Isolation by Distance in Populations with Long-Range Dispersal

Tyler Smith and Daniel B. Weissman

June 24, 2020

1 Abstract

Limited dispersal results in isolation by distance in spatially structured populations, in which individuals found further apart tend to be less related to each other. Models of populations undergoing short-range dispersal predict a close relation between the distance individuals disperse and the length scale over which two sampled individuals are likely to be closely related. In this work, we study the effect of long jumps on patterns of isolation by distance by replacing the typical short-range dispersal kernel with a long-range, power-law kernel. We find that incorporating long jumps leads to a slower decay of relatedness with distance, and that the quantitative form of this slow decay contains visible signatures of the underlying dispersal process.

2 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 [Cayuela et al., 2018, Bradburd and Ralph, 2019, Battey et al., 2020]. 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 [Wright, 1946, Rohlf and Schnell, 1971, Slatkin, 1991]. While in general the spatial pattern of genetic diversity depends on selection [Barton et al., 2013, Allman and Weissman, 2018], 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 [Malécot, 1975, Slatkin, 1993, Slatkin and Arter, 1991].

For populations spread over a fairly continuous range, rather than being clumped into a small number of discrete subpopulations, dispersal is typically assumed to be short range, with displacement approximately following a normal distribution [Barton et al., 2002, Ringbauer et al., 2017]. 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 [Ralph and Coop, 2013, Slatkin and Barton, 1989, Rousset, 2000, Barton et al., 2002, Kimura and Weiss, 1964, Rousset, 1997].

However, many populations exhibit at least occasional long-range dispersal [Aguillon et al., 2017, Brockmann et al., 2006, Atkinson et al., 2002, Dai et al., 2007, Fric and Konvicka, 2007]. In such populations, it is not clear what the pattern of isolation by distance should be. 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 short- or long-range. Even for populations where dispersal is primarily short-range, very rare long jumps are unlikely to be observed directly [Koenig et al., 1996], and can have large effects on the spread of alleles [Hallatschek and Fisher, 2014, Mancinelli et al., 2003, Brockmann and Hufnagel, 2007, Paulose et al., 2019, Paulose and Hallatschek, 2020], so we would also like to be able to infer their pattern and frequency.

In this work, we explore the effects of long-range dispersal in neutrally evolving, demographically stable populations with constant density. Modifying the stepping stone model of Kimura and Weiss [1964], we use a fractional diffusion model [Jespersen et al., 1999] to study coalescence with power law dispersal, following Janakiraman [2017]'s recent results on an analogous problem in chemical physics. Forien [2019] recently proposed a similar model for populations evolving according to a spatial Λ-Fleming-Viot process with long-range dispersal; our analysis may also apply to this model. We find simple expressions for how the pattern of isolation by distance reflects the underlying dispersal process. 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.

3 Model

We consider two lineages sampled in the present a distance x apart, and trace their lineages backwards through time until they coalesce. We assume that lineages follow Lévy flights through an infinite one-dimensional range; we later also consider two-dimensional ranges. Lévy flights are a flexible, mathematically tractable way to model long-range dispersal in which the distribution of jump distances has a power-law tail [Jespersen et al., 1999, Metzler and Klafter, 2000, Metzler et al., 2009]. They occur naturally as the limit of any trajectory composed of many independent, identically distributed jumps, including the special case of diffusive motion in which the jumps have a finite variance.

With this assumption, the displacement y of a lineage from its present position t generations in the past follows a Lévy alpha-stable distribution:

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

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 parameter D_α is a generalized diffusion constant with units of length$^\alpha$/time. It sets the scale of dispersal: at time t, a typical lineage will have a displacement $\overline{y}(t) \sim (D_\alpha t)^{1/\alpha}$. As shown in Fig. 1, the "stability parameter" α, $0 < \alpha \leq 2$, controls the tail of dispersal: for $\alpha < 2$, the probability that at time t a lineage has moved an abnormally long distance $y \gg \overline{y}(t)$ is $K_1(y|t) \propto y^{-1-\alpha}$, a power law tail. In the limiting case $\alpha = 2$, dispersal reduces to ordinary short-range diffusion and K_1 is just a normal distribution with standard deviation $\sqrt{Dt}$.

When the two lineages encounter each other, they coalesce at rate $1/\rho$, where ρ is the density of the population. Technically, for $\alpha \leq 1$ the two lineages will never be at exactly the same position. In two dimensions, all values of α have this problem, including ordinary diffusion [Mörters and Peres, 2010]. So really there must be some small distance δ within which lineages coalesce at rate $\sim 1/(\delta\rho)$. At these small scales, even the model of independent diffusion of lineages breaks down [Barton et al., 2010]. 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:

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

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 $p(t|x)$.

4 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 past $\sim 1/\mu$ generations. In this time, they will disperse a typical distance $\sim \overline{x} \equiv (D_\alpha/\mu)^{1/\alpha}$, 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 falls off exponentially [Barton et al., 2002]:

$$\psi(x) = \frac{e^{-x/\overline{x}}}{4\rho\overline{x}\mu + 1}. \tag{3}$$

Here we generalize (3) to $\alpha < 2$, and find simple approximate expressions for ψ in different parameter regimes, illustrated in Fig. 2, including a universal form for all dispersal kernels at long distances. Intuitively, long-range 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).

4.1 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 within time $\lesssim 1/\mu$ if for one lineage to cover the distance in a long jump. Since such jumps occur at rate $\sim D_\alpha x^{-\alpha-1}\delta$, this occurs with probability $\sim D_\alpha x^{-\alpha-1}\delta/\mu$. The lineages must then coalesce within their neighborhood of $\sim \delta\rho$ individuals before they mutate, which occurs with probability $\sim 1/(\mu\rho\delta)$. We therefore expect that the probability of identity is $\psi \sim D_\alpha x^{-\alpha-1}\delta/\mu/(\mu\rho\delta) = D_\alpha x^{-\alpha-1}/(\mu^2\rho)$, 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) confirms this prediction and determines the prefactor:

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

up to an additional prefactor that enters when $\psi(0)$ is close to one. We confirm (4) with simulations and numerical analysis (Fig. 4).

4.2 Nearby pairs: moderately long-range dispersal

For moderately long-rage dispersal, $1 < \alpha \leq 2$, nearby pairs should be nearly as related as pairs from the same deme: $\psi(x \ll \overline{x}) \approx \psi(0)$. In the time $\sim 1/\mu$ before they mutate, the lineages will disperse over a distance $\sim \overline{x}$. There are $\sim \rho\overline{x}$ individuals in this neighborhood, so the overall average rate of coalescence is $\sim 1/(\rho\overline{x})$. If this is large compared to the mutation rate, then homozygosity will be high, $\psi(0) \approx 1$, while if coalescence is slow compared to mutation, homozygosity will be low, $\psi(0) \sim 1/(\mu\rho\overline{x})$. A more detailed analysis (see Methods) yields numerical factors, and the leading distance dependence was found by Janakiraman [2017]:

$$\psi(x \ll \overline{x}) \approx (2\alpha\sin(\pi/\alpha)\rho\overline{x}\mu + 1)^{-1}\left[1 - \frac{\alpha\sin(\pi/\alpha)}{\Gamma(\alpha)\cos(\pi(1-\alpha/2))}\left(\frac{x}{\overline{x}}\right)^{\alpha-1}\right]. \tag{5}$$

We confirm (5) with simulations and numerical analysis (Fig. 4). The scaling in (5) reflects the superdiffusive dispersal of the lineages.

4.3 Nearby pairs: very long-range dispersal

For very long-range dispersal in which the mean jump size diverges, $\alpha < 1$, nearby lineages that do not coalesce very quickly are likely to disperse across the whole range before coalescing [Palyulin et al., 2014]. This "now-or-never" dynamic has the interesting effect of making the local homozygosity independent of the mutation rate (Fig. 5), since the main competition is between coalescence and long-range dispersal rather than between coalescence and mutation. Intuitively, the pair of lineages take time $t_0 \sim x^\alpha/D_\alpha$ to disperse across the distance between them. From that time on, they are roughly evenly spread over a range $\sim (D_\alpha t)^{1/\alpha}$, and so coalesce at rate $\sim (D_\alpha t)^{-1/\alpha}/\rho$. Integrating this rate over time starting from t_0 out to $\sim 1/\mu$, we find that $\psi \sim 1/(\rho D_\alpha x^{1-\alpha})$, with the upper limit of integration only negligibly decreasing ψ. So ψ again follows a power law, although a different one from the long-distance $1/x^{1+\alpha}$. We calculate ψ more carefully in the Methods to find:

$$\psi(\delta \ll x \ll \overline{x}) \approx \frac{\Gamma(1-\alpha)\sin(\pi\alpha/2)}{2\pi} \frac{1}{\rho D_\alpha x^{1-\alpha}}, \tag{6}$$

again up to an additional prefactor that enters when $\psi(0)$ is close to one. We confirm (6) with simulations and numerical analysis (Fig. 4 and Fig. 5).

