A peer-reviewed version of this preprint was published in PeerJ on 29 March 2016.

View the peer-reviewed version (peerj.com/articles/1848), which is the preferred citable publication unless you specifically need to cite this preprint.

Furstenau TN, Cartwright RA. 2016. The effect of the dispersal kernel on isolation-by-distance in a continuous population. PeerJ 4:e1848 https://doi.org/10.7717/peerj.1848

The effect of the dispersal kernel on isolation-by-distance in a continuous population

Tara N Furstenau, Reed A Cartwright

Under models of isolation-by-distance, population structure is determined by the probability of identity-by-descent between pairs of genes according to the geographic distance between them. Well established analytical results indicate that the relationship between geographical and genetic distance depends mostly on the neighborhood size of the population, $N_b = 4\pi\sigma^2 D_e$, which represents a standardized measure of gene flow. To test this prediction, we model local dispersal of haploid individuals on a two-dimensional torus using seven dispersal kernels: Rayleigh, Exponential, Half-normal, Triangular, Gamma, Lomax and Pareto. When neighborhood size is held constant, the distributions produce similar patterns of isolation-by-distance, confirming predictions. Considering this, we propose that the triangular distribution is the appropriate null distribution for isolation-by-distance studies. Under the triangular distribution, dispersal is uniform within an area of $4\pi\sigma^2$ (i.e. the neighborhood area), which suggests that the common description of neighborhood size as a measure of a local panmictic population is valid for popular families of dispersal distributions. We further show how to draw from the triangular distribution efficiently and argue that it should be utilized in other studies in which computational efficiency is important.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

The effect of the dispersal kernel on isolation-by-distance in a continuous population

Tara N. Furstenau[1] and Reed A. Cartwright[1]

[1]The Biodesign Institute and School of Life Sciences, Arizona State University, Tempe, AZ 85287-5301, USA

Corresponding Author:
Reed A. Cartwright
PO Box 875301, Tempe, AZ, 85287-5301, USA
Email Address: Cartwright@asu.edu

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

The effect of the dispersal kernel on isolation-by-distance in a continuous population

Tara N. Furstenau[1] **and Reed A. Cartwright**[1]

[1]**The Biodesign Institute and School of Life Sciences, Arizona State University, Tempe, AZ 85287-5301, USA**

ABSTRACT

Under models of isolation-by-distance, population structure is determined by the probability of identity-by-descent between pairs of genes according to the geographic distance between them. Well established analytical results indicate that the relationship between geographical and genetic distance depends mostly on the neighborhood size of the population, $N_b = 4\pi\sigma^2 D_e$, which represents a standardized measure of gene flow. To test this prediction, we model local dispersal of haploid individuals on a two-dimensional landscape using seven dispersal kernels: Rayleigh, exponential, half-normal, triangular, gamma, Lomax and Pareto. When neighborhood size is held constant, the distributions produce similar patterns of isolation-by-distance, confirming predictions. Considering this, we propose that the triangular distribution is the appropriate null distribution for isolation-by-distance studies. Under the triangular distribution, dispersal is uniform within an area of $4\pi\sigma^2$ (i.e. the neighborhood area), which suggests that the common description of neighborhood size as a measure of a local panmictic population is valid for popular families of dispersal distributions. We further show how to draw from the triangular distribution efficiently and argue that it should be utilized in other studies in which computational efficiency is important.

Keywords: neighborhood size, identity-by-descent, kinship coefficients, triangular distribution, correlograms, simulation, fine scale, individual based

INTRODUCTION

For many populations, individuals do not exist in discrete patches or demes; instead they are spread across a continuous landscape. Although there are no barriers separating individuals, dispersal distances are often limited, and individuals that are near one another in space will be more similar genetically than individuals further apart. This phenomenon is known as isolation-by-distance and introduces a spatial component that should be considered when studying population genetic processes (Jongejans et al., 2008). Unfortunately, incorporating multiple dimensions of space at fine scales into analytical models is often analytically intractable (Epperson et al., 2010). Therefore, many researchers have turned to spatially-explicit, individual-based computer simulations which offer a more flexible way to incorporate spatial complexity into biological models (e.g. Barton et al., 2013; Cartwright, 2009; Epperson, 2003; Novembre and Stephens, 2008; Rousset, 2004; Slatkin, 1993).

A dispersal kernel describes the distribution of Euclidean distances between birth site and reproduction site. Ideally, when modeling dispersal, the dispersal distribution would be selected based on how well it fits the dispersal kernel estimated from natural populations. Classically, dispersal has been modeled as a diffusion process with Gaussian displacement; however, the observed dispersal kernels in many species tend to be more leptokurtic with a higher probability of short and long distance dispersal (Bateman, 1950). In plants, the shape of the dispersal kernel near the origin depends on the mechanism of dispersal; for example, there may be a high peak near the origin for gravity or animal dispersal whereas there may be a minimum near the origin for wind dispersal (Barluenga et al., 2011; Clark et al., 2005). Wide-tailed dispersal kernels, with a higher probability of long-distance dispersal, are a good fit to many empirical data sets (Klein et al., 2006; Bullock and Clarke, 2000; Clark et al., 2005; Gonzàlez-Martìnez et al., 2006; Martìnez and Gonzàlez-Taboada, 2009). Such long-distance dispersal events have a large impact when modeling a number of population processes including rate of expansion (Kot et al., 1996; Clark et al.,

2001), response to environmental changes (Nathan et al., 2011), local adaptation (Kuparinen et al., 2010), speciation (Hoelzer et al., 2008), and the spatial distribution of genetic diversity (Ibrahim et al., 1996; Bialozyt et al., 2006).

While the shape of the dispersal kernel impacts many population processes at different scales, it remains unclear how it effects patters of isolation-by-distance within a continuous population. It has been argued that the number of long-distance dispersal events will not have a noticeable effect because new long-distance alleles are more likely to be lost due to drift than become established at the new location (Ibrahim et al., 1996; Epperson, 2007). On the other hand, the shape of the dispersal kernel near the origin may have a significant impact on the overall rate of migration. In a plants, this could result in a higher probability of self-fertilization and/or a reduction in the number of successful offspring when there is density dependent regulation (Barluenga et al., 2011; Moyle, 2006; Howe et al., 1985).

Isolation-by-distance theory predicts that the actual shape of the dispersal distribution does not have a significant effect on the evolution of isolation-by-distance. Instead, levels of isolation-by-distance are determined by the mean squared distance of dispersal (i.e. non-central second moment of Euclidean distance), denoted by $2\sigma^2$ (Barton et al., 2013; Malécot, 1969; Rousset, 1997, 2004; Wright, 1946). Therefore, it should be possible to select any dispersal distribution as long as $2\sigma^2$ stays constant. Unfortunately, Rousset (1997, 2000) points out that the relationship is more complicated and does depends on the shape of the dispersal distribution for short distances, but for many classes of distributions, $2\sigma^2$ is likely the only parameter that matters.

The shape of the dispersal kernel is often not considered when modeling isolation-by-distance in continuous population. Here we develop a spatially-explicit, individual-based model to simulate local dispersal in a continuous population to determine if patterns of isolation-by-distance vary based on the shape of several different dispersal distributions: Rayleigh, half-normal, exponential, triangular, gamma, Lomax and Pareto. Each dispersal distribution has a different shape but they can be parameterized such that their non-central second moment is $2\sigma^2$. If the simulations reveal a similar pattern of isolation-by-distance across all dispersal distributions, we can conclude that, for a wide range of dispersal distributions, $2\sigma^2$ is the main determining factor of genetic differentiation with distance in a continuous population. Consequently, when designing isolation-by-distance simulations, researchers may choose a dispersal distribution based on computational needs instead of biological fit.

