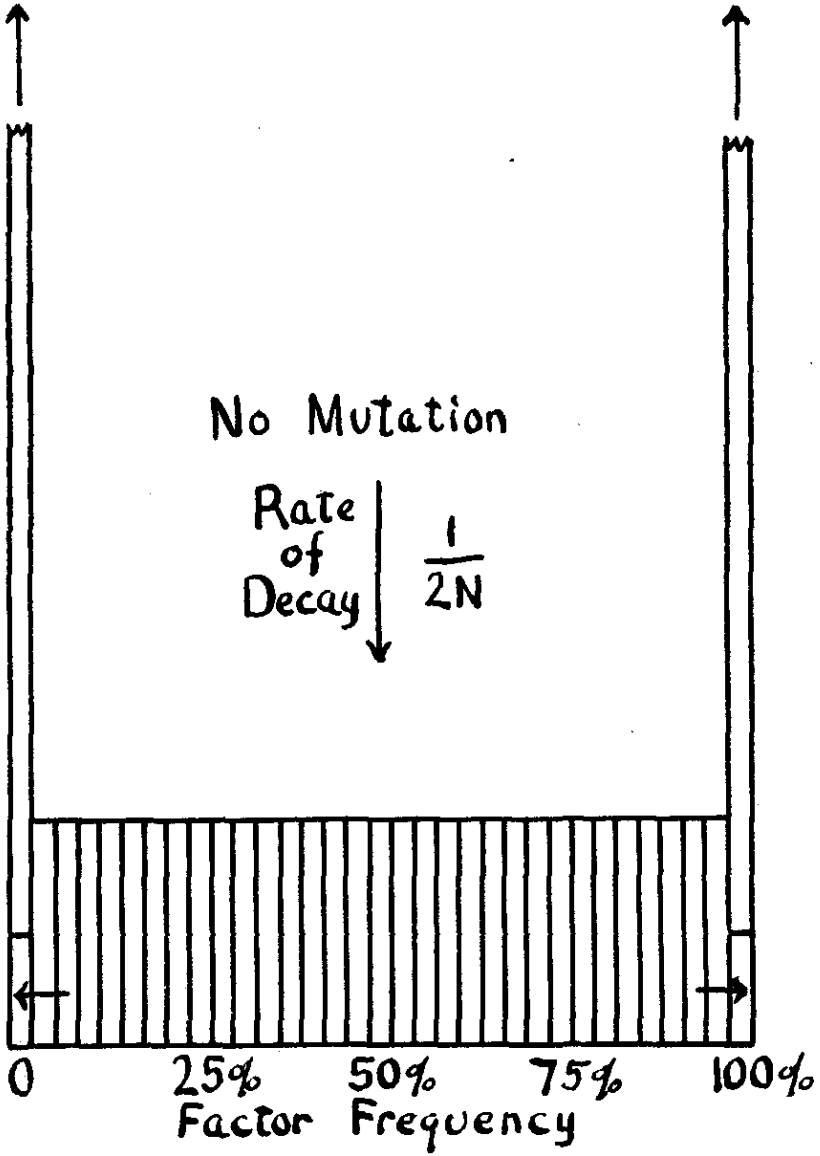
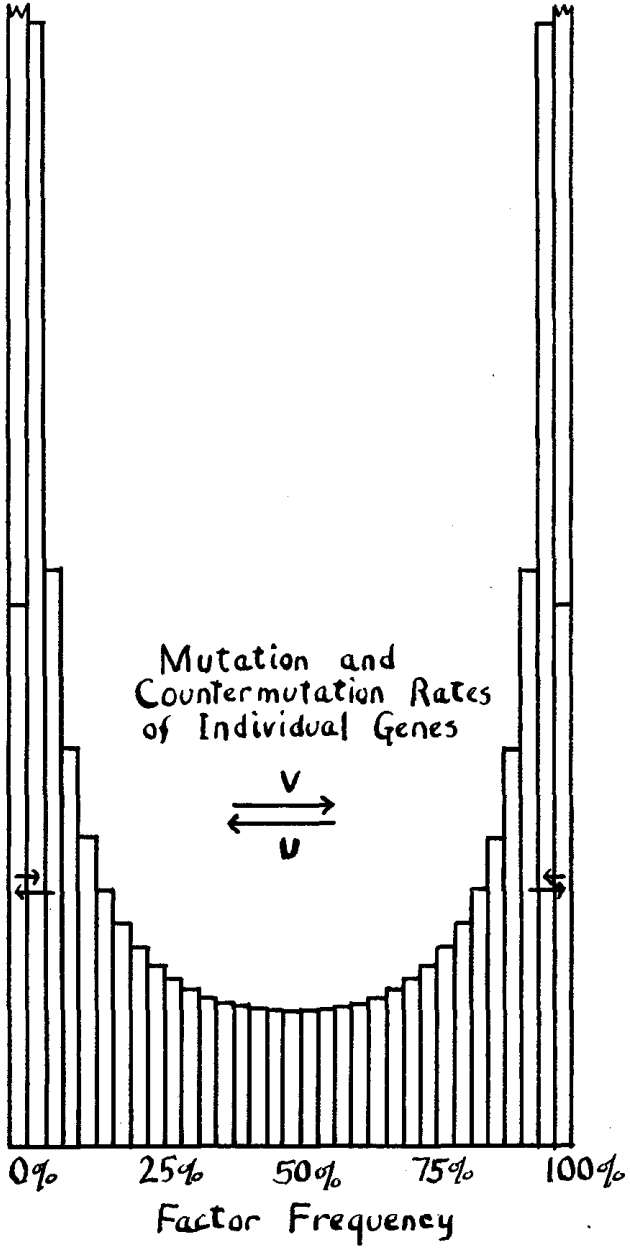