The power law in (6) makes it diverge at very short distances, where it breaks down. Instead, for individuals within the same deme, $x < \delta$, ψ flattens out. Roughly speaking, individuals coalescence at rate $1/(\rho\delta)$ and disperse outside of coalescence range at rate $\sim D_\alpha \delta^{-\alpha}$. When coalescence is faster, homozygosity is high, $\psi(0) \approx 1$, while when dispersal is faster it is low, $\psi \sim 1/(\rho\delta)/(D_\alpha\delta^{-\alpha}) = 1/(\rho D_\alpha \delta^{1-\alpha})$. A more careful calculation gives (see Methods):

$$\psi(x \ll \delta) \approx 1 \Big/ \left[1 + \frac{2^{(\alpha+3)/2}\pi}{\Gamma(1/2-\alpha/2)} \rho D_\alpha \delta^{1-\alpha}\right], \tag{7}$$

although these numerical factors depend on the details of the model and are only valid when $\psi(0)$ is small. We confirm (7) with simulations (Fig. 5).

4.4 Finite Variance Dispersal

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. 6, for nearby pairs, $x \ll \overline{x}$, relatedness decays exponentially, as in the case of purely diffusive motion. For distant pairs, $x \gg \overline{x}$, coalescence tends to occur via one long jump, and the homozygosity shows the power law jump tail $\propto x^{-\alpha-1}$ (Fig. 6), following the same behavior as the Lévy flights described above.

4.5 Two Dimensions

While we have focused above on the case of one spatial dimension, the same intuition applies in two dimensions. In some ways the results are even simpler – the same expressions apply for both $\alpha < 1$ and $1 < \alpha < 2$ (Fig. 7). For distant individuals, $x \gg \overline{x}$, the dynamics of coalescence are again dominated by large jumps in a single generation, and mean homozygosity again has the same power law tail as the dispersal kernel, which is now $\alpha + 2$:

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

For nearby individuals, $\delta \ll x \ll \overline{x}$, all values of $\alpha < 2$ display behavior similar to that of $\alpha < 1$ in one dimension. Coalescence tends to occur via one "quick jump", rather than many small jumps, and the competition is with jumping very far away rather than with mutating. The same intuitive argument used above applies, now with the lineages spread over an area of $\sim (D_\alpha t)^{2/\alpha}$ and therefore coalescing at rate $\sim (D_\alpha t)^{-2/\alpha}/\rho$, for a probability of identity $\psi \sim 1/(\rho D_\alpha x^{2-\alpha})$, independent of the mutation rate. We calculate this more carefully in the Methods, finding:

$$\psi(\delta \ll x \ll \overline{x}) \approx \frac{\Gamma(1-\frac{\alpha}{2})}{\Gamma(\frac{\alpha}{2})2^\alpha\pi} \frac{1}{\rho D_\alpha x^{2-\alpha}}. \tag{9}$$

Again, (9) must be cut off at very small distances, $x \lesssim \delta$, at which point homozygosity flattens out to its $x = 0$ value:

$$\psi(x \ll \delta) \approx 1 \Big/ \left[1 + \frac{2^{2+\alpha/2}\pi}{\Gamma(1-\alpha/2)} \rho D_\alpha \delta^{2-\alpha}\right]. \tag{10}$$

Again, the numerical factors in (10) depend on the details of the model and are only valid when $\psi(0) \ll 1$, but the basic scaling is general and arises from the race between coalescence at rate $\sim 1/(\rho\delta^2)$ and dispersing outside coalescence range at rate $\sim D_\alpha \delta^{-\alpha}$.

5 Discussion

Limited dispersal produces a correlation between spatial and genetic distance [Wright, 1946, Malécot, 1975, Slatkin, 1991, 1993, Slatkin and Arter, 1991]. While previous models have generally only considered diffusive dispersal, dispersal can be long-range in many natural populations [Aguillon et al., 2017, Brockmann et al., 2006, Atkinson et al., 2002, Dai et al., 2007]. We therefore generalize classic diffusive models of isolation by distance by allowing dispersal distance to have a power law tail. We find that this leads to much more long-range relatedness than diffusive dispersal, with relatedness having the same power law tail in distance as the dispersal kernel. This is true even for steep power law dispersal 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 dispersal inference typically assume either short-range, diffusive motion [Rousset, 1997, 2000, Robledo-Arnuncio and Rousset, 2010, Ringbauer et al., 2017, Bradburd et al., 2018] (perhaps with recent long-range admixture [Bradburd et al., 2016]) or a small number of discrete demes [Slatkin, 1991, Whitlock and McCauley, 1999, Rousset and Leblois, 2011, Petkova et al., 2016, Al-Asadi et al., 2019, Lundgren and Ralph, 2019]. Our results open the way to methods for inferring more general, realistic dispersal patterns. One key open question is whether it is possible to reasonably detect the genetic traces of rare long-range dispersal in natural populations, and if so how well the form of long-range dispersal (e.g., the tail exponent α) can be determined. Aguillon et al. [2017] found clear patterns of isolation by distance across scales between 500 meters and 10 kilometers in Florida Scrub-Jays, and directly measured long-range 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.

In addition to 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 lengthscale 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 [Weissman and Hallatschek, 2017]. This will be valid as long as recombination is rare relative to mutation, or even if recombination is frequent as long as "μ" is understood to mean the sum of the block mutation and recombination rates. 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 [Harris and Nielsen, 2013].

Our analysis has focused on the lineages of a particular pair of individuals sampled a fixed distance apart in an infinitely large habitat. We can also consider a pair of individuals sampled from random locations from a habitat with finite length L; the mean coalescence time between them would then be the "effective population size". The pair will typically be sampled a distance $\sim L$ from each other, and so it will typically take a time $\sim L^\alpha/D_\alpha$ for their lineages to overlap in space. At this point the ancestry is effectively well-mixed, and it takes time proportional to the total population size $\sim L^d\rho$, 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 $L^\alpha/D_\alpha \gtrsim L^d\rho$ spatial structure substantially increases genetic diversity. For short-range dispersal, $\alpha = 2$, structure will be strong in a one dimensional habitat of length $L \gtrsim D\rho$ [Maruyama, 1971], while in two dimensions its strength depends only on the local neighborhood size $D\rho$ [Maruyama, 1972]. The amount of population structure thus either increases with the spatial extent of the population (at fixed density) or is insensitive to it. With long-range dispersal, however, we see a new qualitative pattern. For $\alpha < d$, i.e., for any amount of long-range dispersal in two dimensions, the effect of structure on overall diversity counterintuitively becomes *weaker* as the range

size L grows, because L^α/D_α grows more slowly than $L^d\rho$. It seems plausible that in many natural populations, spatial structure may be strong over short time scales (as measured by $\psi(x)$ at large μ) while having little effect on the average coalescence time.

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 null 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 long-range 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 short but can be correlated across many generations [Zaburdaev et al., 2015]. Such an effect can be produced at the level of alleles by hitchhiking on beneficial substitutions [Allman and Weissman, 2018]. But this should be readily distinguishable from neutral long-range dispersal by considering the distribution of relatedness across multiple individuals and loci – hitchhiking will produce long-range 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.

