

# The effect of the dispersal distribution on isolation by distance in continuous populations

Tara Furstenau

Population Genetics  
School of Life Sciences  
Arizona State University  
[tfursten@asu.edu](mailto:tfursten@asu.edu)

April 22, 2014



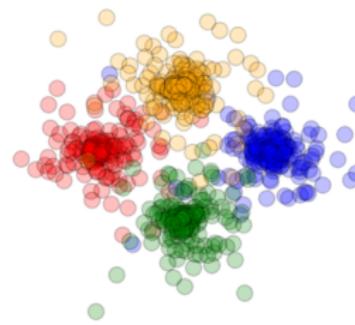
# Isolation by distance in continuous populations

- In natural populations, parent-offspring dispersal tends to be spatially limited.



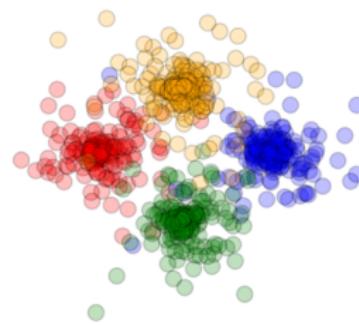
# Isolation by distance in continuous populations

- In natural populations, parent-offspring dispersal tends to be spatially limited.
- Consequently, the probability that two alleles from nearby individuals are descended from the same ancestor (identical by descent) is greater than under random mating.



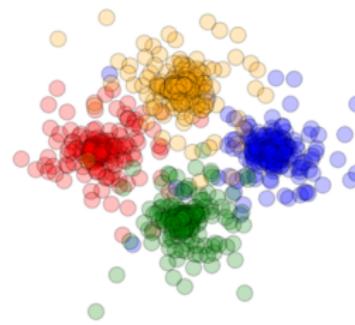
# Isolation by distance in continuous populations

- In natural populations, parent-offspring dispersal tends to be spatially limited.
- Consequently, the probability that two alleles from nearby individuals are descended from the same ancestor (identical by descent) is greater than under random mating.
- Over time, limited dispersal produces genetic correlations among spatially proximal individuals.



# Isolation by distance in continuous populations

- In natural populations, parent-offspring dispersal tends to be spatially limited.
- Consequently, the probability that two alleles from nearby individuals are descended from the same ancestor (identical by descent) is greater than under random mating.
- Over time, limited dispersal produces genetic correlations among spatially proximal individuals.
- Wright (1943) named this phenomenon isolation by distance.

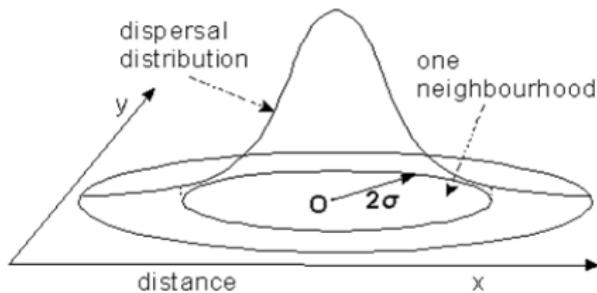


# Neighbourhood size partially quantifies fine-scale genetic structure

- Wright introduced the concept of a genetic neighbourhood to describe a local area where individuals may have the same probability of identity by descent.

# Neighbourhood size partially quantifies fine-scale genetic structure

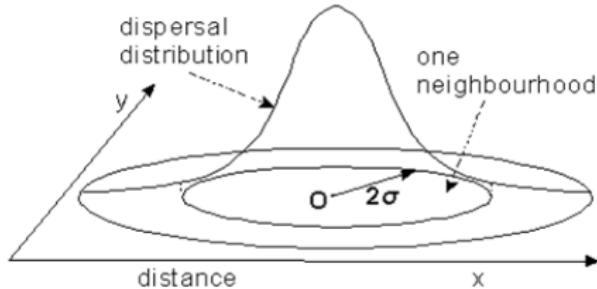
- The spatial dimensions of a neighborhood are a function of the variance of the parent-offspring dispersal distribution.