Drift and Mutation

January 25, 2024

Hancock



Source: Wright, 1931

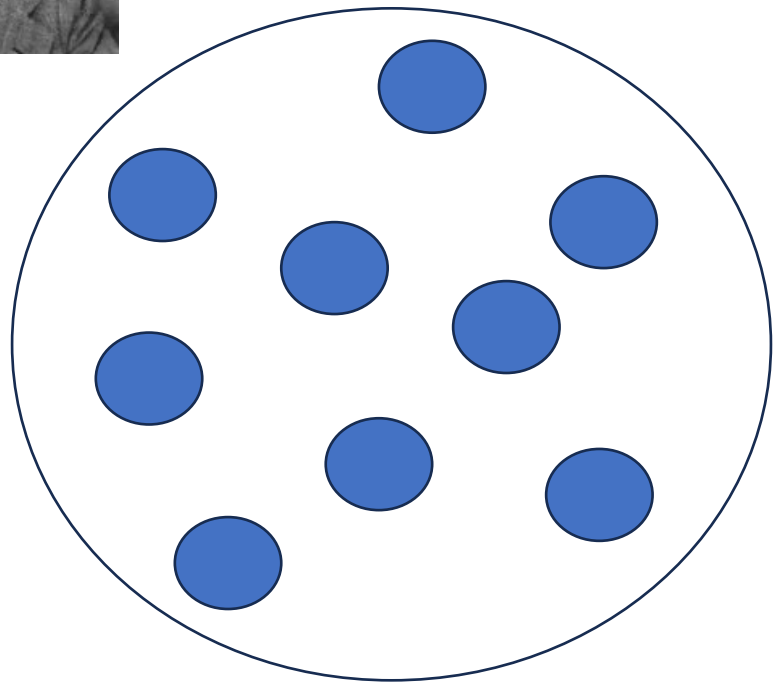
Two different views

Ronald Fisher

- Populations are large and structure is minimal
- Dynamics in small populations ***are un***important

Sewall Wright

- Populations are highly structured
- Dynamics in small populations ***are important***



“Cumulative accidents of sampling”

The very large animals of one strain (No. 13) had such short legs that they seemed to glide on the floor like oversized planarians. The small animals of strain No. 2 had legs as long or longer than the preceding and ran well off the floor. Those of strain No. 13 had rounded noses and bent ears. Those of No. 2 had pointed noses and erect ears. Those of strains No. 39 had notably swayed backs. Strain 35 had protruding eyes; strain 13 sunken ones. […]There were notable differences in temperament. The pigs of strain 13 could be picked up like sacks of meal while those of strains 2 and 35 would struggle and kick a hole in one’s wrist unless picked up properly.

