FANTASTIC FORCES

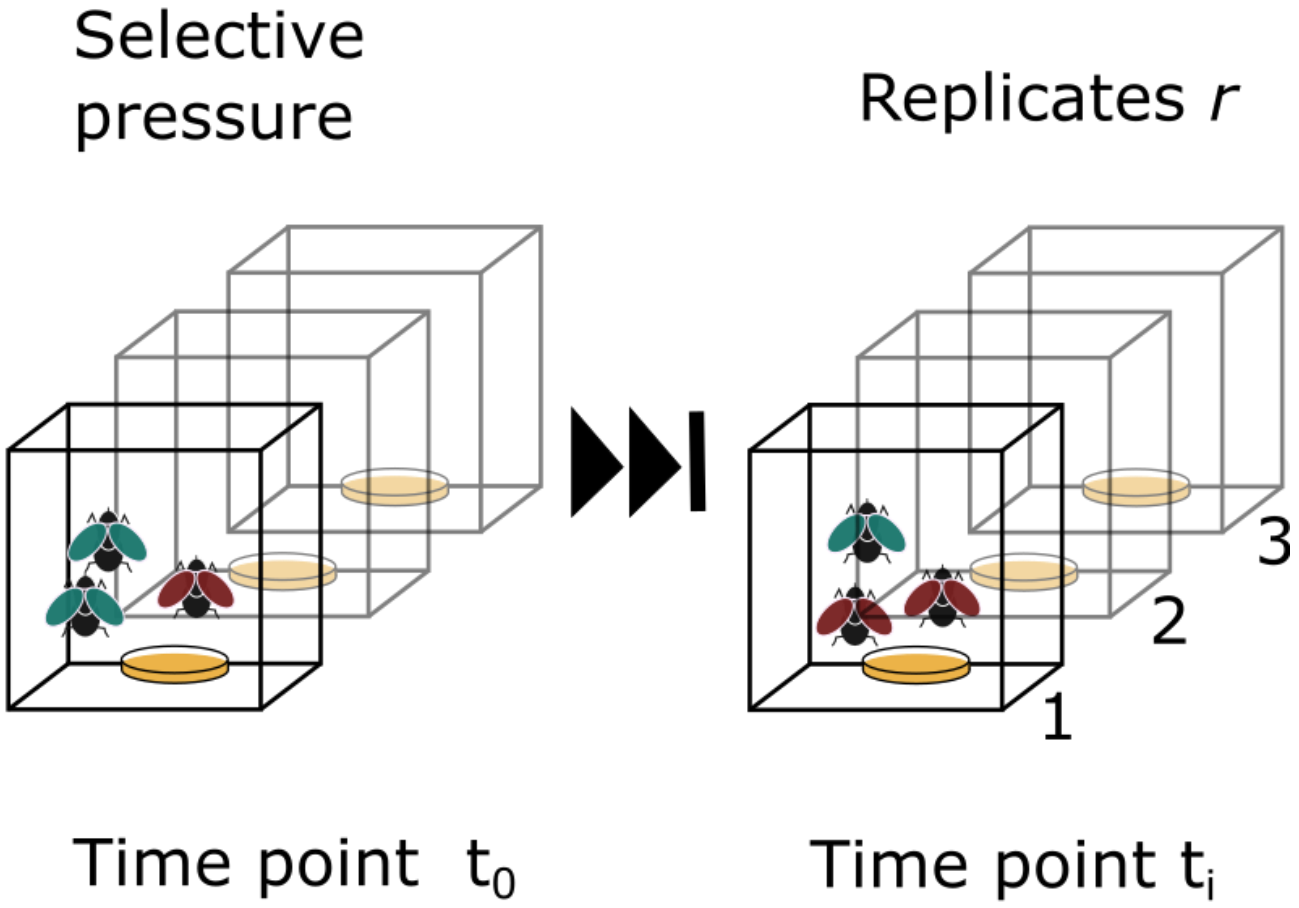
5th - 7th June, 2019

University of St