

Received: 17 October 2020

Accepted: 22 March 2021

DOI: 10.1111/2041-210X.13612

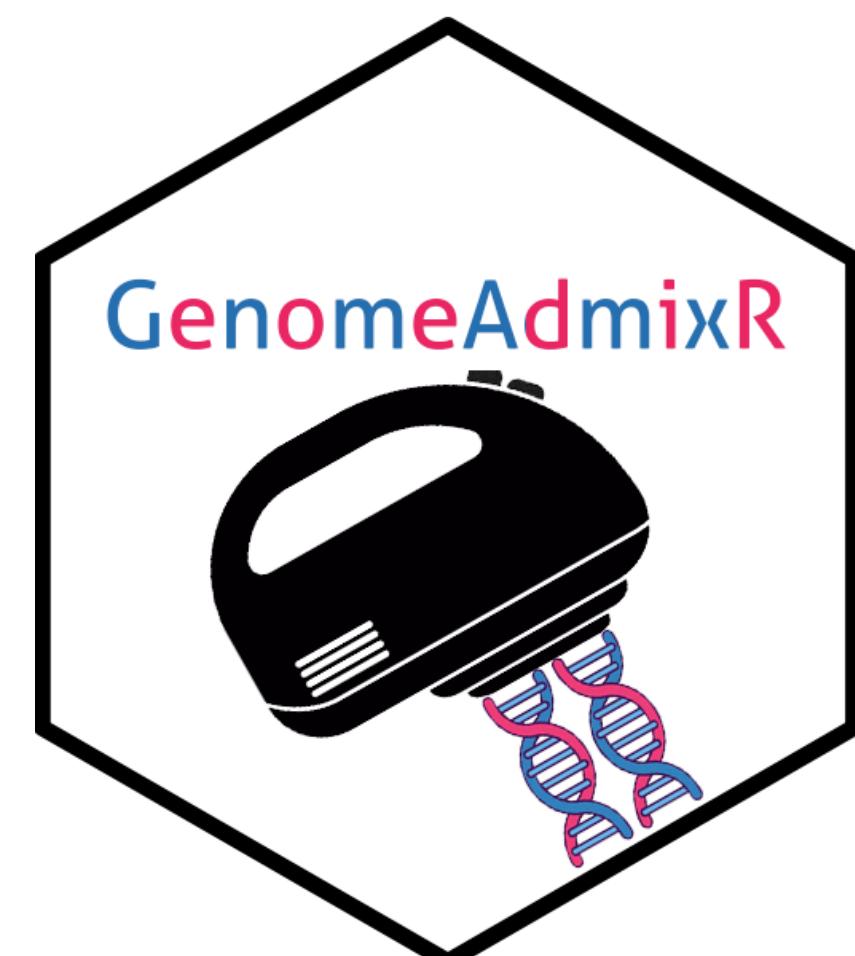
**APPLICATION**

Methods in Ecology and Evolution

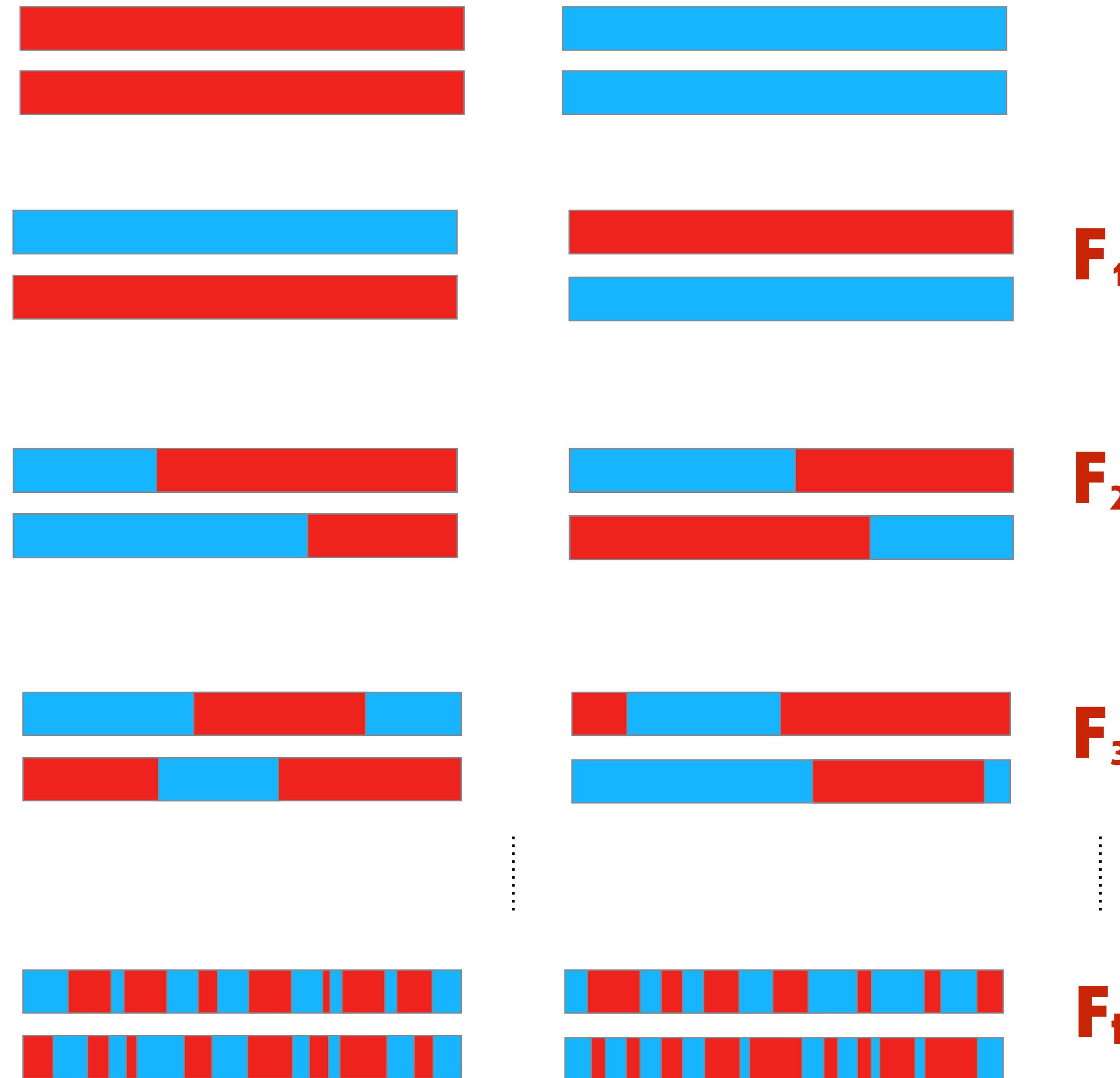