- 1977 Speech Sewall Wright (source: https://genestogenomes.org/sewall-wright-evolving-mendel/)

Sewell Wright’s view was influenced by his studies in animal breeding and guinea pigs

Wright's world view and his influence on population genetics

- Wright thought that natural populations were made up of a complex network of subdivided demes
- Therefore, Wright thought that the dynamics that occurred in small populations were important more generally
- Wright is given credit for developing the mathematical **theory of genetic drift** and the effective population size (N_e)
- He was also responsible for developing the fixation indices (F-statistics), including the inbreeding coefficient, F, and the measure of population differentiation, F_{ST}

We will discuss N_e, the inbreeding coefficient and F_{ST} in future lectures

What is the probability that two “gene copies” have the same state in the next generation?

Step 1. Choose the first gene copy at random.

What is the probability that two “gene copies” have the same state in the next generation?

Step 1. Choose the first gene copy at random.

There are two ways the second gene copy could have the same state:

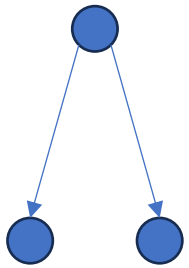
What is the probability that two “gene copies” have the same state in the next generation?

Step 1. Choose the first gene copy at random.

There are two ways the second gene copy could have the same state:

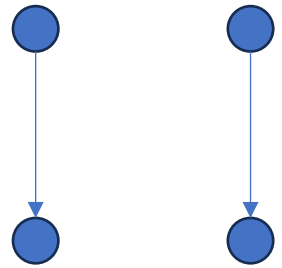
Generation 0

Generation 1



$\frac{1}{2N}$

OR



$\left(1-\frac{1}{2N}\right)\mathcal{G}$

Probability that two gene copies are the same because they derive from the same parental allele (IBD)

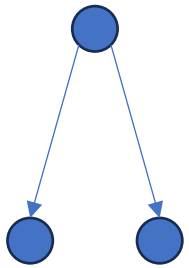
What is the probability that two gene copies have the same state in the next generation?

Step 1. Choose the first gene copy at random.

There are two ways the second gene copy could have the same state:

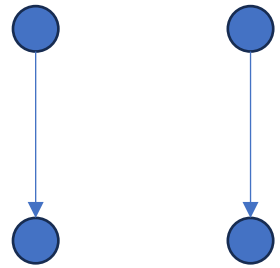
Generation 0

Generation 1



$\frac{1}{2N}$

OR



$\left(1-\frac{1}{2N}\right)\mathcal{G}$

Probability that the gene copy is not IBD but has the same state (IBS),

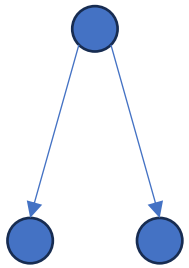
What is the probability that two gene copies have the same state in the next generation?

Step 1. Choose the first gene copy at random.

There are two ways the second gene copy could have the same state:

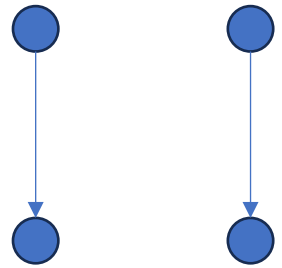
Generation 0

Generation 1



$\frac{1}{2N}$

OR



$\left(1-\frac{1}{2N}\right)\mathcal{G}$

Probability that the gene copy is not IBD but has the same state (IBS),

where $\mathcal{G}$ = probability that two randomly selected gene copies have the same state in this generation

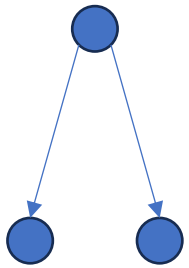
What is the probability that two gene copies have the same state in the next generation, $\mathcal{G}'$?

Step 1. Choose the first gene copy at random.

There are two ways the second gene copy could have the same state:

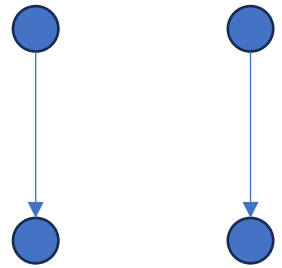
Generation 0

Generation 1



$\frac{1}{2N}$

OR



$\left(1-\frac{1}{2N}\right)\mathcal{G}$

So, the probability that two gene copies chosen at random have the same state is:

$$\mathcal{G}' = \frac{1}{2N} + \left(1-\frac{1}{2N}\right)\mathcal{G}$$

What is the probability that two gene copies at a locus have different states in the next generation ($\mathcal{H}'$)?

