that appears during a range expansion to remain and/or become prominent in the population by means of increasing the survival probability of mutants relative to wildtypes in saturated areas. Incorporating local competition into the model could lead to surprising new outcomes. Ref. [74] recently explored how a slower-expanding microbial strain can beat a faster strain and take over the front of diffusive range expansions. In 2D expansions analogous to what we report here, the slower strain invades by forming a dented sector that eats into the faster strain's sector. Intuition about the "fastest runner winning the race" breaks down because the slower strain has a local advantage over the faster strain and can invade laterally into what would otherwise be the faster strain's territory, despite the slower strain's decreased radial expansion. An exciting future research topic will be to incorporate local competition effects, as that could illuminate pathways to evolution for populations that are unlikely or unable to evolve via global fitness differences alone.

Mutations can also influence fitness by changing the dispersal of those who carry them. This scenario has clear ecological motivations, such as plants evolving to produce seeds that are more easily carried long distances by the wind. Changes to dispersal rather than (or in addition to) fecundity could readily be incorporated into our simulations to study evolution in populations with mutations that affect dispersal. One should carefully consider the populations they hope to model with such changes, and carefully account for any costs associated with changes that affect dispersal. For example, cane toads have evolved longer legs which accelerate their invasion of Australia [10], but there is a metabolic cost to fueling the bigger muscles in longer legs that leads to

81

diminishing returns on fitness beyond some threshold leg size. This sort of trade-off must be acknowledged in studies that hope to model results of physical changes in organisms.

Another interesting future direction would be to study outcomes in populations where multiple mutations could appear stochastically during the expansion rather than having exactly one mutation appear during the expansion at a predetermined time as we have done here. Our simulations model the spread of just one allele in an otherwise homogeneous population, such as the simplified picture of the alpha variant becoming the prominent strain of SARS-CoV-2 in the relatively early stages of the Covid-19 pandemic when most people got infected with the original strain of the virus. But that picture is obviously not the full story; the delta variant later became the prominent strain, then the omicron variant quickly spread worldwide. What were the factors that caused those strains to spread around the world so quickly? How did the relative fitness of each strain combine with the random location that it first appeared and the dispersal by infected people to overcome all other pre-existing and parallel-appearing strains of the virus? This work is a minimal case that will build intuition and basic knowledge to later inform studies of populations where multiple fitness-affecting mutations could appear as the population expands into new territory.

3.5   Supplementary Information

3.5.1   Simulation details.   Simulations begin with 100 individuals randomly placed near the origin. The x and y positions of each individual were random draws from Gaussian distributions with mean zero and a capacity-dependent standard deviation; we used $\sigma = 0.4 r_i$ for the simulations with $K = 100$ reported throughout the main text. We sought to create a compact initial population so that individuals would immediately feel the effects of density-dependent competition rather than a sparsely inhabited landscape where there was little competition and unconstrained growth for the first several time steps. The landscape is large enough to be functionally boundless; the periodic boundary conditions have no effect.

Local population density measurements begin by the focal individual summing up the competition strengths exerted by everyone within distance $3r_i$; those competition strengths are computed according to the Gaussian competition function Eq. (3.2). That total competition strength is then divided by the integral of the interaction function over the entire interaction region:

$$\int_0^{2\pi} \int_0^{3r_i} r \exp\left[-r^2/\left(2r_i^2\right)\right] \, dr \, d\theta \approx 6.21339$$

The integral of the interaction function represents something like the total "interaction field" present around the focal individual [75], and dividing the total competition strength felt by the focal individual by the integrated interaction function returns a population density value in the correct units that accounts for increased competition between nearby individuals relative to that between distant pairs of individuals.

Dispersal distances were drawn using inverse transform sampling. Recall, the jump kernel is

$$J(r) = \begin{cases} A/r_i & r \leq r_i \\ (A/r_i)(r/r_i)^{-(\mu+1)} & r > r_i \end{cases}$$

where the normalization constant is $A = \frac{\mu}{\mu+1}$; A also equals the total dispersal probability with r_i. The sampling procedure begins by drawing a random number X from the uniform distribution between zero and one. If $X \leq A$, the offspring will land within r_i of their parent. The dispersal distance is computed by integrating the jump kernel from zero until the distance D at which the total integrated probability equals X. For short-range dispersal draws when $X \leq A$, this gives

$$X = \int_0^D A/r_i \, dr = \frac{AD}{r_i} \rightarrow D = \frac{Xr_i}{A}$$

$X > A$ corresponds to a long-range dispersal event drawn from the tail of the distribution; now we must integrate both terms of the jump kernel. This gives

$$X = A + Ar_i^\mu \int_{r_i}^D r^{-(\mu+1)} \, dr = A - \frac{Ar_i^\mu}{\mu} \left(D^{-\mu} - r_i^{-\mu}\right)$$

which leads to a distance of

$$D = \left(\frac{r_{\mathrm{i}}^{-\mu}(A + A\mu - X\mu)}{A} \right)^{-1/\mu}$$

The time and memory requirements vary between simulation parameters; the following amounts of each reflect the resources we requested for each individual simulation run on UO's computing cluster. Recall that mutation loss immediately ends simulations; simulations where the mutation is immediately lost require little time or memory. In general, populations with higher values of the local carrying capacity K require more memory, and those with narrower jump kernels (higher μ) require more time. We requested 700 Gb in our batch submission files when running the simulations with $K = 100$ reported here. The actual memory used was often far less, but that request gave us enough of a buffer to run high-memory simulations to completion without running out of memory. We requested up to 2.5 hours for the fast simulations at $\mu = 1$ and 36 hours at $\mu = 2.5$. Again, these requests often overshoot the actual duration of simulations in order to make sure we don't run out of time on the computers when simulations happen to be longer than usual.

3.5.2 Well-mixed expectations. Spatial structure becomes irrelevant in a well-mixed environment. All offspring are expected to have access to the nutrients and other resources they need, so all offspring survive. The population grows exponentially, approximately doubling in size every generation.

More specifically, wildtypes are expected to produce one offspring per time step. Parent and offspring are assumed to survive since the well-mixed environment eliminates spatial structure and therefore also the density-dependent competition described throughout this paper. The wildtype population grows as

$$w(t) = w_0 2^{t-1}$$

where the initial wildtype population is $w_0 = 100$ and $t - 1$ reflects that time starts at generation 1.
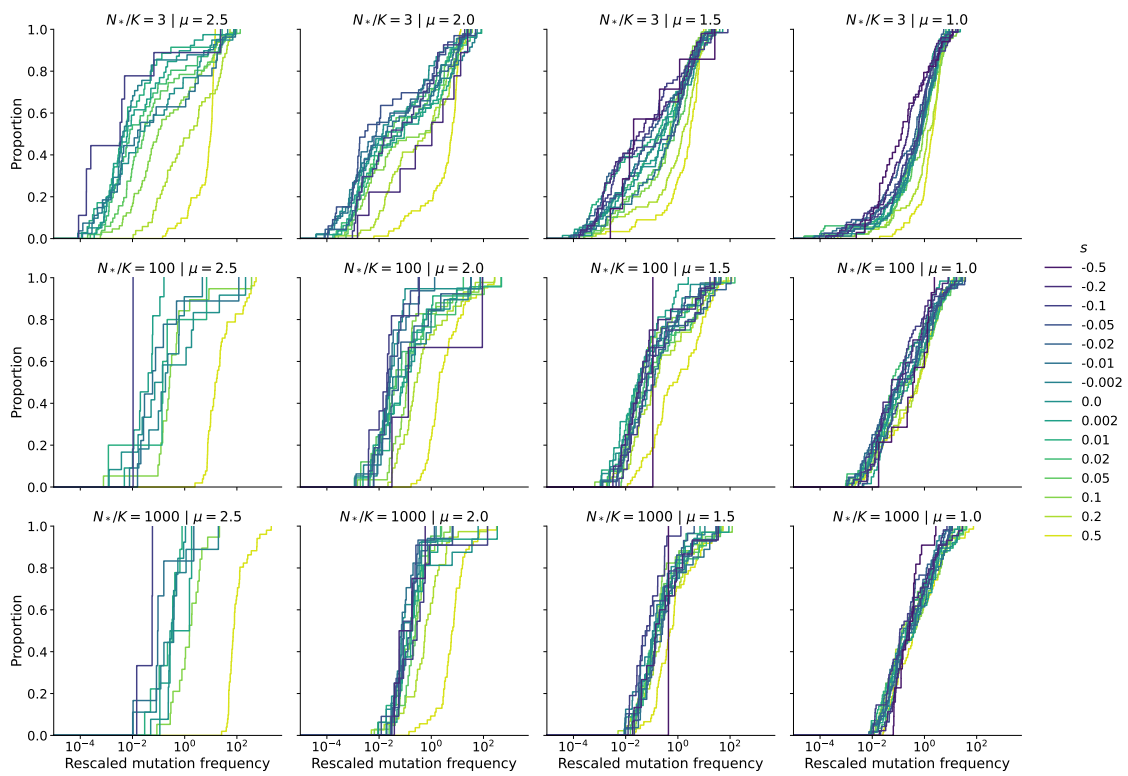
Figure 25. ECDFs of the frequencies of surviving mutations, rescaled by the expected frequency under a well-mixed model. The unscaled frequency distributions are shown in Fig. 21

The mutation affects offspring production; mutants are expected to produce $1 + s$ offspring per time step. The mutant population grows as

$$m(\tau) = (2 + s)^\tau$$

where τ is the number of generations elapsed since the mutation appeared, starting from 0 when the mutation appears in one individual.

We computed the expected frequencies for the rescaling of Fig. 25 by considering the exponential growth of wildtype and mutant populations separately. No one dies due to density-dependent competition since we're in a well-mixed environment, so we can simply compute the wildtype growth trajectory $w(t)$ and mutant growth trajectory $m(\tau)$, determine the relationship between t and τ by seeing when $w(t)$ exceeds N_*, find the first time step T when the total population $w+m$ exceeds 10^7, and then compute the expected mutation frequency as $m/(w + m)$.