Wright (1946) uses the term "neighborhood" to describe a local population from which parents are randomly drawn. He measures the effective size of the neighborhood, N_b, as the inverse of the probability that two gametes at the same location came from the same parent. Assuming dispersal is normally distributed along each axis, he calculated that $N_b = 4\pi\sigma^2 D_e$, where D_e is the effective density of individuals, and $2\sigma^2$ is the mean squared distance of dispersal. — In his model this captures 86.5% of parents of central individuals. — Although Wright assumed Gaussian dispersal, his formula can be used to calculate N_b for many different dispersal models. N_b is important because it helps define the rate of decay of genetic similarity over spatial distance, i.e. the amount of isolation-by-distance in a population (Rousset, 1997, 2000; Barton et al., 2013).

If a neighborhood is supposed to represent a local panmictic unit, then in the ideal model parents should be chosen uniformly from a circle of radius 2σ centered on an offspring, and the Euclidean distance between parents and offspring should follow a triangular distribution: $f(r;\sigma) = r/(2\sigma^2)$, where $2\sigma^2$ is again the non-central second moment. This type of neighborhood is similar to the neighborhood defined in the spatially continuous Λ-Fleming-Voit disc model in which a number of parents, v, are chosen uniformly at random from a disc with radius r to replace a fraction u of the population (Barton et al., 2013). In this model, neighborhood size is defined by the ratio v/u and the individuals occupying the disc constitute a panmictic population. If 100% of the population is replaced ($u = 1$), the definition of neighborhood size reduces to the number of individuals competing for the central location.

Below, we demonstrate that patterns of isolation-by-distance are equivalent for different dispersal kernels with the same second moment, and discuss the use of the triangular distribution to model dispersal in a continuous population.

METHODS

Simulation

In our individual-based simulation, a population exists on a 100×100 rectangular lattice. Individuals are uniformly spaced with a single individual per cell. Each individual contains one haploid locus. The initial

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

population of 10,000 individuals each carry a unique allele. Generations are discrete, and individuals reproduce by generating 15 clonal offspring that experience mutations according to the infinite alleles model at rate μ. All starting and mutant alleles were selectively neutral.

The offspring disperse from the parent cell following a given dispersal distribution. The boundaries of the landscape are absorbing so when offspring disperse off of the lattice they are lost. When offspring land on a valid destination cell, they are immediately accepted or rejected using a reservoir sampling method to avoid storing them all in memory (Vitter, 1985). Using this method we were able to record some information about the probability of identity-by-descent of offspring competing for the same cell by keeping track of two randomly chosen offspring.

When dispersal is complete, there will be a maximum of one uniformly sampled offspring per cell and that offspring will become a parent in the next generation. We determined that when each parent generated 15 offspring, the number of empty cells per generation was negligible so we could assume a constant homogeneous population density.

Modeling Dispersal

The simulation is spatially-explicit with space represented on a rectangular lattice. Due to the discrete nature of the lattice, the dispersal kernels will be discretized approximations of continuous distributions (Chesson and Lee, 2005; Chipperfield et al., 2011). The dispersal kernel function, $f(r,\theta;\sigma)$, takes a parameter σ and returns continuous polar coordinates. The σ parameter is the square root of one-half the second moment of dispersal distance. The polar coordinates include the angle, $\theta \in [0,2\pi]$, which is uniformly distributed to ensure isotropic dispersal and distance, $r > 0$, which is drawn from a continuous distribution.

Once the angle and distance are drawn, the final position is determined by converting the polar coordinates into rectangular coordinates and adding them to the parent's position. The new coordinates are then rounded to determine the integer coordinates of the destination cell. This dispersal scheme is similar to the centroid-to-area approximation of continuous dispersal kernels described by Chipperfield et al. (2011), which showed minimal deviation from expectations especially when cell length is less than the expected distance.

We looked at seven different dispersal distance kernels (Table 1): Rayleigh, exponential, half-normal, triangular, gamma, Pareto, and Lomax. We chose these distributions because they provide a range of shapes for short, intermediate, and long distance dispersal.

The Rayleigh is a distribution of Euclidean distances that result from bivariate normal displacement along the x and y axis. The Rayleigh distribution follows the assumptions of Wright (1946)'s two-dimensional isolation-by-distance model.

The exponential distribution is more leptokurtic with higher probability of dispersal at short and long distances and less at intermediate distances. The exponential tail is the boundary that separates truly heavy-tailed distributions with potentially infinite higher moments from distributions with all moments finite. The distinction is important because leptokurtic, heavy-tailed dispersal kernels are typically a better fit to observed dispersal in nature (Clark, 1998).

The half-normal distribution is equivalent to a normal distribution that has been folded over the y-axis. In this case, Euclidean distance is simply the absolute value of normally distributed random variables. The half-normal is a monotonically decreasing distribution with a convex shoulder near zero. This distribution has a higher probability of dispersal at intermediate distances compared to the exponential.

The triangular distribution is typically defined using three points: a lower limit, a, an upper limit, b, and a mode, c. Here we use a special case of the triangular distribution where $a = 0$ and $b = c = 2\sigma$. The triangular distribution is the only one of our distributions that has a finite range, $r \in [0,2\sigma]$. In two-dimensions, the triangular dispersal kernel returns positions that are uniformly sampled from a circle with area $4\pi\sigma^2$.

Unlike the previous single parameter distributions, the final three distributions have an additional α shape parameter. The gamma distribution is equivalent to the exponential distribution when $\alpha = 1$, and as α increases the distribution becomes more symmetrical with a higher probability for intermediate distances and a lower probability for short distances.

The Lomax and Pareto distributions are both heavy-tailed power-law distributions. The n-th moments are finite only when $\alpha > n$. The support for the Pareto distribution, $r \in [x_{min}, +\infty)$, begins at a parameter $x_{min} > 0$. The Lomax distribution is a special case of the Pareto distribution that has been shifted so that

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

the support begins at zero. We chose values of α between 2 and 3 so that the second moment of the distribution would be finite but higher moments are infinite.

The dispersal function is executed over 100-billion times per simulation, and thus it was important to make the implementation as efficient as possible. With this aim in mind, we used an xorshift algorithm for uniform pseudo-random number generation and the ziggurat rejection sampling algorithm to draw values from exponential and normal distributions when applicable (Marsaglia and Tsang, 2000; Marsaglia, 2003). The bounded range of the triangular distribution allowed us to simulate dispersal using an efficient discrete sampling method. See the appendix for a description of the algorithm.

Analysis

Simulations were run for each of the seven dispersal distributions under 4 levels of dispersal (σ = 1, 1.5, 2 and 4) with a mutation rate of $\mu = 10^{-4}$. The simulations were run for 2 million generations after a burn-in period of 10,000 generations to allow the populations to reach a mutation-drift equilibrium. After the burn-in, data was collected every 1,000 generations to decrease correlations between samples. A straight transect of 50 individuals was sampled from the center of the landscape to avoid measuring edge effects.

For each sampled transect, all possible pairs of individuals were placed into distance classes based on the distance between them. The number of pairs that shared an identical allele was determined and recorded as a proportion of the total number of pairs in the distance class. The probabilities for each distance class were then averaged over all sampled generations. Under this sampling scheme, it should be pointed out that the number of pairs per distance class decreases as distance increases so in distance class 50 there is only one pair sampled per generation.

The parameters for each dispersal distribution were calculated so that $E[X^2] = 2\sigma^2$; the calculations are reflected in the probability distribution functions in Table 1. Due to the discrete nature of the lattice, some parameters values were adjusted slightly until the simulations produced an average squared distance between parent and offspring, s^2, that was within 5% of the expected value, σ^2. Three of the distributions require a second α parameter. For the gamma distribution we used $\alpha = 1, 2, 4, 8$. For the Lomax and Pareto distributions we used $\alpha = 2.4, 2.6, 2.8, 3.0$ all of which are infinite in the 3rd and higher moments.