6 Methods

6.1 Simulation Methods

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 same dispersal simulations can be used for calculating the homozygosity and coalescence time distribution for multiple choices of ρ and μ. Another advantage 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,

$$X_{t+1} = X_t + \Delta X_t, \tag{11}$$

where X_t represents the signed distance between two lineages at a given time

(ignoring coalescence., i.e., assuming $\rho \to \infty$), 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:

$$K(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \exp\left(-iky - 2D_\alpha |k|^\alpha\right). \tag{12}$$

Note that this differs from (1) because $\Delta t = 1$, and there is an extra factor of two because ΔX_t is the sum of the two lineages' independent jumps. To simulate steeper tails with $\alpha > 2$, we use an F-distribution, defined below in (15). We use the GNU Scientific Library's efficient pseudorandom generators for both stable distributions and the F-distribution [Galassi et al., 2009]; because these are available only for the one-dimensional distributions, all our simulations are in one dimension.

For each simulated trajectory $\{x_t\}$, we then compute the conditional distribution of coalescence times $p(t|\{x_{t' \le t}\})$ and conditional mean homozygosity $\psi(x|\{x_{t' \le t}\})$:

$$p(t \ge 1|\{x_{t' \le t}\}) = \left(1 - e^{-\frac{1}{\rho} R(x_t)}\right) \exp\left[-\frac{1}{\rho} \sum_{t'=1}^{t-1} R(x_{t'})\right], \tag{13}$$

$$\psi(x|\{x_{t' \le t}\}) = \sum_{t=1}^{\infty} p(t|\{x_{t' \le t}\}) e^{-2\mu t}. \tag{14}$$

We start (13) and (14) at $t = 1$ because we assume that the individuals are sampled immediately after dispersal, so no coalescence takes place at $t = 0$. $R(x)$ in (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 discussion issues with the microscopic interpretation of this model after we introduce our analytical model below.

To obtain the unconditioned values $p(t|x)$ and $\psi(x)$, we then average (13) and (14) across all simulated trajectories. All error bars in plots show 68% confidence intervals, as determined by the percentile bootstrap with 10,000 bootstrap samples [Davison and Hinkley, 1997]. 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 use, 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 [Chernick, 2011].

We set $\delta = 0.5$ for all simulations. For the simulations of mean homozygosity ψ shown in Fig. 4 and Fig. 6, we simulate 250,000 independent runs of 1000 generations each for each combination of initial 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 initial 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. 5 and Fig. 9, 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. 8, we set initial 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 $\alpha \leq 1$.

6.1.1 Dispersal kernel for $\alpha > 2$

To simulate dispersal kernels with tail exponents $\alpha > 2$, we draw ΔX_t from a Fisher F-distribution:

$$K(y) = \frac{\Gamma(2\alpha)}{\Gamma^2(\alpha)\omega}(y/\omega)^{\alpha-1}\left(1 + (y/\omega)\right)^{-2\alpha}. \tag{15}$$

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:

$$D = \frac{c^2}{2} = \left(\frac{\alpha(2\alpha-1)}{4(\alpha-2)(\alpha-1)^2} + \frac{\alpha^2}{4(\alpha-1)^2}\right)\omega^2.$$

6.2 Analytical model in one dimension

We want to find a tractable analytical approximation to the model described above. This can be done if we consider only the dynamics on times long compared to a single time step, so that we can treat time as continuous. Then we can again let X_t be the signed distance between the pair of lineages assuming that they are completely independent and cannot coalesce, i.e., $\rho \to \infty$. Since the dispersal kernel is the sum of dispersal over a continuum of infinitesimal time steps, it should approach a stable distribution, i.e., the lineages should follow Lévy flights. This means that the distribution $K(y,t)$ of X_t can be written in terms of its characteristic function:

$$K(y|t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, \exp\left(-iky - 2D_\alpha t|k|^\alpha\right). \tag{16}$$

The factor of two relative to the single-lineage distribution K_1 (1) is because K is the sum of the motion of two independent lineages.

For $\alpha > 1$, 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 $\frac{1}{\rho}\delta(X_t)$. The problem then becomes mathematically identical to that considered by Janakiraman [2017], and we will follow her approach, along with deriving new asymptotic approximations. For $\alpha \leq 1$, however, they will never coincide [Palyulin et al., 2014], and we must allow coalescence to take place at a finite distance. Let the coalescence kernel be some normalized distribution $\mathcal{N}(x)$ 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 [2019] avoids this issue by using a spatial Λ-Fleming-Viot model, in which dispersal and coalescence are produced by the same long-range events, but we wish to keep long-range 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 is then:

$$p(t|x) = E\left[\frac{1}{\rho}\mathcal{N}(X_t)\exp\left(-\frac{1}{\rho}\int_0^t d\tau\,\mathcal{N}(X_\tau)\right)\middle|\, X_0 = x\right], \tag{17}$$

and the probability of identity is its Laplace transform:

$$\psi(x) = \int_0^\infty dt\, p(t|x)e^{-2\mu t}. \tag{18}$$

When X_t follows a diffusion, (18) can be written as a Feynman-Kac (diffusion) equation for ψ [Barton et al., 2002, Allman and Weissman, 2018]. For $\alpha < 2$, this generalizes to a fractional differential equation:

$$0 = 2D_\alpha\left(\frac{\partial^2}{\partial x^2}\right)^{\alpha/2}\psi(x) - 2\mu\psi(x) + \frac{1}{\rho}\mathcal{N}(x)\left(1-\psi(x)\right), \tag{19}$$

where $\left(\frac{\partial^2}{\partial x^2}\right)^{\alpha/2}$ is a Riesz fractional derivative, defined by its Fourier transform $\mathcal{F}\left\{\left(\frac{\partial^2}{\partial x^2}\right)^{\alpha/2} f\right\}(k) = -|k|^\alpha\,\mathcal{F}\{f\}(k)$ [Metzler et al., 2009, Carmi et al., 2010, Janakiraman, 2017]. It is therefore simpler to consider the Fourier transform of (19):

$$0 = -(2D_\alpha|k|^\alpha + 2\mu)\widehat{\psi}(k) + \frac{1}{\rho}\mathcal{F}\left\{\mathcal{N}(x)(1-\psi(x))\right\}(k), \tag{20}$$

where $\widehat{\psi}$ is the Fourier transform of ψ.

In general, Fourier transforms of products such as $\mathcal{F}\left\{\mathcal{N}(x)(1-\psi(x))\right\}$ do not have a simple form in Fourier space. But we can make the ansatz that $1-\psi(x)$ is nearly constant over all x values where $\mathcal{N}(x)$ is non-negligible. This is exactly true in continuous space for $\alpha > 1$, where we can take $\mathcal{N}$ to be a δ-function. For $\alpha \leq 1$, it will necessarily be accurate in the limit $\psi(0) \ll 1$. In these regimes, we verify that our approximation is consistent below and check it

with simulations (Fig. 4, Fig. 5, and Fig. 9). However, for $\alpha \leq 1$ and $1-\psi(0) \ll 1$, the expression is inaccurate; we discuss this below. With this ansatz, the Fourier transform is simply $\mathcal{F}\{\mathcal{N}(x)(1-\psi(x))\}(k) \approx \widehat{\mathcal{N}}(k)(1-\psi(0))$, and (20) simplifies to:

$$\frac{\widehat{\psi}(k)}{1-\psi(0)} \approx \frac{\widehat{\mathcal{N}}(k)}{2\rho(\mu + D_\alpha |k|^\alpha)}. \tag{21}$$

(21) can be rewritten to emphasize the connection to the dispersal kernel K. First, note that the Fourier transform of K is its characteristic function: $\widehat{K}(k,t) = \exp(-2D_\alpha t|k|^\alpha)$. Taking the Laplace transform, which we will denote with a tilde, gives:

$$\begin{aligned}\widehat{\widetilde{K}}(k, 2\mu) &\equiv \int_0^\infty dt\, e^{-2\mu t - 2D_\alpha t|k|^\alpha} \\ &= \frac{1}{2(\mu + D_\alpha |k|^\alpha)}.\end{aligned}$$

We see that the denominator in (21) is coming from $\widehat{\widetilde{K}}$:

$$\widehat{\psi}(k) \approx \frac{1-\psi(0)}{\rho} \widehat{\mathcal{N}}(k) \widehat{\widetilde{K}}(k, 2\mu). \tag{22}$$

The natural interpretation of (22) is that identity by descent is a kind of convolution of coalescence ($\frac{1}{\rho}(1-\psi(0))\mathcal{N}$) and dispersal ($K$), integrated over a time set by the inverse of the mutation rate. (22) is defined for any dispersal kernel, not just stable ones, and we conjecture that it will still hold more generally.