$$N_b = 4\pi\sigma^2$$

# Neighbourhood size partially quantifies fine-scale genetic structure

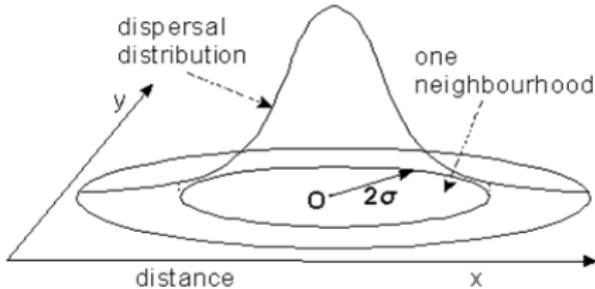
- The spatial dimensions of a neighborhood are a function of the variance of the parent-offspring dispersal distribution.
- Wright assumed that dispersal distances follow a zero-mean, normal distribution with variance  $\sigma^2$  along an axis.



$$N_b = 4\pi\sigma^2$$

# Neighbourhood size partially quantifies fine-scale genetic structure

- The spatial dimensions of a neighborhood are a function of the variance of the parent-offspring dispersal distribution.
- Wright assumed that dispersal distances follow a zero-mean, normal distribution with variance  $\sigma^2$  along an axis.
- The neighbourhood is then a circle with radius  $2\sigma$  and an area of  $A = 4\pi\sigma^2$  that contains 86.5% of parents of central individuals.



$$N_b = 4\pi\sigma^2$$

## The Question:

How well does neighborhood size predict patterns of isolation by distance when  $\sigma^2$  remains the same but the dispersal distribution is not normal?

# What is $\sigma^2$ ?

The  $\sigma^2$  in the neighbourhood size is the variance of dispersal measured relative to a single axis.

Assuming symmetric dispersal in two dimensions,  $x$  is the signed distance along one axis (axial distance):

$$E(x) = E(y) = 0 \text{ and}$$

$$\sigma^2 = \text{Var}(x) = E(x^2) - [E(x)]^2 = E(x^2)$$

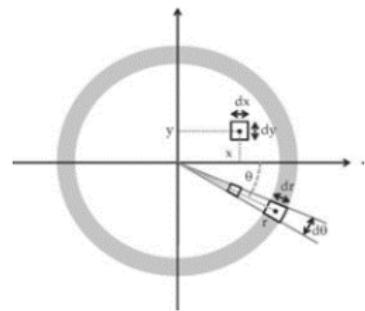
Euclidean distance,  $d$ , in two dimensions (absolute radial distance):

$$E(d^2) = E(x^2) + E(y^2) = 2E(x^2)$$

Therefore:

$$\sigma^2 = \frac{1}{2}E(d^2) = E(x^2)$$

$\sigma^2$  is not the variance of Euclidean dispersal distances.



# Dispersal

- Given the  $\sigma$  parameter, the dispersal function returns polar coordinates where  $\theta$  is a uniform direction ( $\frac{1}{2\pi}$ ) and  $r$  is the distance, determined by the given probability distribution.

## Example

$$E[X^2] = \int_0^\infty x^2 e^{-\lambda x} = \dots = \frac{2}{\lambda^2}$$

Set second moment of distribution equal to  $2\sigma^2$  and solve for  $\lambda$ .

$$\frac{2}{\lambda^2} = 2\sigma^2$$

$$\lambda = \frac{1}{\sigma}$$

# Dispersal

- Given the  $\sigma$  parameter, the dispersal function returns polar coordinates where  $\theta$  is a uniform direction ( $\frac{1}{2\pi}$ ) and  $r$  is the distance, determined by the given probability distribution.
- The parameter for each distribution is adjusted so that they all produce the same second moment of Euclidean distance.

## Example

$$E[X^2] = \int_0^\infty x^2 e^{-\lambda x} = \dots = \frac{2}{\lambda^2}$$

Set second moment of distribution equal to  $2\sigma^2$  and solve for  $\lambda$ .

$$\frac{2}{\lambda^2} = 2\sigma^2$$

$$\lambda = \frac{1}{\sigma}$$

# Dispersal

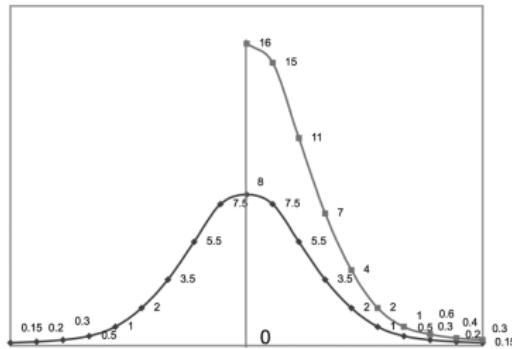
## Half-normal

$$f(r, \theta | \sigma) = \frac{1}{2\pi} \frac{\sqrt{2}}{\sigma \sqrt{\pi}} e^{\frac{-x^2}{2\sigma^2}}, x \geq 0$$