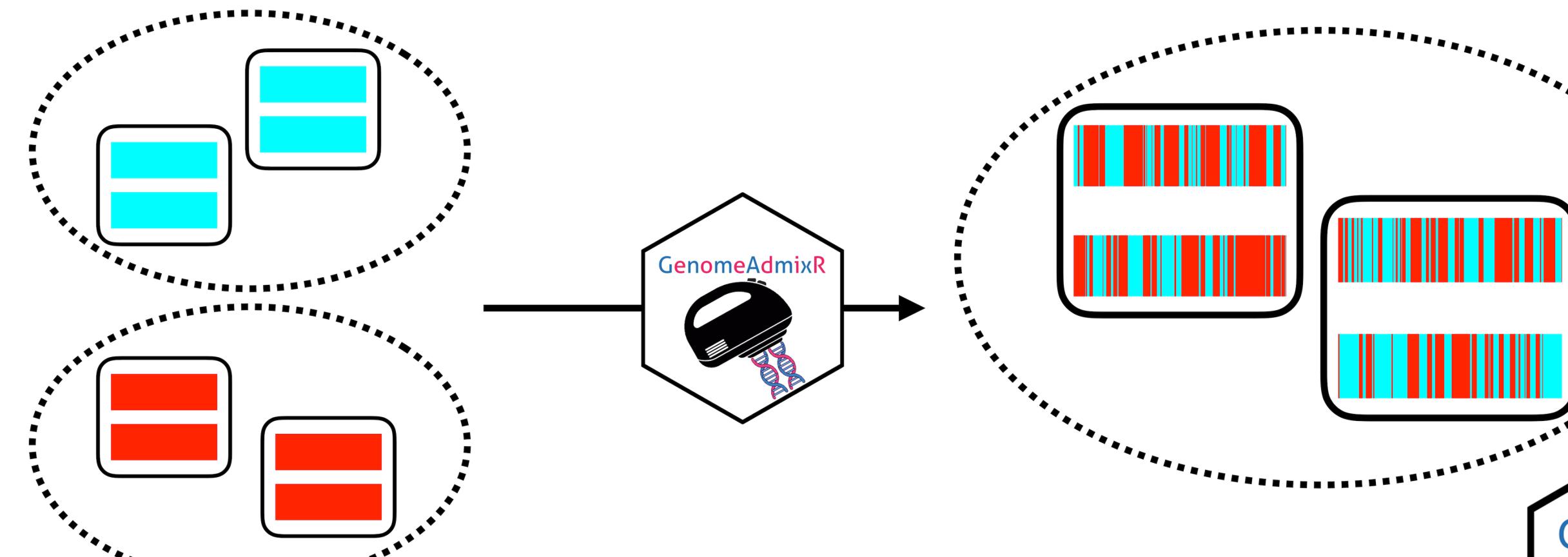
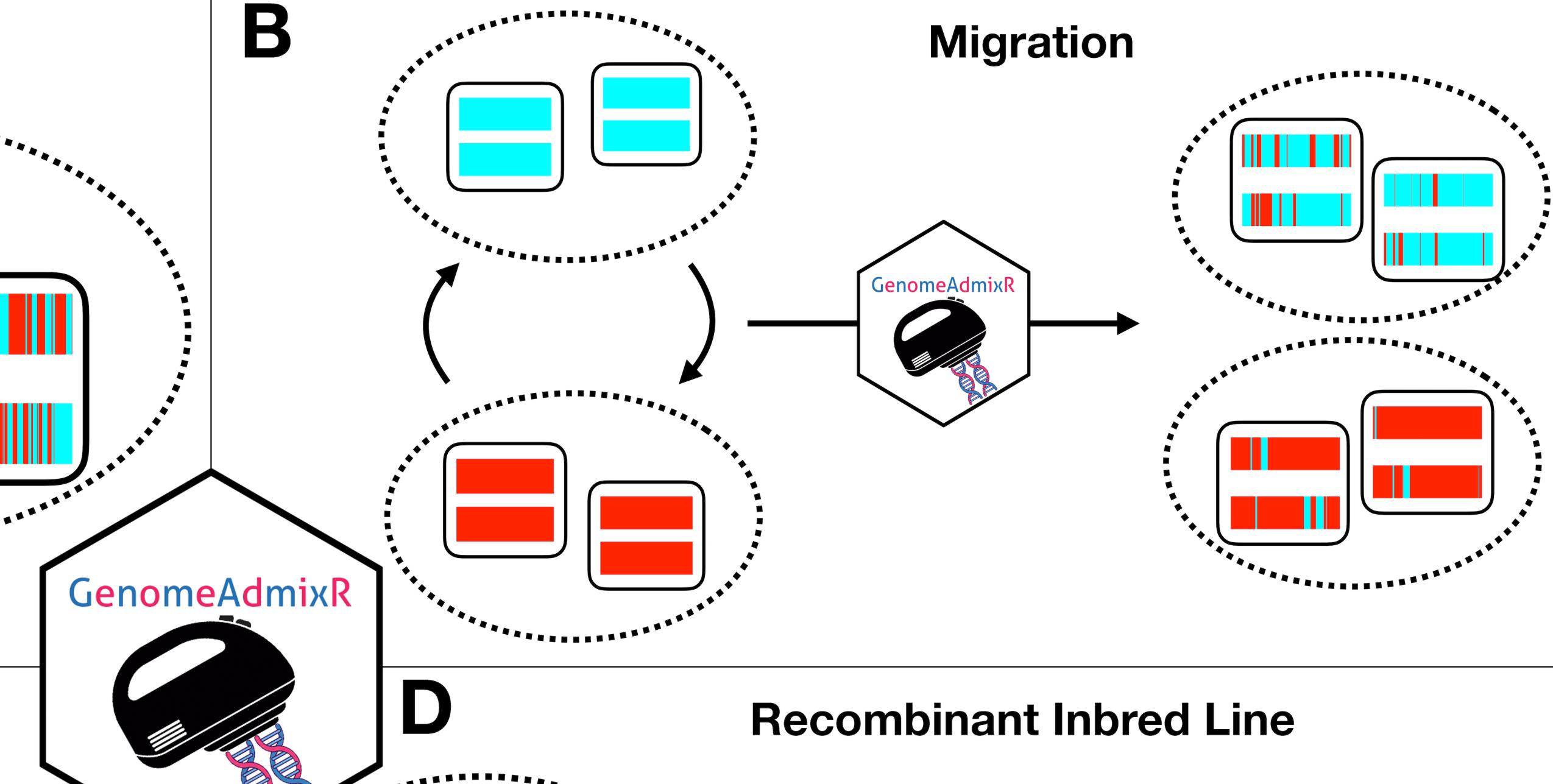
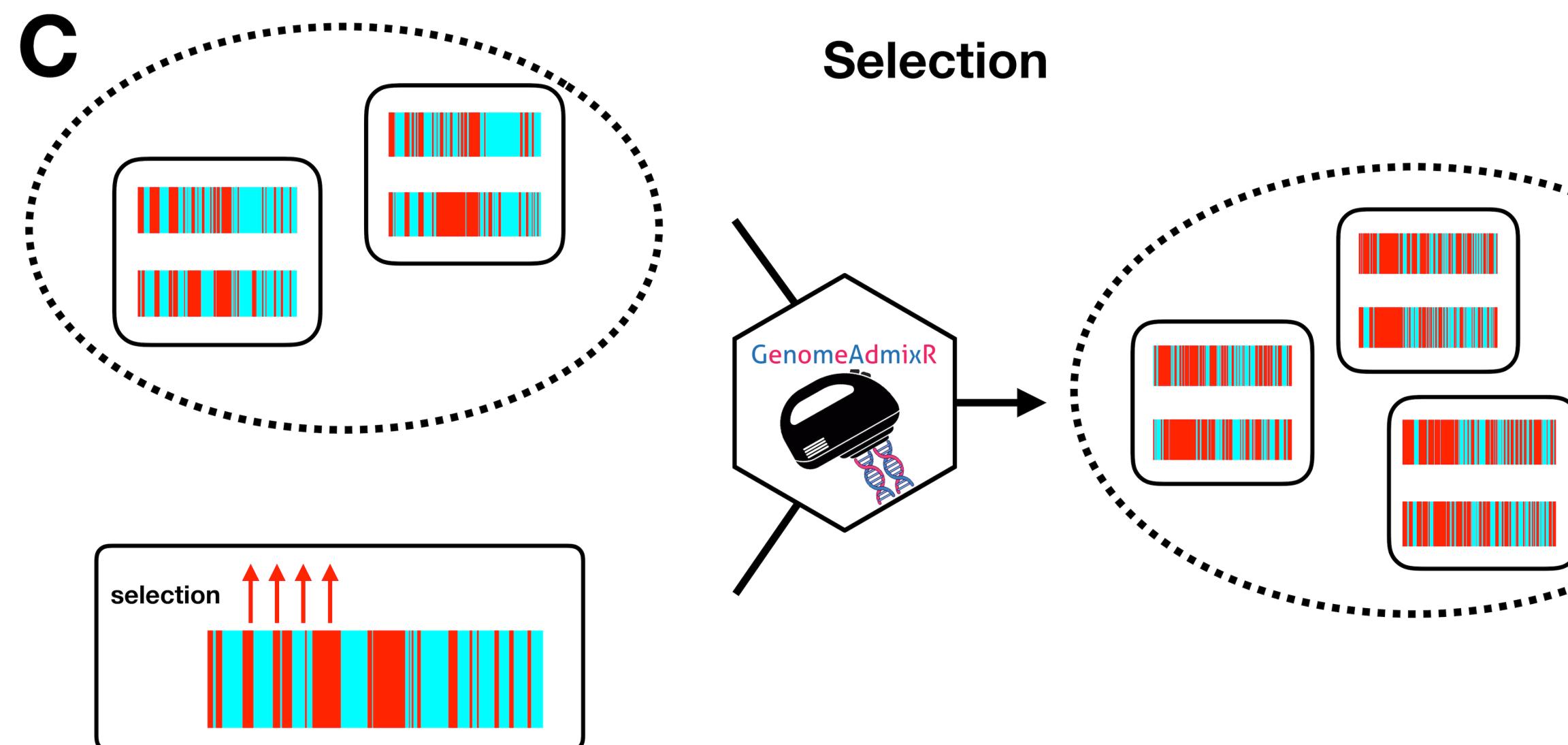
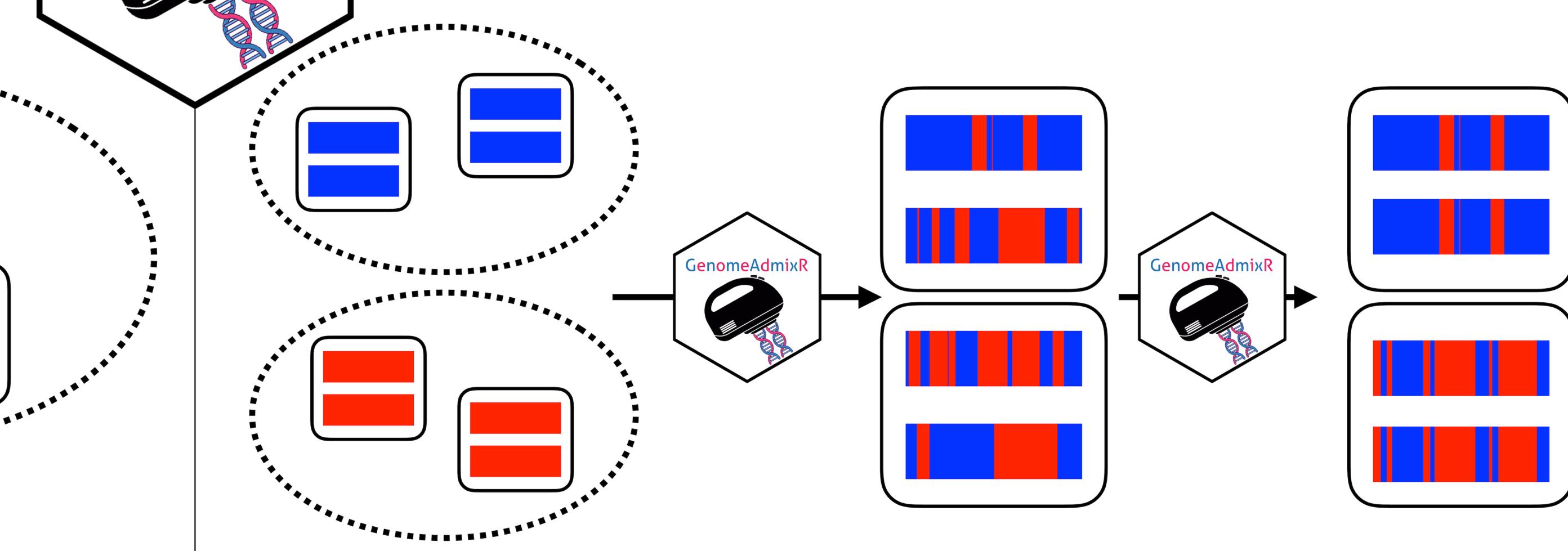
# Individual-based simulations of genome evolution with ancestry: The GENOMEADMIXR R package