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 $\widehat{\mathcal{N}}(k) = \exp(-\delta^2 k^2/2)$. Then we can invert the Fourier transform in (21):

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{2\pi\rho} \int_0^\infty dk \frac{\cos(kx) e^{-\delta^2 k^2/2}}{\mu + D_\alpha k^\alpha}, \tag{23}$$

which can be re-expressed in dimensionless units as

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{2\pi\rho\mu\overline{x}} \int_0^\infty d\kappa \frac{\cos(\kappa x/\overline{x}) e^{-(\delta/\overline{x})^2 \kappa^2/2}}{1+\kappa^\alpha}. \tag{24}$$

Examining (23), 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, leaving (in dimensionless form):

$$\frac{\psi(x \gg \delta)}{1-\psi(0)} \approx \frac{1}{2\pi\rho\mu\overline{x}} \int_0^\infty d\kappa \frac{\cos(\kappa x/\overline{x})}{1+\kappa^\alpha}. \tag{25}$$

Dropping the normal factor also allows us to trivially invert the Fourier transform in (22) and see that ψ is just proportional to the Laplace transform of K:

$$\psi(x \gg \delta) \approx \frac{1-\psi(0)}{\rho} \widetilde{K}(x, 2\mu). \tag{26}$$

Again, although we have only shown that (26) holds for stable dispersal kernels, it can be defined for any dispersal kernel, and we conjecture that it will hold more generally (see $\alpha > 2$ section).

We can solve (24) 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 [1997] for short-range dispersal. Ψ is closely related to Rousset [2000]'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:

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

6.2.1 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 [2018], 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 (26) and evaluating the Laplace transform gives (4):

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{\Gamma(\alpha+1)}{2\pi} \sin\left(\frac{\pi\alpha}{2}\right) \frac{D_\alpha}{\rho\mu^2 x^{\alpha+1}}. \tag{27}$$

6.2.2 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 $D = \mathrm{Var}(K)/2$. At long distances $y \gg \sqrt{Dt}$, however, the tail will still be dominated by the probability of taking a single large jump [Vezzani et al., 2019], so for $x \gg \overline{x}$, we will have $K(x|t \lesssim 1/\mu) \approx K(x|1)t$. Plugging this into (26) and evaluating the Laplace transform gives:

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

For the F-distribution kernel (15) used in the simulations, this is

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{\Gamma(2\alpha)}{4\Gamma^2(\alpha)} \frac{D_\alpha}{\rho\mu^2 x^{\alpha+1}}, \tag{28}$$

where we have defined $D_\alpha \equiv \omega^\alpha$/generation. (28) is confirmed by simulations (Fig. 6). We can then use the classic diffusive expression for $\psi(0)$ to get an explicit expression for probability of identity at large distances:

$$\psi(x \gg \overline{x}) \approx \left(1 + \frac{1}{4\rho\overline{x}\mu}\right)^{-1} \frac{\Gamma(2\alpha)}{4\Gamma^2(\alpha)} \frac{D_\alpha}{\rho\mu^2 x^{\alpha+1}}. \tag{29}$$

6.2.3 Moderately long-range dispersal, $1 < \alpha < 2$

For $\alpha > 1$, (25) and (26) are exact for all x. Evaluating (25) for $x = 0$ gives $\psi(0)$:

$$\psi(0) = \frac{1}{2\alpha \sin(\pi/\alpha)\rho\overline{x}\mu + 1}. \tag{30}$$

Plugging (30) into (27) lets us solve for $\psi(x)$ at large distances $x \gg \overline{x}$:

$$\psi(x \gg \overline{x}) \approx \left(1 + \frac{1}{2\alpha \sin(\pi/\alpha)\rho\overline{x}\mu}\right)^{-1} \sin\left(\frac{\pi\alpha}{2}\right) \frac{\Gamma(\alpha+1)}{2\pi} \frac{D_\alpha}{\rho\mu^2 x^{\alpha+1}}. \tag{31}$$

For $0 < x \ll \overline{x}$, Janakiraman [2017] (Eq. (C1)) found that to leading order ψ falls off as:

$$\psi(x \ll \overline{x}) \approx \psi(0)\left[1 - \frac{\alpha \sin(\pi/\alpha)}{\Gamma(\alpha)\cos(\pi(1-\alpha/2))}\left(\frac{x}{\overline{x}}\right)^{\alpha-1}\right]. \tag{32}$$

When $\alpha = 2$, the above expression is equivalent to the classic diffusive result (3) for $x \ll \overline{x}$, which can be found by integrating (19) with $\delta = 0$.

6.2.4 Very long-range dispersal, $\alpha < 1$

For $\alpha < 1$, the finite width δ of the coalescence kernel is important for determining $\psi(0)$. Dropping the cosine factor in (24) gives:

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

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, i.e., there is a large probability that individuals from the same deme are differentiated even for infinitesimal mutation rates:

$$\psi(0) \approx 1 \Big/ \left[1 + \frac{2^{(\alpha+3)/2}\pi}{\Gamma(1/2-\alpha/2)} \rho D_\alpha \delta^{1-\alpha}\right] \tag{33}$$

Very long-range dispersal of nearby lineages causes them to quickly wander away from each other, and for infinite range size many pairs will never coalesce. While

(33) is only accurate for $\psi(0) \ll 1$, the independence from mutation rate should persist even for large $\psi(0)$.

Plugging (33) for $\psi(0)$ into (27) gives an explicit expression for the probability of identity of distant pairs with $x \gg \overline{x}$:

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

For pairs that are nearby but still well outside of coalescence range, $\delta \ll x \ll \overline{x}$, the integral in (25) 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 (33) gives an explicit expression for ψ:

$$\psi(\delta \ll x \ll \overline{x}) \approx \left(1 + \frac{\Gamma(1/2 - \alpha/2)}{2^{(\alpha+3)/2}\pi\rho D_\alpha \delta^{1-\alpha}}\right)^{-1} \frac{\Gamma(1-\alpha)\sin(\pi\alpha/2)}{2\pi} \frac{x^{\alpha-1}}{\rho D_\alpha}. \quad (34)$$

6.2.5 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:

$$\psi(0) \approx \left(1 + \frac{2\pi\rho D_1}{\ln(\overline{x}/\delta)}\right)^{-1}. \quad (35)$$

Recall that the ansatz we used to derive (35) is only justified when $\psi(0) \ll 1$.

For pairs that are nearby but still well outside of coalescence range, $\delta \ll x \ll \overline{x}$, (25) gives:

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1 - \psi(0)} \approx \frac{\ln(\overline{x}/x) - \gamma}{2\pi\rho D_1}. \quad (36)$$

Plugging the expression (35) for $\psi(0)$ into (36) and (27) gives explicit expressions for $\psi(x)$ at both short and long distances:

$$\psi(\delta \ll x \ll \overline{x}) \approx \frac{\ln(\overline{x}/x)}{2\pi\rho D_1 + \ln(\overline{x}/\delta)} \quad (37)$$

$$\psi(x \gg \overline{x}) \approx \frac{1}{2\pi\rho D_1 + \ln(\overline{x}/\delta)} \left(\frac{\overline{x}}{x}\right)^2. \quad (38)$$

6.3 Analytical model in two dimensions

For a two-dimensional Lévy flight, the dispersal kernel takes the form of an isotropic stable distribution [Zolotarev, 1981]:

$$K(y|t) = \frac{1}{2\pi}\int_0^\infty dk\, kJ_0(ky)\exp\left(-2D_\alpha t k^\alpha\right), \tag{39}$$

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. (39) is the two-dimensional inverse Fourier transform (equivalently, the inverse zeroth-order Hankel transform) of the characteristic function $\widehat{K}(k|t) = \exp\left(-2D_\alpha t k^\alpha\right)$, where k is the radial coordinate in two-dimensional k-space. At large distances, $y \gg (D_\alpha t)^{1/\alpha}$, K has a power-law tail [Nolan, 2013]:

$$K(y \gg (D_\alpha t)^{1/\alpha}|t) \approx \frac{\alpha^2\Gamma(\alpha/2)^2}{2^{1-\alpha}\pi^2}\sin\left(\frac{\pi\alpha}{2}\right)\frac{D_\alpha t}{x^{\alpha+2}}. \tag{40}$$

In two dimensions, we must allow coalescence to take place at a finite distance for all α [Mörters and Peres, 2010]. 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 $\frac{1}{\rho}\mathcal{N}(X_t)$.