Where  $\sigma = \sigma \sqrt{2}$

The half-normal distribution is a folded normal distribution with mean zero.

$$Y = |X|$$

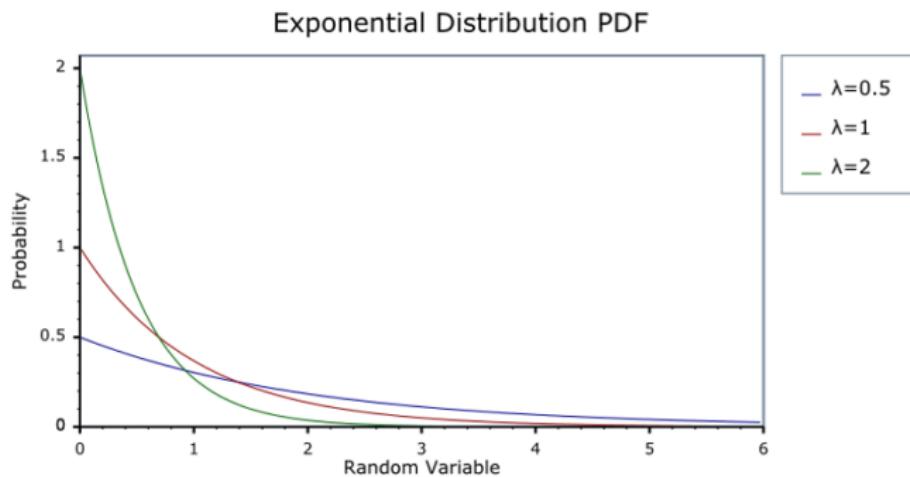


# Dispersal

## Exponential

$$f(r, \theta | \lambda) = \frac{1}{2\pi} \lambda e^{-\lambda x}, x \geq 0$$

Where  $\lambda = \frac{1}{\sigma}$



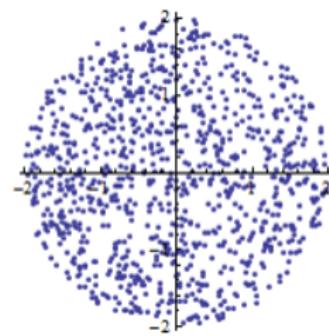
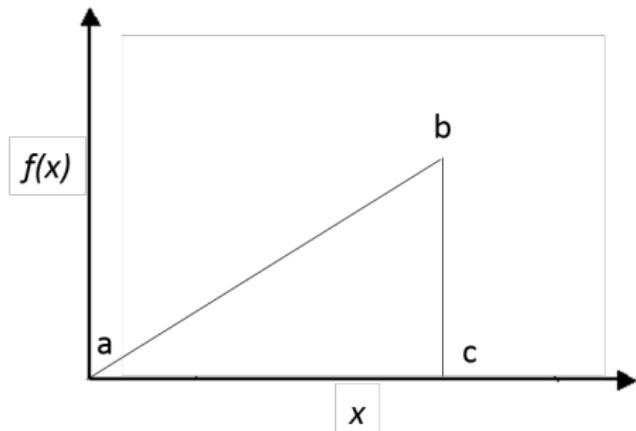
# Dispersal

## Triangular

$$f(r, \theta | c) = \frac{1}{2\pi} \frac{2x}{c^2}, 0 \leq x \leq c$$

Where  $c = 2\sigma$

A special case of the triangular distribution where  $a = 0$  and  $b = c$ .  
Only distribution that does not have an infinite tail.



