

Simulating sweeps and adaptation in evolution experiments using SLiM, part II

Changyi Xiao & Neda Barghi

Overview

Day1

Detecting
selection target

Reconstruct
haplotype blocks

Drift

statistical
methods
(CMH-test
Chisq-test)

Linkage
(correlated
frequency
change)

characterize
selection (e.g.
selection
coefficient)

Day 2

SLiM
simulation

Model Fitting

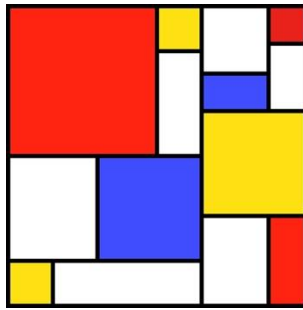
Model
selection

connection
between
simulation
and
experiments

summary
statistics (e.g.
Jaccard index)

Simulations of E&R Under Neutrality and Selection - SLiM

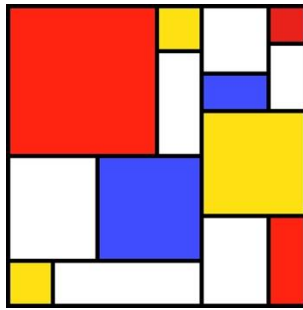
Neda Barghi & Changyi Xiao



SLiM

forward genetic simulation package
from the Messer Lab at Cornell University

SLiM - Materials

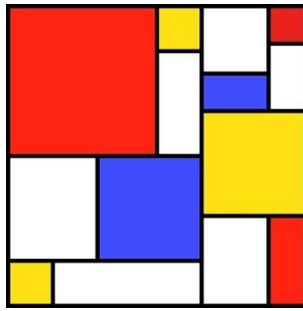


- www.messerlab.org/slim/
- SLiM Manual
- Eidos Manual
- Reference Sheets
- Workshop (www.benhaller.com/workshops/workshops.html)
- Mailing Lists (slim-announce & slim-discuss)

What is SLiM?

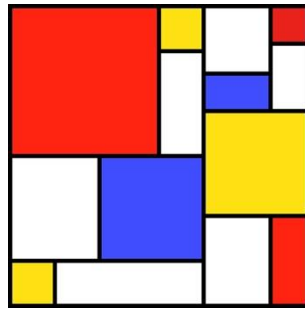
- Forward genetic simulations
 - Forward simulation
 - Works on the genomic level
 - Individual-based modeling
- Fitting empirical population genomic data
- Analysing theoretical evolutionary models
- Developing statistical methods

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
- Universally used

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^{*,1} and Philipp W. Messer^{*,1}

¹Department 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. Haller¹ | Jared Galloway² | Jerome Kelleher³ |
Philipp W. Messer^{1,*} | Peter L. Ralph^{2,*}