Under isolation-by-distance, individuals geographically near one another will tend to be genetically similar, and this similarity will decrease as the distance between pairs of individuals increases. Therefore, isolation-by-distance is described by constructing correlograms of genetic similarity between individuals versus the distance between them. Genetic similarity can be measured using identity-by-descent, identity-by-state, relatedness, conditional kinship, or F-coefficients and can be based on coalescent times, an ancestral population, or the current population (Hardy and Vekemans, 1999; Hardy, 2003; Malécot, 1969; Rousset, 1997, 2002; Wang, 2014). For two-dimensional populations, genetic similarity is often plotted against the log-distance separating pairs because theory predicts that this relationship is approximately linear (Barton et al., 2013; Hardy and Vekemans, 1999; Rousset, 2000).

We recorded the probability of identity-by-descent for pairs of individuals in each distance class. Under the infinite alleles model, pairs of individuals were considered identical-by-descent if they shared the same allele. The probability of identity-by-descent in each distance class depends on the mutation rate; the probability will be greater when there are fewer alleles. For more consistent results that are nearly independent of mutation rate, the probability of identity is often calculated as a ratio that measures genetic similarity (or differentiation) relative to a particular reference group. We calculated the kinship coefficient which measures the correlation of genetic similarity between pairs of individuals a certain distance apart relative to the genetic similarity in the whole sample.

$$F_r = \frac{p_{ij} - \bar{p}}{1 - \bar{p}} \approx \frac{E[T] - E_{ij}[T]}{E[T]} \tag{1}$$

Here p_{ij} is the probability of identity-by-descent between haploid individuals i and j at distance r and $\bar{p}$ is the probability of identity-by-descent between random haploid individuals in the current sample (Hardy and Vekemans, 1999). The kinship coefficient is related to differences in the expected coalescent times, T, between a specific pair of individuals and a random pair in the population (Barton et al., 2013). Kinship coefficients were calculated for each transect and then averaged across transects for each distance class. Since this statistic is highly dependent on the sampling scheme, we sampled the same transect in all simulations.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

We also calculated the a_r parameter of Rousset (2000):

$$a_r = \frac{p_0 - p_{ij}}{1 - p_0} \tag{2}$$

which measures genetic differentiation over distance relative to the probability of identity-by-descent within a location. The a_r parameter is independent of sampling scheme, but it does depend on the level of local identity-by-descent, p_0, in the population such that a_r approaches infinity as p_0 approaches one (Vekemans and Hardy, 2004). Typically, p_0 is estimated from the amount of autozygosity in the population; however, we estimated p_0 as the probability that an individual shared an allele with one of the offspring that it competed with for the cell, which is suitable for haploid organisms and better fits its definition (Vekemans and Hardy, 2004).

For each simulation, we calculated the average number of unique alleles in a 50-individual transect ($\bar{k}$) and the average squared distance between parents and offspring ($2s^2$). Using $\bar{k}$, we estimated the population-level diversity, $\hat{\theta}_k$ (Ewens, 2004, eq. 9.32) and estimated effective haploid population size as $\hat{N}_e = \hat{\theta}_k / 2\mu$ and effective density as $\hat{D}_e = \hat{N}_e / A$, where $A = 10{,}000$.

Finally, we estimated haploid neighborhood size using two different methods. First we used our estimated demographic parameters to calculate neighborhood size as the product $\hat{N}_b = 4\pi s^2 \hat{D}_e$. We then estimated neighborhood size using the regressions of both F_r and a_r on the log of distance. The slope of the a_r regression is an estimate of $1/2\pi\sigma^2 D_e$ and the slope of F_r regression is an estimate of $-(1-F_0)/2\pi\sigma^2 D_e$ (Barton et al., 2013; Hardy and Vekemans, 1999; Rousset, 2000).

RESULTS

Behavior of Dispersal Distributions

Figure 1 shows the empirical cumulative distributions generated from 10,000 simulated dispersal events from each distribution. The probability of not dispersing from the original cell is indicated by the height of the left-most horizontal line for each distribution. The more leptokurtic distributions (exponential, gamma-1 and Lomax) with a high probability peak near zero have a much higher probability of not dispersing from the original cell, especially when σ is low. The Pareto distribution, which has a fat tail but has been shifted so it does not have a peak at zero, has a very low probability of not dispersing. Under the gamma distribution as the α parameter increases, the probability of remaining at the origin decreases; when $\alpha = 8$ the probability is nearly zero for all values of σ.

The average squared parent-offspring dispersal distance, s^2, observed for each distribution was very similar with a relative error of less than 5% from the expected σ^2 value (Table 2); however, the distribution of these values over sampled generations varied (Fig. 2A). Expectedly, the thin tailed or no-tail (triangular) dispersal distributions have the smallest variance because their properties are easier to represent with a small number of samples. The Lomax distribution has the highest variance with the median falling slightly below the expected value.

Figure 2B shows the distribution of the average cubed parent-offspring dispersal distances, s^3, for each transect. The theoretical third moment of the Lomax and Pareto distributions is infinite, while it is not possible to simulate this on a finite landscape, we do observe values of s^3 that are several orders of magnitude larger than distributions with finite third moments. The distribution of s^2 and s^3 for the Lomax and Pareto distributions both have a large positive skew.

Allelic Diversity

The distribution of the number of unique alleles is similar for most of the dispersal kernels with the median falling near the expected value under the infinite alleles model (Fig. 3). The expected number of alleles under the infinite alleles model (gray horizontal line) is equal to $\sum_{i=0}^{n-1} \theta/(\theta+i) = 7.03$ where $n = 50$ is the number of individuals in the sampled transect. The Lomax distributions have a higher median number of alleles at lower values of σ but this gets closer to the expected value when $\sigma > 2$. The average diversity is also slightly elevated for the exponential and gamma-1 simulations.

Differences in effective population size between simulations can be measured by comparing the number of unique alleles observed in the transects. Different dispersal kernels produce similar levels of diversity, except for the Lomax distributions which have a higher θ_k and consequently a larger effective population size (Table 2).

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

Spatial Autocorrelation and Isolation-by-Distance

To quantify the patterns of isolation-by-distance, we first measured the average probability of identity-by-descent for each sampled population as a function of distance. All dispersal kernels produced very similar patterns of isolation-by-distance especially for larger distance classes (Fig. 4). The probability of identity-by-descent is higher at small distance when σ is small but the relationship flattens out when $\sigma = 4$. Differences between the different dispersal distributions become apparent when the distance between individuals is small. The more leptokurtic dispersal distributions have a steeper incline as distance decreases and they have a higher probability of autozygosity at distance class zero. The plots for the triangular distribution nearly perfectly overlap the plots for the Rayleigh distribution in all cases.

Because the probability of identity-by-descent is sensitive to differences in the number of alleles present in the sample, we also calculated the pairwise kinship coefficient over the log of distance (Fig. 5). The kinship coefficient shows how much more or less similar pairs of individuals in a given distance class are compared to the sample as a whole. The kinship coefficient is nearly independent of differences in allele number and there is much better overlap of the plots for the different dispersal distributions. When the kinship coefficient is plotted against the log of distance there is a negative linear relationship over a certain range of distances (Hardy and Vekemans, 1999). The slope of this linear range is also similar across distributions for each value of σ.

Finally, we plot Rousset (2000)'s a_r parameter against the log of distance. There is a positive linear relationship between a_r and the log of distance (Fig. 6). The slope of a_r is fairly similar among the dispersal distributions for a given value of σ. However, there is less overlap in the plots for the different dispersal distributions because the overall amount of differentiation varies.