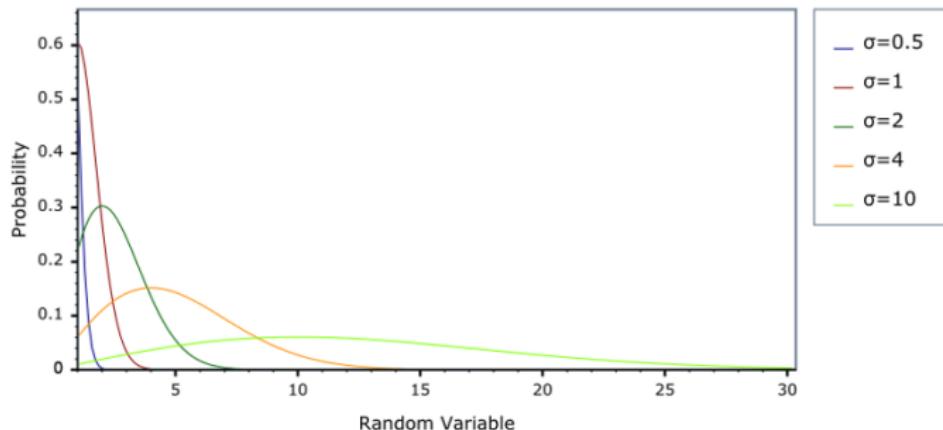
# Dispersal

## Rayleigh

$$f(r, \theta | \sigma) = \frac{1}{2\pi} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, x \geq 0$$

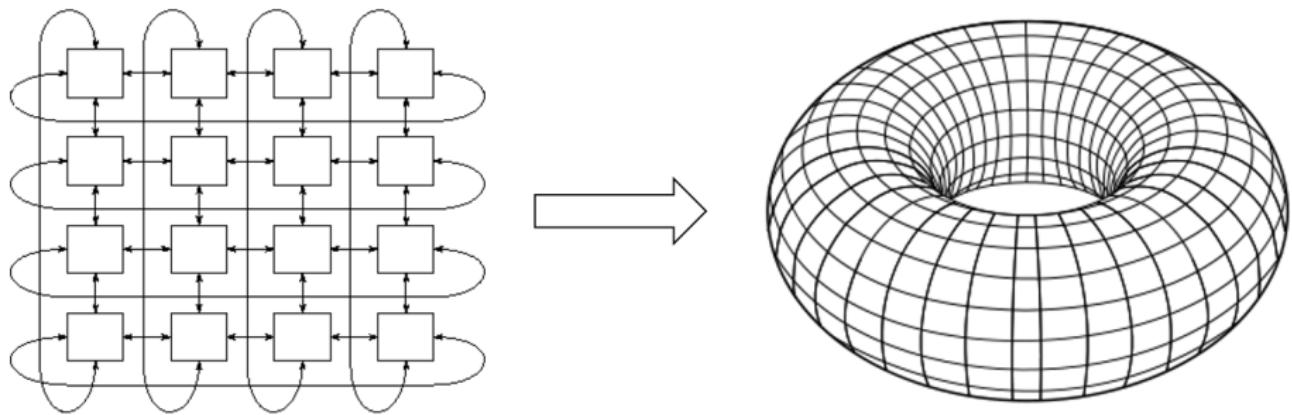
$R = \sqrt{X^2 + Y^2}$ , where  $X \sim N(0, \sigma^2)$  and  $Y \sim N(0, \sigma^2)$  are independent normal random variables.

Rayleigh Distribution PDF



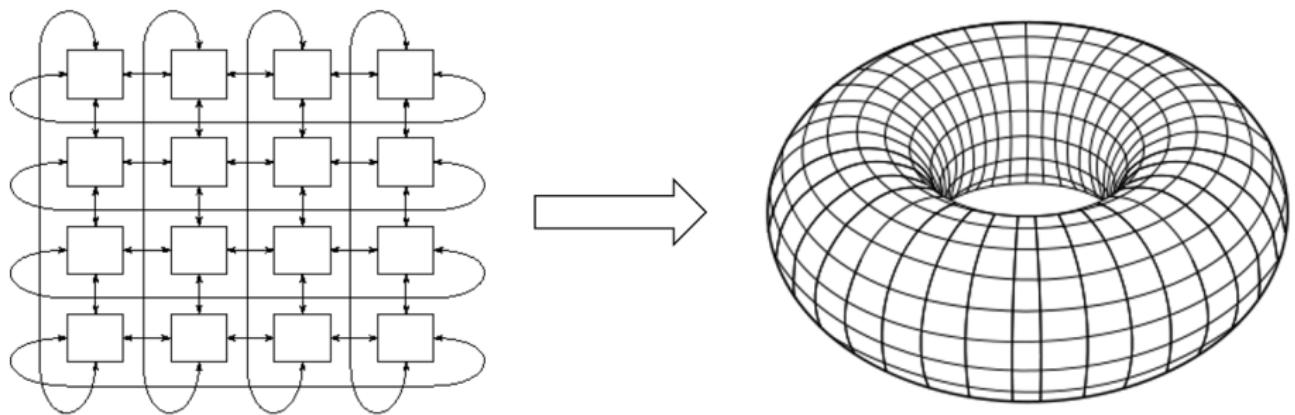
# Simulation

- The population exists on a  $100 \times 100$  toroidal lattice.