The rescaled ECDFs of Fig. 25 reveal that well-mixed models accurately predict the frequencies of surviving mutations in populations with broad jump kernels; see how the distributions fall on top of one another and rise sharply near the expected frequency (where the rescaled frequency equals one) in the panels with $\mu = 1$. Agreement with the well-mixed model tells us that the spatial landscape doesn't play a big role in the outcomes of *surviving* mutations.

The empirical distributions spread apart from each other and away from the well-mixed expectation when the jump kernels are narrower. The shift away from the expected frequency tells us that now the environment and the population's ability to disperse across it are playing an important role. Furthermore, there seems to be an interplay between dispersal and fitness effect; strongly beneficial mutations have higher frequencies than other mutations and can even reach expected frequencies orders of magnitude above the well-mixed expectation in qualitative agreement with the trends reported by Ref. [69].
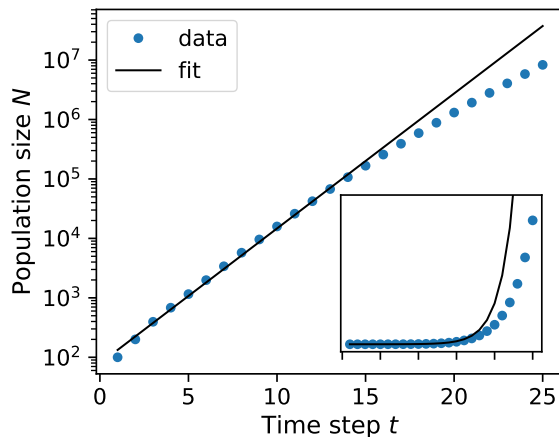
Figure 26. An example fit used to estimate an effective exponential growth rate for neutral expansions with $\mu = 1$ and $K = 100$. We fit for the early growth rate by using only the first 15 time steps in the fits. The data shown are the average population growth across 176 independent simulations. The same data and fit are plotted in linear space in the inset, and the x axis ticks on the inset match those on the main x axis.

3.5.3 Fits and statistics. An estimate of the population's exponential growth rate is required to compute a new mutation's fixation probability according to the procedure derived by Ref. [36]. Population growth approaches exponential as $\mu \to 0$, but in general growth follows a stretched exponential form for $\mu < d$ [21]. The fastest expansions we simulated had $\mu = 1$. We estimated an effective exponential growth rate using the average population growth by finding the best fit line in semilog space, where the growth rate is extracted as the slope of $\log N(t) = At + B$. We used data from the first 15 time steps in order to capture the average growth experienced by mutations that appear at various values of N_*, thus giving us a reasonable growth rate to plug into Eq. (3.5). An example of this procedure is shown in Fig. 26. We determined this level of accuracy to be satisfactory considering that the model of Ref. [36] does not exactly apply to our populations and we're solely using their fixation probability as a benchmark to calibrate our expectations and interpretations.

s threshold	N_*/K	μ	p value	adjusted p value
0.01	3	1.0	8.482500e-01	8.851304e-01

s threshold	N_*/K	μ	p value	adjusted p value
0.01	3	1.5	3.439800e-01	4.586400e-01
0.01	3	2.0	4.238400e-01	5.237400e-01
0.01	3	2.5	2.251700e-01	3.524400e-01
0.01	100	1.0	3.437100e-01	4.586400e-01
0.01	100	1.5	4.338800e-01	5.237400e-01
0.01	100	2.0	1.521900e-01	2.632800e-01
0.01	100	2.5	2.000000e-01	3.200000e-01
0.01	1000	1.0	7.838800e-01	8.423785e-01
0.01	1000	1.5	9.333300e-01	9.464755e-01
0.01	1000	2.0	6.565700e-01	7.306560e-01
0.01	1000	2.5	2.594200e-01	3.811886e-01
0.02	3	1.0	1.257600e-01	2.321723e-01
0.02	3	1.5	1.454600e-01	2.618280e-01
0.02	3	2.0	4.202100e-01	5.237400e-01
0.02	3	2.5	3.535500e-01	4.628291e-01
0.02	100	1.0	4.364500e-01	5.237400e-01
0.02	100	1.5	4.912300e-01	5.704606e-01
0.02	100	2.0	4.479800e-01	5.287633e-01
0.02	100	2.5	2.000000e-01	3.200000e-01
0.02	1000	1.0	7.415300e-01	8.089418e-01
0.02	1000	1.5	9.652600e-01	9.652600e-01
0.02	1000	2.0	8.289100e-01	8.776694e-01
0.02	1000	2.5	2.594200e-01	3.811886e-01
0.05	3	1.0	2.575500e-03	6.622714e-03
0.05	3	1.5	8.720700e-03	1.902698e-02
0.05	3	2.0	5.376600e-02	1.075320e-01
0.05	3	2.5	7.885500e-03	1.774238e-02
0.05	100	1.0	5.161700e-01	5.899086e-01
0.05	100	1.5	3.062300e-01	4.240108e-01
0.05	100	2.0	2.850600e-01	4.104864e-01
0.05	100	2.5	2.000000e-01	3.200000e-01
0.05	1000	1.0	6.596200e-01	7.306560e-01
0.05	1000	1.5	9.249800e-01	9.464755e-01
0.05	1000	2.0	2.989300e-01	4.220188e-01
0.05	1000	2.5	2.594200e-01	3.811886e-01

s threshold	N_*/K	μ	p value	adjusted p value
0.10	3	1.0	1.889000e-12	8.000471e-12
0.10	3	1.5	1.797100e-06	5.881418e-06
0.10	3	2.0	2.883200e-04	7.688533e-04
0.10	3	2.5	8.462700e-12	3.206918e-11
0.10	100	1.0	1.657900e-02	3.510847e-02
0.10	100	1.5	4.343300e-01	5.237400e-01
0.10	100	2.0	2.210700e-04	6.121938e-04
0.10	100	2.5	2.836900e-03	6.808560e-03
0.10	1000	1.0	7.432900e-02	1.446402e-01
0.10	1000	1.5	7.676400e-02	1.454476e-01
0.10	1000	2.0	1.535800e-01	2.632800e-01
0.10	1000	2.5	3.272700e-06	9.818100e-06
0.20	3	1.0	6.729100e-52	9.689904e-51
0.20	3	1.5	1.208500e-31	1.087650e-30
0.20	3	2.0	2.817100e-13	1.267695e-12
0.20	3	2.5	2.833000e-30	2.266400e-29
0.20	100	1.0	1.324100e-08	4.766760e-08
0.20	100	1.5	2.016300e-04	5.806944e-04
0.20	100	2.0	4.252200e-12	1.700880e-11
0.20	100	2.5	2.836900e-03	6.808560e-03
0.20	1000	1.0	2.371500e-02	4.878514e-02
0.20	1000	1.5	4.208000e-03	9.773419e-03
0.20	1000	2.0	5.832600e-08	1.999749e-07
0.20	1000	2.5	3.272700e-06	9.818100e-06
0.50	3	1.0	5.232300e-128	3.767256e-126
0.50	3	1.5	2.188700e-81	7.879320e-80
0.50	3	2.0	3.061900e-57	5.511420e-56
0.50	3	2.5	2.999000e-71	7.197600e-70
0.50	100	1.0	7.463900e-28	5.374008e-27
0.50	100	1.5	3.243000e-24	2.122691e-23
0.50	100	2.0	7.095900e-34	8.515080e-33
0.50	100	2.5	1.608200e-18	8.906954e-18
0.50	1000	1.0	2.430100e-17	1.249766e-16
0.50	1000	1.5	3.875000e-17	1.860000e-16
0.50	1000	2.0	5.526900e-32	5.684811e-31

s threshold	N_*/K	μ	p value	adjusted p value
0.50	1000	2.5	7.038700e-22	4.223220e-21

Table 3. Results of all K-sample Anderson Darling tests. Each individual test interrogated the empirical distributions of surviving mutation frequencies for all mutations with $|s| \leq s$ threshold. The data from Table 2 appears here as the entries with s threshold $=$ 0.05. Adjusted p values were computed using the FDR-controlling procedure of Benjamini and Hochberg [72].

We used the `ad.test()` function in the R library `kSamples` to perform the Anderson-Darling tests mentioned in Section 3.3. The p values reported in Tables 2 and 3 were adjusted using the R function `p.adjust()` with `method = 'fdr'`. The code to run those tests and adjust p values is included in the GitHub repository associated with this paper.

The p values of Table 2 report results of tests on the empirical distributions of mutation frequencies when the mutations have $|s| \leq 0.05$. The p values of all tests are reported in Table 3. The p values show a transition from being consistently unable to reject the null hypothesis when s threshold < 0.05 to consistently rejecting it when s threshold > 0.05.