Andrews, Scotland

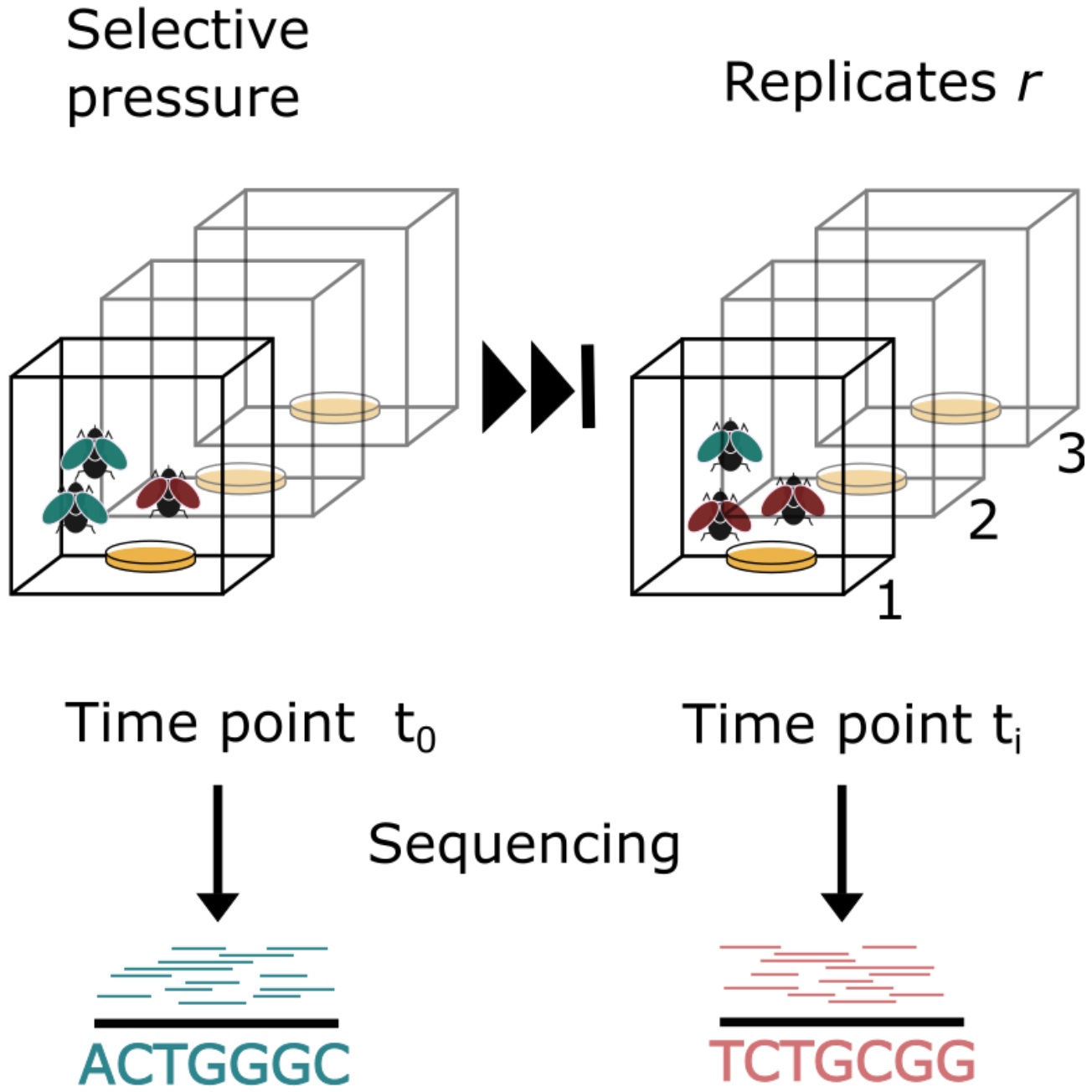
Tutorial 2: Detecting selection for dummies