Thijs Janzen<sup>1,2</sup>  | Fernando Diaz<sup>3</sup> 



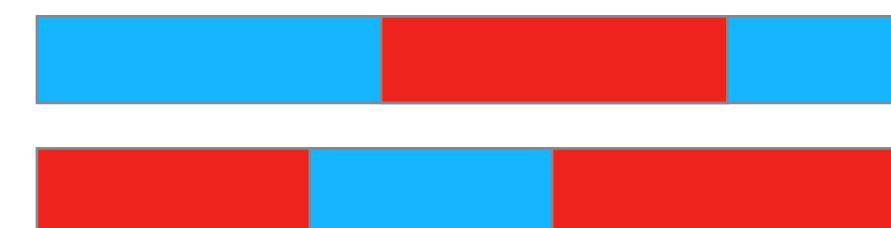
# Admixture



**A****Hybridization****B****Migration****C****Selection****D****Recombinant Inbred Line**

# How to use GenomeAdmixR

## Ancestry module



## Sequence module

atataataaattttttcccc  
aaatgccagtttttcgct

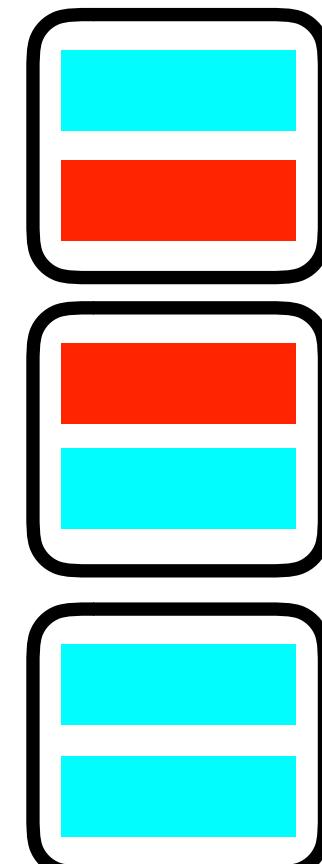
## Selection

[location, waa, WAa, WAA, A]