3.5.4 Visualizing effect of kernel exponent on mutation frequencies.

Ref. [23] discovered that the coarsening of space leads to a decay of neutral diversity during range expansions with $d < \mu < d + 1$, whereas frequent long-range dispersal events preserve diversity when $\mu < d$. We report another observable result of the coarsening of space here. Fig. 20 and Fig. 22 show that new mutations are more likely to survive in populations with broader jump kernels (lower μ). However, out of all the surviving mutations, the highest frequencies are often achieved by mutations in populations with narrower jump kernels as shown in Fig. 27. This is because the coarsening of space allows the mutant population to carve out and control a larger chunk of the habitat for itself, while populations with broader jump kernels have more dispersal across the landscape and finer spatial "speckles" composed of one genotype or the other.
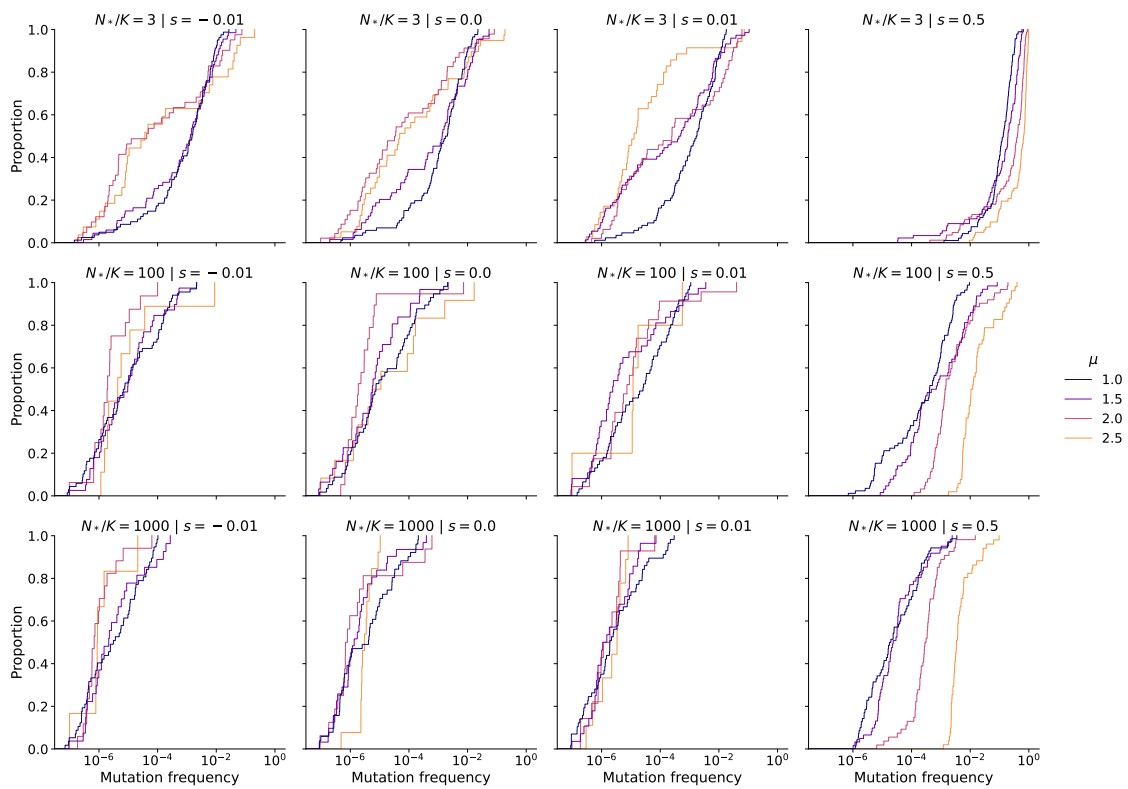
Figure 27. ECDFs of the frequencies of surviving mutations. This is an alternative view of some of the data plotted in Fig. 21.

CHAPTER IV

ESTIMATING DISPERSAL PARAMETERS USING NEURAL NETWORKS

4.1 Introduction

In this chapter, we present preliminary results on using genomic data to infer information about a population's jump kernel. The results shown here serve as a proof of concept that these approaches can yield useful results in populations with long-range dispersal, something which was not previously known, and highlight several exciting directions for future research. This work was done in collaboration with Chris Smith.

The previous chapters used spatially explicit but genomically simplistic simulations to gain insights about populations with long-range dispersal of offspring. There are important gains to be made on the genomic side of the aforementioned simulations: a connection between actual biological populations and simplified theory such as what has been presented here and in the literature will require stepping up the biological complexity of simulation models.

Furthermore, all previous work on populations with long-range dispersal by power law jump kernels takes the jump kernel as a given [21, 22, 23, 34]. The kernel exponent is an input to the simulation or model, and the evolved population is the output. However, we often learn about biological populations in nature by investigating populations in their current state, i.e. after evolution has acted on the population, with no ability to look back in time to see how a population evolved from its initial to its current state. With that being the case, it would be useful to invert the structure of previous work: can we use observations of an evolved population to learn its jump kernel? Such information would enable predictions regarding evolutionary phenomena that the population may experience, like whether or not neutral diversity would be preserved during expansions into new territories [23], or could give clues about how to sample the population in search of alleles undergoing soft sweeps throughout the habitat [22]. It could also reveal how the population might respond to range shifts forced by climate change [11].

Ref. [38] showed that convolutional neural networks (CNNs) can accurately perform the closely related task of inferring the standard deviation of Gaussian jump kernels. We extend their results to show that networks with the same architecture can also be trained to infer the kernel exponent when the training data and labels come from populations with power law jump kernels. We then build further by training networks to learn typical and standard deviations in dispersal distances without assuming any particular jump kernel and describe attempts to classify between Gaussian and power law jump kernels. We hope that the methods introduced below will continue to be developed into tools that can be used to learn dispersal information from samples of natural populations.

4.2 Background

Our use of neural networks here builds on a growing body of recent work that leverages machine learning for population genetic inferences. Supervised machine learning is a great choice for such applications because models can be trained using simulated data, where researchers know the correct answers, before being used on data sampled from natural populations where they are trying to learn previously unknown things. Machine learning models have been shown to be able to perform well in a variety of tasks in population genetics [76, 77, 78, 79, 80, 81, 82]. In particular, deep neural networks have grown to be increasingly popular recently [39, 40, 41, 42, 43, 44] in part because they can take in large numbers of features without requiring any sort of compression, enabling predictions using raw genotype data as model inputs and bypassing the need to compute population genetic summary statistics, which might empower models to extract additional relevant information from the input that would otherwise be lost during compression.

Recent work has trained neural networks to learn from correlations between genotypes and location; for example, Ref. [42] showed they can be trained to predict an individual's location using their genomic data. Of particular interest here, Ref. [38]

93

trained a convolutional neural network to learn about a population's jump kernel
from a modest number of genomic samples. In the following sections, we introduce the
methods of Ref. [38], describe how we adapt their software to populations with long-range
dispersal, and finally discuss preliminary results from models that extend and generalize
their techniques.

4.3   Introducing disperseNN methods

Ref. [38] developed a software tool called disperseNN that used convolutional
neural networks to learn about dispersal from genomic samples. disperseNN was trained
on data generated by 2D continuous-space SLiM simulations [33] which themselves were
based on the simulation model presented in Ref. [42]. We provide a concise summary
here but encourage interested readers to refer to the original papers for all the details of
their simulation models. Mating pairs produce a Poisson-distributed number of offspring;
the mean of the distribution is set to $1/L$, where L is the expected lifetime when the
population is at equilibrium. Offspring dispersal distances in x and y are draws from
a Gaussian distribution with mean zero and standard deviation σ; the length scales of
mate choice and spatial competition are also set by σ. Simulations are spatially explicit
and genomically realistic, with density-dependent competition throughout the landscape,
genomes of length 10^8 base pairs, and a recombination rate of 10^{-8} crossovers per base
pair. Simulations begin at approximately the global carrying capacity with individuals
randomly located throughout the landscape and then run for 10^5 time steps before being
terminated. Simulations generate tree sequences [59], which record the true ancestry at
all loci of all individuals alive at the end of simulations. Tree sequences are recapitated
using coalescent simulations [45], and then neutral mutations are placed on the tree
sequences in order to create the diversity from which the network will learn.

With neutral mutations atop tree sequences, the training data is ready to be fed
into the network. The disperseNN input data consists of two parts. First, a genotype
matrix that encodes the allelic identities of a modest number of randomly sampled

94

individuals at a user-defined number of single nucleotide polymorphisms (SNPs). Ref. [38] showed good results sampling as few as ten individuals, and often used 5,000 SNPs. Second, the distance between the furthest-separated pair of individuals in the sample is given to the network so it knows the spatial scale of the sampling area. The network is trained to map the genotype matrix and size of the sampling area onto a predicted value of σ, the standard deviation of the Gaussian jump kernel. It learns by first sending the genotype matrix through several successive convolution and pooling layers, where the number of convolution + pooling steps depends on the number of SNPs. The convolutional layers are one-dimensional, spanning all individuals (columns) and two SNPs (rows) of the genotype matrix. Simultaneously seeing the allelic identities of all sampled individuals gives the network a view of how much diversity there is within the sampling area and provides important clues about how dispersal is driving spatial mixing of genotypes.

After the convolutional portion of the network, the intermediate tensor is flattened and sent through three dense layers. The size of the sampling area is concatenated with the output from the third dense layer and then the resulting tensor is sent through a fourth dense layer. Finally, a dense layer with a single output gives the estimate of σ. Since σ is proportional to the expected dispersal distance, disperseNN is learning the typical dispersal rate of new offspring.

Typically 20% of the training data is set aside to use as the validation set at the end of each training epoch. The model trains and learns from the rest of the training data, and only sees the validation set at the end of each epoch for benchmarking performance on data that isn't learned from during training and as a way of monitoring the network for overfitting. A complete summary of the network architecture is presented in Ref. [38] and all simulation and network code related to disperseNN is available at https://github.com/kr-colab/disperseNN.

4.4 Inferring the kernel exponent

A simplistic view of disperseNN is that it is a software that has been trained to use genomic data and the size of the sampling area to estimate some dispersal parameter. That dispersal parameter must be the one and only parameter necessary to specify the jump kernel; the network would be unable to learn if, for example, it was attempting to learn the shape parameter of gamma-distributed dispersal distances while the scale parameter was also varying and thus confounding the signal. The original disperseNN software was trained using simulations with Gaussian-distributed dispersal distances, and learned to predict the one parameter necessary to specify the jump kernel: the standard deviation.

The work presented throughout this dissertation focuses on populations with jump kernels that decay as power laws at long distances. Power law kernels are another family of jump kernels with just one parameter, namely the exponent μ that sets the weight of the power law tail according to $J(r) \sim 1/r^{\mu+1}$. When the kernel exponent μ is below a dimension-dependent threshold, power law jump kernels drive accelerating expansions that are dominated by rare long-range dispersal events [21]. Long-range dispersal leads to remarkably different growth dynamics during range expansions [21] and leaves behind distinct spatial-genetic signatures following many generations of evolution in a static range [22] when compared with populations that are only capable of short-range dispersal. The latter case resembles the simulations used to train the original disperseNN software, where many generations of evolution took place in a static landscape and offspring dispersed according to Gaussian jump kernels. We now report attempts to adapt the disperseNN network architecture to infer the kernel exponent from genomic samples of populations where offspring disperse according to power law jump kernels.