For all $\alpha < 2$, the solution for ψ in two dimensions can be written as a fractional differential equation [Chen et al., 2012]:

$$0 = 2D_\alpha\left(\frac{\partial^2}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x}\right)^{\frac{\alpha}{2}}\psi(x) - 2\mu\psi(x) + \frac{1}{\rho}\mathcal{N}(x)\left(1-\psi(x)\right), \tag{41}$$

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 $\mathcal{F}\left\{\left(\frac{\partial^2}{\partial x^2} + \frac{1}{x}\frac{\partial}{\partial x}\right)^{\alpha/2}\right\}(k) = -|k|^\alpha\,\mathcal{F}\{f\}(k)$ [Kwaśnicki, 2017, Lischke et al., 2020]. 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 Fourier transform of (41) has the same form as the one-dimensional equation (20):

$$0 = -(2D_\alpha k^\alpha + 2\mu)\widehat{\psi}(k) + \frac{1}{\rho}\mathcal{F}\left\{\mathcal{N}(x)(1-\psi(x))\right\}(k). \tag{42}$$

The expressions (21) and (22) for $\widehat{\psi}(k)$ therefore remain the same. Note however that their interpretation is different: if we want to transform back to real space, we must use the two-dimensional inverse Fourier transform. Applying it to (21) gives:

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{4\pi\rho}\int_0^\infty dk\frac{kJ_0(kx)e^{-\delta^2k^2/2}}{\mu + D_\alpha k^\alpha} \tag{43}$$

$$= \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa\frac{\kappa J_0(\kappa x/\overline{x})e^{-(\delta/\overline{x})^2\kappa^2/2}}{1+\kappa^\alpha}, \tag{44}$$

where again we make the ansatz 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 needs to be adjusted for $1-\psi(0) \ll 1$. The analysis of (44) 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.

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

For pairs sampled from the same location, $x = 0$, the Bessel function in (44) is simply equal to one and can be dropped:

$$\begin{aligned}\frac{\psi(0)}{1-\psi(0)} &\approx \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa e^{-(\delta/\overline{x})^2\kappa^2/2}}{1+\kappa^\alpha} \\ &\approx \frac{\Gamma(1-\alpha/2)}{2^{2+\alpha/2}\pi\rho D_\alpha \delta^{2-\alpha}},\end{aligned} \tag{45}$$

where in the last line we have assumed that $\delta \ll \overline{x}$. Intuitively, (45) 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, (45) is only accurate for $\psi(0) \ll 1$.

Solving (45) for ψ gives:

$$\psi(0) \approx \left(1 + \frac{2^{2+\alpha/2}\pi}{\Gamma(1-\alpha/2)}\rho D_\alpha \delta^{2-\alpha}\right)^{-1}. \tag{46}$$

For $\alpha = 2$, integrating (44) with $x = 0$ recovers the classic diffusive result in two dimensions, which we expect to hold for pairs in contact when $\alpha \geq 2$ [Barton et al., 2002]:

$$\psi(0) \approx \frac{\ln(\overline{x}/\delta)}{\ln(\overline{x}/\delta) + 4\pi\rho D_2}. \tag{47}$$

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

For pairs that are outside coalescence range, $x \gg \delta$, oscillations in the Bessel function suppress the contribution of $k \gtrsim 1/x$ to (43), and the normal factor can be neglected. This is a somewhat subtle point for $\alpha \leq 1/2$, for which the amplitude of the integrand grows with k (or does not decay for $\alpha = 1/2$). But the oscillations are indeed sufficiently fast so that the contributions to the integral cancel, as can be checked by numerical integration. One simple way to see that this is at least plausible is to note that at $t = 0$ (39) must reduce to $K(y|0) = \delta(y)$, i.e., the oscillations cancel the integral entirely even though the integrand grows like $\sqrt{k}$.

Without the normal factor, (44) is:

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{\kappa J_0(\kappa x/\overline{x})}{1+\kappa^\alpha}. \tag{48}$$

For nearby pairs $x \ll \overline{x}$, the integral in (48) 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:

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{\Gamma(1-\alpha/2)}{\Gamma(\alpha/2)2^\alpha\pi\rho D_\alpha} x^{\alpha-2}. \tag{49}$$

The convergence of (48) to (49) 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 (46) for $\psi(0)$ into (49) lets us solve for ψ:

$$\psi(\delta \ll x \ll \overline{x}) \approx \left(1+\frac{\Gamma(1-\alpha/2)}{2^{2+\alpha/2}\pi\rho D_\alpha \delta^{2-\alpha}}\right)^{-1} \frac{\Gamma(1-\alpha/2)}{\Gamma(\alpha/2)2^\alpha\pi\rho D_\alpha} x^{\alpha-2}, \tag{50}$$

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 $\alpha = 2$ and $\delta \ll x \ll \overline{x}$ we can recover the known result for diffusive motion by approximating (44) as

$$\frac{\psi(\delta \ll x \ll \overline{x})}{1-\psi(0)} \approx \frac{1}{4\pi\rho\mu\overline{x}^2}\int_0^\infty d\kappa \frac{J_0(\kappa x/\overline{x})\kappa}{1+\kappa^2}. \tag{51}$$

Integrating (51) confirms that we find logarithmic scaling of $\psi(x)$ at short distances [Barton et al., 2002]:

$$\frac{\psi(x)}{1-\psi(0)} \approx \frac{1}{4\pi\rho D_2}\ln(\overline{x}/x), \tag{52}$$

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

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

Since, as discussed in the previous section, the δ-dependent terms can be dropped for $x \gg \delta$, (26) still applies, just as it did in one dimension. The probability of identity by descent for distant pairs $x \gg \overline{x}$ can then immediately be read off from (26), now using the tail of the two-dimensional dispersal kernel (40) for $\alpha < 2$:

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{\alpha^2\Gamma(\alpha/2)^2}{2^{3-\alpha}\pi^2}\sin\left(\frac{\pi\alpha}{2}\right)\frac{D_\alpha}{\rho\mu^2}x^{-\alpha-2}. \tag{53}$$

Plugging in (46) for $\psi(0)$ lets us solve for ψ:

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

When $\alpha = 2$, we instead recover classic expression for two dimensional diffusive motion at large distances [Barton et al., 2002]:

$$\frac{\psi(x \gg \overline{x})}{1-\psi(0)} \approx \frac{1}{4\rho D_2} \frac{\exp(-x/\overline{x})}{\sqrt{2\pi x/\overline{x}}}. \tag{54}$$

6.4 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 [1971], XIII.5, Theorem 4).

6.4.1 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 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 $\psi(0)$ (35) and (46) 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 (27) and (53) to obtain:

$$\begin{aligned} p(t \ll x^\alpha/D_\alpha|x) &\approx 2d^\alpha \left(\frac{\Gamma(1+\alpha/d)}{\pi}\right)^d \sin\left(\frac{\pi\alpha}{2}\right) \frac{D_\alpha t}{\rho x^{\alpha+d}} \\ &\approx \frac{1}{\rho} K(x|t). \end{aligned} \tag{55}$$

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. (55) 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. 4 and Fig. 6.

6.4.2 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. 8.

