Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Tyler B. Smith Date

Information Flow in Spatially Structured Populations

By

Tyler B. Smith

Doctor of Philosophy

Physics

Daniel Weissman, Ph.D.

Advisor

Gordon Berman, Ph.D.

Committee Member

Katia Koelle, Ph.D.

Committee Member

Ilya Nemenman, Ph.D.

Committee Member

Connie Roth, Ph.D.

Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

Date

Information Flow in Spatially Structured Populations

By

Tyler B. Smith

Advisor: Daniel B. Weissman, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Physics

Summer, 2021

Abstract

Information Flow in Spatially Structured Populations

By Tyler B. Smith

Spatial structure has a strong effect on biological systems at many scales, with information flow being limited by the rate at which individual constituents of the system move through space.

At the population scale, limited dispersal of individuals and the accumulation of mutations results in isolation by distance, in which individuals found further apart tend to be less related, as measured by the proportion of their genomes shared identical-by-descent. Classic models assume dispersal distances are drawn from a thin-tailed distribution and predict that relatedness should decrease exponentially as the separation between pairs becomes large. We study the effect of heavy-tailed dispersal on patterns of isolation by distance and find that a power-law dispersal kernel leads to power-law decay of relatedness at large distances and either power-law or logarithmic decay at short distances depending on the exact form of the kernel. The model is then used to solve the inverse problem of inferring dispersal from empirical isolation by distance curves.

At the cellular level, limited exchange and degradation of messenger molecules between cells bound the precision of gradient sensing. Precision increases with the length of a cell collective in the gradient direction up to this bound and then saturates. Intuition from studies of concentration sensing suggests that precision should also increase with detector length in the direction transverse to the gradient, since then spatial averaging should reduce the noise. However, here we show that, unlike for concentration sensing, the precision of gradient sensing decreases with transverse length for the simplest gradient sensing model, local excitation–global inhibition. The reason is that gradient sensing ultimately relies on a subtraction of measured concentration values. While spatial averaging indeed reduces the noise in these measurements, it also reduces the covariance between the measurements, which results

in the net decrease in precision. We demonstrate how a recently introduced gradient sensing mechanism, regional excitation–global inhibition, overcomes this effect and recovers the benefit of transverse averaging.

Work on an unrelated problem in quantum mechanics is also presented, and a self-contained introduction is given in the relevant chapter.

Information Flow in Spatially Structured Populations

By

Tyler B. Smith
B.S., University of South Florida, Tampa, Florida, 2015

Advisor: Daniel B. Weissman, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Physics

Summer, 2021

Contents

List of Figures

Acknowledgments

I would like to thank everyone who has supported and encouraged me during my time at Emory. While it would be impossible to list every single person who has helped me along the way, I can at least acknowledge some of them here.

First of all, I want to thank Daniel Weissman, my advisor, for years of research training and patient mentorship. I learned a great deal from Daniel about genetics and statistics, and even more about clear and careful thinking. He showed me that being a theorist is less about complicated equations and exact solutions than it is about stripping away details and understanding a problem in the simplest possible terms. He also encouraged me to take on problems that at first seemed too challenging to solve, and I hope to carry the confidence I developed in his group into all my future pursuits. It would be difficult to overstate his impact on my thinking.

I would also like to give my sincere thanks to the other members of my committee: Ilya Nemenman, Katia Koelle, Gordon Berman, and Connie Roth. Their guidance and feedback over the past few years have been beyond helpful. In addition to my committee, I would like to thank Professor Ajit Srivastava for mentorship during my first two years in the graduate program and many thought-provoking discussions on the connections between physics and geometry. I would also like to thank the staff members of the physics department, as well as the students and post-docs in both the physics and biology departments, for making my time at Emory both memorable and enjoyable. Special thanks to Weissman group members Qihan Liu, Linnea Bavik, Brent Allman and Rohan Mehta for putting up with my attempts at humor and making the last year of remote work during the coronavirus pandemic as engaging as possible.

Finally, I would like to thank my friends and family for all the support and encouragement they have given me during my time at Emory. I especially want to thank Ojay Liburd for helping me navigate and enjoy Atlanta, my uncle James Schockett for invaluable advice on weathering the challenges of a PhD program, and my parents and sister for a lifetime of love and support despite my many eccentricities. Most importantly, I want to thank my wife, Mallory Morrison, for enduring all the challenges of the last few years with me and exercising a truly supernatural amount of patience along the way. There is no way I could have made it to this point without her.

Tyler B. Smith
Emory University, Atlanta, USA
Summer 2021

Chapter 1

Introduction

1.1 Information flow via stochastic transport

Spatial structure can have strong effects on the behavior of both physical and biological systems, with distance mitigating the strength of interactions and the rate at which they can occur. In biology, spatial structure is known to affect processes at all scales, from ecological and population-scale dynamics to the function of cells and the spread of pathogens within a single individual [20, 67, 95, 113]. The dynamics of spatially structured systems can often be described in terms of the transport of individual constituents, which can possess both stochastic and deterministic components. Limitations on the rate of transport can limit the strength and rate of interactions, as well as the flow of signals and information [23, 80, 112].

Stochastic transport, i.e. random motion with displacements drawn from a probability distribution, plays a particularly important role across all of these scales [62]. At the molecular level, the dynamics of many systems are dominated by diffusion due to thermal fluctuations [21]. Even for systems in which a steady state exists, important properties are often determined by the rate of diffusion and the size of fluctuations about this equilibrium [27, 39].

At the population level, phenomenological models of plant, animal, and microbial movement often make similar assumptions of stochastic transport, with lineages within the population undergoing random walks through space over many genera-

tions [116]. While steady states for models describing the genetic structure of spatial populations often exist, deviations from these predictions due to finite population size and random motion make the inference of a population's dynamics from its genetic composition a significant challenge [30].

In this dissertation, we study problems at both scales. At the molecular scale, we consider models of cell-cell communication that use the diffusion of messenger molecules to detect chemical concentration gradients. We discuss the relevance of these models to chemotaxis below, and we show how both the noise in concentration and the noise in communication combine to limit the precision with which cells can sense chemical gradients.

At the population scale, we generalize classic models of equilibrium genetic structure in spatial populations to include the effects of superdiffusive, rather than just diffusive, movement. We give the necessary background on these models below, and we will show in a later chapter that this superdiffusive motion has important effects on genetic structure. We also develop an inference scheme to estimate the underlying parameters of the model using genomic data sampled from real populations.

For both systems we investigate, the random movement of individuals through space is an important means of information transmission. In addition to the rate of stochastic transport, decay rates due to either particle degradation (for molecules) or genetic mutations (for lineages) are important as well, as they bound the typical scale of information flow. We will see that the length scale set by the balance between stochastic transport and degradation limits both the precision of cellular gradient sensing and the genetic similarity of individuals sampled throughout space.

1.2 Gradient sensing and chemotaxis

At the molecular level, concentration sensing is one component of chemotaxis, which is the process of cells following chemical gradients. After detection, cells polarize and move up the gradient in an attempt to migrate towards sources of concentration [83]. The chemical of interest could be a nutrient, as is often the case for bacteria, or it could be a morphogen that governs the pattern of tissue development.

Chemotaxis along morphogen gradients is a critical part of embryonic development in mammals, and also aids in immune response and tissue regeneration in adults [97]. In addition to development and regeneration, chemotaxis plays an important role in the dissemination of cancerous tumors [143]. While prokaryotic organisms like bacteria use the change of chemical concentration over time to detect gradients while moving, eukaryotic cells instead use spatial averaging to allow for the detection of gradients while remaining stationary [125].

The formation of morphogen gradients can be modeled via a differential equation with terms that describe the diffusion and degradation of concentration [163]. Assuming the degradation term is linear, we expect the steady-state concentration profile to decay exponentially with distance, with the length scale of decay set by the balance between dispersal and degradation. If we allow for nonlinear degradation or drift due to cell lineage transport, concentration can decay according to a power law, rather than an exponential. We will see in the next section that analogous effects at the population level can lead to similar power-law profiles in models of isolation by distance.

In order to probe the limits of sensing precision, we focus on the case of a weak, constant gradient such that the concentration profile can be treated as linear. This can be thought of as the exponential profile described above in the limit of strong dispersal and weak degradation. Once a linear concentration profile has been established, we can model strategies of gradient detection that cells and cell collectives may choose to implement. A lower bound on the error in gradient measurements is set by the noise in an individual concentration measurement. A measurement of the gradient is really the difference between two measurements of concentration, and for a linear detector that measures concentration at its front and back, the expected value of this difference increases linearly with the size of the detector. The noise in the gradient, on the other hand, depends only on the average concentration and is independent of the total length of the detector. The classic theory of Berg and Purcell then predicts that the signal to noise ratio should increase quadratically with the length of the detector, and should be inversely proportional to the average (background) concentration [22, 112]. Note that this model assumes that two measurements of concentration can

be immediately compared, i.e., that information can be communicated between the front and the back of the detector instantaneously. In reality, information must be communicated between the two ends in finite time through some process that may introduce additional error.

To study the effects of communication noise on the sensing of weak chemical gradients, we consider a generalized model that allows for diffusive communication across a cell collective based on the principle of local excitation and global inhibition (LEGI) [84, 96]. In the LEGI model, there are two species of messenger molecules. One is confined to a given cell, while the other can be exchanged among cells. Both messengers are produced in each cell at a rate proportional to the local external concentration, but because of the exchange of the "global" species, we expect its count in a given cell to reflect the average concentration over many cells, rather than just the local concentration of the current cell. Because messenger molecules also degrade at a known rate, the distance a global messenger molecule typically travels before degrading is given by the ratio of its exchange and degradation rates. By subtracting the "global" count from the "local" count, individual cells within the detector determine whether their local concentration is above or below the average concentration found over cells within the length given by the exchange-degradation ratio. Again, the balance of diffusion and degradation sets the characteristic length scale, but rather than representing the gradient of the concentration profile, this length scale can be seen as a communication length, setting limits to the precision with which chemical concentration gradients can be measured by cells. The prediction for the signal to noise ratio of the one-dimensional LEGI model has the same form as the Berg-Purcell expression, but we replace the actual length of the detector with the communication length when the detector is sufficiently large. Above the communication length, the precision of gradient sensing saturates with detector length.

1.3 Isolation by distance

At the population level, stochastic transport is known as dispersal, and we can model the genetic diversity of spatial populations via rate equations for the allele

frequencies at all points in space [89]. Just as in the LEGI model at the molecular scale, we again have both exchange (dispersal) and degradation (mutation), with the resulting equations being linear if the effects of natural selection are negligible [91]. Spatial correlations in allele frequency can be interpreted as the fraction of genes we expect to be shared between two individuals sampled at a given separation [91]. It can in fact be shown that the stepping stone model governing allele frequencies is equivalent to a model of "coalescence", in which we consider two individuals sampled at present day, and trace the trajectories of their ancestors back through time until the time of their most recent common ancestor (TMRCA) [15]. While counterintuitive at first, this backwards time formulation is extremely helpful given that it restricts the analysis to the ancestral lineages of sampled individuals, rather than the entire population.

Using the backwards time formulation, known as coalescent theory, we can obtain a probability distribution over possible TMRCA for two individuals in a population [138]. If a population reproduces sexually, recombination will lead to different parts of the genome following distinct trajectories back through time. Thus, in general, pairs will have many "most recent common ancestors", and multiple TMRCA that vary across the genome. From this distribution of TMRCA, or "coalescence times", we can make predictions about levels of relatedness in the present-day population. Assuming mutations occur at a constant rate μ, the probability of identity (fraction of identical genes) is simply the generating function of the coalescence time distribution [15]. Though we focus on the case of spatially structured populations here, it is instructive to review the results of coalescent theory for a pair of individuals in a well-mixed population without spatial structure. We can think of this as the limit of a small spatial range in which two individuals are always close enough to coalesce in the previous generation, i.e., to share a parent. In such a population, the probability of coalescence follows an exponential distribution (backwards) in time, with an average coalescence time proportional to the total size of the population [122]. The resulting probability of identity for any pair sampled from the population decreases linearly with the product of mutation rate and population size (provided that this product is small). As the population becomes larger, the typical time to coalescence becomes

longer, and we expect more mutations to accumulate between individuals. Likewise, if the mutation rate becomes larger, we expect more mutations to accumulate for a given population size and coalescence time distribution.

For populations in which the range size is large, limited dispersal of individuals can lead to the distribution of coalescence times, and thus the probability of identity, depending on the spatial separation between pairs [20]. The tendency of genetic identity to decrease with spatial separation is known as "isolation by distance". Just as for concentration profiles at the molecular level, if dispersal is diffusive and the mutation rate is constant, the resulting isolation by distance profile is exponential in one dimension, with the characteristic length scale of decay set by the ratio of the dispersal constant and the mutation rate [15].

While we will focus on neutral populations in this work, if selection is present, some individuals in the population will produce more offspring than others. Fitness effects result in what are known as selective sweeps, in which individuals possessing beneficial mutations produce many offspring that sweep through space as a travelling wave [18]. Looking backwards in time, the effect of these sweeps (assuming recombination is fast) can be accounted for by adding a drift term that pulls individuals at present day back towards the locations of fit ancestors. Just as in the case of concentration with drift due to cellular lineage transport, this drift term will lead to isolation by distance decaying according to a power law instead of an exponential [4].

Even without selection, power laws can arise in neutral populations if diffusive dispersal is replaced with superdiffusive motion, which is "faster" than diffusion in the sense that the typical displacement distance increases superlinearly with time. At the molecular scale, superdiffusive motion is typically modeled via Lévy walks, which assume particles move at a constant velocity, or continuous-time random walks (CTRW), models that assume trajectories consist of instantaneous jumps and random waiting times between jumps [107, 172]. CTRW models used to model morphogen gradient formation at the molecular level predict that the steady-state profile remains exponential even when motion becomes superdiffusive [92]. Subdiffusive CTRW models make similar predictions: concentration takes the form of an exponential with a time-dependent characteristic length scale [78].

Superdiffusive motion in population genetics tends to be modeled as a Lévy flight rather than a Lévy walk or CTRW [19, 40, 75, 114]. Lévy flights are simpler mathematically, and equate to replacing the Gaussian dispersal kernel describing Brownian motion with a heavy-tailed, power-law kernel [86, 108, 119]. This model can also be seen as a special case of a CTRW in which waiting times between jumps are constant, rather than being random variables drawn from a power-law distribution. While Lévy flights can have unbounded velocities and divergent mean squared displacement, we will show that the resulting population-genetic models can predict patterns of isolation by distance for discrete-time simulations on finite spatial ranges. We can thus think of Lévy flights as convenient approximations of truncated power-law distributions with more physically reasonable properties. In fact, many real populations appear to have heavy-tailed dispersal distributions that are better described by a power law than a thin-tailed distribution [2, 7, 11, 34, 45, 46, 48, 63, 159, 167]. It was shown by Nagylaki that, for a subset of Lévy flights in one dimension, the resulting isolation by distance profile has a well defined steady state, and that the steady-state profile has a power-law, rather than exponential, tail [114].

1.4 The present work

In this work, we extend previous results for models of cell-cell communication and isolation by distance in order to better understand the effects of dimensionality and superdiffusive motion on information flow in spatially structured biological systems. In both cellular models of gradient sensing and population-level models of isolation by distance, we find that increasing dimensionality can lead to qualitative changes in the behavior of the system.

For cellular models of gradient sensing, we show that the effects of spatial averaging can be counterintuitive. In two dimensions, increasing the width of a detector transverse to a concentration gradient can decrease the precision with which the gradient is measured. While transverse averaging will always improve concentration sensing, gradient sensing depends on the covariance between two distinct concentrations measurements, and the decrease of this covariance with transverse averaging

decreases precision. In order to reverse this effect, the local reporter molecule must also be allowed to disperse. We study this generalization with the recently developed regional excitation global inhibition (REGI) model and show that, for REGI, the scaling of precision with transverse detector width is non-monotonic and leads to optimal detector shapes in two and three dimensions

Most of the work in this dissertation focuses on population-genetic models of isolation by distance, and for these models, we extend previous results for Lévy flight dispersal to arbitrary power-law dispersal kernels in both one and two dimensions. We show that there is an important interplay between dispersal and dimension, with the underlying dynamics of lineages changing dramatically at a dimension-dependent value of the power-law dispersal exponent.

The model developed for isolation by distance with power-law dispersal is then used to solve the inverse problem of inferring dispersal from empirical isolation by distance curves. We show that our inference scheme performs well with simulated data and apply the method to natural populations to infer the relevant parameters and detect signatures of heavy-tailed dispersal. We also extend the model to predict multilocus summaries of genetic similarity, namely the size and number of long blocks of shared sequence along the genome [131, 134]. We then develop an inference scheme for this generalization and test its performance with simulated data before applying the method to real populations.

In addition to the work described above, Chapter 5 of this dissertation describes results on an unrelated problem in quantum mechanics. A self-contained introduction to that work is given in section 5.1.

Chapter 2

Role of Spatial Averaging in Multicellular Gradient Sensing

Material presented in this chapter was published in Smith et al. [151]

2.1 Introduction

Determining the strength and direction of a chemical concentration gradient is an essential task for a diverse array of biological processes. Gradient sensing underlies the polarization of single cells, the orientation and migration of cells and cell collectives, and the changes in tissue morphology that occur during embryogenesis and the subsequent development of an organism [29, 51, 52, 64, 81, 101, 103, 123, 129, 139, 153, 156]. Experiments have shown that cells are remarkably precise gradient sensors [52, 139], and a large amount of effort has gone into understanding the mechanisms of, and the limits to, biological gradient sensing [22, 53, 54, 70, 79, 87, 96, 112, 123].

At its core, gradient sensing requires the comparison of concentration measurements between the "front" and the "back" of a detector. Front and back here are defined with respect to the gradient direction, and the detector here is a single cell or a group of cells. If the front and back are more separated, then the concentration measurements are more different from each other, which improves the determination of the gradient. This implies that detectors that are longer in the gradient direc-

tion have a higher gradient sensing precision [53, 54, 70, 79]. This argument neglects the fact that information must be communicated between different parts of a detector, especially if the detector is multicellular. Recently we derived the limits to the precision of gradient sensing including communication, and we found that for a one-dimensional (1-D) detector, the precision indeed increases with detector length, but then saturates due to the fact that communication introduces its own noise [52, 112]. Nonetheless, the precision of gradient sensing increases or saturates with the length of a 1-D detector aligned with the gradient; it does not decrease.

Yet biological detectors are not 1-D in general. Two-dimensional (2-D) detectors include the quasi-cylindrical arrangement of cell nuclei during the early stages of Drosophila development [71] and the planar arrangement of epithelial cell layers [103]. Three-dimensional (3-D) detectors include single cells and the multicellular tips of growing epithelial ducts [56], as well as border cells exhibiting collective guidance in Drosophila [29]. This raises the question of what effect the dimensions transverse to the gradient direction have on the precision of gradient sensing.

Intuition about this question can be drawn from the similar task of sensing the value of a concentration (as opposed to sensing its difference between two points in space, i. e., the gradient). If the concentration profile is uniform in space, then the precision of concentration sensing benefits from increasing the detector length in any direction. The reason is that communication with other parts of the detector, or *spatial averaging*, does not change the mean of a particular measurement within the detector, but it does reduce the noise [22, 53, 54]. Even if the concentration profile is graded, but the goal is still concentration (rather than gradient) sensing, as in stripe formation in early Drosophila development, the precision still benefits from spatial averaging [55].[1] The benefit is especially clear in a direction transverse to the gradient direction: once again, spatial averaging in this direction does not change the mean of a particular measurement, but it does reduce the noise. These considerations, drawn from the problem of concentration sensing, suggest that the precision of gradient

[1]The distinction between gradient sensing, and concentration sensing with a graded profile, is a subtle but important one, and is further discussed in Results section 1 and the Discussion.

sensing should also increase with the length of a detector in a direction transverse to the gradient.

Here we investigate theoretically and computationally the precision of gradient sensing for 2-D and 3-D detectors. We start with one of the simplest models of gradient sensing, the local excitation–global inhibition (LEGI) model [87, 96]. This is an accepted basic model when gradient sensing is adaptive (that is, background concentration largely does not effect the gradient sensing). Surprisingly, in contrast to the case of concentration sensing, we find that the precision of gradient sensing decreases with the length of the detector in a direction transverse to the gradient direction. The reason is that gradient sensing fundamentally relies on a subtraction of concentration measurements, e.g. between the front and back of the detector. While spatial averaging reduces the intrinsic noise in these measurements, which increases precision, it also reduces the covariance between the measurements, which decreases precision. We demonstrate that the latter effect dominates, such that the net result is a decrease in precision with transverse detector size. Then we show that this decrease can actually be overcome by a gradient-sensing strategy that we recently introduced, termed regional excitation–global inhibition (REGI) [112]. We demonstrate that REGI retains a high covariance between measurements and restores the benefit of transverse averaging. Using a REGI-based model, we compute the optimal 2-D and 3-D detector shapes, which arise from an interplay of the effects of transverse averaging on both the signal and the noise of gradient detection. We argue that these shapes are consistent with the shapes of the multicellular tips of epithelial ducts, suggesting that this and other similarly shaped gradient-sensing systems benefit from spatial averaging in all dimensions.

2.2 Methods

As in previous work [52, 112], we consider the local excitation–global inhibition (LEGI) model of multicellular gradient sensing, which is a minimal, adaptive, spatially extended model of gradient sensing. We consider a signal concentration profile c that varies linearly in a particular direction in 3-D space, with concentration gradient

g (Fig. 2.1A, C). In the nth cell, both a local molecular species X and a global molecular species Y are produced at a rate β and degraded at a rate μ. The production rate is also proportional to the number of signal molecules in the cell's vicinity $c_n a^3$, where a is the cell diameter. Whereas the local species X is confined to each cell, the global species Y is exchanged between neighboring cells at a rate γ_y (Fig. 2.1C; note that although all cells are producing X and Y molecules, we show as examples in Fig. 2.1A and C only those molecules originating from the rightmost, middle cell). Conceptually, X measures the local concentration of signal molecules, while Y represents their spatially-averaged concentration. As in [52, 112] we consider the linear response regime, in which the dynamics of the local and global species satisfy the stochastic equations

$$\begin{aligned} \frac{dx_n}{dt} &= \beta(c_n a^3) - \mu x_n + \eta_n, && (2.1) \\ \frac{dy_n}{dt} &= \beta(c_n a^3) - \mu y_n + \gamma_y \sum_{n' \in \mathcal{N}(n)} (y_{n'} - y_n) + \xi_n \\ &= \beta(c_n a^3) - \mu \sum_{n'} M^y_{nn'} y_{n'} + \xi_n. && (2.2) \end{aligned}$$

Here $M^y_{nn'} \equiv (1+|\mathcal{N}_n|\gamma_y/\mu)\delta_{nn'} - (\gamma_y/\mu)\delta_{n' \in \mathcal{N}_n}$ is the connectivity matrix for the global species that accounts for degradation and molecule exchange. $\mathcal{N}_n$ and $|\mathcal{N}_n|$ denote the indices and the number of nearest neighbors of cell n, respectively. The intrinsic noise terms η_n and ξ_n correspond to the Poissonian production, degradation, and exchange reactions [52]. The signal c_n also fluctuates, which introduces extrinsic noise. As described below, in this work we assume that these fluctuations are slow compared to the downstream signal processing, which is equivalent to assuming either slow diffusion of signaling molecules or instantaneous downstream processing, and leads to Poisson-distributed signal molecule counts [52].

In the LEGI paradigm, X excites a downstream species while Y inhibits it. If the cell is at the higher edge of the gradient, then the local concentration (X) is higher than the spatial average (Y), and the excitation exceeds the inhibition. While such comparison of the excitation and the inhibition can be done by many different molecular mechanisms [87], we consider here the limit of shallow gradients, where the

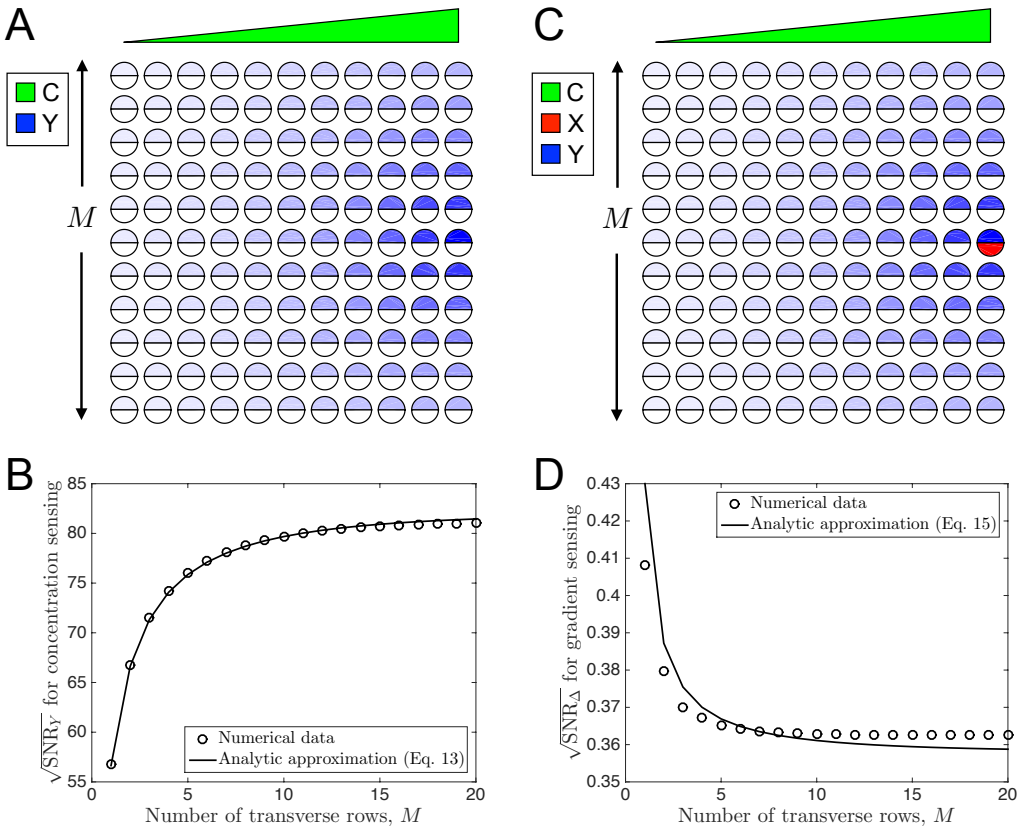

Figure 2.1: Spatial averaging transverse to a gradient improves concentration sensing, but worsens gradient sensing. (A) A 2-D array of cells is exposed to a concentration profile C that varies linearly in the horizontal direction (green wedge). In each cell, Y molecules are produced in proportion to the local C value. Y molecules are also exchanged between neighboring cells, providing the spatial averaging. Thus Y is the readout for the average concentration in the vicinity of a particular cell. Blue indicates the mean number of Y molecules $\bar{y}$ in each cell that have originated from the rightmost, middle cell. (B) The signal-to-noise ratio (SNR) for y increases with the number M of rows of cells added transverse to the gradient direction. (C) As in A, but with an additional internal species X. The molecules are also produced in proportion to the local C value, but they are not exchanged between cells. Red indicates the mean number of X molecules $\bar{x}$ in each cell that have originated from the rightmost, middle cell. The difference $\Delta = x - y$ provides the readout for the gradient (LEGI). (D) In contrast to B, the SNR for Δ decreases with the number of transverse rows M. In B and D, the numerical results are compared with the theoretical approximations (see Eqs. 2.13 and 2.15, respectively) and agree at small M as expected. Parameters are similar to the experiments in [52]: $\bar{c}_N = 1.25$ nM, $g = 0.5$ nM/mm, $a = 10$ μm, $n_y = \sqrt{\gamma_y/\mu} = 4$, $N = 50$ cells per row, and $G = \beta/\mu = 10$. In B and D the numerical value of $\bar{y}_N$ is used in the approximations.

comparison is equivalent to subtracting Y from X [52]. This difference, $\Delta_n = x_n - y_n$, is the readout of the model. If Δ_n is positive, the nth cell is further up the gradient than average; if Δ_n is negative, the nth cell is further down the gradient than average. In this work, we always focus on the readout Δ_N of the cell highest up the gradient, which we denote as the Nth cell.

We assume that the cells do not average concentrations of the signal C and the messenger molecules X and Y over time (though generalizations with averaging are certainly possible [112]). Then the precision of gradient sensing is given by the square root of the instantaneous signal-to-noise ratio (SNR) for the readout, $\mathrm{SNR}_\Delta = (\bar{\Delta}_N/\delta\Delta_N)^2$, where the mean and variance are given by [52]

$$\bar{\Delta}_N = \bar{x}_N - \bar{y}_N, \tag{2.3}$$

$$\bar{x}_N = Ga^3\bar{c}_N, \tag{2.4}$$

$$\bar{y}_N = Ga^3\sum_n K_n^y \bar{c}_{N-n}, \tag{2.5}$$

and

$$(\delta\Delta_N)^2 = (\delta x_N)^2 + (\delta y_N)^2 - 2\mathrm{cov}(x_N, y_N), \tag{2.6}$$

$$(\delta x_N)^2 = \bar{x}_N + G^2a^3\bar{c}_N, \tag{2.7}$$

$$(\delta y_N)^2 = \bar{y}_N + G^2a^3\sum_n (K_n^y)^2\bar{c}_{N-n}, \tag{2.8}$$

$$\mathrm{cov}(x_N, y_N) = G^2a^3K_0^y\bar{c}_N, \tag{2.9}$$

respectively. Here $K_n^y \equiv (M^y)^{-1}_{N,N-n}$ is the communication kernel, and $G \equiv \beta/\mu$ is the gain. The first terms in Eqs. 2.7 and 2.8 correspond to intrinsic noise, while the second terms correspond to extrinsic noise and assume that the diffusion of the signal is slow [52]. Computing the precision for a given configuration of cells only requires inverting the connectivity matrix M^y. That is, because our model is linear, both the means and (co)variances are obtained directly by matrix inversion, and no stochastic simulations are performed.

In the recently introduced regional excitation–global inhibition (REGI) model [112], the local species X is also exchanged among cells, but at a lower rate $\gamma_x < \gamma_y$.

Then Eq. 2.1 becomes analogous to Eq. 2.2, and Eqs. 2.4, 2.7, and 2.9 are replaced by

$$\bar{x}_N = Ga^3 \sum_n K_n^x \bar{c}_{N-n}, \tag{2.10}$$

$$(\delta x_N)^2 = \bar{x}_N + G^2 a^3 \sum_n (K_n^x)^2 \bar{c}_{N-n}, \tag{2.11}$$

$$\mathrm{cov}(x_N, y_N) = G^2 a^3 \sum_n K_n^x K_n^y \bar{c}_{N-n}, \tag{2.12}$$

respectively, where $K_n^x \equiv (M^x)^{-1}_{N,N-n}$ is the communication kernel for the local species, and $M^x_{nn'} \equiv (1 + |\mathcal{N}_n|\gamma_x/\mu)\delta_{nn'} - (\gamma_x/\mu)\delta_{n' \in \mathcal{N}_n}$. Once more, computing the precision for a given configuration of cells in the REGI model only requires inverting the connectivity matrices M^x and M^y. While diffusion of X decreases $\bar{x}_N$ at the Nth cell, and hence decreases the difference $\bar{\Delta}_N$, it also averages X over a larger volume, hence decreasing its noise. As shown in Ref. [112], under a broad range of conditions, the decrease in the noise dominates, and the overall precision of the REGI model is higher than that of LEGI.

2.3 Results

2.3.1 Concentration sensing precision increases with transverse detector size

Before investigating gradient sensing, we focus on the simpler problem of concentration sensing. In the local excitation–global inhibition (LEGI) model, both X and Y provide readouts of the local concentration, while their difference Δ provides a readout of the gradient. The concentration readout provided by Y is spatially averaged, whereas the concentration readout provided by X is not. Even if the signal profile is graded, X and Y are concentration readouts if viewed independently (with different spatial averaging), not gradient readouts. For example, during Drosophila development, the morphogen profiles are graded, but individual nuclei in the embryo measure (and threshold) the local concentration, possibly with some spatial averaging [55, 71, 94, 152].

How does the precision of concentration sensing depend on transverse detector size? To answer this question, we focus on the spatially averaged concentration readout Y. We consider a linear signal profile with gradient g and compute the SNR of Y in the Nth cell, as we vary the number M of rows of cells in a direction transverse to the gradient (Fig. 2.1A). We see in Fig. 2.1B (circles) that the precision of concentration sensing increases with M. The reason is that adding rows of cells transverse to the gradient allows for Y molecules to be exchanged between rows (in addition to along each row). This does not change the mean $\bar{y}_N$ due to the translational symmetry in the transverse direction. However, it does reduce the variance, since the global species Y is now averaged over more cells. The net effect is an increase in the SNR beyond what is allowed by longitudinal averaging.

We can elucidate the effect of spatial averaging more quantitatively by appealing to the expression for the variance in Y (Eq. 2.8). In a single row of cells, and in the limit of many cells ($N \gg 1$) and fast communication ($\gamma_y \gg \mu$), the kernel reduces to $K^y_n \approx e^{-n/n_y}/n_y$, where $n_y \equiv \sqrt{\gamma_y/\mu}$ sets the effective length scale of communication [52]. Approximating the sum in Eq. 2.8 as an integral that extends to infinity (since N is large) obtains $(\delta y^{(1)}_N)^2 \approx \bar{y}_N + G^2 a^3 \bar{c}_{N-n_y/2}/(2n_y)$ for a single row. In the case of M rows, Y is averaged with the communication kernel K^y over M cells transverse to the gradient. This will not affect the intrinsic component of the variance (since the mean is unchanged), but the extrinsic component will be reduced according to $(\delta y^{(M)}_N)^2_{\text{ext}} = \sum_{m=-M/2}^{M/2} (K^y_m)^2 (\delta y^{(1)}_N)^2_{\text{ext}}$. We again make the exponential kernel approximation, this time normalizing over the finite domain of size M, giving $K^y_m \approx e^{-|m|/n_y} / \sum_{m'=-M/2}^{M/2} e^{-|m'|/n_y}$. Finally, again approximating the sums as integrals, we obtain

$$(\delta y_N)^2 \approx \bar{y}_N + \frac{G^2 a^3 \bar{c}_{N-n_y/2}}{2n_y} \frac{n_y(1-e^{-M/n_y})}{[2n_y(1-e^{-M/(2n_y)})]^2}. \tag{2.13}$$

In deriving this approximation, we have neglected effects that the transverse edges have on the exponential shape of the kernel, as well as correlations between exchange reactions parallel and perpendicular to the gradient. Nonetheless, the SNR calculated using this approximation is compared with the numerical result in Fig. 2.1B, and we

see that the agreement is excellent. In the limit $M \ll n_y$, Eq. 2.13 simplifies to

$$(\delta y_N)^2 \approx \bar{y}_N + \frac{G^2 a^3 \bar{c}_{N-n_y/2}}{2 n_y M}. \tag{2.14}$$

We see that for a small number of rows, the averaging over rows is nearly uniform, and the extrinsic component of the variance is reduced by a factor of M, as expected.

2.3.2 Gradient sensing precision decreases with transverse detector size

We now turn our attention to gradient sensing. How does the precision of gradient sensing depend on transverse detector size? To answer this question for a linear signal profile, we compute the SNR of the gradient readout Δ_N as a function of the number M of rows of cells in a direction transverse to the gradient (Fig. 2.1C). We see in Fig. 2.1D that the precision of gradient sensing decreases with M (circles). This is in contrast to the precision of concentration sensing, which increases with M (Fig. 2.1B).

To understand why the precision of gradient sensing decreases with M, we once again consider the mean and the variance of the readout. The mean $\bar{\Delta}_N = \bar{x}_N - \bar{y}_N$ does not change with M because neither $\bar{x}_N$ nor $\bar{y}_N$ changes with M. However, the variance $(\delta\Delta_N)^2 = (\delta x_N)^2 + (\delta y_N)^2 - 2\mathrm{cov}(x_N, y_N)$ changes with M due to two effects. First, the variance in the global species $(\delta y_N)^2$ decreases with M due to spatial averaging, as discussed in the previous section. Second, the covariance $\mathrm{cov}(x_N, y_N)$ also decreases with M because Y is exchanged with a larger number of cells, whereas X is not exchanged, so the two covary more weakly. The effects have opposite signs. To understand which effect dominates, we again appeal to analytic approximation. For a single row of cells, under the exponential kernel approximation, the covariance in Eq. 2.9 reduces to $\mathrm{cov}^{(1)}(x_N, y_N) \approx G^2 a^3 \bar{c}_N / n_y$. For M rows of cells, since only Y is exchanged, the covariance is reduced according to $\mathrm{cov}^{(M)}(x_N, y_N) = K^y_{m=0} \mathrm{cov}^{(1)}(x_N, y_N)$. Making the same approximation as above for K^y_m of an exponential in a finite domain, we

obtain

$$(\delta\Delta_N)^2 \approx (\delta x_N)^2 + \bar{y}_N + \frac{G^2a^3\bar{c}_{N-n_y/2}}{2n_y}\frac{n_y(1-e^{-M/n_y})}{[2n_y(1-e^{-M/(2n_y)})]^2} - 2G^2a^3\frac{\bar{c}_N}{n_y}\frac{1}{2n_y(1-e^{-M/(2n_y)})}. \quad (2.15)$$

The SNR calculated using this approximation is compared with the numerical result in Fig. 2.1D, and we see good agreement. In the limit $M \ll n_y$, Eq. 2.15 simplifies to

$$(\delta\Delta_N)^2 \approx (\delta x_N)^2 + \bar{y}_N + \frac{G^2a^3\bar{c}_{N-n_y/2}}{2n_yM} - 2\frac{G^2a^3\bar{c}_N}{n_yM} \quad (2.16)$$

$$= (\delta x_N)^2 + \bar{y}_N - \frac{G^2a^3[4\bar{c}_N - \bar{c}_{N-n_y/2}]}{2n_yM}. \quad (2.17)$$

Eq. 2.16 shows that in this limit of near-uniform averaging, both (i) the extrinsic component of the variance in Y and (ii) the covariance are reduced by a factor of M, as expected. Furthermore, because the Nth cell is at the highest concentration, we have $\bar{c}_N > \bar{c}_{N-n_y/2}$, and we see that Eq. 2.17 is an increasing function of M. Thus, this limit elucidates the fact that the decrease of the covariance dominates over the decrease of the variance in Y, causing the variance of Δ_N to increase with M for all parameter values. Because the mean $\bar{\Delta}_N$ does not change with M, we conclude that the precision of gradient sensing decreases with transverse detector size.

2.3.3 REGI mechanism recovers the benefit of transverse averaging

In the previous section we saw that the precision of gradient sensing using the LEGI model (local messenger X is not exchanged among the cells) decreases with the size of a detector in a direction transverse to the gradient, due to the fact that the covariance between the subtracted variables decreases with the transverse size. For the REGI model, exchange of the X molecules has an additional effect beyond increasing the sensing precision for 1-D line of cells [112]: it increases the covariance of X and Y, compared to the LEGI mechanism. Indeed, now both X and Y are downstream signals from some of the same external ligand molecules. Since the

decrease of gradient sensing precision with transverse detector size is due to the loss of covariance (Fig. 2.1D), this raises the question of whether the REGI strategy can overcome this effect and allow gradient sensing precision to benefit from transverse averaging.

To answer this question, we once again consider a linear signal profile, and we compute the SNR of the gradient readout Δ_N under the REGI model (see Methods), as a function of the number M of rows of cells in a direction transverse to the gradient (Fig. 2.2A). We see in Fig. 2.2B that for a sufficiently large value of $n_x \equiv \sqrt{\gamma_x/\mu}$, which sets the lengthscale of spatial averaging for the local species, the precision of gradient sensing increases with M. This is in contrast to the case of LEGI, for which the precision decreases with M (Fig. 2.1D and black circles in Fig. 2.2B). Therefore, the recovery of covariance between X and Y in the REGI mechanism avoids the loss of gradient sensing precision and restores the benefit of transverse averaging.