The first step in adapting disperseNN to power law jump kernels was to generate new training data. This was a reasonably straightforward task since the SLiM simulation

code was made available on GitHub along with the rest of the disperseNN code base. Our hope was to make as few changes to the simulation code and network architecture as possible. Fortunately, all that needed to be done was change the jump kernel from Gaussian to power law, decide on a training distribution from which to draw values of μ, and then run a few hundred simulations on the computing cluster.

We chose a jump kernel of the following form:

$$
J(r) = \begin{cases} A/\sigma & r \leq \sigma \\ (A/\sigma)(r/\sigma)^{-(\mu+1)} & r > \sigma \end{cases} \tag{4.1}
$$

where the normalization constant is $A = \frac{\mu}{\mu+1}$ and σ is the length scale of spatial competition and mate choice. We chose a jump kernel that is featureless out to distance σ to model species where offspring are equally likely to draw dispersal distances anywhere between zero and the length scales of competition and mate choice (the only pre-existing length scales in the population), and decreasingly likely to end up further away. Combining the length scale of dispersal with that of the other spatial interactions is a simple and sensible choice that prevents us from having to introduce more parameters into our model and follows the approach taken by previous work, including our predecessors in using this particular simulation model [38, 42]. We draw the dispersal distance following the inverse transform sampling procedure outlined in the previous chapter, then draw the direction uniformly at random.

We chose values of μ to span the different growth regimes for populations in two dimensional landscapes. Kernel exponents were drawn at random between 1 and 3.3 to give us good coverage in each of the stretched exponential, power law, and diffusive growth regimes; keeping $\mu > 1$ ensures a finite mean dispersal distance and was hoped to be far enough from well-mixed that the network could still see and learn from spatial structure. We kept all other parameters fixed when generating the training data. The most important quantity we held constant was the value of σ, which we kept at 0.75 for

97

all simulations. This guaranteed that dispersal distances were small enough relative to the landscape width (50 SLiM distance units in each dimension) that the population would be far from well-mixed. We generated a training data set of 759 simulations and a test set of 372 simulations with kernel exponents between 1 and 3.3. We were also interested in whether or not the model would be able to generalize to kernel exponents beyond the training distribution, so we ran 191 simulations with exponents between 3.5 and 5 and 183 simulations with exponents between 0.4 and 1.

Ref. [38] released a pretrained model as part of the disperseNN code base. Our first test of disperseNN for power law jump kernels involved seeing if the pretrained model could detect any changes to the dispersal rate at different values of μ. This was to interrogate what one may see if they erroneously used the pretrained model instead of training a new model themselves with simulations that more accurately represent their organism. disperseNN was trained to learn the dispersal rate from genomic samples of individuals in Gaussian-dispersing populations; we can compare model performance by integrating the jump kernel (Eq. (4.1)) to compute the expected dispersal distance in our simulations:

$$\mathbb{E}(r) = \int_0^\infty r J(r)\, dr = \frac{A}{\sigma} \left[\int_0^\sigma r\, dr + \sigma^{\mu+1} \int_\sigma^\infty r^{-\mu}\, dr \right] = \frac{\mu\sigma}{2(\mu-1)}. \qquad (4.2)$$

Computing the expected dispersal distance reveals two key takeaways. First, we quantitatively see how the expected distance decreases with increasing kernel exponent (with a narrower and narrower jump kernel). Second, we see that the expected jump distance diverges for $\mu \leq 1$: we cannot compute an expected jump distance for the broadest kernels.

We asked the pretrained model to predict σ using genomic samples from all 1505 available simulations with power law dispersal. The model correctly picks up on the scale of most dispersal events, but is insensitive to any changes in kernel exponent until the jump kernel becomes so broad as to have a divergent expected distance, see
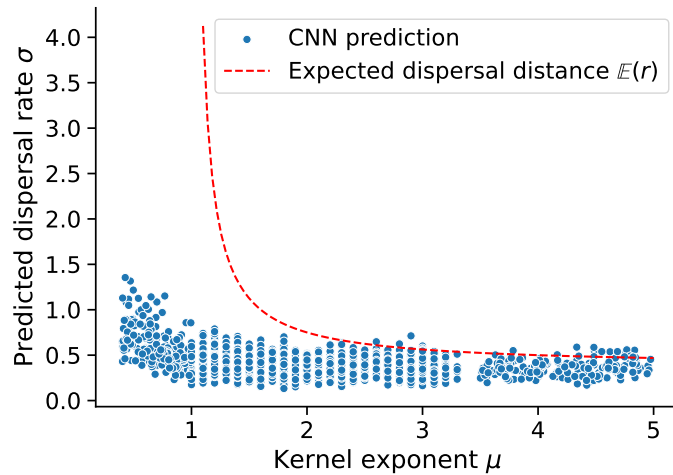
98

Figure 28. Predicted dispersal rate from the pretrained disperseNN model plotted against kernel exponent for all 1505 available simulations with power law dispersal. The red dashed line marks the expected dispersal distance as a function of kernel exponent, the solution of Eq. (4.2). Recall that the expected dispersal distance diverges for $\mu \leq 1$.

Fig. 28. The red dashed line is the expected jump distance (the solution to Eq. (4.2)) for $\mu \geq 1.1$. The pretrained model predictions get reasonably close to the true value for the narrowest jump kernels we used, but fail dramatically for broader kernels. Clearly it's important to train the model using data that looks like what we want to eventually generate predictions from.

We should note that this comparison is admittedly imperfect. The pretrained model was designed to provide order of magnitude estimates, and was trained on simulation data where several parameters varied. An "ideal" misspecification analysis would be to generate new training data with the same fixed parameters we used but have Gaussian rather than power law dispersal, train a disperseNN network to estimate the dispersal rate from that data, and then predict the dispersal rate from our power law simulations as we have done in Fig. 28. This could enable better understanding of the consequences of misspecifying the jump kernel in simulations that otherwise closely match what we ran. However, instead of pressing harder on misspecification analysis, we
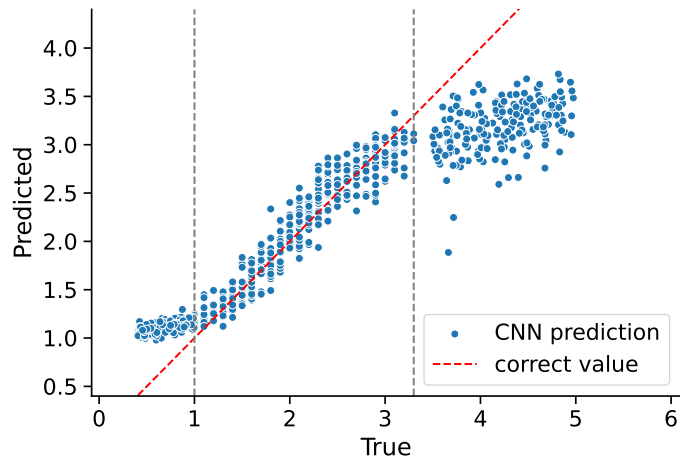
Figure 29. Predicted versus true kernel exponent using test data that the model did not see during training. The red dashed line marks the correct value, and the vertical dashed lines mark the bounds of the training distribution.

proceeded by attempting to train disperseNN networks to learn about power law jump kernels.

Our next step was to see if a new model with disperseNN architecture could be trained to infer the kernel exponent using genomic samples from populations with power law dispersal when we give the model relevant training labels (kernel exponents). This new model performed quite well on test data that resembled what it saw during training, achieving a mean relative absolute error (MRAE) of 7% on predictions where the kernel exponent fell within the training distribution; see Fig. 29.

The model failed to recognize kernel exponents beyond the training distribution in either direction; see the roughly flat clusters of predictions outside of the vertical reference lines in Fig. 29. The MRAE was 24% for samples with kernel exponents above 3.5 and 70% for those with kernel exponents less than 1. On one hand, this should be no surprise since an inability to generalize beyond training data is a standard feature of machine learning models. However, it was a surprise that the model showed no signs of recognizing the rapidly increasing spatial mixing at the lowest values of μ. This serves as yet another example of the importance of using training distributions that capture

all regimes of interest; the network is unable to learn beyond examples it saw during training.

4.5 Learning dispersal parameters without assuming a particular jump kernel

The previous section showed promising results using a CNN to learn the kernel exponent when we assume and train using power law jump kernels, building on recent literature showing that CNNs can infer the dispersal rate from genomic samples of populations with Gaussian jump kernels [38]. In this section, we describe efforts to continue down the path of inferring dispersal parameters from genomic data, but now without assuming any particular jump kernel. We hope to train a model capable of answering three questions: What is the typical dispersal distance? How much variability is there about that typical distance? Is there any dispersal much longer than that expected from the typical range characterized by the mean and the variance? The initial idea was to build up a data set consisting of simulations with Gaussian and power law jump kernels, and then train one model capable of performing both regression tasks *and* classifying between the two jump kernels. We only have preliminary results to share at the time of this writing, but hope the project will continue and grow into a useful tool for learning about dispersal from actual biological samples.