Carolina Barata @BarataCarol

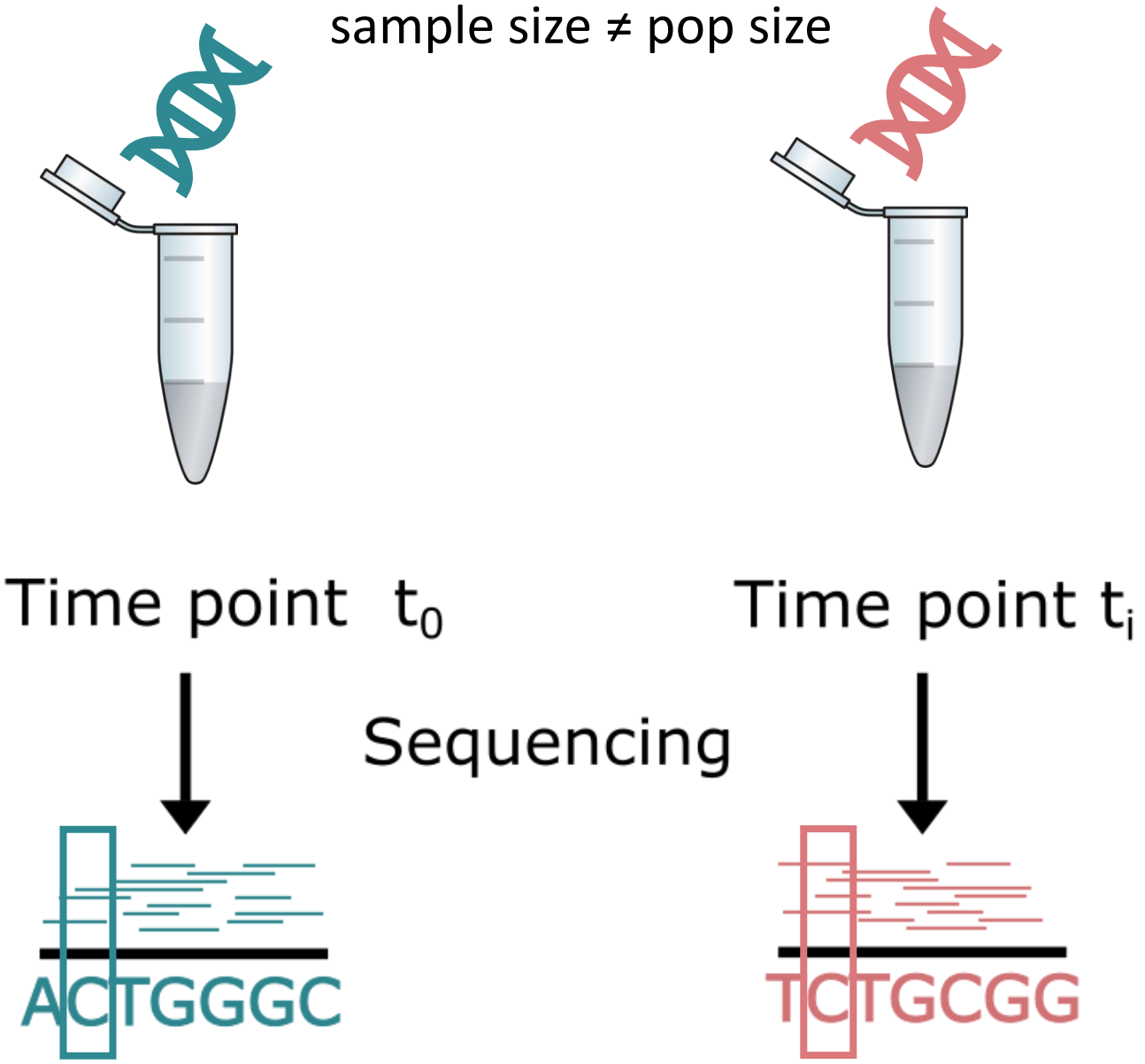
Evolve & Re-sequence



Evolve & Re-sequence



E&R + Pool-seq


sample size ≠ pop size
Time point t_0
Time point t_i
Sequencing
ACTGGGC
TCTGCGG

BAIT-ER

Bayesian inference

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

Allele frequency trajectories

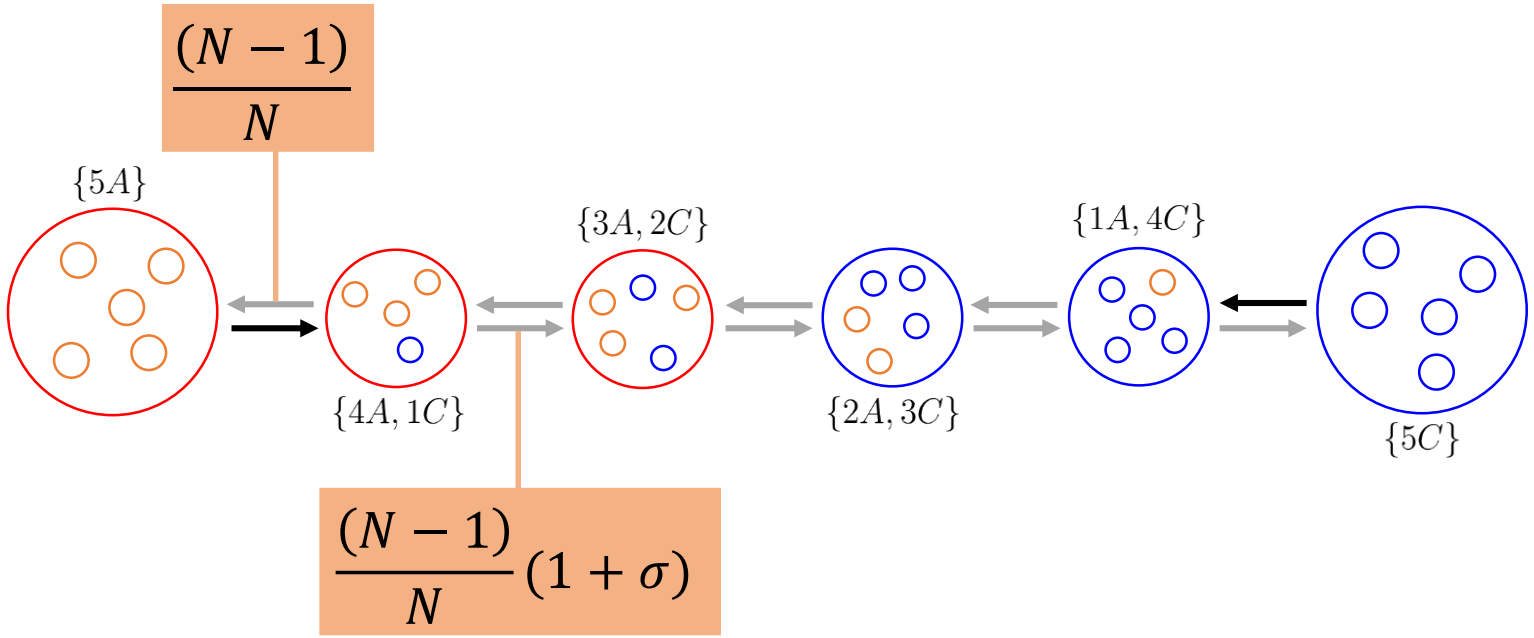
Evolve & Re-sequence experiments

Borges, Barata and Kosiol *(in prep)*

Why BAIT-ER?

- Mechanistic model
  1. Sequencing depth noise
  2. Replicate populations

- Moran model



Tutorial

Goal: estimate selection coefficients

1) Simulate allele frequency trajectories

2) Real time series data

Tutorial

Software

Download BAIT-ER:

https://github.com/mrborges23/Bait-ER

Then:

or

Tutorial

Software

```
install.packages(“Rcpp”)
install.packages(“RcppArmadillo”)
install.packages(“RcppParallel”)

library("Rcpp")
library("RcppArmadillo")
library("RcppParallel")
```

Tutorial

Software

```
sourceCpp("Bait-ER Code/read_sync_file.cpp")

sourceCpp("Bait-ER Code/allele_trajectory_simulator.cpp")

sourceCpp("Bait-ER Code/reads_to_moran_states.cpp")

sourceCpp("Bait-ER Code/sigma_posterior.cpp")

sourceCpp("Bait-ER Code/sites_sigma_posterior.cpp")
```

(1) Simulations

- Demographic:
  1. N_e
  2. Initial allele frequency
  3. True σ (0 or 10)