For $\alpha < d$, we can simply take the inverse Laplace and Fourier transforms of (22) 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 (22) can be neglected and $p(t|x)$ is simply given by the inverse Laplace transform of (26):

$$p(t \gg \delta^\alpha / D_\alpha | x) \approx \frac{1 - \psi(0)}{\rho} K(x|t) \tag{56}$$

$$\approx \frac{1 - \psi(0)}{\alpha (2\pi)^{d-1} \rho} \Gamma(d/\alpha) (2 D_\alpha t)^{-d/\alpha} \text{ for } t \gg x^\alpha / D_\alpha. \tag{57}$$

Integrating yields the cumulative distribution:

$$P(t \gg \delta^\alpha / D_\alpha | x) = P(\infty | x) - \frac{1 - \psi(0)}{(d - \alpha)(2\pi)^{d-1}} \frac{\Gamma(d/\alpha)(2D_\alpha)^{-d/\alpha}}{\rho \, t^{d/\alpha - 1}} \text{ for } t \gg x^\alpha / D_\alpha, \tag{58}$$

where $P(\infty|x) = \lim_{\mu \to \infty} \psi(x)$ is given by (7), (34), (46), or (50), 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 [1971], XIII.5, Theorem 2) to convert (35), (37), (47), and (52) to expressions for the cumulative distribution P:

$$P(t \gg x/D_d | x) \approx \begin{cases} \left[1 + 2d\pi\rho D_d / \ln(2D_d t/\delta)\right]^{-1} & \text{for } x \ll \delta \\ \ln(2D_d t/x) / \left[2d\pi\rho D_d + \ln(2D_d t/\delta)\right] & \text{for } x \gg \delta. \end{cases} \tag{59}$$

We can then differentiate to find the density p:

$$p(t \gg x/D_d | x) \approx \begin{cases} t^{-1} \left(2d\pi\rho D_d\right) / \left[2d\pi\rho D_d + \ln(2D_d t/\delta)\right]^2 & \text{for } x \ll \delta \\ t^{-1} \left(\ln(x/\delta) + 2d\pi\rho D_d\right) / \left[2d\pi\rho D_d + \ln(2D_d t/\delta)\right]^2 & \text{for } x \gg \delta. \end{cases} \tag{60}$$

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 $d = 1 < \alpha \leq 2$, since the leading behavior of the cumulative distribution is trivial, $\lim_{\mu \to 0} \psi(x) = \lim_{t \to \infty} P(t|x) = 1$, we must instead consider the complementary cumulative distribution, $\overline{P}(t|x) \equiv 1 - P(t|x)$. Its Laplace transform is:

$$\widetilde{\overline{P}}(2\mu) = \frac{1}{2\mu} - \widetilde{P}(2\mu|x)$$

$$= \frac{1}{2\mu} \left[1 - \psi(x)\right].$$

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 (30) and (32), 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:

$$\overline{P}(t \gg x^\alpha/D_\alpha|x) \approx \alpha \sin\left(\frac{\pi}{\alpha}\right) \frac{\rho(2D_\alpha)^{1/\alpha}}{t^{1-1/\alpha}}. \tag{61}$$

Differentiating (61) yields the density $p(t|x)$:

$$p(t \gg x^\alpha/D_\alpha|x) \approx (\alpha - 1) \sin\left(\frac{\pi}{\alpha}\right) \frac{\rho(2D_\alpha)^{1/\alpha}}{t^{2-1/\alpha}}, \tag{62}$$

in agreement with Janakiraman [2017]'s Eq. 19.

For $\alpha = 2$, (61) and (62) simplify to the classic diffusive results:

$$\overline{P}(t \gg x^2/D|x) \approx 2\rho\sqrt{2D/t}, \tag{63}$$

$$p(t \gg x^2/D|x) \approx \rho\sqrt{2D/t^3}. \tag{64}$$

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. 8, bottom right).

6.5 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 [Felsenstein, 1975, Barton et al., 2010]. 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 [Barton et al., 2002]. Forien [2019] has recently given an example of a biologically consistent forward-time, continuous-space model that we believe would be described by our results on macroscopic scales. We also believe that our results would 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. 9, 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 coalesce 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:

$$\begin{aligned}\psi(x \ll c) &\approx \sum_{t=1}^{\infty} \frac{K(0|t)}{\rho} \\ &= \frac{\Gamma(1/\alpha)\zeta(1/\alpha)}{\alpha} \frac{1}{\rho c},\end{aligned} \tag{65}$$

where ζ is the Riemann zeta function. Fig. 9 shows that (65) accurately describes the simulations.

References

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.

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

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

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

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

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.

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.

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.

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

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.

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.

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

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

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

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.

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.

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

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.

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

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

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

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.

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.

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.

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.

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

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

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.

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.

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.

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

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.

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.

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

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.

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

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

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

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.

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

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

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.

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.

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.

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.

D. Petkova, J. Novembre, and M. Stephens. Visualizing spatial population structure with estimated effective migration surfaces. *Nature genetics*, 48(1): 94, 2016.

P. Ralph and G. Coop. The geography of recent genetic ancestry across Europe. *PLoS biology*, 11(5):e1001555, 2013.

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.

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.

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

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

F. Rousset. Genetic differentiation between individuals. *J Evol Biol*, 13:58–62, 2000.

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.

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

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

M. Slatkin and H. E. Arter. Spatial autocorrelation methods in population genetics. *The American Naturalist*, 138(2):499–517, 1991.

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.

A. Vezzani, E. Barkai, and R. Burioni. Single-big-jump principle in physical modeling. *Physical Review E*, 100(1):012108, 2019.

D. B. Weissman and O. Hallatschek. Minimal-assumption inference from population-genomic data. *eLife*, 6:e24836, 2017.

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.

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

V. Zaburdaev, S. Denisov, and J. Klafter. Lévy walks. *Reviews of Modern Physics*, 87(2):483, 2015.

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.

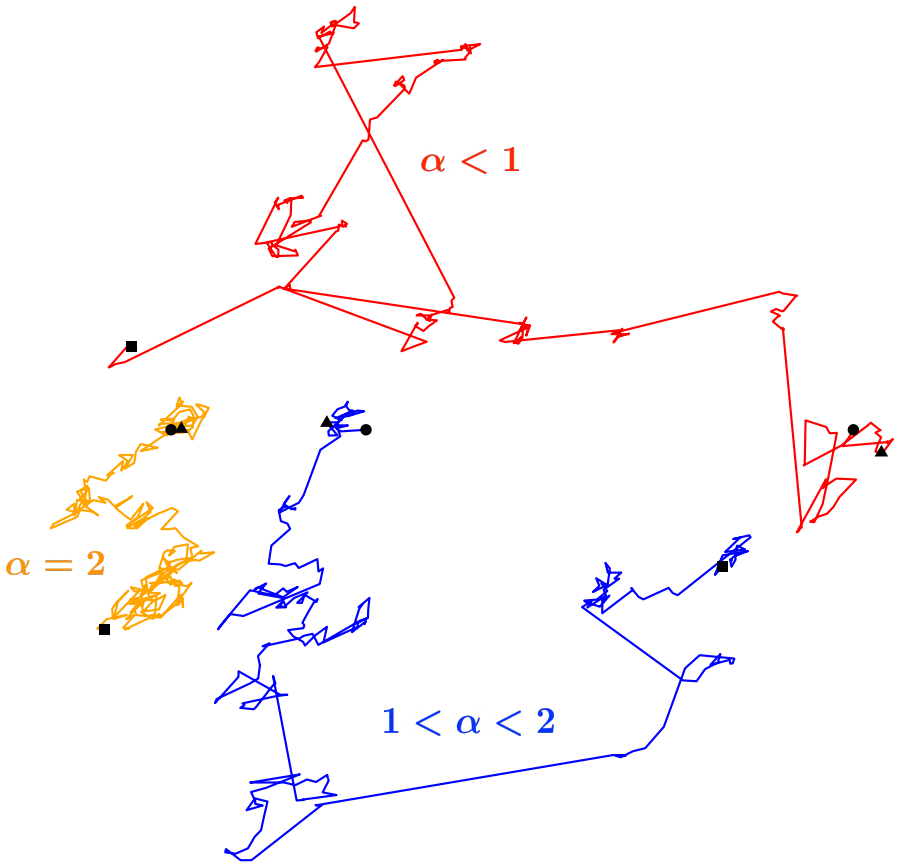

Figure 1: **α controls the size and number of long-range jumps.** Two-dimensional Lévy flights with α values 0.95, 1.5, and 2. When $\alpha = 2$, Lévy flight motion reduces to normal diffusion without any long-range jumps. Lévy flights with $\alpha < 2$ have divergent mean squared displacement and a power law dispersal kernel proportional to $x^{-\alpha-1}$. For $\alpha < 1$, the mean displacement also diverges, and large jumps become noticeably more prevalent than for $1 < \alpha < 2$. Circles mark the beginnings of the trajectories, while squares mark the ends. The generalized dispersal constants D_α are chosen such that the flights all have the same characteristic displacement at the time step marked by triangles. On shorter time scales, the diffusive trajectory tends to have the largest displacement, while on longer time scales the trajectory with $\alpha < 1$ tends to have the largest displacement.