4.5.1 Simulation methods. Simulations generally follow the protocol developed by Refs. [38, 42] and outlined in Section 4.3, but there are changes to the choice of jump kernel, what data is recorded during simulations, and how data is logged after simulations. Again there is a model input called σ which is the length scale of mate choice and spatial competition. We want to build towards a model that can make predictions agnostic of jump kernel, so these simulations may use Gaussian or power law dispersal kernels. The decision is made at runtime by a computational coin flip (i.e. by testing `runif(1) > 0.5`); Gaussian simulations use σ as the standard deviation for jump distance draws in x and y just like the original disperseNN methods, while power
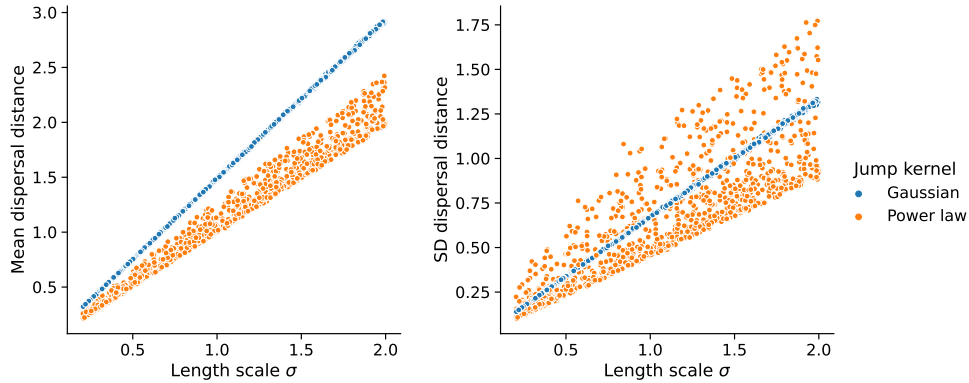
101

Figure 30. Sample mean and sample standard deviation versus the length scale σ, where the samples are the dispersal distances of all individuals alive in the final time step of simulations. The color of the points indicate whether Gaussian or power law jump kernels were used in that simulation. Each point corresponds to one simulation.

law simulations use σ as the cutoff between uniform probability and power law decay as spelled out in Eq. (4.1).

A key change from previous models is that now we record the offspring dispersal distances during simulations in such a way that the information is tied to the relevant individual. Individuals disperse a random distance from their mother's position when they are born, and we 'tag' individuals with the average of the distances between their position and that of each of their parents. We compute the sample mean and standard deviation of the dispersal distances of everyone alive in the final time step of simulations, giving us typically about 20,000 dispersal distance samples from any single simulation at the parameters we used. These methods give us a way to record, and eventually train to learn, dispersal distances as an emergent property rather than using simulation input values like σ or μ directly as model targets as described in the previous sections. Simulations produce two outputs: a CSV file including parameter values, which jump kernel was used, and sample mean and standard deviation in dispersal distances from everyone alive in the final simulation time step; and tree sequences containing the ancestry of all individuals alive in the final time step.
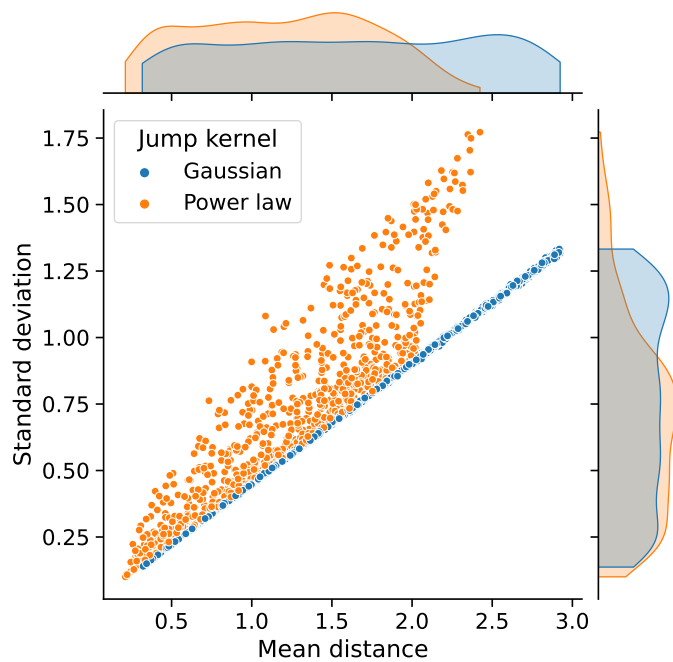
Figure 31. Phase plot showing sample mean and sample standard deviation of the dispersal distances of all individuals alive in the final time step of simulations, where the color of the points indicates whether Gaussian or power law jump kernels were used in that simulation. Each point corresponds to one simulation. The marginal plots at the top and right show kernel density estimates of the variable values to give a sense of what the distributions of each separate quantity look like in the data.

We ended up with a training set of 1953 simulations: 993 with Gaussian dispersal and 960 with power law dispersal. We drew values of σ uniformly at random between 0.2 and 2, and kernel exponents between 2 and 5 (where the mean and variance are both finite). σ is the independent variable that sets the scale for all spatial interactions; its effect on typical dispersal distances in the training data is depected in Fig. 30. We chose these parameter distributions with a few things in mind. First, we hoped to keep dispersal distances generally well below our chosen landscape width of 50 SLiM distance units in either direction; this was based on the disperseNN network being unable to learn the dispersal rate when there's too much spatial mixing (when dispersal distances are an appreciable fraction of the landscape size). Second, we wanted to create a non-trivial problem and see if a network could still learn. Great separation in phase space between the Gaussian and power law simulations might make the regression and particularly the classification between jump kernels easy. We hoped our training data, visualized as a phase space in Fig. 31, would provide a challenging but still solvable problem.

4.5.2 Training the network. With the data in hand, we next began adapting the disperseNN code base and attempting to train models. We sought to make no changes to the network architecture and as few changes as possible to the rest of the code base in an effort to build on knowledge and experience from training networks to accurately learn dispersal rates and kernel exponents, and with an eye on possibly merging useful changes into the main disperseNN repository in the future. The classification between jump kernels ended up being a surprisingly difficult problem; we begin by discussing only results from the regression outputs and later return to classification.

The one obvious change we had to make to the network architecture was the number of outputs. The original disperseNN network had just one output, which predicted a continuous quantity; we needed two such outputs. Adding another output to the model fortunately required just a few lines of additional code since the original

104

software was written using Keras' functional API [83]. We used the same loss function (mean squared error) and optimizer (Adam) as in the original output since this regression task is very similar to the original.

We improved model performance by updating how genomic samples are shown to the network. Our network uses the same two inputs as the original disperseNN software. The first input is the genotype matrix, a 2D array with shape (m, n) where m is the number of SNPs and n is the number of individuals sampled; we typically used $m = 5000$ SNPs and $n = 100$ individuals. The second model input is the size of the sampling area. The order of the n individuals making up the columns of the genotype matrix was random in the original disperseNN software; however, those individuals carry useful information about the spatial distribution of genotypes across the landscape, especially when considering that we tell the network the size of the sampling area. In general, nearby pairs of individuals are more closely related and more similar genetically than distant pairs of individuals [84, 85], even when power law dispersal can transport individuals long distances across the landscape [61] (this statement views isolation by distance from the opposite perspective – relatedness by proximity?). Additionally, recent research [22] has studied the spatial distributions of alleles undergoing soft sweeps when power law dispersal can transmit them long distances across a landscape and found that power law jump kernels leave distinct signatures in the spatial distributions of genotypes. We sought to leverage this correlation between location and genotype by sorting individuals by spatial location when constructing the genotype matrix.

We used multidimensional scaling [86, 87, 88] to sort individuals by location. Multidimensional scaling is a dimensionality reduction technique that projects high-dimensional data into a lower-dimensional space while preserving distances between pairs of data points from the original high-dimensional space as well as possible. We used it to project our original positions on the 2D landscape into a new 1D landscape, then used the order of individuals in the new 1D landscape to set the order of the columns in the
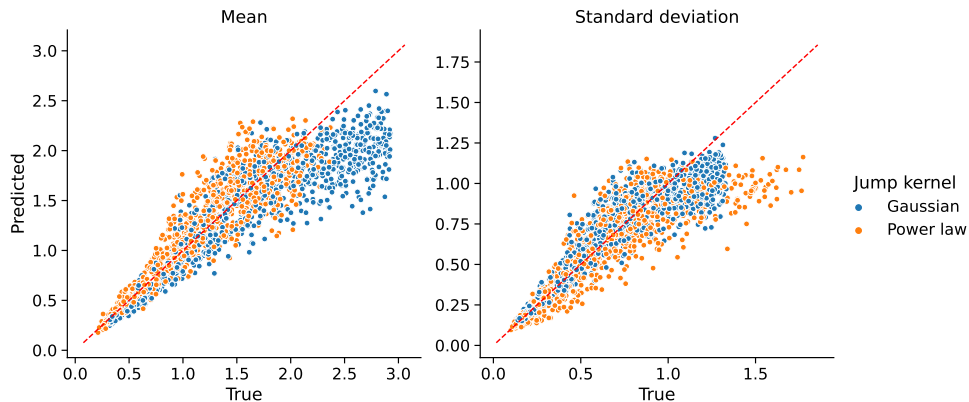
Figure 32. Predicted versus true mean and standard deviations in dispersal distances. The red dashed line marks where the predicted value equals the true value. Each point is the model prediction output using data from one simulation as input; these are predictions for the full training set of 1953 simulations.

genotype matrix. We chose to sort using multidimensional scaling in order to "let the data speak for itself" in determining the appropriate sorting rather than an arbitrary decision to sort by x or y positions which may not be a useful descriptor for a given random sample of individuals. This information about who was close to whom in an area of known size proved to be quite useful to the network; implementing multidimensional scaling led to a 50% reduction in loss on the validation set compared with networks trained using unsorted samples.

4.5.3 Regression results. The model does a decent job learning to predict both of the continuous outputs, achieving an MRAE of about 15% for both the mean and standard deviation targets; see the model predictions in Fig. 32. Note that the accuracy metrics reported throughout this section will refer to model accuracy when predicting on the entire training set rather than an unseen test set. We don't believe we have found the best possible model or model architecture yet, so have not yet generated a test set for final model validation. Recall, though, that part of the training data is set aside for model validation at the end of each epoch. The model doesn't get to learn from the validation set, only benchmark its performance on data it isn't using for training. The

model is saved to file at the end of epochs where the *validation* loss decreases; that is, the model's performance improves even on data that it isn't using for training. We do not continue saving models when they overfit to the training data, with continually decreasing training loss at the expense of stagnant or increasing validation loss. Based on saving the model with the best observed validation loss, we suspect the accuracy metrics we report mimic what would be seen if we generated a new test set for additional model performance benchmarking. Furthermore, the n individuals randomly sampled from each tree sequence during the prediction step are very unlikely to be the same n individuals chosen from each simulation during training, so we have no reason to worry about the model predicting on training data that it has memorized.