To understand this effect quantitatively, we turn once more to analytic approximation. The variance in X (Eq. 2.11) and Y (Eq. 2.8) for M rows of cells will be approximated by expressions of the form of Eq. 2.13. The covariance (Eq. 2.12) for a single row of cells under the exponential kernel approximation for both X and Y is $\mathrm{cov}^{(1)}(x_N, y_N) \approx G^2 a^3 \bar{c}_{N-\bar{n}}/(n_x + n_y)$, where $\bar{n} \equiv n_x n_y/(n_x + n_y)$. For M rows of cells the covariance is therefore $\mathrm{cov}^{(M)}(x_N, y_N) = \sum_{m=-M/2}^{M/2} K_m^x K_m^y \mathrm{cov}^{(1)}(x_N, y_N)$. Again approximating K_m^x and K_m^y as exponentials in a finite domain, we obtain

$$\begin{aligned}(\delta\Delta_N)^2 \;\approx\; & \bar{x}_N + \frac{G^2 a^3 \bar{c}_{N-n_x/2}}{2n_x} \frac{n_x(1-e^{-M/n_x})}{[2n_x(1-e^{-M/(2n_x)})]^2} \\ & + \bar{y}_N + \frac{G^2 a^3 \bar{c}_{N-n_y/2}}{2n_y} \frac{n_y(1-e^{-M/n_y})}{[2n_y(1-e^{-M/(2n_y)})]^2} \\ & - 2\frac{G^2 a^3 \bar{c}_{N-\bar{n}}}{n_x+n_y} \frac{\bar{n}(1-e^{-M/(2\bar{n})})}{2n_x(1-e^{-M/(2n_x)})n_y(1-e^{-M/(2n_y)})}. \end{aligned} \tag{2.18}$$

This approximation has assumed $\{n_x, n_y\} \gg 1$, and we find that for $n_y = 4$ the SNR calculated using this approximation agrees very well with the numerical result for $n_x \geq 1$; see Fig. 2.2B. In the limit $M \ll n_x$, Eq. 2.18 simplifies to

$$(\delta\Delta_N)^2 \approx \bar{x}_N + \frac{G^2 a^3 \bar{c}_{N-n_x/2}}{2n_x M} + \bar{y}_N + \frac{G^2 a^3 \bar{c}_{N-n_y/2}}{2n_y M} - 2\frac{G^2 a^3 \bar{c}_{N-\bar{n}}}{(n_x+n_y)M}, \tag{2.19}$$

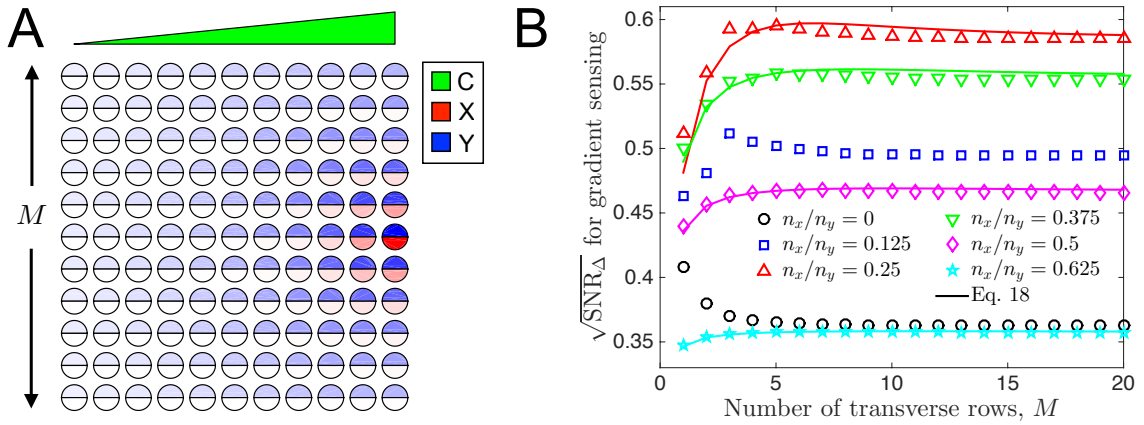

Figure 2.2: The regional excitation–global inhibition (REGI) strategy allows cells to exploit transverse spatial averaging for gradient sensing. (A) As in Fig. 2.1C, but for REGI. X molecules are exchanged between neighboring cells, at a lower rate than Y molecules. The difference $\Delta = x - y$ still provides the readout for the gradient. (B) In contrast to Fig. 2.1D, for sufficiently large communication length n_x the SNR *increases* with the number of transverse rows M, before ultimately decreasing, which leads to an optimum as a function of M. Since $n_x = 0$ (LEGI) and $n_x = n_y = 4$ (no sensing) are suboptimal, a global optimum emerges over both M and n_x. Parameters are as in Fig. 2.1, with $n_x = 1$ in A, which is near its optimal value as seen in B. In B the numerical value of $\bar{y}_N$ is used in the approximations.

in which the extrinsic components of the variances and the covariance are reduced by M, as expected. If we further assume that the gradient is shallow compared to the background concentration ($ag \ll \bar{c}_N$), we may approximate $\bar{c}_{N-n_x/2} \approx \bar{c}_{N-n_y/2} \approx \bar{c}_{N-\bar{n}} \approx \bar{c}_N$, yielding

$$(\delta\Delta_N)^2 \approx \bar{x}_N + \bar{y}_N + \frac{G^2 a^3 \bar{c}_N}{n_y M}\left[\frac{1}{2\rho} + \frac{1}{2} - \frac{2}{\rho+1}\right], \tag{2.20}$$

where $\rho \equiv n_x/n_y$. The expression in brackets in Eq. 2.20 is positive for all $0 < \rho < 1$, which demonstrates analytically that in this limit the variance in the readout decreases with M, and therefore that the REGI strategy restores the benefit of transverse averaging.

We also see in Fig. 2.2B that a maximal precision emerges in the REGI model as a function of M at a particular number of rows M^*. This maximum is due to the fact that the exchange of X, which causes an increase in precision with M, and the

exchange of Y, which causes a decrease in precision with M, occur on different length scales, $n_x < n_y$. Indeed, we see that as n_x increases, the location of the maximum M^* increases concomitantly. Additionally, we see that the maximal precision value first increases with n_x, then decreases with n_x, leading to an optimal value n_x^*. This is due to the previously understood tradeoff that is introduced when n_x increases: on the one hand the variance of X is reduced, which increases precision; on the other hand, the means of X and Y are more similar, which decreases the precision [112]. Here this tradeoff is modified by the additional benefit of increasing n_x, namely that it increases the covariance of X and Y in the transverse direction, and thus further reduces the noise in gradient sensing.

2.3.4 Emergence of optimal detector shapes in two and three dimensions

The emergence of an optimal number of transverse rows of cells, seen in the previous section, raises the more general question of whether there is an optimal detector shape for spatially extended gradient sensing. This question has relevance for both 2-D and 3-D multicellular geometries involved in gradient sensing. Is the optimal detector shape more "hairlike", to maximize its extent in the gradient direction, or more "globular", to exploit potential benefits of extending along the transverse direction?

To address this question, we perform a controlled optimization for both 2-D and 3-D multicellular geometries. For a fixed number of cells $N = 50$, we confine cells to an elliptical (2-D) or ellipsoidal (3-D) envelope, and compute the precision of gradient sensing as a function of the ellipse axis parameters (LEGI), as well as the ratio of averaging length scales n_x/n_y (REGI), exhaustively exploring substantial ranges of both. In addition to the extra shape parameter, there is one more important difference between the 2-D and 3-D cases: in the 2-D case, we assume that every cell detects signal molecules, since we imagine that these molecules diffuse in the 3-D bulk, while the cells form a sensory sheet exposed to the bulk. In contrast, in the 3-D case, we assume that only the surface cells detect signal molecules, whereas cells that are

blocked on all six sides by neighboring cells are "shielded" and thus do not detect signal molecules (although all cells still communicate via molecule exchange). The optimal detector shapes determined by such exhaustive search for the REGI model are shown in Fig. 2.3A, for 2-D (top) and 3-D (bottom).

To explain why these optimal shapes emerge, we present the precision of gradient sensing as a function of the control parameters. First we investigate the behavior of the LEGI model in 2-D (Fig. 2.3B). The control parameter is N_g, the (projected) number of cells in the gradient direction, which is set uniquely in 2-D by the ratio of the ellipse axis parameters. Small $N_g \to 1$ corresponds to a chain of cells transverse to the gradient, while large $N_g \to N$ corresponds to a chain of cells parallel to the gradient. The small "stair steps" in the curves are due to the numerical task of fitting the discrete multicellular square lattice within the continuous elliptical envelope. We see that the precision vanishes at $N_g = 1$, as expected, since in our model a single cell cannot perform gradient detection. The precision is near maximal at $N_g = N$. This trend is analogous to that seen for LEGI in Fig. 2.1D, where here $N/N_g \sim M$ is the analog of the number of transverse rows. However, unlike in Fig. 2.1D, we see in Fig. 2.3B that there is a weak optimum at an intermediate value of N_g. This is due to a difference between the protocols of adding rows of cells (Fig. 2.1D) and reshaping a fixed number of cells (Fig. 2.3B). Adding rows does not change $\bar{\Delta}_N$. In contrast, as seen in the inset of Fig. 2.3B, reshaping changes $\bar{\Delta}_N$. The reason is that elliptical configurations (like Fig. 2.3A, top) are not translationally symmetric in the transverse direction. In particular, a large density of cells in the middle of the configuration is a sink for molecules of Y. This decreases the mean number of Y in the rightmost cell, $\bar{y}_N$, which weakly increases the signal $\bar{\Delta}_N = \bar{x}_N - \bar{y}_N$ at intermediate values of N_g (Fig. 2.3B inset), and therefore increases the precision (Fig. 2.3B). Finally, we see that the precision increases with the gain G, as expected, and that the increase saturates with G, since then the variance of X and Y is dominated entirely by extrinsic, and not intrinsic, noise (see Methods).

Next we investigate the behavior of REGI in 2-D (Fig. 2.3C). Once again the control parameter is N_g. Additionally, at every N_g we optimize the local species' averaging length scale n_x (generally we find an optimal value between $\sim 0.1 n_y$ and

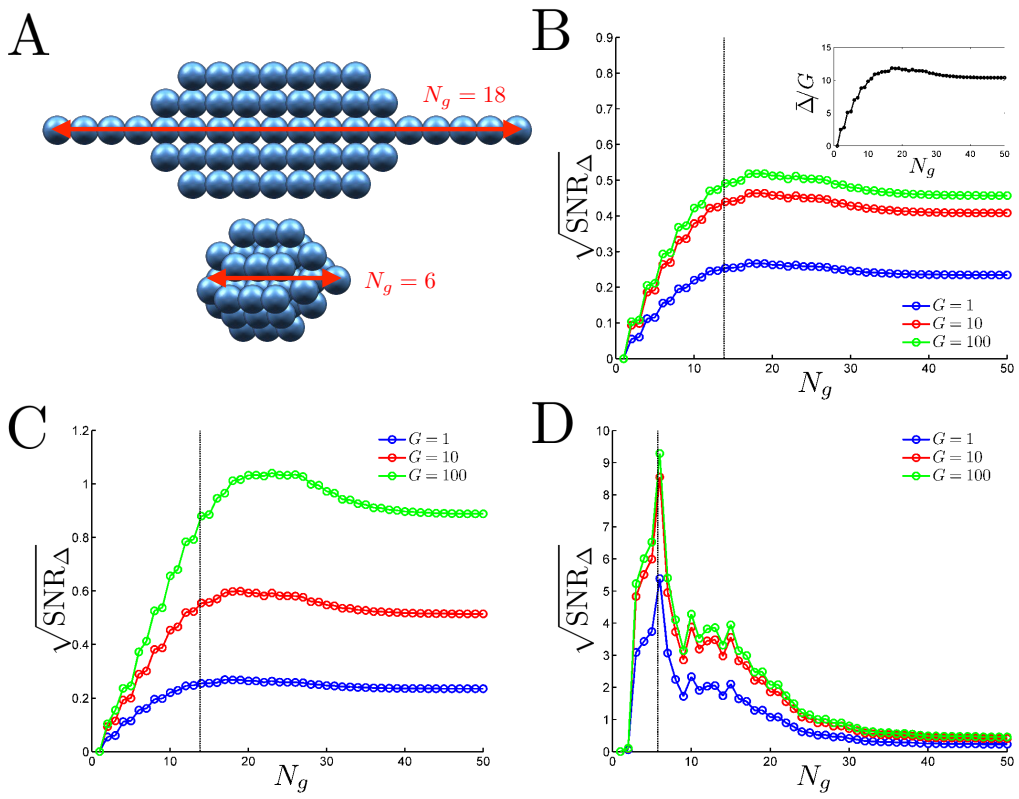

Figure 2.3: Optimal gradient sensing by 2-D and 3-D detectors. (A) Optimal elliptical (2-D, top) or ellipsoidal (3-D, bottom) configurations of $N = 50$ cells for the REGI model The number of cells in the gradient direction for each shape is $N_g = 18$ (top) and $N_g = 6$ (bottom). Cells are depicted as spheres, even though in the 3-D configuration (bottom) only the outermost cells sense the signal; the rest are shielded. Gradient sensing precision is optimized at the rightmost cell, and the signal profile increases linearly to the right. We see that the optimal shapes are "globular", not "hairlike", especially in 3-D. (B) Precision vs. N_g (the projected number of cells in the gradient direction) for the LEGI model in 2-D, for various gains G. Inset: mean readout $\bar{\Delta}$ normalized by G (all three curves overlap and are colored black). (C) As in B, but for REGI. The additional REGI parameter n_x is optimized over at each N_g value, and the optimal precision is shown. At the observed optima in C, these values are $n_x^*/n_y = 0.09$ ($G = 1$), 0.30 ($G = 10$), and 0.53 ($G = 100$). (D) As in B but for 3-D. Internal cells are shielded and do not sense, but do communicate. Ellipsoid axes transverse to gradient are equal. Optimal $n_x^* = 0$ for all N_g. Curve jaggedness arises due to numerical effects of fitting a cubic lattice of cells in a smooth ellipsoidal envelope. Black vertical dashed lines correspond to a perfect circle (B, C) or sphere (D). Parameters are as in Fig. 2.1.

$\sim 0.5 n_y$, see Fig. 2.3). We see in Fig. 2.3C that the trend of precision versus N_g is similar to that of the LEGI model (Fig. 2.3C), but with two key differences. First, the precision is higher for REGI than for LEGI. This is due to regional averaging reducing the variance of the local species, as was known previously for the 1-D model [112]. Second, the optimum in the precision as a function of N_g is more pronounced for REGI than for LEGI. This is because the region surrounding the optimum corresponds to near-circular ellipses, where considerable transverse averaging occurs. As shown in the previous section, transverse averaging increases precision in the REGI model. Overall, the optimal structure (Fig. 2.3A, top) is closer to a "globular" circle than to "hairlike" chain (compare locations of the optima to the dashed vertical line in Fig. 2.3C, which corresponds to a perfect circle). Therefore, we see that optimal gradient sensing by a 2-D structure benefits from an elliptical shape in which transverse averaging occurs.

Finally, we investigate the behavior of REGI in 3-D (Fig. 2.3D). Here there are two control parameters: the number of cells in the gradient direction N_g, and the asymmetry of the ellipsoid in the two directions transverse to the gradient. Generally we find that the optimal shape at a fixed N_g displays symmetry in the two transverse directions, and therefore we impose this symmetry explicitly and focus on the control parameter N_g. As before, at every N_g we optimize the local species' averaging length scale n_x. Importantly, in the 3-D geometry, we find that the optimal value at every N_g is $n_x^* = 0$, corresponding to no averaging of the local species (an effective LEGI model). This is due to the shielding of internal cells: since internal cells do not detect signal molecules, averaging of the local species would dramatically reduce the mean local readout, making it far less than the actual local signal value at the edge cell. This would severely reduce the mean $\bar{\Delta}_N$, and thus the precision. The dependence of precision on N_g is shown in Fig. 2.3D. The additional jaggedness is again due to the incommensurate nature of the cubic cell lattice with the smooth ellipsoidal envelope, here amplified due to the additional dimension. We see in Fig. 2.3D that there is again an optimum. In fact, it is much more pronounced than in 2-D: the overall value of the precision is ten-fold higher than in 2-D. This is again due to the shielding of internal cells: the global species Y is averaged among internal cells that do not produce it, which sharply decreases $\bar{y}_N$, and thereby increases $\bar{\Delta}_N$ and thus

the precision.[2] Overall, the optimal structure is very "globular" (Fig. 2.3A, bottom). Indeed, it is almost a sphere (compare the optima to the dashed vertical line in Fig. 2.3D). We conclude that, due to the combined effects of spatial averaging and shielding, the optimal 3-D detector of linear gradients extends significantly in all three spatial dimensions. The effects of shielding are expanded upon in the Discussion.

2.4 Discussion

We have investigated theoretically and computationally the ways in which the precision of spatially extended, multi-component gradient sensing is affected by detector geometry. Using a minimal model of adaptive gradient sensing (LEGI), we have found that, unlike for concentration sensing, the precision of gradient sensing decreases with the size of the detector in a direction transverse to the gradient. This is due to the competing effects of noise reduction and a reduction of the covariance between concentrations subtracted to estimate the gradient. We have demonstrated that a simple modification of LEGI (REGI) restores the covariance and recovers the benefit of transverse averaging for gradient sensing. The result is that the optimal detectors in 2-D and 3-D are more globular than hairlike.

Our study elucidates the important roles of spatial averaging in gradient sensing, which are several-fold. First, there is spatial averaging along the gradient. In both LEGI and REGI, the global species Y is averaged along the gradient. For a linear signal profile, this averaging both increases the signal $\bar{\Delta}^2$, and decreases the noise $(\delta\Delta)^2$. Therefore, it is optimal for Y to be averaged along the gradient to as large an extent as possible. Second, in the REGI model, the local species X is also averaged along the gradient. This decreases the signal but also decreases the noise [112]. Therefore, there is often an optimal ratio n_x/n_y of the spatial extents of the averaging. Third, there is spatial averaging transverse to the gradient. In the LEGI model, only Y is

[2]Note that this particular effect of shielding will result in the value of $\bar{\Delta}_N$ being positive in every edge cell, instead of only the edge cells at the high end of the gradient. The sensory outcomes are still biased, but are less adaptive, similar to "tug-of-war" chemotaxis mechanisms that have been proposed [35].

averaged transverse to the gradient. In a translationally symmetric geometry, this does not change the signal, but it changes the noise by both decreasing the variance of Y and decreasing the covariance between X and Y. These have opposite effects on the precision. For LEGI, the latter dominates, decreasing the precision. Therefore, transverse averaging is detrimental for gradient sensing. However, in the REGI model, X is also averaged transverse to the gradient. Once again, in a translationally symmetric geometry, this does not change the signal with respect to REGI in 1-D, but it decreases the noise, both by further reducing the variance in X and by restoring a larger covariance between X and Y. Therefore, transverse averaging is beneficial for REGI-type gradient sensing. These roles of spatial averaging are modified in geometries without translational symmetry as we discussed above. However, the net result remains the same: the optimal 2-D and 3-D REGI-type gradient detectors are globular, benefitting from extensive spatial averaging in the transverse directions.

Our study also reveals the effects of shielding of signal from the inner cells in a 3-D geometry. Shielding amplifies the effect of spatial averaging, since the measurements performed by edge cells, which detect signal, are averaged with those of their interior neighbors, which do not detect signal. This amplification increases the signal in a particular edge cell, but makes the system less adaptive, since every edge cell has an above-average readout. With shielding, a more appropriate measure of the sensory outcome might therefore be the difference in readouts between cells up and down the gradient, e. g. $\bar{\Delta}_N - \bar{\Delta}_1$. This measure is likely to depend nontrivially on internal and geometric parameters such as n_x and M, and will likely result in a nontrivial optimal local averaging length scale, $n_x^* \neq 0$. Another possibility is that gain G should be different in Eqs. 2.4 and 2.5, compensating for the two messenger molecules averaging over different numbers of neighbors that do not detect the ligand. We leave both of these interesting explorations for future investigations.

In this work, we have emphasized the distinction between (i) concentration sensing within a graded concentration profile and (ii) gradient sensing. For example, in Drosophila development, individual nuclei in the embryo measure (and are thought to threshold) the local concentration, even though the morphogen gradient is graded [55, 71, 94, 152]. This is an example of concentration sensing. In contrast, gradient

sensing, as explored here, is the task of obtaining an internal readout of the *difference* in local signal concentrations at two or more different points in space. In other words, unlike concentration sensing, gradient sensing determines the direction in which the concentration changes, and it allows subsequent directional polarization of the sensor. This definition of gradient sensing, by construction, is adaptive: the readout does not depend on the background concentration. Systems that respond adaptively and directionally to chemical gradients, such as amoeba [158] and epithelial cell groups [52], are performing gradient sensing. Because concentration sensing and gradient sensing are distinct, it may not be so surprising that transverse averaging has very different effects on them: the precision of concentration sensing increases with the transverse size, whereas the precision of LEGI gradient sensing decreases with the transverse size (Fig. 2.1).

How do our results compare to experimental systems? A well-studied example of a natural gradient-sensing system is the growth factor-directed extension of mammary epithelial ducts [56, 81]. Gradient sensing in this system has been shown to be multicellular and adaptive [52]. *In vivo*, the extension is led by an "end bud" of cells at the duct tip. These tips can form either long hairlike structures or coalesce into nearly spherical globules, as was observed in organotypic studies with different chemical and genetic perturbations [52]. Long hairs could act as "feelers" for the duct, sampling a long swath of the environment in the gradient direction. However, our analysis predicts that such hairlike morphologies are suboptimal, and the globular bud shape, as in Fig. 2.3A, would produce a better precision. In agreement with the prediction, the end buds in wildtype mice are nearly spherical, and the globule is often wider than the duct itself [81]. Similarly, neither chemotaxing amoeba [158] and neutrophils [123], nor growing neurons [139] form very thin hairlike protrusions to facilitate sensing. Instead they keep the aspect ratio of the gradient sensing part of the protrusions closer to one, again supporting our findings. Further, in Drosophila border cell migration, another example of directional collective cell behavior, groups of cells travel as a sphere in a confined space, where it would have been easier to travel as a chain [29]. All of these examples provide indirect evidence that transverse averaging is used in multiple biological contexts. While direct tests of effects of

transverse averaging have not been done, they are certainly possible. Indeed, as mentioned above, different perturbations to organotypic epithelial cultures result in them assuming different geometric shapes [52]. Thus it should be possible to measure the accuracy of sensing (and the subsequent organoid polarization) as a function of the shape. Such experiments would allow direct testing of our main prediction that transverse averaging leads to more accurate directional sensing outcomes, especially in REGI-type models.

Chapter 3

Isolation by Distance in Populations with Power-law Dispersal

The material presented in this chapter can be found in Smith and Weissman [150] and is under review at Genetics.

3.1 Introduction

Direct measurement of dispersal in natural populations is often difficult or impossible due to practical difficulties in tracking large numbers of individuals over long periods of time. It is often more feasible to instead infer dispersal from spatial patterns of genetic diversity [20, 30, 38]. Populations with limited dispersal should exhibit "isolation by distance": the more distant individuals are from each other in space, the less related they tend to be [137, 146, 169]. While in general the spatial pattern of genetic diversity depends on selection [4, 18], for populations evolving neutrally the strength of isolation by distance is simply determined by the balance between dispersal and mutation, and thus if the mutation rate is known, the dispersal rate can be inferred directly [100, 147, 148].

For populations spread over a fairly continuous range, rather than being clumped

into a small number of discrete subpopulations, dispersal is often assumed to be thin-tailed, with displacement approximately following a normal distribution [15, 134]. If dispersal is unbiased and homogeneous, it is then characterized by a single parameter, the dispersal rate D (the diffusion constant). Pairwise genetic similarity is predicted to decay exponentially with distance, with a decay rate of $\sqrt{\mu/D}$, where μ is the mutation rate [15, 89, 131, 140, 141, 149].

However, many populations exhibit a heavy-tailed dispersal distribution that decays very slowly with distance [2, 7, 11, 34, 45, 46, 48, 63, 159, 167]. This heavy tail can allow for dispersal events that are 10 or more times the "typical" jump size given by the standard deviation of the distribution, and for extremely heavy tails the standard deviation itself becomes infinite. The presence of jumps that span a huge range of sizes leads to a qualitatively different form of motion than one would expect from the "diffusive" picture that is typical of thin-tailed dispersal distributions [108]. Even for populations where dispersal is primarily thin-tailed, very rare long jumps that are unlikely to be observed directly [90] can have large effects on the population's evolution [28, 33, 57, 75, 82, 102, 126, 127, 168], so we would also like to be able to infer their pattern and frequency. For many other populations, particularly non-animal ones, very little is known about dispersal, and we would like to know how to use genetic data to determine if it is thin- or heavy-tailed [38, 115].

In this work, we further explore the effects of heavy-tailed (power-law) dispersal in neutrally evolving, demographically stable populations with constant density. This problem was previously studied by Nagylaki [114] in one dimension for moderately heavy-tailed dispersal with finite mean distance and by Chave and Leigh Jr [40] in two dimensions for Cauchy-distributed dispersal. Recently, Janakiraman [85] studied an analogous problem in chemical physics in the same regime as Nagylaki [114] and found complementary results. We unify and extend this work to cover arbitrary power-law dispersal tails in both one and two spatial dimensions and find expressions for how the pattern of isolation by distance reflects the underlying dispersal process at both large and small distances. We also find how the distribution of time to the most recent common ancestor of a pair of individuals depends on the distance between them. Barton et al. [19] and more recently Forien [60] have considered a similar model for

populations evolving according to a spatial Λ-Fleming-Viot process with power-law dispersal. This model differs from these previously considered models in that lineages undergoing power-law dispersal can be treated as independent. Our analysis may apply to an alternative spatial Λ-Fleming-Viot model in which reproduction events encompass the entire range, and the probability of inclusion in an event decays with the distance from its origin according to a power-law.

3.2 Model

We consider two individuals sampled in the present a distance x apart, and trace their lineages backward through time until they coalesce. We assume that individuals disperse through an infinite range according to a distribution that falls off as a power law at long dispersal distances. While our model assumes infinite range size, we will show via simulation that all of our predictions hold for finite ranges when the range size is sufficiently large. For concreteness, we will mostly consider lineages that follow Lévy flights, a flexible, mathematically tractable way to model dispersal with power-law tails [86, 107, 108]. Lévy flights can be seen as a generalization of diffusive motion. Apart from the special case of classic diffusion (for which the power-law tail disappears), these trajectories include rare long-range jumps and are governed by power-law kernels with infinite variance. We also consider finite variance power-law dispersal kernels that cannot be described by a Lévy flight.

For a lineage following a Lévy flight in two dimensions, the probability of dispersing to a particular point a distance y away follows an isotropic stable distribution [174]:

$$K_1(y|t) = \frac{1}{2\pi}\int_0^\infty dk\, kJ_0(ky)\exp\left(-D_\alpha t k^\alpha\right), \tag{3.1}$$

where J_0 is the zeroth Bessel function and t is the number of generations in the past. The parameter D_α is a generalized diffusion constant. It sets the scale of dispersal: at time t, a typical lineage will have a displacement proportional to $(D_\alpha t)^{1/\alpha}$. As shown in Fig. 3.1, the "stability parameter" α controls the size and number of long-range jumps: the maximum of t dispersal events is proportional to $t^{1/\alpha}$, and for $\alpha < 2$,

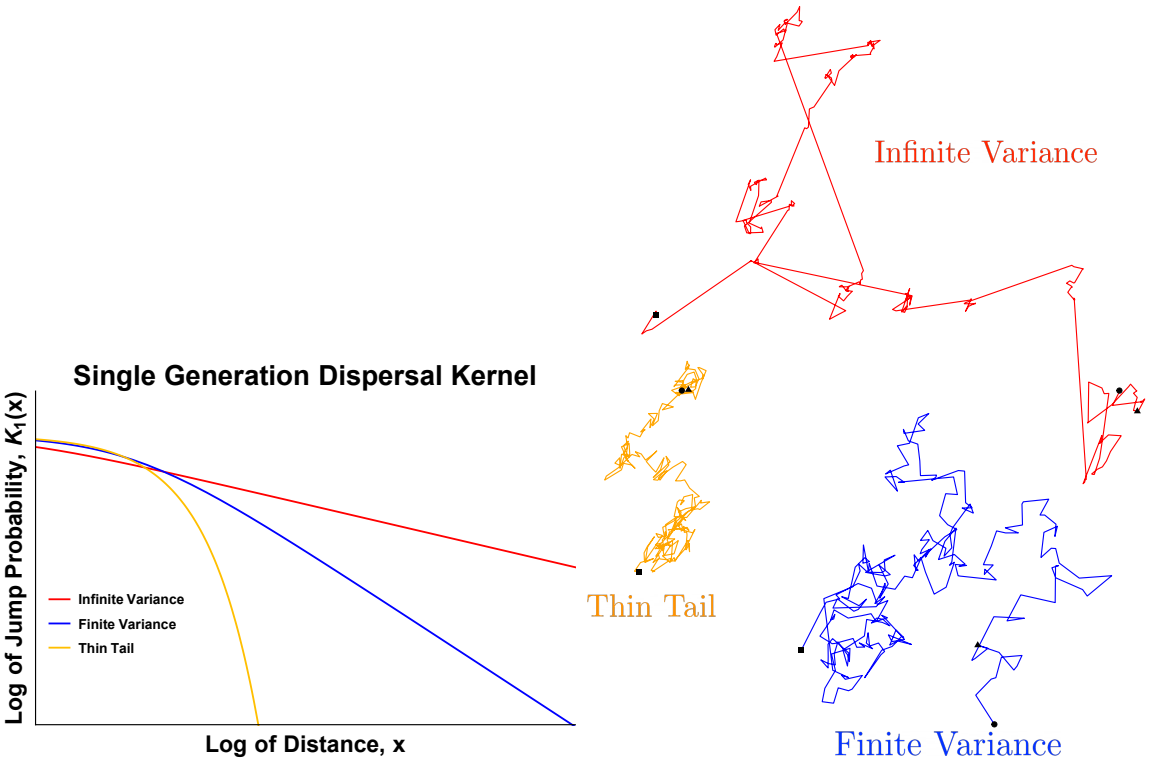

Figure 3.1: **The tail of the dispersal distribution controls the size and number of long-range jumps.** We demonstrate the effect of varying the dispersal distribution tail by generating random walks using both thin-tailed and heavy-tailed (power-law) kernels. **Left: Single generation dispersal distributions used to generate random walks.** The orange curve represents the jump probability for a thin-tailed dispersal distribution, while the blue and red curves show jump probabilities for finite-variance and infinite variance power-law dispersal kernels. **Right: Two-dimensional random walks with jumps drawn from thin-tailed (orange) and power-law (blue and red) distributions.** When the dispersal distribution is thin-tailed, the motion reduces to normal diffusion without any long-range jumps. When the dispersal distribution has a power-law tail, trajectories can jump large distances in a single time step, dramatically changing the rate at which lineages move through space. If the power-law tail is very broad, trajectories will have divergent mean squared displacement, and large jumps become noticeably more prevalent than for steep power laws with finite variance. Circles mark the beginnings of the trajectories, triangles mark the positions after 10 jumps, and squares mark the ends. On shorter time scales, the diffusive trajectory tends to have the largest displacement, while on longer time scales the infinite variance trajectory in red tends to have the largest displacement.

Symbol	Definition
x	Distance between samples
ρ	Population density
α	Stability parameter of dispersal tail
D_α	Generalized dispersal constant
μ	Mutation rate
ψ	Probability of identity
$\overline{x} = (D_\alpha/\mu)^{1/\alpha}$	Characteristic length scale of identity

the probability that a lineage has moved to a particular point at an abnormally long distance $y \gg (D_\alpha t)^{1/\alpha}$ is proportional to $y^{-2-\alpha}$, a power-law tail. In the limiting case $\alpha = 2$, dispersal reduces to ordinary thin-tailed diffusion and (3.1) is just a normal distribution.

When the two lineages encounter each other, they coalesce at rate proportional to $1/\rho$, where ρ is the density of the population. In two dimensions, two lineages of infinitesimal size will never be at exactly the same position [111]. So really there must be some small distance δ within which lineages coalesce at a rate that is approximately $1/(\delta^2\rho)$. At these small scales, even the model of independent diffusion of lineages breaks down [16]. But we will see below that this coalescence length scale does not affect isolation by distance on larger scales $x \gg \delta$.

We are interested in the probability ψ of identity by descent of our sample pair as a function of the distance between them, x, which we will also refer to as the homozygosity or relatedness. If the time to their most recent common ancestor is T and the mutation rate is μ, then ψ is given by:

$$\psi(x) = E\left[e^{-2\mu T}|x\right]. \tag{3.2}$$

Although usually it is ψ rather than the coalescence time T itself that is directly observable, T is important for, e.g., determining whether it is reasonable to assume stable demography, so we will also find expressions for its distribution $p(t|x)$.

3.3 Results

In this section, we will describe our main results and provide brief sketches of the logic behind key features. Roughly speaking, the basic intuition is that the sampled pair will be identical if their lineages coalesce within the approximately the past $1/\mu$ generations. In this time, they will disperse a typical distance of order $\overline{x} \equiv (D_\alpha/\mu)^{1/\alpha}$, so this is the key length scale over which identity decays: pairs separated by $x \ll \overline{x}$ should be relatively closely related, while identity between pairs separated by $x \gg \overline{x}$ should be rare.

For the classic case of diffusive motion ($\alpha = 2$), this length scale is $\overline{x} = \sqrt{D/\mu}$, and the probability of identity in two dimensions falls off logarithmically for $x \ll \overline{x}$ and exponentially for $x \gg \overline{x}$ [15]:

$$\psi(x) \approx \begin{cases} \frac{\ln(\overline{x}/\delta)}{\ln(\overline{x}/\delta)+4\pi\rho D_2} & \text{for } x \ll \delta \ll \overline{x} \\ \frac{1-\psi(0)}{4\pi\rho D_2}\ln(\overline{x}/x) & \text{for } \delta \ll x \ll \overline{x} \\ \frac{1-\psi(0)}{4\rho D_2}\frac{\exp(-x/\overline{x})}{\sqrt{2\pi x/\overline{x}}} & \text{for } x \gg \overline{x} \gg \delta. \end{cases} \tag{3.3}$$

Here we generalize (3.3) to $\alpha \neq 2$, and find simple approximate expressions for ψ in different parameter regimes, illustrated in Fig. 3.2, including a universal form for all power-law dispersal kernels at long distances. Intuitively, power-law dispersal broadens the distribution of coalescence times for pairs at a given separation x, creating more overlap in the distributions for different x values (Fig. 3.3).

3.3.1 Distant pairs

For distant samples, $x \gg \overline{x}$, we expect substantial isolation by distance. For the pair to coalesce, their lineages must approach within δ of each other. The most likely way for this to happen before a mutation occurs is for one lineage to cover the distance in a long jump. Since such jumps occur at a rate of $D_\alpha x^{-\alpha-2}\delta$, this occurs with approximate probability $D_\alpha x^{-\alpha-2}\delta/\mu$. The lineages must then coalesce within

their neighborhood of about $\delta^2\rho$ individuals before they mutate, which occurs with probability $1/(\mu\rho\delta^2)$. We therefore expect that the probability of identity is $\psi(x) \approx D_\alpha x^{-\alpha-2}\delta^2/\mu/(\mu\rho\delta^2) = D_\alpha x^{-\alpha-2}/(\mu^2\rho)$, i.e., that there is a power-law dependence of identity on distance, with the same exponent as that of dispersal. A more careful calculation (see Methods) yields the precise result for a two-dimensional Lévy flight when $\alpha < 2$:

$$\psi(x) \propto \frac{(1-\psi(0))\, D_\alpha}{\rho\mu^2 x^{2+\alpha}} \qquad \text{for } x \gg \overline{x} \gg \delta. \tag{3.4}$$

While the prefactors in (3.4) depend on our exact choice of dispersal distribution, the form of the expression does not. For any power-law dispersal kernel, we predict a corresponding power-law tail in $\psi(x)$ with matching exponent (see Methods). We confirm (3.4) with simulations and numerical analysis (Fig. 3.4). As shown in Fig. 3.5, even steep power-law kernels with $\alpha > 2$ and finite variance lead to a matching power law in $\psi(x)$.

3.3.2 Nearby pairs: broad power-law kernels with infinite variance

For very heavy-tailed dispersal in which the mean squared jump size diverges, $\alpha < 2$, nearby lineages are likely to either coalesce very quickly or to disperse across the whole range before coalescing [25, 124]. This "now-or-never" dynamic has the interesting effect of making the local probability of identity by descent independent of the mutation rate, since the main competition is between coalescence and heavy-tailed dispersal rather than between coalescence and mutation. Intuitively, the pair of lineages take time of order x^α/D_α to disperse across the distance between them. From that time on, they are roughly evenly spread over a range of size $(D_\alpha t)^{2/\alpha}$, and so coalesce at a rate of about $(D_\alpha t)^{-2/\alpha}/\rho$. Integrating this rate over time starting from x^α/D_α out to $1/\mu$, we find that $\psi(x) \approx 1/(\rho D_\alpha x^{2-\alpha})$, with the upper limit of integration only negligibly decreasing $\psi(x)$. So $\psi(x)$ again follows a power law, although a different one from the long-distance $1/x^{2+\alpha}$. We calculate $\psi(x)$ more

carefully in the Methods to find:

$$\psi(x) \approx \frac{\Gamma(1-\alpha/2)}{2^{1+\alpha}\pi\Gamma(\alpha/2)} \frac{1-\psi(0)}{\rho D_\alpha x^{2-\alpha}} \qquad \text{for } \delta \ll x \ll \overline{x}. \tag{3.5}$$

While (3.5) is for Lévy flights, we expect this expression to hold for any broad power-law dispersal distribution, so long as many dispersal events occur before the mutation timescale $1/\mu$. This is because the sum of many draws from an infinite-variance distribution follows a Lévy stable distribution for small displacements, just as the sum of many draws from a finite-variance distribution follows a normal distribution [121]. Note that (3.5) reduces to Eq. (A6) in Chave and Leigh Jr [40] when $\alpha = 1$. We confirm (3.5) with simulations (Fig. 3.4).

The power law in (3.5) makes it diverge at very short distances, where it breaks down. Instead, for individuals within the same deme, $x < \delta$, $\psi(x)$ flattens out. Roughly speaking, individuals coalesce at rate $1/(\rho\delta^2)$ and disperse outside of coalescence range at rate of about $D_\alpha\delta^{-\alpha}$. When coalescence is faster, probability of identity is high, $\psi(0) \approx 1$, while when dispersal is faster it is low, $\psi(0) \approx 1/(\rho\delta^2)/(D_\alpha\delta^{-\alpha}) = 1/(\rho D_\alpha\delta^{2-\alpha})$. A more careful calculation gives (see Methods):

$$\psi(x) \approx 1 \Big/ \left[1 + \frac{2^{2+\alpha/2}\pi}{\Gamma(1-\alpha/2)} \rho D_\alpha \delta^{2-\alpha}\right] \qquad \text{for } x \ll \delta \ll \overline{x}\,, \tag{3.6}$$

although these numerical factors depend on the details of the coalescence kernel.

3.3.3 Nearby pairs: steep power-law kernels with finite variance

In addition to considering Lévy flight dispersal kernels with $0 < \alpha \leq 2$, we consider F-distribution kernels (see Methods) with steeper power-law tails ($\alpha > 2$). These have finite variance and approach a diffusion after infinitely many steps, but at any finite time will be different, particularly in the tail. As shown in Fig. 3.5, for nearby pairs, $x \ll \overline{x}$, relatedness decays logarithmically according to (3.3), as in the case of purely diffusive motion.

3.4 Discussion

Limited dispersal produces a correlation between spatial and genetic distance [100, 146–148, 169]. While most previous models have only considered diffusive dispersal, dispersal can be heavy-tailed in many natural populations. Nagylaki [114] was the first to generalize classic diffusive models of isolation by distance by allowing dispersal distance to have a power-law tail (with $1 < \alpha < 2$). This groundbreaking work has largely been neglected; it was last cited by Chave and Leigh Jr [40], who extended the results to two dimensions for the special case of Cauchy flights ($\alpha = 1$) in a paper modeling ecological diversity. Recent studies suggest that heavy-tailed dispersal may in fact be common, and we hope that this paper will reintroduce these classic results to population genetics now that the field may have sufficient data to apply them [2, 7, 11, 34, 45, 46, 48, 63, 159, 167]. We also extend this previous work by considering Lévy flights for all $\alpha \leq 2$ in both one and two dimensions, as well as steeper power-law kernels ($\alpha > 2$) with finite variance. We find that, for all α, power-law dispersal leads to much more heavy-tailed relatedness than diffusive dispersal, with relatedness having the same power-law tail in distance as the dispersal kernel. This is true even for steep kernels with finite variance. In this case, even though a diffusive approximation can fit the pattern of isolation by distance between nearby individuals, it will greatly underestimate the degree of relatedness between distant individuals.

Standard methods for inferring dispersal from pairwise measures of relatedness or autocorrelations in allele frequency typically assume either thin-tailed, diffusive motion [32, 134, 135, 140, 141] (perhaps with recent long-range admixture [31]) or a small number of discrete demes [3, 99, 128, 142, 146, 165]. Methods using cline theory to infer dispersal from the width of hybrid zones make similar assumptions about the motion of lineages being diffusive [12, 13, 38, 65, 133, 154], and methods for non-stable demographies based on historical biogeography or coalescent theory tend to also assume a small number of discrete demes [77, 132, 144]. Other genetic methods such as parentage analysis are better equipped to infer heavy-tailed dispersal on continuous ranges, but these techniques require exhaustive sampling of the population to ensure that the parents of each individual can be located [1, 9, 88, 162]. More recent

methods for pollen dispersal have been developed that allow for the inference of heavy-tailed dispersal without the need for exhaustive sampling, but knowledge of maternal genotypes for all sampled individuals is still required [8, 136]. For plant species where this data is available, our results could serve as the basis for complementary inference methods. While the pollen dispersal methods are focused on inferring the dispersal kernel over a single generation, isolation by distance reflects the history of dispersal over many generations, so a comparison of the results could reveal changes in dispersal over time. For species where no such pedigree data is available, continuous-space inference methods based on the model developed here could allow for the presence (or absence) of heavy-tailed dispersal to be inferred for the first time.