Estimated Neighborhood Size

As mentioned previously, the populations with Lomax dispersal tend to have a greater number of unique alleles and this translates to higher $\hat{\theta}_k$, higher effective population size, and ultimately higher effective density. Because s^2 was similar for each dispersal kernel, this means that the estimated neighborhood size using demographic estimates, $\hat{N}_{b(\theta_k)}$, was slightly higher for the Lomax dispersing populations (Table 2). Otherwise, the estimates for the other dispersal distributions are similar and close to the expected values.

Isolation-by-distance theory predicts that the slope of the regression of certain measures of genetic similarity and the log of distance will provide an estimate of $4\pi\sigma^2 D$ for haploid populations in two-dimensions. We calculated the slope over a distance range of 5 to 35 for each group of simulations. Both of the slope methods using a_r and F_r statistics provide identical neighborhood size estimates so only the estimates using a_r are reported in Table 2 as $\hat{N}_{b(a_r)}$. The estimates for the Lomax dispersal populations appear to be much closer to the expected value and more similar to the other populations.

DISCUSSION

Approximating continuous dispersal on a discrete lattice will introduce obvious biases when the dispersal distance is small compared to the scale of the lattice nodes (Chipperfield et al., 2011). This bias can be seen in Fig. 1 by the jagged nature of the empirical cumulative distribution (ECDF) (especially when σ is small) compared to the CDF of the continuous distribution. In the simulation, the distance between nodes is one lattice unit so dispersal has to exceed at least a distance of 0.5 lattice units to leave the original cell. For Lomax simulations with small σ, the high probability density near zero falls rapidly before a distance of 0.5 lattice units has been reached. This means that the majority of dispersal events do not leave the parent cell. The Pareto and Lomax distributions share a similar shape and a wide tail, but unlike the Lomax distribution, the mode of the Pareto is greater than zero and almost all dispersal events leave the original cell. We refer back to the differences between the Lomax and the Pareto when we discuss whether we can differentiate results that are specific to dispersal with a high peak at zero or are more general to wide-tailed dispersal.

Allelic diversity is near the expected value predicted by the infinite alleles model for most distributions. The Lomax distributions tend to have a higher number of alleles up until $\sigma = 4$. This appears to be in agreement with Maruyama (1972) which showed that the effective population size is larger than the census size when $\sigma < 1$ which is the case in many of the Lomax simulations (Fig. 2). Because the median allele number for the Pareto simulations falls near the expected value, it seems likely that the higher allelic diversity in the Lomax simulations is due to the high probability of not dispersing. This is supported by the fact that the average diversity is slightly higher for the exponential and gamma-1 as well. When

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

dispersal is unlikely to occur outside of the original cell, the number of migrants is low and the pool of offspring before competition will consist mostly of offspring from the same parent. It is unlikely that migrants will become established at their new location after competition and thus more alleles will be maintained.

Much of the theory of isolation-by-distance in continuous populations is based on infinite or periodic lattice models. Here we simulated dispersal in a continuous population occupying a finite lattice with absorbing boundaries to better understand the effect of the dispersal kernel on isolation-by-distance models on a more natural landscape. As expected under isolation-by-distance, the probability of identity-by-descent between neutral alleles in pairs of individuals decreases as the distance between them increases. When neighborhood size is small, the relationship is very pronounced with a high initial probability that quickly declines. As neighborhood size increases ($\sigma = 4$), this relationship nearly disappears. This is similar to two-dimensional stepping stone models that show strong differentiation between populations when $Nm << 1$ and little differentiation when $Nm > 4$ (Kimura and Maruyama, 1971).

Simulations with our different dispersal kernels show a strikingly similar pattern of isolation-by-distance. However, theory predicts that when distance is small, deviation in the shape of the dispersal kernel relative to the Rayleigh distribution will become important (Rousset, 1997, 2000). This is evident in our results when we compare the probabilities of identity-by-descent at small distances between the different dispersal kernels. When the dispersal kernel is leptokurtic, the probability is higher between individuals occupying the same location and it is slightly lower for short distances compared to the Rayleigh results. The pattern of identity-by-descent in other distributions, including the triangular are nearly identical to the Rayleigh. The situation is similar for the pairwise kinship except there is even greater similarity between the different dispersal kernels.

Rousset (2008) makes it clear that the increase of genetic differentiation with distance is robust to the shape of the dispersal kernel but the overall magnitude of differentiation will depend on the shape of the kernel. Looking at the relationship between a_r and the log of distance for our simulations, we can see that the slope for each distribution is similar for larger distance values but the plots are shifted up or down depending on kurtosis. Compared to the other wide tailed distributions, the Pareto distribution is not shifted upward due to the lack of dispersal at the origin. The a_r statistic is a ratio that compares the amount genetic differentiation between individuals at certain distance to the differentiation within a single individual. When the probability of identity-by-descent within an individual is high, the differentiation between neighbors will appear much higher due to a steep initial drop in identity. As a result, the a_r statistic will be greater for leptokurtic distributions even if the actual probability of identity is similar to other distributions.

As expected, the neighborhood-size estimates are similar to the expected value for all simulations. Neighborhood size was slightly higher for the Lomax simulations when using allele diversity to estimate effective density. Otherwise, the slopes of the regression methods were similar and thus predicted similar neighborhood sizes. This reconfirms that neighborhood size is a robust descriptor of the decrease of genetic identity with distance. It also seems clear that fat-tailed dispersal kernels do not have much of an effect in isolated continuous populations.

The triangular dispersal model can serve as a null model for the probability that two lineages will meet and coalesce in a previous generation. Identity-by-descent may be defined as the total probability of coalescence between the current generation, t_0, and a generation at some time t in the past (Rousset, 2002). When a population is not panmictic due to limited dispersal, the time to coalescence depends on the probability that the two lineages will move close enough together so that there is some probability that they shared a parent in the previous generation. When the dispersal kernel has an infinite tail, there is always some small probability that two individuals coalesce even if they are very far apart. Under triangular dispersal, the probability that two individuals will coalesce in the previous generation is $1/(4\pi\sigma^2 D)$ if they are separated by a distance less than 2σ and zero otherwise.

The triangular distribution has not been considered a reasonable distribution to use for modeling biological dispersal. However, as discussed previously, it arises from the simple assumption that dispersal is locally panmictic, making it potentially useful. When we compared the triangular distribution against more popular dispersal models, there were no significant differences between the resulting patterns of isolation-by-distance. Furthermore, the bounded range of the triangular distribution allows us to simulate dispersal more efficiently than other dispersal kernels. The algorithm we provide in the Appendix C allows us to bypass the costly conversion from polar to Cartesian coordinates by using a discrete sampling

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

method. This method requires some complicated pre-processing but it allows us to draw a relative dispersal position in constant time by using alias method look-up tables (Vose, 1991). Both these theoretical and computation concerns suggest that triangular distributions should be included in the molecular ecologists toolkit.

Our results suggest that the relationship between probability of identity-by-descent and distance is similar for a wide range of dispersal kernels in a continuous population. These results should not be taken to mean that it is always safe to ignore the shape of the dispersal kernel. The shape of the tail, in particular, can impact many population processes but it does not appear to have a strong effect on isolation-by-distance in a finite, isolated population. Because speed is an important factor in deploying isolation-by-distance simulations in analytical contexts, e.g. approximate Bayesian computation, we recommend using the triangular distribution when long distance dispersal and other features of the dispersal kernel can safely be ignored.

DATA ACCESSIBILITY

Source code for simulations can be found at `https://github.com/tfursten/IBD/tree/vpub`. Simulation parameters and results can be accessed at `http://dx.doi.org/10.6084/m9.figshare.1611097`.

ACKNOWLEDGMENTS

The authors would like to thank R. Schwartz, D. Winter and S. Wu for helpful comments, and K. Dai for programing tips. The authors declare that they have no competing interests. The authors received funding from the School of Life Sciences, Arizona State University for this work.