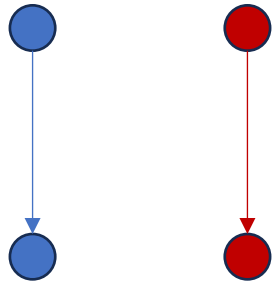
(Assuming no mutation), gene copies can only differ in the next generation if they differed in the previous generation, so:

What is the probability that two gene copies at a locus have different states in the next generation ($\mathcal{H}'$)?

(Assuming no mutation), gene copies can only differ in the next generation if they differed in the previous generation, so:

Generation 0

Generation 1



$$\left(1 - \frac{1}{2N}\right)\mathcal{H}$$

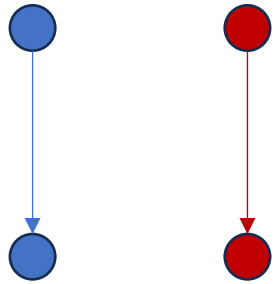
Probability that the a gene copy chosen at random from the population is different

What is the probability that two gene copies at a locus have different states in the next generation ($\mathcal{H}'$)?

(Assuming no mutation), gene copies can only differ in the next generation if they differed in the previous generation, so:

Generation 0

Generation 1



$$\left(1 - \frac{1}{2N}\right)\mathcal{H}$$

Probability that the an allele chosen at random from the population is different,

where $\mathcal{H}$ = probability two randomly selected alleles have a different state in this generation

What is the probability that two gene copies at a locus have different states in the next generation ($\mathcal{H}'$)?

(Assuming no mutation), gene copies can only differ in the next generation if they differed in the previous generation, so:

So, the probability that two alleles chosen at random are different is:

$$\mathcal{H}' = \left(1 - \frac{1}{2N}\right)\mathcal{H}$$

What is the probability that two gene copies at a locus have different states in the next generation ($\mathcal{H}'$)?

(Assuming no mutation), gene copies can only differ in the next generation if they differed in the previous generation, so:

$$\mathcal{H}' = \left(1 - \frac{1}{2N}\right)\mathcal{H}$$

Since all genotypes are either heterozygous or homozygous, $\mathcal{H}' + \mathcal{G}'$ = 1 and $\mathcal{H}'$= 1 - $\mathcal{G}'$

How does heterozygosity ($\mathcal{H}$) change over time?