Evolutionary Modeling in SLiM 3 for Beginners

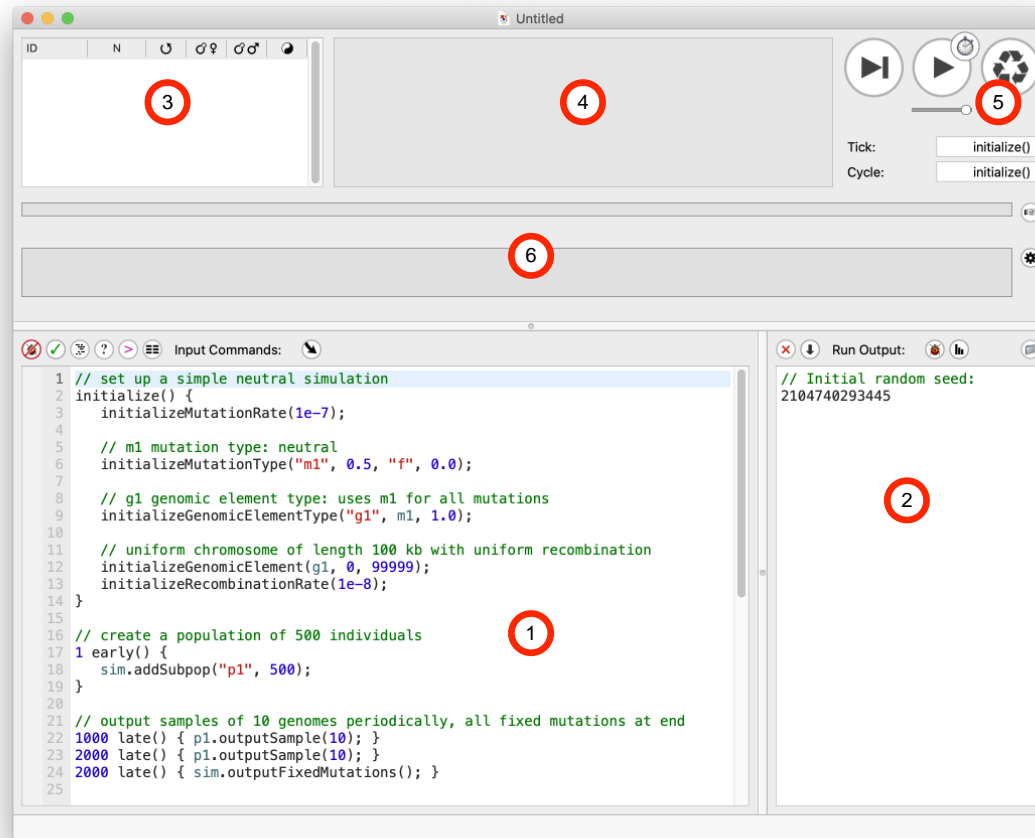
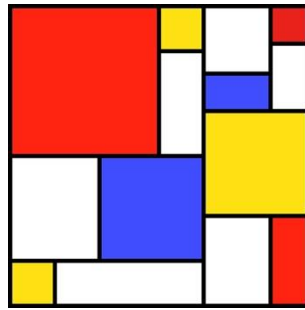
Benjamin C. Haller^{*,1} and Philipp W. Messer¹

¹Department 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 panel
- 2. The output panel.
- 3. The population view.
- 4. The individual view.
- 5. The tick controls.
- 6. The chromosome view.

Figure from the SLiM-Manual.

Five selection models

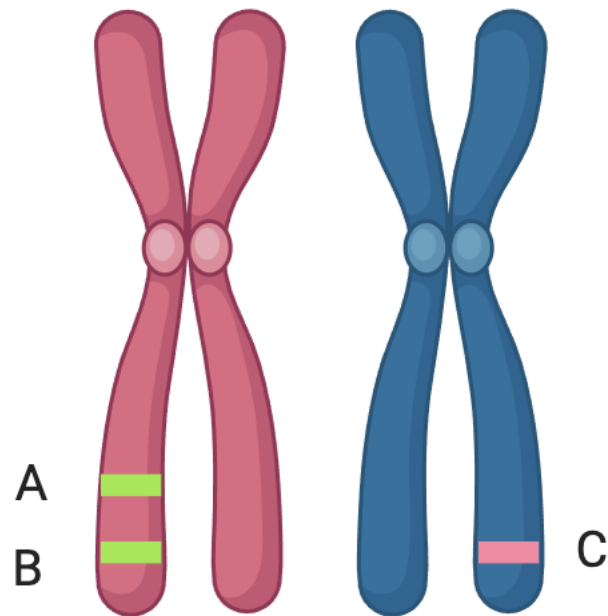
- Selective sweep + LE
- Selective sweep + LD
- Redundancy
- QTL + LE
- QTL + LD

The haplotype data (input)