The model is able to learn dispersal distances from genotype data even when the jump kernel is unknown; it is also able to estimate typical variability in dispersal distances. The presence of multiple jump kernels in the training data seems to break down any highly accurate mapping between genotypes and typical dispersal distances, but we find that there are optimistic interpretations of these results. The model can learn useful information from genotypes even when there are multiple "pathways" (jump kernels) leading to the relevant dispersal parameters.

4.5.4 Classification results. While the model achieved respectable performance predicting the two continuous output targets, it failed to accurately classify between jump kernels. This came as a surprise since long-range dispersal by power law jump kernels has been shown to promote spatial mixing of alleles [22], and we suspected that the increased spatial mixing in simulations with power law jump kernels would make the classification easy. No models we trained ever got much better than 50% accuracy on the training set, and predictions using even our best models were heavily biased towards guessing mostly one jump kernel or the other. It seemed that our models were failing to learn this task.
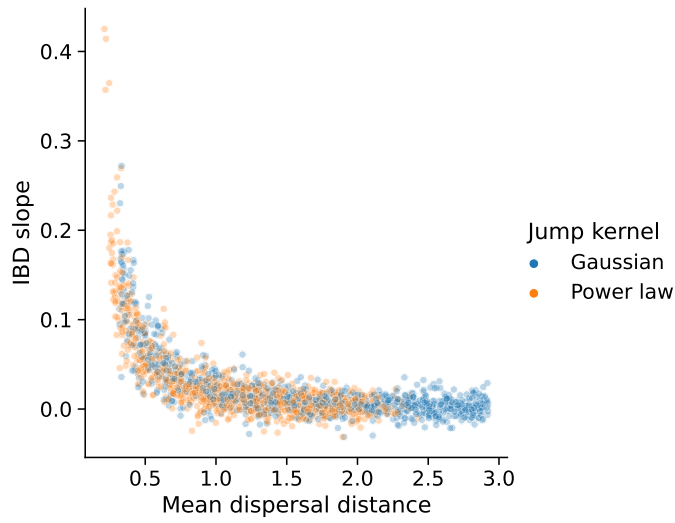
Figure 33. Slope of the isolation by distance model best fit line plotted against sample mean dispersal distance from the final time step of simulations. Points are colored by jump kernel and semitransparent in order to show that many points from either jump kernel fall right on top of one another.

The model architecture was developed by Ref. [38] to accurately predict the dispersal rate in populations with Gaussian dispersal kernels, and we showed in the previous sections that it can learn other continuous targets related to dispersal, but those facts don't necessarily mean that the same architecture will be able to classify between two jump kernels. It could be that a classifier CNN may need different structure or more/different inputs to perform well, or maybe an entirely different model architecture is required for this task.

We began to explore if we could find relevant information in the genomic data to give helpful hints to the network. In particular, we hoped that computing summary statistics using the tree sequences could illuminate quantities that might be useful as other inputs to the neural network to aid the classification. If we found multiple statistics to be descriptive and generally different between the two jump kernels, we could maybe even train a much simpler logistic regression model to accurately discriminate between the two classes. We found no differences between the two classes when we computed nucleotide diversity π [89]. We also explored computing isolation by distance (IBD)

following the procedure outlined by Ref. [90]. We compute genetic distance between every pair of individuals and then find the best fit line between genetic distance and geographic distance; the relevant summary statistic is the slope of the line, which tells us how genetic distance changes with geographic distance. Since power law kernels promote spatial mixing [22] and have been shown to reduce IBD relative to Gaussian jump kernels [61], we expected to see generally lower slopes to the IBD line for power law jump kernels compared with Gaussian jump kernels at a given typical dispersal distance. Instead, we see no difference in isolation by distance between the two jump kernels; see Fig. 33.

4.6   Discussion

The disperseNN software was optimized to learn the dispersal rate from genomic samples of populations that obey Gaussian jump kernels [38]; we explored other applications by training to infer the kernel exponent from simulations with power law jump kernels (Section 4.4) and learn mean and standard deviations in dispersal distances without assuming any particular jump kernel (Section 4.5). These preliminary investigations suggest several possible avenues of future research.

We achieved good accuracy when inferring the kernel exponent for values within the training distribution (MRAE 7%; Fig. 29), but that result comes from somewhat pristine training data where the only quantity that varied was the kernel exponent and test data that had exactly the same conditions as the training data. A more complete analysis would involve varying other quantities such as the local carrying capacity or the length scale σ, or misspecifying some quantities between the training and test sets, in order to learn how different values and/or misspecifications affect model performance. The work done by Ref. [38] could be used as a roadmap for possible next steps.

We showed that disperseNN-like networks can learn the dispersal rate and standard deviation in dispersal distances even when there are multiple jump kernels present in the training data. While we couldn't match the accuracy of networks trained to infer specific parameters from assumed jump kernels, we hope the generality of this

109

kernel-agnostic method may help it eventually become a useful tool to researchers. Few biological species have known jump kernels, and measuring typical dispersal distances empirically can be challenging, so we believe the ability to get reasonably accurate estimates of typical dispersal distances and variability from a handful of genomic samples without having to assume any particular dispersal kernel could someday be a useful step forward.

The classification results presented in the previous section were a surprising failure (to us). We expected the model to be able to classify between the two jump kernels fairly well, and had visions of maybe someday using this network as a sort of model selection algorithm that pointed out worthwhile directions of future study depending on which jump kernel it inferred. If it inferred power law, perhaps one would proceed by trying to infer the kernel exponent; if it inferred Gaussian, no more work is required since the model also would've inferred the dispersal rate and typical variability. It would be great to figure out why the model was unable to learn this task, or at least experiment with different architectures to see if anything else looks promising. One possibility is that we restricted kernel exponents to too narrow a range when generating the training data. The expected jump distance from the power law jump kernel diverges for $\mu \leq 1$ (Eq. (4.2)) and the variance diverges for $\mu \leq 2$; perhaps a model could be highly sensitive to the differences between Gaussian and some power law kernels if such broad power law kernels were included.

Exploring the classification failure led us to begin exploring the tree sequences for any noteworthy changes in population genetic summary statistic values between the two jump kernels; we suspect that this in itself could be a worthwhile research target. We have almost 2000 simulations, roughly evenly split between Gaussian and power law jump kernels, with typical dispersal distances being about the same between the two but the power law jump kernel permitting occasional long-range dispersal events. Populations evolved for 10^5 time steps regardless of jump kernel; dispersal across the

landscape was the only difference between the two classes of simulations. Is there any sign of that difference in the tree sequences? Do the genomes or some summary statistics somehow quietly encode a difference between Gaussian and power law jump kernels? Or maybe more broadly, between Gaussian and any fat-tailed jump kernel? Such investigations could lead to powerful inferences and novel insights about relationships between genealogical structure and long-range dispersal.

CHAPTER V

CONCLUSIONS

In this chapter, we briefly summarize the key findings of the previous three chapters and call out various avenues of possible future research. While we have answered several questions with the results presented in this dissertation, many exciting questions remain.

5.1   Modeling expansions driven by long-range dispersal

In Chapter II, we explored the consequences of breaking the instant local saturation and founder takes all assumptions that previous lattice-based studies of populations with power law jump kernels [21, 22, 23] relied on. We found that quantitative measures of population growth rate and the evolution of neutral diversity are sensitive to the degree to which the aforementioned assumptions were violated. For example, increased amounts of long-range dispersal slow down local saturation and lead to greater levels of local heterozygosity while accelerating growth of the population as a whole.

One possible avenue of future research would be to design theoretical models of expanding populations that account for non-instantaneous local saturation. We showed in Chapter II that qualitative trends match the predictions of lattice-based models; perhaps a relatively minor correction to those models could lead to quantitative agreement with more realistic spatially explicit simulations such as what was presented throughout this dissertation. Models that account for the finite duration of local dynamics would be an important step towards being able to predict the growth and spread of actual biological populations as they invade new territories.

5.2   Evolution during expansions driven by long-range dispersal

In Chapter III, we studied the fate of fitness-affecting mutations that appear during range expansions driven by long-range disperal. We showed that mutation effect sizes have no measurable impact on mutation outcomes until the mutations become

extremely beneficial or deleterious. "Survival of the luckiest" dynamics generally win out over "survival of the fittest" for the wide range of most likely effect sizes.

Throughout that work, we used a common definition of "fitness-affecting." We considered mutations that only affected offspring production rate, and implemented that change by changing the mean of the Poisson distribution from which mutant offspring numbers were drawn each simulation time step. However, there are various other pathways to changing fitness, and exploring any of them could lead to novel and interesting results.

We think of changing the offspring production rate as a "global" fitness change; it has no effect on local interactions between individuals, it just changes the growth rate of the mutant population relative to wildtypes. Another fascinating global change to study would be mutations that affect dispersal. Naively, it seems that faster dispersal into new territories should simply be better, but there is often a cost associated with improved dispersal. This cost may be minimal in e.g. plants where small changes to the form of their seeds or pollen could facilitate further wind transport. However, in animal populations where individuals must actively cover the dispersal distance themselves, the metabolic cost becomes a macroscopic factor. For example, cane toads have evolved longer legs during their invasion of Australia which allow them to travel further and faster [10]. Longer legs enable faster movement but bigger muscles require more food and water; legs that are too long may actually hinder the invasion by making it harder for toads to obtain all the resources they need for survival. Future investigations that study changes to dispersal would need to consider costs associated with changing dispersal and find a way of balancing the cost of dispersal with the benefit of being the first to access previously uninhabited territories.

There is also room for work to be done in the realm of mutations with local fitness effects. It was recently shown that slower-expanding alleles can take over the front in diffusive expansions when the slower allele has a local competitive advantage

113

that allows it to invade laterally into the faster allele's territory [74]. We showed in Chapter III (Fig. 19) that juveniles are more and more likely to be born into saturated areas as range expensions proceed, which makes it difficult for new mutations to establish a solid local footing and escape drift; this problem is especially dire for mutations that appear in populations with narrower jump kernels. Mutations that appear later are often immediately lost because they are surrounded and outcompeted by wildtypes. Including a local competitive advantage could drastically improve the outcomes of mutations that appear in saturated areas by giving the mutants a better chance of surviving than the wildtypes around them. This change could easily be incorporated into the SLiM script that ran the simulations presented throughout Chapter III.