[0.5, 1, 1+h\*s, 1+s, 0]

**multiple genes can be under selection**

Population of fixed size N



Assign fitness

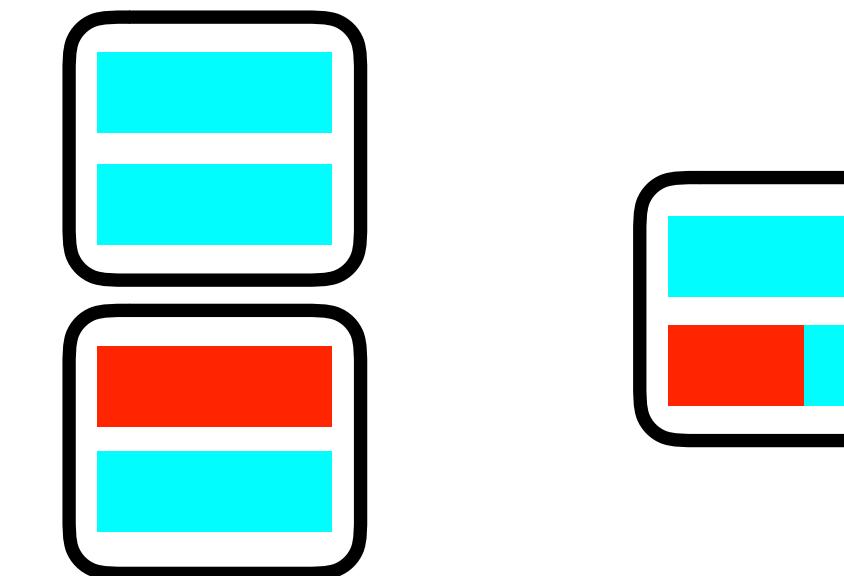
1

1

1

Repeat N times

- .....
- Draw parents prop. to fitness
- Generate offspring



Replace parents with offspring

# Comparison modules

Ancestry module

```
> markers <- seq(from = 0, to = 1, length.out = 100)

> pop_10 <- simulate_admixture(
  module = ancestry_module(
    number_of_founders = 2,
    markers = markers),
  total_runtime = 10,
  pop_size = 100)

> ld_10 <- calculate_ld(pop_10, markers)

> het_10 <-
  calculate_heterozygosity(pop_10$population,
  locations = markers)
```