One key open question is to what extent it is possible to detect the genetic traces of rare heavy-tailed dispersal in natural populations, and if so how well the form of heavy-tailed dispersal (e.g., the tail exponent α) can be determined. Austerlitz et al. [8] were able to detect heavy-tailed pollen dispersal in *Sorbus torminalis* and *Dinizia excelsa* tree populations using parentage analysis and the seed-specific TwoGener method, with α estimates for both species being extremely close to $\alpha = 1$. The success of TwoGener suggests that similar patterns should be detectable from the genomes of the trees themselves. Aguillon et al. [2] found clear patterns of isolation by distance across scales between 500 meters and 10 kilometers in Florida Scrub-Jays, and directly measured heavy-tailed dispersal; a good first test of an inference method would be to apply it to such a dataset to see if it can recover the known dispersal pattern.

Along with predicting characteristic scaling of identity by descent with distance, our results predict characteristic scalings with the mutation rate μ, and also a scaling of the typical length scale of identity $\overline{x}$ with μ. While mutation rate cannot be scanned directly as distance can, μ here should be understood as referring to the mutation rate in a block of non-recombining genome, and so a wide range of μ values can be scanned by considering identity by descent in blocks of varying size [164]. This will be valid as long as recombination is rare relative to mutation.This suggests that it should be possible to measure identity by descent statistics corresponding to μ values ranging over five orders of magnitude in a single sample [76].

Our analysis has focused on pairs of lineages in an infinite habitat, but we have shown via simulation that these results can hold for habitats of finite length L provided that this length is sufficiently large. To better understand the effect of finite range size, we can consider a pair of individuals sampled from random locations within a habitat of length L; the mean coalescence time between them would then be the "effective population size". The pair will typically be sampled a distance of about L from each other, and so it will typically take a time of order L^α/D_α for their lineages to overlap in space. At this point the ancestry is effectively well-mixed, and coalescence takes time proportional to the total population size $L^d\rho$, where $d = 1$ or 2 is the dimension of the habitat. For $L^\alpha/D_\alpha \ll L^d\rho$, the mixing time has little effect, while for L^α/D_α of order $L^d\rho$ or greater, the population becomes spatial structured, i.e., the non-negligible mixing time associated with crossing the range leads to a higher mean coalescence time than one would expect in the panmictic limit. For thin-tailed dispersal, $\alpha = 2$, structure will be strong in a one-dimensional habitat of length greater than about $D\rho$ [104], while in two dimensions its strength depends only on the local neighborhood size $D\rho$ [105]. The amount of population structure thus either increases with the spatial extent of the population (at fixed density) or is insensitive to it. With heavy-tailed dispersal, however, we see a new qualitative pattern. For $\alpha < d$, i.e., for broad power laws, the effect of structure on mean time to coalescence counterintuitively becomes *weaker* as the range size L grows, because L^α/D_α grows more slowly than $L^d\rho$. We thus expect the panmictic result to be an accurate prediction of the mean coalescence time for populations with very heavy-tailed dispersal ($\alpha < d$) and large range size.

While the size of the range has a strong effect on the average time to coalescence, it does not necessarily have a strong effect on $\psi(x)$ and related measures of genetic diversity. We can understand this phenomenon by considering the separation of coalescence timescales: lineages tend to either coalesce quickly or wander away from each other and avoid coalescing until a time set by the range size [17, 166]. If the timescale of mutation, $1/\mu$, is much smaller than the time set by the range size, $\psi(x)$ is not affected by coalescence events that occur on these long timescales and behaves as if the range and effective population size are infinite. For populations

with low mutation rates such that $1/\mu$ is larger than or comparable to this long timescale, finite range size may have a significant effect on $\psi(x)$. Assuming the mixing time is negligible, $L^\alpha/D_\alpha \ll L^d\rho$, the homozygosity of pairs that "wander off" can be approximated by the panmictic result $\psi_{\text{panmictic}}$. We can thus estimate the effect of finite range size by adding a term to ψ that is proportional to this result: $\psi_{\text{finite}}(x) \approx \psi(x)_{\text{infinite}} + (1 - \psi(x)_{\text{infinite}})\,\psi_{\text{panmictic}}$, where the infinite-range $\psi(x)_{\text{infinite}}$ can be interpreted as the probability of "coalescing quickly".

Our use of stable distributions for the dispersal kernel has been partly motivated by the fact that any isotropic single-generation dispersal kernel will converge to a stable one if it is repeated over many independent generations. But as we have noted, this is only true asymptotically, and in any real population there will be correlations across generations, spatial inhomogeneities, shifts in dispersal over time, limits due to finite range size, and many other effects that cannot be captured by a stable distribution. It is therefore better to see it as a simple reference model, one step closer to reality than the purely diffusive one, that can serve as a background against which to measure all these other effects.

What other processes could produce similar patterns to heavy-tailed dispersal? One obvious one is if individuals are performing something more like a "Lévy walk" than a Lèvy flight, in which dispersal in any one generation is thin-tailed but can be correlated across many generations [172]. Such an effect can be produced at the level of alleles by hitchhiking on beneficial substitutions [4]. But this should be readily distinguishable from neutral heavy-tailed dispersal by considering the distribution of relatedness across multiple individuals and loci – hitchhiking will produce heavy-tailed relatedness at the same few loci across all individuals, whereas neutral effects will be more evenly distributed. It is an open question whether other neutral processes, in particular demographic fluctuations, might produce similar patterns.

3.5 Methods

3.5.1 Simulation methods in two dimensions

All simulation code and displayed data are available at https://github.com/weissmanlab/Long_Range_Dispersal. We simulate our model in two stages. First, for each value of present-day separation x, dispersal constant D_α, and tail parameter α, we simulate dispersal of the lineages, ignoring coalescence and mutation. Then, for each value of ρ and μ, we calculate the expected homozygosity and coalescence time distribution for each simulated trajectory. We then average over many independent trajectories. A major advantage of this two-part method is that the second part of the method, in which conditional expectations are calculated for previously generated paths, is entirely deterministic. This reduces computational costs and noise in the estimations.

We simulate lineage motion using a discrete time random walk,

$$X_{t+1} = X_t + \Delta X_t, \tag{3.7}$$

where X_t represents the position of a lineage at a given time (ignoring coalescence., i.e., assuming $\rho \to \infty$), and the displacement, ΔX_t, is a vector of integer valued random variables drawn from the dispersal distribution at each integer time t. We use the GNU Scientific Library's efficient pseudorandom generators for both stable distributions and the F-distribution [66]; because these are available only for the one-dimensional distributions, we draw radial distances using the one-dimensional distributions and then select a random direction in which to move. For dispersal, we primarily use one-dimensional Lévy alpha-stable distributions to randomly draw distances.In continuous space, ΔX_t would have distribution:

$$K_1(y) = \frac{1}{2\pi^2|y|}\int_{-\infty}^{\infty} dk \exp\left(-iky - D_\alpha|k|^\alpha\right), \tag{3.8}$$

where the probability of a displacement depends only on its magnitude, y. To enforce our condition of discrete dispersal distances, we then round ΔX_t to the nearest pair of integers, i.e., the closest point in $\mathbb{Z}^2$. To simulate steeper tails with $\alpha > 2$, we use an F-distribution for radial distances, defined below in (3.15).

For each pair of simulated trajectories $\{x_t\}$, we then compute the path-specific distribution of coalescence times $p(\{x_{t'\leq t}\})$, i.e., the probability of coalescing at and not before time t, and the path-specific mean homozygosity $\psi(\{x_{t'\leq\infty}\})$, i.e., the probability that lineages following these exact trajectories have not mutated before coalescence:

$$p(\{x_{t'\leq t}\}) = \left(1 - e^{-\frac{1}{\rho}\delta_{x_1x_2}}\right) \exp\left[-\frac{1}{\rho}\sum_{t'=1}^{t-1}\delta_{x_1x_2}\right] \qquad \text{for } t > 1, \tag{3.9}$$

$$\psi(\{x_{t'\leq\infty}\}) = \sum_{t=1}^{\infty} p(\{x_{t'\leq t}\})e^{-2\mu t}. \tag{3.10}$$

We start (3.9) and (3.10) at $t = 1$ because we assume that the individuals are sampled immediately after dispersal, so no coalescence takes place at $t = 0$. $\delta_{x_1x_2}$ in (3.9) is the Kronecker delta function:

$$\delta_{x_1x_2} \equiv \begin{cases} 1 & \text{if } x_1 = x_2 \\ 0 & \text{otherwise.} \end{cases}$$

For every time-step the lineages spend in this region, there is a probability of coalescence $1 - e^{-\frac{1}{\rho}}$.

We then average (3.9) and (3.10) across all simulated trajectories with present-day separation x to obtain $p(t|x)$ and $\psi(x)$. All error bars in plots show 68% confidence intervals, as determined by the percentile bootstrap with 1000 bootstrap samples [47]. At large distances, the distribution of the probability of identity across sample trajectories is highly skewed, with most trajectories having very low probabilities of identity, but a few having the lineages rapidly jump close to each other and having a high probability of identity. This means that we cannot quantify the uncertainty in our estimates using, for example, the standard error of the mean, but it also means that we must simulate many independent trajectories to get good enough coverage for the bootstrap to be accurate [43].

For the simulations of mean homozygosity ψ shown in Fig. 3.4 and Fig. 3.5, we simulate 10,000,000 independent runs of 1000 generations each for each combination of present-day separation x and tail parameter α. We set the dispersal constant D_α

indirectly by setting the characteristic spread c of each lineage after one generation ($t = 1$), $c = (D_\alpha)^{1/\alpha}$, to be fixed at $c = 10$ for $\alpha \leq 2$ and $c = 10\sqrt{2}$ for $\alpha > 2$ (see below for the definition of c for $\alpha > 2$). We also apply periodic boundary conditions, with the range size extending from -5000 to 5000 along both dimensions of the discrete lattice. Our choice of range size is significantly larger than the maximum value used for present-day separation between pairs ($x = 237$). For the largest mutation rate considered, $\mu = 1$, coalescence before mutation is extremely rare, and so we increase the number of independent runs to 100 million (with the number of generations reduced to ten).

3.5.2 Simulation methods in one dimension

Our one-dimensional simulation methods are identical to those used in two dimensions, except that space is now taken to be continuous rather than discrete. We again simulate lineage motion using a discrete time random walk,

$$X_{t+1} = X_t + \Delta X_t, \tag{3.11}$$

where X_t now represents the signed distance between two lineages at a given time, and the step size, ΔX_t, is a real valued random variable drawn from the dispersal distribution at each integer time t. For dispersal, we primarily use Lévy alpha-stable distributions, so ΔX_t has distribution:

$$K(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \exp\left(-iky - 2D_\alpha |k|^\alpha\right). \tag{3.12}$$

Note that this differs from $K_1(y)$ by an extra factor of two because ΔX_t is the sum of the two lineages' independent jumps, i.e., K is the convolution of K_1 with itself. To simulate steeper tails with $\alpha > 2$, we again use an F-distribution, defined below in (3.15).

For each simulated trajectory $\{x_t\}$, we then compute the path-specific distribution

of coalescence times $p(\{x_{t'\leq t}\})$ and path-specific mean homozygosity $\psi(\{x_{t'\leq\infty}\})$:

$$p(\{x_{t'\leq t}\}) = \left(1 - e^{-\frac{1}{\rho}R(x_t)}\right) \exp\left[-\frac{1}{\rho}\sum_{t'=1}^{t-1} R(x_{t'})\right], \qquad \text{for } t > 1 \tag{3.13}$$

$$\psi(\{x_{t'\leq\infty}\}) = \sum_{t=1}^{\infty} p(\{x_{t'\leq t}\})e^{-2\mu t}. \tag{3.14}$$

$R(x)$ in (3.13) is a rectangular function representing a uniform rate of coalescence of all lineages within a distance δ:

$$R(x) \equiv \begin{cases} \frac{1}{2\delta} & \text{if } |x| < \delta \\ 0 & \text{otherwise.} \end{cases}$$

For every time-step the lineages spend in this region, there is a probability of coalescence $1 - e^{-\frac{1}{2\rho\delta}}$. We discuss issues with the microscopic interpretation of this model after we introduce our analytical model below.

Unconditioned values $p(t|x)$ and $\psi(x)$ are again obtained by averaging across all simulated trajectories. Error bars in plots show 68% confidence intervals, as determined by the percentile bootstrap with 10,000 bootstrap samples.

We set $\delta = 0.5$ for all simulations in one dimension. For the simulations of mean homozygosity ψ shown in Fig. 3.8 and Fig. 3.5, we simulate 250,000 independent runs of 1000 generations each for each combination of present-day separation x and tail parameter α. We set the dispersal constant D_α indirectly by setting the characteristic spread c of two lineages after one generation ($t = 1$), $c = (2D_\alpha)^{1/\alpha}$, to be fixed at $c = 250$ for $\alpha < 2$, and $c = 179.68$ for $\alpha = 2.05$ (see below for the definition of c for $\alpha > 2$). For the largest present-day separations, $x = e^{10}$ and e^{11}, coalescence within 1000 generations is very rare, so we increase the number of runs to 1.25×10^6. For Fig. 3.6 and Fig. 3.10, we choose D_α such that $c = 0.2$, and simulate 10,000 independent runs of length 1000 generations each.

For the simulations of the cumulative distribution of coalescence times $P(t)$ shown in Fig. 3.9, we set present-day separation $x = 0$ and generate 10,000 independent trajectories of 1.5 million generations each for each combination of c and tail parameter α. We set $c = 3.59$ for $\alpha > 2$, $c = 5$ for $1 < \alpha < 2$, and $c = 1$ for $\alpha \leq 1$.

Dispersal kernel for $\alpha > 2$

To simulate dispersal kernels with tail exponents $\alpha > 2$, we draw ΔX_t from a two-sided Fisher F-distribution with $d_1 = 2$ and $d_2 = 2\alpha$:

$$K(y) = (2\omega)^{-1} \left(1 + \alpha^{-1} \left(|y|/\omega\right)\right)^{-\alpha-1}. \tag{3.15}$$

At long times, the displacement distribution approaches that of a diffusive kernel, with dispersal constant equal to half the mean squared single-generation displacement of one lineage:

$$D = \frac{c^2}{2} = \frac{\alpha^2}{2\left(\alpha-2\right)\left(\alpha-1\right)}\omega^2.$$

In two dimensions, we use the one-sided analog of (3.15) to draw radial distances, and then draw a direction from a uniform distribution. The resulting distribution for a single lineage is:

$$K_1(y) = (2\pi\omega|y|)^{-1} \left(1 + \alpha^{-1} \left(|y|/\omega\right)\right)^{-\alpha-1}, \tag{3.16}$$

with the dispersal constant now defined as:

$$D = c^2 = \frac{\alpha^2}{2\left(\alpha-2\right)\left(\alpha-1\right)}\omega^2.$$

3.5.3 Analytical model in one dimension

Generic dispersal

We want to find a tractable analytical approximation to the model described above. For recurrent motion, the lineages will sometimes be in exactly the same place, and we can model coalescence with a δ distribution, i.e., as taking place at rate $\frac{1}{\rho}\delta(X_t)$. For transient motion, however, they will never coincide [124], and we must allow coalescence to take place at a finite distance. Let the coalescence kernel be some probability density $\mathcal{N}(x)$ symmetric about $x = 0$ and with width $\sim \delta$, with coalescence taking place at rate $\frac{1}{\rho}\mathcal{N}(X_t)$. The δ-distribution is just the limit of $\mathcal{N}$ as δ

goes to 0, so we can treat the two cases together. Forien [60] avoids this issue by using a spatial Λ-Fleming-Viot model, in which dispersal and coalescence are produced by the same heavy-tailed process, but this leads to dispersal distances for an individual's offspring being strongly correlated rather than independent [19]. We instead choose to keep power-law dispersal as a distinct process from short-range coalescence. As mentioned in the Simulation Model section, this creates issues with the microscopic interpretation of the model, which we discuss below in "Breakdown of models at small scales".

The coalescence time distribution, i.e., the probability density of coalescence times for lineages with initial displacement x, is then:

$$p(t|x) = E\left[\frac{1}{\rho}\mathcal{N}(X_t)\exp\left(-\frac{1}{\rho}\int_0^t d\tau\,\mathcal{N}(X_\tau)\right)\middle|\, X_0 = x\right], \tag{3.17}$$

and the probability of identity is its Laplace transform:

$$\psi(x) = \int_0^\infty dt\, p(t|x)e^{-2\mu t}. \tag{3.18}$$

There are several different ways to derive an explicit expression for ψ from (3.18), including balancing mutation, coalescence, and dispersal over an infinitesimal time step [15, 100] or, for Lévy flights, using a fractional diffusion equation [85] (see Appendix). Here we start with a generalization of Barton and Wilson [14]'s expression for $p(t|x)$ that is valid for any two-lineage dispersal kernel K, which is defined as the convolution of K_1 with itself. Assuming that $\mathcal{N}(x) = \delta(x)$:

$$p(t|x) = \frac{1}{\rho}K(x|t) - \int_0^t dt'\, p(t-t'|0)\frac{1}{\rho}K(x|t'). \tag{3.19}$$

To interpret (3.19), notice that the first term is the probability of coalescing at time t neglecting the possibility that the lineages have coalesced more recently. The second term corrects for these more recent coalescences: for every trajectory where the lineages coincide at $t' < t$, we subtract off the probability that the lineages would coalesce at t' and then again exactly at t. Notice that we do not need to correct again for lineages that coincide three times, at $t'' < t' < t$: the factor of p guarantees that each trajectory is weighted appropriately.

We can immediately find a simple expression for ψ for recurrent dispersal by taking the Laplace transform of (3.19):

$$\psi(x) = \frac{1-\psi(0)}{\rho}\widetilde{K}(x, 2\mu), \tag{3.20}$$

where tilde denotes the Laplace transform. Plugging in $x = 0$, we can solve (3.20) for $\psi(0)$ and express $\psi(x)$ purely in terms of the dispersal kernel:

$$\psi(x) = \frac{\widetilde{K}(x, 2\mu)}{\rho + \widetilde{K}(0, 2\mu)}. \tag{3.21}$$

Note that, for diffusive dispersal, (3.20) reduces to the classical Wright-Malécot formula for isolation by distance [15].

For transient dispersal, we must consider a coalescence kernel of finite width, and (3.19) generalizes to:

$$\begin{aligned} p(t|x) = &\int dy \frac{1}{\rho}\mathcal{N}(y)K(x-y|t) \\ &- \int_0^t dt' \int dy\, p(t-t'|y)\frac{1}{\rho}\mathcal{N}(y)K(x-y|t'). \end{aligned} \tag{3.22}$$

(3.22) is exactly the same as (3.19) except that now we must integrate over possible locations y of coalescence at both t and t'. Taking the Laplace transform of (3.22) now gives:

$$\psi(x) = \frac{1}{\rho}\int dy(1-\psi(y))\mathcal{N}(y)\widetilde{K}(x-y, 2\mu). \tag{3.23}$$

To simplify (3.23), we can make the approximation that $1-\psi(y)$ is nearly constant over all separations $|y| \lesssim \delta$ where $\mathcal{N}(y)$ is non-negligible, allowing us to pull it out of the integral:

$$\psi(x) \approx \frac{1-\psi(0)}{\rho}\int dy\mathcal{N}(y)\widetilde{K}(x-y, 2\mu). \tag{3.24}$$

This approximation will necessarily be accurate when identity is low, $\psi(0) \ll 1$ because $1-\psi$ will be close to 1 for all y. However, for $1-\psi(0) \ll 1$, the approximation can become inaccurate; we discuss this below. At long distances $x \gg \delta$, $\widetilde{K}$ will also be roughly constant in the integral, and we simply recover (3.20), although now only as an approximation:

$$\psi(x \gg \delta) \approx \frac{1-\psi(0)}{\rho}\widetilde{K}(x, 2\mu). \tag{3.25}$$

We see that the details of the short-range behavior only affect the long-range probability of identity by descent through the overall factor $1-\psi(0)$ [15]. Mathematically, the main challenge is to find simple expressions for $\psi(0)$ and especially $\widetilde{K}$.

Because (3.25) is invalid for $x=0$, we cannot solve it directly for $\psi(0)$ as we could with (3.20), and so we must also work with (3.24). We can simplify the convolution in (3.24) by taking the spatial Fourier transform $\mathcal{F}\{\cdot\}$:

$$\widehat{\psi}(k) \approx \frac{1-\psi(0)}{\rho}\widehat{\mathcal{N}}(k)\widehat{\widetilde{K}}(k,2\mu). \tag{3.26}$$

where $\widehat{\psi}$ and $\widehat{\widetilde{K}}$ are the Fourier transforms of ψ and $\widetilde{K}$.

Lévy flight dispersal

For Lévy flights, the characteristic function is $\widehat{K}(k|t) = \exp(-2D_\alpha t|k|^\alpha)$ and the Fourier-Laplace transform is $\widehat{\widetilde{K}}(k,2\mu) = 1/(2\mu+2D_\alpha|k|^\alpha)$. (3.26) for $\widehat{\psi}$ is correspondingly simple:

$$\frac{\widehat{\psi}(k)}{1-\psi(0)} \approx \frac{\widehat{\mathcal{N}}(k)}{2\rho(\mu+D_\alpha|k|^\alpha)}. \tag{3.27}$$

To get an explicit expression for ψ, we need to specify a form for the coalescence kernel $\mathcal{N}$. We will use a normal distribution with standard deviation δ, which has the simple Fourier transform $\widehat{\mathcal{N}}(k) = \exp(-\delta^2k^2/2)$. Then we can invert the Fourier transform in (3.27):

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{2\pi\rho}\int_0^\infty dk \frac{\cos(kx)e^{-\delta^2k^2/2}}{\mu+D_\alpha k^\alpha}, \tag{3.28}$$

which can be re-expressed in dimensionless units as

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{2\pi\rho\mu\overline{x}}\int_0^\infty d\kappa \frac{\cos(\kappa x/\overline{x})e^{-(\delta/\overline{x})^2\kappa^2/2}}{1+\kappa^\alpha}. \tag{3.29}$$

Examining (3.28), we see that the power-law tail in the integrand can be cut off either when oscillations in the cosine factor become rapid at $k\sim 1/x$ or by the normal factor at $k\sim 1/\delta$. As long as we are sampling pairs that are outside the immediate range of coalescence, $x \gg \delta$, the former cutoff will happen at lower k, and therefore

the normal factor can be neglected (by setting $\delta = 0$), leaving (in dimensionless form):

$$\frac{\psi(x \gg \delta)}{1-\psi(0)} \approx \frac{1}{2\pi\rho\mu\overline{x}} \int_0^\infty d\kappa \frac{\cos(\kappa x/\overline{x})}{1+\kappa^\alpha}. \tag{3.30}$$

(3.30) can equivalently be derived directly from (3.25) by substituting in the Lévy flight dispersal kernel and writing $\widetilde{K}$ as the inverse Fourier transform of $\widehat{\widetilde{K}}$.

We can solve (3.29) for $\psi(x)$ by first evaluating it at $x = 0$ to find $\psi(0)$; we do this below. But it is interesting that the ratio $\Psi(x) \equiv \psi(x)/(1-\psi(0))$ has the simplest relationship to the underlying parameters, as shown by Rousset [140] for short-range dispersal. Ψ is closely related to Rousset [141]'s statistic a_r: $a_r = \Psi(0) - \Psi(r)$. It is also related to the expected pairwise F_{ST} between demes separated by x:

$$E\left[F_{\mathrm{ST}}(x)\right] = \frac{\Psi(0)-\Psi(x)}{2+\Psi(0)-\Psi(x)}.$$

Probability of identity for distant pairs $x \gg \overline{x}$, $\alpha < 2$

For large $x \gg (D_\alpha t)^{1/\alpha}$, the dispersal kernel has a simple asymptotic form for $\alpha < 2$ (Nolan [121], Theorem 1.12):

$$K(x \gg (D_\alpha t)^{1/\alpha}|t) \approx \frac{2\Gamma(\alpha+1)}{\pi} \sin\left(\frac{\pi\alpha}{2}\right) \frac{D_\alpha t}{x^{\alpha+1}}.$$

Plugging this into (3.25) and evaluating the Laplace transform gives the probability of identity for distant pairs, which was originally found by Nagylaki [114]:

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{\Gamma(\alpha+1)}{2\pi} \sin\left(\frac{\pi\alpha}{2}\right) \frac{D_\alpha}{\rho\mu^2 x^{\alpha+1}}. \tag{3.31}$$

Probability of identity for distant pairs $x \gg \overline{x}$, $\alpha > 2$

There is no stable distribution with $\alpha > 2$, but in discrete-time models such as the one we use in our simulations, we can consider single-generation jump kernels $K(y|1)$ with power-law tails with $\alpha > 2$. These will approach a diffusion with diffusion constant $D = \mathrm{Var}(K)/4$. At long distances $y \gg \sqrt{Dt}$, however, the tail will still be dominated by the probability of taking a single large jump [161], so for $x \gg \overline{x}$, we will

have $K(x|t \lesssim 1/\mu) \approx K(x|1)t$. Plugging this into (3.25) and evaluating the Laplace transform gives:

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{K(x|1)}{4\rho\mu^2}. \tag{3.32}$$

For the F-distribution kernel (3.15) used in the simulations, this is

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{D_\alpha}{8\rho\mu^2(x/\alpha)^{\alpha+1}}, \tag{3.33}$$

where we have defined $D_\alpha \equiv \omega^\alpha$/generation, i.e., D_α has the same value as ω^α, but its dimensions are now length$^\alpha$/time. (3.33) is confirmed by simulations (Fig. 3.5). We can then use the classic diffusive expression for $\psi(0)$ to get an explicit expression for probability of identity at large distances:

$$\psi(x \gg \overline{x}) \approx \left(1 + \frac{1}{8\rho\overline{x}\mu}\right)^{-1} \frac{D_\alpha}{4\rho\mu^2(x/\alpha)^{\alpha+1}}. \tag{3.34}$$

Moderately heavy-tailed dispersal, $1 < \alpha < 2$

For $\alpha > 1$, (3.20) and (3.30) are exact for all x (when $\delta = 0$). Evaluating (3.30) for $x = 0$ gives $\psi(0)$:

$$\psi(0) = \frac{1}{2\alpha\sin(\pi/\alpha)\rho\overline{x}\mu + 1}. \tag{3.35}$$

Plugging (3.35) into (3.31) lets us solve for $\psi(x)$ at large distances $x \gg \overline{x}$:

$$\psi(x \gg \overline{x}) \approx \frac{\sin(\pi\alpha/2)\Gamma(\alpha+1)/(2\pi)}{1 + 1/(2\alpha\sin(\pi/\alpha)\rho\overline{x}\mu)} \frac{D_\alpha}{\rho\mu^2 x^{\alpha+1}}. \tag{3.36}$$

For $0 < x \ll \overline{x}$, Janakiraman [85] (Eq. (C1)) found that to leading order ψ falls off as:

$$\psi(x \ll \overline{x}) \approx \psi(0)\left[1 - \frac{\alpha\sin(\pi/\alpha)}{\Gamma(\alpha)\cos(\pi(1-\alpha/2))}\left(\frac{x}{\overline{x}}\right)^{\alpha-1}\right]. \tag{3.37}$$

When $\alpha = 2$, the above expression is equivalent to the classic diffusive result for $x \ll \overline{x}$, which can be found by integrating (3.23) with $\delta = 0$:

$$\psi(x) = \frac{e^{-x/\overline{x}}}{4\rho\overline{x}\mu + 1}. \tag{3.38}$$

Very heavy-tailed dispersal, $\alpha < 1$

For $\alpha < 1$, the finite width δ of the coalescence kernel is important for determining $\psi(0)$. Setting $x = 0$ in (3.29) gives:

$$\begin{aligned}\frac{\psi(0)}{1-\psi(0)} &\approx \frac{1}{2\pi\rho\mu\overline{x}}\int_0^\infty d\kappa \frac{e^{-(\delta/\overline{x})^2\kappa^2/2}}{1+\kappa^\alpha} \\ &\approx \frac{\Gamma(1/2-\alpha/2)}{2^{(\alpha+3)/2}\pi\rho D_\alpha \delta^{1-\alpha}},\end{aligned}$$

where in evaluating the integral we have assumed that $\delta \ll \overline{x}$, i.e., that the mutation rate is not extremely large. We see that on small scales, the probability of identity by descent is independent of the mutation rate (Fig. 3.6), i.e., there is a large probability that individuals from the same deme are differentiated even for infinitesimal mutation rates:

$$\psi(0) \approx 1 \Big/ \left[1 + \frac{2^{(\alpha+3)/2}\pi}{\Gamma(1/2-\alpha/2)}\rho D_\alpha \delta^{1-\alpha}\right] \tag{3.39}$$

Very heavy-tailed dispersal of nearby lineages causes them to quickly wander away from each other, and for infinite range size many pairs will never coalesce. While (3.39) is only accurate for $\psi(0) \ll 1$, the independence from mutation rate should persist even for large $\psi(0)$.

Plugging (3.39) for $\psi(0)$ into (3.31) gives an explicit expression for the probability of identity of distant pairs with $x \gg \overline{x}$:

$$\psi(x \gg \overline{x}) \approx \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)/(2\pi)}{1+\Gamma(1/2-\alpha/2)/(2^{(\alpha+3)/2}\pi\rho D_\alpha\delta^{1-\alpha})}\frac{D_\alpha}{\rho\mu^2 x^{\alpha+1}}.$$

For pairs that are nearby but still well outside of coalescence range, $\delta \ll x \ll \overline{x}$, the integral in (3.30) is dominated by $\kappa \gg 1$ and is approximately:

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{\Gamma(1-\alpha)\sin(\pi\alpha/2)}{2\pi}\frac{x^{\alpha-1}}{\rho D_\alpha}.$$

Again, the probability of identity is independent of the mutation rate to lowest order. Substituting in (3.39) gives an explicit expression for ψ:

$$\psi(\delta \ll x \ll \overline{x}) \approx \frac{\Gamma(1-\alpha)\sin(\pi\alpha/2)/(2\pi)}{1+\Gamma(1/2-\alpha/2)/(2^{(\alpha+3)/2}\pi\rho D_\alpha\delta^{1-\alpha})}\frac{x^{\alpha-1}}{\rho D_\alpha}. \tag{3.40}$$

While $\psi(x)$ is independent of μ only for $\alpha < 1$, note that the rate at which $\psi(x)$ changes, $\partial_x\psi(x)$, is independent of μ for all $\alpha \leq 2$ when ρ is large (in one dimension).

Marginal case $\alpha = 1$

The analysis of the marginal case $\alpha = 1$ is essentially the same as for $\alpha < 1$ above, but we have separated it out because the form of the final expressions is very different. As with $\alpha < 1$, the finite coalescence width δ is important for $x = 0$:

$$\begin{aligned}\frac{\psi(0)}{1-\psi(0)} &\approx \frac{1}{2\pi\rho D_1}\int_0^\infty d\kappa \frac{e^{-(\delta/\overline{x})^2\kappa^2/2}}{1+\kappa} \\ &= \frac{2\ln(\overline{x}/\delta) + \ln 2 - \gamma}{4\pi\rho D_1},\end{aligned}$$

where $\gamma \approx 0.58$ is Euler's constant. Again assuming $\delta \ll \overline{x} = D_1/\mu$, the constant terms in the numerator can be neglected and ψ is approximately:

$$\psi(0) \approx \left(1 + \frac{2\pi\rho D_1}{\ln(\overline{x}/\delta)}\right)^{-1}. \tag{3.41}$$

Recall that the approximation we used to derive (3.41) ((3.24)) is only justified when $\psi(0) \ll 1$.

For pairs that are nearby but still well outside of coalescence range, $\delta \ll x \ll \overline{x}$, (3.30) gives:

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{\ln(\overline{x}/x) - \gamma}{2\pi\rho D_1}. \tag{3.42}$$

Plugging the expression (3.41) for $\psi(0)$ into (3.42) and (3.31) gives explicit expressions for $\psi(x)$ at both short and long distances:

$$\psi(\delta \ll x \ll \overline{x}) \approx \frac{\ln(\overline{x}/x)}{2\pi\rho D_1 + \ln(\overline{x}/\delta)} \tag{3.43}$$

$$\psi(x \gg \overline{x}) \approx \frac{1}{2\pi\rho D_1 + \ln(\overline{x}/\delta)}\left(\frac{\overline{x}}{x}\right)^2. \tag{3.44}$$

3.5.4 Analytical model in two dimensions

Generic dispersal

For generic dispersal, the solution for ψ in two dimensions can again be found from (3.23), now with the integral over two spatial dimensions. The Fourier transform $\widehat{\psi}$

has the same form as the one-dimensional equation (3.26):

$$\widehat{\psi}(k) \approx \frac{1-\psi(0)}{\rho}\widehat{\mathcal{N}}(k)\widehat{\widetilde{K}}(k,2\mu), \tag{3.45}$$

where again we make the approximation that $1-\psi(x)$ is approximately constant over the x values where $\mathcal{N}(x)$ is non-negligible. This is again accurate for $\psi(0) \ll 1$, but may need to be adjusted for $1-\psi(0) \ll 1$. While (3.45) looks exactly like the one-dimensional expression (3.26), its interpretation is different: k is now the radial coordinate in two-dimensional k-space, and if we want to transform back to real space, we must use the two-dimensional inverse Fourier transform. For pairs that are far outside coalescence range, $x \gg \delta$, the simple relation (3.25) between $\psi(x)$ and $\widetilde{K}(x,2\mu)$ still holds.

Lévy flight dispersal

For a two-dimensional Lévy flight, the dispersal kernel takes the form of an isotropic stable distribution [174]:

$$K(y|t) = \frac{1}{2\pi}\int_0^\infty dk\, kJ_0(ky)\exp\left(-2D_\alpha t k^\alpha\right), \tag{3.46}$$

where $K(y|t)$ is the probability density of being at a particular point a distance y away from the position at time 0, and J_0 is the zeroth Bessel function of the first kind. (3.46) is the two-dimensional inverse Fourier transform (equivalently, the inverse zeroth-order Hankel transform) of the characteristic function $\widehat{K}(k|t) = \exp\left(-2D_\alpha t k^\alpha\right)$. The Fourier-Laplace transform is again $\widehat{\widetilde{K}}(k,2\mu) = 1/(2\mu + 2D_\alpha|k|^\alpha)$. At large distances, $y \gg (D_\alpha t)^{1/\alpha}$, K has a power-law tail [120]:

$$K(y \gg (D_\alpha t)^{1/\alpha}|t) \approx \frac{\alpha^2\Gamma(\alpha/2)^2}{2^{1-\alpha}\pi^2}\sin\left(\frac{\pi\alpha}{2}\right)\frac{D_\alpha t}{y^{\alpha+2}}. \tag{3.47}$$

In two dimensions, we must allow coalescence to take place at a finite distance for all α [111]. For the coalescence kernel, we use an isotropic normal distribution $\mathcal{N}(x)$ with mean zero and standard deviation δ, with coalescence taking place at rate

$\frac{1}{\rho}\mathcal{N}(X_t)$. Inverting the Fourier transform in (3.45) then gives:

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{4\pi\rho}\int_0^\infty dk \frac{kJ_0(kx)e^{-\delta^2k^2/2}}{\mu + D_\alpha k^\alpha} \tag{3.48}$$

$$= \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})e^{-(\delta/\overline{x})^2\kappa^2/2}}{1+\kappa^\alpha}, \tag{3.49}$$

The analysis of (3.49) parallels that of the one-dimensional case, but all $\alpha < 2$ can be treated together for all distances x, not just $x \gg \overline{x}$, and so we can conduct one unified analysis moving from short distances to long ones.

Probability of identity for co-located pairs, $x = 0$

For pairs sampled from the same location, $x = 0$, the Bessel function in (3.49) is simply equal to one and can be dropped:

$$\frac{\psi(0)}{1-\psi(0)} \approx \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa e^{-(\delta/\overline{x})^2\kappa^2/2}}{1+\kappa^\alpha}$$

$$\approx \frac{\Gamma(1-\alpha/2)}{2^{2+\alpha/2}\pi\rho D_\alpha \delta^{2-\alpha}}, \tag{3.50}$$

where in the last line we have assumed that $\delta \ll \overline{x}$. Intuitively, (3.50) can be understood as roughly the ratio between the time to coalesce, i.e., the neighborhood size $\sim \rho\delta^2$ and the time $\sim \delta^\alpha/D_\alpha$ that the lineages will spend in the same neighborhood before jumping apart. Note that mutation does not enter: in two dimensions, all $\alpha < 2$ act like $\alpha < 1$ does in one dimension, where locally mutation is irrelevant. Again, (3.50) is only accurate for $\psi(0) \ll 1$.

Solving (3.50) for ψ gives:

$$\psi(0) \approx \left(1 + \frac{2^{2+\alpha/2}\pi}{\Gamma(1-\alpha/2)}\rho D_\alpha \delta^{2-\alpha}\right)^{-1}. \tag{3.51}$$

For $\alpha = 2$, integrating (3.49) with $x = 0$ recovers the classic diffusive result in two dimensions, which we expect to hold for pairs in contact when $\alpha \geq 2$ [15]:

$$\psi(0) \approx \frac{\ln(\overline{x}/\delta)}{\ln(\overline{x}/\delta) + 4\pi\rho D_2}. \tag{3.52}$$

Probability of identity for separated but nearby pairs, $\delta \ll x \ll \overline{x}$

For pairs that are outside coalescence range, $x \gg \delta$, we can find ψ from (3.25):

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{1+\kappa^\alpha}. \tag{3.53}$$

This looks different from the one-dimensional equation (3.28) because now we had to apply the two-dimensional inverse Fourier transform to $\widehat{\widetilde{K}}$ to obtain $\widetilde{K}$. For nearby pairs $x \ll \overline{x}$, the integral in (3.53) is dominated by $\kappa \gg 1$ and for $\alpha < 2$ we can approximate the denominator in the integrand as $1+\kappa^\alpha \approx \kappa^\alpha$, giving:

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{\Gamma(1-\alpha/2)}{\Gamma(\alpha/2)2^{1+\alpha}\pi\rho D_\alpha} x^{\alpha-2}. \tag{3.54}$$

The convergence of (3.53) to (3.54) is however quite slow in $x/\overline{x}$ when α is close to 0 or 2. For example, for $x/\overline{x} = 0.01$, the two expressions differ by $\approx 30-40\%$ for $\alpha = 0.25$ and $\alpha = 1.75$, and only approach to within 10% of each other at extreme values of $x/\overline{x}$ ($\approx 10^{-5}$ and $\approx 10^{-4}$ for $\alpha = 0.25$ and $\alpha = 1.75$, respectively).

Plugging (3.51) for $\psi(0)$ into (3.54) lets us solve for ψ:

$$\psi(\delta \ll x \ll \overline{x}) \approx \left(1 + \frac{\Gamma(1-\alpha/2)}{2^{2+\alpha/2}\pi\rho D_\alpha \delta^{2-\alpha}}\right)^{-1} \frac{\Gamma(1-\alpha/2)}{\Gamma(\alpha/2)2^{1+\alpha}\pi} \frac{x^{\alpha-2}}{\rho D_\alpha}. \tag{3.55}$$

We see that in two dimensions, relatedness at short distances is independent of μ to leading order for all $\alpha < 2$. However, the slow convergence mentioned above means that for most biologically reasonable parameter values, this should be interpreted as meaning that the dependence on mutation rate is weak rather than negligible.

For $\alpha = 1$ and $\delta \ll x \ll \overline{x}$ we recover Eq. (A6) of Chave and Leigh Jr [40] for Cauchy dispersal. Note that they consider distances large compared to the typical single-generation dispersal distance, $c \equiv (2D_\alpha)^{1/\alpha}$, but small compared to $\overline{x}$, and thus our result for $\delta \ll x \ll \overline{x}$ is consistent with their findings.

For $\alpha = 2$ and $\delta \ll x \ll \overline{x}$ we can recover the known result for diffusive motion by approximating (3.49) as

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{J_0(\kappa x/\overline{x})\kappa}{1+\kappa^2}. \tag{3.56}$$

Integrating (3.56) confirms that we find logarithmic scaling of $\psi(x)$ at short distances [15]:

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{4\pi\rho D_2}\ln(\overline{x}/x), \tag{3.57}$$

which we expect to hold at $\delta \ll x \ll \overline{x}$ for all $\alpha \geq 2$.