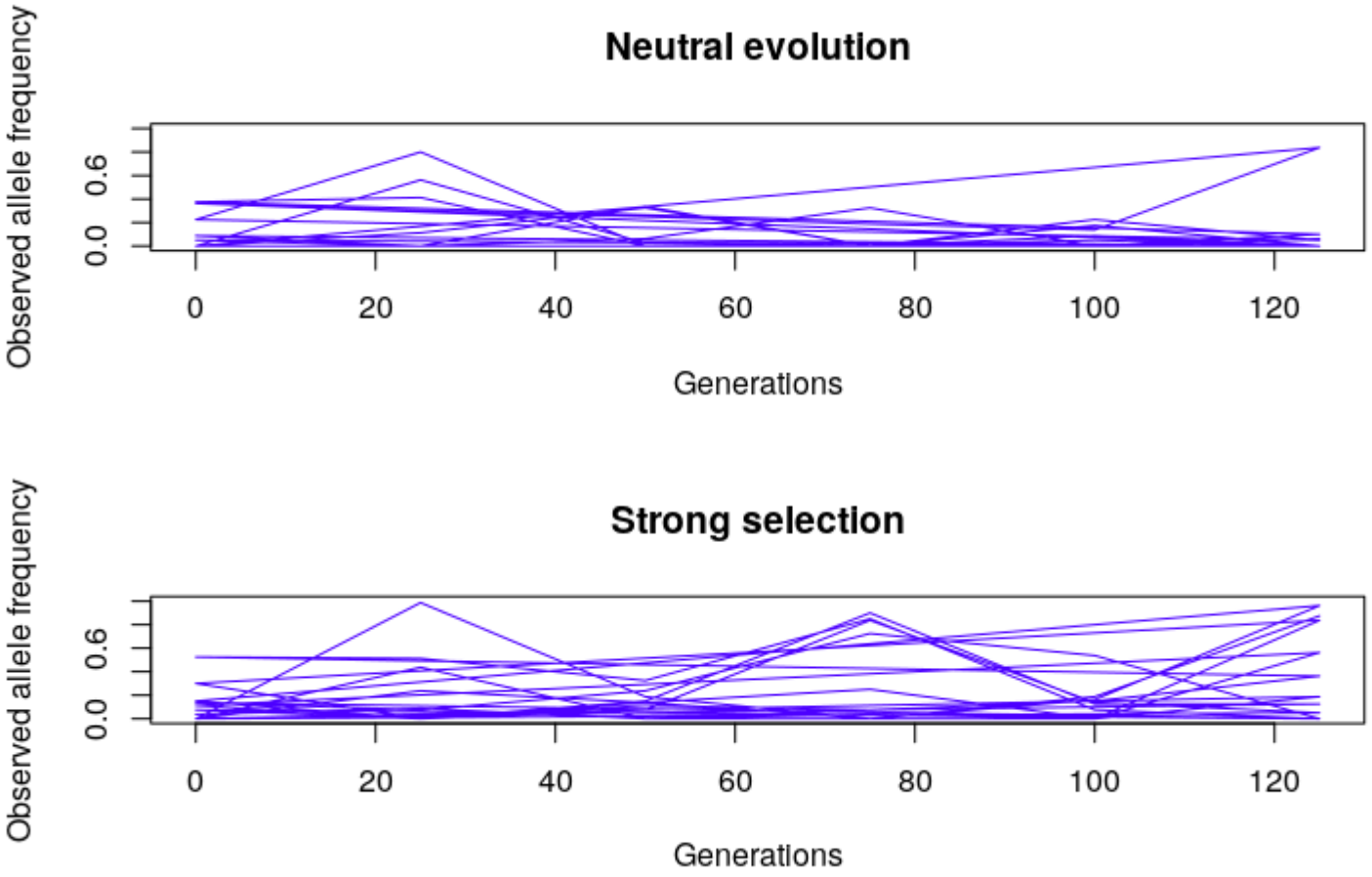
- Experimental
  1. # replicates
  2. Time scheme
  3. Coverage