- 189 Iso-female lines
- Crossing with the reference strain(M252)
- Individual sequencing
- Vcf format

```
##fileformat=VCFv4.2
##fileDate=20230320
##source=SLiM
##INFO=<ID=MID,Number=.,Type=Integer,Description="Mutation ID in SLiM">
##INFO=<ID=S,Number=.,Type=Float,Description="Selection Coefficient">
##INFO=<ID=DOM,Number=.,Type=Float,Description="Dominance">
##INFO=<ID=PO,Number=.,Type=Integer,Description="Population of Origin">
##INFO=<ID=TO,Number=.,Type=Integer,Description="Tick of Origin">
##INFO=<ID=MT,Number=.,Type=Integer,Description="Mutation Type">
##INFO=<ID=AC,Number=.,Type=Integer,Description="Allele Count">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=MULTIALLELIC,Number=0,Type=Flag,Description="Multiallelic">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT i0 i1 i2 i3 i4 i5 i6 i7 i8 i9 i10
1 549212 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 0|0 0|0 0|0
1 700853 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 707846 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 739322 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 746042 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 746164 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 748488 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 749128 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 751110 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 751163 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 751391 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 752526 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 755156 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 755202 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 755252 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 755520 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 755690 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
1 755721 . A C 1000 PASS MID=1;S=0.0000150881509433584;DOM=0.5;PO=0;TO=1;MT=1;DP=1000 GT 0|0 1|1 1|1 1|1
```

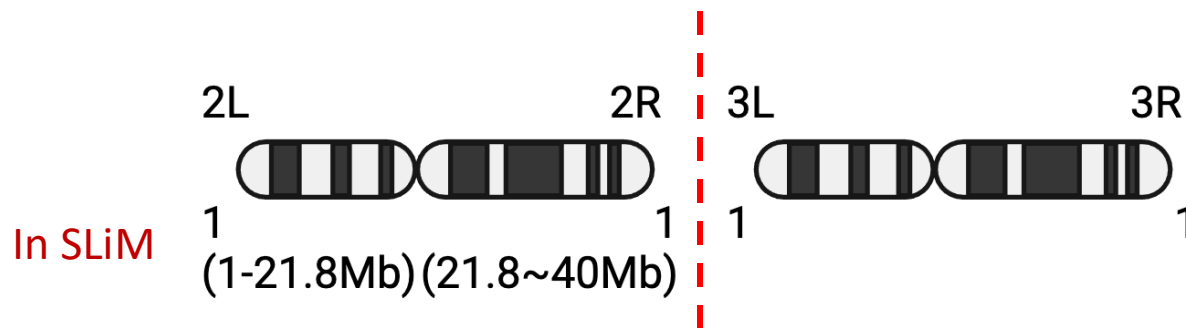
Linkage in the simulation



- A,B inherited together (unless recombination happen)
- Inheritance of C is independent of A and B
- Linkage equilibrium
 - Each SNPs is independent of each other
- Linkage disequilibrium:
 - SNPs located on the limited number of chromosome
 - recombination map of D. simulans

Simulation of multiple chromosomes in SLiM

- SLiM doesn't take the chromosome ID into consideration
 - Adjust the position based on the chromosome
 - Introduce 1bp with free recombination(0.5) between chromosome



	cM/Mb
39780001	2.63375167013952
39880001	2.34266934230936
39980001	2.01997126280032
40060001	1.7532025116829
40060002	5e+07
40160002	3.27309372567106
40260002	3.36055497885853
40360002	3.44449727356264
40460002	3.52014345700354
40560002	3.59418181843053