Probability of identity by descent for distant pairs, $x \gg \overline{x}$

The probability of identity by descent for distant pairs $x \gg \overline{x} \gg \delta$ can be immediately be read off from (3.25) by substituting in the tail of the two-dimensional dispersal kernel (3.47) for $\alpha < 2$:

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{\alpha^2\Gamma(\alpha/2)^2}{2^{3-\alpha}\pi^2}\sin\left(\frac{\pi\alpha}{2}\right)\frac{D_\alpha}{\rho\mu^2}x^{-\alpha-2}. \tag{3.58}$$

Plugging in (3.51) for $\psi(0)$ lets us solve for ψ:

$$\psi(x \gg \overline{x}) \approx \frac{\alpha^2\Gamma(\alpha/2)/[2^{3-\alpha}\pi\Gamma(1-\alpha/2)]}{1+\Gamma(1-\alpha/2)/[2^{2+\alpha/2}\pi\rho D_\alpha\delta^{2-\alpha}]}\frac{D_\alpha x^{-\alpha-2}}{\rho\mu^2}.$$

When $\alpha = 2$, we instead recover classic expression for two-dimensional diffusive motion at large distances [15]:

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{1}{4\rho D_2}\frac{\exp(-x/\overline{x})}{\sqrt{2\pi x/\overline{x}}}. \tag{3.59}$$

For our simulations, rather than using a true two-dimensional stable distribution, we use radial draws from a one-dimensional stable distribution and then pick a direction at random. The resulting dispersal kernel is shown in (3.8). At large distances, $x \gg \overline{x}$, we can apply (3.32) to find that the tail expression for IBD (when α ¡ 2) is simply $(\pi x)^{-1}$ times (3.31):

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{\Gamma(\alpha+1)}{2\pi^2}\sin\left(\frac{\pi\alpha}{2}\right)\frac{D_\alpha}{\rho\mu^2 x^{\alpha+2}}. \tag{3.60}$$

For the finite variance 2D kernel (3.16) used when α ¿ 2, we can again apply (3.32) to find the tail expression for IBD:

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{D_\alpha}{4\pi\alpha\rho\mu^2(x/\alpha)^{\alpha+2}}, \tag{3.61}$$

where D_α is defined as ω^α for the single lineage kernel (3.16).

While the above expressions are accurate in continuous time, we add an extra factor to adjust for discrete time in the tail expressions of Fig. 3.4 where $\mu = 1$. This factor f is the ratio between the discrete time sum of te^{-2t} from $t = 1$ to $t = \infty$ and the continuous time integral of te^{-2t} from $t = 0$ to $t = \infty$: $f \sim 0.724$.

3.5.5 Coalescence time distribution

In this section we will find asymptotic expressions for the coalescence time distribution. As stated in the Results, intuitively we can think of the probability of identity ψ as measuring the probability of the pair of lineages coalescing $\lesssim 1/(2\mu)$ generations ago. We can make this statement more rigorous using the Hardy-Littlewood Tauberian theorem connecting the long (short) time probability of coalescence to the small (large) mutation rate limit of ψ. It states that a function $f(t)$ has the limiting behavior $f(t) \to \frac{1}{\Gamma(\beta)} t^{\beta-1} L(t)$ as $t \to \infty$ ($t \to 0$), where L is a slowly varying function and $\beta > 0$, if and only if its Laplace transform $\widetilde{f}(2\mu)$ has the limiting behavior $\widetilde{f}(2\mu) \to (2\mu)^{-\beta} L(1/(2\mu))$ as $\mu \to 0$ ($\mu \to \infty$) (Feller [58], XIII.5, Theorem 4).

Recent times

First we will consider the limit of recent times, $t \to 0$ / $\mu \to \infty$. For pairs sampled within coalescence range, $x \lesssim \delta$, by definition the (density of the) coalescence time distribution approaches $p(t|x) \sim 1/(\rho\delta^d)$, up to numerical factors that depend on the details of the coalescence kernel. Here d is the dimensionality of the range, $d = 1$ or 2. For pairs sampled well outside coalescence range, $x \gg \delta$, we can assume that $x \gg \overline{x}$ as well, since $\overline{x} = (D_\alpha/\mu)^{1/\alpha} \to 0$ as $\mu \to \infty$. We can also assume that $1 - \psi(0) \to 1$ is independent of μ to leading order. (For $\alpha < d$ our expressions for $\psi(0)$ (3.41) and (3.51) are also independent of μ and non-zero, but these are only valid when $\overline{x} \gg \delta$, i.e., when μ is not arbitrarily large.) We can therefore apply the Tauberian theorem

to (3.31) and (3.58) to obtain:

$$\begin{aligned} p(t \ll x^\alpha/D_\alpha|x) &\approx 2d^\alpha \left(\frac{\Gamma(1+\alpha/d)}{\pi}\right)^d \sin\left(\frac{\pi\alpha}{2}\right) \frac{D_\alpha t}{\rho x^{\alpha+d}} \\ &\approx \frac{1}{\rho} K(x|t). \end{aligned} \tag{3.62}$$

Our heuristic derivation in the Results section essentially proceeded in the opposite direction, starting from $p(t \ll x^\alpha/D_\alpha|x) \approx \frac{1}{\rho}K(x|t)$ and then deriving $\psi(x)$ from that. (3.62) is thus essentially just a restatement of our expressions for the tail of ψ, and its accuracy can be seen from the same simulation results shown in Fig. 3.8 and Fig. 3.5.

Long times

While there is a single unified expression for p in the $t \to 0$ limit, corresponding to the single expression for ψ in the $x \to \infty$ limit, for the opposite limit, $t \to \infty$ / $\mu \to 0$, we must treat different values of α separately, just as we did for ψ at small x. We verify our results with simulations, shown in Fig. 3.9. Note that these expressions will only hold at long times on an infinite range. For any range of finite size, the right tail of the coalescence time distribution will decay exponentially, as in the case of a panmictic population [166].

For $\alpha < d$, we can simply take the inverse Laplace and Fourier transforms of (3.26) to find p, because $1 - \psi(0)$ is independent of μ to leading order. Since we are concerned with times long compared to the time for the lineages to traverse the coalescence zone, $t \gg \delta^\alpha/D_\alpha$, the normal factor in (3.26) can be neglected and $p(t|x)$ is simply given by the inverse Laplace transform of (3.25):

$$p(t \gg \delta^\alpha/D_\alpha|x) \approx \frac{1-\psi(0)}{\rho} K(x|t) \tag{3.63}$$

$$\approx \frac{1-\psi(0)}{\alpha 2^{d-1}\pi\rho} \frac{\Gamma(d/\alpha)}{(2D_\alpha t)^{d/\alpha}} \text{ for } t \gg \frac{x^\alpha}{D_\alpha}. \tag{3.64}$$

Integrating (3.64) yields the cumulative distribution for $t \gg x^\alpha/D_\alpha$:

$$P\left(t \gg \frac{x^\alpha}{D_\alpha}\middle| x\right) \approx P(\infty|x) - \frac{\Gamma(d/\alpha)(1-\psi(0))(2D_\alpha)^{-d/\alpha}}{(d-\alpha)2^{d-1}\pi\rho t^{d/\alpha-1}}, \tag{3.65}$$

where $P(\infty|x) = \lim_{\mu\to 0}\psi(x)$ is given by (3.39), (3.40), (3.51), or (3.55), depending on x and d.

For the marginal case $\alpha = d$, we can use a slightly different statement of the Tauberian theorem that applies for $\beta = 0$ (Feller [58], XIII.5, Theorem 2) to convert (3.41), (3.43), (3.52), and (3.57) to expressions for the cumulative distribution P:

$$P(t \gg x/D_d|x) \approx \begin{cases} \left[1 + \frac{2d\pi\rho D_d}{\ln(2D_d t/\delta)}\right]^{-1} & \text{for } x \ll \delta \\ \frac{\ln(2D_d t/x)}{2d\pi\rho D_d + \ln(2D_d t/\delta)} & \text{for } x \gg \delta. \end{cases} \tag{3.66}$$

We can then differentiate to find the density p:

$$\begin{aligned} p\left(t \gg \frac{x}{D_d}\middle| x\right) \approx & \frac{2d\pi\rho D_d}{t\left[2d\pi\rho D_d + \ln(2D_d t/\delta)\right]^2} \\ & \times \begin{cases} 1 & \text{for } x \ll \delta \\ 1 + \frac{\ln(x/\delta)}{2d\pi\rho D_d} & \text{for } x \gg \delta. \end{cases} \end{aligned} \tag{3.67}$$

Note that in two dimensions, $\alpha = d$ represents the diffusive limit, and we expect these expressions for the marginal case to hold for all $\alpha \geq 2$.

For $d = 1$ and $1 < \alpha \leq 2$, since the cumulative distribution approaches one at large times, $\lim_{\mu\to 0}\psi(x) = \lim_{t\to\infty} P(t|x) = 1$, we must instead consider the complementary cumulative distribution, $\overline{P}(t|x) \equiv 1 - P(t|x)$. Its Laplace transform is:

$$\begin{aligned} \widetilde{\overline{P}}(2\mu) &= \frac{1}{2\mu} - \widetilde{P}(2\mu|x) \\ &= \frac{1}{2\mu}\left[1 - \psi(x)\right]. \end{aligned}$$

We can now apply the Tauberian theorem to $\overline{P}$ and $\widetilde{\overline{P}}$. Since we are taking the $\mu \to 0$ limit, we have $\overline{x} \to \infty$, and we need only consider $\psi(x \ll \overline{x})$. Inspecting (3.35) and (3.37), we see that they have the limit:

$$\frac{1}{2\mu}\left[P(\infty|x) - \psi(x)\right] \to \alpha\sin(\pi/\alpha)\rho\overline{x} \text{ as } \mu \to 0.$$

Since $\overline{x} = (D_\alpha/\mu)^{1/\alpha}$, $\overline{P}$ has the limit:

$$\overline{P}(t \gg x^\alpha/D_\alpha|x) \approx \alpha\sin\left(\frac{\pi}{\alpha}\right)\frac{\rho(2D_\alpha)^{1/\alpha}}{t^{1-1/\alpha}}. \tag{3.68}$$

Differentiating (3.68) yields the density $p(t|x)$:

$$p(t \gg x^\alpha/D_\alpha|x) \approx (\alpha - 1)\sin\left(\frac{\pi}{\alpha}\right)\frac{\rho(2D_\alpha)^{1/\alpha}}{t^{2-1/\alpha}}, \tag{3.69}$$

in agreement with Janakiraman [85]'s Eq. 19.

For $\alpha = 2$, (3.68) and (3.69) simplify to the classic diffusive results:

$$\overline{P}(t \gg x^2/D|x) \approx 2\rho\sqrt{2D/t}, \tag{3.70}$$

$$p(t \gg x^2/D|x) \approx \rho\sqrt{2D/t^3}. \tag{3.71}$$

For $\alpha > 2$, we expect the coalescence rate $p(t|x)/\overline{P}(t|x)$ to behave similarly at long times, since the dispersal approaches a diffusion. But the distribution P may be different, due to differences in the probability of early coalescence (Fig. 3.9, bottom right).

3.5.6 Breakdown of models at small scales

Great care must be taken in defining coalescent models in continuous space in order to guarantee that they have a consistent forward-time biological interpretation [16, 59]. We have not done this, and therefore the microscopic behavior of our models does not correspond to any biological population. However, the behavior at large scales (time long compared to one generation, distance long compared to the coalescence scale δ and the typical single-generation dispersal distance $c \equiv (2D_\alpha)^{1/\alpha}$) should still be realistic [15]. We have also shown via simulation that our results describe a stepping-stone model of discrete demes of size $\sim \rho\delta^d$ separated by distance $\sim \delta$.

The key place in which the microscopic details matter even for large distances and long times is the factor $1 - \psi(0)$ which appears in many of our expressions. As discussed above, for $\alpha < d$ even here the microscopic details are not necessarily important, but for $\alpha \geq d$ they are. Practically speaking, this quantity would typically have to simply be measured in a population or else treated as a fitting parameter when matching the large-scale predictions to data.

At a microscopic level, we expect that our continuous-time analytic model should deviate from discrete-time models such as the one we use in our simulations. As shown in Fig. 3.10, this becomes apparent for $\alpha < 1$ in one dimension (or more generally, $\alpha < d$). The two differ at scales smaller than the typical single-generation dispersal distance, $x < c = (2D_\alpha)^{1/\alpha}$, when this scale is large compared to the coalescence scale, $c \gg \delta$. In continuous time, nearby pairs with $x \ll c$ would be able to coalesce at times smaller than a single generation, $t \ll 1$. But in discrete time no pairs can coalesce until $t = 1$, by which time the dispersal kernel $K(x|1)$ is roughly flat out to $x \lesssim c$, and probability of identity thus becomes approximately constant for $x \lesssim c$. (For $\alpha \geq d$, the continuous-time model already predicts that ψ should be changing slowly at $x \ll \overline{x}$, and therefore we do not expect a disagreement with the discrete-time model.) Recall that our discrete-time model assumes no coalescence at $t = 0$ even for lineages starting at $x < \delta$; if we were to change this, ψ would discontinuously jump up to a second, higher plateau for $x < \delta$.

We can estimate the discrete-time value of $\psi(x \ll c)$ from a heuristic argument, at least when $\psi \ll 1$. In the absence of coalescence, the probability of the lineages being within coalescence range of each other in generation $t \geq 1$ is $\approx (2\delta)K(x|t) \approx (2\delta)K(0|t)$. For $\psi \ll 1$, including the possibility of coalescence will only slightly decrease this probability. Given that the lineages are in coalescence range, they coalesce with probability $1/(2\delta\rho)$. So in any one generation the probability of coalescence is $\approx K(0|t)/\rho$ and we can find ψ by summing over all generations:

$$\begin{aligned}\psi(x \ll c) &\approx \sum_{t=1}^{\infty} \frac{K(0|t)}{\rho} \\ &= \frac{\Gamma(1/\alpha)\zeta(1/\alpha)}{\alpha} \frac{1}{\rho c},\end{aligned} \tag{3.72}$$

where ζ is the Riemann zeta function. Fig. 3.10 shows that (3.72) accurately describes the simulations.

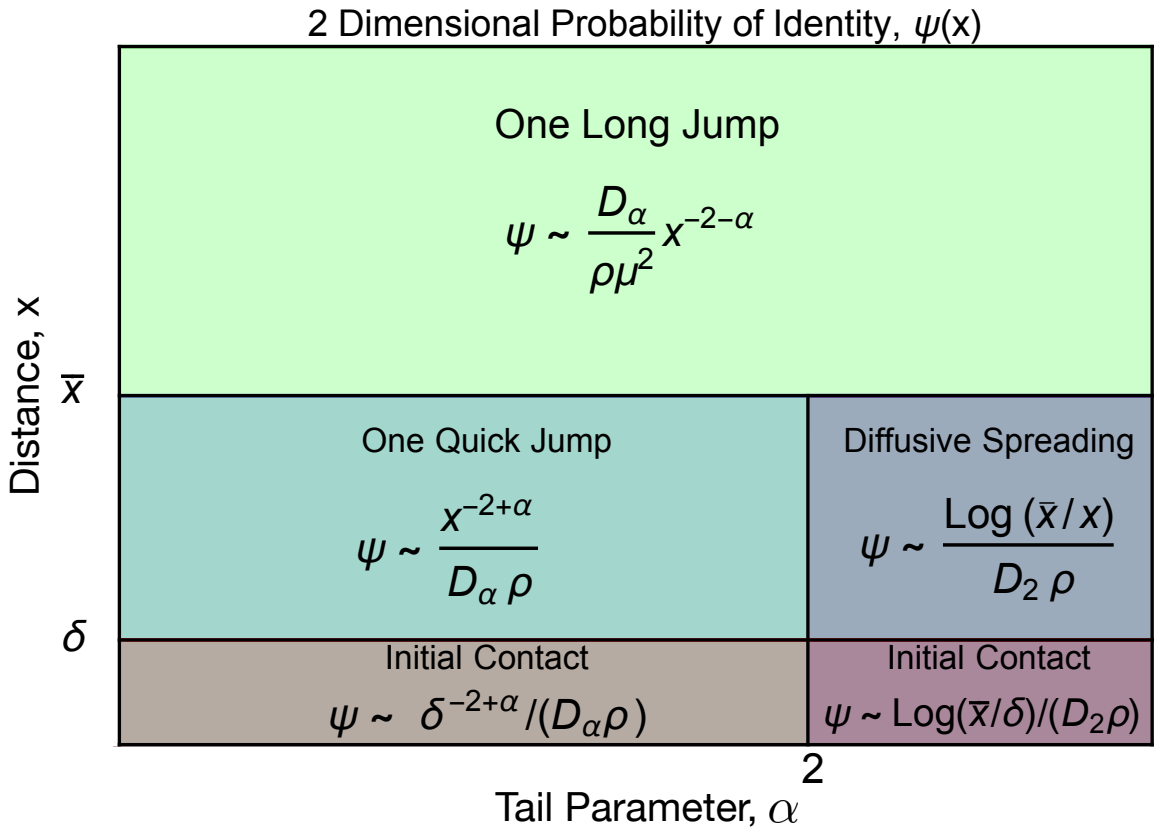

Figure 3.2: **For power-law dispersal, the form of isolation by distance in two dimensions is universal at long distances.** Approximate form for the probability of identity as a function of distance, $\psi(x)$, for different dispersal kernels α. Different regimes of the parameter space are separated by solid lines, and labelled by their qualitative dynamics. Coalescence for distant pairs, $x \gg \overline{x}$, typically occurs via one long jump, which leads to the power-law scaling at large distances predicted by (3.4). Nearby pairs, $x \ll \overline{x}$, typically either coalesce very quickly or disperse far away from each other, so the probability of identity is nearly independent of the mutation rate, as shown in (3.5). This quick coalescence is effectively diffusive for $\alpha > 2$, while for $\alpha < 2$, it is typically driven by a single jump. We use "$\sim$" to denote proportionality in the limit of large population density where $\psi(0) \ll 1$.

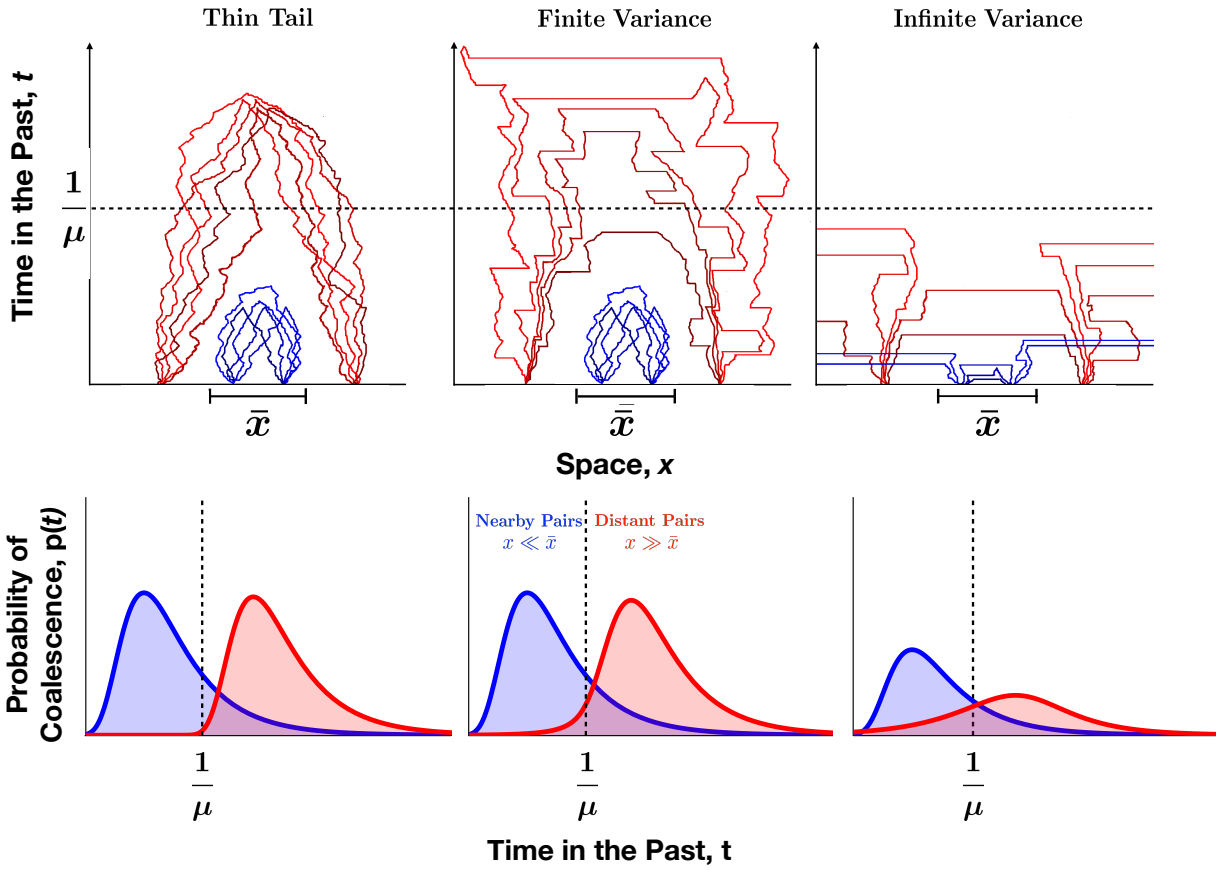

Figure 3.3: **Long-range jumps affect when and where lineages coalesce.** Qualitative illustrations of lineage dynamics and coalescence time distributions for each of the three dispersal regimes in two dimensions. Typical histories are shown for nearby samples ($x \ll \overline{x}$, blue) and distant samples ($x \gg \overline{x}$, red). **Left**: For thin-tailed dispersal distributions, motion is effectively diffusive and separation x is a relatively good predictor of coalescence time. **Center**: For steep power-law dispersal distributions with finite variance, large jumps broaden the spatial and temporal ranges over which lineages coalesce. Lineages at large separations $x \gg \overline{x}$ are occasionally able to coalesce at times comparable to $1/\mu$, while lineage dynamics at short distances are indistinguishable from thin-tailed dispersal. **Right:** For broad power-law dispersal distributions with infinite variance, large jumps are common. This allows for the rapid coalescence of lineages at both small and large distances but also lets lineages jump very far away from each other and avoid coalescing until a much later time set by the range size (not shown).

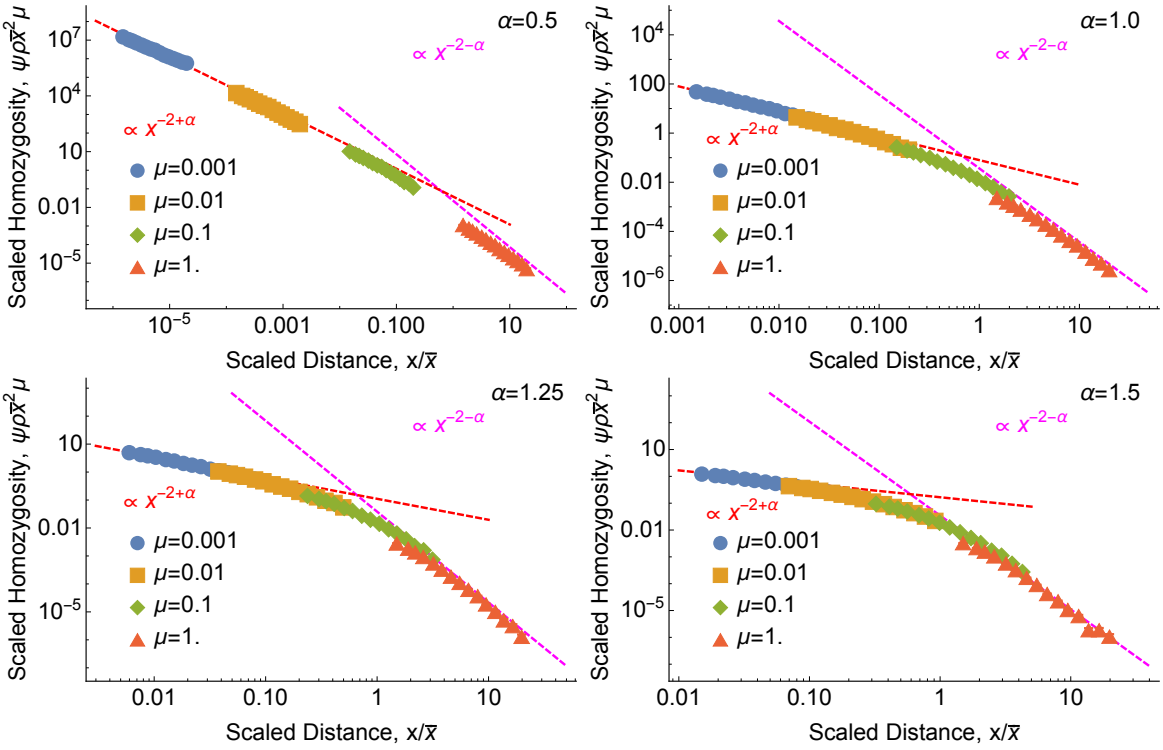

Figure 3.4: **Isolation by distance in two dimensions follows the same power law as dispersal.** Each panel shows the scaled probability of identity between a sampled pair of individuals, $\psi\rho\overline{x}^2\mu$, as a function of the scaled distance $x/\overline{x}$ between them. Points show discrete-space simulation results and magenta lines show the power law that emerges at large distances (3.4) (see (3.60) for prefactors). Red curves show the asymptotic behavior predicted at short distances by (3.5). For all plots, $\rho = 1$ and error bars show 68% percentile bootstrap confidence intervals (see Methods).

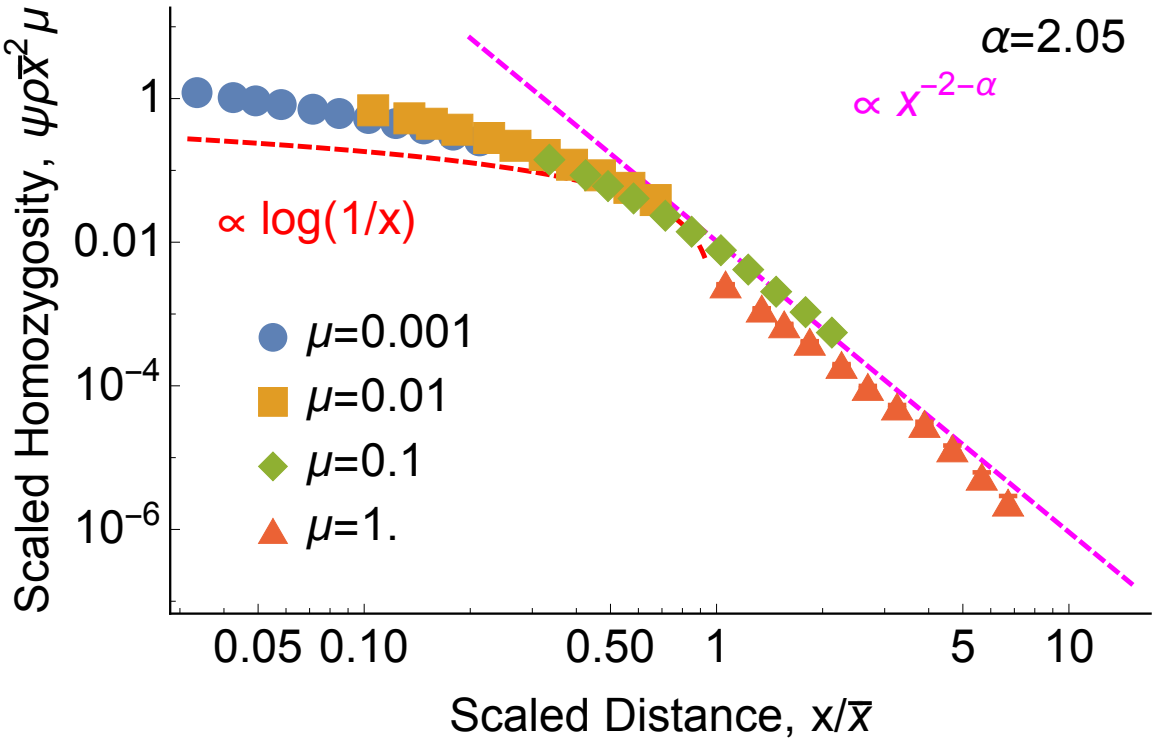

Figure 3.5: **Even for $\alpha > 2$, relatedness still follows the same power law as dispersal, rather than the diffusive prediction.** Points show discrete-space simulation results with $\rho = 1$. Since the dispersal kernel has finite variance, it approaches a diffusion, and at short distances $x \ll \overline{x}$ the probability of identity can be approximated by the continuous-space diffusive prediction (3.3) (red curve). But at long distances $x \gg \overline{x}$, relatedness is driven by rare long-range jumps and therefore has the same power-law tail as dispersal, (3.61).

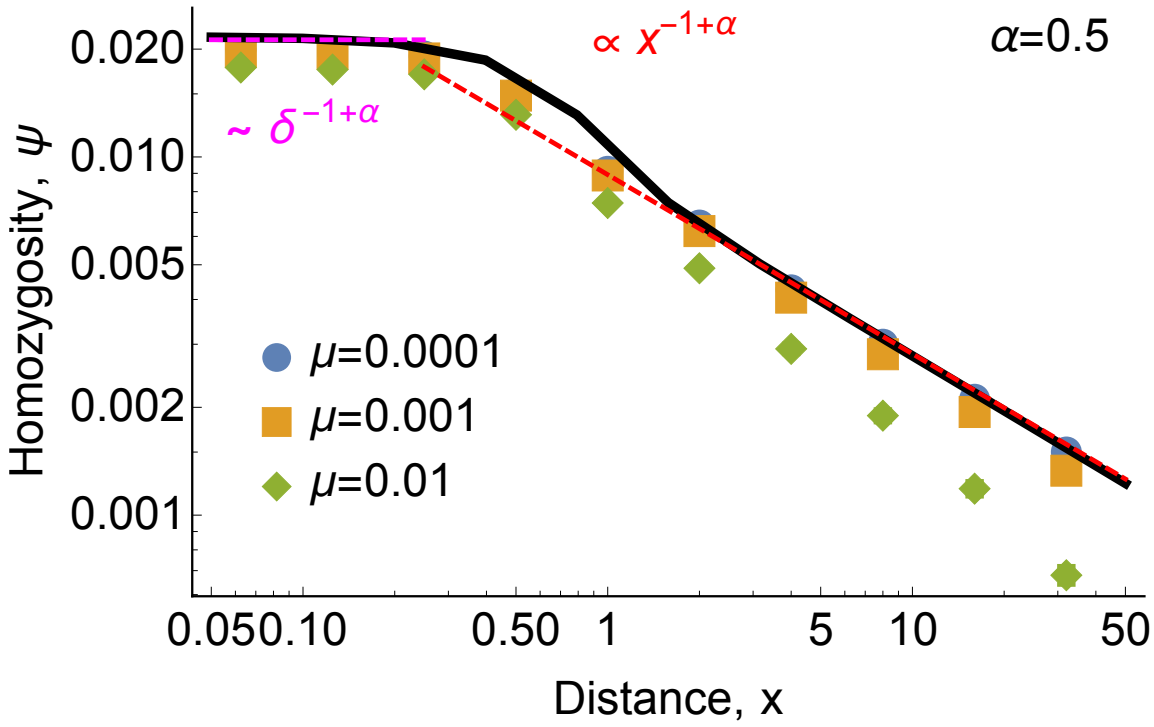

Figure 3.6: **For very heavy-tailed dispersal, $\alpha < 1$, relatedness at short distances is independent of mutation rate.** Nearby lineages at $x \ll \overline{x}$ either coalesce quickly and are identical, or jump very far away from each other and never coalesce. Points show continuous-space simulation results, and red and magenta lines show the asymptotic predictions of (3.40) and (3.39), respectively. The black curve shows a numerical solution of $\psi(x)$ calculated from (3.28) with $\mu = 10^{-4}$. $\rho = 100$ in all plots, and data with $\rho = 10$ and $\rho = 1$ (not shown) yield indistinguishable plots.

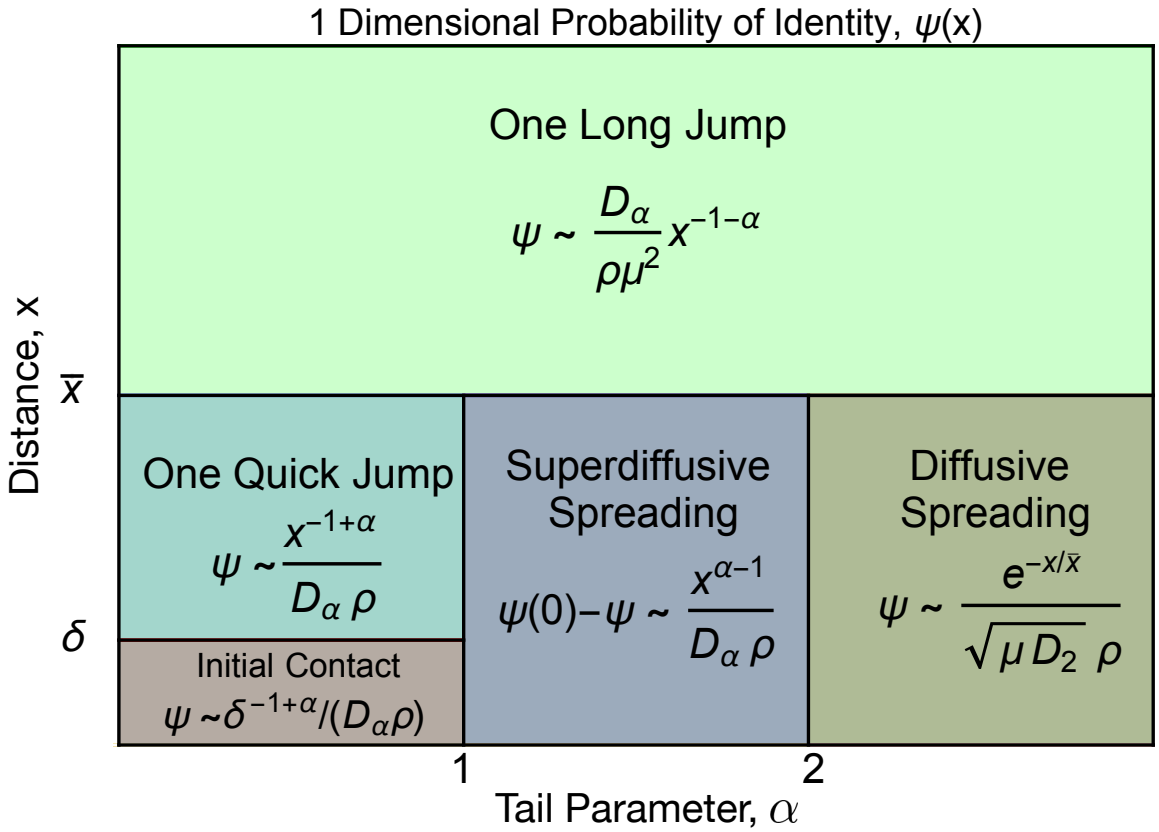

Figure 3.7: **For power-law dispersal, the form of isolation by distance in one dimension is universal at long distances and varies at short distance.** Approximate form for the probability of identity as a function of distance, $\psi(x)$, for different dispersal kernels α. Different regimes of the parameter space are separated by solid lines and labelled by their qualitative dynamics. Coalescence for distant pairs, $x \gg \overline{x}$, where $\overline{x} = (D_\alpha/\mu)^{1/\alpha}$ is the characteristic length scale of identity, occurs via one long jump for all α, leading to the power-law scaling at large distances predicted by (3.31). Coalescence for nearby pairs, $x \ll \overline{x}$, depends on the value of α considered. For $\alpha > 2$, the motion of lineages across short distances is diffusive and ψ scales exponentially, as shown in (3.38). For $1 < \alpha < 2$, short distances are covered via many small jumps, but lineages spread faster than they would under diffusion, leading to the broader scaling found in (3.37). For $\alpha < 1$, even short distances are covered by one quick jump, leading to the power law shown in (3.40). Lineages that do not coalesce quickly (at $t \ll 1/\mu$) will likely never coalesce, and probability of identity is limited by δ, rather than μ, as shown in (3.39). We use "$\sim$" to denote proportionality in the limit of large population density where $\psi(0) \ll 1$.

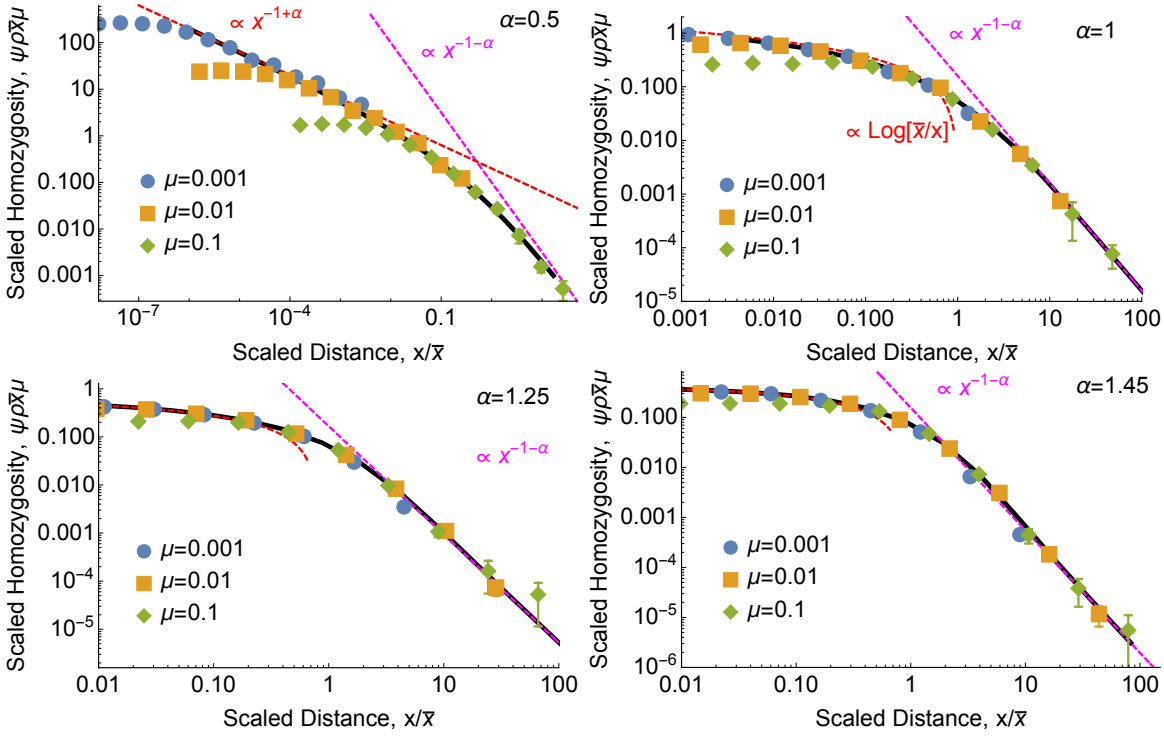

Figure 3.8: **Isolation by distance in one dimension follows the same power law as dispersal.** Each panel shows the scaled probability of identity between a sampled pair of individuals, $\psi\rho\overline{x}\mu$, as a function of the scaled distance $x/\overline{x}$ between them. Points show simulation results, black curves show numerical solutions of $\psi(x)$ calculated from (3.28) with $\delta = 0$ and $1 - \psi(0)$ set to 1, and magenta lines show the power law that emerges at large distances (3.31). Red curves show the asymptotic behavior predicted at short distances by (3.40) ($\alpha < 1$), (3.37) ($1 < \alpha < 2$), and (3.43) ($\alpha = 1$). For all plots, error bars show 68% percentile bootstrap confidence intervals (see Methods). $\rho = 100$ in all plots, and data with $\rho = 10$ and $\rho = 1$ (not shown) yield indistinguishable plots.

