

# Simulations of E&R Under Neutrality and Selection- SLiM

Hannah Götsch & Yiwen Chen

# This Afternoon

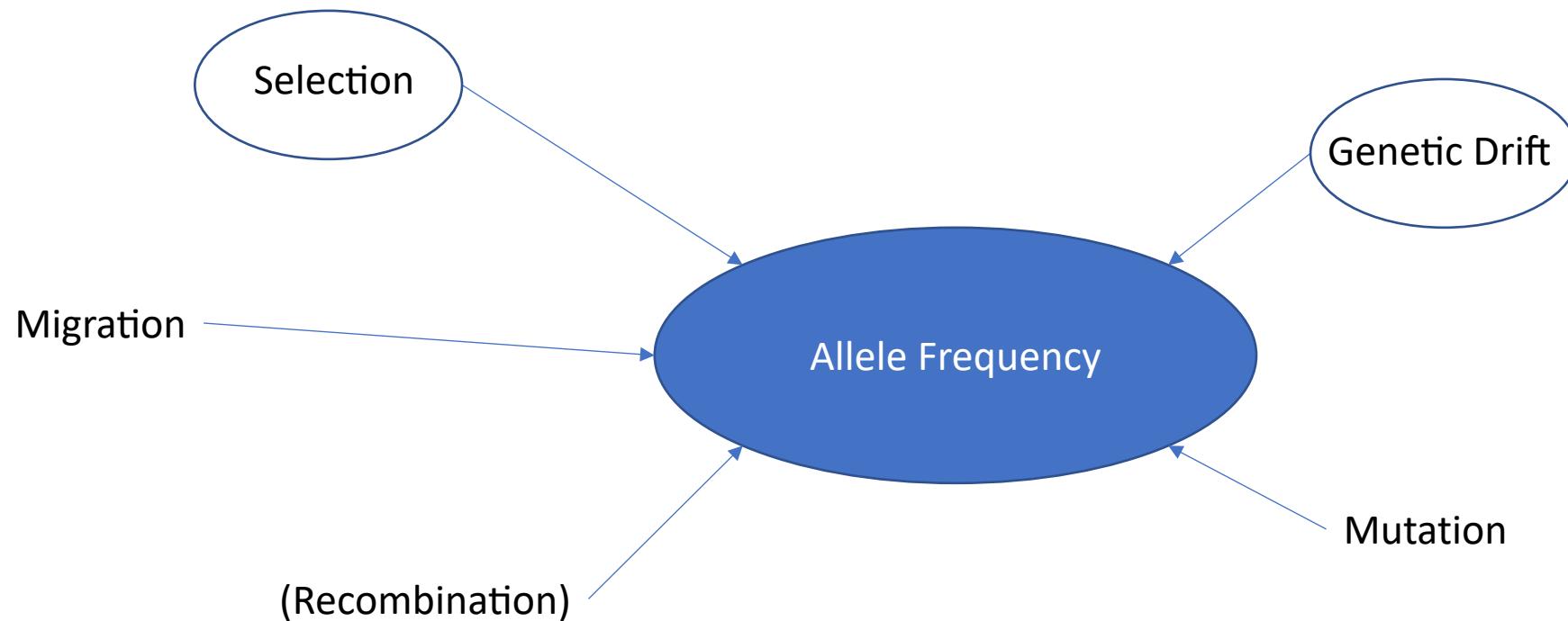
## 1. Neutrality

- Brief introduction to genetic drift and the Wright-Fisher model
- Brief introduction to SLiM
- First Task: Simulate drift on one locus
- Second Task: Simulate drift on multiple loci

## 2. Selection

- Directional, stabilizing and diversifying selection
- Third Task: Simulate directional selection
- Fourth Task: Simulate stabilizing selection
- Fifth Task: Simulate diversifying selection

# Evolutionary Forces



# Genetic Drift

# Genetic Drift

- Changes in allele frequencies due to stochastic (independent of external and heritable factors) sampling variation (in offspring number) in a finite population
- Several models have been proposed. The most commonly used models are the Wright-Fisher model and the Moran model.