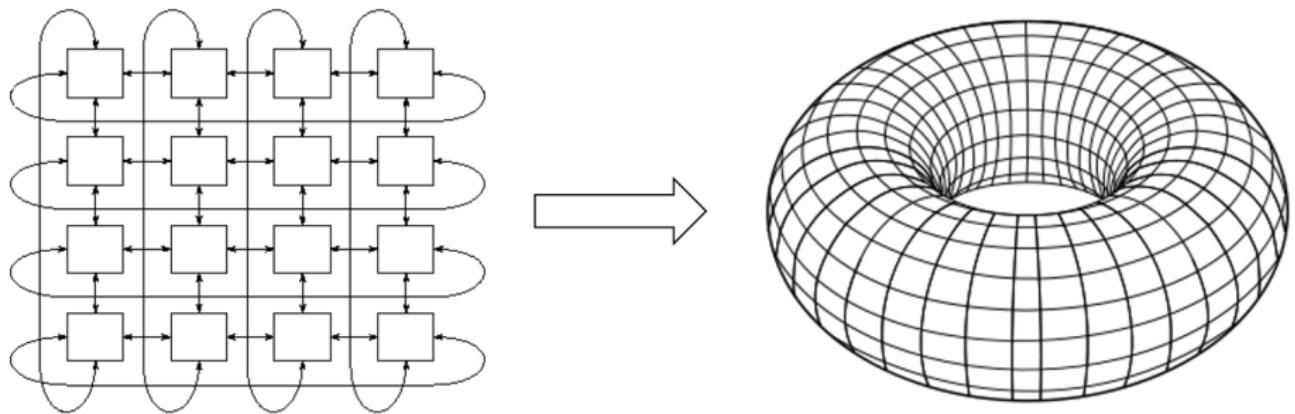
# Simulation

- The population exists on a  $100 \times 100$  toroidal lattice.
- Each cell is occupied by a haploid individual with a single locus.



# Simulation

- The population exists on a  $100 \times 100$  toroidal lattice.
- Each cell is occupied by a haploid individual with a single locus.
- The population is initialized so that each individual carries a unique allele.



# Generations

- During each generation cycle, every individual reproduced by generating  $N = 15$  clonal copies.

# Generations

- During each generation cycle, every individual reproduced by generating  $N = 15$  clonal copies.
- Those offspring experienced mutations according to the Infinite Alleles Model at rate  $\mu = 10^{-4}$ .

## Generations

- During each generation cycle, every individual reproduced by generating  $N = 15$  clonal copies.
- Those offspring experienced mutations according to the Infinite Alleles Model at rate  $\mu = 10^{-4}$ .
- The offspring were dispersed from their cell according to the distribution that was set for the simulation.

## Generations

- During each generation cycle, every individual reproduced by generating  $N = 15$  clonal copies.
- Those offspring experienced mutations according to the Infinite Alleles Model at rate  $\mu = 10^{-4}$ .
- The offspring were dispersed from their cell according to the distribution that was set for the simulation.
- In the case where multiple offspring landed on the same cell, one individual was uniformly selected using a reservoir sampling technique to optimize memory usage.

## Generations

- During each generation cycle, every individual reproduced by generating  $N = 15$  clonal copies.
- Those offspring experienced mutations according to the Infinite Alleles Model at rate  $\mu = 10^{-4}$ .
- The offspring were dispersed from their cell according to the distribution that was set for the simulation.
- In the case where multiple offspring landed on the same cell, one individual was uniformly selected using a reservoir sampling technique to optimize memory usage.
- When dispersal was completed the offspring occupying the cells became the parents of the next generation.

## Data Collection

- The simulations were run for a burn-in period prior to data collection to allow the populations to stabilize.

## Data Collection

- The simulations were run for a burn-in period prior to data collection to allow the populations to stabilize.
- After the burn-in, data from the population were collected every 1000 generations for a total of 100 samples.

## Data Collection

- The simulations were run for a burn-in period prior to data collection to allow the populations to stabilize.
- After the burn-in, data from the population were collected every 1000 generations for a total of 100 samples.
- The population was sampled by recording a transect of  $N = 100$  individuals straight across a row of the grid.