Sequence module

```
> data("dgrp2.3R.5k.data")

> markers <- sort(sample(dgrp2.3R.5k.data$markers, 100))

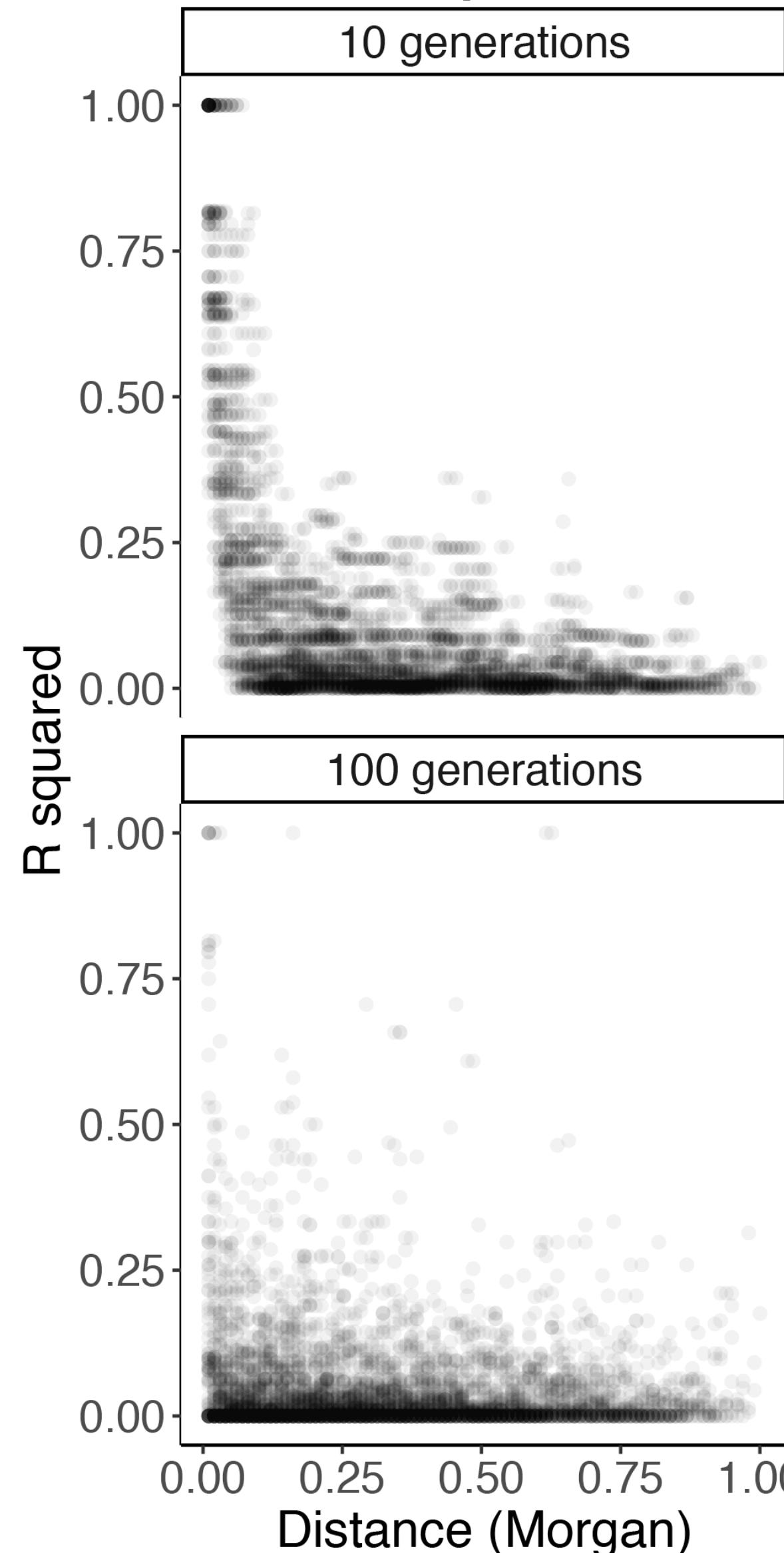
> pop_10 <- simulate_admixture(
  module = sequence_module(
    input_data = dgrp2.3R.5k.data,
    markers = markers),
  total_runtime = 10,
  pop_size = 100)

> ld_10 <- calculate_ld(pop_10, markers)

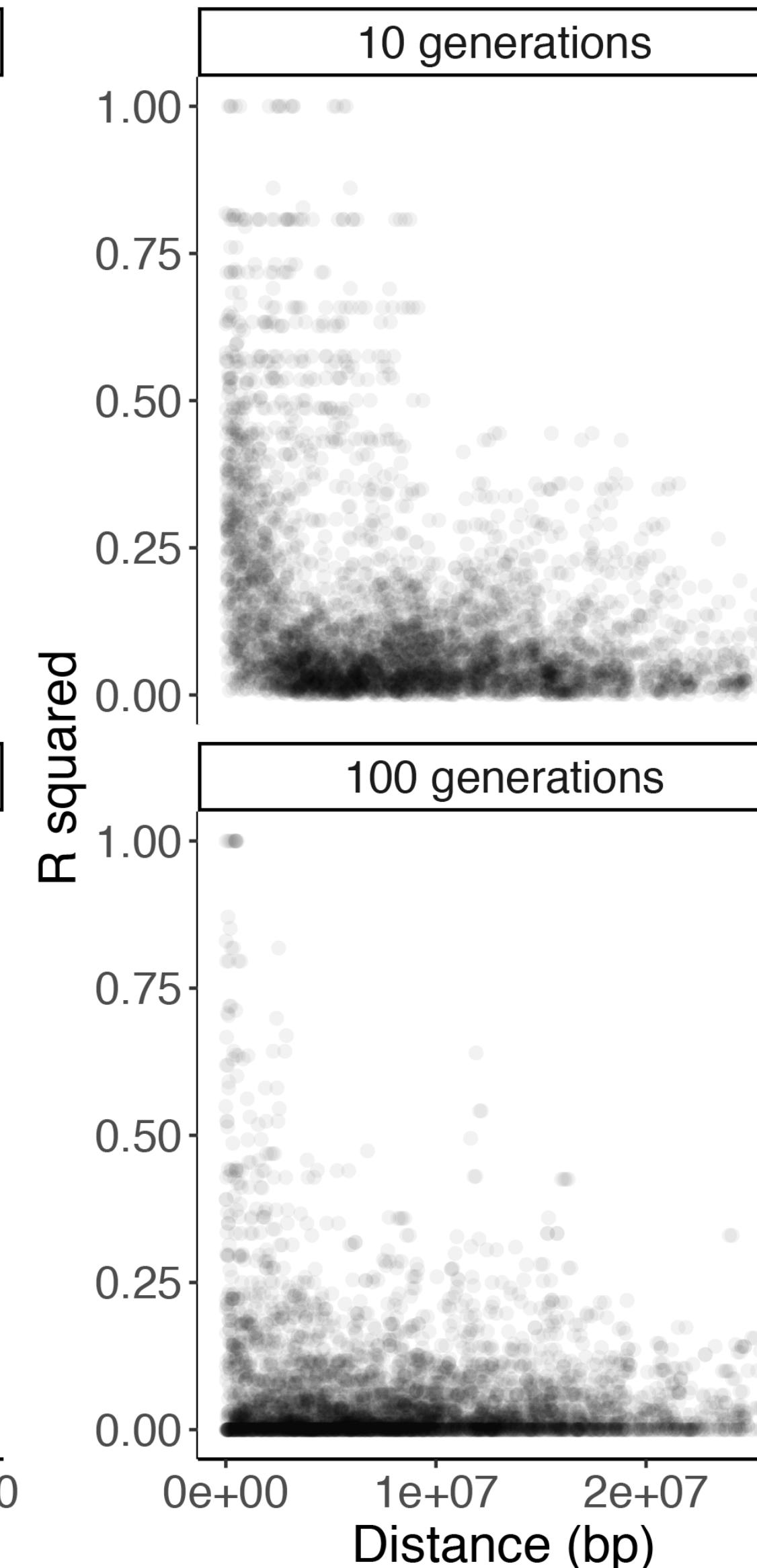
> het_10 <- calculate_heterozygosity(pop_10$population,
  locations = markers)
```