# The Wright-Fisher Model – A Neutral Model

- Discrete time, non-overlapping generations (i.e. all individuals in the population are replaced by their descendants)
- Single locus with two alleles
- Population of constant size with  $N$  diploid individuals ( $2N$  chromosomes in the population) or  $2N$  haploid individuals → no dominance
- Alleles in the next generation are a random sample of alleles in the current generation → no selection ( $s = 0$ ), individuals are hermaphrodites

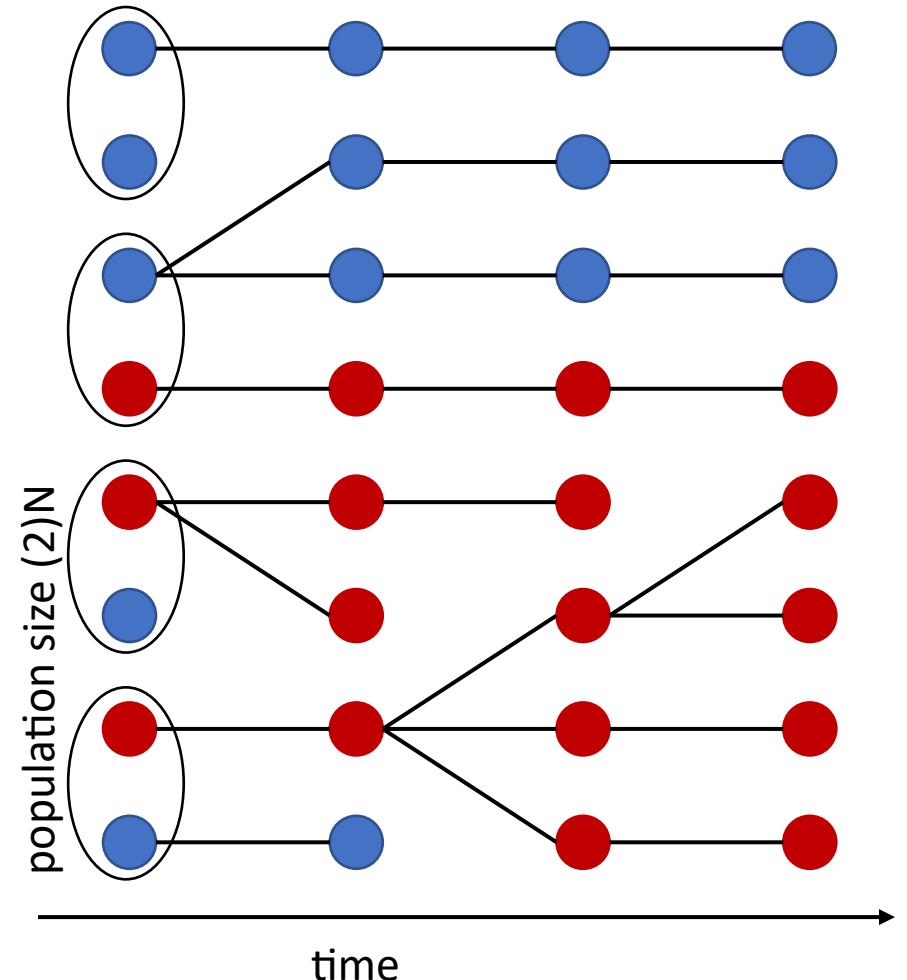
# The WF Model – Drift as Binomial Sampling

The binomial distribution:

$$P(X' = k \mid 2N, p) = \binom{2N}{k} p^k (1 - p)^{2N-k}$$

for  $k = 0, 1, 2, \dots, 2N$ ,

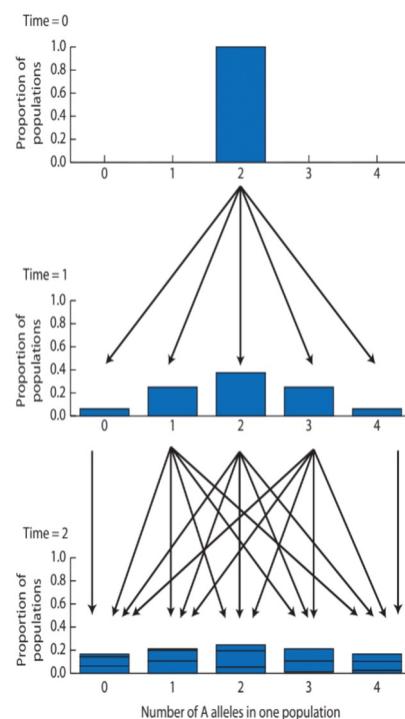
where  $\binom{2N}{k} = \frac{(2N)!}{k!(2N-k)!}$