5.3   Using machine learning to infer dispersal parameters from genomic data

In Chapter IV, we shared preliminary results that highlight multiple directions of future work. Fleshing out any of these projects would be a valuable step towards connecting the growing body of theoretical and computational work about populations with power law dispersal to samples that could be taken from actual biological populations.

First, we showed that the disperseNN architecture [38] can learn to infer the kernel exponent when trained to do so using simulations with power law dispersal kernels. Those results had ideal training conditions, where the only thing varying between simulations was the kernel exponent. A fuller understanding of the network's abilities and weaknesses would require exploring distributions of training parameters rather than keeping everything fixed except the kernel exponent. It would also be helpful to misspecify parameters between training and test data sets, following what was done by Ref. [38], to learn more about which parameters the model is most sensitive to and which are somewhat less important.

Second, we trained models with the same architecture but two outputs rather than one to learn mean and standard deviations of dispersal distances from genomic

114

samples without any assumptions about the jump kernel itself; we had a mix of Gaussian and power law jump kernel simulations in the training data so we could learn kernel-agnostic dispersal parameters from mixed data and also attempt to train a classifier capable of discriminating between the two jump kernels. The grand vision for this project was to train a network capable of answering three questions. 1. What is the expected/typical dispersal distance? 2. How much variability is there about that typical distance? 3. Is there any long-range dispersal (i.e. anything much longer than typical plus normal variability)? We showed some promising results, but they come with many of the same aforementioned caveats about relatively pristine training data. We varied the length scales of dispersal, competition, and mate choice, and drew kernel exponents from a distribution of possible values where appropriate, but have not yet explored varying nuisance parameters or misspecifying anything; we have also not yet shown that such networks are capable of learning to classify between the two jump kernels.

Beyond those obvious next steps, it would be interesting to see if a network could be trained for similar regression tasks and a general yes/no classification regarding whether or not there is long distance dispersal rather than our current classification between two jump kernels. This would require at least two notable tasks. The first is deciding how to determine the yes/no long distance dispersal classification. We explored using the sample excess kurtosis from the dispersal distances of everyone alive at the end of simulations to generate those labels and found that a value of one cleanly separated Gaussian and power law dispersal samples. The second would be deciding which additional jump kernels to use. It may be ideal to choose one that can allow for long-range dispersal at some parameter combinations but not at others, such as the gamma distribution at various combinations of its two parameters. This could create data sets with a smooth transition from short-range to long-range dispersal where the boundary between the two classes could be defined by such measurements as the nature of advancement of the expanding front (constant-speed front for short-range, accelerating

front for long-range [91]), the convergence of the mean or variance of the jump kernel, or the excess kurtosis of a sample of dispersal distances. The network may object to the fuzzy boundary between classes by returning nearly identical 50/50 class probabilities for samples drawn from populations near the border, but having reliable predictions regarding whether or not a population exhibits long-range dispersal could be a valuable tool for predicting how the population mixes across its range or spreads into new habitats when new areas become accessible.

REFERENCES CITED

[1] Jean Clobert, Michel Baguette, Tim G Benton, and James M Bullock. *Dispersal ecology and evolution*. Oxford University Press, 2012.

[2] Ran Nathan. Long-Distance Dispersal of Plants. *Science*, 313(5788):786–788, August 2006.

[3] Casper H. A. van Leeuwen, José L. Tella, and Andy J. Green. Editorial: Animal-mediated dispersal in understudied systems. *Frontiers in Ecology and Evolution*, 7, 2020.

[4] H. Brunner, R. V. Harris, and R. L. Amor. A note on the dispersal of seeds of blackberry (rubus procerus p.j. muell.) by foxes and emus. *Weed Research*, 16(3):171–173, 1976.

[5] Janet L Fryer. Rubus armeniacus, r. bifrons: Himalayan blackberry. `www.fs.usda.gov/database/feis/plants/shrub/rubspp/all.html`, 2021. Accessed 2023-04-04.

[6] Scott Hotaling, Daniel H. Shain, Shirley A. Lang, Robin K. Bagley, Lusha M. Tronstad, David W. Weisrock, and Joanna L. Kelley. Long-distance dispersal, ice sheet dynamics and mountaintop isolation underlie the genetic structure of glacier ice worms. *Proceedings of the Royal Society B*, 286(1905), jun 2019.

[7] James M. Bullock, Dries Bonte, Gesine Pufal, Carolina da Silva Carvalho, Daniel S. Chapman, Cristina García, Daniel García, Erik Matthysen, and Maria Mar Delgado. Human-mediated dispersal and the rewiring of spatial networks. *Trends in Ecology Evolution*, 33(12):958–970, 2018.

[8] Matteo Chinazzi, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, Cécile Viboud, Xinyue Xiong, Hongjie Yu, M. Elizabeth Halloran, Ira M. Longini, and Alessandro Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, 368(6489):395–400, apr 2020.

[9] Kara J. Pitman, Jonathan W. Moore, Matthias Huss, Matthew R. Sloat, Diane C. Whited, Tim J. Beechie, Rich Brenner, Eran W. Hood, Alexander M. Milner, George R. Pess, Gordan H. Reeves, and Daniel E. Schindler. Glacier retreat creating new pacific salmon habitat in western north america. *Nature Communications*, 12(1):6816, 2021.

[10] Benjamin L. Phillips, Gregory P. Brown, Jonathan K. Webb, and Richard Shine. Invasion and the evolution of speed in toads. *Nature*, 439(7078):803–803, feb 2006.

[11] Gian Reto Walther, Eric Post, Peter Convey, Annette Menzel, Camille Parmesan, Trevor J.C. Beebee, Jean Marc Fromentin, Ove Hoegh-Guldberg, and Franz Bairlein. Ecological responses to recent climate change, mar 2002.

[12] Laurent Excoffier, Matthieu Foll, and Rémy J. Petit. Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):481–501, 2009.

[13] Christopher A. Edmonds, Anita S. Lillie, and L. Luca Cavalli-Sforza. Mutations arising in the wave front of an expanding population. *Proc. Natl. Acad. Sci. U.S.A.*, 101(4):975–979, jan 2004.

[14] Seraina Klopfstein, Mathias Currat, and Laurent Excoffier. The fate of mutations surfing on the wave of a range expansion. *Mol. Biol. Evol.*, 23(3):482–490, mar 2006.

[15] Oskar Hallatschek, Pascal Hersen, Sharad Ramanathan, and David R. Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proc. Natl. Acad. Sci. U.S.A.*, 104(50):19926–19930, dec 2007.

[16] Laurent Excoffier and Nicolas Ray. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, 23(7):347–351, jul 2008.

[17] Oskar Hallatschek and David R. Nelson. Gene surfing in expanding populations. *Theoretical population biology*, 73(1):158–170, feb 2008.

[18] Oskar Hallatschek and David R. Nelson. Life at the front of an expanding population. *Evolution*, 64(1):193–206, jan 2010.

[19] Mark Kot, Mark A. Lewis, and P. Van Den Driessche. Dispersal data and the spread of invading organisms. *Ecology*, 77(7):2027–2042, 1996.

[20] James M. Bullock, Laura Mallada González, Riin Tamme, Lars Götzenberger, Steven M. White, Meelis Pärtel, and Danny A.P. Hooftman. A synthesis of empirical plant dispersal kernels. *Journal of Ecology*, 105(1):6–19, jan 2017.

[21] Oskar Hallatschek and Daniel S. Fisher. Acceleration of evolutionary spread by long-range dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, 111(46):E4911–E4919, nov 2014.

[22] Jayson Paulose, Joachim Hermisson, and Oskar Hallatschek. Spatial soft sweeps: Patterns of adaptation in populations with long-range dispersal. *PLoS Genetics*, 15(2), 2019.

[23] Jayson Paulose and Oskar Hallatschek. The impact of long-range dispersal on gene surfing. *Proceedings of the National Academy of Sciences*, page 201919485, mar 2020.

[24] Sewall Wright. EVOLUTION IN MENDELIAN POPULATIONS. *Genetics*, 16(2):97–159, 03 1931.

[25] Richard A. Nichols and Godfrey M. Hewitt. The genetic consequences of long distance dispersal during colonization. *Heredity*, 72(3):312–317, 1994.

[26] Kamal M Ibrahim, Richard A Nichols, and Godfrey M Hewitt. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, 77(3):282–291, sep 1996.

[27] Valérie Le Corre, Nathalie Machon, Rémy J. Petit, and Antoine Kremer. Colonization with long distance seed dispersal and genetic structure of maternally inherited genes in forest trees: A simulation study. *Genet. Res.*, 69(2):117–125, apr 1997.

[28] M. A. Lewis and S. Pacala. Modeling and analysis of stochastic invasion processes. *Journal of mathematical biology*, 41(5):387–429, 2000.

[29] Luzie U. Wingen, James K.M. Brown, and Michael W. Shaw. The population genetic structure of clonal organisms generated by exponentially bounded and fat-tailed dispersal. *Genetics*, 177(1):435–448, sep 2007.

[30] Nicolas Ray and Laurent Excoffier. A first step towards inferring levels of long-distance dispersal during past expansions. *Molecular Ecology Resources*, 10(5):902–914, sep 2010.

[31] C. E.G. Amorim, T. Hofer, N. Ray, M. Foll, A. Ruiz-Linares, and L. Excoffier. Long-distance dispersal suppresses introgression of local alleles during range expansions. *Heredity*, 118(2):135–142, feb 2017.

[32] C J Battey, Peter L Ralph, and Andrew D Kern. Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data. *Genetics*, 215(1):193–214, May 2020.

[33] Benjamin C Haller and Philipp W Messer. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, 36(3):632–637, mar 2019.

[34] Nathan Villiger and Jayson Paulose. The influence of explicit local dynamics on range expansions driven by long-range dispersal. *G3 Genes—Genomes—Genetics*, 13(5), 03 2023. jkad066.