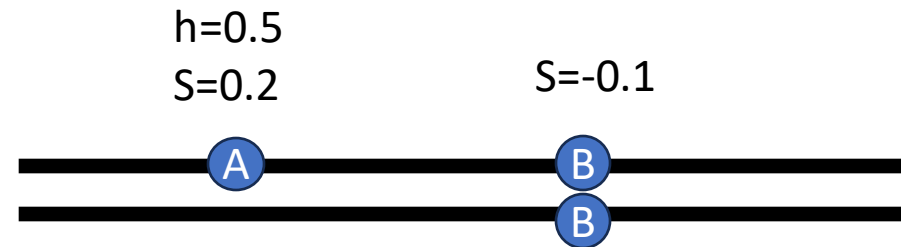
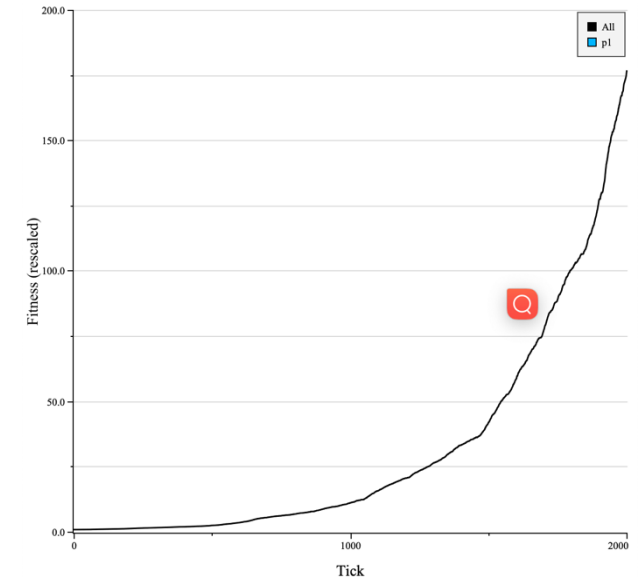
Selective sweep

- Each locus is assigned with selection coefficients (s). Allele with higher s result to higher fitness additive effect (h=0.5)
w = individual fitness

- $W_{AA}=1+s$
- $W_{Aa}=1+0.5*s$
- $W_{aa}=1$

$$\text{Fitness of individual} = \prod_{i=0}^n W_i$$

*n represent number of loci of that individual



$$W_{Aa}=1+0.5*0.2=1.1$$

$$W_{BB}=1+(-0.1)=0.9$$

$$W=W_{Aa}*W_{BB}=0.99$$

Model I: selective sweep + LE

- Census population of 1250
- 60 generations
- Selective sweep
- Free recombination($r=0.5$)
 - Each segregating is inherited independently

```
slim -d "rep=1" 01sweep_le.slim
```

Model II: selective sweep + LD

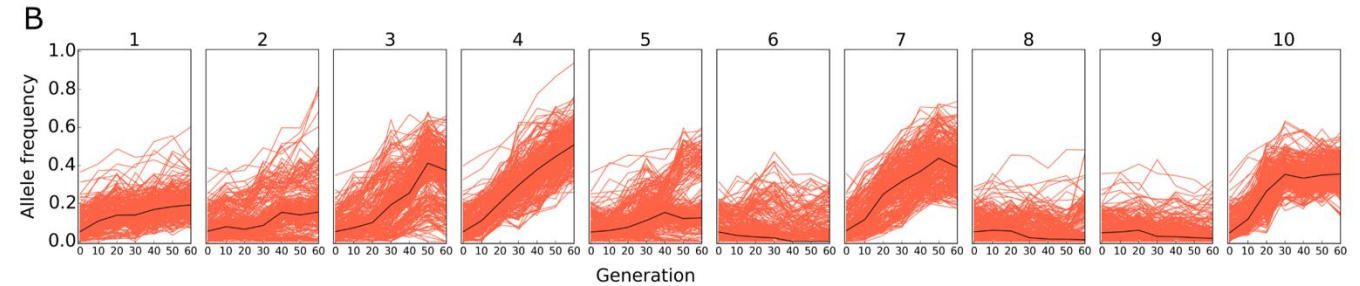
- Census population of 1250
- 60 generations
- Selective sweep
- Recombination map of *D. simulans* (materials/rec_map_slim.txt)

```
slim -d "rep=1" 02sweep_ld.slim
```