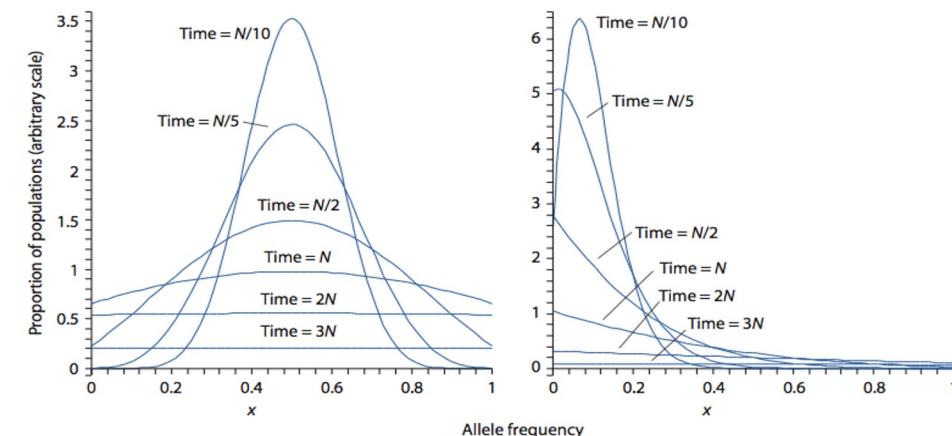
# “Estimate” Drift – Two Approaches

## Markov-Chain Model

- Using the binomial distribution
- Track the frequency changes of an allele by estimating the transition matrix

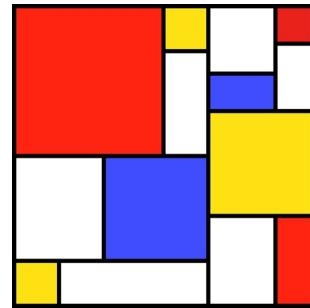


## Diffusion Approximation



**Figure 3.13** Probability densities of allele frequency for many replicate populations predicted using the diffusion equation. The initial allele frequency is 0.5 on the left and 0.1 on the right. Each curve represents the probability that a single population would have a given allele frequency after some interval of time has passed. The area under each curve is the proportion of alleles that are not fixed. Time is scaled in multiples of the effective population size,  $N$ . Both small and large populations have identically shaped distributions, although small populations reach fixation and loss in less time than large populations. The populations that have reached fixation or loss are not shown for each curve. For a color version of this image see Plate 3.13.

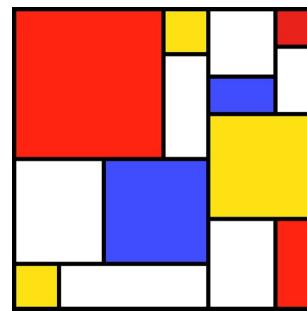
Figures from Hamilton, M. (2012). *Population genetics*. Chichester: Wiley-Blackwell.



# SLiM

# Selection on Linked Mutations

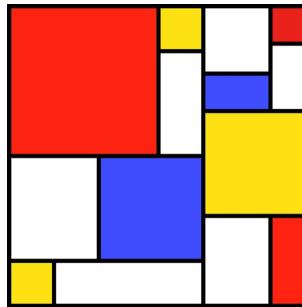
forward genetic simulation package  
from the Messer Lab at Cornell University



# SLiM - Materials

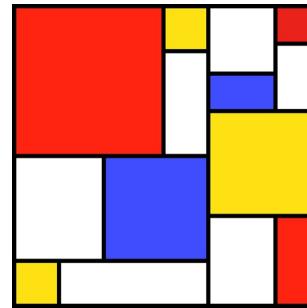
[www.messerlab.org/slim/](http://www.messerlab.org/slim/)

- SLiM Manual
- Eidos Manual
- Reference Sheets
- Workshop ([www.benhaller.com/workshops/workshops.html](http://www.benhaller.com/workshops/workshops.html))
- Mailing Lists (slim-announce & slim-discuss)



# Why SLiM?

- Open source on GitHub, GPL license
- Interactive & graphical (SLiMgui)
- Fast
- Flexible & customizable with Eidos
  - Eidos is similar to R in syntax & function names, but ends statements with semicolons and is zero-based



# How to use SLiM?

- Initial modeling in SLiMgui
  - SLiM Recipes
- Production runs (on a cluster)
- Post-run analysis in Eidos, Python, or R

(we use version 4.0.1 of SLiM)

## SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model

Benjamin C. Haller<sup>\*,1</sup> and Philipp W. Messer<sup>\*,1</sup>

<sup>1</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY

\*Corresponding authors: E-mails: bhaller@benhaller.com; messer@cornell.edu.