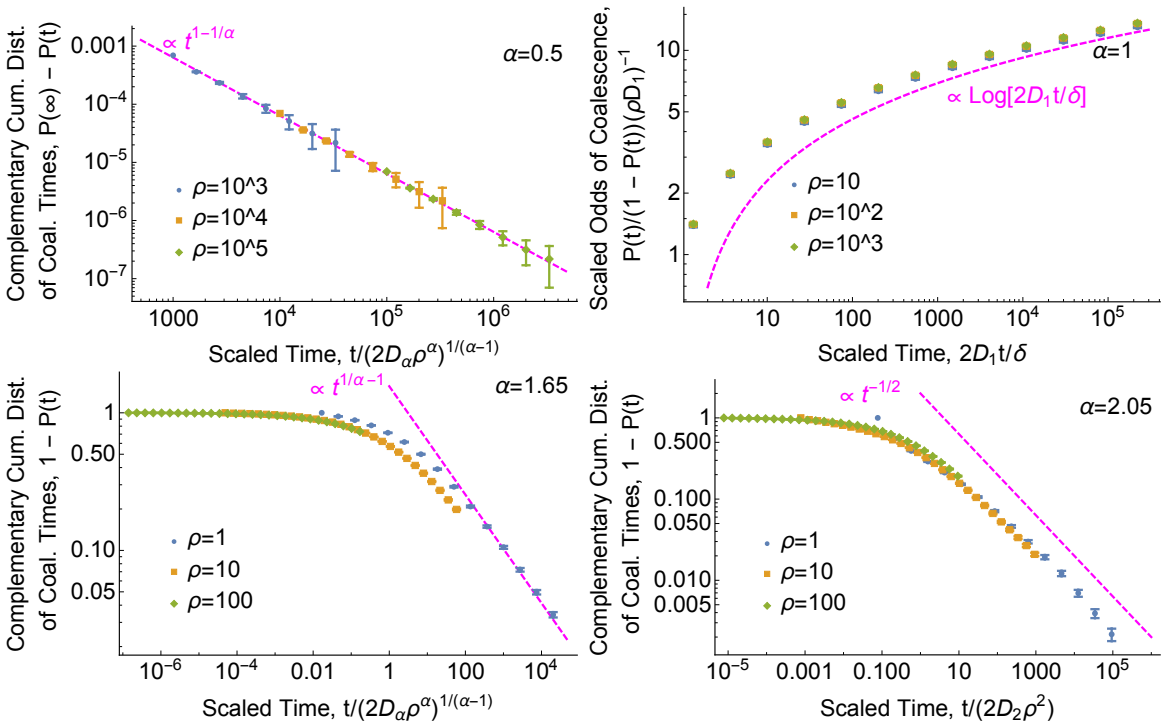

Figure 3.9: **The distribution of coalescence times has a power-law tail.** Points show one-dimensional simulation results. Dashed magenta curves show the asymptotic predictions (in order of increasing α) (3.65), (3.66), (3.68), and (3.70). Time is scaled to dimensionless units. See Simulation Methods section for D_α values. We show statistics based on the cumulative distribution $P(t)$ rather than the density $p(t)$ because simulation estimates for the latter are very noisy. **Top left: for $\alpha < 1$ in one dimension, the distribution of coalescence times is proportional to the probability of lineages being nearby,** $K(0|t) \propto t^{1-1/\alpha}$**.** Plot shows $P(\infty) - P(t)$ rather than $1 - P(t)$ because lineages can disperse infinitely far away from each other and avoid coalescing entirely, i.e., $P(\infty) < 1$. We use the simulated value of $P(t = 10^6)$ to approximate $P(\infty)$. This empirical value deviates from the continuous-time prediction (3.39) by $\approx 30\%$ due to differences in the amount of coalescence in the first few generations (see "Breakdown of models at small scales"). **Top right: the distribution of coalescence times has a logarithmic tail for $\alpha = 1$ in one dimension.** In this marginal case, lineages do eventually coalesce even in infinite ranges, but can take extremely long to do so. **Bottom left: for $1 < \alpha < 2$, the distribution of coalescence times in one dimension decays more quickly than the probability of lineages being nearby.** The coalescence time distribution has a power-law tail, $p(t|x) \propto t^{1/\alpha-2}$. This deviation from the scaling of the dispersal kernel at long times is due to the high probability of previous coalescence events. **Bottom right: for $\alpha > 2$, the coalescence time distribution may approach the diffusive limit.** The scaling of $1 - P$ appears to be close to that of the diffusive prediction, (3.70), but there is at least a difference in prefactor, perhaps again due to different probabilities of coalescence at very recent times. Present-day separation x was set to zero for all simulation results shown.

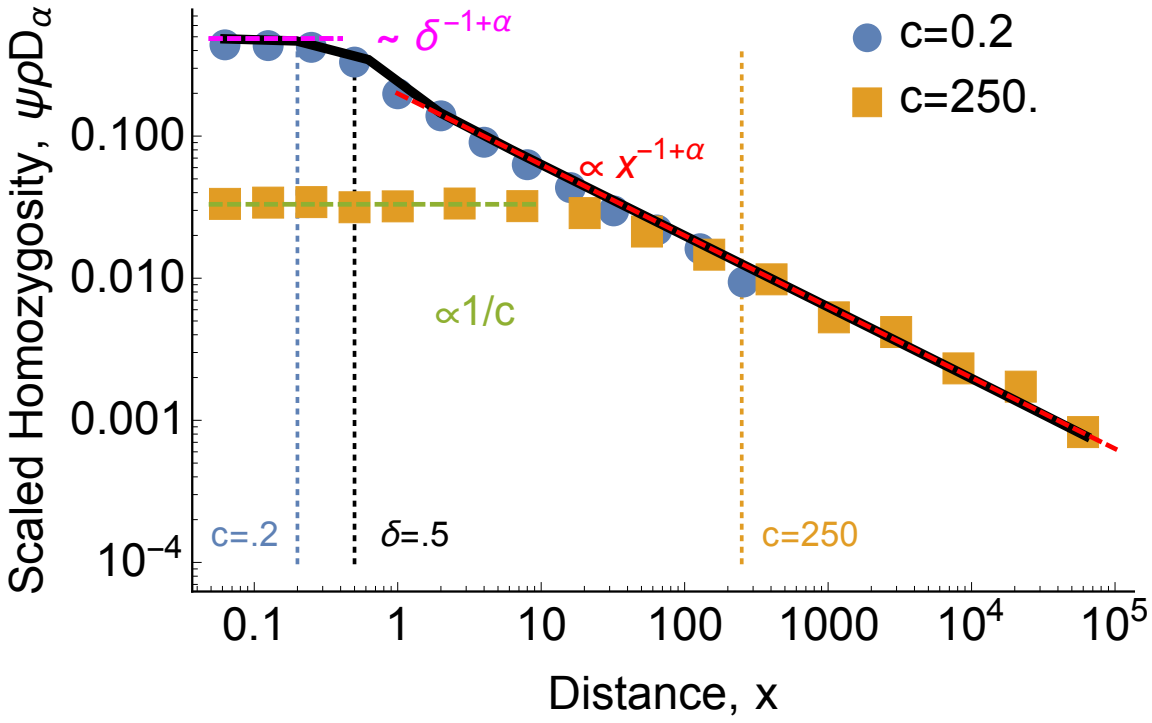

Figure 3.10: **For very heavy-tailed dispersal, $\alpha < 1$, continuous-time and discrete-time models differ at short distances.** Scaled probability of identity ψ as a function of distance x for $\alpha = 0.5$, $\delta = 0.5$, and $\rho = 100$. Points show discrete-time simulation results. For the continuous-time model, the black curve shows the result of numerically integrating (3.28), while the dashed red and magenta lines show the asymptotic approximations (3.40) and (3.39), respectively. The continuous-time model predicts that ψ should only plateau within the coalescence distance δ, but for distance between δ and the typical single-generation dispersal distance c, the change in ψ is driven by the probability of coalescing at $0 < t \ll 1$. In the discrete-time model, these lineages have to wait until $t = 1$ to coalesce, leading to a lower, broader plateau, given by (3.72) (dashed green line). This discrepancy only exists for $\delta < x \ll c$, i.e., if $c < \delta$ then the discrete-time and continuous-time models agree (blue points).

Chapter 4

Inferring Power-law Dispersal from Patterns of Isolation by Distance

4.1 Introduction

Direct measurement of dispersal in spatially structured populations is often a prohibitively difficult task, with tagging and tracking typically only being feasible for select populations under extended observation. As a result of these challenges, indirect inferences of dispersal from readily available genomic data have become popular, with a large class of methods being developed for populations of varying demographic and spatial structure [20, 38, 140, 141, 165]. For demographically stable populations in continuous space, dispersal and neighborhood size are often inferred from the rate at which genetic identity decreases with the spatial separation between sampled pairs. The fact that limited dispersal leads to the decay of genetic correlations with increasing separation is known as isolation by distance [30, 91, 137, 146, 169].

If we compare sufficiently long blocks of sequence, the rates of mutation and recombination will capture short timescales over which the demography and dispersal are likely constant, and the effects of selection can be ignored [17, 76, 131]. As long as the blocks are not too large, the timescales involved will not reflect the recent pedigree or reproduction process either, and thus the local details of coalescence can be ignored as well [134]. For these moderately large blocks, the resulting patterns of

isolation by distance can be described by classic diffusive models that are governed by just three parameters: the population density, dispersal constant and mutation rate [15, 89, 131, 149]. Recently developed continuous-space models derived using more rigorous assumptions make similar predictions and are again described by only a few parameters [19]. Using these models, both the fraction of identical sequence blocks of a fixed size and the distribution of blocks of varying size can be used to infer dispersal. Dispersal is assumed to be thin-tailed, and the dynamics over many generations can be characterized by a single parameter, the dispersal constant D [15, 134]. Pairwise genetic similarity is predicted to decay logarithmically at short distances and exponentially at large distances, with a decay rate of $\sqrt{\mu/D}$, where μ is the mutation rate.

While the assumption of diffusive dispersal is often reasonable, many populations appear to instead display signatures of a heavy-tailed dispersal distribution that decays very slowly with distance [2, 7, 11, 34, 45, 46, 48, 63, 159, 167]. There are many other populations for which little about dispersal is known, and we would like to be able infer dispersal from genetic data in a way that can distinguish thin- and heavy-tailed kernels [150]. A heavy-tailed kernel allows for dispersal events that are orders of magnitude larger than the "typical" jump size given by the standard deviation of the distribution, and for extremely heavy tails the standard deviation itself becomes infinite. The presence of jumps that span a huge range of sizes leads to a qualitatively different form of motion than the diffusion that is typical of thin-tailed dispersal distributions [108].

In this work, we use our recently developed model of isolation by distance in populations with a heavy-tailed (power-law) dispersal kernel [150] to generalize existing methods of dispersal inference. We also use our previous results on the probability of identity and the distribution of coalescence times for pairs undergoing heavy-tailed dispersal to predict the size and number of long shared sequence blocks between pairs. We develop inference schemes based on both the probability of identity and the number of long shared sequence blocks between pairs and apply these methods to both real and simulated sequence data. We test our methods on simulated data and data from Florida scrub-jays at Archbold Biological Station, and apply them to data from

European humans in the POPRES dataset [2, 117, 131].

4.2 Model

We consider two individuals sampled in the present a distance x apart, and trace their lineages back in time. We assume individuals disperse a random distance every generation according to an isotropic dispersal kernel, and that the range is infinite in size. We consider dispersal kernels that are normal, as well as kernels with a power-law tail [150]. Our primary inference scheme will assume that lineages follow Lévy flights, a model of dispersal in which trajectories can include rare long-range jumps [86, 107, 108]. The dispersal kernel of a Lévy flight follows a stable distribution [174]:

$$K_1(y|t) = \frac{1}{2\pi}\int_0^\infty dk\, k J_0(ky)\exp\left(-D_\alpha t k^\alpha\right), \tag{4.1}$$

where J_0 is the zeroth Bessel function and t is the number of generations in the past. The generalized diffusion constant D_α determines the rate at which lineages spread: just as the mean squared displacement is set by the diffusion constant for Brownian motion, Lévy flights generalize this relation, with $\langle x^\alpha\rangle = D_\alpha t$ for $\alpha \leq 2$. For $\alpha = 2$, Lévy flights reduce to ordinary diffusion with a normal dispersal distribution. For $\alpha < 2$ however, lineages can take long jumps due to the tail of the kernel following a power law proportional to $y^{-2-\alpha}$, and the characteristic rate of spread, $(D_\alpha t)^{1/\alpha}$, is faster than that found for diffusion. We also consider finite variance t-distribution dispersal kernels with $\alpha > 2$ (see Methods). Even for these steep power laws, the tail of the kernel is proportional to $y^{-2-\alpha}$ [150].

When the two lineages meet in space, they coalesce at a rate proportional to $1/\rho$, where ρ is the population density. We assume that ρ is a constant, which is equivalent to assuming local density-dependent regulation of reproduction [59]. In two dimensions, two lineages of infinitesimal size will never be at exactly the same position [111], and we assume there is a local scale δ below which nearby lineages are able to coalesce. Our continuous-space model can be interpreted as the limit of stepping-stone model in which individuals occupy discrete demes [15]: every small

patch of space has approximately $\rho\delta^2$ individuals, and the rate of coalescence for nearby pairs is roughly $1/(\delta^2\rho)$. For $x \gg \delta$, the value of δ will not affect patterns of genetic diversity, and we assume all sampled pairs used for inference meet this criteria.

The lineages each acquire neutral mutations at rate μ. The probability of identity in state for a pair sampled at separation x is therefore [15]:

$$\psi(x) = E\left[e^{-2\mu T}|x\right], \tag{4.2}$$

where the expectation is over the time T to their most recent common ancestor. Roughly speaking, the basic intuition is that the sampled pair will be identical if their lineages coalesce within approximately the past $1/\mu$ generations. In this time, they will disperse a typical distance of order $\overline{x} \equiv (D_\alpha/\mu)^{1/\alpha}$, so this is the key length scale over which identity decays: pairs separated by $x \ll \overline{x}$ should be relatively closely related, while identity between pairs separated by $x \gg \overline{x}$ should be rare [150]. For Lévy flights with $\alpha \leq 2$, there is no closed-form expression for $\psi(x)$. Identity in state for $x \gg \delta$ is instead given by the following integral (see Methods):

$$\psi(x) = \frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{1+\kappa^\alpha}. \tag{4.3}$$

In sexual populations, the effective mutation rate is often quite small compared to the rate of recombination, and even distantly related individuals are identical by state at most sites along their genomes. While identity in state hardly varies at all within these populations, identity by descent, defined as the proportion of the genome inherited from the same ancestors, can still vary considerably with the distance between sampled pairs [2]. While our model was originally developed to predict identity in state, in the $\mu = 0$ limit $\psi(x)$ reduces to the probability of coalescence for sampled pairs. For $\alpha \geq 2$, $\psi(x) = 1$ when $\mu = 0$, and the model is trivial. For $\alpha < 2$ however, our infinite-range model shows a drastic separation of coalescence timescales [150]: pairs either coalesce quickly, and are effectively identical by descent everywhere along the genome, or wander away from each other and never coalesce. In real populations on finite ranges, "never coalescing" corresponds to coalescing after the mean coalescence time for a random pair. In this case, $\psi(x)$ predicts the probability of very recent

shared ancestry, and thus identity by descent, for sampled pairs [150]:

$$\psi_{IBD}(x) \propto \frac{1-\psi_{IBD}(0)}{\rho D_\alpha x^{2-\alpha}} \qquad \text{for } \alpha < 2,\ \mu = 0. \tag{4.4}$$

In addition to the probability of identity, we can consider the number of long shared sequence blocks between pairs [17]. Assuming that long blocks are broken up due to recombination and that the mutation rate is negligibly small in comparison, the block density for N_c chromosomes of total map length G is given by the following expectation [134]:

$$E[N_L(x)] = E\left[N_c\left(4\left(\frac{G}{N_c}-L\right)T^2+4T\right)e^{-2LT}|x\right], \tag{4.5}$$

where L is the length of the block in Morgans and we assume all N_c chromosomes are the same length. The intuition here is that the size of a block determines the rate at which we expect recombination to break it up. Pairs that coalesce within approximately the past $1/L$ generations should possess many blocks of length L, while pairs that coalesce after $1/L$ should have very few blocks of this size. Since lineages disperse a typical distance of $(D_\alpha/L)^{1/\alpha}$ in $1/L$ generations, this is the length scale over we which we expect the number of long shared sequence blocks to decay. For Lévy flights with $\alpha \leq 2$, there is again no closed form expression for the block density. For $x \gg \delta$, the following integral expression is valid for large blocks and very long chromosomes (see Methods):

$$E[N_L(x)] \approx \frac{G}{2\pi\rho c^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/c)}{\left(L+\kappa^\alpha\right)^3}, \tag{4.6}$$

where $c \equiv (D_\alpha)^{1/\alpha}$ is the typical distance an individual disperses in a single generation.

4.3 Inference scheme

We use both the pairwise probability of identity, $\psi(x)$, and the number of long shared sequence blocks, $E[N_L(x)]$, to infer dispersal from spatially labelled genomic data. For both quantities, we use maximum likelihood estimation to find the dispersal distribution of best fit from a parameterized family of distributions, namely Lévy alpha-stable distributions with $\alpha \leq 2$, or t-distributions with $\alpha > 2$.

We use (4.3) for $\psi(x)$ to infer dispersal from identity in state, treating α, $\overline{x}$ and the prefactor $\rho\mu/(1-\psi(0))$ as independent parameters. We assume no linkage between sites and treat the number of independent sites per pair, N_{sites}, as a fixed parameter that must be specified before estimating dispersal. All sites are modelled as tracts of sequence with the same base-pair length, and the effects of recombination within a site are assumed to be negligible. Identity for each independent site is assumed to be binary (1 or 0), and is treated as a Bernoulli process with mean $\psi(x)$. The mean probability of identity across all sites along the genome is also assumed to be $\psi(x)$, with homozygosity for pairs following a binomial distribution with N_{sites} trials. The mean probability of identity for all pairs at a given separation distance also follows a binomial distribution with mean $\psi(x)$, and we denote the total number of sites in a given distance bin (N_{sites} times the number of pairs) as $N(x_i)$. The resulting likelihood function for $\psi(x)$ is then:

$$L_\psi(\alpha, \overline{x}, \frac{\rho\mu}{1-\psi(0)}) = \prod_{x_i} N(x_i)! \frac{\psi(x_i)^{N(x_i)\hat{\psi}(x_i)}(1-\psi(x_i))^{N(x_i)\left(1-\hat{\psi}(x_i)\right)}}{\left(N(x_i)\hat{\psi}(x_i)\right)!\left(N(x_i)\left(1-\hat{\psi}(x_i)\right)\right)!}, \tag{4.7}$$

where $\hat{\psi}(x_i)$ is the average identity in state for pairs in a given distance bin, and the product is over all bins. The likelihood function in (4.7) has three free parameters: α, $\overline{x}$, and $\rho\mu/\left(1-\psi(0)\right)$.

To reduce the number of free parameters in the inference problem (see Methods), we assume that $N(x_i)$ in each bin is large and approximate the binomial distribution for $\psi(x)$ as a normal distribution. Our model for $\psi(x)$ also assumes that $\psi(x) \ll 1$ [150], and we can thus approximate the variance in identity as $\psi(x)/N(x_i)$. The resulting log-likelihood function is:

$$LL_\psi(\alpha, \overline{x}) = -\frac{1}{2}\sum_{x_i}\left[\frac{N(x_i)}{\psi(x_i)}\left(\psi(x_i)-\hat{\psi}(x_i)\right)^2 + \log\left(\frac{2\pi\psi(x_i)}{N(x_i)}\right)\right]. \tag{4.8}$$

While (4.8) technically has three free parameters, we can reduce this to two: the optimal value of $\rho\mu/\left(1-\psi(0)\right)$ can be determined analytically for fixed α and $\overline{x}$ (see Methods). We use the procedures described in Optimization methods to find the maximum likelihood estimates of the parameters and their confidence intervals.

For our model of identity by descent, $\mu = 0$ and $x \ll \overline{x}$ for all sampled pairs, and $\psi_{IBD}(x)$ takes the simple asymptotic form given in (4.4). In this limit, the three parameters α, $\rho\mu/\left(1-\psi_{IBD}(0)\right)$ and $\overline{x}$ reduce to just two: α and $\rho\mu\overline{x}^{\alpha}/\left(1-\psi_{IBD}(0)\right) \equiv \rho D_{\alpha}/\left(1-\psi_{IBD}(0)\right)$. Varying $\overline{x}$ in this regime changes the corresponding optimal value of $\rho\mu/\left(1-\psi_{IBD}(0)\right)$ in a way that keeps $\psi_{IBD}(x)$ and the log-likelihood constant, leading to ridges and degenerate optima in (4.8). We thus develop a separate dispersal inference scheme for $\psi_{IBD}(x)$, which we apply to data from the Florida scrub-jay population at Archbold Biological Station [2].

While the inference method based on identity in state assumed a fixed number of independent sites along the genome, this is likely not accurate for the available scrub-jay data. The large number of sampled SNPs should lead to significant correlations in SNP trajectories over the relatively short timescales of interest. Even if the alleles demonstrate linkage equilibrium, the number of meioses over 5 to 10 generations will not generate enough distinct spatial trajectories for each SNP to disperse independently. Rather than try to model the noise in identity by descent from first principles, we bin the scrub-jay pairs into 10 meter distance bins and use the empirical variation within each bin to estimate the noise in mean identity. Concretely, we use a weighted least squares expression based on binned mean identity as our log-likelihood function:

$$LL_{\psi_{IBD}}(\alpha) = -\frac{1}{2}\sum_{x_i}\left(\psi_{IBD}(x_i) - \hat{\psi}_{IBD}(x_i)\right)^2/\sigma^2_{SEOM}(x_i), \tag{4.9}$$

where $\sigma_{SEOM}(x_i)$ is the standard error of the mean in a given distance bin. While $\psi_{IBD}(x)$ has two independent parameters, α and $\rho D_{\alpha}/\left(1-\psi_{IBD}(0)\right)$, we can use the simple form of the likelihood function to find the optimal value of $\rho D_{\alpha}/(1-\psi_{IBD}(0))$ analytically for any fixed α (see Methods). We again use the procedures discussed in Optimization methods to find the maximum likelihood estimates for α and $\rho D_{\alpha}/(1-\psi_{IBD}(0))$ and the associated confidence intervals.

When inferring dispersal from long shared sequence blocks, we use the expressions in Methods that have the same basic form as (4.6), treating α, $c \equiv (D_{\alpha})^{1/\alpha}$ and ρ as independent parameters. We assume that all detected blocks are completely independent segments of the genome with their own trajectories and coalescence

times. We thus model the number of blocks for pairs at a given separation as a Poisson process with mean $E[N_L(x)]\Delta_L$, where Δ_L is the "bin width" that denotes the range of block sizes aggregated into a single block count. For parameter estimates, we again fit data using a composite likelihood approach, binning both distance and map length. The resulting log-likelihood function again has three parameters, α, c and ρ:

$$LL_{N_L}(\alpha, c) = \sum_{x_i} N_{\text{pairs}}(x_i) \sum_{L_i} \left[\log(E[N_L(x_i)]\Delta_L)\hat{N}_L(x_i)\Delta_L - E[N_L(x_i)]\Delta_L \right], \tag{4.10}$$

where $N_{\text{pairs}}(x_i)$ is the total number of pairs at distance x_i, and $\hat{N}_L(x_i)\Delta_L$ is the average number of blocks per pair in a given bin. Note that we have omitted the constant factorial term in the log-likelihood for simplicity. We can again reduce the three free parameters in (4.10) to two by finding the optimal ρ value analytically for fixed α and c (see Methods). Just as for inference using identity, we rely on the procedures discussed in Optimization methods to find maximum likelihood estimates and confidence intervals for α, c, and ρ.

4.4 Application to simulated data

We used the backwards-time simulations described in Methods to calculate isolation by distance patterns for a two-dimensional stepping-stone model on a finite range. For both $\psi(x)$ and $E[N_L(x)]$, we varied α in our simulations between 0.5 and 2 and fit the resulting data to the Lévy flight model described above. For $E[N_L(x)]$, we also simulate steep, finite-variance power laws with $\alpha > 2$, which we fit to a distinct t-distribution model (see Methods).

For both $\psi(x)$ and $E[N_L(x)]$, we find good agreement between the "true" input parameters of the discrete-space simulations and the estimated parameters found via our continuous-space inference methods when $\alpha \leq 2$. For all simulated data fit with the site-based Lévy flight model, we set the number of total independent sites at each separation, $N(x_i)$, to 10^6. For the block-based Lévy flight model, 10^5 independent pairs were simulated for each separation distance and block length. While these sam-

ple sizes are large, they are not unreasonable. In the data from natural populations analyzed in the following sections, we find a large number of total pairs and thus expect an even larger number of total sites: more than 10^5 pairs were sampled from the scrub-jay population and over 10^6 pairs are present in the POPRES dataset [2, 131].

While we see a slight bias in estimates of ρ for $E[N_L(x)]$, we believe this is an artifact of our discrete-space simulation method (see Methods). True values of α are contained within the 95 percent confidence intervals for all simulations with $\alpha \leq 2$, and estimates of the length scale ($\overline{x}$ or c) show a similar degree of accuracy (Fig. 4.5 and Fig. 4.7). Likelihood heatmaps for $\psi(x)$ show a slow variation of log-likelihood with both parameters, while the variation in log-likelihood is much more drastic for $E[N_L(x)]$ (Fig. 4.6 and Fig. 4.8). We attribute this difference to the fact that there are far more shared blocks per distance bin than identical sites with our choice of simulation parameters. Focusing on a single block size, we see that the number of blocks per pair and cM is of order 10^{-2} at the shortest distances, and thus the total number of blocks within a single bin is of order 10^3. For $\psi(x)$, the total number of identical sites is $N_{\text{sites}}\psi(x)$, which is of order 10^2 in a single bin. Given that the signal to noise ratio is proportional to the number of identical sites or blocks per bin, it is not surprising that this difference affects the precision of inference. If we had instead considered a smaller range of blocks on shorter genomes, this effect may well have been reversed.

For $\psi(x)$, we see that the best fit curves for $\alpha = 0.5$ and $\alpha = 1.0$ are pure power laws and that the corresponding likelihood heatmaps have long, horizontal ridges that extend to large values of $\overline{x}$. As explained in the previous section, this is due to the asymptotic form of $\psi(x)$ for $x \ll \overline{x}$ [150]. We see another pure power law in the plot of $E[N_L(x)]$ for $\alpha = 0.5$, but no corresponding horizontal ridge in the likelihood heatmap. This is because the block density still depends on $c \equiv (D_\alpha)^{1/a}$ at short distances $x \ll c$.

4.5 Application to Florida scrub-jay data

We used our predictions for $\psi_{IBD}(x)$ to infer dispersal for the population of Florida scrub-jays at Archbold Biological Station. This dataset, provided by Aguillon et al. [2], contains 130,618 spatially labelled pairs and measurements of pairwise identity by descent based on 7,483 autosomal SNPs at which the jays were genotyped. Identity by descent for a pair is defined as the proportion of the genome inherited from the same ancestors. This quantity is estimated by comparing identity in state for a given pair to the average level of identity in state for all pairs in the population. Pairs with above-average levels of identity in state likely share common ancestry in the more recent past than pairs with average or below average relatedness, thus allowing the inference of kinship from genomic data. This raw measure of above-average identity is further transformed by PLINK to obtain an estimate of identity by descent that is always between 0 and 1 [130].

Maximum likelihood parameter estimates were found for both the full dataset and a truncated dataset representing pairs with separations between 200 meters and 2 km. Truncating the data allows us to exclude short distances at which the local details of coalescence and dispersal may deviate from the assumptions made by our simple model. It also allows us to exclude large distances near the edge of the range in which jays were sampled. While the sampling bounds set by Archbold Biological Station do not represent true bounds on the population range for the jays, they do determine the baseline level of identity in state used to determine identity by descent for all pairs. If this baseline is above the true average identity of the population, identity by descent at these large distances will go undetected due to the inherent sampling limitations. For the full dataset, we found an α value of 1.69 and a value of 10.01 for $\rho D_\alpha / (1 - \psi(0))$. For the truncated dataset, we find an α value of 1.46 and a $\rho D_\alpha / (1 - \psi(0))$ value of 5.35. Plots of the data and fits are displayed in Fig. 4.1 and Fig. 4.2.

In addition to indirect inference of dispersal via isolation by distance, we can leverage the direct measurements of dispersal available for Florida scrub-jays to estimate the dispersal kernel. We use the data provided by Aguillon et al. [2] for the radial

distribution of dispersal distances ($2\pi y$ times our definition of the dispersal kernel in (4.1)) and fit this one-dimensional (radial) histogram to a one-dimensional stable distribution (Fig. 4.1). For the full and truncated datasets we find best-fit α values of 1.29 and 1.60 using this more direct method.

4.6 Application to European POPRES data

We applied our $E[N_L(x)]$ based inference method to data provided by Ralph and Coop [131] on the number of long shared sequence blocks between pairs of humans within the Population Reference Sample (POPRES) dataset [117]. We analyzed this data with both our Lévy flight and t-distribution dispersal inference codes and considered both the full dataset and the subset of pairs within Eastern Europe. Our criteria for selecting pairs within Eastern Europe is identical to that used by Ringbauer et al. [134]. For both the full and Eastern European datasets, we inferred dispersal from long blocks above a certain cutoff size. We set the maximum block size to 20 cM, varied the minimum block length between 4, 6, and 8 cM and found that the choice of minimum block length can have a significant effect on the estimated parameters. We binned both separation distance and block size during inference, with distances being binned in 10 km increments and blocks being binned every 0.25 cM for the Lévy flight code (1 cM for the t-distribution code). We also enforce a minimum ρ value of 10^{-3} km^{-2} during inference.

While we have analyzed the number of shared blocks for all European pairs, we note that our model fails to explain the *lack* of isolation by distance for small blocks in the full dataset (Fig. 4.3). We should thus be cautious about drawing biological conclusions from this analysis. Indeed, our assumption of a continuous spatial range with no boundaries or fragmentation is clearly violated when considering all of Europe. As noted by Ringbauer et al. [134], the assumption of a continuous range without fragments or boundaries is better suited to Eastern Europe, and we believe that our model and analysis should be most relevant for this subset of the POPRES data.

In our analysis of both the full POPRES dataset and the Eastern European subset, we find that dispersal is best described by a t-distribution when a 4 cM block

length cutoff is used. Model selection (AIC) favors the t-distribution model over both the Lévy-stable and diffusive models of isolation by distance. While a direct comparison of the t-distribution and stable distribution models is not feasible (due to the difference in bin sizes used during inference), the diffusive model was fit to data using both choices of bin size. The *difference* in AIC scores is significantly larger when comparing the t-distribution model to the diffusive model for both the full and Eastern European datasets. While the steep power-law model that assumes t-distribution dispersal outperforms the broad power-law model with Lévy-stable dispersal, the best-fit α value for Eastern Europe is pinned to the model's lower bound of $\alpha = 2.01$, which suggests that a broader t-distribution may outperform both steep t-distributions and Lévy-stable kernels. Our spline-based numerical methods fail for these broad t-distributions, and fitting these kernels is thus beyond the scope of the present work.

When a larger minimum cutoff of 6 cM or 8 cM is used, we find that models including long-range (power-law) dispersal again provide a better description of the data than the classic diffusive model. A 6 cM cutoff produces best-fit α values of 0.93 and 1.05 for the full and Eastern European datasets, whereas an 8 cM cutoff yields α values of 0.69 and 1.09. While our long-range model outperforms the classic short-range model, the fit for larger block cutoffs is less accurate than that found for the 4 cM cutoff, and the amount of noise present in the count data for large blocks makes the exact form of the isolation by distance pattern difficult to discern. Much of the noise in count data for larger blocks can be attributed to the small number of blocks detected at these sizes. More data would be necessary to draw definitive conclusions about the form of isolation by distance and the presence (or absence) of long-range dispersal on the timescales associated with these larger blocks.

Surprisingly, the estimated value of ρ for the Eastern European subset differs by an order of magnitude from the estimate found by Ringbauer et al. [134] for the same (diffusive) model. We attribute this discrepancy to a difference in optimization procedures. A non-negligible number of false positives in the block data may also contribute to the difference in inferred parameters [131]. Maximum likelihood estimates for all Eastern European data between 4 and 20 cM can be found in

Fig. 4.4. All other maximum likelihood estimates, as well as heatmaps and AIC values, can be viewed at https://github.com/weissmanlab/Dispersal_Inference/blob/master/IBD_block_data/aggregated_popres_plots.pdf.

4.7 Discussion

Here we have extended the model developed in Smith and Weissman [150] of isolation by distance in populations with long-range (power-law) dispersal to predict the number of long shared sequence blocks between pairs. We also used these expressions for the number of shared blocks, as well as previously derived expressions for the probability of identity, to develop methods for inferring long-range dispersal from genomic data. We showed that these inference methods perform well for data generated from coalescent simulations of neutral populations, and then applied the methods to a population of Florida scrub-jays at Archbold Biological Station and to human samples found within the European POPRES dataset [2, 131].

For the Florida scrub-jay data, we detect strong genetic signatures of long-range dispersal and find that the observed pattern of isolation by distance is better described by a model with a power-law kernel than the classic diffusive model. Direct fits of empirical dispersal data similarly show that the dynamics of the scrub-jay population are better described by a model that includes long-range dispersal. Detecting long-range dispersal in a population where it is known to be present is a good first test for this new inference method, and passing this test provides evidence for the method's efficacy.

It is also interesting to note that long-range dispersal provides a mechanism for significant levels of genetic diversity even when the effective mutation rate is very small [150]. Lineages undergoing long-range migration can demonstrate an extreme separation of timescales in which pairs either coalesce very quickly or wander away from each other and avoid coalescing for many generations. In our infinite-range Lévy flight model, these wandering lineages literally never coalesce, and $\psi(x)$ is non-constant even for $\mu = 0$, as the number of pairs that coalesce before "wandering off"

decreases with separation. In real populations on finite-ranges, lineages that wander off do eventually coalesce, with the typical coalescence time set by the density of the population and the size of the entire range. While the effects of a finite range size make our infinite-range prediction for $\psi(x)$ an inaccurate model of identity in state, it still captures the probability of very recent shared ancestry, and can thus be used to model identity by descent. Aguillon et al. [2] showed that the pattern of isolation by distance found in Florida scrub-jays can be described by a modified Malécot model that assigns pairs a binary level of identity based on whether a long-range migration event occurs before coalescence. Our Lévy flight model is conceptually similar: when the mutation rate is very small, pairs that coalesce quickly are assigned a value of 1 and pairs that wander off are assigned a value of 0. While the model developed by Aguillon et al. [2] uses the known dispersal distribution up to 5 km to model short-range movements and treats long-range jumps out of Archbold Biological Station as an entirely separate process, here we model both short-range and long-range dispersal events with a single power-law kernel.

While long-range dispersal helps to explain the slow decay in identity with distance, the fact that the fits found by Aguillon et al. [2] are far better than those found here make it clear that incorporating the small-scale and sex-specific details of dispersal is necessary to explain scrub-jay isolation by distance patterns, and that classic models and the extensions developed here are too simple to provide a full description of real populations. While simplistic, these models can be fit to data without knowledge of the true dispersal distribution or pedigree, and can thus be used to infer dispersal in populations where it hasn't been measured directly. Even if the exact form of the dispersal distribution can't be inferred with such simple models, it may be possible to determine whether dispersal is long- or short-range.

Genomic signatures of long-range dispersal have also been detected for European humans in the POPRES dataset, but our findings should be seen as suggestive rather than definitive. Models of isolation by distance that include long-range dispersal may in fact provide a better description of the patterns of human ancestry within Europe, but further analysis (ideally with a larger number of sampled pairs) is needed before any conclusions can be drawn. Given the drastic changes in human travel over the

past 10 generations, it seems plausible that long-range movement of individuals may have left visible signatures in the genomes of present-day humans. Indeed, previous studies of current human dispersal suggest that a power-law kernel may be much more realistic than a diffusive or thin-tailed model [34]. It is also likely that the simple models discussed here fail to capture key aspects of human populations over recent timescales. Rapid population growth certainly invalidates the assumption of demographic equilibrium, and models that allow for a time-dependent population density may be more appropriate for this dataset [134].

One significant limitation of our inference method is that we must assume $\psi(x) \ll 1$. This is only a mild constraint when fitting to relative measures of relatedness like identity by descent, but it does rule out fitting absolute measures of relatedness like identity in state for a large number of populations. While the effective mutation rate can in principle be increased by comparing larger tracts of the genome [76], there are still many existing datasets that do not contain the data needed to perform such a rescaling. Given that the assumptions of continuous space and constant population density are also tenuous [59], generalizing the current scheme to one that is mathematically consistent and rigorous may also be worthwhile. A clear candidate for a more rigorous replacement is the spatial Λ-Fleming-Viot (SLFV) model, which is flexible enough to include both long-range dispersal and long-range reproduction events [19]. The SLFV model has proven effective for inference when dispersal is diffusive, and variants of the model that include power-law dispersal and reproduction have recently been developed [60, 73]. How well the SLFV model might perform relative to the dispersal inference method presented here is an open question.

4.8 Methods

4.8.1 Lévy flight model

The expected homozygosity is the Laplace transform of the distribution of coalescence times [15]:

$$\psi(x) = \int_0^\infty dt\, p(t|x) e^{-2\mu t}. \tag{4.11}$$

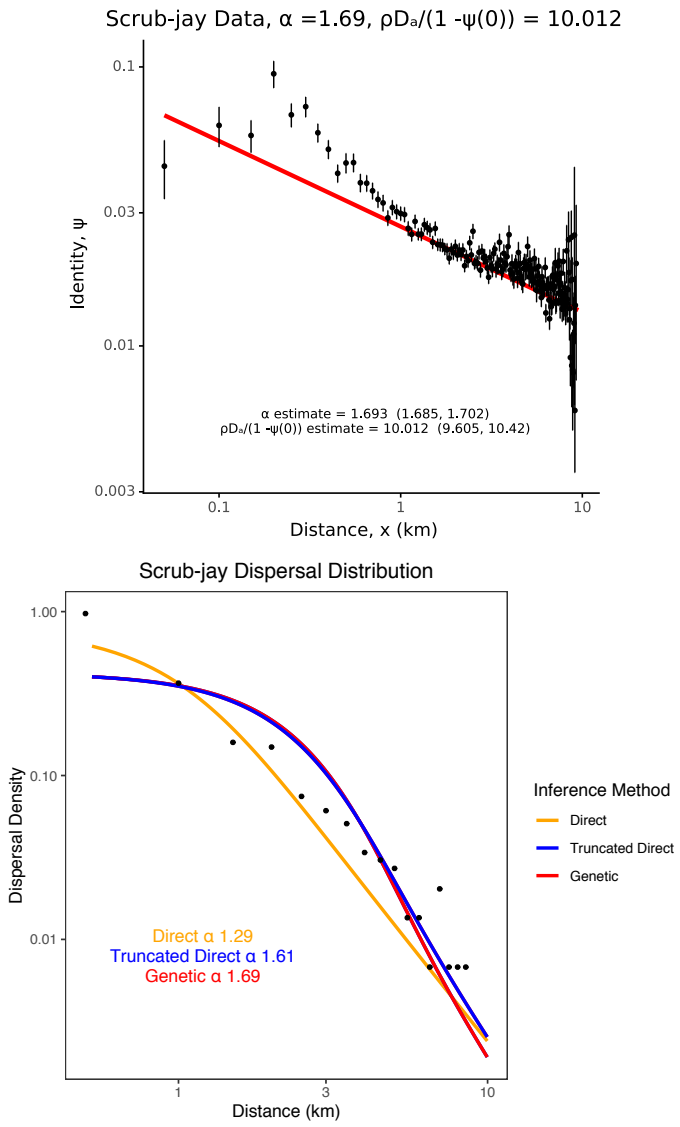

Figure 4.1: **Top: Florida scrub-jays display genetic signatures of heavy-tailed dispersal.** While we see systematic deviations between the data and our model at short distances, at larger distances identity by descent decays slowly and shows good agreement with our power-law model. Points show empirical measurements from the scrub-jay dataset. Error bars show the standard error of the mean in each 50 km distance bin. The red curve shows the maximum likelihood fit of the asymptotic Lévy flight model with $\mu = 0$. Parameter estimates for the model are inset in the figure. **Bottom: Lévy flight dispersal shows good agreement with the known scrub-jay dispersal distribution.** Points show the histogram of measured dispersal distances for jays at Archbold Biological Station. The curves are best-fit stable distributions using the full dispersal dataset (orange), the long-distance dispersal data at 500 meters or greater (blue), and the genomic data (red). We see that the inferred α using the dispersal data is reasonably close to the α value of 1.69 found via our genetic method. Using the long-distance dispersal data only, the inferred α using direct and genetic methods are nearly identical. Details of the inference procedures can be found in the text.