Model III: redundancy

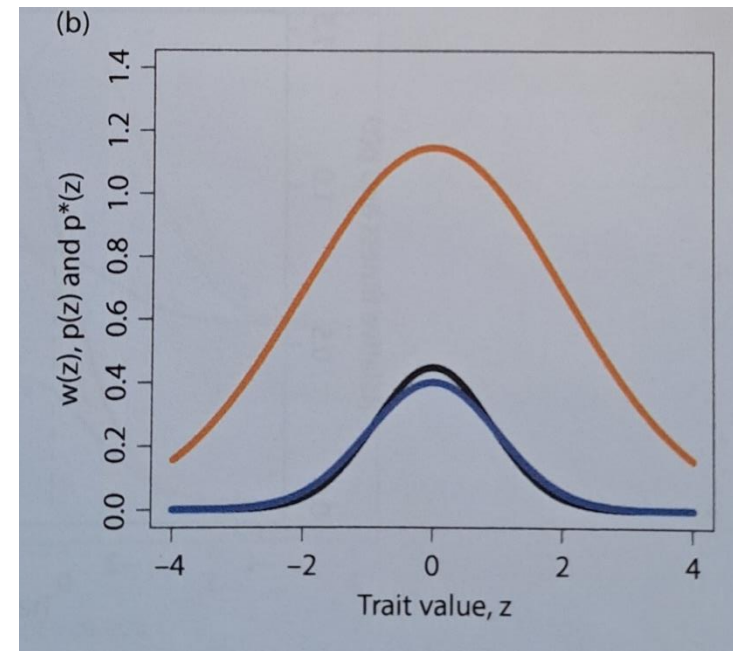
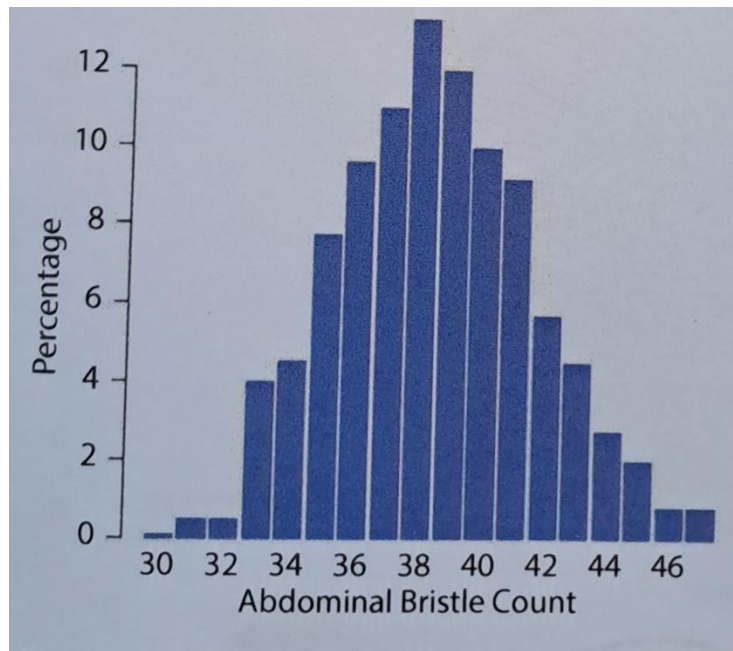
(Barghi et al, 2019)

- Census population of 1250
- 60 generations
- Selective sweep
- Recombination map of *D. simulans* (materials/rec_map_slim.txt)
- Random subset of contributing loci respond to selection