[35] Kirill S Korolev, Melanie J I Muller, Nilay Karahan, Andrew W Murray, Oskar Hallatschek, and David R Nelson. Selective sweeps in growing microbial colonies. *Physical Biology*, 9(2), apr 2012.

[36] Sarah P Otto and Michael C Whitlock. The Probability of Fixation in Populations of Changing Size. *Genetics*, 146(2):723–733, 06 1997.

[37] Judith R. Miller. Survival of mutations arising during invasions. *Evolutionary Applications*, 3(2):109–121, 2010.

[38] Chris C R Smith, Silas Tittes, Peter L Ralph, and Andrew D Kern. Dispersal inference from population genetic variation using a convolutional neural network. *Genetics*, 04 2023.

[39] Sara Sheehan and Yun S. Song. Deep learning for population genetic inference. *PLOS Computational Biology*, 12(3):1–28, 03 2016.

[40] Andrew D Kern and Daniel R Schrider. diploS/HIC: An Updated Approach to Classifying Selective Sweeps. *G3 Genes—Genomes—Genetics*, 8(6):1959–1970, 06 2018.

[41] Lex Flagel, Yaniv Brandvain, and Daniel R Schrider. The Unreasonable Effectiveness of Convolutional Neural Networks in Population Genetic Inference. *Molecular Biology and Evolution*, 36(2):220–238, 12 2018.

[42] CJ Battey, Peter L Ralph, and Andrew D Kern. Predicting geographic location from genetic variation with deep neural networks. *eLife*, 9:e54507, jun 2020.

[43] Jeffrey R Adrion, Jared G Galloway, and Andrew D Kern. Predicting the Landscape of Recombination Using Deep Learning. *Molecular Biology and Evolution*, 37(6):1790–1808, 02 2020.

[44] Graham Gower, Pablo Iáñez Picazo, Matteo Fumagalli, and Fernando Racimo. Detecting adaptive introgression in human evolution using convolutional neural networks. *eLife*, 10:e64669, may 2021.

[45] Franz Baumdicker et al. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3):iyab229, 2022.

[46] Paul A Delcourt and Hazel R Delcourt. *Long-term forest dynamics of the temperate zone*. Springer (New York, NY), 1987.

[47] C. D. Thomas and J. J. Lennon. Birds extend their ranges northwards. *Nature*, 399(6733):213, may 1999.

[48] Louis D. Zeidberg and Bruce H. Robison. Invasive range expansion by the Humboldt squid, Dosidicus gigas, in the eastern North Pacific. *Proc. Natl. Acad. Sci. U.S.A.*, 104(31):12948–12950, jul 2007.

[49] S. D. Ling. Range expansion of a habitat-modifying species leads to loss of taxonomic diversity: A new and impoverished reef state. *Oecologia*, 156(4):883–894, jul 2008.

[50] Colin Robertson, Trisalyn A. Nelson, Dennis E. Jelinski, Michael A. Wulder, and Barry Boots. Spatial–temporal analysis of species range expansion: The case of the mountain pine beetle, Dendroctonus ponderosae. *J. Biogeogr.*, 36(8):1446–1458, aug 2009.

[51] Vitor Sousa, Stephan Peischl, and Laurent Excoffier. Impact of range expansions on current human genomic diversity. *Curr. Opin. Genet. Dev.*, 29:22–30, 2014.

[52] R. Bialozyt, B. Ziegenhagen, and R. J. Petit. Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, 19(1):12–20, 2006.

[53] K. S. Korolev, Mikkel Avlund, Oskar Hallatschek, and David R. Nelson. Genetic demixing and evolution in linear stepping stone models. *Reviews of Modern Physics*, 82(2):1691–1718, 2010.

[54] Joseph Felsenstein. A Pain in the Torus: Some Difficulties with Models of Isolation by Distance. *The American Naturalist*, 109(967):359–368, May 1975.

[55] Nicholas H. Barton, Jerome Kelleher, and Alison M. Etheridge. A new model for extinction and recolonization in two dimensions: Quantifying Phylogeography. *Evolution*, 64(9):2701–2715, 2010.

[56] 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, jan 2013.

[57] Bryan K Epperson. *Geographical genetics (MPB-38)*. Princeton University Press, 2003.

[58] Peter Ralph and Graham Coop. Parallel Adaptation: One or Many Waves of Advance of an Advantageous Allele? *Genetics*, 186(2):647–668, oct 2010.

[59] Jerome Kelleher, Kevin R. Thornton, Jaime Ashander, and Peter L. Ralph. Efficient pedigree recording for fast population genetics simulation. *PLOS Computational Biology*, 14(11):1–21, 11 2018.

[60] N. Berestycki, A. M. Etheridge, and A. Véber. Large scale behaviour of the spatial $\varLambda$-Fleming–Viot process. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49(2):374–401, 2013.

[61] Tyler B Smith and Daniel B Weissman. Isolation by distance in populations with power-law dispersal. *G3 Genes—Genomes—Genetics*, 13(4), 01 2023.

[62] Alexander J Nicholson. An outline of the dynamics of animal populations. *Australian journal of Zoology*, 2(1):9–65, 1954.

[63] Anne E. Ballmann, Miranda R. Torkelson, Elizabeth A. Bohuski, Robin E. Russell, and David S. Blehert. DISPERSAL HAZARDS OF PSEUDOGYMNOASCUS DESTRUCTANS BY BATS AND HUMAN ACTIVITY AT HIBERNACULA IN SUMMER. *Journal of Wildlife Diseases*, 53(4):725–735, 10 2017.

[64] Flavia Schlichta, Stephan Peischl, and Laurent Excoffier. The Impact of Genetic Surfing on Neutral Genomic Diversity. *Molecular Biology and Evolution*, 39(11), 11 2022.

[65] Antoine Moinet, Flavia Schlichta, Stephan Peischl, and Laurent Excoffier. Strong neutral sweeps occurring during a population contraction. *Genetics*, 220(4), 02 2022.

[66] Adam Eyre-Walker and Peter D. Keightley. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8):610–618, 2007.

[67] Justin M. J. Travis, Tamara Münkemüller, Olivia J. Burton, Alex Best, Calvin Dytham, and Karin Johst. Deleterious Mutations Can Surf to High Densities on the Wave Front of an Expanding Population. *Molecular Biology and Evolution*, 24(10):2334–2343, 08 2007.

[68] S. Peischl, I. Dupanloup, M. Kirkpatrick, and L. Excoffier. On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, 22(24):5972–5982, 2013.

[69] Matti Gralka, Fabian Stiewe, Fred Farrell, Wolfram Mobius, Bartlomiej Waclaw, and Oskar Hallatschek. Allele surfing promotes microbial adaptation from standing variation. *Ecology Letters*, 19(8):889–898, 2016.

[70] P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1):60â€"71, 1958.

[71] W J Ewens. The probability of survival of a new mutant in a fluctuating environment. *Heredity*, 22(3):438–443, 1967.

[72] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[73] F. W. Scholz and M. A. Stephens. K-sample anderson-darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.

[74] Hyunseok Lee, Jeff Gore, and Kirill S. Korolev. Slow expanders invade by forming dented fronts in microbial colonies. *Proceedings of the National Academy of Sciences*, 119(1):e2108653119, 2022.

[75] Benjamin C Haller and Philipp W Messer. Slim manual. `https://github.com/MesserLab/SLiM/releases/download/v4.0.1/SLiM_Manual.pdf`, 2022. Accessed 2022-11-01.

[76] Pierre Pudlo, Jean-Michel Marin, Arnaud Estoup, Jean-Marie Cornuet, Mathieu Gautier, and Christian P. Robert. Reliable ABC model choice via random forests. *Bioinformatics*, 32(6):859–866, 11 2015.

[77] Daniel R. Schrider and Andrew D. Kern. S/hic: Robust identification of soft and hard sweeps using machine learning. *PLOS Genetics*, 12(3):1–31, 03 2016.

[78] Mehreen R Mughal and Michael DeGiorgio. Localizing and Classifying Adaptive Targets with Trend Filtered Regression. *Molecular Biology and Evolution*, 36(2):252–270, 11 2018.

[79] Lauren Alpert Sugden, Elizabeth G. Atkinson, Annie P. Fischer, Stephen Rong, Brenna M. Henn, and Sohini Ramachandran. Localization of adaptive variants in human genomes using averaged one-dependence estimation. *Nature Communications*, 9(1):703, 2018.

[80] Daniel R. Schrider, Julien Ayroles, Daniel R. Matute, and Andrew D. Kern. Supervised machine learning reveals introgressed loci in the genomes of drosophila simulans and d. sechellia. *PLOS Genetics*, 14(4):1–29, 04 2018.

[81] Arun Durvasula and Sriram Sankararaman. A statistical model for reference-free inference of archaic local ancestry. *PLOS Genetics*, 15(5):1–18, 05 2019.

[82] Fernando A. Villanea and Joshua G. Schraiber. Multiple episodes of interbreeding between neanderthal and modern humans. *Nature Ecology & Evolution*, 3(1):39–44, 2019.

[83] François Chollet et al. Keras. `https://keras.io`, 2015.

[84] Sewall Wright. Isolation By Distance Under Diverse Systems of Mating. *Genetics*, 31(1):39–59, 01 1946.

[85] F. James Rohlf and Gary D. Schnell. An investigation of the isolation-by-distance model. *The American Naturalist*, 105(944):295–324, 1971.

[86] J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2):115–129, 1964.

[87] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[88] Ingwer Borg and Patrick Groenen. *Modern Multidimensional Scaling: Theory and Applications (Springer Series in Statistics)*. 08 2005.

[89] M Nei and W H Li. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10):5269–5273, 1979.

[90] Francois Rousset. Genetic differentiation between individuals. *Journal of Evolutionary Biology*, 13(1):58–62, 2000.

[91] Denis Mollison et al. The rate of spatial propagation of simple epidemics. In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, volume 3, pages 579–614. Berkeley and Los Angeles, University of California Edinburgh, UK, 1972.