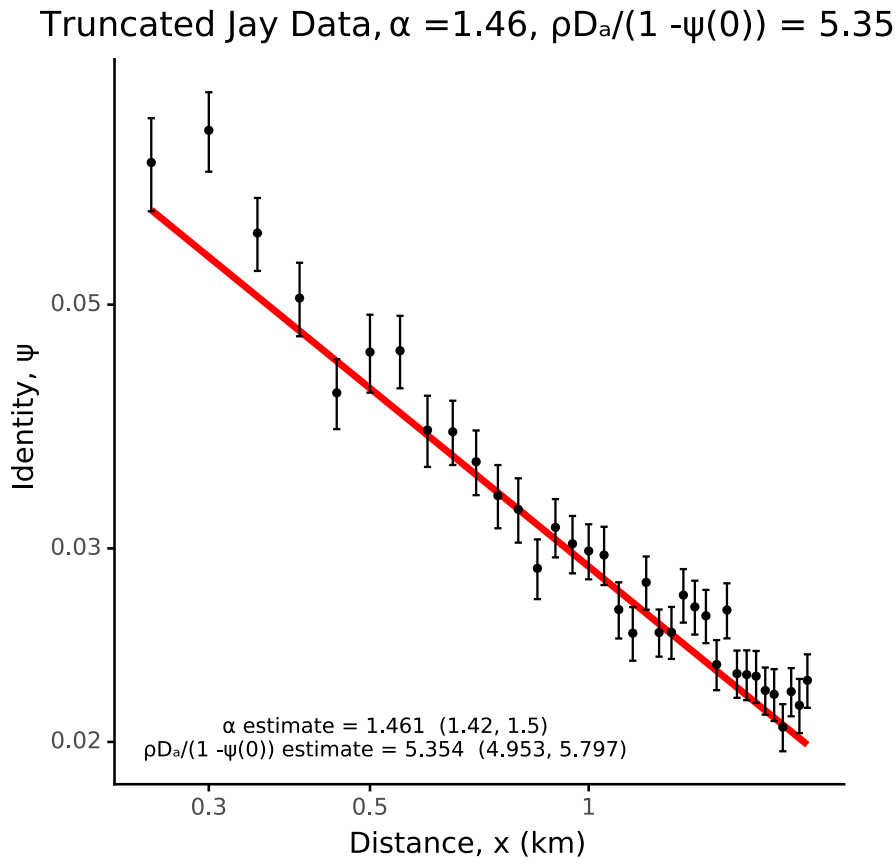

Figure 4.2: **We find the strongest signatures of long-range dispersal in the scrub-jay isolation by distance data between 200 meters and 2 kilometers.** As explained in the text, truncating the data allows us to omit the local details of dispersal and coalescence. It also allows us to exclude long-distance data that is affected by the finite size of the sampling range. This subset of the data demonstrates power-law like scaling over distances spanning a full order of magnitude. Our power-law model thus provides an excellent description of isolation by distance in this regime. Points show empirical measurements from the scrub-jay dataset. Error bars show the standard error of the mean in each 50 km distance bin. The red curve shows the maximum likelihood fit of the asymptotic Lévy flight model with $\mu = 0$. Parameter estimates for the model are inset in the figure.

As shown in Smith and Weissman [150], the following expression for ψ holds in two dimensions when $\psi \ll 1$:

$$\psi(x) \approx \frac{1-\psi(0)}{4\pi\rho}\int_0^\infty dk \frac{kJ_0(kx)}{\mu + D_\alpha k^\alpha} \tag{4.12}$$

$$= \frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{1+\kappa^\alpha}. \tag{4.13}$$

Note that we are assuming $\delta \ll x$ for all data such that the local details of coalescence can be ignored and that $\psi(0)$ is an unknown constant.

The block density can be defined for a single chromosome by the following expectation [134]:

$$E[N_L(x)] = \int_0^\infty dt\ \left(4Gt^2 + 4t\right) p(t|x) e^{-2Lt}, \tag{4.14}$$

where G is the map length of the genome. The block density for N_c chromosomes of total map length G is then given by:

$$E[N_L(x)] = \int_0^\infty dt\, N_c \left(4\left(\frac{G}{N_c} - L\right)t^2 + 4t\right) p(t|x) e^{-2Lt}, \tag{4.15}$$

where we assume all N_c chromosomes are the same length. Following the diffusive model developed by Ringbauer et al. [134], we assume we're working with long blocks (and thus short timescales) so that we can approximate the coalescence time distribution as the dispersal kernel divided by ρ:

$$E[N_L(x)] \approx \int_0^\infty dt\, N_c \left(4\left(\frac{G}{N_c} - L\right)t^2 + 4t\right) \frac{K(x|t)}{\rho} e^{-2Lt}. \tag{4.16}$$

We can then get approximate expressions for $E\left[Te^{-2LT}\right]$ and $E\left[T^2e^{-2LT}\right]$, and thus the block density, by differentiating (4.12) in the large block (short time) limit where $p(t|x) \approx K(x|t)/\rho$ and $1-\psi(0) \approx 1$:

$$E\left[Te^{-2LT}\right] \approx \frac{1}{8\pi\rho}\int_0^\infty dk \frac{kJ_0(kx)}{\left(L + D_\alpha k^\alpha\right)^2} \tag{4.17}$$

$$= \frac{1}{8\pi\rho L^2\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{\left(1+\kappa^\alpha\right)^2}, \tag{4.18}$$

$$E\left[T^2 e^{-2LT}\right] \approx \frac{1}{8\pi\rho}\int_0^\infty dk \frac{kJ_0(kx)}{\left(L + D_\alpha k^\alpha\right)^3} \tag{4.19}$$

$$= \frac{1}{8\pi\rho L^3 \overline{x}^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{\left(1+\kappa^\alpha\right)^3}. \tag{4.20}$$

Note that we have replaced the mutation rate μ in (4.12) with the map length of a shared block in Morgans, L. $\overline{x}$ is now defined as $(D_\alpha/L)^{1/\alpha}$.

We can leverage the simple form of the integral expressions (4.13), (4.18), and (4.20) by evaluating the dimensionless integrals (without their dimensionful prefactors) over a 2D grid of α and $x/\overline{x}$ values and fitting splines to these grids. We can then add the dimensionful prefactors of the integrals to the spline predictions to get the final numeric evaluations of ψ or $E[N_L(x)]$. This approach of using precomputed splines in place of direct numerical integration greatly improves the speed at which these expressions can be evaluated. Three splines are used at different distance scales for each quantity: one between $10^{-5}\ \overline{x}$ and $10^{-3}\ \overline{x}$, another between $10^{-3}\ \overline{x}$ and $10^{-1}\ \overline{x}$, and a third spline from $10^{-1}\ \overline{x}$ to $100\ \overline{x}$. The integral evaluations used to fit the two splines at short distances were computed with Mathematica, and the evaluations for the large distance spline were computed using the oscillatory quadrature method within Python's mpmath package. As explained below, asymptotic expressions are used at distances below $10^{-5}\ \overline{x}$ and above $100\ \overline{x}$.

4.8.2 T-distribution model

For the t-distribution inference code, we assume $\alpha > 2$ and define the single-generation dispersal distribution for a lineage as:

$$K_1(y) = \frac{\alpha}{2\pi c^2}\left(1 + y^2/c^2\right)^{-(\alpha+2)/2} \qquad \text{for } \alpha > 2. \tag{4.21}$$

Note that we have rescaled the distribution from its typical definition, which can be recovered by changing variables to $Y = \sqrt{\alpha}y$. The two-dimensional Fourier transform is:

$$\hat{K}_1(k) = \frac{\alpha\pi(ck)^{\alpha/2}\left(I_{-\alpha/2}(ck) - I_{\alpha/2}(ck)\right)}{2^{1+\alpha/2}\Gamma(1+\alpha/2)\sin(\pi\alpha/2)}, \tag{4.22}$$

where k represents the radial coordinate (magnitude of the frequency vector) in Fourier space and I is a modified Bessel function of the first kind.

For non-stable distributions, we must use a discrete-time model of coalescence, as all Markovian dispersal distributions become stable in the limit of continuous time [119]. Using the expressions derived for a generic dispersal distribution in Smith and Weissman [150], we find that ψ for our t-distribution dispersal kernel takes the following form:

$$\psi(x) \approx \frac{1-\psi(0)}{2\pi\rho} \int_0^\infty kJ_0(kx) \left(1 - e^{-2L} \left(\frac{\alpha\pi(ck)^{\alpha/2}\left(I_{-\alpha/2}(ck) - I_{\alpha/2}(ck)\right)}{2^{1+\alpha/2}\Gamma(1+\alpha/2)\sin(\pi\alpha/2)}\right)^2\right)^{-1} dk. \tag{4.23}$$

We can also rewrite (4.23) in terms of the dimensionless distance x/c:

$$\psi(x) \approx \frac{1-\psi(0)}{2\pi\rho c^2} \int_0^\infty \kappa J_0(\kappa x/c) \left(1 - e^{-2L} \left(\frac{\alpha\pi(\kappa)^{\alpha/2}\left(I_{-\alpha/2}(\kappa) - I_{\alpha/2}(\kappa)\right)}{2^{1+\alpha/2}\Gamma(1+\alpha/2)\sin(\pi\alpha/2)}\right)^2\right)^{-1} d\kappa, \tag{4.24}$$

which is the expression we use to create the spline grid for this model.

As explained in the previous section, the derivatives of ψ with respect to L can be used to approximate the block density. For the t-distribution block density model, we use the first and second derivatives of (4.24) to fit splines for inference. We evaluate these block density expressions on a three-dimensional grid of parameter values, with x/c between 0 and 100, α between 2.01 and 7, and L between .04 and .2. For all $\alpha < 4$, an additional set of splines were fit with the same L values and x/c between 100 and 250. Using the spline numerics for $E[N_L(x)]$, we apply the likelihood function described in (4.10) when fitting to data. The fact that map length is constrained between 4 and 20 cM makes the three-dimensional grid evaluations feasible for block density. Since the mutation rate μ has no such bounds, the t-distribution model is used only for block inference.

Fits to simulated data with $\alpha > 2$ were performed using the t-distribution model described above (see Simulation Methods subsection of Methods). The performance of this inference method for large α was found to be poorer than that of the Lévy flight methods for $\alpha \leq 2$. One key difference between the two regimes is that all

power-law dispersal distributions with $\alpha > 2$ begin to resemble a normal distribution at short distances after multiple generations, and thus reliable information about the tail parameter α can only be obtained from the tail of an isolation by distance curve. For Lévy flight dispersal, however, we see signatures of α at all distances, not just the (often noisy) tail of the isolation by distance curve. It is thus not surprising that inference for $\alpha > 2$ is significantly more difficult than the case of $\alpha \leq 2$. While α cannot be reliably inferred with our t-distribution model, we find that diffusive simulations results in an optimal α value of 7, which is the maximum allowed value in our method. This is expected: t-distributions approach a normal distribution as the tail parameter α becomes large. Simulations with steep power-law dispersal ($2 < \alpha < 6$) instead prefer lower α values below 3. We can thus use our t-distribution to detect the presence of long-range dispersal: an estimated α less than 7 suggests that heavy-tailed dispersal may in fact be present. We can also compare estimates made for natural populations to the estimates made using diffusive and Lévy flight models. If the t-distribution model provides a better fit (AIC value) than the other two models, we claim that dispersal is heavy-tailed with $\alpha > 2$, but do attempt to estimate α with any further precision.

4.8.3 Asymptotic expansions at long and short distances

For $\alpha < 2$, $1-\psi(0)$ is independent of μ to leading order, and we can obtain $p(t|x)$ for $x \gg \delta$ by taking the inverse Laplace transform of (4.13) [150]:

$$p(t|x) \approx \frac{1-\psi(0)}{\rho} K(x|t), \tag{4.25}$$

where the two-lineage dispersal kernel K has the same definition as K_1 given in (4.1), but with D_α replaced with $2D_\alpha$. The distance coordinate x now represents the present-day separation, rather than locations, of two lineages.

We can approximate $\psi(x)$ for both $x \ll \overline{x}$ and $x \gg \overline{x}$ by expanding the characteristic function for K given in (4.1) and Laplace transforming each term in the expansion. At large distances, $x \gg (D_\alpha t)^{1/\alpha}$, expanding the kernel yields:

$$p(t|x) \approx \frac{1-\psi(0)}{\rho} \sum_{1}^{n=\infty} \frac{(-1)^{n+1} 2^{n\alpha} \Gamma(1+\frac{n\alpha}{2})^2 \sin(\frac{n\pi\alpha}{2})}{n! \pi^2 (2D_\alpha t)^{2/\alpha}} \left(\frac{x}{(2D_\alpha t)^{1/\alpha}} \right)^{-2-n\alpha}. \tag{4.26}$$

By Laplace transforming each term in the series, we find an expression for ψ when $x \gg \overline{x}$:

$$\psi(x \gg \overline{x}) \approx \frac{1-\psi(0)}{\rho\mu\overline{x}^2} \sum_{1}^{n=\infty} \frac{(-1)^{n+1} 2^{n\alpha-1} \Gamma(1+\frac{n\alpha}{2})^2 \sin(\frac{n\pi\alpha}{2})}{\pi^2} \left(\frac{x}{\overline{x}}\right)^{-2-n\alpha}. \tag{4.27}$$

At short distances $x \ll \overline{x}$, we can split (4.13) into two integrals, one from 0 to 1 and the other from 1 to ∞. As we are interested in very small distances, we keep only the terms that are constant or divergent as x approaches zero. For the first integral from 0 to 1, to leading order we find:

$$\frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2} \int_0^1 d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{1+\kappa^\alpha} \approx \frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2} \left(\frac{F(1/2+1/\alpha) - F(1/\alpha)}{2\alpha}\right), \tag{4.28}$$

where F is the digamma function. For the integral from 1 to ∞, we expand the denominator $(1+\kappa^\alpha)^{-1}$ in powers of $\kappa^{-\alpha}$, which is equivalent to expanding the integrand of the dimensionful expression (4.12) about $\mu = 0$. The resulting expression is:

$$\frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2} \int_1^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{1+\kappa^\alpha} \approx \frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2} \sum_0^\infty (-1)^n \int_1^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{\kappa^{\alpha(n+1)}}. \tag{4.29}$$

Using Mathematica's AsymptoticIntegrate function to solve for the leading-order behavior of each term in the series, we find:

$$\frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2} \int_1^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{1+\kappa^\alpha} \approx$$
$$\frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2} \left(\sum_0^\infty \frac{(-1)^n}{-2+(n+1)\alpha} + \sum_{n=0}^{\lfloor 2/\alpha - 1 \rfloor} \frac{(-1)^n \Gamma\left(1-\frac{(n+1)\alpha}{2}\right)}{2^{-1+(n+1)\alpha}\Gamma\left(\frac{(n+1)\alpha}{2}\right)} \left(\frac{x}{\overline{x}}\right)^{(n+1)\alpha-2} \right), \tag{4.30}$$

where the second series contains singular terms that occur only for $n < (2/\alpha - 1)$. Summing the first series and adding (4.30) to (4.28), we find that the expression for the full integral is:

$$\psi(x \ll \overline{x}) \approx$$
$$\frac{1-\psi(0)}{4\pi\rho\mu\overline{x}^2} \left(\frac{\pi \csc(2\pi/\alpha)}{\alpha} + \sum_{n=0}^{\lfloor 2/\alpha - 1 \rfloor} \frac{(-1)^n \Gamma\left(1-\frac{(n+1)\alpha}{2}\right)}{2^{-1+(n+1)\alpha}\Gamma\left(\frac{(n+1)\alpha}{2}\right)} \left(\frac{x}{\overline{x}}\right)^{(n+1)\alpha-2} \right). \tag{4.31}$$

Note that the expression for ψ has poles at integer values of $2/\alpha$. In practice, we round alpha to 4 decimal places and shift by a factor of $10^{-4}\pi$ to avoid the poles of the series, and we find that the resulting asymptotics agree with our spline numerics and are effective for inference. Continuous optimization methods in Python perform well despite this regulation of alpha.

We use similar series expressions for the block density when $\alpha < 2$, which can be obtained by differentiating (4.31) above. We also assume that the true population has discrete generations, and thus modify our continuous-time block density model by assuming a constant value below $x = c/100$. Assuming the smallest blocks used for inference have map length $L = .04$, we find that the asymptotic expressions become unnecessary for $\alpha \geq 1/2$, where splines down to $10^{-5}\overline{x}$ cover all distance above the $c/100$ threshold. For smaller values of α, the asymptotic expressions are used to ensure accuracy down to the $c/100$ threshold.

For $\alpha > 2$, the t-distribution model naturally has discrete generations, and we set the block density to a constant value for $x < c/10$. Asymptotic power-law expressions derived from (4.21) are used for all $x > 100\overline{x}$ and all x that exceed the maximum value of x/c on the spline grid [150].

For $\alpha = 2$, we use the known logarithmic expression for $\psi(x)$ at short distances [15, 150]. At large distances, ψ and its derivatives decay exponentially, and thus all quantities of interest are effectively zero for $x/\overline{x} > 100$.

4.8.4 Simulation methods

All simulation code and displayed data are available at `https://github.com/weissmanlab/Dispersal_Inference`. We simulate our model in two stages. First, for each value of present-day separation x, dispersal constant D_α, and tail parameter α, we simulate dispersal of the lineages, ignoring coalescence and mutation. Then, for each value of ρ and μ, we calculate the expected homozygosity and coalescence time distribution for each simulated trajectory. We then average over many independent trajectories. A major advantage of this two-part method is that the second part of the method, in which conditional expectations are calculated for previously generated

paths, is entirely deterministic. This reduces computational costs and noise in the estimations.

We simulate lineage motion using a discrete time random walk,

$$X_{t+1} = X_t + \Delta X_t, \tag{4.32}$$

where X_t represents the position of a lineage at a given time (ignoring coalescence., i.e., assuming $\rho \to \infty$), and the displacement, ΔX_t, is a vector of integer valued random variables drawn from the dispersal distribution at each integer time t. We use the GNU Scientific Library's efficient pseudorandom generators for both stable distributions and the F-distribution [66]; because these are available only for the one-dimensional distributions, we draw radial distances using the one-dimensional distributions and then select a random direction in which to move. For dispersal, we primarily use one-dimensional Lévy alpha-stable distributions to randomly draw distances. To simulate steeper tails with $\alpha > 2$, we use an F-distribution for radial distances. In continuous space, ΔX_t would have distribution:

$$K_1(y) = \frac{1}{2\pi^2|y|}\int_{-\infty}^{\infty} dk \exp\left(-iky - D_\alpha|k|^\alpha\right) \qquad \text{for } \alpha \leq 2, \tag{4.33}$$

$$K_1(y) = \frac{\Gamma(2\alpha)}{2\pi\omega\Gamma(\alpha)^2}|y/\omega|^{\alpha-2}\left(1+|y/\omega|\right)^{-2\alpha} \qquad \text{for } \alpha > 2, \tag{4.34}$$

where the probability of a displacement depends only on its magnitude, y. As explained below, ω for $\alpha > 2$ is chosen such that the scaled F-distribution has the desired mean squared displacement. To enforce our condition of discrete dispersal distances, we then round ΔX_t to the nearest pair of integers, i.e., the closest point in $\mathbb{Z}^2$. Note that this discretization procedure will affect the resulting coalescence and dispersal process. Estimates for ρ using our inference method will be slightly higher than the ρ parameter used in simulations due to this effect. Preliminary simulations on continuous space showed no such bias.

For each pair of simulated trajectories $\{x_t\}$, we then compute the path-specific distribution of coalescence times $p(\{x_{t'\leq t}\})$, i.e., the probability of coalescing at and not before time t. We also compute the path-specific mean homozygosity $\psi(\{x_{t'\leq\infty}\})$, i.e., the probability that lineages following these exact trajectories have not mutated

before coalescence, or the path-specific number of long shared sequence blocks per pair and cM $E[N_L](\{x_{t'\leq\infty}\})/100$:

$$p(\{x_{t'\leq t}\}) = \left(1 - e^{-\frac{1}{\rho}\delta_{x_1x_2}}\right) \exp\left[-\frac{1}{\rho}\sum_{t'=1}^{t-1}\delta_{x_1x_2}\right] \qquad \text{for } t > 1, \tag{4.35}$$

$$\psi(\{x_{t'\leq\infty}\}) = \sum_{t=1}^{\infty} p(\{x_{t'\leq t}\})e^{-2\mu t}, \tag{4.36}$$

$$E[N_L](\{x_{t'\leq\infty}\})/100 = \sum_{t=1}^{\infty} N_c\left(4\left(\frac{G}{N_c} - L\right)t^2 + 4t\right)p(\{x_{t'\leq t}\})e^{-2\mu t}/100. \tag{4.37}$$

We start (4.35) and (4.36) at $t = 1$ because we assume that the individuals are sampled immediately after dispersal, so no coalescence takes place at $t = 0$. $\delta_{x_1x_2}$ in (4.35) is the Kronecker delta function:

$$\delta_{x_1x_2} \equiv \begin{cases} 1 & \text{if } x_1 = x_2 \\ 0 & \text{otherwise.} \end{cases}$$

For every time-step the lineages spend in this region, there is a probability of coalescence $1 - e^{-\frac{1}{\rho}}$.

We then draw binomial and Poisson-distributed random variables for each trajectory using (4.36) and (4.37) as means. Draws are averaged across all simulated trajectories with present-day separation x to obtain $\psi(x)$ and $E[N_L(x)]/100$.

We simulate 1,000,000 independent runs of 100 generations each for $\psi(x)$ for each present-day separation x and tail parameter α. All $\psi(x)$ simulations shown use a mutation rate μ of 0.1. For $E[N_L(x)]$, we generated separate trajectories for every 1 cM bin of block density, and used 100,000 trajectories per bin for the data fit using the Lévy flight inference code ($\alpha \leq 2$). For the t-distribution block inference code, we performed simulations with $\alpha \geq 2$ and generated 25000 trajectories per bin. All trajectories used for $E[N_L(x)]$ were also 100 generations in length. We set the dispersal constant D_α indirectly by setting the characteristic spread c of each lineage after one generation ($t = 1$), $c \equiv (D_\alpha)^{1/\alpha}$, to be fixed at $c = 5$. For $\alpha > 2$, c is the root mean squared displacement of the dispersal distribution. We also apply periodic

boundary conditions, with the range size extending from -5000 to 5000 along both dimensions of the discrete lattice. Our choice of range size is significantly larger than the maximum values used for present-day separation between pairs, which are $x = 591$ for all t-distribution simulations and $x = 237$ for all Lévy flight simulations.

4.8.5 Optimization methods

The simple forms of the likelihood functions (4.8) and (4.10) allow us to reduce these three-dimensional parameter estimation problems to two: for every pair (α, $\overline{x}$) or (α, c), the optimal value of the additional parameter $\rho\mu/\left(1-\psi(0)\right)$ (or ρ) can be found analytically. Setting the derivative of the log-likelihood with respect to $\frac{\rho\mu}{1-\psi(0)}$ (or ρ) to zero results in a quadratic equation, and the single positive root is the unique solution for the optimal value of the parameter (holding the other parameters constant). This dimensional reduction, along with the use of splines to approximate numerical integration, greatly reduces the computational cost of parameter inference and makes fitting large datasets with high spatial resolution feasible on a single CPU core (see Lévy flight model and T-distribution model sections of Methods).

Initial estimates of $\overline{x}$ (or c) are first found via exponential or logistic fits to data. Using these initial estimates, we evaluate the likelihood for a given dataset over a two-dimensional grid of α and $\overline{x}$ (or c) values. α is varied from .25 to 2.0 in increments of .05, and $\overline{x}$ (or c) is varied from one-tenth to ten times the initial estimate in increments of one-tenth of the initial estimate. We take the 5 points of best fit on the parameter-space grid as starting points for further optimization using SciPy's implementation of the BFGS algorithm in Python. The best parameter values found over the course of these 5 optimization runs are used to set a new initial estimate for $\overline{x}$ (or c), and this process is reiterated a fixed number of times, after which the best parameter values found are used as the final parameter estimates. Five iterations are used for the identity inference code based on $\psi(x)$, and two iterations are used for the block density inference code involving $E[N_L(x)]$ (though this number is increased to four in our analysis of the POPRES data). Confidence intervals for α and $\overline{x}$ (or c) are then computed using the likelihood ratio test on the two-dimensional grid of log-likelihood

evaluations. Confidence intervals for $\frac{\rho\mu}{1-\psi(0)}$ (or ρ) are set by either the range of values found on the grid or the range given by the Fisher information matrix (whichever is larger). For the block density model that assumes t-distribution dispersal with $\alpha > 2$, we use the procedure described above, but with α varied between 2.1 and 7.0 in increments of 0.1.

For the asymptotic model of identity by descent used to analyze the Florida scrub-jay population, we used the same basic scheme described above, but with a reduced number of dimensions. The optimal value of $\rho D_\alpha / (1 - \psi(0))$ for (4.9) can be determined analytically for fixed α, and we vary α from .25 to 1.99 in increments of .01 to get an initial grid of likelihood values. We again take the 5 points of best fit on the grid as starting points for further optimization using SciPy's BFGS algorithm, and confidence intervals for α are once again found using the likelihood ratio test or the Fisher information matrix (whichever is larger). Confidence intervals for $\rho D_\alpha / (1 - \psi(0))$ are similarly set by either the range of values found on the grid or the range given by the Fisher information matrix (whichever is larger).

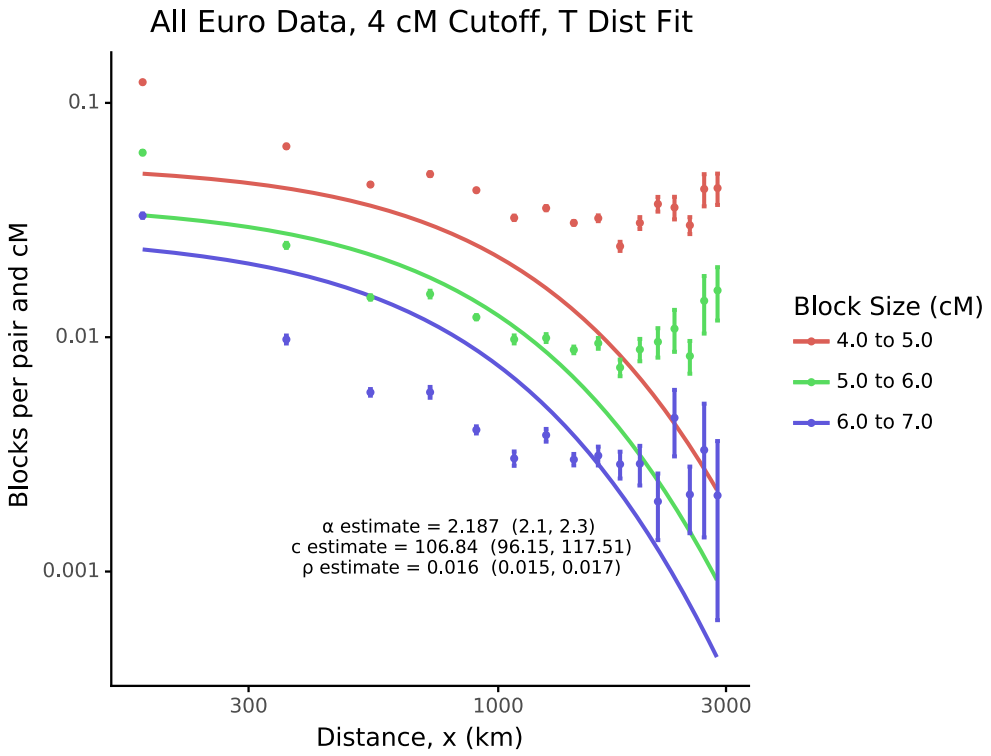

Figure 4.3: **For all European humans within the POPRES dataset, the lack of isolation by distance obstructs dispersal inference.** We see that the number of shared sequence blocks plateaus at large distances. This results in our model of isolation by distance and any associated inferences being inaccurate. The plot shows blocks between 4 and 7 cM using distance bins of width 180 km. Error bars show the standard deviation of the mean for the number of blocks per pair in each bin. Parameter estimates are inset in the figures above, and the details of the inference procedures are described in the text.

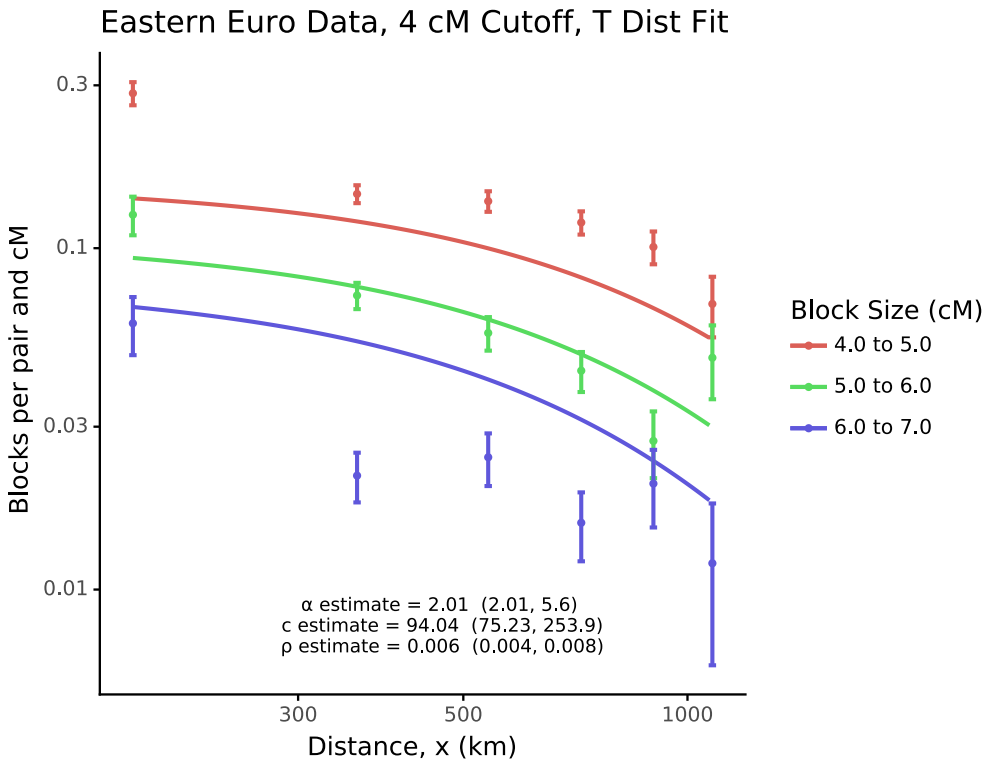

Figure 4.4: **For Eastern European humans, we detect genomic signatures of long-range dispersal within the POPRES dataset.** Using a 4 cM minimum block length, we find that our t-distribution model provides a better description of Eastern European isolation by distance than the classic diffusive model, with AIC scores for the models being 10845 and 10850 respectively. The plot shows blocks between 4 and 7 cM using distance bins of width 180 km. Error bars show the standard deviation of the mean for the number of blocks per pair in each bin. Parameter estimates are inset in the figures above, and the details of the inference procedures are described in the text.

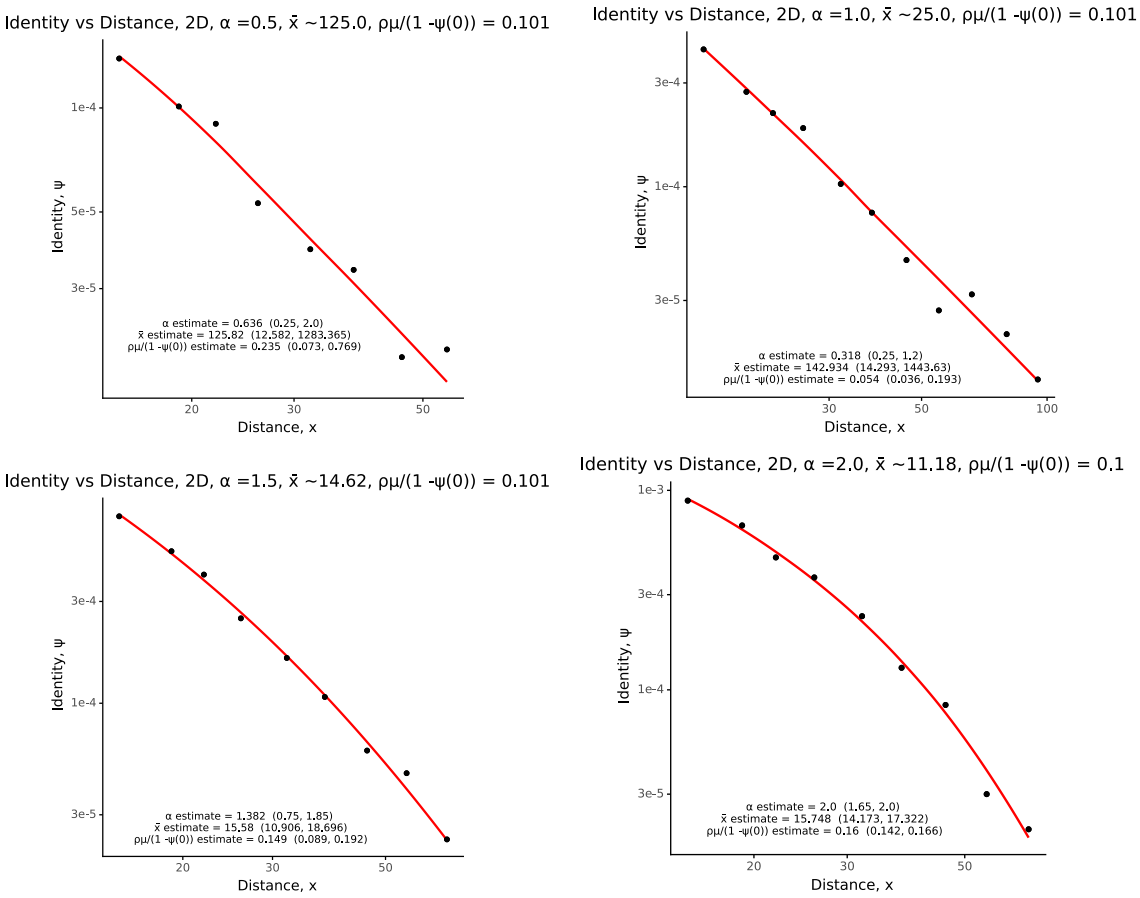

Figure 4.5: **Long-range dispersal inference via probability of identity performs well for simulated data.** Points show discrete-space simulation results. Curves show the best-fit curves of the continuous-space Lévy flight model. We see that our inference method based on the probability of identity $\psi(x)$ performs well for power-law dispersal ($\alpha < 2$) and diffusive dispersal ($\alpha = 2$). $\rho = 1$, $c = 5$, and $\mu = 0.1$ for all simulated data shown, and 95 percent confidence intervals for all estimated parameters are displayed in each panel. Actual parameter values used in simulations are shown in the panel titles.

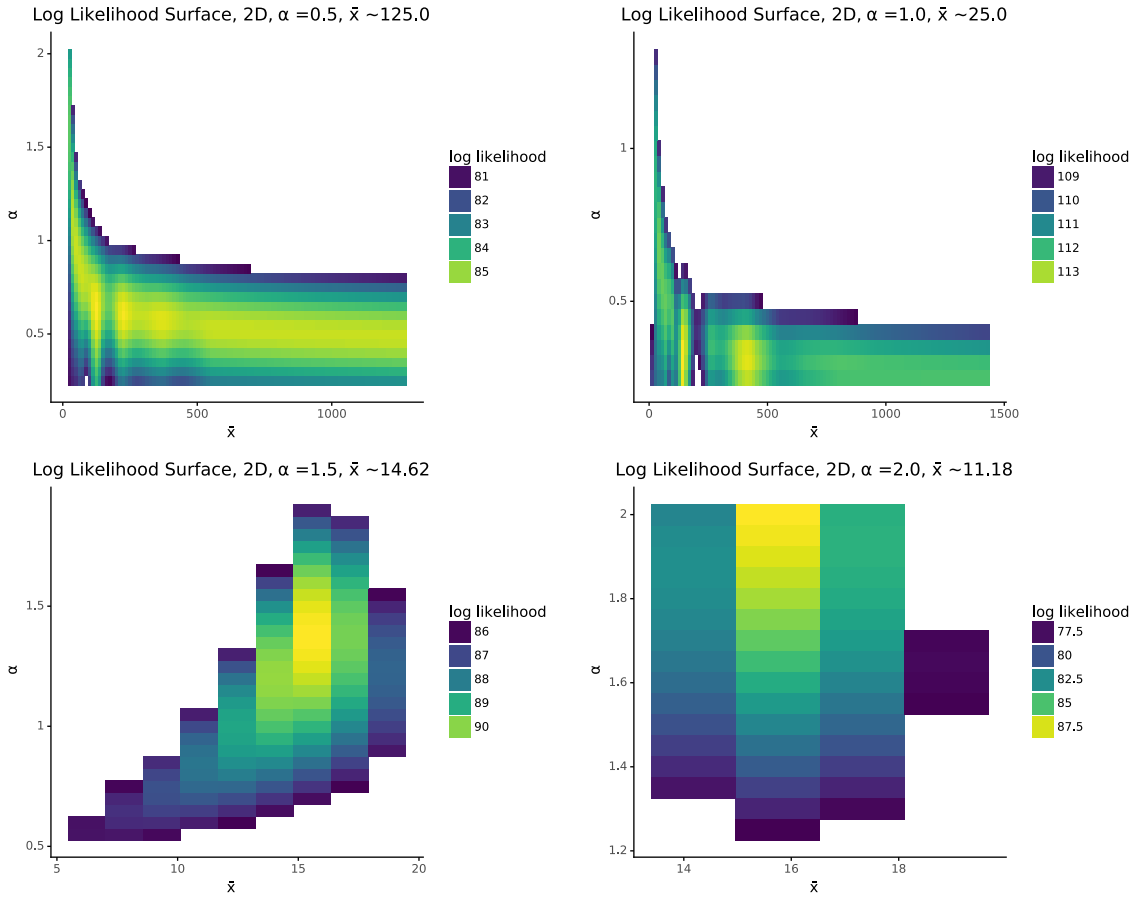

Figure 4.6: **Parameter-space likelihood heatmaps for ψ fit to simulated data.** Heatmaps show log-likelihood evaluated over a discrete grid of α and $\overline{x}$ values. Brighter colors indicate a higher value of log-likelihood. Actual parameter values used in simulations are shown in the panel titles.

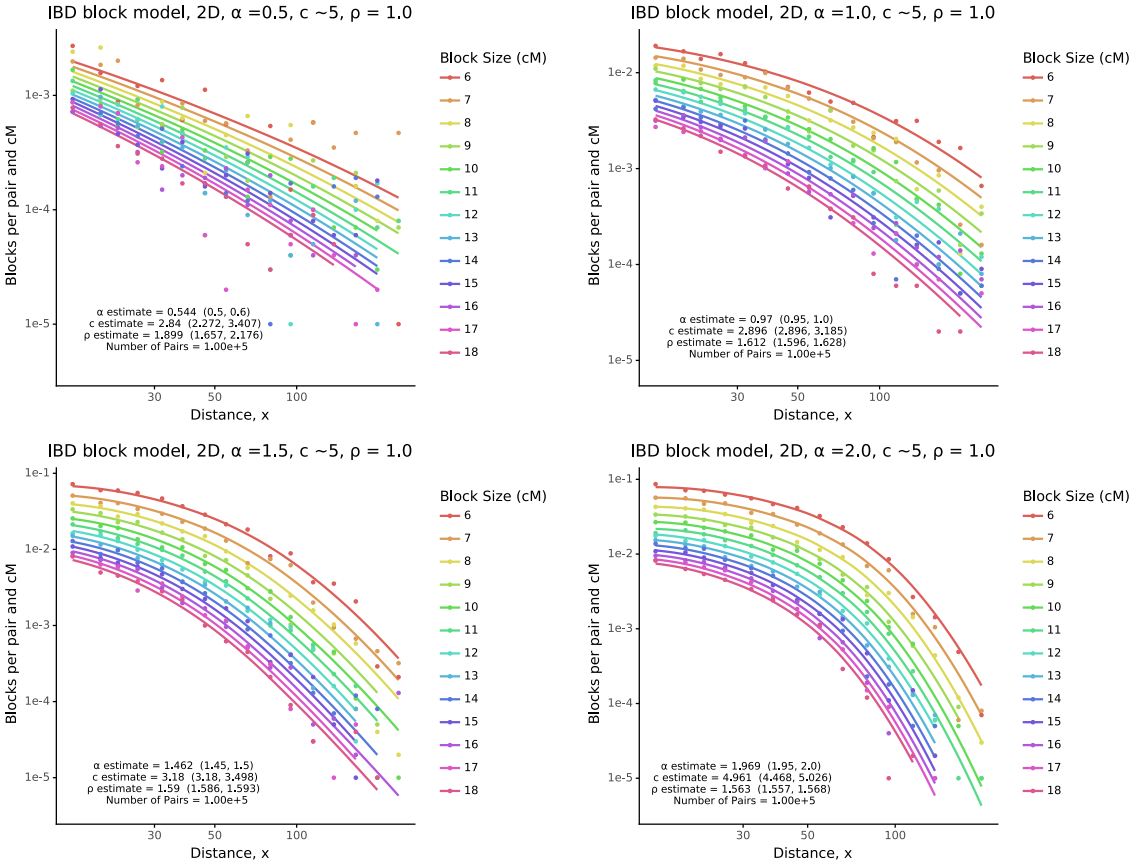

Figure 4.7: **Long-range dispersal inference via long shared sequence blocks performs well for simulated data.** Points show simulation results. Curves show Lévy flight model predictions using the estimated parameters. Colors of points and curves indicated the size of the blocks in centimorgans. We see that our inference method based on the block density $E[N_L(x)]$ performs well for power-law dispersal ($\alpha < 2$) and diffusive dispersal ($\alpha = 2$). $\rho = 1$ and $c = 5$ for all simulated data shown. 95 percent confidence intervals for all estimated parameters are displayed in each panel. Actual parameter values used in simulations are shown in the panel titles.