Associate editor: Ryan Hernandez

Received: 10 September 2018 | Revised: 6 November 2018 | Accepted: 9 November 2018  
DOI: 10.1111/1755-0998.12968

### RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

## Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes

Benjamin C. Haller<sup>1</sup> | Jared Galloway<sup>2</sup> | Jerome Kelleher<sup>3</sup> |

Philipp W. Messer<sup>1,\*</sup> | Peter L. Ralph<sup>2,\*</sup> |

## Evolutionary Modeling in SLiM 3 for Beginners

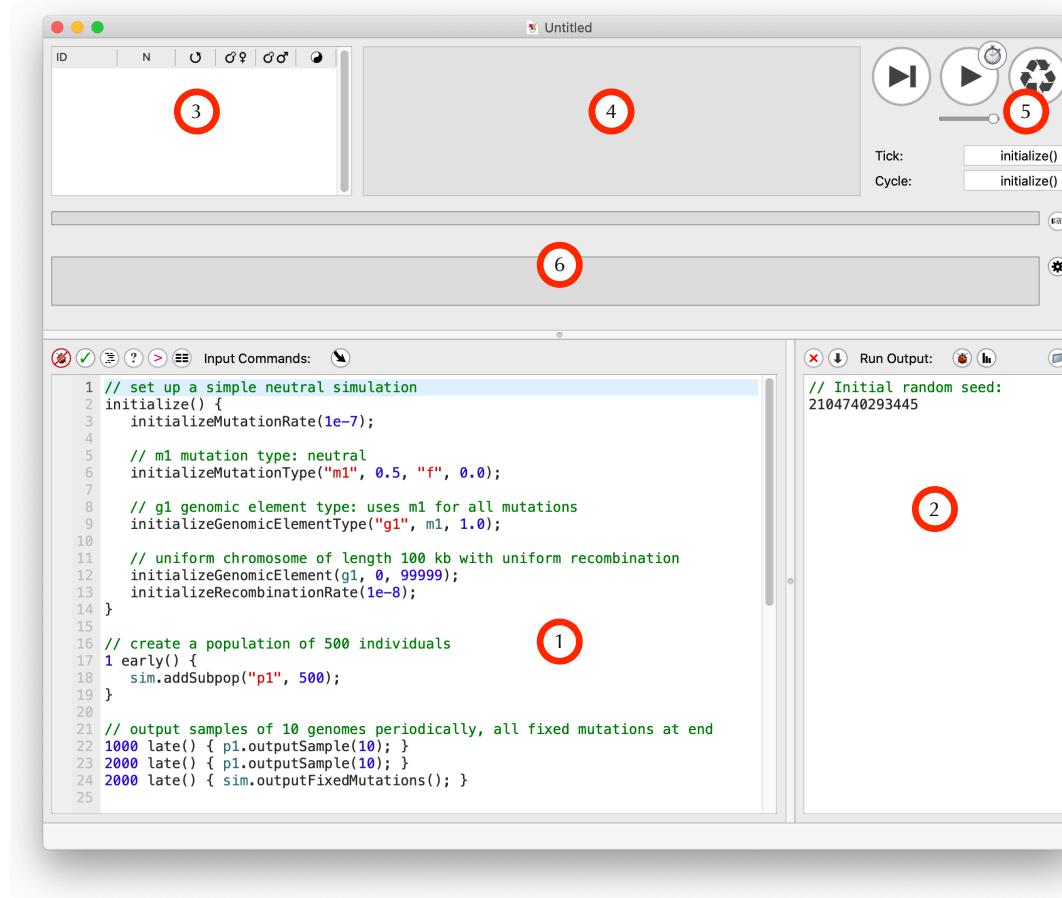
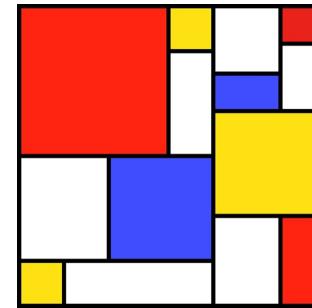
Benjamin C. Haller<sup>\*,1</sup> and Philipp W. Messer<sup>1</sup>

<sup>1</sup>Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY

\*Corresponding author: E-mail: bhaller@benhaller.com.