$$\mathcal{H}' = \left(1 - \frac{1}{2N}\right)\mathcal{H}$$

$$\mathcal{H}' = \mathcal{H} - \frac{\mathcal{H}}{2N}$$

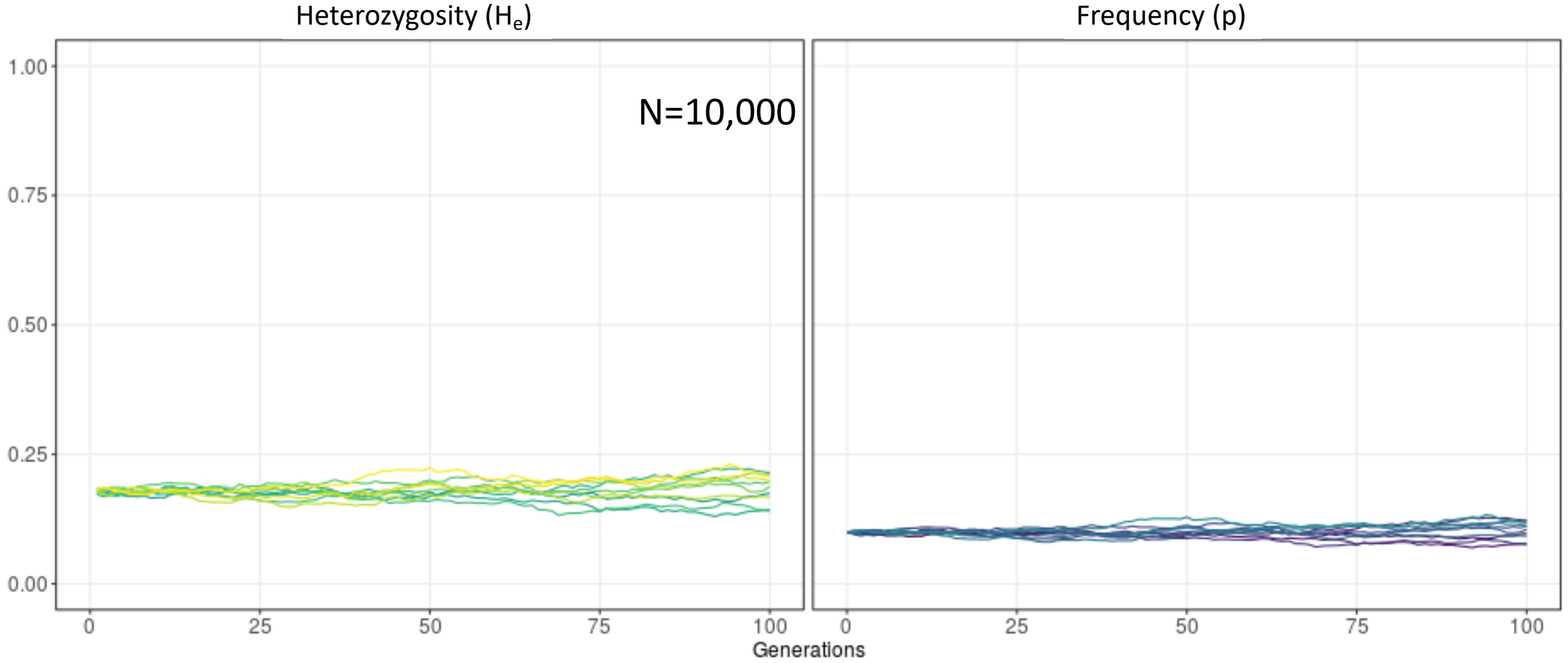
$\mathcal{H}$ decreases at a rate of 1/2N, and

$$\Delta\mathcal{H} = \mathcal{H}' - \mathcal{H} = -\frac{1}{2N}\mathcal{H}$$

$\mathcal{H}$ at time t is:

$$\mathcal{H}_t = \mathcal{H}_0(1 - \frac{1}{2N})^t$$

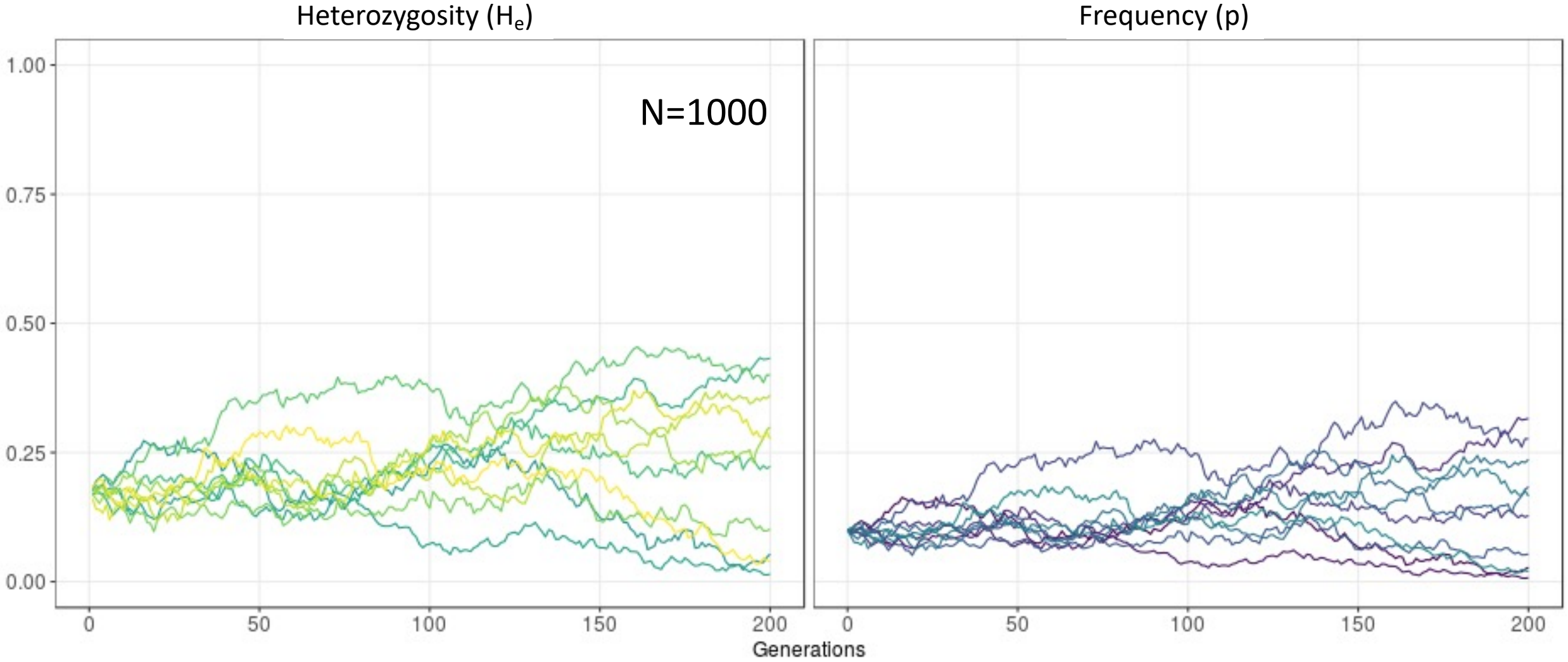
When population size is large, variation can be maintained for many generations