## Linkage Disequilibrium

### Ancestry module

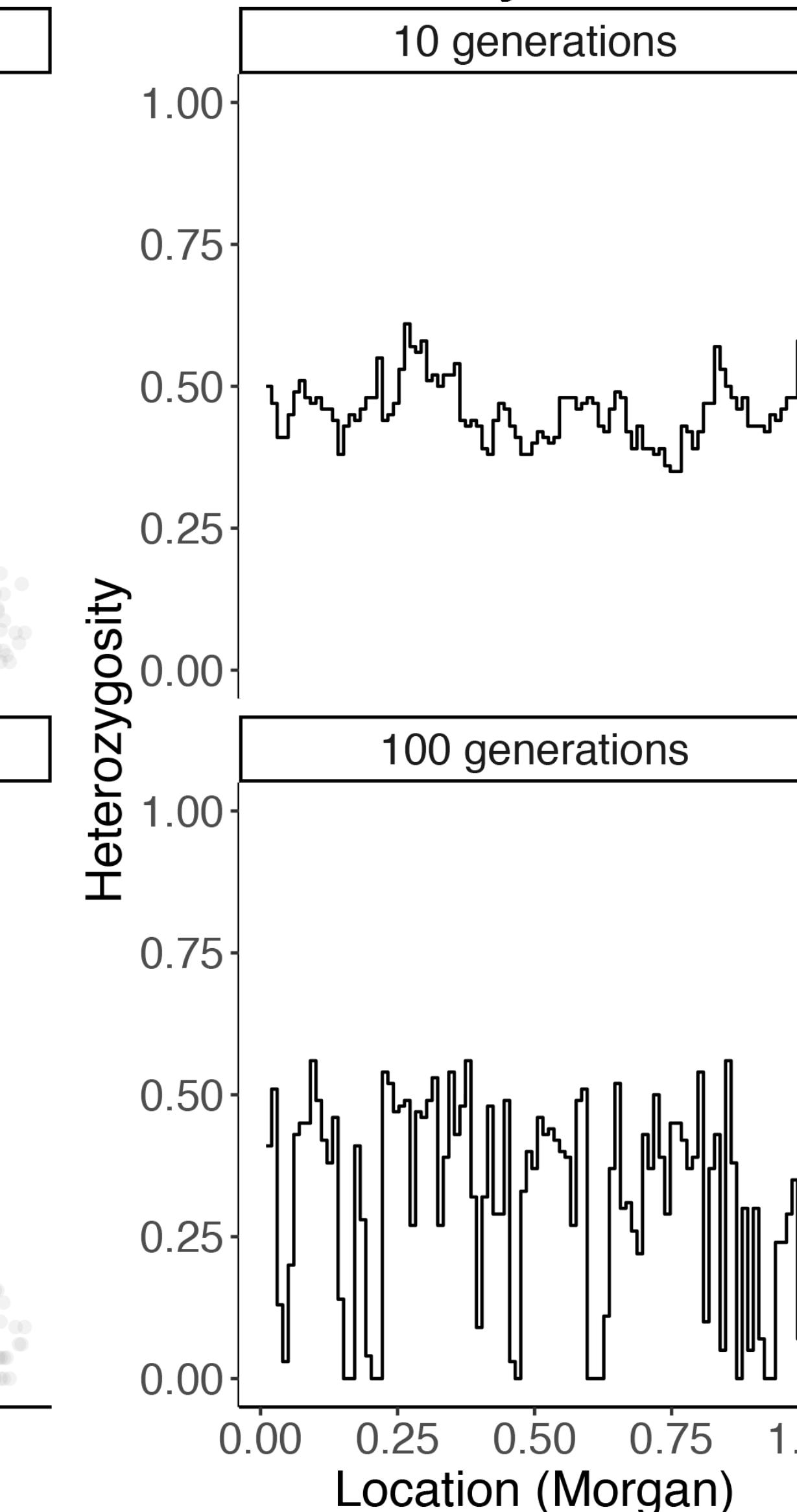


### Sequence module

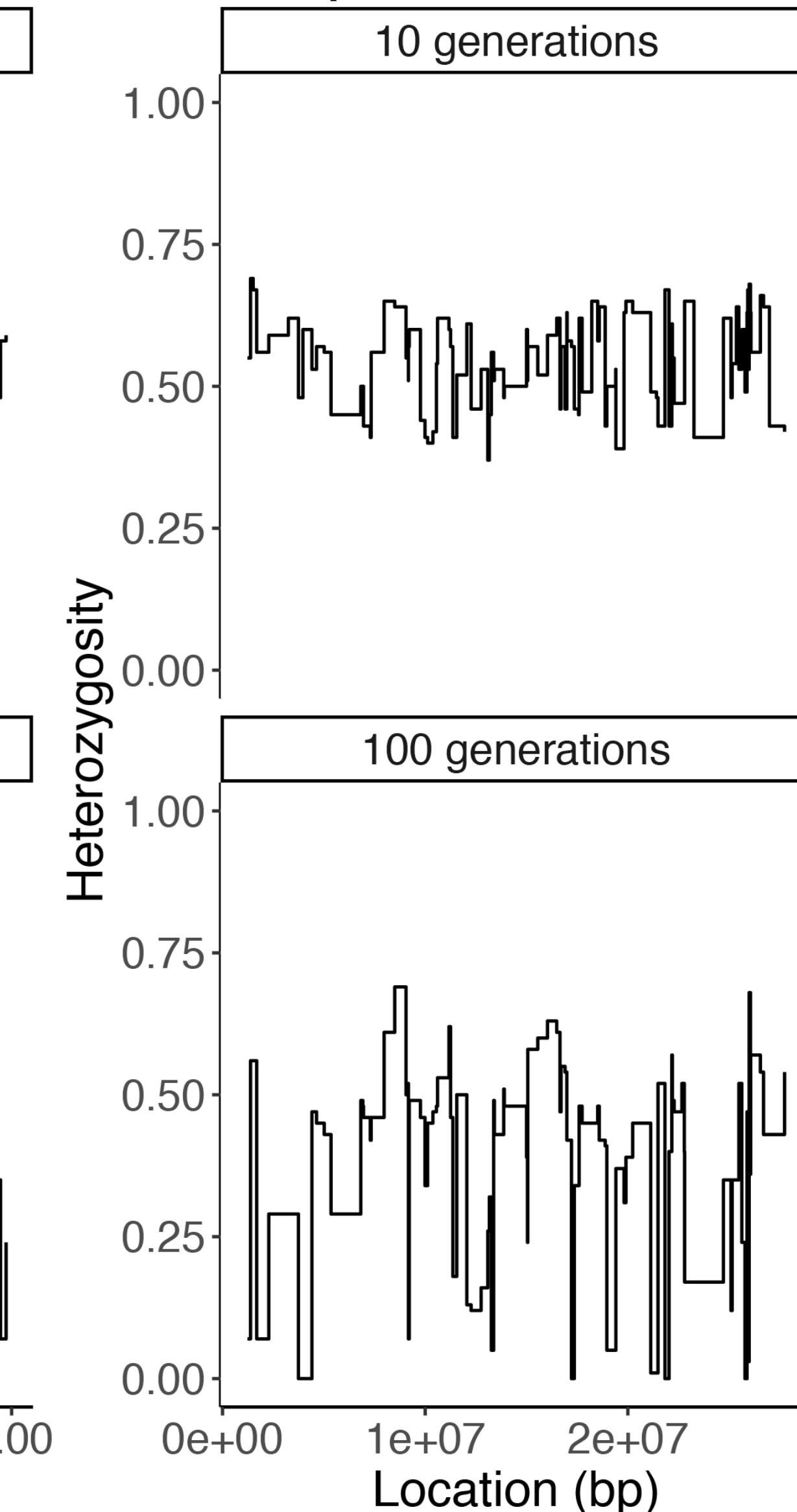


## Heterozygosity

### Ancestry module

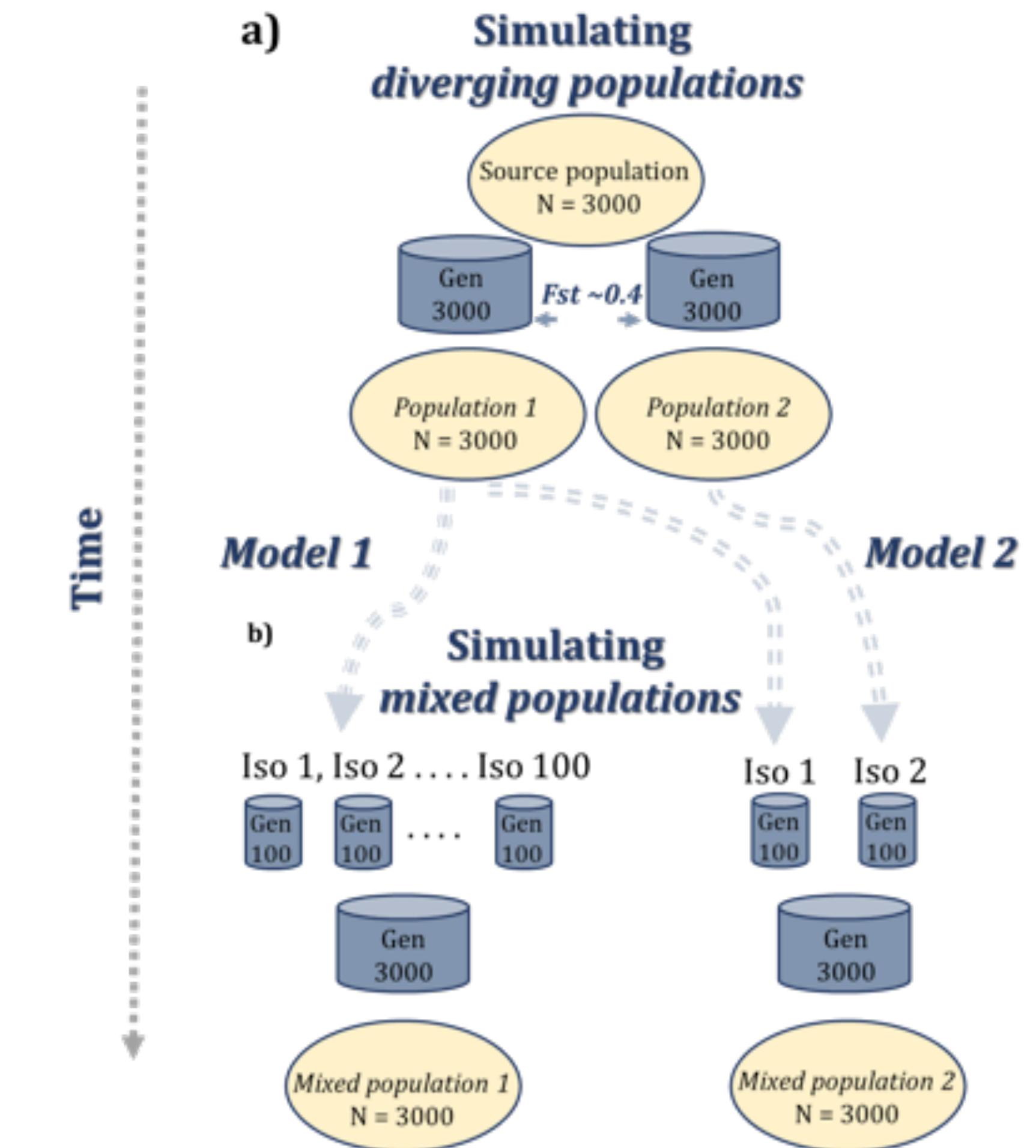


### Sequence module

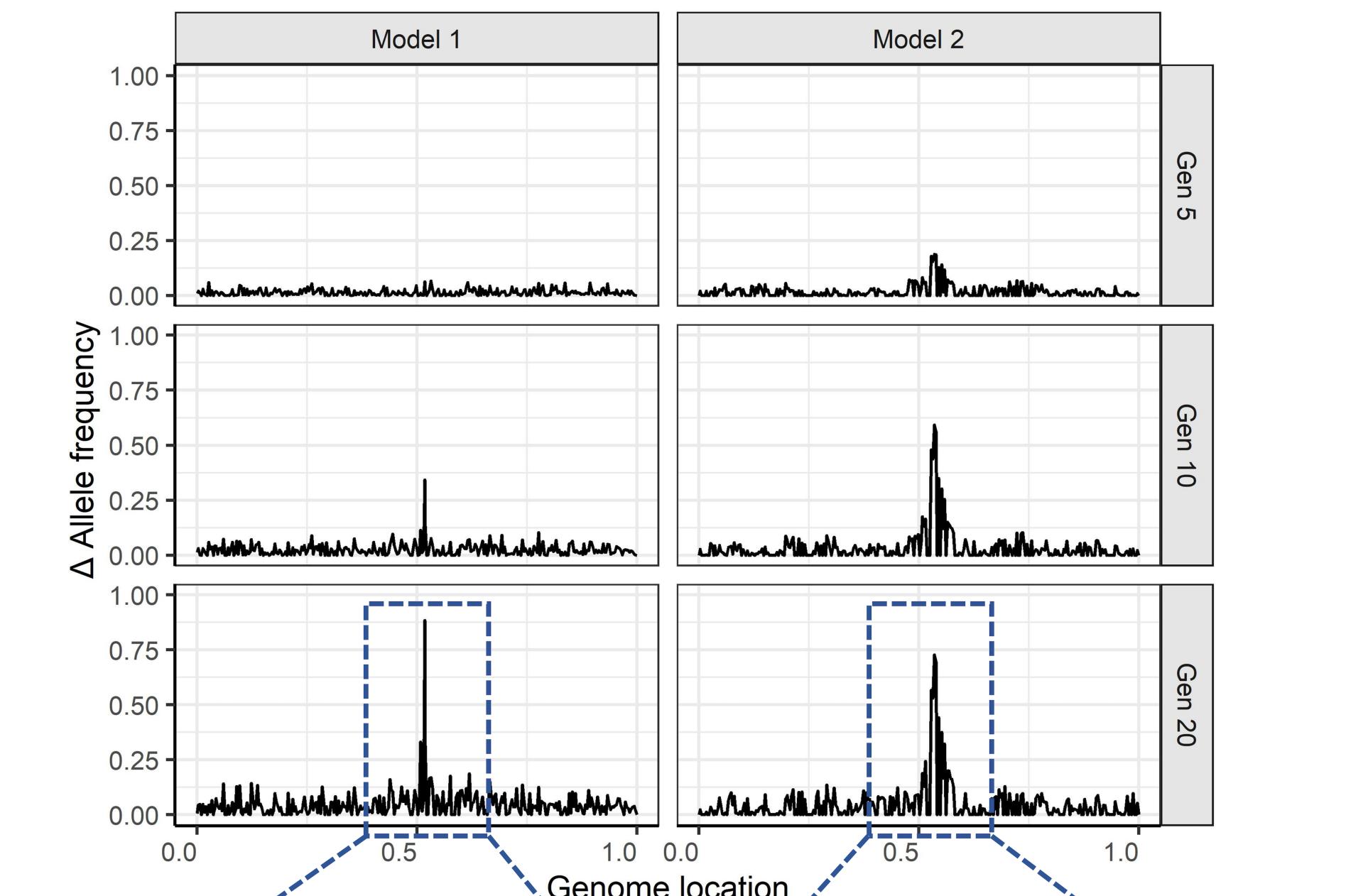


# Real life application: evolve & resequence

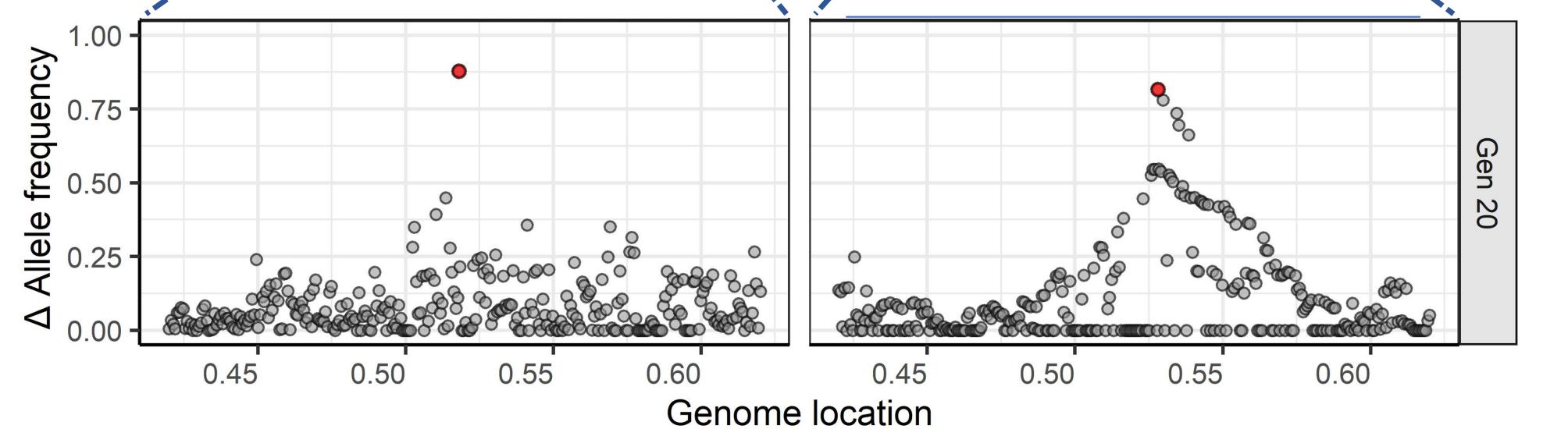