Associate editor: Ryan Hernandez

# SLiMgui Simulation Window



1. The scripting pane
2. The output pane.
3. The population view.
4. The individual view.
5. The tick controls.
6. The chromosome view.

Figure from the SLiM-Manual.

# First Task

Simulate Drift on One Locus

# Simulate Drift on One Locus

- Neutral Wright-Fisher model
- Single locus – two alleles
- Fixed mutant frequency in the beginning, no additional mutations

## Exercises:

1. Plot the allele frequency trajectory in SLiM. Vary the starting allele frequency ( $p$ ), the population size ( $N$ ) and the number of generations ( $t$ ).
2. When is an allele more likely to get fixed or lost? What is the effect of  $p$  and  $N$ ?
3. When do you expect an allele to remain the longest in the population (assuming the same  $N$ )?

# Multiple Replicates

## Exercises:

1. Plot the allele frequency trajectories in SLiM.
2. Try to understand how the mean and variance of drift are affected by the following parameters:
  - Starting allele frequency ( $p$ )
  - Population size ( $N$ )
  - Number of generations ( $t$ )
3. How is the observed heterozygosity affected by the same parameters?

# Second Task

Simulate Drift on Multiple Loci

# Simulate Drift on Multiple Locus

- Neutral Wright-Fisher model
- Multiple loci

## Exercises:

1. Plot the allele frequency trajectories in SLiM.
2. Plot the SFS in SLiM. How does it change with time?

# Selection

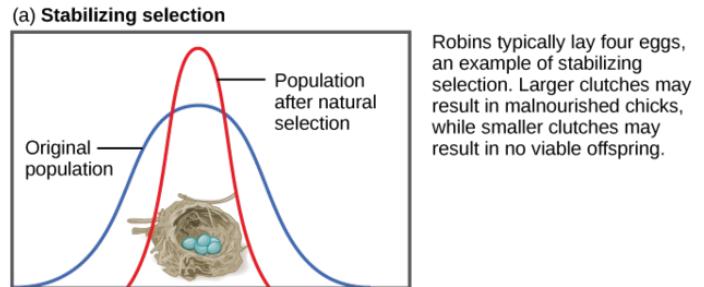
# Natural Selection

- Each locus is assigned with selection coefficients ( $s$ ). Allele with higher  $s$  result to higher individual fitness
  - h... dominance coefficient
  - w... individual fitness
    - $W_{AA}=1+s$
    - $W_{Aa}=1+hs$
    - $W_{aa}=1$

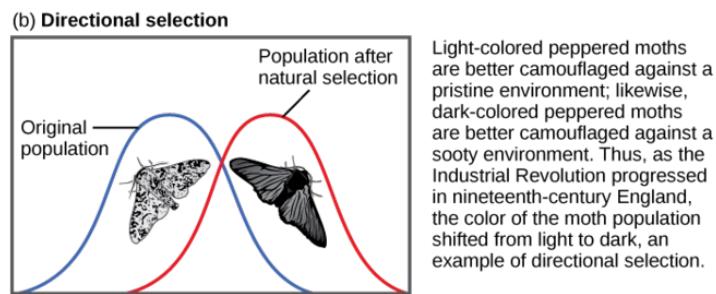
# Natural Selection

- Selection acts on phenotype, mostly not on a single locus:
  - ✓ Higher fecundity is fitter than lower fecundity
  - ✗ AA on rs12345678 is fitter than Aa
- Combining all loci together => phenotype of an individual
- Quantitative models of selection:
  - Fitness functions that define the relationship between phenotype and fitness:
    - Directional selection
    - Disruptive selection
    - Stabilizing selection