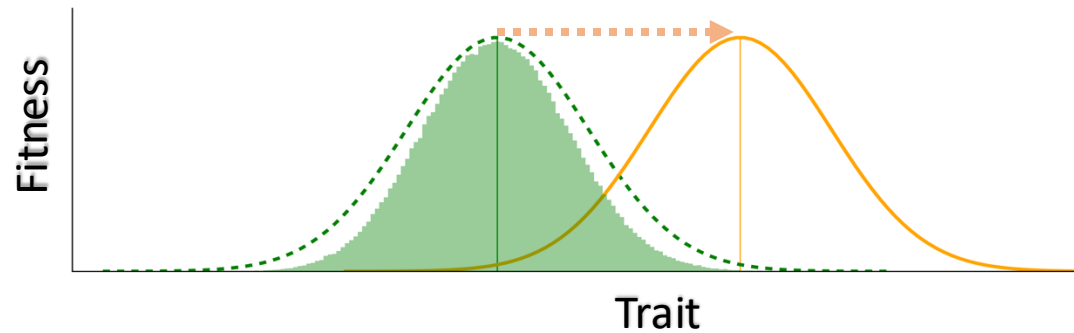


```
slim -d "rep=1" -d "num_mut=40" 03redundancy.slim
```

Quantitative traits under stabilizing selection



Adaptation of a quantitative trait after a shift in trait optimum



Current optimum

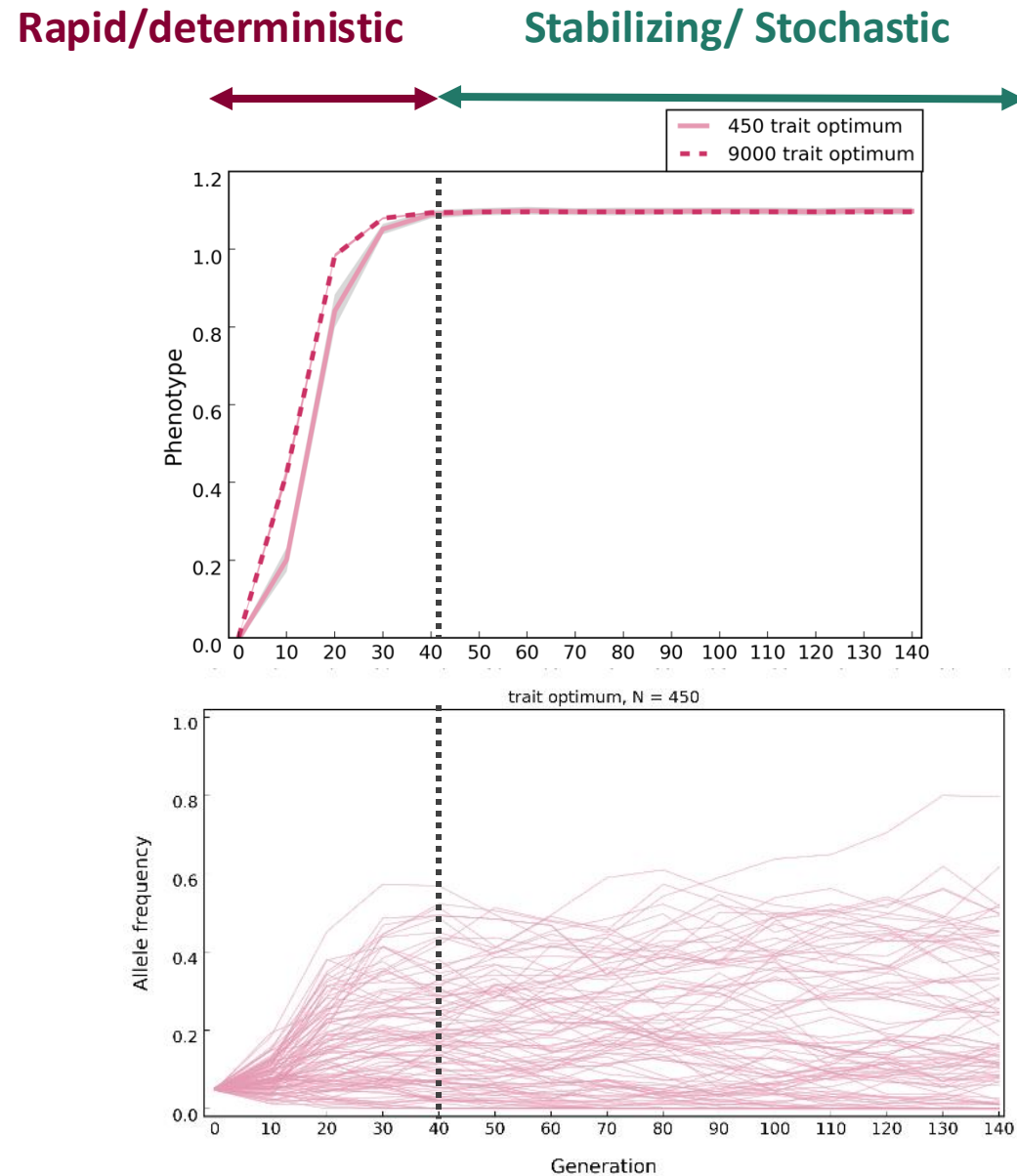
New optimum

$$w = e^{\frac{-(z-z_0)^2}{2Vs}}$$

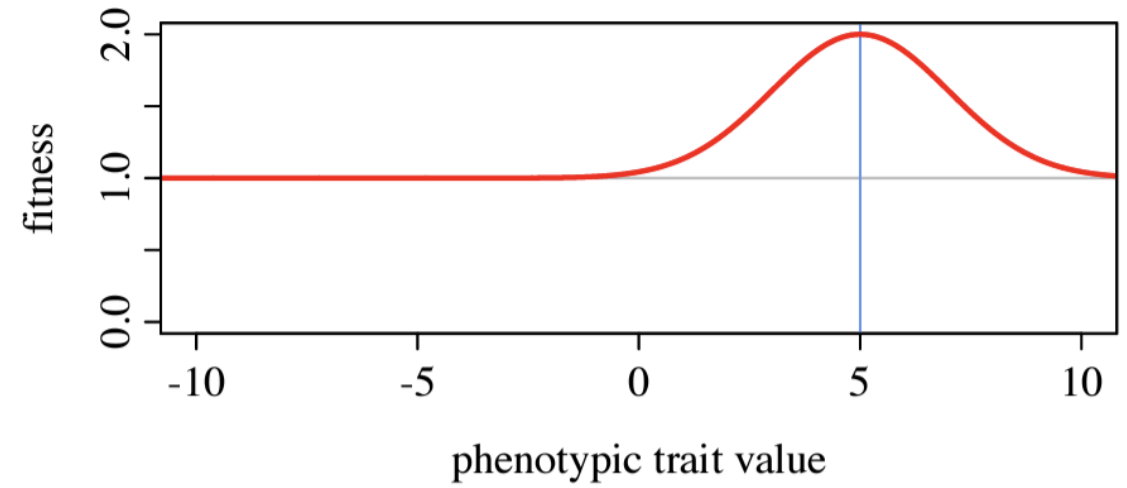
Fitness

Variance of fitness function

Patterns of phenotypic/genomic change during polygenic adaptation



Stabilizing



- Multi-loci
- single population
- Instead, we:
 - Calculate the phenotype of an individual
 - Acquire the fitness distribution in the population
 - Selection act on phenotype through the fitness function, ultimately changing the genotypes.
- Peak of fitness when phenotype = 5 (a hypothetical situation)

Model IV: QTL + LE

- Census population of 1250
- 60 generations
- Stabilizing selection
- Free recombination

```
slim -d "rep=1" 04qtl_le.slim
```

Model V: QTL + LD

- Census population of 1250
- 60 generations
- Stabilizing selection
- Recombination map of *D. simulans* (materials/rec_map_slim.txt)

```
slim -d "rep=1" 05qtl_ld.slim
```