## Data Collection

- The simulations were run for a burn-in period prior to data collection to allow the populations to stabilize.
- After the burn-in, data from the population were collected every 1000 generations for a total of 100 samples.
- The population was sampled by recording a transect of  $N = 100$  individuals straight across a row of the grid.
- Each possible pairing of sampled individuals was made and then sorted into classes based on the minimum distance between them.

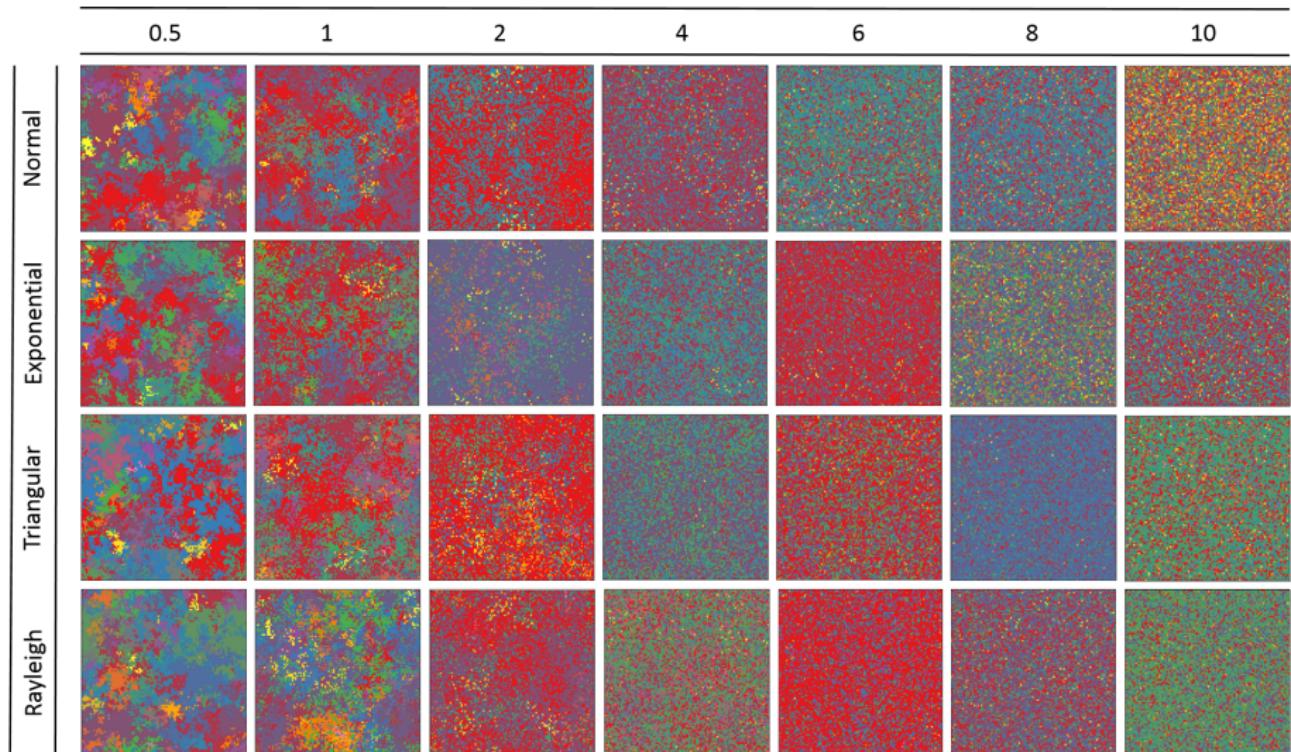
## Data Collection

- The simulations were run for a burn-in period prior to data collection to allow the populations to stabilize.
- After the burn-in, data from the population were collected every 1000 generations for a total of 100 samples.
- The population was sampled by recording a transect of  $N = 100$  individuals straight across a row of the grid.
- Each possible pairing of sampled individuals was made and then sorted into classes based on the minimum distance between them.
- For each distance class, the proportion of pairs of individuals that had identical alleles determined the probability of identity by descent.