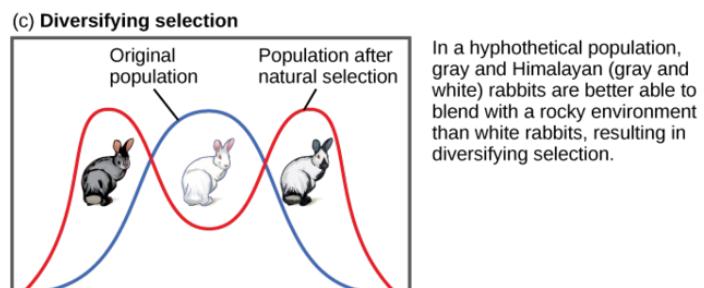
# Natural Selection



Robins typically lay four eggs, an example of stabilizing selection. Larger clutches may result in malnourished chicks, while smaller clutches may result in no viable offspring.



Light-colored peppered moths are better camouflaged against a pristine environment; likewise, dark-colored peppered moths are better camouflaged against a sooty environment. Thus, as the Industrial Revolution progressed in nineteenth-century England, the color of the moth population shifted from light to dark, an example of directional selection.



In a hypothetical population, gray and Himalayan (gray and white) rabbits are better able to blend with a rocky environment than white rabbits, resulting in diversifying selection.

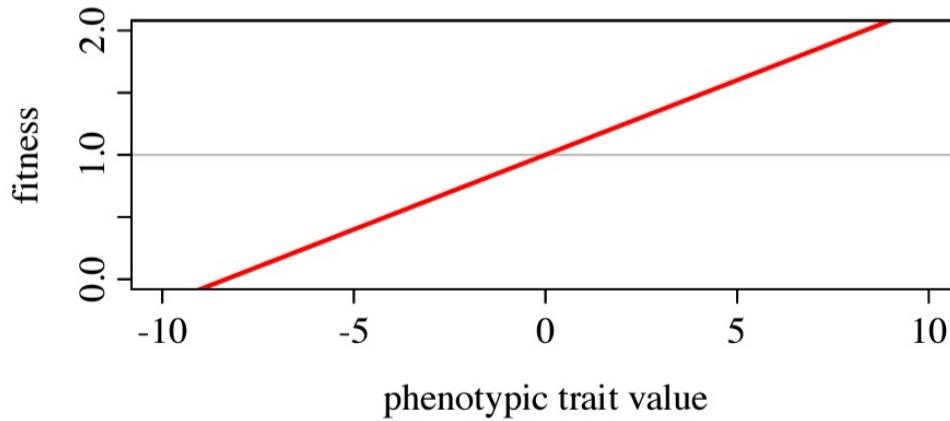
# Third Task

Simulate Directional Selection

# Directional Selection

## **Directional\_selection.slim**

- Multi-loci
- Multi population
- Each locus doesn't have any selection coefficient, they are all neutral in that sense.
- Instead, we:
  1. Calculate the phenotype of an individual
  2. Acquire the fitness distribution in the population
  3. Selection act on phenotype through the fitness function, ultimately changing the genotypes.
- Linear relationship between phenotype and fitness



# Directional Selection

## Exercises:

1. Plot the Fitness ~ Time trajectories in SLiM.
2. Try to understand how the level and speed of fitness increase are affected by the following parameters:
  - Effect size of QTLs (5<sup>th</sup> parameter passed to `initializeMutationType()`)
  - Population size (N)
  - Recombination rate (r)
3. Next, check subpopulation fitness distribution, do you notice a different behavior among our subpopulations? Why is it? Does the variance of fitness distribution in a population change?