Simulations, $p_{t=0}$=0.1, N=10K, 200 generations

https://cjbattey.shinyapps.io/driftR/

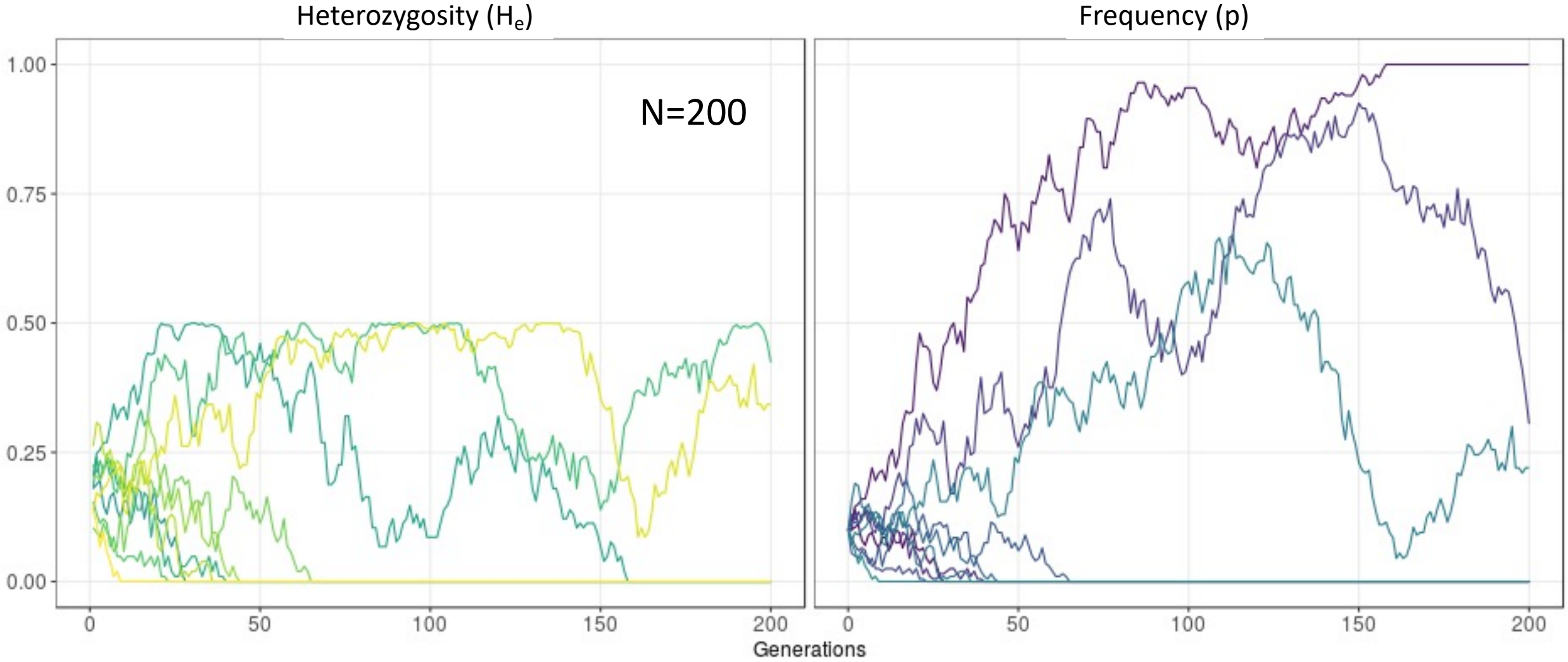
When population size is large, variation can be maintained for many generations