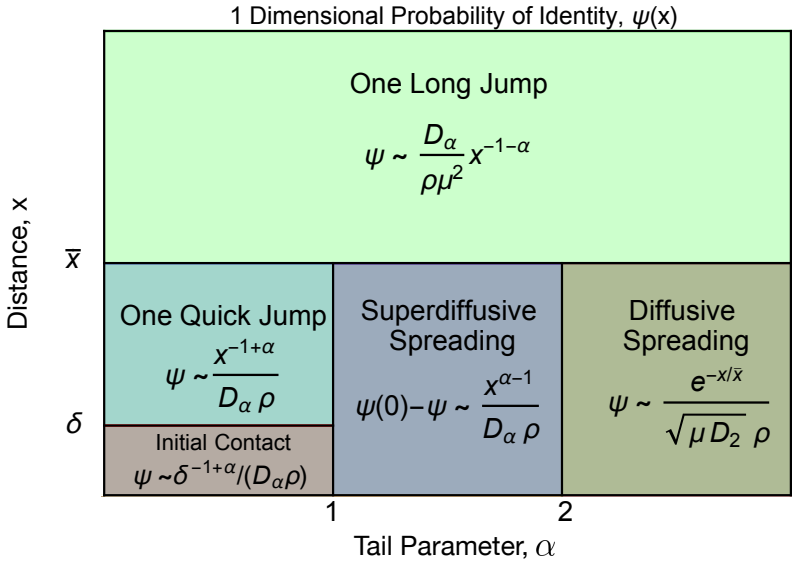

Figure 2: **The form of isolation by distance in one dimension is universal at long distances and varies at short distance.** Approximate form for the probability of identity as a function of distance, $\psi(x)$, for different dispersal kernels α. Different regimes of the parameter space are separated 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 (4). 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). 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 (5). For $\alpha < 1$, even short distances are covered by one quick jump, leading to the power law shown in (6). 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 (7). We use "$\sim$" to denote proportionality in the limit of large population density where $\psi(0) \ll 1$.

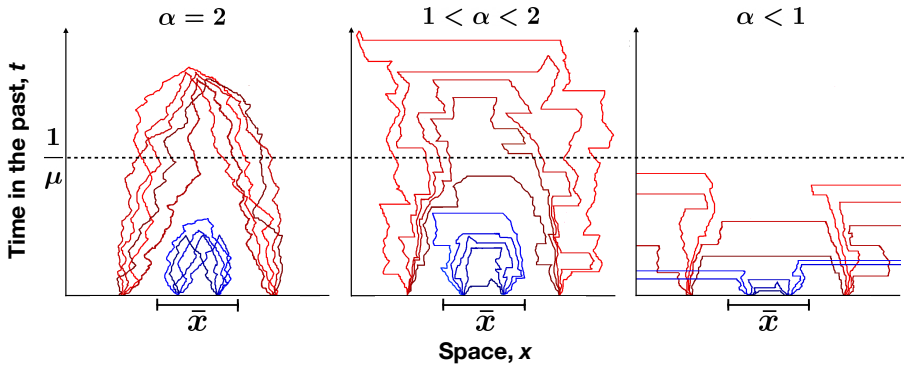

Figure 3: **Long-range jumps affect when and where lineages coalesce.** Qualitative illustrations of lineage dynamics for each of the three α regimes in 1 dimension. Typical histories are shown for nearby samples ($x \ll \overline{x}$, blue) and distant samples ($x \gg \overline{x}$, red). **Left**: For diffusive motion, $\alpha = 2$, initial separation x is a relatively good predictor of coalescence time. **Center**: For moderately long-range dispersal, $1 < \alpha < 2$, 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 $1/\mu$. **Right:** For very long-range dispersal, $\alpha < 1$, large jumps are common. This allows for rapid coalescence of lineages at both small and large distances, but also lets lineages jump very far away from each other and avoid coalescing.

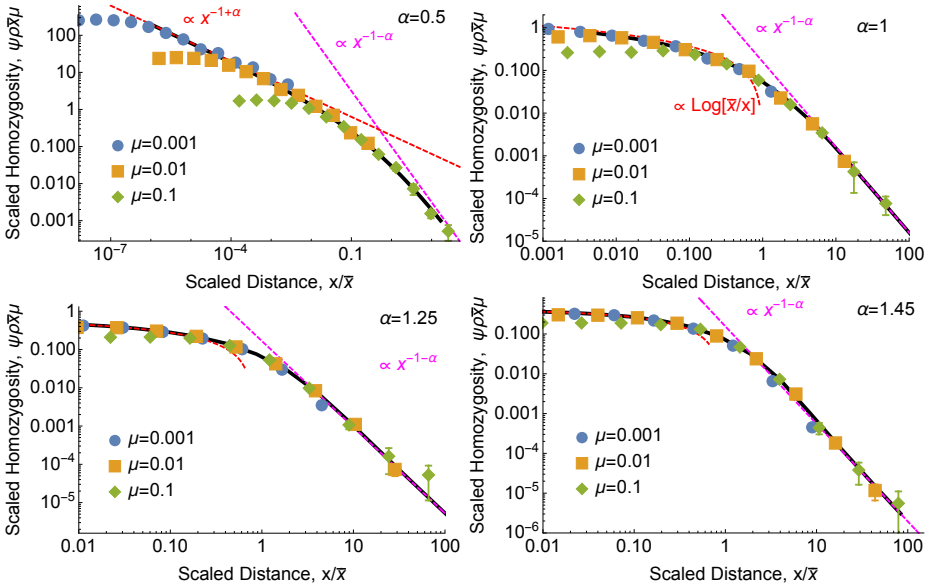

Figure 4: **Isolation by distance follows the same power law as dispersal.** Each panel shows the scaled probability of identity between a sampled pair of individuals, $\psi\rho\overline{x}\mu$, as a function of the scaled distance $x/\overline{x}$ between them. Points show simulation results, black curves show numerical solutions of $\psi(x)$ calculated from (23) with $\delta = 0$, and magenta lines show the power law that emerges at large distances (4). Red curves show the asymptotic behavior predicted at short distances by (6) ($\alpha < 1$), (5) ($1 < \alpha < 2$), and (37) ($\alpha = 1$). 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.

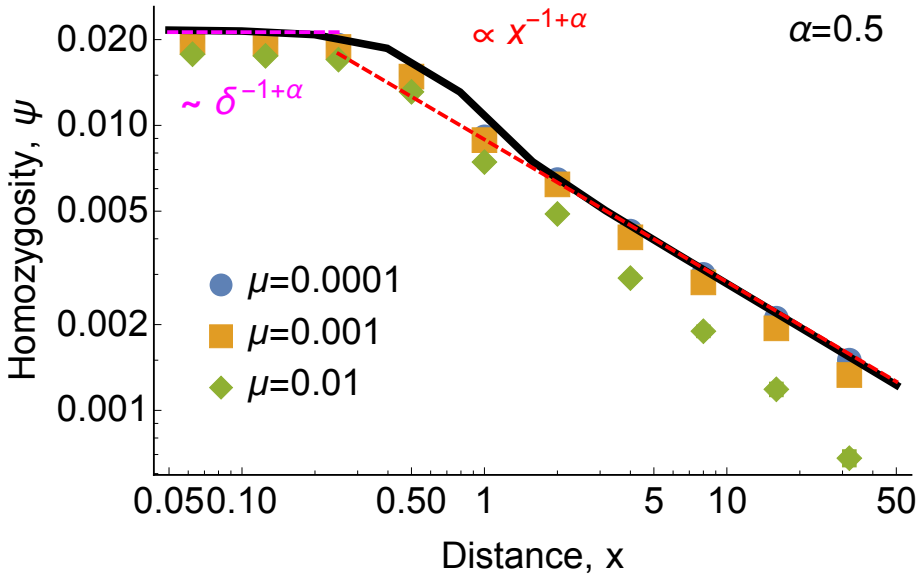

Figure 5: **For very long-range dispersal, $\alpha < 1$, relatedness at short distances is independent of mutation rate.** Nearby lineages at $x \ll \overline{x}$ either coalesce quickly and are identical, or jump very far away from each other and never coalesce. Points show simulation results, and red and magenta lines show the asymptotic predictions of (6) and (7), respectively. The black curve shows a numerical solution of $\psi(x)$ calculated from (23) with $\mu = 10^{-4}$. $\rho = 100$ in all plots, and data with $\rho = 10$ and $\rho = 1$ (not shown) yield indistinguishable plots.