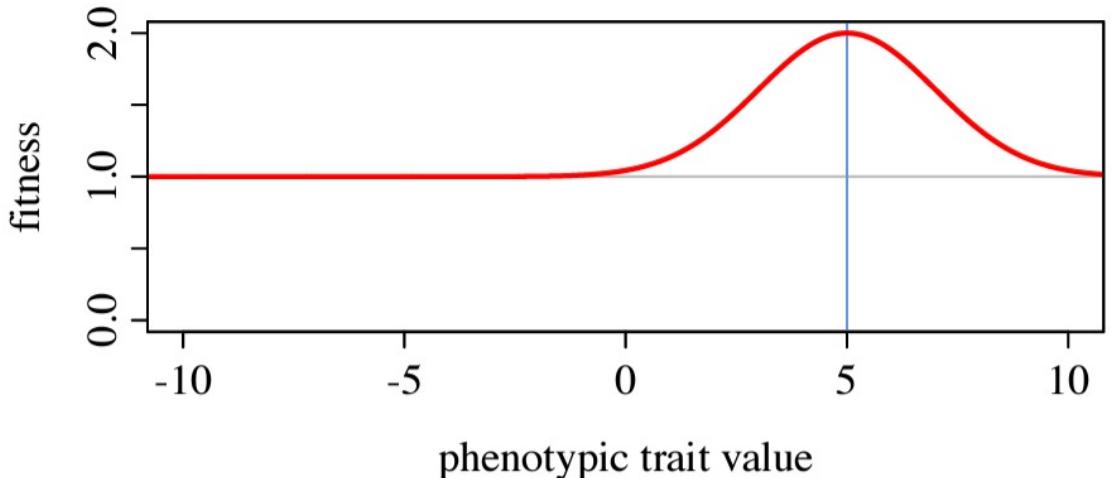
# Fourth Task

Simulate Stabilizing Selection

# Stabilizing Selection

## **Stabilizing\_selection.slim**

- Multi-loci
- Multi population
- Each locus doesn't have any selection coefficient, they are all neutral in that sense.
- Instead, we:
  1. Calculate the phenotype of an individual
  2. Acquire the fitness distribution in the population
  3. Selection act on phenotype through the fitness function, ultimately changing the genotypes.
- Peak of fitness when phenotype = 5 (a hypothetical situation)



# Stabilizing Selection

## Exercises:

1. Plot the Fitness ~ Time trajectories in SLiM
  - What is different from the directional selection?
2. Plot subpopulation fitness distribution, did they stop at some point? How many m2 type mutation is in the population reached in the trait optimum?
3. Play around with the effect size of m2, population size and scaling

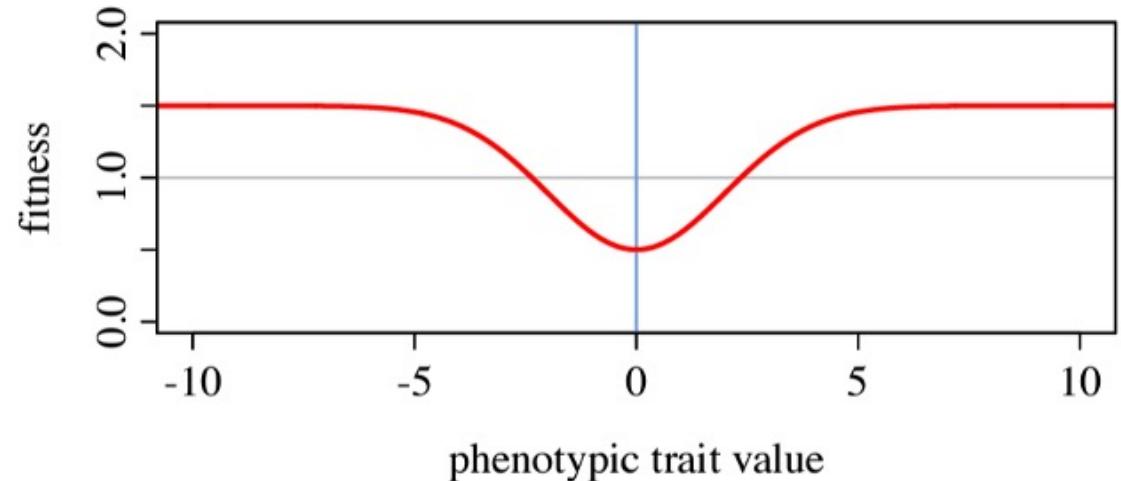
# Fifth Task

Simulate Diversifying Selection

# Diversifying Selection

## **Diversifying\_selection.slim**

- Multi-loci
- Multi population
- Each locus doesn't have any selection coefficient, they are all neutral in that sense.
- Instead, we:
  1. Calculate the phenotype of an individual
  2. Acquire the fitness distribution in the population
  3. Selection act on phenotype through the fitness function, ultimately changing the genotypes.
- Valley of fitness when phenotype = 0 (a hypothetical situation)



# Diversifying Selection

## Exercises:

1. Plot the Fitness ~ Time trajectories in SLiM
  - Looks like stabilizing selection, why?
2. Check the auto-generated Rplot, how does the phenotypic change look?
  - Copy the plotting function to the stabilizing selection script to have a similar view