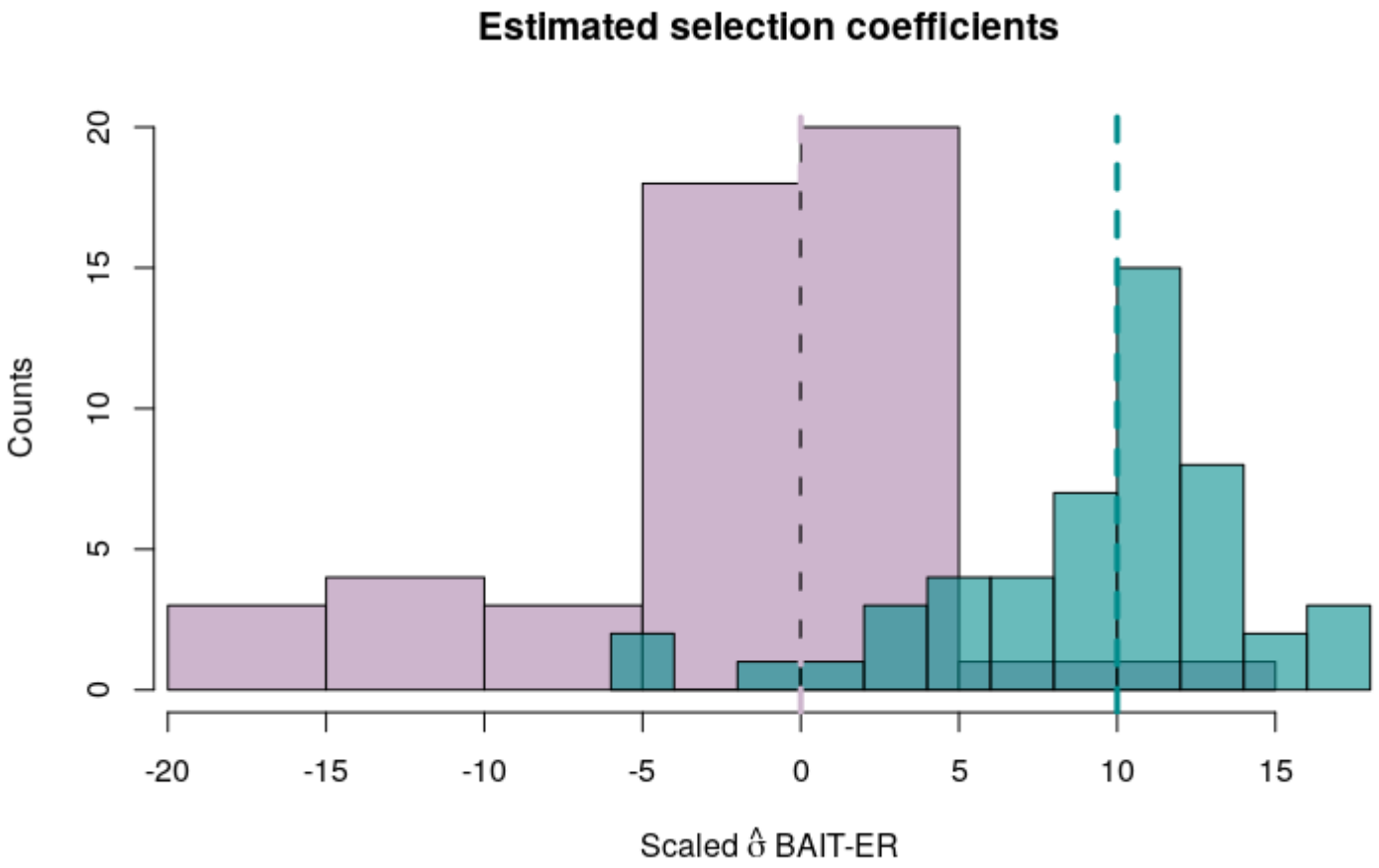
(1) Simulations

- Demographic:
  1. N_e = 500
  2. Initial allele frequency = 10%
  3. True σ (0 or 10)

- Experimental
  1. # replicates = 5
  2. Time scheme = 6 time points
  3. Coverage = 80x

(1) Simulations





(1) Simulations

(1) Simulations

- Demographic:
  1. N_e = 500
  2. Initial allele frequency = 10%
  3. True σ (0 or 10)

- Experimental
  1. # replicates = 5
  2. Time scheme = 6 time points
  3. Coverage = 80x

What if we change the starting allele frequency?

- Demographic:
  1. N_e = 100
  2. Initial allele frequency = 10%
  3. True σ (0 or 10)

- Experimental
  1. # replicates = 5
  2. Time scheme = 6 time points
  3. Coverage = 80x

(1) Simulations

What if we change the effective population size?

- Demographic:
  1. N_e = 500
  2. Initial allele frequency = 50%
  3. True σ (0 or 10)

- Experimental
  1. # replicates = 5
  2. Time scheme = 6 time points
  3. Coverage = 80x

(1) Simulations

(2) Real time series data

Barghi *et al* 2019

PLOS BIOLOGY

RESEARCH ARTICLE

Genetic redundancy fuels polygenic adaptation in *Drosophila*

Neda Barghi[1], Raymond Tobler[1,2¤a], Viola Nolte[1], Ana Marija Jakšić[1,2¤b], François Mallard[1¤c], Kathrin Anna Otte[1], Marlies Dolezal[1,3], Thomas Taus[1,2], Robert Kofler[1], Christian Schlötterer[1]*