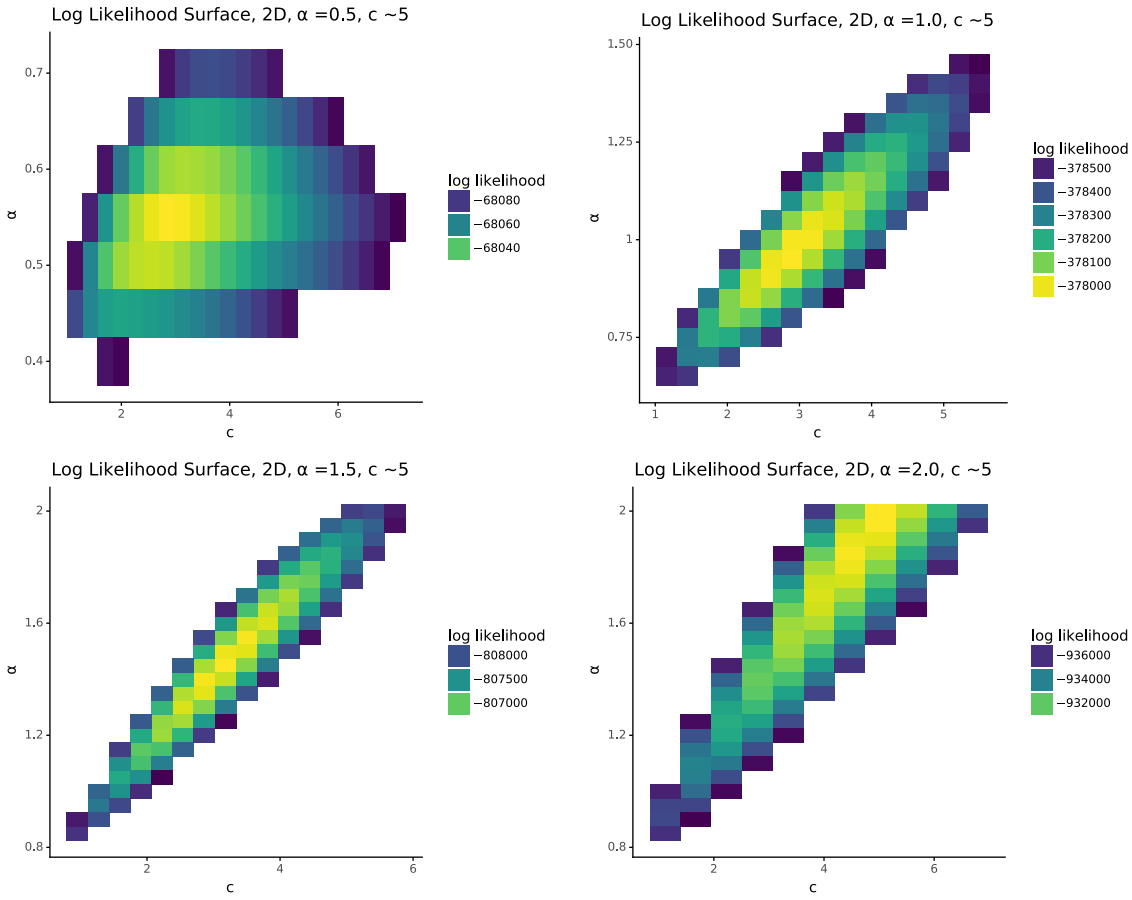

Figure 4.8: **Likelihood heatmaps of the Lévy flight block density model fit to simulated data.** Heatmaps show log-likelihood evaluated over a discrete grid of α and c values. Brighter colors indicate a higher value of log-likelihood. Actual parameter values used in simulations are shown in the panel titles.

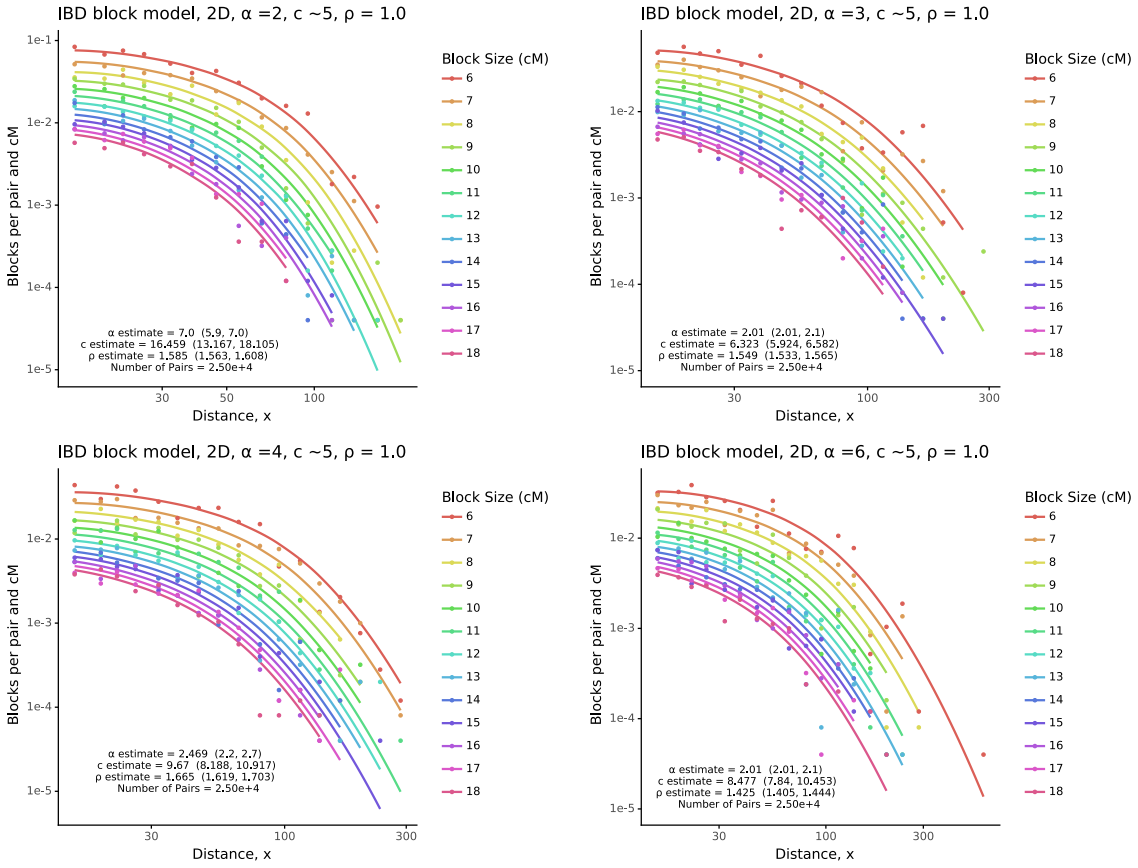

Figure 4.9: **Plots of the t-distribution block density model fit to simulated data.** Points show simulation results. Curves show block density model predictions using the estimated parameters. Colors of points and curves indicated the size of the blocks in centimorgans. We see that our inference method based on the t-distribution model of block density $E[N_L(x)]$ performs poorly for $\alpha > 2$, but does assign diffusive motion the largest possible α value, $\alpha = 7$. This is expected, as a t-distribution approaches a normal distribution as the power-law exponent α becomes large. We us the t-distribution model to detect the presence or absence of long-range dispersal with $\alpha > 2$, but do not attempt to infer an exact value of α. $\rho = 1$ and $c = 5$ for all simulated data shown. 95 percent confidence intervals for all estimated parameters are displayed in each panel. Actual parameter values used in simulations are shown in the panel titles.

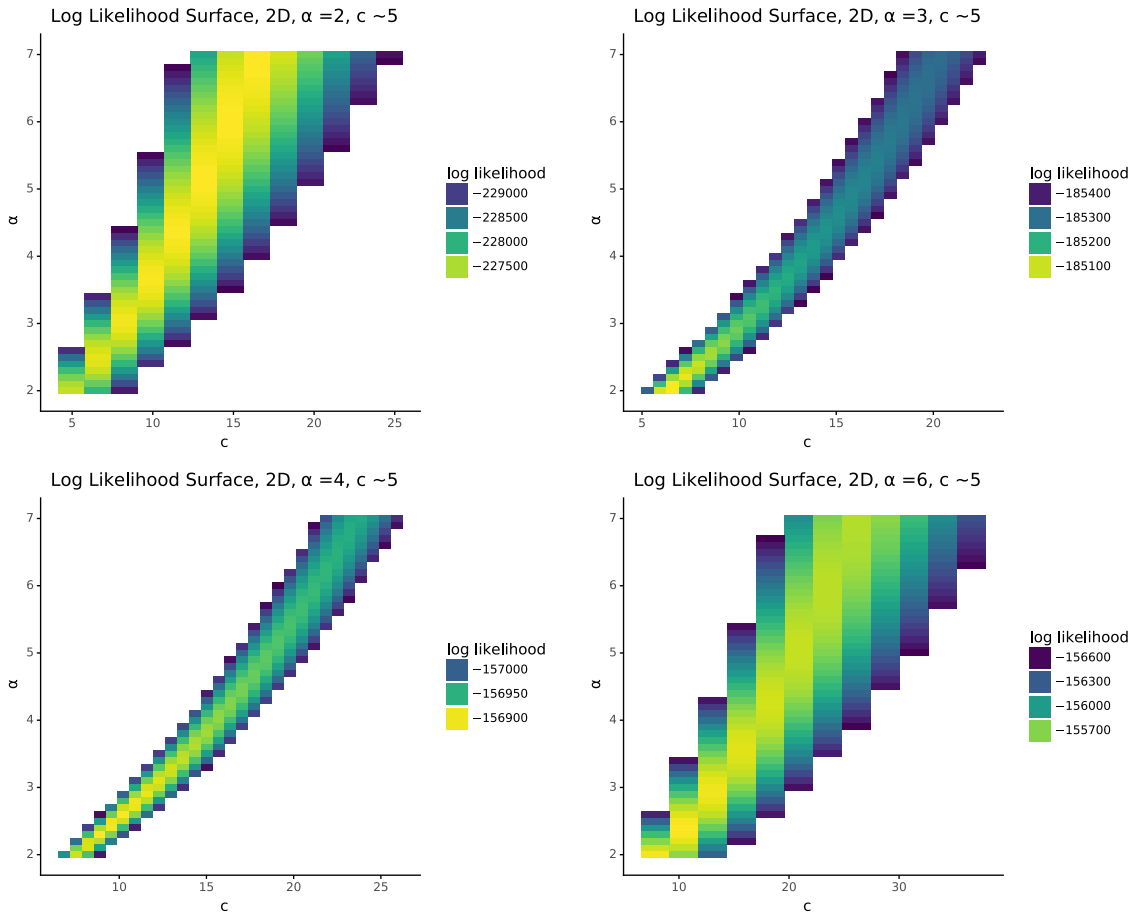

Figure 4.10: **Likelihood heatmaps of the t-distribution block density model fit to simulated data.** Heatmaps show log-likelihood evaluated over a discrete grid of α and c values. Brighter colors indicate a higher value of log-likelihood. Actual parameter values used in simulations are shown in the panel titles.

Chapter 5

Quantum Geometry and Semiclassical Electron Dynamics

The work presented in this chapter was performed under the supervision of Professor Ajit Srivastava

5.1 Introduction

Berry curvature serves as a unifying concept in solid-state physics, helping to explain orbital magnetization, the quantum Hall effect, and many other seemingly disparate phenomena [171]. For crystalline systems, Berry curvature exists for the Bloch waves that describe the eigenstates of a single electron [6]:

$$|\psi_{n\mathbf{k}}(\mathbf{r})\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{n\mathbf{k}}(\mathbf{r})\rangle, \tag{5.1}$$

where $\mathbf{r}$ represents position and $\mathbf{k}$ represents crystal momentum, which is a conserved quantity in periodic systems. Crystal momentum parameterizes the Bloch waves, and the variation of these wave functions endows the momentum space with a particular geometric structure. The hallmark of this geometric structure is a momentum space Aharanov Bohm effect known as the Berry phase, with the Berry curvature playing the role of a momentum space magnetic field [171].

A natural approach to understanding the role Berry curvature plays in the electronic properties of solids is the semiclassical formulation of electron dynamics, which considers the motion of a wave packet that is localized in both position and momentum space [157]. The wave function describing the packet can be expressed as a superposition of Bloch waves [6]:

$$|W\rangle = \int \mathrm{d}^3\mathbf{k}\; w(\mathbf{k}-\mathbf{k}_c)e^{i\mathbf{k}\cdot\mathbf{r}}|u_0(\mathbf{k})\rangle, \tag{5.2}$$

where $w(\mathbf{k}-\mathbf{k}_c)$ is a rapidly decaying function centered around its peak at $\mathbf{k}_c$. The dynamics of this packet can then be derived from an effective Lagrangian [171]:

$$\mathcal{L} = \langle W|i\partial_t - H|W\rangle, \tag{5.3}$$

where H is the Hamiltonian for a single electron within the crystal. The resulting equations of motion for the packet are:

$$\dot{\mathbf{k}} = \frac{\partial V(\mathbf{r})}{\partial \mathbf{r}} - \dot{\mathbf{r}} \times \mathbf{B}, \tag{5.4}$$

$$\dot{\mathbf{r}} = \frac{\partial \mathcal{E}(\mathbf{k})}{\partial \mathbf{k}} - \dot{\mathbf{k}} \times \mathbf{\Omega}, \tag{5.5}$$

where $\mathcal{E}$ is the energy of a Bloch wave in the lowest band and $\mathbf{\Omega}$ is the Berry curvature, which is defined as the curl of the Berry connection [24]:

$$\mathcal{A}(\mathbf{k}) = \langle u_0(\mathbf{k})|i\nabla_{\mathbf{k}}|u_0(\mathbf{k})\rangle. \tag{5.6}$$

We see that nonzero Berry curvature leads to a linear response that is orthogonal to the applied electric field. This anomalous velocity can be viewed as the momentum space "dual" of the Lorentz force [44]. In addition to the Berry curvature, there exists a natural metric in momentum space known as the quantum geometric tensor:

$$g_{ij}dk^i dk^j = 1 - |\langle u_0(\mathbf{k})|u_0(\mathbf{k}+\mathbf{dk})\rangle|^2. \tag{5.7}$$

This metric measures the distance in Hilbert space between Bloch waves in the same band with different values of crystal momentum [5]. The quantum geometric tensor (QGT) has been shown to play a significant role in the presence of current noise,

orbital magnetization, the shifting of exciton energy levels, the fractional quantum hall effect, and quantum phase transitions [69, 74, 118, 155, 173]. Though apparently quite significant, the QGT and its role in solid-state physics are still not well understood.

Normally the semiclassical equations of motion are expressed to first-order in the electric field, and any higher-order corrections to the Berry curvature and dispersion are ignored. Recently though, second-order extensions of the semiclassical theory were studied for a static electric field and were found to result in a nonlinear hall response [68, 69]. In this letter, motivated by the fact that the quantum metric is a vacuum solution of the Einstein field equations in Hilbert space, we show that the second-order semiclassical equations of motion extend the first-order position-momentum duality by accounting for the curvature of momentum space:

$$\dot{k}_l = \frac{\partial V(\mathbf{r})}{\partial \mathbf{r}^l} - (\dot{\mathbf{r}} \times \mathbf{B})_l, \tag{5.8}$$

$$\dot{r}_l \approx \frac{\partial \mathcal{E}(\mathbf{k})}{\partial \mathbf{k}^l} - (\dot{\mathbf{k}} \times \boldsymbol{\Omega})_l - \frac{1}{\mathcal{E}} \Gamma_{jil} \dot{k}^i \dot{k}^j, \tag{5.9}$$

where Γ_{jil} is the Christoffel symbol representing the Levi-Civita connection of the metric [145]. While (5.5) is dual to the Lorentz force in flat spacetime, (5.9) is the momentum space dual of the Lorentz force in curved spacetime [109]. As shown in Fig. 5.1, this Christoffel term is what generalizes motion along straight lines in flat space to motion along geodesics in curved space.

We will also show that the stress-energy tensor for mixed states can be nonzero and that its trace can be expressed in terms of the entropy of the quantum state:

$$T \approx -\frac{R}{2k_B} S(\mathbf{k}) - \frac{1}{k_B} \Delta_{\mathbf{k}} S(\mathbf{k}). \tag{5.10}$$

(5.10) can be seen as the momentum space analog of Poisson's equation, which emerges from the Newtonian limit of general relativity [109].

5.2 Corrections to the band structure

To determine the response of the band structure to a constant electric field, we consider the Hamiltonian for a single electron in the dipole gauge:

$$H(\mathbf{k}) = H_0(\mathbf{k}) + \mathbf{E}\cdot\mathbf{r},$$

where $H_0(\mathbf{k})$ represents the electron Hamiltonian in the absence of an electric field. The perturbative correction to a two band Hamiltonian due to $\mathbf{E}\cdot\mathbf{r}$ can be represented as:

$$H' \approx \begin{pmatrix} \mathbf{E}\cdot\langle u_0(\mathbf{k})|\mathbf{r}|u_0(\mathbf{k})\rangle & \mathbf{E}\cdot\mathcal{A}_{01}(\mathbf{k}) \\ \mathbf{E}\cdot\mathcal{A}_{10}(\mathbf{k}) & \mathbf{E}\cdot\langle u_1(\mathbf{k})|\mathbf{r}|u_1(\mathbf{k})\rangle \end{pmatrix}, \tag{5.11}$$

where $\mathcal{A}_{10}$ and $\mathcal{A}_{01}$ are the interband Berry connections [61]. Note that the diagonal elements of H' are ambiguous in this gauge due to the periodicity of the Bloch waves, but that this ambiguity will not affect the resulting second-order wave packet corrections.

We can now use time independent perturbation theory in tandem with (5.11) to compute the first-order correction to the Berry connection and the second-order correction to the energy [68]:

$$a_i' = \frac{-2g_{ij}E^j}{\mathcal{E}_1 - \mathcal{E}_0},$$

$$\mathcal{E}'' = \frac{-g_{ij}E^iE^j}{\mathcal{E}_1 - \mathcal{E}_0}.$$

Using these corrections, we can determine the second-order semiclassical Lagrangian and equations of motion of a Bloch electron.

5.3 Second-order semiclassical equations of motion

5.3.1 Effective Lagrangian

The wave function describing the packet can be expressed as a superposition of Bloch states: $|W\rangle = \int \mathrm{d}^3\mathbf{k}\; w(\mathbf{k}-\mathbf{k}_c)e^{i\mathbf{k}\cdot\mathbf{r}}|u_0(\mathbf{k})\rangle$. Writing out the full effective

Lagrangian for a semiclassical wave packet, $L = \langle W|i\partial_t - H|W\rangle$, we find [68, 171]:

$$\mathcal{L} = (\mathcal{A}_i(\mathbf{k}) + a_i'(\mathbf{k}))\,\dot{k}^i + \dot{r}_i k^i - \mathcal{E}(\mathbf{k}) - \dot{r}_i A^i(\mathbf{r}, t) + \frac{g_{ij}\dot{k}^i\dot{k}^j}{2\mathcal{E}}. \tag{5.12}$$

We see that the momentum-space component of the Lagrangian is the dual of the Lagrangian for a charged particle in an electromagnetic field in curved space. To second-order, the resulting equations of motion are [68]:

$$\dot{\mathbf{r}} = -\dot{\mathbf{k}} \times \tilde{\mathbf{\Omega}}(\mathbf{k}) + \nabla_{\mathbf{k}}\tilde{\mathcal{E}}(\mathbf{k}), \tag{5.13}$$

where $\tilde{\Omega}$ is the Berry curvature corrected to first-order, and $\tilde{\mathcal{E}}$ is the energy corrected to second-order. Given that the first-order response due to Berry curvature dominates any second-order anomalous velocity orthogonal to the applied field, the correction due to a' can be neglected. The resulting equation of motion can then be written as $\dot{\mathbf{r}} \approx -\dot{\mathbf{k}} \times \mathbf{\Omega}(\mathbf{k}) + \nabla_{\mathbf{k}}\tilde{\mathcal{E}}(\mathbf{k})$. Noting that $(\nabla_{\mathbf{k}}\mathcal{E}'')_l = -\frac{2}{\mathcal{E}_1 - \mathcal{E}_0}\Gamma_{jil}E^iE^j$, where Γ_{jil} is the Christoffel symbol associated with the Levi-Civita connection of the quantum metric, we can write $\mathcal{E}_1 - \mathcal{E}_0$ as $2\mathcal{E}$ and recover (5.9).

5.3.2 Toy model for geodesic equation

We consider a two-band model of gapped graphene with the following Hamiltonian [155]:

$$H(\mathbf{k}) = \begin{pmatrix} \Delta & ab(k_x - ik_y) \\ ab(k_x + ik_y) & -\Delta \end{pmatrix}. \tag{5.14}$$

The energy in each band is:

$$\mathcal{E}(\mathbf{k}) = \pm(a^2b^2k_x^2 + a^2b^2k_y^2 + \Delta^2)^{1/2}. \tag{5.15}$$

The Berry curvature in each band has only a Z component (pointing out of the plane) [170] :

$$\Omega(\mathbf{k}) = \mp\frac{a^2b^2\Delta}{2(a^2b^2k_x^2 + a^2b^2k_y^2 + \Delta^2)^{3/2}}, \tag{5.16}$$

and the components of the metric are found to be [118]:

$$g_{ij}(\mathbf{k}) = \frac{\left((a^2b^2k_x^2 + a^2b^2k_y^2 + \Delta^2)\delta_{ij} - a^2b^2k_ik_j\right)}{4(a^2b^2k_x^2 + a^2b^2k_y^2 + \Delta^2)^2}. \tag{5.17}$$

We can use the metric to directly compute the Christoffel symbols of the first kind. Assuming the applied field is in the x direction, the response to the field is determined by $\Gamma_{k_xk_xk_x} = \frac{1}{2}\partial_{k_x}g_{xx}$ and $\Gamma_{k_xk_xk_y} = \frac{1}{2}\partial_{k_y}g_{xx}$, which can be expressed as:

$$\begin{aligned}\Gamma_{k_xk_xk_x} &= \frac{-a^2b^2k_x(a^2b^2k_y^2 + \Delta^2)}{2(a^2b^2k_x^2 + a^2b^2k_y^2 + \Delta^2)^3}, \\ \Gamma_{k_xk_xk_y} &= \frac{-a^2b^2k_y(a^2b^2k_y^2 - a^2b^2k_x^2 + \Delta^2)}{4(a^2b^2k_x^2 + a^2b^2k_y^2 + \Delta^2)^3}.\end{aligned} \tag{5.18}$$

$\Gamma_{k_xk_xk_x}$ determines the parallel component of the second-order geodesic response, while $\Gamma_{k_xk_xk_y}$ controls the geodesic response orthogonal to the applied field.

For $\mathbf{k}^2 \ll \Delta^2$, the strength of the geodesic response increases linearly with the magnitude of the momentum vector $\mathbf{k}$. This can be contrasted with the response due to Berry curvature, which is effectively constant in this regime. The response due to the dispersion will also depend linearly on $\mathbf{k}$ in this limit, but will be independent of the field strength. The change in the anomalous velocity with $\mathbf{k}$ and $\mathbf{E}$ should thus reveal the role of the QGT connection in the second-order response of electrons.

5.4 Momentum-space Einstein Field Equations

5.4.1 Pure states and vacuum EFE

When quantum states are parameterized by crystal momentum, the momentum space inherits the metric of the underlying Hilbert space. The space of all pure states in Hilbert space can be seen as a high dimensional sphere, while the space of physically distinguishable states is the quotient space obtained by identifying all states on the sphere that differ by a phase factor [5, 42]. This space of physically distinct quantum states is known as the projective Hilbert space of states, and has the

geometry of a complex projective space. All such complex projective spaces possess a canonical Riemannian metric, known as the Fubini-Study metric [110]. This Fubini-Study metric is Einstein, meaning it has a Ricci tensor proportional to itself. As a result, the quantum metric is a vacuum solution of the Einstein field equations in the projective Hilbert space [26]:

$$R_{\mu\nu} - \tfrac{1}{2} R\, g_{\mu\nu} + \Lambda g_{\mu\nu} = 0, \tag{5.19}$$

where $R_{\mu\nu}$ is the Ricci tensor, R is the scalar curvature, and Λ is the cosmological constant.

The momentum space can be thought of as being embedded in the Hilbert space: it parameterizes a submanifold of all possible quantum states $|u_0(\mathbf{k})\rangle$. It also inherits the metric of the Hilbert space, defined via the overlaps of k-dependent Bloch functions: $g_{ij}dk^i dk^j = 1 - |\langle u_0(\mathbf{k})|u_0(\mathbf{k}+\mathbf{dk})\rangle|^2$.

While the momentum space inherits the Hilbert-space metric, the curvature of this subspace may be different due to the reduction in dimension. A similar situation arises in general relativity: the intrinsic curvature on a hypersurface may be different than the curvature of the full spacetime [10]. The momentum-space stress-energy can thus be nonzero due to the extrinsic curvature associated with the embedding of momentum space in Hilbert space. We will see below that this is not the only source of stress-energy in momentum space.

5.4.2 Mixed states, Bures metric and the source of EFE

Here we show that the Bures metric for mixed states can have a nonzero stress-energy tensor even when the pure state metric is a vacuum solution of the EFE. Since all two-dimensional metrics have vanishing stress-energy by construction, we show this for an $N > 2$ band quantum system in which the Bloch waves have three dimensions of crystal momentum (see Supplement).

First we consider the density matrix of a mixed quantum state:

$$\rho = \sum_{n=0}^{N} p_n(\mathbf{k})|u_n(\mathbf{k})\rangle\langle u_n(\mathbf{k})|, \tag{5.20}$$

where $p_n = \frac{e^{-\beta\mathcal{E}_n(\mathbf{k})}}{\mathcal{Z}(\mathbf{k})}$ are the Boltzmann weights for a system at finite temperature. Assuming the probabilities $p_n(\mathbf{k})$ change slowly in k-space and that all the probabilities above the ground state are small, we find that the Bures metric $\bar{g}_{ij}$ takes the following form (see Supplement) [50]:

$$\bar{g}_{ij}dk^i dk^j = e^{-\frac{S(\mathbf{k})}{k_B}} g_{ij}dk^i dk^j. \tag{5.21}$$

We see that the Bures metric differs from the Fubini-Study metric by a conformal scale factor due to entropy. The above equation (5.21) is completely general: it holds for any number of parameters and any number of bands.

If the pure state momentum-space metric is a solution of the vacuum EFE, the trace of the stress-energy tensor for the Bures metric is found to be (see Appendix):

$$T = -\frac{R}{2k_B}S(\mathbf{k}) - \frac{1}{k_B}\Delta_{\mathbf{k}}S(\mathbf{k}). \tag{5.22}$$

More generally, the stress energy will contain additional terms that describe the embedding of momentum-space within the full Hilbert space, but we see that, for mixed states, the stress-energy tensor depends on the entropy and its momentum-space Laplacian.

The stress-energy in (5.22) is analogous to the weak field limit of general relativity in which Newtonian gravity and Poisson's equation are recovered [109]. The Bures metric responds to changes in entropy, just as the spacetime metric responds to changes in the distribution of matter.

5.4.3 Entropy maximization

We can understand the conformal scaling in (5.21) by viewing the resulting change in the Christoffel connection and geodesics [37]:

$$\bar{\Gamma}^k{}_{ij} = \Gamma^k_{ij} - \frac{1}{2}\delta^k_i\partial_j S/k_B - \frac{1}{2}\delta^k_j\partial_i S/k_B + \frac{1}{2}g_{ij}\nabla^k S/k_B. \tag{5.23}$$

The mixed state connection includes additional terms driving geodesics along the gradient of entopy. We can thus understand the paths of least distance for mixed states as having two components: one attempting to minimize the distance associated with the underlying pure states, and another trying to maximize the entropy of the mixed state.

5.4.4 Illustrative example: momentum-space EFE for 3D Dirac fermion

Assuming the charge carrier behaves as a three-dimensional Dirac fermion, we can use the momentum-space Dirac equation as its effective Hamiltonian [49]:

$$H = \begin{bmatrix} \Delta \mathbf{I} & \vec{\sigma} \cdot \mathbf{k} \\ \vec{\sigma} \cdot \mathbf{k} & -\Delta \mathbf{I} \end{bmatrix}. \tag{5.24}$$

The resulting Bloch function is then described by a four-spinor:

$$|u_0(\mathbf{k})\rangle = \frac{|\mathcal{E}(\mathbf{k})|+\Delta}{\sqrt{\mathbf{k}^2+(|\mathcal{E}(\mathbf{k})|+\Delta)^2}} \begin{bmatrix} \frac{\vec{\sigma}\cdot\mathbf{k}}{|\mathcal{E}(\mathbf{k})|+\Delta}\chi_s \\ \chi_s \end{bmatrix}, \tag{5.25}$$

where the state has been normalized to one and $|\mathcal{E}(\mathbf{k})|$ is the absolute value of the energy of the band, $\mathcal{E}(\mathbf{k}) = -\sqrt{\Delta^2 + \mathbf{k}^2}$. While we deal with the negative energy solution here, it is straightforward to show that our expression for the metric also applies to the positive energy band. Note that χ_s is an arbitrary two spinor, and the ground state of the Dirac Hamiltonian is thus degenerate. As shown by Matsuura and Ryu [106], we find the following expression for the quantum metric when $\mathbf{k}^2 \ll \Delta^2$:

$$g_{ij}dk^i dk^j \approx \frac{\delta_{ij}dk^i dk^j}{4\Delta^2} \qquad \text{for } \mathbf{k}^2 \ll \Delta^2. \tag{5.26}$$

In this limit, the pure state metric becomes completely flat and is thus a vacuum solution of the EFE. For a mixed state with finite entropy, the stress-energy is then described by (5.22) with $R = 0$:

$$T \approx -\frac{1}{k_B}\Delta_{\mathbf{k}} S(\mathbf{k}). \tag{5.27}$$

We see that the analogy between the weak field limit of GR and our momentum-space expressions for small entropy becomes even more striking in this regime, with the entropy and gravitational potential acting as a source of stress-energy in momentum space and spacetime respectively.

5.5 Discussion

In this work, we have highlighted the duality between the momentum-space equation for the velocity of a Bloch electron and the position-space equation for the Lorentz force in curved spacetime. While the role of the metric in the second-order dynamics of a Bloch electron has been shown previously, the geodesic nature of these expressions has not previously been explored [68]. The fact that the momentum space Berry connection affects the dynamics of charge carriers is already well accepted, but from the expressions derived here, it is clear that the Levi-Civita connection of the metric plays an important role as well. Deviations from flat momentum space via the Christoffel term in (5.9) can be thought of as arising from the general covariance of the equations of motion under transformations: just as the Berry connection is necessary to account for gauge transformations, the Levi-Civita connection is necessary to account for coordinate transformations [145].

In addition to extending the duality present in the semiclassical formulation of electron dynamics, we have shown that the presence of nonzero entropy for mixed states can lead to the emergence of a stress-energy tensor in momentum space. We have quantified the role of entropy in the creation of momentum space stress-energy, and have shown that the stress-energy is in part due to the k-space Laplacian of the entropy, just as the real space stress-energy is due to the Laplacian of the gravitational potential in the Newtonian limit of general relativity [109]. We can thus view our expression (5.10) for stress-energy in the limit of low entropy as the momentum space analog of the Newtonian limit of general relativity. The momentum space analog of the gravitational potential is the von Neumann entropy of a mixed quantum state, which supports previous speculations on a potentially deep relationship between entropy and gravity [160].

While the analytic expressions developed here further our understanding of Bloch band geometry and its effect on wave packet dynamics, more work needs to be done to verify their validity via simulation and experiments. Another natural question, not studied here, is how finite temperature fluctuations affect the dynamics of a wave packet. As entropy deforms the metric and further curves the momentum space, it

may be possible to detect these finite temperature effects in the trajectories of Bloch electrons.

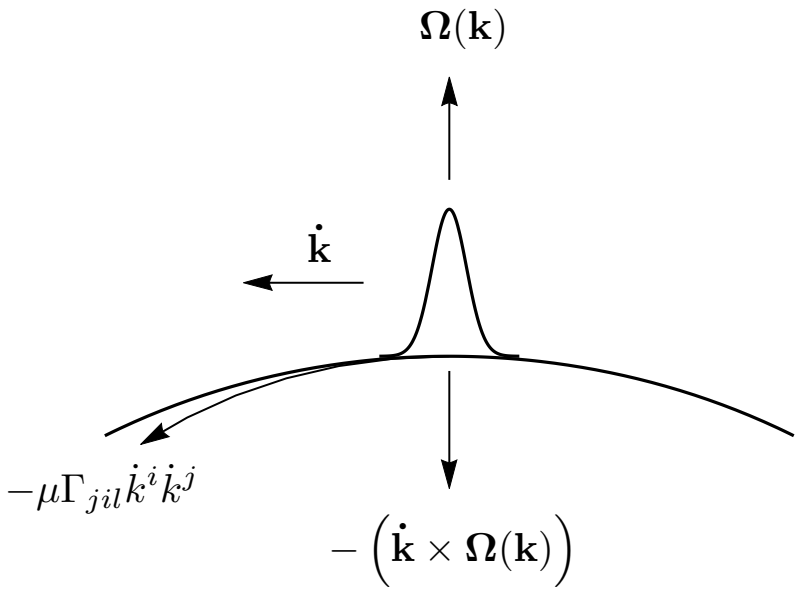

Figure 5.1: **Velocity in momentum space is analogous to force in real space.** The coupling between the electric field and Berry curvature creates a Lorentz force-like term driving the electron in a direction orthogonal to the applied field, while the curvature of the momentum space drives the electron in the direction of k-space geodesics.

Chapter 6

Summary

At both the molecular and population level, stochastic transport and degradation combine to limit the scale of information flow in spatially structured biological systems. At the population level, we extended classic models of isolation by distance to include the effects of heavy-tailed dispersal by replacing the typical thin-tailed dispersal kernel with a power-law dispersal distribution. We found that power-law dispersal leads to power-law decay of relatedness at large distances, and can produce either power-law or logarithmic decay at short distances depending on the exponent of the power law in the dispersal kernel and the dimensionality of space.

The power-law isolation by distance model was then used to solve the inverse problem of inferring dispersal from empirical isolation by distance curves. Our Python dispersal inference code was shown to perform well for simulated data and was then applied to data from a natural population of Florida Scrub-jays, in which we found signatures of heavy-tailed dispersal. We also extended the model to predict the size and number of long shared sequence blocks and developed an inference scheme for this generalization. This second code was used to analyze sequence block data from European humans, and additional signatures of heavy-tailed dispersal were detected for this population.

We also investigated models of cellular gradient sensing in two and three spatial dimensions and found that the effects of spatial averaging can be counterintuitive. We showed that, unlike for concentration sensing, the precision of gradient sens-

ing decreases with transverse length for the simplest gradient sensing model, local excitation–global inhibition. In two dimensions, increasing the width of a detector transverse to a concentration gradient can decrease the precision with which the gradient is measured. While transverse averaging will always improve concentration sensing, gradient sensing depends on the covariance between two distinct concentrations measurements, and the decrease of this covariance with transverse averaging decreases precision. In order to reverse this effect, the local reporter molecule must also be allowed to disperse. We studied this generalization with the recently developed regional excitation-global inhibition (REGI) model and show that, for REGI, the scaling of precision with transverse detector width is non-monotonic and leads to optimal detector shapes in two and three dimensions.

Appendix A

Alternative Derivations of the Probability of Identity by Descent

A.1 Starting from a recursion equation

(3.26) can also be derived starting from a recursion equation for ψ requiring that it remain constant over an infinitesimal timestep dt [15, 100]:

$$\psi(x) = \frac{dt}{\rho}(1-\psi(x))\mathcal{N}(x) + e^{-2\mu dt}\int dy\, \psi(x-y)K(y|dt). \tag{A.1}$$

(A.1) is saying that at equilibrium the local increase in identity due to coalescence (first term) must be balanced by the loss of identity due to mutation and the spreading of identity due to dispersal (both included in the second term).

Taking the spatial Fourier transform $\mathcal{F}\{\cdot\}$ of (A.1) simplifies the second term at the expense of complicating the first:

$$\widehat{\psi}(k) = \frac{dt}{\rho}\mathcal{F}\{(1-\psi(x))\mathcal{N}(x)\}(k) + e^{-2\mu dt}\widehat{\psi}(k)\widehat{K}(k|dt).$$

Solving for $\widehat{\psi}$ gives:

$$\begin{aligned}\widehat{\psi}(k) &= \frac{\mathcal{F}\{(1-\psi(x))\mathcal{N}(x)\}(k)}{\rho}\frac{dt}{1-e^{-2\mu dt}\widehat{K}(k|dt)} \\ &= \frac{\mathcal{F}\{(1-\psi(x))\mathcal{N}(x)\}(k)}{\rho}\sum_{j=0}^{\infty} dt\, e^{-2\mu j dt}\widehat{K}(k|dt)^j,\end{aligned} \tag{A.2}$$

where in the second line we can take the Taylor series expansion because $e^{-2\mu dt} < 1$ and $\widehat{K}(k|dt) \leq 1$ since it is a characteristic function. Assuming dispersal is Markovian, we can simplify (A.2) by noting that $\widehat{K}(k|dt)^j = \widehat{K}(k|jdt)$, i.e., the distribution after time jdt is just the convolution of j dispersal steps of time dt each. Using this, we can convert the sum into an integral to find (3.26):

$$\begin{aligned}\widehat{\psi}(k) &= \frac{\mathcal{F}\{(1-\psi(x))\mathcal{N}(x)\}(k)}{\rho}\widehat{\widehat{K}}(k,2\mu) \\ &\approx \frac{1-\psi(0)}{\rho}\widehat{\mathcal{N}}(k)\widehat{\widehat{K}}(k,2\mu),\end{aligned}$$

where in the second line we have used the same approximation that $1-\psi(x) \approx 1-\psi(0)$ for $|x| \lesssim \delta$ that we used in the main text.

A.2 Fractional diffusion equation

For Lévy flight dispersal, (3.27) can also be derived using a fractional diffusion equation. When X_t follows a diffusion, (3.18) can be written as a Feynman-Kac (diffusion) equation for ψ [4, 15]. For $\alpha < 2$, this generalizes to a fractional differential equation:

$$0 = 2D_\alpha \left(\frac{\partial^2}{\partial x^2}\right)^{\alpha/2} \psi(x) - 2\mu\psi(x) + \frac{1}{\rho}\mathcal{N}(x)\left(1-\psi(x)\right), \tag{A.3}$$

where $\left(\frac{\partial^2}{\partial x^2}\right)^{\alpha/2}$ is a Riesz fractional derivative, defined by its Fourier transform $\mathcal{F}\left\{\left(\frac{\partial^2}{\partial x^2}\right)^{\alpha/2} f\right\}(k) = -|k|^\alpha \mathcal{F}\{f\}(k)$ [36, 85, 108]. It is therefore simpler to consider the Fourier transform of (A.3), which is equivalent to (3.27):

$$\begin{aligned}0 &= -(2D_\alpha|k|^\alpha + 2\mu)\widehat{\psi}(k) + \frac{1}{\rho}\mathcal{F}\left\{\mathcal{N}(x)(1-\psi(x))\right\}(k) \\ &\approx -(2D_\alpha|k|^\alpha + 2\mu)\widehat{\psi}(k) + \frac{1-\psi(0)}{\rho}\widehat{\mathcal{N}}(k).\end{aligned} \tag{A.4}$$

For all $\alpha < 2$, the solution for ψ in two dimensions can be also written as a fractional differential equation [41]:

$$0 = 2D_\alpha \left(\frac{\partial^2}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x}\right)^{\frac{\alpha}{2}} \psi(x) - 2\mu\psi(x) + \frac{1}{\rho}\mathcal{N}(x)\left(1-\psi(x)\right), \tag{A.5}$$

where $\left(\frac{\partial^2}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x}\right)^{\frac{\alpha}{2}}$ is a fractional Laplacian, defined by its Fourier transform $\mathcal{F}\left\{\left(\frac{\partial^2}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x}\right)^{\alpha/2}\right\} = -|k|^{\alpha}\mathcal{F}\{f\}(k)$ [93, 98]. Note that the rotational symmetry of the problem allows us to write the Laplacian in terms of just the radial coordinate x, and ignore the angular coordinate. The two-dimensional Fourier transform of (A.5) has exactly the same form as (A.4), although again the interpretation is different. k is now the radial coordinate in two-dimensional k-space, i.e., the magnitude of the two-dimensional frequency vector.

Appendix B

Perturbative Corrections and Geometric Calculations

B.1 Perturbative corrections to energy and Berry connection

We begin with a Hamiltonian parameterized by crystal momentum and perturb it with a potential term due to a constant electric field:

$$H(\mathbf{k}) = H_0(\mathbf{k}) + \mathbf{E} \cdot \mathbf{r}. \tag{B.1}$$

Note that we have set the magnitude of the electron charge e to one for simplicity. The perturbative correction to our 2 band Hamiltonian due to a constant electric field can be represented as:

$$H' \approx \begin{pmatrix} \mathbf{E} \cdot \langle u_0(\mathbf{k})|\mathbf{r}|u_0(\mathbf{k})\rangle & \mathbf{E} \cdot \mathcal{A}_{01}(\mathbf{k}) \\ \mathbf{E} \cdot \mathcal{A}_{10}(\mathbf{k}) & \mathbf{E} \cdot \langle u_1(\mathbf{k})|\mathbf{r}|u_1(\mathbf{k})\rangle \end{pmatrix}. \tag{B.2}$$

In this gauge we can use time independent perturbation theory to compute the quantites that produce second-order effects in the semiclassical theory: the first-order correction to the Berry connection and the second-order correction to the energy [68].