Simulations, $p_{t=0}$=0.1, N=1000, 200 generations

https://cjbattey.shinyapps.io/driftR/

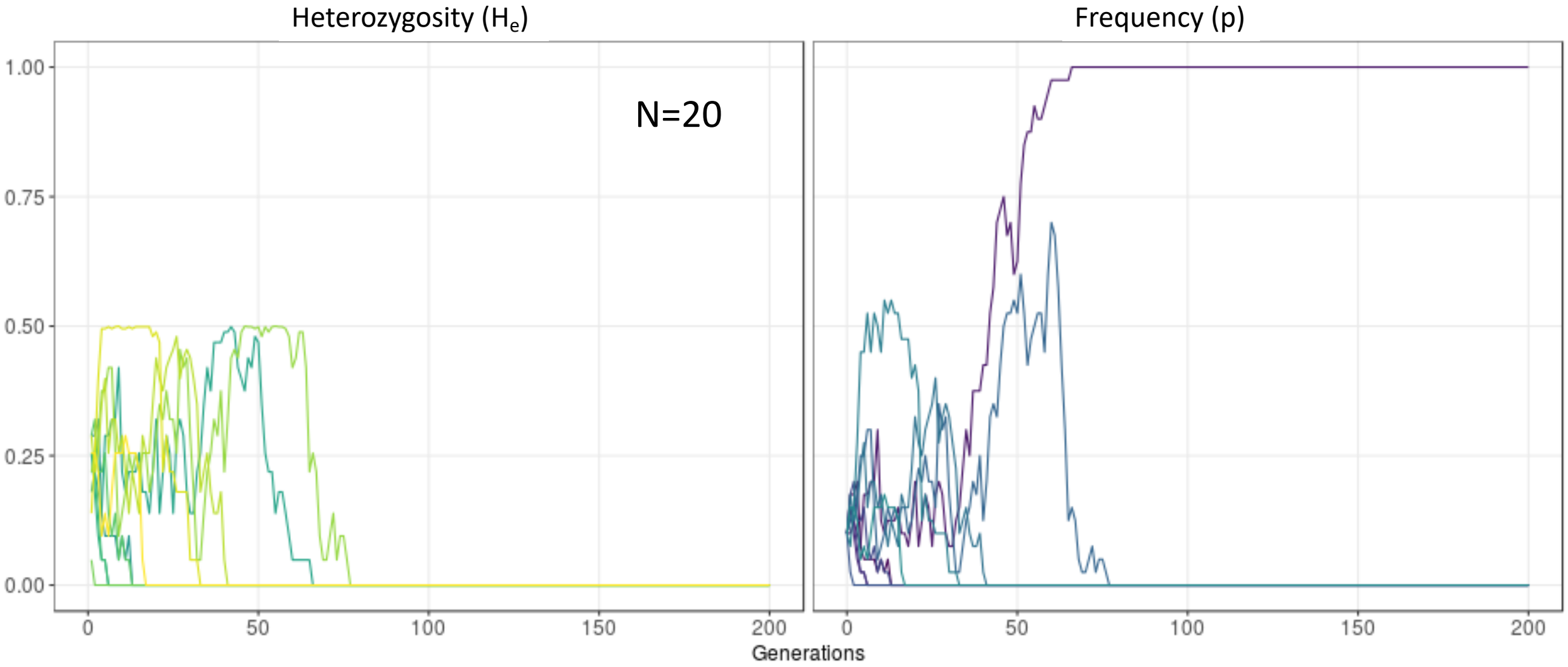
As population size is reduced, alleles are more likely to fix and heterozygosity often decreases