- link observed variation with genetic variation: identify underlying genes
- starting point:
  - model I: individuals sampled from wild population
  - model II: individuals sampled from multiple populations
- Then, the mixed population is exposed to (artificial) selection, and the resulting selected population is sequenced to identify genomic changes.



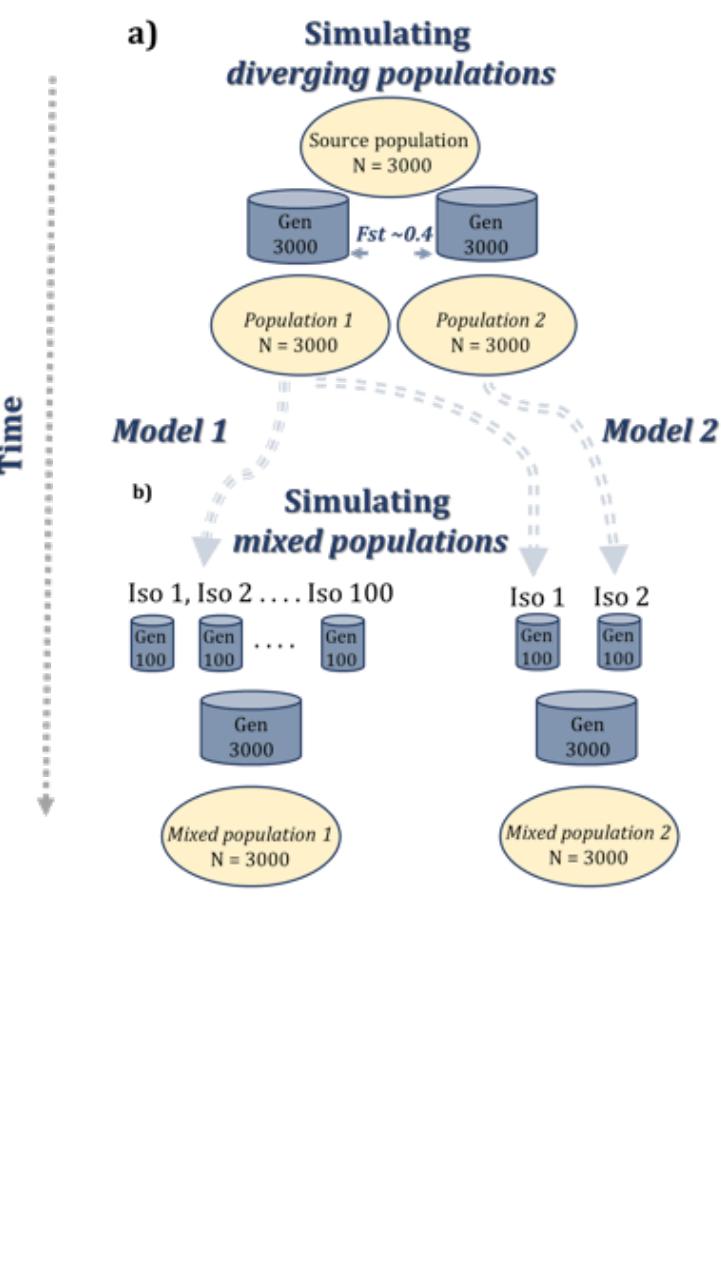
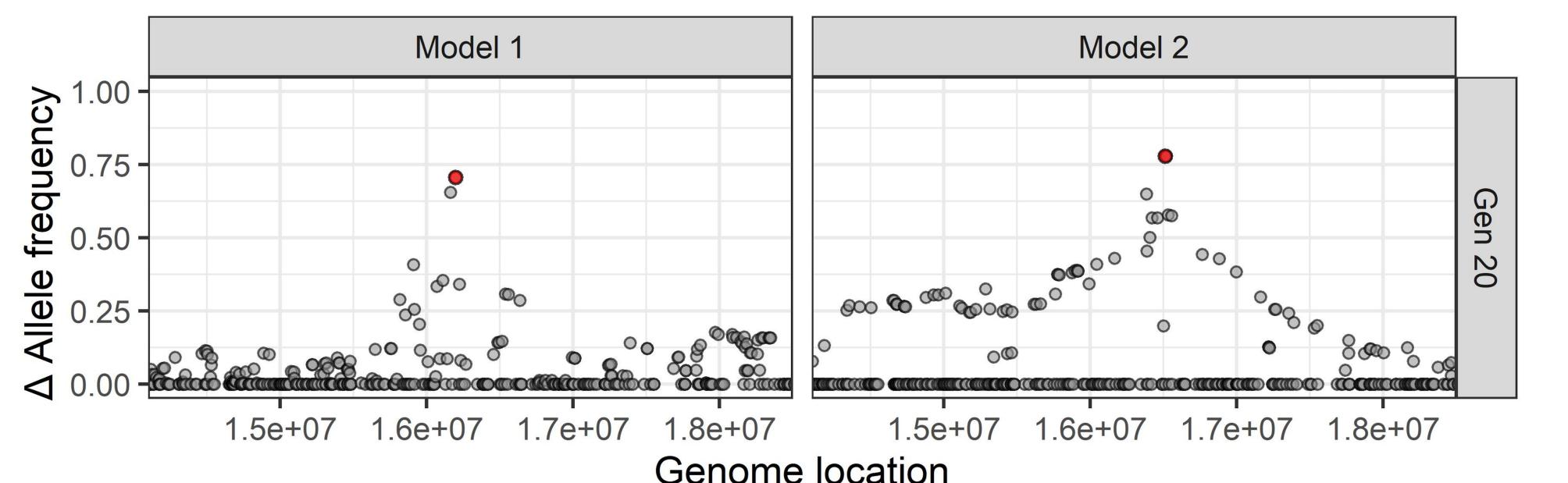
# Results



a) Ancestry module



b) Sequence module



# Concluding remarks

- GenomeAdmixR uses:
  - Rcpp
  - RcppParallel
  - Tidyverse: data as tibbles, plots as ggplot objects
- Development takes place here: <https://github.com/thijsjanzen/GenomeAdmixR>

Thank you!