(2) Barghi *et al* 2019

Drosophila simulans

Replicates = 10

Sampling scheme = {0, 10, 20, 30, 40, 50, 60}

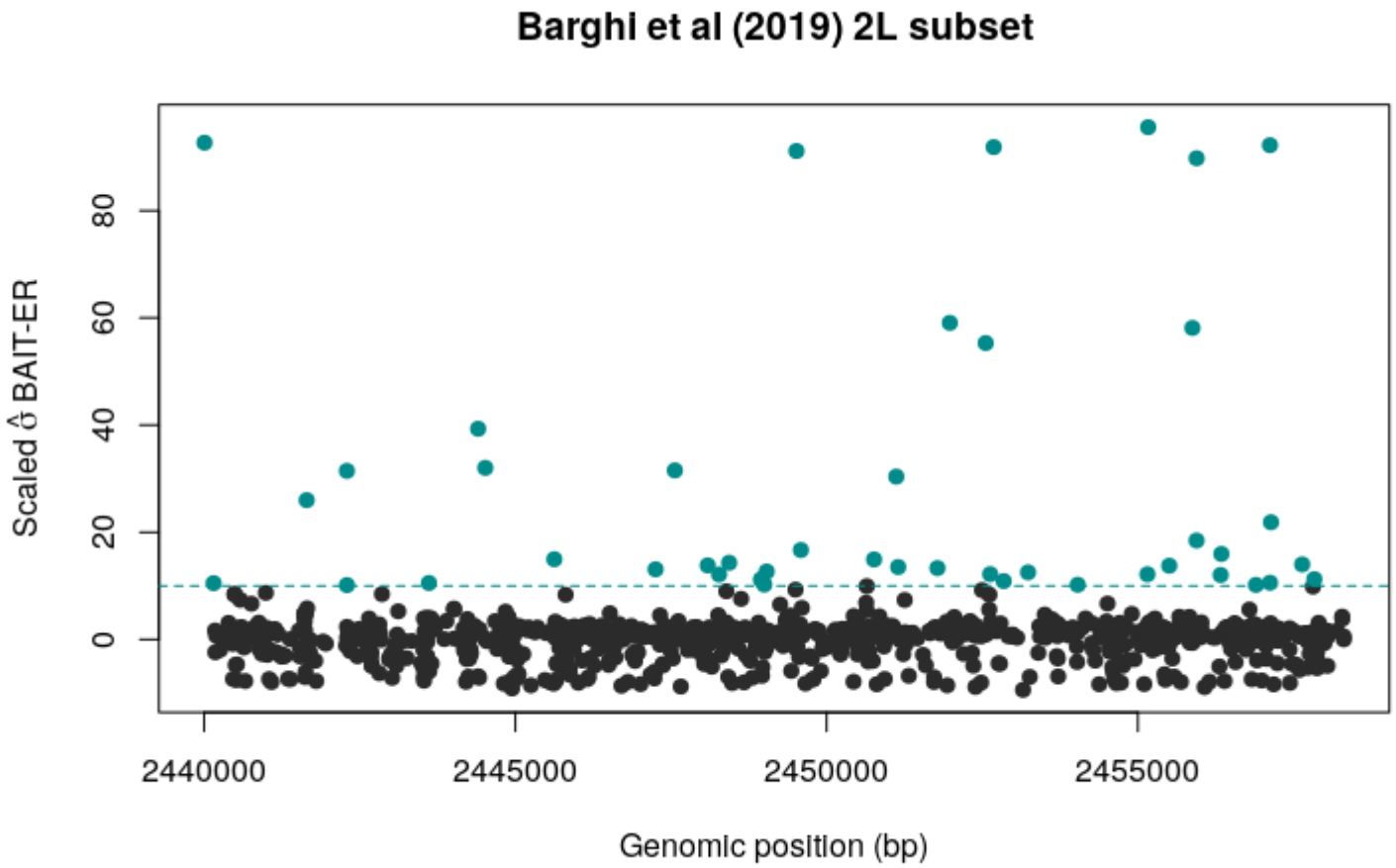
$N_e \approx 300$

.sync file structure

2L 2439984 T 0:194:5:0:0:0 (…)

(2) Barghi *et al* 2019

1) Read in the data using `read_sync_file()`

2) Estimate selection coefficients with `sites_sigma_posterior()`

(2) Barghi *et al* 2019