REFERENCES

Barluenga, M., Austerlitz, F., Elzinga, J. A., Teixeira, S., Goudet, J., and Bernasconi, G. (2011). Fine-scale spatial genetic structure and gene dispersal in *Silene latifolia*. *Heredity*, 106:13–24.

Barton, N. H., Etheridge, A. M., Kelleher, J., and Véber, A. (2013). Inference in two dimensions: Allele frequencies versus lengths of shared sequence blocks. *Theoretical Population Biology*, 87:105–119.

Bateman, A. J. (1950). Is gene dispersal normal? *Heredity*, 4:353–363.

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

Brent, R. P. (2007). Some long-period random number generators using shifts and xors. *ANZIAM Journal*, 48:C118–C202.

Bullock, J. M. and Clarke, R. T. (2000). Long distance seed dispersal by wind: Measuring and modelling the tail of the curve. *Oecologia*, 124:506–521.

Cartwright, R. A. (2009). Antagonism between local dispersal and self-incompatibility systems in a continuous plant population. *Molecular Ecology*, 18:2327–2336.

Chesson, P. and Lee, C. T. (2005). Families of discrete kernels for modeling dispersal. *Theoretical Population Biology*, 67:241–256.

Chipperfield, J. D., Holland, E. P., Dytham, C., Thomas, C. D., and Hovestadt, T. (2011). On the approximation of continuous dispersal kernels in discrete-space models. *Methods in Ecology and Evolution*, 2:668–681.

Clark, C. J., Poulsen, J. R., Bolker, B. M., Connor, E. F., and Parker, V. T. (2005). Comparative seed shadows of bird-, monkey-, and wind-dispersed trees. *Ecology*, 86:2684–2694.

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

Clark, J. S., Lewis, M., and Horvath, L. (2001). Invasion by extremes: Population spread with variation in dispersal and reproduction. *The American Naturalist*, 157:537–554.

Epperson, B. K. (2003). *Geographical Genetics*. Princeton University Press.

Epperson, B. K. (2007). Plant dispersal, neighbourhood size and isolation by distance. *Molecular Ecology*, 16:3854–3865.

Epperson, B. K., Mcrae, B. H., Scribner, K., Cushman, S. A., Rosenerg, M. S., Forin, M.-J., James, P. M. A., Murphy, M., Manel, S., Legendre, P., and Dale, M. R. T. (2010). Utility of computer simulations in landscape genetics. *Molecular Ecology*, 19:3549–3564.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

Ewens, W. J. (2004). *Mathematical Population Genetics I. Theoretical Introduction*. Springer Science+Business Media, Inc., 2 edition.

Gonzàlez-Martìnez, S. C., Burczyk, J., Nathan, R., Nanos, N., Gil, L., and Alìa, R. (2006). Effective gene dispersal and female reproductive success in Mediterranean maritime pine (*Pinus pinaster* Aiton). *Molecular Ecology*, 15:4577–4588.

Hardy, O. J. (2003). Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology*, 12(6):1577–1588.

Hardy, O. J. and Vekemans, X. (1999). Isolation by distance in a continuous population: Reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, 83:145–154.

Hoelzer, G. A., Drewes, R., Meier, J., and Doursat, R. (2008). Isolation-by-distance and outbreeding depression are sufficient to drive parapatric speciation in the absence of environmental influences. *PLoS Computational Biology*, 4:e1000126.

Howe, H. F., Schupp, E. W., and Westley, L. C. (1985). Early consequences of seed dispersal for a neotropical tree (*Virola surinamensis*). *Ecology*, 66:781–791.

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

Jongejans, E., Skarpaas, O., and Shea, K. (2008). Dispersal, demography and spatial population models for conservation and control managment. *Perspectives in Plant Ecology, Evolution and Systematics*, 9:153–170.

Kimura, M. and Maruyama, T. (1971). Pattern of neutral polymorphism in a geographically structured population. *Genetical Research*, 18:125–131.

Klein, E. K., Lavigne, C., Picault, H., Renard, M., and Gouyon, P.-H. (2006). Pollen dispersal of oilseed rape: Estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology*, 43:141–151.

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

Kuparinen, A., Savolainen, O., and Schurr, F. M. (2010). Increased mortality can promote evolutionary adaptation of forest trees to climate change. *Forest Ecology and Management*, 259:1003–1008.

L'Ecuyer, P. and Simard, R. (2007). TestU01: A C library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, 33(4).

Malécot, G. (1969). *The Mathematics of Heredity*. W. H. Freeman and Company.

Marsaglia, G. (2003). Xorshift RNGs. *Journal of Statistical Software*, 8.

Marsaglia, G. and Tsang, W. W. (2000). The ziggurat method for generating random variables. *Journal of Statistical Software*, 5.

Martìnez, I. and Gonzàlez-Taboada, F. (2009). Seed disperal patterns in a temperate forest during a mast event: Performance of alternative dispersal kernels. *Oecologia*, 159:389–400.

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

Moyle, L. C. (2006). Correlates of genetic differentiation and isolation by distance in 17 congeneric *Silene* species. *Molecular Ecology*, 15:1067–1081.

Nathan, R., Horvitz, N., He, Y., Kuparinen, A., Schurr, F. M., and Katul, G. G. (2011). Spread of North American wind-dispersed trees in future environments. *Ecology Letters*, 14:211–219.

Novembre, J. and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics*, 40(5):646–649.

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

Rousset, F. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology*, 13:58–62.

Rousset, F. (2002). Inbreeding and relatedness coefficients: What do they measure? *Heredity*, 88:371–380.

Rousset, F. (2004). *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press.

Rousset, F. (2008). Demystifying Moran's I. *Heredity*, 100:231–232.

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

Vekemans, X. and Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

plant populations. *Molecular Ecology*, 13:921–935.

Vitter, J. S. (1985). Random sampling with a reservoir. *ACM Transactions on Mathematical Software*, 11:37–57.

Vose, M. D. (1991). A linear algorithm for generating random numbers with a given distribution. *IEEE Transactions on Software Engineering*, 17:972–975.

Wang, J. (2014). Marker-based estimates of relatedness and inbreeding coefficients: An assessment of current methods. *Journal of Evolutionary Biology*, 27:518–530.

Wright, S. (1946). Isolation by distance under diverse systems of mating. *Genetics*, 31:39–59.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

A XORSHIFT RANDOM NUMBER GENERATOR

Xorshift is a type of pseudo-random number generator that relies on exclusive-or and bitshift operators (Marsaglia, 2003). Xorshift is one of the most efficient, high-quality random-number generators known. Our implementation is a 64-bit xorshift with shift parameters (5, 15, 27) added to a Weyl series to decrease bit correlations (Brent, 2007). It passes the BigCrush tests in the TestU01 suite (L'Ecuyer and Simard, 2007).

B GENERATING FROM A TRIANGULAR DISTRIBUTION

Inverse sampling can be used to generate values from a triangular distribution. — Note that we are only working with monotonically increasing triangular distributions and not more general formulations. — If u is uniformly distributed in $(0,1)$, the value $d = 2s\sqrt{u}$ has a triangular distribution with parameter s. However, a modified rejection sampling algorithm is faster. If u_1 and u_2 are independent and uniformly distributed in $(0,1)$, then $d = 2s\max(u_1,u_2)$ also has a triangular distribution. Because we can generate 32-bit values for both u_1 and u_2 from a single 64-bit random number, this second algorithm is more efficient than the first. While it is possible to construct a ziggurat algorithm (Marsaglia and Tsang, 2000) for a triangular distribution, our second algorithm is more efficient because it involves fewer steps and never rejects.