External resources

- SLiM Homepage:
 - <https://messerlab.org/slim/>
- SLiM Manual:
 - https://github.com/MesserLab/SLiM/releases/download/v4.0.1/SLiM_Manual.pdf
- SLiM workshop material:
 - <http://benhaller.com/workshops/workshops.html>
- Recipe inside SLiM GUI

Overview

Day1

Detecting
selection target

Reconstruct
haplotype blocks

Drift

statistical
methods
(CMH-test
Chisq-test)

Linkage
(correlated
frequency
change)

characterize
selection (e.g.
selection
coefficient)

Day 2

SLiM
simulation

Model Fitting

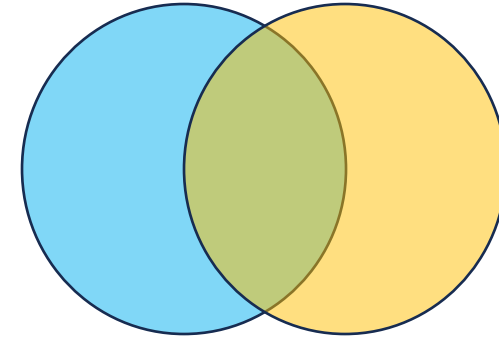
Model
selection

connection
between
simulation
and
experiments

summary
statistics (e.g.
Jaccard index)

Bridging the empirical and simulation

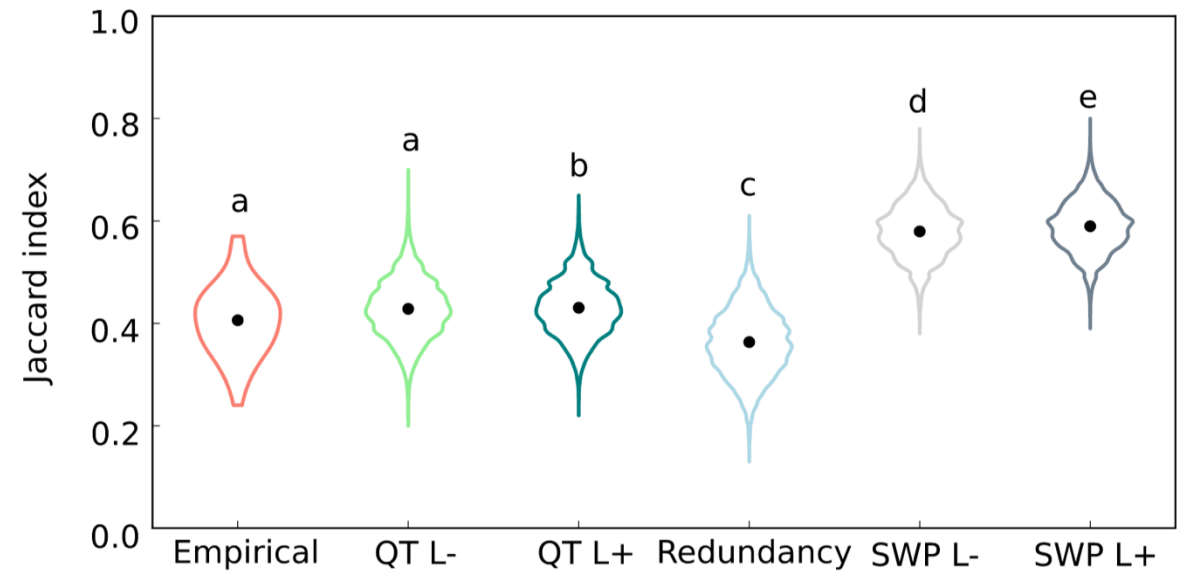
- Which models fits the empirical data best?
- Jaccard index between biological replicates
 - Measurement of parallelism



$$\text{Jaccard index} = \frac{A \cap B}{A \cup B}$$

- Detect SNPs responding to selection in each replicate
 - e.g. allele frequency larger than 0.1
- Estimate Jaccard index between replicates within each model
- Compare the Jaccard index between models

Jaccard index



Hope you had fun!

Any questions?