We first find the first-order correction to the wave function, which is required to

find the correction to the Berry connection [72]:

$$|u_0'(\mathbf{k})\rangle = \frac{-\mathbf{E}\cdot\langle u_1(\mathbf{k})|\mathbf{r}|u_0(\mathbf{k})\rangle}{\mathcal{E}_1-\mathcal{E}_0}|u_1(\mathbf{k})\rangle. \tag{B.3}$$

Noting that the quantity $\langle u_1|\mathbf{r}|u_0\rangle$ can be rewritten as the interband Berry connection, $\mathcal{A}_{10} = \langle u_1|i\nabla_{\mathbf{k}}|u_0\rangle$, we can re-express the above correction to the ground-state wave function as [69]:

$$|u_0'(\mathbf{k})\rangle = \frac{-\mathbf{E}\cdot\mathcal{A}_{10}(\mathbf{k})}{\mathcal{E}_1-\mathcal{E}_0}|u_1(\mathbf{k})\rangle. \tag{B.4}$$

.

We can then find the first-order correction to the Berry connection, which is defined as [68]:

$$a'(\mathbf{k}) = \langle u_0(\mathbf{k})|i\nabla_{\mathbf{k}}|u_0'(\mathbf{k})\rangle + c.c. \tag{B.5}$$

Using (B.4) for $|u_0(\mathbf{k})'\rangle$, (B.5) becomes:

$$a'(\mathbf{k}) = \frac{-\mathcal{A}_{01}(\mathbf{k})\mathcal{A}_{10}(\mathbf{k})\cdot\mathbf{E}}{\mathcal{E}_1-\mathcal{E}_0} + c.c. \tag{B.6}$$

Noting that the outer product $\mathcal{A}_{01}\mathcal{A}_{10}$ is equivalent to the quantum metric tensor, we can re-expresss the above equation in terms of the metric [68]:

$$a_i' = \frac{-2g_{ij}E^j}{\mathcal{E}_1-\mathcal{E}_0}. \tag{B.7}$$

We can again apply the relations between the interband elements of $\mathbf{r}$, the interband Berry connection, and the quantum metric to the standard perturbation theory formulas to get a geometric expression for the second-order energy correction [68]:

$$\mathcal{E}'' = \frac{-g_{ij}E^iE^j}{\mathcal{E}_1-\mathcal{E}_0}. \tag{B.8}$$

B.2 Calculating the metric and stress-energy for mixed states

The quantum metric for pure states is an Einstein metric in any dimension, and is thus a vacuum solution of the Einstein field equations [26]. Here we show the that

the Bures metric for mixed states need not be a vacuum solution, and can instead satisfy the Einstein field equations with a nonzero stress-energy tensor. Given that the two-dimensional case is trivial, with all metrics satisfying the vacuum equations with zero cosmological constant, we show this for a three dimensional momentum space. First we consider the density matrix of a mixed state:

$$\rho = \sum_{n=0}^{N} p_n(\mathbf{k}) |u_n(\mathbf{k})\rangle\langle u_n(\mathbf{k})|, \tag{B.9}$$

where we define the Boltzmann weights as $p_n = \frac{e^{-\beta\mathcal{E}_n(\mathbf{k})}}{\mathcal{Z}(\mathbf{k})}$. The difference between mixed state density matrices at nearby points in momentum space, $d\rho$, is found to be:

$$\langle u_i(\mathbf{k})|d\rho|u_j(\mathbf{k})\rangle = (d\mathbf{k}\cdot\nabla_{\mathbf{k}} p_i)\,\delta_{ij} + i(p_i - p_j)\mathcal{A}_{ij}\cdot d\mathbf{k}, \tag{B.10}$$

where $\mathcal{A}_{ij}$ is the interband Berry connection.

Using (B.10) we can calculate the distance between mixed states as defined by the Bures metric [50]:

$$\bar{g}_{ij}dk^i dk^j = \frac{1}{2}\sum_{j,k=0}^{N} \frac{|\langle j|d\rho|k\rangle|^2}{p_j + p_k}, \tag{B.11}$$

which can be re-expressed as:

$$\bar{g}_{ij}dk^i dk^j = \frac{1}{2}\sum_{j,k=0}^{N} \frac{|\mathbf{dk}\cdot\nabla_{\mathbf{k}} p_j|^2}{2p_j}\delta_{jk} + \frac{(p_j - p_k)^2}{p_j + p_k}|\mathcal{A}_{jk}\cdot d\mathbf{k}|^2. \tag{B.12}$$

Assuming the probabilities $p_n(\mathbf{k})$ change slowly in k-space, the above expression reduces to:

$$\bar{g}_{ij}dk^i dk^j \approx \frac{1}{2}\sum_{j,k=0}^{N} \frac{(p_j - p_k)^2}{p_j + p_k}|\mathcal{A}_{jk}\cdot d\mathbf{k}|^2. \tag{B.13}$$

Further assuming that all bands above the lowest band have degenerate energies reduces (B.13) to:

$$\bar{g}_{ij}dk^i dk^j \approx p_0\left(\frac{(1 - p_1/p_0)^2}{1 + p_1/p_0}\right) g_{ij}dk^i dk^j, \tag{B.14}$$

where we have once again re-expressed the outer product of the interband Berry connections as the pure state quantum metric [69]. When $p_1/p_0 \ll 1$, we can approximate

the prefactor $p_0\left(\frac{(1-p_1/p_0)^2}{1+p_1/p_0}\right)$ in (B.14) as $p_0 = e^{-S/k_B}$:

$$\bar{g}_{ij}dk^i dk^j \approx e^{-S/k_B} g_{ij}dk^i dk^j. \tag{B.15}$$

We see that the mixed state metric differs from the pure state metric by a conformal scale factor due to entropy. Because of this simple conformal relation, the scalar curvature of the Bures metric can be expressed in terms of the scalar curvature of the pure state metric as [26]:

$$\bar{R} = e^{S/k_B}\left(R + 2\Delta_{\mathbf{k}} S/k_B - \sum_i |\partial_{k_i} S/k_B|^2\right), \tag{B.16}$$

where $\Delta_{\mathbf{k}}$ is the Laplace-Beltrami operator associated with the curved momentum space. Note that we are using the physics convention in which Δ has a positive sign rather than the math convention that includes an extra factor of -1.

Using the trace of the Einstein field equations, we can find the trace of the stress tensor:

$$-\frac{1}{2}R - \frac{1}{2}R' + 3\Lambda = T, \tag{B.17}$$

where we have set the typical prefactor of the stress-energy Tensor, $\frac{8\pi G}{c^4}$, to one. The trace of the stress-energy tensor for the Bures metric is thus:

$$T = -\frac{1}{2}e^{S/k_B}\left(R + 2\Delta_{\mathbf{k}} S/k_B - \sum_i |\partial_{k_i} S/k_B|^2\right) + \frac{1}{2}R. \tag{B.18}$$

Assuming S and its first derivatives are small, (B.18) reduces to:

$$T \approx -\frac{R}{2k_B} S(\mathbf{k}) - \Delta_{\mathbf{k}} S(\mathbf{k})/k_B. \tag{B.19}$$

B.3 Calculating the Metric and Curvature for a Dirac Fermion

Assuming the charge carrier behaves as a three-dimensional Dirac fermion, we can use the momentum-space Dirac equation as its effective Hamiltonian. The resulting

Bloch function is then described by a four-spinor:

$$|u_0(\mathbf{k})\rangle = \frac{|\mathcal{E}(\mathbf{k})|+\Delta}{\sqrt{\mathbf{k}^2+(|\mathcal{E}(\mathbf{k})|+\Delta)^2}} \begin{bmatrix} \frac{\vec{\sigma}\cdot\mathbf{k}}{|\mathcal{E}(\mathbf{k})|+\Delta}\chi_s \\ \chi_s \end{bmatrix}, \tag{B.20}$$

where the state has been normalized to one and $|\mathcal{E}(\mathbf{k})|$ is the absolute value of the energy of the band, $\mathcal{E}(\mathbf{k}) = -\sqrt{\Delta^2+\mathbf{k}^2}$. Note that χ_s is an arbitrary two spinor, and the ground state of the Dirac Hamiltonian is thus degenerate. While the non-abelian quantum geometric tensor is needed to describe the complete case of degenerate bands, here we focus on the momentum space distance between electron states with the same spin polarization, χ_s. In this case, the distance in momentum space can be determined from the standard abelian quantum geometric tensor, which itself can be determined from the overlap of the Bloch states via $g_{ij}dk^i dk^j = 1 - |\langle u_0(\mathbf{k})|u_0(\mathbf{k}+\mathbf{dk})\rangle|^2$.

Noting the Pauli matrix identity $(\vec{a}\cdot\vec{\sigma})\left(\vec{b}\cdot\vec{\sigma}\right) = \left(\vec{a}\cdot\vec{b}\right) I + i\left(\vec{a}\times\vec{b}\right)\cdot\vec{\sigma}$ and defining $\hat{\mathbf{s}}$ as the unit vector in the direction of spin polarization, the overlap integral $\langle u_0(\mathbf{k})|u_0(\mathbf{k}')\rangle$ is found to be:

$$\langle u_0(\mathbf{k})|u_0(\mathbf{k}')\rangle = \frac{(|\mathcal{E}(\mathbf{k})|+\Delta)(|\mathcal{E}'(\mathbf{k}')|+\Delta)}{\sqrt{\left(\mathbf{k}^2+(|\mathcal{E}(\mathbf{k})|+\Delta)^2\right)\left(\mathbf{k}'^2+(|\mathcal{E}'(\mathbf{k}')|+\Delta)^2\right)}}\left(1+\frac{\mathbf{k}\cdot\mathbf{k}'+i(\mathbf{k}\times\mathbf{k}')\cdot\hat{\mathbf{s}}}{(\mathcal{E}'(\mathbf{k}')+\Delta)(\mathcal{E}(\mathbf{k})+\Delta)}\right). \tag{B.21}$$

For the squared magnitude of the overlap we find:

$$|\langle u_0(\mathbf{k})|u_0(\mathbf{k}')\rangle|^2 = \frac{(|\mathcal{E}(\mathbf{k})|+\Delta)^2(|\mathcal{E}'(\mathbf{k}')|+\Delta)^2+2(|\mathcal{E}(\mathbf{k})|+\Delta)(|\mathcal{E}'(\mathbf{k}')|+\Delta)+(\mathbf{k}\cdot\mathbf{k}')^2+((\mathbf{k}\times\mathbf{k}')\cdot\hat{\mathbf{s}})^2}{\left(\mathbf{k}^2+(|\mathcal{E}(\mathbf{k})|+\Delta)^2\right)\left(\mathbf{k}'^2+(|\mathcal{E}'(\mathbf{k}')|+\Delta)^2\right)}, \tag{B.22}$$

which leads to:

$$1-|\langle u_0(\mathbf{k})|u_0(\mathbf{k}')\rangle|^2 = \frac{|\mathbf{k}\times\mathbf{k}'|^2-((\mathbf{k}\times\mathbf{k}')\cdot\hat{\mathbf{s}})^2+|(|\mathcal{E}'(\mathbf{k}')|+\Delta)\mathbf{k}-(|\mathcal{E}(\mathbf{k})|+\Delta)\mathbf{k}'|^2}{\left(\mathbf{k}^2+(|\mathcal{E}(\mathbf{k})|+\Delta)^2\right)\left(\mathbf{k}'^2+(|\mathcal{E}'(\mathbf{k}')|+\Delta)^2\right)}. \tag{B.23}$$

The quantity above can then be re-expressed as:

$$1-|\langle u_0(\mathbf{k})|u_0(\mathbf{k}')\rangle|^2 = \frac{|\frac{\mathbf{k}}{|\mathcal{E}(\mathbf{k})|+\Delta}\times\frac{\mathbf{k}'}{|\mathcal{E}'(\mathbf{k}')|+\Delta}|^2-\left((\frac{\mathbf{k}}{|\mathcal{E}(\mathbf{k})|+\Delta}\times\frac{\mathbf{k}'}{|\mathcal{E}'(\mathbf{k}')|+\Delta})\cdot\hat{\mathbf{s}}\right)^2+\left|\frac{\mathbf{k}}{|\mathcal{E}(\mathbf{k})|+\Delta}-\frac{\mathbf{k}'}{|\mathcal{E}'(\mathbf{k}')|+\Delta}\right|^2}{\left(1+\frac{\mathbf{k}^2}{(|\mathcal{E}(\mathbf{k})|+\Delta)^2}\right)\left(1+\frac{\mathbf{k}'^2}{(|\mathcal{E}'(\mathbf{k}')|+\Delta)^2}\right)}. \tag{B.24}$$

We assume $\mathbf{k}^2 \ll \Delta^2$ and expand $1 - |\langle u_0(\mathbf{k})|u_0(\mathbf{k}')\rangle|^2$ in powers of $\frac{\mathbf{k}}{\Delta}$ and $\frac{\mathbf{k}'}{\Delta}$. Keeping only terms of lowest order, we find:

$$1 - |\langle u_0(\mathbf{k})|u_0(\mathbf{k}')\rangle|^2 \approx \left|\frac{\mathbf{k}}{2\Delta} - \frac{\mathbf{k}'}{2\Delta}\right|^2 \qquad \text{for } \mathbf{k}^2 \ll \Delta^2. \tag{B.25}$$

To obtain the metric, we can assume that the separation vector $\mathbf{k}' - \mathbf{k} = \mathbf{dk}$, i.e., that the separation of states in momentum space is infinitesimal. We then find an expression for the quantum geometric tensor when $\mathbf{k}^2 \ll \Delta$:

$$g_{ij}dk^i dk^j \approx \frac{\delta_{ij}dk^i dk^j}{4\Delta^2} \qquad \text{for } \mathbf{k}^2 \ll \Delta^2. \tag{B.26}$$

We see that the metric is independent of the spin polarization of the fermion states. (B.26) agrees with the expression found previously by Matsuura and Ryu [106] for the quantum metric of the Dirac fermion.

Bibliography

[1] W. Adams, A. Griffin, and G. Moran. Using paternity analysis to measure effective pollen dispersal in plant populations. *The American Naturalist*, 140 (5):762–780, 1992.

[2] S. M. Aguillon, J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, G. Coop, and N. Chen. Deconstructing isolation-by-distance: the genomic consequences of limited dispersal. *PLoS genetics*, 13(8):e1006911, 2017.

[3] H. Al-Asadi, D. Petkova, M. Stephens, and J. Novembre. Estimating recent migration and population-size surfaces. *PLoS genetics*, 15(1):e1007908, 2019.

[4] B. E. Allman and D. B. Weissman. Hitchhiking in space: Ancestry in adapting, spatially extended populations. *Evolution*, 72(4):722–734, 2018.

[5] J. Anandan and Y. Aharonov. Geometry of quantum evolution. *Physical review letters*, 65(14):1697, 1990.

[6] N. W. Ashcroft, N. D. Mermin, et al. Solid state physics [by] neil w. ashcroft [and] n. david mermin., 1976.

[7] R. Atkinson, C. Rhodes, D. Macdonald, and R. Anderson. Scale-free dynamics in the movement patterns of jackals. *Oikos*, 98(1):134–140, 2002.

[8] F. Austerlitz, C. W. Dick, C. Dutech, E. K. Klein, S. Oddou-Muratorio, P. E. Smouse, and V. L. Sork. Using genetic markers to estimate the pollen dispersal curve. *Molecular ecology*, 13(4):937–954, 2004.

[9] C. F. Bacles and R. A. Ennos. Paternity analysis of pollen-mediated gene flow for *Fraxinus excelsior* L. in a chronically fragmented landscape. *Heredity*, 101 (4):368–380, 2008.

[10] J. C. Baez and J. P. Muniain. *Gauge fields, knots and gravity*, volume 4. World Scientific Publishing Company, 1994.

[11] M. Baguette. Long distance dispersal and landscape occupancy in a metapopulation of the cranberry fritillary butterfly. *Ecography*, 26(2):153–160, 2003.

[12] N. Barton. Multilocus clines. *Evolution*, pages 454–471, 1983.

[13] N. H. Barton and G. M. Hewitt. Analysis of hybrid zones. *Annual review of Ecology and Systematics*, 16(1):113–148, 1985.

[14] N. H. Barton and I. Wilson. Genealogies and geography. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 349(1327): 49–59, 1995.

[15] N. H. Barton, F. Depaulis, and A. M. Etheridge. Neutral evolution in spatially continuous populations. *Theoretical population biology*, 61(1):31–48, 2002.

[16] N. H. Barton, J. Kelleher, and A. M. Etheridge. A new model for extinction and recolonization in two dimensions: Quantifying phylogeography. *Evolution*, 64(9):2701 – 2715, 2010.

[17] N. H. Barton, A. M. Etheridge, J. Kelleher, and A. Véber. Inference in two dimensions: allele frequencies versus lengths of shared sequence blocks. *Theoretical population biology*, 87:105–119, 2013.

[18] N. H. Barton, A. M. Etheridge, J. Kelleher, and A. Véber. Genetic hitchhiking in spatially extended populations. *Theoretical Population Biology*, 87:75 – 89, 2013. ISSN 0040-5809. doi: 10.1016/j.tpb.2012.12.001.

[19] N. H. Barton, A. M. Etheridge, and A. Véber. Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, 2013 (01):P01002, 2013.

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

[21] H. C. Berg. *Random walks in biology*. Princeton University Press, 1993.

[22] H. C. Berg and E. M. Purcell. Physics of chemoreception. *Biophysical journal*, 20(2):193–219, 1977.

[23] T. W. Berngruber, S. Lion, and S. Gandon. Spatial structure, transmission modes and the evolution of viral exploitation strategies. *PLoS Pathog*, 11(4): e1004810, 2015.

[24] M. V. Berry. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 392(1802):45–57, 1984.

[25] J. Bertoin. *Lévy processes*, volume 121. Cambridge university press Cambridge, 1996.

[26] A. L. Besse. *Einstein manifolds*. Springer Science & Business Media, 2007.

[27] W. Bialek. *Biophysics: searching for principles*. Princeton University Press, 2012.

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

[29] A. Bianco, M. Poukkula, A. Cliffe, J. Mathieu, C. M. Luque, T. A. Fulga, and P. Rørth. Two distinct modes of guidance signalling during collective migration of border cells. *Nature*, 448(7151):362–365, 2007.

[30] G. S. Bradburd and P. L. Ralph. Spatial population genetics: it's about time. *Annual Review of Ecology, Evolution, and Systematics*, 2019.

[31] G. S. Bradburd, P. L. Ralph, and G. M. Coop. A spatial framework for understanding population structure and admixture. *PLoS genetics*, 12(1):e1005703, 2016.

[32] G. S. Bradburd, G. M. Coop, and P. L. Ralph. Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1):33–52, 2018.

[33] D. Brockmann and L. Hufnagel. Front propagation in reaction-superdiffusion dynamics: Taming Lévy flights with fluctuations. *Physical review letters*, 98 (17):178301, 2007.

[34] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462, 2006.

[35] B. A. Camley, J. Zimmermann, H. Levine, and W.-J. Rappel. Emergent collective chemotaxis without single-cell gradient sensing. *Physical review letters*, 116(9):098101, 2016.

[36] S. Carmi, L. Turgeman, and E. Barkai. On distributions of functionals of anomalous diffusion paths. *Journal of Statistical Physics*, 141(6):1071–1092, 2010.

[37] D. F. Carneiro, E. A. Freiras, B. Gonçalves, A. G. de Lima, and I. L. Shapiro. On useful conformal tranformations in general relativity. *arXiv preprint gr-qc/0412113*, 2004.

[38] H. Cayuela, Q. Rougemont, J. G. Prunier, J.-S. Moore, J. Clobert, A. Besnard, and L. Bernatchez. Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Molecular ecology*, 27(20): 3976–4010, 2018.

[39] D. Chandler. Introduction to modern statistical. *Mechanics. Oxford University Press, Oxford, UK*, 40, 1987.

[40] J. Chave and E. G. Leigh Jr. A spatially explicit neutral model of β-diversity in tropical forests. *Theoretical population biology*, 62(2):153–168, 2002.

[41] X. Chen, Y. Hu, and J. Song. Feynman-Kac formula for fractional heat equation driven by fractional white noise. *arXiv preprint arXiv:1203.0477*, 2012.

[42] R. Cheng. Quantum geometric tensor (fubini-study metric) in simple quantum system: a pedagogical introduction. *arXiv preprint arXiv:1012.1337*, 2010.

[43] M. R. Chernick. *Bootstrap methods: A guide for practitioners and researchers*, volume 619. John Wiley & Sons, 2011.

[44] M. Claassen, C. H. Lee, R. Thomale, X.-L. Qi, and T. P. Devereaux. Position-momentum duality and fractional quantum hall effect in chern insulators. *Physical review letters*, 114(23):236802, 2015.

[45] J. S. Clark. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *The American Naturalist*, 152(2):204–224, 1998.

[46] X. Dai, G. Shannon, R. Slotow, B. Page, and K. J. Duffy. Short-duration daytime movements of a cow herd of African elephants. *Journal of Mammalogy*, 88(1):151–157, 2007.

[47] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge University Press, 1997.

[48] C. Devaux, C. Lavigne, F. Austerlitz, and E. Klein. Modelling and estimating pollen movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. *Molecular Ecology*, 16(3):487–499, 2007.

[49] P. A. M. Dirac. *The principles of quantum mechanics*. Number 27. Oxford university press, 1981.

[50] J. Dittmann. Explicit formulae for the bures metric. *Journal of Physics A: Mathematical and General*, 32(14):2663, 1999.

[51] E. Donà, J. D. Barry, G. Valentin, C. Quirin, A. Khmelinskii, A. Kunze, S. Durdu, L. R. Newton, A. Fernandez-Minan, W. Huber, et al. Directional tissue migration through a self-generated chemokine gradient. *Nature*, 503(7475): 285–289, 2013.

[52] D. Ellison, A. Mugler, M. D. Brennan, S. H. Lee, R. J. Huebner, E. R. Shamir, L. A. Woo, J. Kim, P. Amar, I. Nemenman, et al. Cell–cell communication enhances the capacity of cell ensembles to sense shallow gradients during morphogenesis. *Proceedings of the National Academy of Sciences*, 113(6):E679–E688, 2016.

[53] R. G. Endres and N. S. Wingreen. Accuracy of direct gradient sensing by single cells. *Proceedings of the National Academy of Sciences*, 105(41):15749–15754, 2008.

[54] R. G. Endres and N. S. Wingreen. Maximum likelihood and the single receptor. *Physical review letters*, 103(15):158101, 2009.

[55] T. Erdmann, M. Howard, and P. R. Ten Wolde. Role of spatial averaging in the precision of gene expression patterns. *Physical review letters*, 103(25):258101, 2009.

[56] A. J. Ewald, A. Brenot, M. Duong, B. S. Chan, and Z. Werb. Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. *Developmental cell*, 14(4):570–581, 2008.

[57] J. Fayard, E. K. Klein, and F. Lefèvre. Long distance dispersal and the fate of a gene from the colonization front. *Journal of evolutionary biology*, 22(11): 2171–2182, 2009.

[58] W. Feller. *An introduction to probability theory and its applications*, volume 2. Wiley, 2nd edition, 1971.

[59] J. Felsenstein. A pain in the torus: some difficulties with models of isolation by distance. *The American Naturalist*, 109(967):359–368, 1975.

[60] R. Forien. Isolation by distance patterns arising from short range and long range dispersal–a forwards in time approach. *arXiv preprint arXiv:1907.07930*, 2019.

[61] F. Freimuth, S. Blügel, and Y. Mokrousov. Geometrical contributions to the exchange constants: Free electrons with spin-orbit interaction. *Physical Review B*, 95(18):184428, 2017.

[62] E. Frey and K. Kroy. Brownian motion: a paradigm of soft matter and biological physics. *Annalen der Physik*, 14(1-3):20–50, 2005.

[63] Z. Fric and M. Konvicka. Dispersal kernels of butterflies: power-law functions are invariant to marking frequency. *Basic and Applied Ecology*, 8(4):377–386, 2007.

[64] P. Friedl and D. Gilmour. Collective cell migration in morphogenesis, regeneration and cancer. *Nature reviews Molecular cell biology*, 10(7):445–457, 2009.

[65] P.-A. Gagnaire, T. Broquet, D. Aurelle, F. Viard, A. Souissi, F. Bonhomme, S. Arnaud-Haond, and N. Bierne. Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary applications*, 8(8):769–786, 2015.

[66] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual*. Network Theory Ltd., 3rd edition, 2009.

[67] M. E. Gallagher, C. B. Brooke, R. Ke, and K. Koelle. Causes and consequences of spatial within-host viral spread. *Viruses*, 10(11):627, 2018.

[68] Y. Gao, S. A. Yang, and Q. Niu. Field induced positional shift of bloch electrons and its dynamical implications. *Physical review letters*, 112(16):166601, 2014.

[69] Y. Gao, S. A. Yang, and Q. Niu. Geometrical effects in orbital magnetic susceptibility. *Physical Review B*, 91(21):214405, 2015.

[70] G. J. Goodhill and J. S. Urbach. Theoretical analysis of gradient detection by growth cones. *Journal of neurobiology*, 41(2):230–241, 1999.

[71] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek. Probing the limits to positional information. *Cell*, 130(1):153–164, 2007.

[72] D. J. Griffiths and D. F. Schroeter. *Introduction to quantum mechanics*. Cambridge University Press, 2018.

[73] S. Guindon, H. Guo, and D. Welch. Demographic inference under the coalescent in a spatial continuum. *Theoretical population biology*, 111:43–50, 2016.

[74] F. Haldane. Geometrical description of the fractional quantum hall effect. *Physical review letters*, 107(11):116801, 2011.

[75] O. Hallatschek and D. S. Fisher. Acceleration of evolutionary spread by long-range dispersal. *Proceedings of the National Academy of Sciences*, 111(46): E4911–E4919, 2014.

[76] K. Harris and R. Nielsen. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS genetics*, 9(6), 2013.

[77] J. Hey. Isolation with migration models for more than two populations. *Molecular biology and evolution*, 27(4):905–920, 2010.

[78] G. Hornung, B. Berkowitz, and N. Barkai. Morphogen gradient formation in a complex environment: an anomalous diffusion model. *Physical Review E*, 72 (4):041916, 2005.

[79] B. Hu, W. Chen, W.-J. Rappel, and H. Levine. Physical limits on cellular sensing of spatial gradients. *Physical review letters*, 105(4):048104, 2010.

[80] R. R. Hudson et al. Gene genealogies and the coalescent process. *Oxford surveys in evolutionary biology*, 7(1):44, 1990.

[81] R. J. Huebner and A. J. Ewald. *Cellular foundations of mammary tubulogenesis*, volume 31. 2014.

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

[83] P. A. Iglesias and P. N. Devreotes. Navigating through models of chemotaxis. *Current opinion in cell biology*, 20(1):35–40, 2008.

[84] P. A. Iglesias and A. Levchenko. Modeling the cell's guidance system. *Science Signaling*, 2002(148):re12–re12, 2002.

[85] D. Janakiraman. Lévy flights in the presence of a point sink of finite strength. *Physical Review E*, 95(1):012154, 2017.

[86] S. Jespersen, R. Metzler, and H. C. Fogedby. Lévy flights in external force fields: Langevin and fractional Fokker-Planck equations and their solutions. *Physical Review E*, 59(3):2736, 1999.

[87] A. Jilkine and L. Edelstein-Keshet. A comparison of mathematical models for polarization of single eukaryotic cells in response to guided cues. *PLoS Comput Biol*, 7(4):e1001121, 2011.

[88] A. G. Jones and W. R. Ardren. Methods of parentage analysis in natural populations. *Molecular ecology*, 12(10):2511–2523, 2003.

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

[90] W. D. Koenig, D. Van Vuren, and P. N. Hooge. Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in ecology & evolution*, 11(12):514–517, 1996.

[91] K. S. Korolev, M. Avlund, O. Hallatschek, and D. R. Nelson. Genetic demixing and evolution in linear stepping stone models. *Reviews of modern physics*, 82 (2):1691, 2010.

[92] K. Kruse and A. Iomin. Superdiffusion of morphogens by receptor-mediated transport. *New Journal of Physics*, 10(2):023019, 2008.

[93] M. Kwaśnicki. Ten equivalent definitions of the fractional Laplace operator. *Fractional Calculus and Applied Analysis*, 20(1):7–51, 2017.

[94] A. D. Lander. Pattern, growth, and control. *Cell*, 144(6):955–969, 2011.

[95] P. Legendre and M. J. Fortin. Spatial pattern and ecological analysis. *Vegetatio*, 80(2):107–138, 1989.

[96] A. Levchenko and P. A. Iglesias. Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. *Biophysical journal*, 82(1): 50–63, 2002.

[97] H. Levine and W.-J. Rappel. The physics of eukaryotic chemotaxis. *Physics today*, 66(2), 2013.

[98] A. Lischke, G. Pang, M. Gulian, F. Song, C. Glusa, X. Zheng, Z. Mao, W. Cai, M. M. Meerschaert, M. Ainsworth, et al. What is the fractional Laplacian? A comparative review with new results. *Journal of Computational Physics*, 404: 109009, 2020.

[99] E. Lundgren and P. L. Ralph. Are populations like a circuit? comparing isolation by resistance to a new coalescent-based method. *Molecular ecology resources*, 19(6):1388–1406, 2019.

[100] G. Malécot. Heterozygosity and relationship in regularly subdivided populations. *Theoretical population biology*, 8(2):212–241, 1975.

[101] G. Malet-Engra, W. Yu, A. Oldani, J. Rey-Barroso, N. S. Gov, G. Scita, and L. Dupré. Collective cell motility promotes chemotactic prowess and resistance to chemorepulsion. *Current Biology*, 25(2):242–250, 2015.

[102] R. Mancinelli, D. Vergni, and A. Vulpiani. Front propagation in reactive systems with anomalous diffusion. *Physica D: Nonlinear Phenomena*, 185(3-4):175–195, 2003.

[103] M. Mani, S. Goyal, K. D. Irvine, and B. I. Shraiman. Collective polarization model for gradient sensing via dachsous-fat intercellular signaling. *Proceedings of the National Academy of Sciences*, 110(51):20420–20425, 2013.

[104] T. Maruyama. The rate of decrease of heterozygosity in a population occupying a circular or a linear habitat. *Genetics*, 67(3):437 – 454, 1971.

[105] T. Maruyama. Rate of decrease of genetic variability in a two-dimensional continuous population of finite size. *Genetics*, 70(4):639 – 651, 1972.

[106] S. Matsuura and S. Ryu. Momentum space metric, nonlocal operator, and topological insulators. *Physical Review B*, 82(24):245113, 2010.

[107] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics reports*, 339(1):1–77, 2000.

[108] R. Metzler, A. V. Chechkin, and J. Klafter. Lévy statistics and anomalous transport: Lévy flights and subdiffusion. *Encyclopedia of Complexity and Systems Science*, pages 5218–5239, 2009.

[109] C. W. Misner, K. S. Thorne, J. A. Wheeler, et al. *Gravitation*. Macmillan, 1973.

[110] A. Moroianu. *Lectures on Kähler geometry*, volume 69. Cambridge University Press, 2007.

[111] P. Mörters and Y. Peres. *Brownian motion*, volume 30. Cambridge University Press, 2010.

[112] A. Mugler, A. Levchenko, and I. Nemenman. Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proceedings of the National Academy of Sciences*, 113(6):E689–E695, 2016.

[113] C. D. Nadell, K. R. Foster, and J. B. Xavier. Emergence of spatial structure in cell groups and the evolution of cooperation. *PLoS Comput Biol*, 6(3):e1000716, 2010.

[114] T. Nagylaki. The relation between distant individuals in geographically structured populations. *Mathematical Biosciences*, 28(1-2):73–80, 1976.

[115] R. Nathan, G. Perry, J. T. Cronin, A. E. Strand, and M. L. Cain. Methods for estimating long-distance dispersal. *Oikos*, 103(2):261–273, 2003.

[116] R. Nathan, E. Klein, J. J. Robledo-Arnuncio, and E. Revilla. Dispersal kernels. *Dispersal ecology and evolution*, pages 187–210, 2012.

[117] M. R. Nelson, K. Bryc, K. S. King, A. Indap, A. R. Boyko, J. Novembre, L. P. Briley, Y. Maruyama, D. M. Waterworth, G. Waeber, et al. The population reference sample, popres: a resource for population, disease, and pharmacological genetics research. *The American Journal of Human Genetics*, 83(3):347–358, 2008.

[118] T. Neupert, C. Chamon, and C. Mudry. Measuring the quantum geometry of bloch bands with current noise. *Physical Review B*, 87(24):245103, 2013.

[119] J. Nolan. *Stable distributions: models for heavy-tailed data*. Birkhauser New York, 2003.

[120] J. P. Nolan. Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics*, 28(5):2067–2089, 2013.

[121] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhäuser, Boston, 2018. In progress, Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html.

[122] M. Nordborg. Coalescent theory. *Handbook of statistical genetics*, 2004.

[123] M. D. Onsum, K. Wong, P. Herzmark, H. R. Bourne, and A. P. Arkin. Morphology matters in immune cell chemotaxis: membrane asymmetry affects amplification. *Physical biology*, 3(3):190, 2006.

[124] V. V. Palyulin, A. V. Chechkin, and R. Metzler. Lévy flights do not always optimize random blind search for sparse targets. *Proceedings of the National Academy of Sciences*, 111(8):2931–2936, 2014.

[125] C. A. Parent and P. N. Devreotes. A cell's sense of direction. *Science*, 284 (5415):765–770, 1999.

[126] J. Paulose and O. Hallatschek. The impact of long-range dispersal on gene surfing. *Proceedings of the National Academy of Sciences*, 117(14):7584–7593, 2020.

[127] J. Paulose, J. Hermisson, and O. Hallatschek. Spatial soft sweeps: patterns of adaptation in populations with long-range dispersal. *PLoS genetics*, 15(2): e1007936, 2019.

[128] D. Petkova, J. Novembre, and M. Stephens. Visualizing spatial population structure with estimated effective migration surfaces. *Nature genetics*, 48(1): 94, 2016.

[129] S. M. Pocha and D. J. Montell. Cellular and molecular mechanisms of single and collective cell migrations in drosophila: themes and variations. *Annual review of genetics*, 48:295–318, 2014.

[130] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, M. J. Daly, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3):559–575, 2007.

[131] P. Ralph and G. Coop. The geography of recent genetic ancestry across europe. *PLoS Biol*, 11(5):e1001555, 2013.

[132] R. H. Ree and S. A. Smith. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic biology*, 57(1):4–14, 2008.

[133] A. Rieux, T. Lenormand, J. Carlier, L. De Lapeyre de Bellaire, and V. Ravigné. Using neutral cline decay to estimate contemporary dispersal: a generic tool and its application to a major crop pathogen. *Ecology letters*, 16(6):721–730, 2013.

[134] H. Ringbauer, G. Coop, and N. H. Barton. Inferring recent demography from isolation by distance of long shared sequence blocks. *Genetics*, 205(3):1335–1351, 2017.

[135] J. Robledo-Arnuncio and F. Rousset. Isolation by distance in a continuous population under stochastic demographic fluctuations. *Journal of evolutionary biology*, 23(1):53–71, 2010.

[136] J. J. Robledo-Arnuncio, F. Austerlitz, and P. E. Smouse. A new method of estimating the pollen dispersal curve independently of effective density. *Genetics*, 173(2):1033–1045, 2006.

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

[138] N. A. Rosenberg and M. Nordborg. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, 3(5):380, 2002.

[139] W. J. Rosoff, J. S. Urbach, M. A. Esrick, R. G. McAllister, L. J. Richards, and G. J. Goodhill. A new chemotaxis assay shows the extreme sensitivity of axons to molecular gradients. *Nature neuroscience*, 7(6):678–682, 2004.

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

[141] F. Rousset. Genetic differentiation between individuals. *J Evol Biol*, 13:58–62, 2000.

[142] F. Rousset and R. Leblois. Likelihood-based inferences under isolation by distance: two-dimensional habitats and confidence intervals. *Molecular biology and evolution*, 29(3):957–973, 2011.

[143] E. T. Roussos, J. S. Condeelis, and A. Patsialou. Chemotaxis in cancer. *Nature Reviews Cancer*, 11(8):573–587, 2011.

[144] I. Sanmartín, P. Van Der Mark, and F. Ronquist. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography*, 35(3):428–449, 2008.

[145] B. F. Schutz. *Geometrical methods of mathematical physics*. Cambridge university press, 1980.

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

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

[148] M. Slatkin and H. E. Arter. Spatial autocorrelation methods in population genetics. *The American Naturalist*, 138(2):499–517, 1991.

[149] M. Slatkin and N. H. Barton. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, 43(7):1349–1368, 1989.

[150] T. Smith and D. B. Weissman. Isolation by distance in populations with long-range dispersal. *bioRxiv*, 2020.

[151] T. Smith, S. Fancher, A. Levchenko, I. Nemenman, and A. Mugler. Role of spatial averaging in multicellular gradient sensing. *Physical biology*, 13(3):035004, 2016.

[152] T. R. Sokolowski and G. Tkačik. Optimizing information flow in small genetic networks. iv. spatial coupling. *Physical Review E*, 91(6):062710, 2015.

[153] L. Song, S. M. Nadkarni, H. U. Bödeker, C. Beta, A. Bae, C. Franck, W.-J. Rappel, W. F. Loomis, and E. Bodenschatz. Dictyostelium discoideum chemotaxis: threshold for directed motion. *European journal of cell biology*, 85(9-10): 981–989, 2006.

[154] E. E. Sotka and S. R. Palumbi. The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology*, 87(5):1094–1103, 2006.

[155] A. Srivastava and A. Imamoğlu. Signatures of bloch-band geometry on excitons: nonhydrogenic spectra in transition-metal dichalcogenides. *Physical review letters*, 115(16):166802, 2015.

[156] M. D. Sternlicht, H. Kouros-Mehr, P. Lu, and Z. Werb. Hormonal and local control of mammary branching morphogenesis. *Differentiation*, 74(7):365–381, 2006.

[157] G. Sundaram and Q. Niu. Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and berry-phase effects. *Physical Review B*, 59(23):14915, 1999.

[158] K. F. Swaney, C.-H. Huang, and P. N. Devreotes. Eukaryotic chemotaxis: a network of signaling pathways controls motility, directional sensing, and polarity. *Annual review of biophysics*, 39:265–289, 2010.

[159] V. Vallaeys, R. C. Tyson, W. D. Lane, E. Deleersnijder, and E. Hanert. A Lévy-flight diffusion model to predict transgenic pollen dispersal. *Journal of the Royal Society Interface*, 14(126):20160889, 2017.

[160] E. Verlinde. On the origin of gravity and the laws of newton. *Journal of High Energy Physics*, 2011(4):1–27, 2011.

[161] A. Vezzani, E. Barkai, and R. Burioni. Single-big-jump principle in physical modeling. *Physical Review E*, 100(1):012108, 2019.

[162] J. Wang and A. W. Santure. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics*, 181(4):1579–1594, 2009.

[163] O. Wartlick, A. Kicheva, and M. González-Gaitán. Morphogen gradient formation. *Cold Spring Harbor perspectives in biology*, 1(3):a001255, 2009.

[164] D. B. Weissman and O. Hallatschek. Minimal-assumption inference from population-genomic data. *eLife*, 6:e24836, 2017.

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

[166] J. F. Wilkins. A separation-of-timescales approach to the coalescent in a continuous population. *Genetics*, 168(4):2227–2244, 2004.

[167] M. Willson. Dispersal mode, seed shadows, and colonization patterns. *Vegetatio*, 107(1):261–280, 1993.

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

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

[170] D. Xiao, M.-C. Chang, and Q. Niu. Berry phase effects on electronic properties. *arXiv preprint arXiv:0907.2021*, 2009.

[171] D. Xiao, M.-C. Chang, and Q. Niu. Berry phase effects on electronic properties. *Reviews of modern physics*, 82(3):1959, 2010.

[172] V. Zaburdaev, S. Denisov, and J. Klafter. Lévy walks. *Reviews of Modern Physics*, 87(2):483, 2015.

[173] H.-Q. Zhou and J. P. Barjaktarevič. Fidelity and quantum phase transitions. *Journal of Physics A: Mathematical and Theoretical*, 41(41):412001, 2008.

[174] V. M. Zolotarev. Integral transformations of distributions and estimates of parameters of multidimensional spherically symmetric stable laws. In J. Gani

and V. K. Rohatgi, editors, *Contributions to Probability: A Collection of Papers Dedicated to Eugene Lukacs*, pages 283 – 305. Academic Press, 1981.