We compared the speed of these algorithms and a naive rejection sampler using the medium Crush tests (L'Ecuyer and Simard, 2007). This allowed us to compare the speeds of these algorithms in a data-intensive application as well as verify that the algorithms produced independent and identically distributed values from the correct distribution. The 'maximum' algorithm took 1656 seconds to complete, while the 'sqrt' took 1700s and the rejection sampler took 1911s. The maximum algorithm produced faster execution, but only sped up the tests by 3% over sqrt.

C GENERATING DISCRETE TWO-DIMENSIONAL DISPERSAL FROM A TRIANGULAR DISTRIBUTION

We can use the maximum algorithm above to generate the values in polar coordinates and convert them to Cartesian coordinates; however, this requires calculating sine and cosine functions, which we would rather not do. When modeling dispersal on a lattice, the bounded nature of the triangular distribution allows dispersal to be modeled discretely. To discretize this distribution on a rectangular lattice we must determine the probabilities for each cell which are proportional to the area of the cell that is covered by a disk of radius $r = 2\sigma$ (centered on a focal cell). The algorithm described here produces probability tables by calculating the appropriate area for each cell and dividing by the total area. We assume that cells are squares with unit area.

Since the disk is symmetrical, this algorithm may be simplified by calculating areas for quadrant I of the disk and mirroring those values to the other quadrants. We further simplify by calculating approximately half of the areas for quadrant I and mirroring those as well. — Note that this results in cells along the x and y axes having an area of 1/2. — Starting at the center of the focal cell ($y_0 = 0$), we record the top/bottom boundary of each cell along the y-axis up to the radius: $y_1 = 0.5, y_2 = 1.5, \ldots, y_n = n - 0.5$ where $n = \sup_{n\in\mathbb{Z}} y_n \leq r$.

Next we calculate the area of the first column of cells which has a left boundary at $x_0 = 0$ and a right boundary at $x_1 = \min(0.5, r)$:

$$A = \int_{x_0}^{x_1} \sqrt{r^2 - x^x}\, dx$$

Starting with the bottom cell, we check if the area of a cell is less than the area of the column. If so, the cell is completely contained in the disk, and the cell is assigned a weight equal to its area. Its area is then subtracted from the area of the column. We continue this procedure until the the area of last cell is less than the remaining area of the column and assign the final cell a weight equal to the remaining area in the column.

We then move to the next column by setting $x_0 = 0.5$ and $x_1 = 1.5$. However, before we calculate the area, we must check if the edge of the disk passes through the bottom of the top cell. This occurs if $x_1^2 + y^2 > r^2$, where y is the value of the bottom boundary of the cell. When this occurs, we split the

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

column into two smaller columns and each column is processed just like before. We continue calculating the area of subsequent columns until we reach the column that contains the point $\{x,y\} = \left\{r/\sqrt{2}, r/\sqrt{2}\right\}$, which marks the point where the edge of the disk intersects the diagonal. After this column is processed, the weights for these cells can be copied symmetrically. The weight of each cell is divided by the total area of the disk and becomes a probability. These probabilities are then copied symmetrically to the other three quadrants. The completed table of probabilities can then be passed into the alias algorithm for discrete sampling (Vose, 1991).

Our implementation of a discretized triangular kernel can be found in src/disk.h and src/disk.cpp in the source code. Code for generating an alias table can be found in src/aliastable.h.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

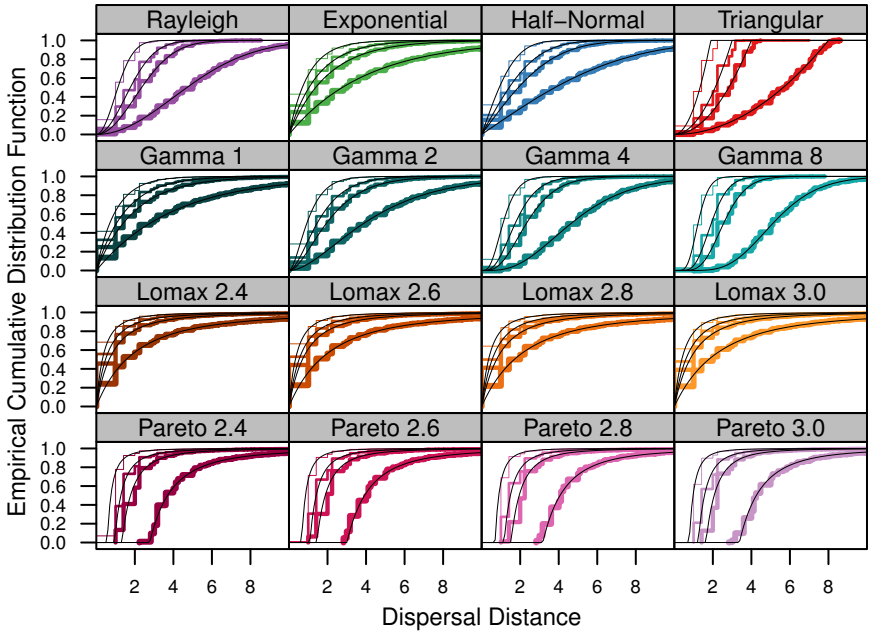


Figure 1. Discretization has a small effect on dispersal distributions. The empirical cumulative distribution function for each dispersal distribution on a discrete lattice compared to the CDF of its continuous counterpart (black line). The different plots in each panel represent simulations run using different σ parameters: 1, 1.5, 2, 4. An increase in the thickness of the line corresponds to increasing σ parameter.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

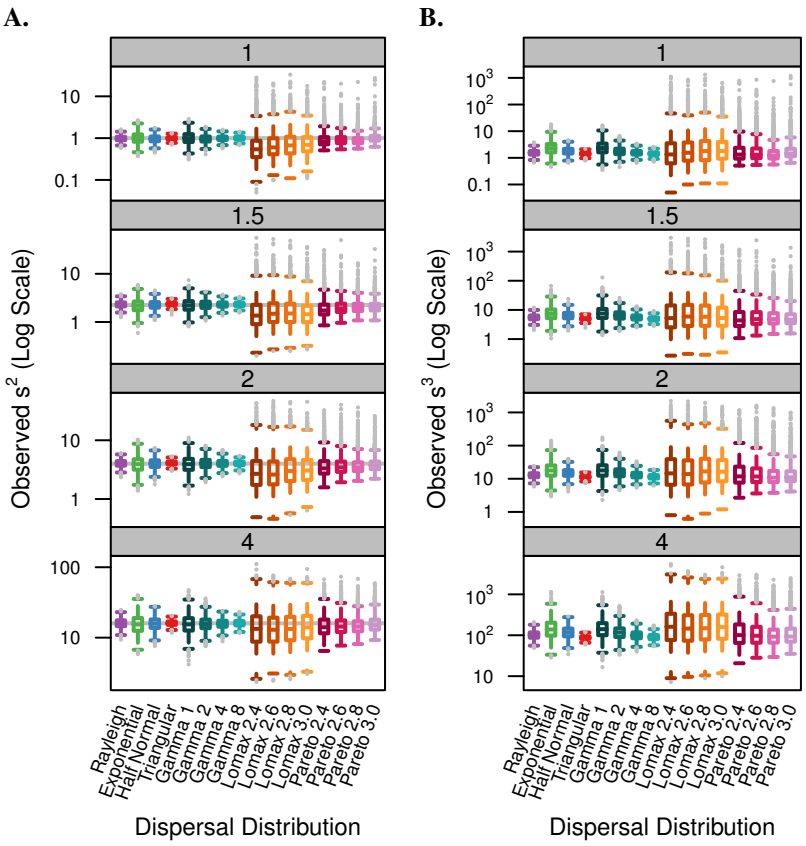


Figure 2. Different dispersal kernels have equivalent second moments but different third moments. Each panel represents groups of simulations run with different σ parameters and contains box-whisker plots summarizing the distribution of the average (A) squared or (B) cubed parent-offspring distance of 2,000 sampled transects. The top and bottom of the boxes represent the 75% and 25% quartiles, while the central bar represents the median. The gray dots outside the whiskers represent outliers. The gray horizontal line in A represents the expected σ^2 value. The observed values are shown on a log scale which is different in some panels.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