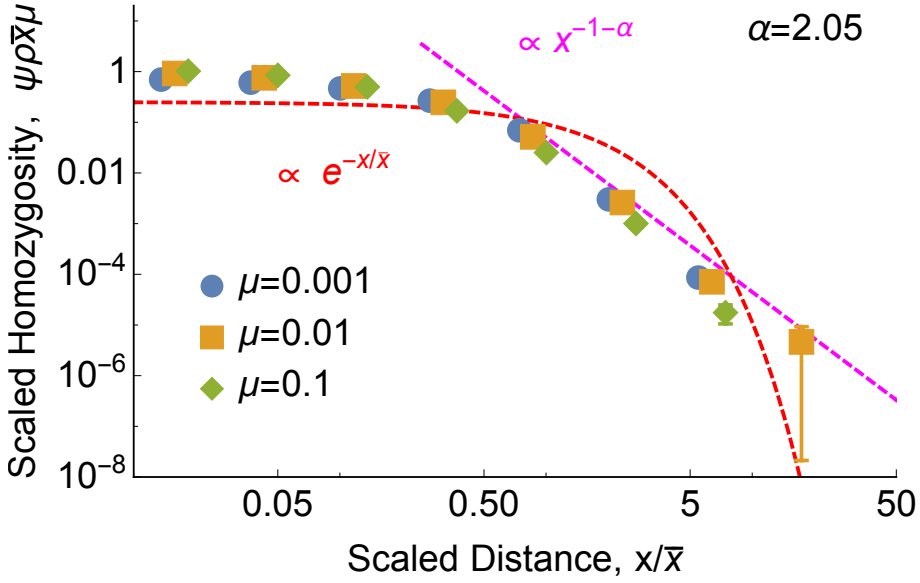

Figure 6: **Even for $\alpha > 2$, relatedness still follows the same power law as dispersal, rather than the diffusive prediction.** Points show simulation results with $\rho = 100$; $\rho = 10$ and $\rho = 1$ yield indistinguishable plots. Since the dispersal kernel has finite variance, it approaches a diffusion, and at short distances $x \ll \overline{x}$ the probability of identity is well-approximated by the diffusive prediction (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, (28) (magenta line).

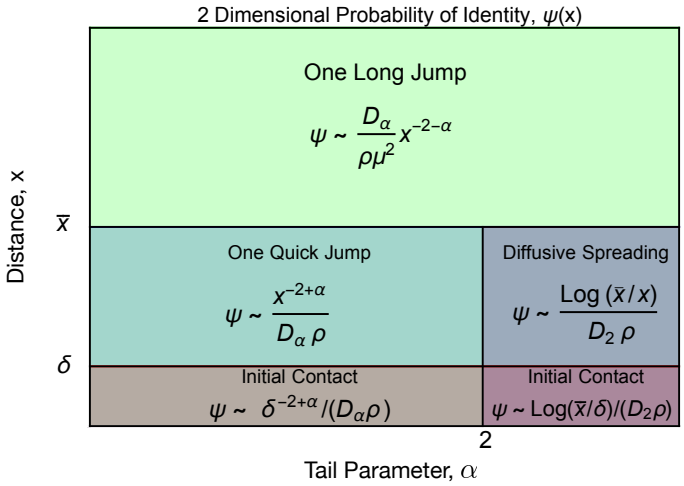

Figure 7: **The form of isolation by distance in two dimensions is universal at long distances.** Approximate form for the probability of identity as a function of distance, $\psi(x)$, for different dispersal kernels α. Different regimes of the parameter space are separated by solid lines, and labelled 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 (8). 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 (9). 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 $\psi(0) \ll 1$.

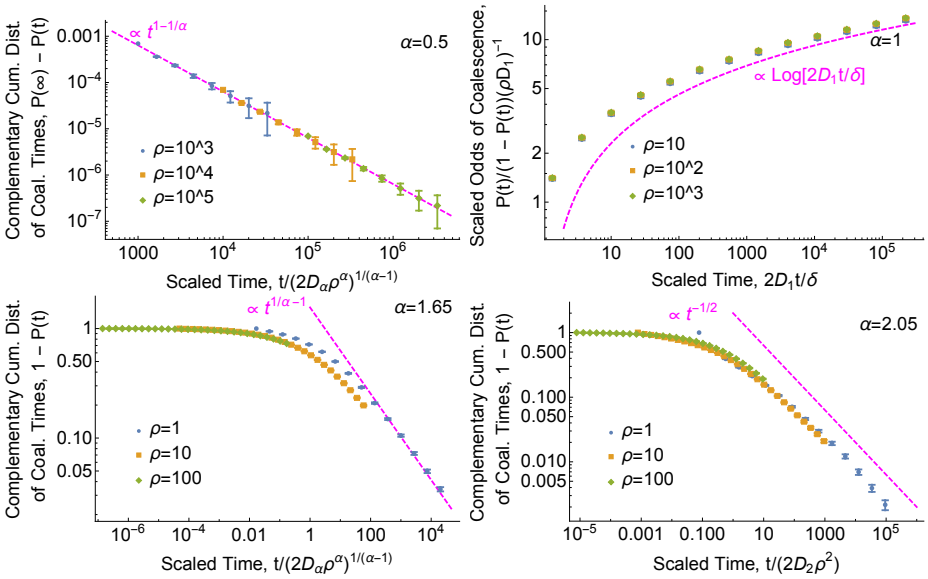

Figure 8: **The distribution of coalescence times has a power law tail.** Points show simulation results. Dashed magenta curves show the asymptotic predictions (in order of increasing α) (58), (59), (61), and (63). Time is scaled to dimensionless units. See Simulation Methods section for D_α values. 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 $\alpha < 1$, 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 (7) 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 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 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, (63), but there is at least a difference in prefactor, perhaps again due to different probabilities of coalescence at very recent times.

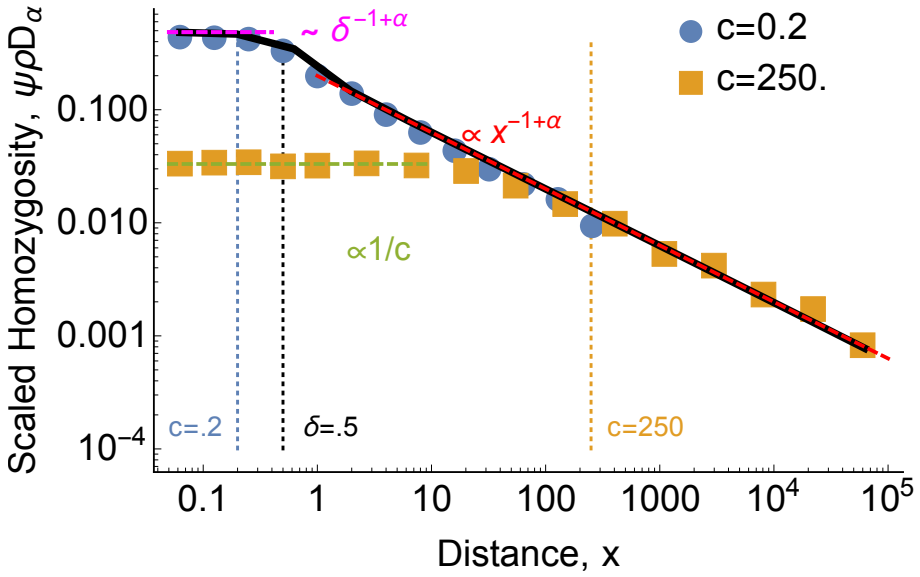

Figure 9: **For very long-range dispersal, $\alpha < 1$, continuous-time and discrete-time models differ at short distances.** Scaled probability of identity ψ as a function of distance x for $\alpha = 0.5$, $\delta = 0.5$, and $\rho = 100$. Points show discrete-time simulation results. For the continuous-time model, the black curve shows the result of numerically integrating (23), while the dashed red and magenta lines show the asymptotic approximations (6) and (7), 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 (65) (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).