## Data Collection

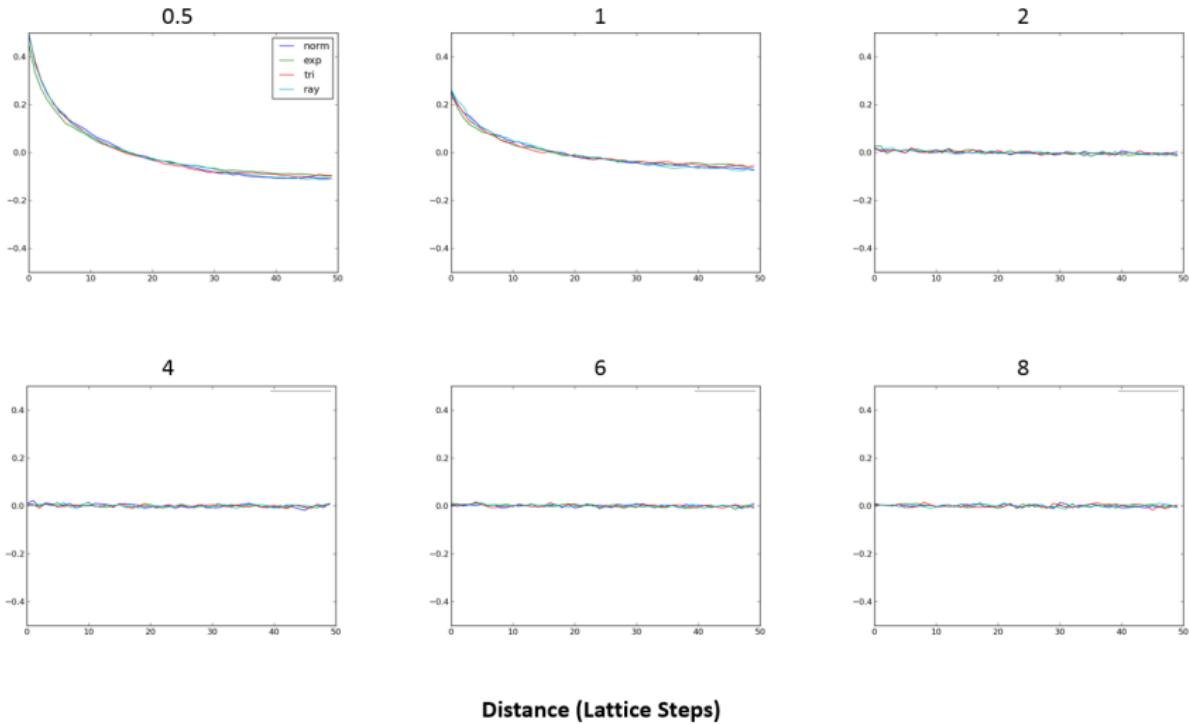
- The simulations were run for a burn-in period prior to data collection to allow the populations to stabilize.
- After the burn-in, data from the population were collected every 1000 generations for a total of 100 samples.
- The population was sampled by recording a transect of  $N = 100$  individuals straight across a row of the grid.
- Each possible pairing of sampled individuals was made and then sorted into classes based on the minimum distance between them.
- For each distance class, the proportion of pairs of individuals that had identical alleles determined the probability of identity by descent.
- The probability of identity was averaged over all 100 generations.

# Results: Spatial distribution of haplotypes



Results: The different distributions produce very similar patterns of isolation by distance

Probability of Identity by Descent



# Conclusions

- Each dispersal distribution produced similar patterns of isolation by distance in the population for  $\sigma$  ranging from 0.5 – 10.

# Conclusions

- Each dispersal distribution produced similar patterns of isolation by distance in the population for  $\sigma$  ranging from 0.5 – 10.
- This suggests that, at least for the conditions simulated here,  $\sigma^2$  is sufficient to determine the pattern of spatial differentiation.

# Conclusions

- Each dispersal distribution produced similar patterns of isolation by distance in the population for  $\sigma$  ranging from 0.5 – 10.
- This suggests that, at least for the conditions simulated here,  $\sigma^2$  is sufficient to determine the pattern of spatial differentiation.
- These results should be verified over a wider range of conditions by varying population size, shape and mutation rate.

# Conclusions

- Each dispersal distribution produced similar patterns of isolation by distance in the population for  $\sigma$  ranging from 0.5 – 10.
- This suggests that, at least for the conditions simulated here,  $\sigma^2$  is sufficient to determine the pattern of spatial differentiation.
- These results should be verified over a wider range of conditions by varying population size, shape and mutation rate.
- For future population simulations it may be possible to model dispersal using the less computationally intensive triangular distribution.