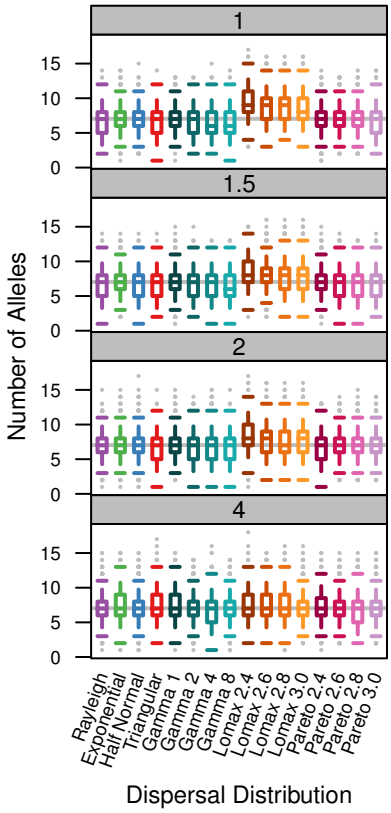


Figure 3. The distribution of unique alleles is similar for most dispersal kernels. Each panel represents simulations run with a the σ parameter provided in the gray box. For each dispersal distribution, the box-whisker plot summarizes the number of unique alleles (k) found in 2,000 50-individual transects. The gray horizontal line represents the expectation under the infinite alleles model. The features of the box-whisker summary are the same as Fig. 2.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

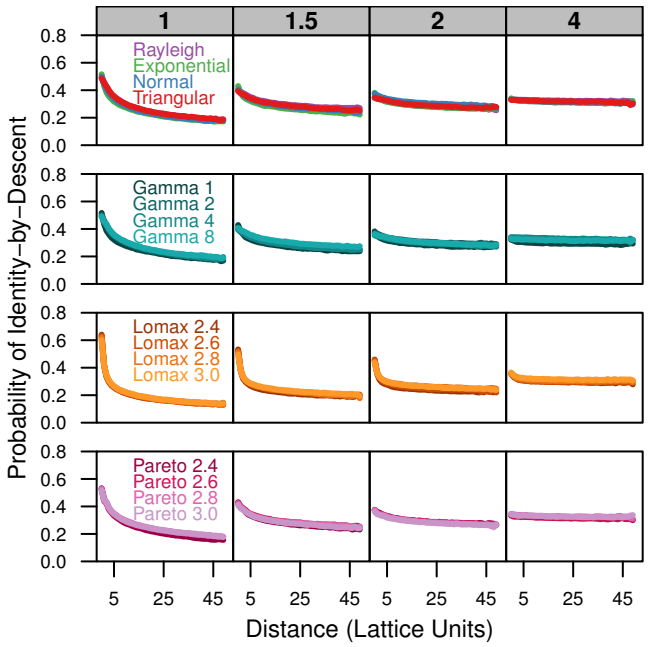


Figure 4. Identity-by-descent is similar between different dispersal models. Each plot shows the average probability of identity-by-descent for pairs of individuals in each distance class. Each panel represents simulations run with different σ parameters (gray box) for different groups of dispersal distributions.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

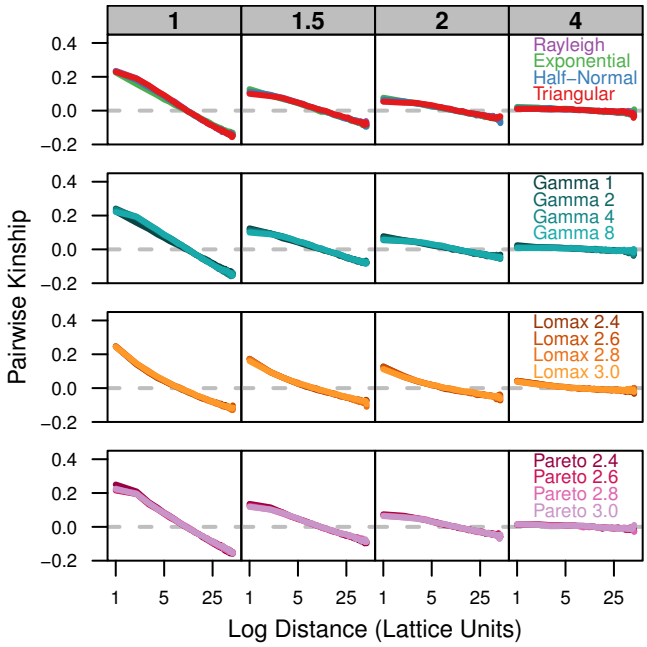


Figure 5. Kinship coefficients are similar between different dispersal models. Each plot shows the average kinship coefficient for pairs of individuals over the log of the distance between them. Each panel represents simulations run with different σ parameters (gray box) for different groups of dispersal distributions. The gray dashed line is at zero so values above the line are more similar than the sample as a whole while values below the line are less similar than the population as a whole.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

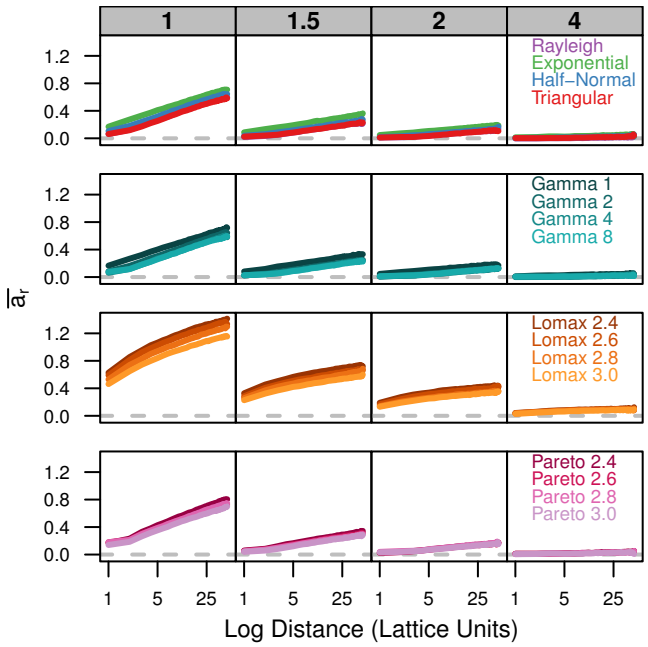


Figure 6. Slopes of genetic differentiation are similar between different dispersal models. Each plot shows the average differentiation, a_r, for pairs of individuals over the log of the distance between them. Each panel represents simulations run with different σ parameters (gray box) for different groups of dispersal distributions.

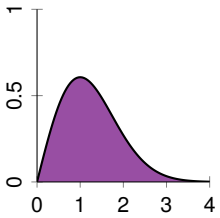
PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

Table 1. Dispersal Kernels. The dispersal function, range and probability density function for $\sigma = 1$.

Rayleigh

$$f(r,\theta;\sigma) = \frac{1}{2\pi}\frac{r}{\sigma^2}e^{\frac{-r^2}{2\sigma^2}}$$

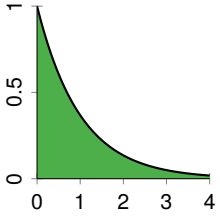
$$r \geq 0$$



Exponential

$$f(r,\theta;\sigma) = \frac{1}{2\pi}\frac{1}{\sigma}e^{\frac{-r}{\sigma}}$$

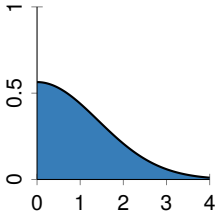
$$r \geq 0$$



Half-Normal

$$f(r,\theta;\sigma) = \frac{1}{2\pi}\frac{1}{\sigma\sqrt{\pi}}e^{\frac{-r^2}{4\sigma^2}}$$

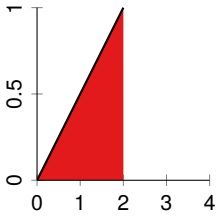
$$r \geq 0$$



Triangular

$$f(r,\theta;\sigma) = \frac{1}{2\pi}\frac{r}{2\sigma^2}$$

$$0 \leq r \leq 2\sigma$$

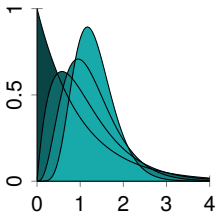


Gamma

$$f(r,\theta;\sigma,\alpha) = \frac{1}{2\pi}\frac{\left(\frac{\sqrt{a(a+1)}}{\sqrt{2}\sigma}\right)^{\alpha}}{\Gamma(a)}r^{\alpha-1}e^{-r\frac{\sqrt{a(a+1)}}{\sqrt{2}\sigma}}$$

$$\alpha = 1,2,4,8$$

$$r \geq 0$$

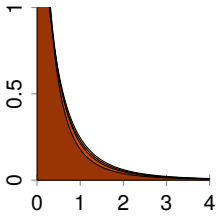


Lomax

$$f(r,\theta;\sigma,\alpha) = \frac{1}{2\pi}\frac{\alpha\left(1+\frac{r}{\sqrt{(\sigma^2(\alpha-2)(\alpha-1)}}\right)^{-(\alpha+1)}}{\sqrt{\sigma^2(\alpha-2)(\alpha-1)}}$$

$$\alpha = 2.4,2.6,2.8,3.0$$

$$r \geq 0$$

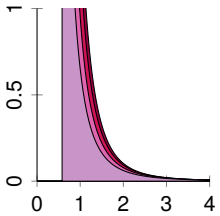


Pareto

$$f(r,\theta;\sigma,\alpha) = \frac{1}{2\pi}\frac{\alpha\sqrt{\frac{2\sigma^2(\alpha-2)}{\alpha}}}{r^{a+1}}$$

$$\alpha = 2.4,2.6,2.8,3.0$$

$$r \geq \sqrt{\frac{2\sigma^2(\alpha-2)}{\alpha}}$$



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015

Table 2. Estimated neighborhood sizes are similar across all dispersal distributions. Estimates of allele diversity, $\hat{\theta}_k$, effective population density, $\hat{D}_e$, dispersal, s^2, and neighborhood size. Neighborhood size is estimated two different ways. $\hat{N}_{b(\theta)}$ is $4\pi s^2 \hat{D}_e$ where $\hat{D}_e$ is estimated from $\hat{\theta}_k$. $\hat{N}_{b(a_r)}$ is twice the inverse of the slope of a_r and the log of distance. The expected neighborhood size ($4\pi\sigma^2 \cdot 1$) is 12.56, 28.28, 50.26, and 201.06 for σ =1, 1.5, 2, and 4, respectively.

	σ									
	1					1.5				
	$\hat{\theta}_k$	$\hat{D}_e$	s^2	$\hat{N}_{b(\theta_k)}$	$\hat{N}_{b(a_r)}$	$\hat{\theta}_k$	$\hat{D}_e$	s^2	$\hat{N}_{b(\theta_k)}$	$\hat{N}_{b(a_r)}$
Ray	1.82	0.91	0.99	11.31	13.07	1.83	0.91	2.33	26.79	31.16
Exp	2.09	1.04	1.04	13.70	14.32	2.04	1.02	2.26	29.04	29.00
Nor	1.94	0.97	0.98	11.94	13.49	1.91	0.95	2.31	27.69	30.61
Tri	1.82	0.91	1.00	11.37	13.58	1.83	0.92	2.36	27.18	31.02
Gam 1	2.07	1.04	1.05	13.63	14.41	2.01	1.00	2.32	29.22	30.07
Gam 2	1.89	0.94	0.98	11.62	12.80	1.85	0.92	2.32	26.98	30.13
Gam 4	1.83	0.92	1.00	11.49	12.88	1.87	0.94	2.32	27.31	28.27
Gam 8	1.80	0.90	1.01	11.45	13.31	1.79	0.90	2.32	26.16	29.91
Lom 2.4	2.97	1.49	1.06	19.70	13.41	2.62	1.31	2.16	35.53	26.65
Lom 2.6	2.88	1.44	0.97	17.61	13.23	2.47	1.24	2.34	36.25	26.11
Lom 2.8	2.73	1.36	1.04	17.78	12.82	2.41	1.21	2.22	33.66	25.30
Lom 3	2.72	1.36	1.00	17.07	14.28	2.36	1.18	2.34	34.71	28.50
Par 2.4	1.98	0.99	0.98	12.18	11.71	1.93	0.97	2.19	26.56	27.12
Par 2.6	1.95	0.98	1.04	12.74	13.82	1.81	0.91	2.28	25.98	27.95
Par 2.8	1.90	0.95	0.97	11.57	12.25	1.85	0.93	2.25	26.16	30.85
Par 3	1.89	0.95	0.99	11.80	13.56	1.89	0.94	2.24	26.54	29.79
	2					4				
	$\hat{\theta}_k$	$\hat{D}_e$	s^2	$\hat{N}_{b(\theta_k)}$	$\hat{N}_{b(a_r)}$	$\hat{\theta}_k$	$\hat{D}_e$	s^2	$\hat{N}_{b(\theta_k)}$	$\hat{N}_{b(a_r)}$
Ray	1.97	0.99	4.07	50.39	58.81	2.02	1.01	16.11	204.93	236.23
Exp	2.02	1.01	4.08	51.88	49.60	2.09	1.05	16.16	212.48	154.94
Nor	1.95	0.97	4.08	49.87	55.00	2.04	1.02	16.04	205.76	189.69
Tri	1.94	0.97	4.11	50.13	54.57	2.09	1.04	16.09	210.87	245.02
Gam 1	2.03	1.01	4.06	51.74	52.25	2.16	1.08	16.15	218.67	257.28
Gam 2	1.89	0.95	4.12	48.88	54.39	2.02	1.01	16.08	204.41	214.04
Gam 4	1.94	0.97	4.08	49.80	55.60	1.98	0.99	15.94	197.97	191.02
Gam 8	1.89	0.94	4.06	48.21	52.47	2.02	1.01	16.11	203.96	231.04
Lom 2.4	2.48	1.24	3.98	62.01	47.94	2.19	1.09	16.06	220.82	180.03
Lom 2.6	2.36	1.18	3.94	58.49	48.10	2.15	1.07	15.45	208.62	219.14
Lom 2.8	2.27	1.13	4.16	59.23	51.08	2.14	1.07	15.81	212.44	241.05
Lom 3	2.24	1.12	3.97	56.05	47.23	2.07	1.04	16.55	215.21	211.19
Par 2.4	1.93	0.97	4.13	50.12	48.20	2.03	1.02	16.04	204.74	192.65
Par 2.6	1.95	0.97	4.11	50.29	51.74	2.03	1.02	15.91	203.23	189.19
Par 2.8	1.98	0.99	4.02	49.95	47.73	1.95	0.97	15.53	189.90	219.90
Par 3.	1.98	0.99	4.10	50.92	49.58	2.01	1.00	16.30	205.48	169.53

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1545v1 | CC-BY 4.0 Open Access | rec: 30 Nov 2015, publ: 30 Nov 2015