Simulations, $p_{t=0}$=0.1, N=200, 200 generations

https://cjbattey.shinyapps.io/driftR/

In a very small population, variation is lost from the population



Simulations, $p_{t=0}$=0.1, N=20, 200 generations

https://cjbattey.shinyapps.io/driftR/

In case you would like to run the drift simulation on your own

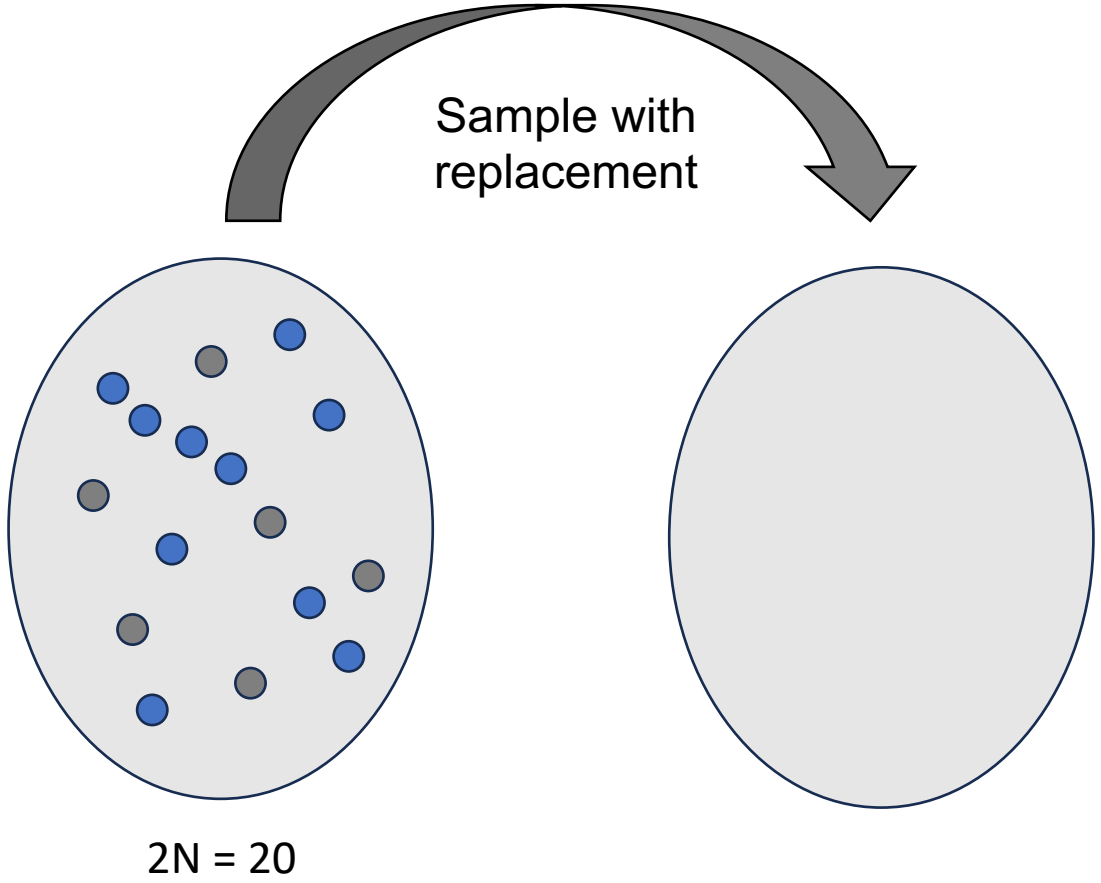
- https://cjbattey.shinyapps.io/driftR/
- If you want to play with this app at home, simplify the simulations by setting:
- Mutation rate = 0
- Starting allele frequency A = 0.1
- Fitness = 1 for all genotypes
- Migration rate = 0
- Number of populations = 10 (default)

Remember that these simulations involve taking a random sample, so the results are stochastic. You can try running the same simulations many times or with many populations to see how they change

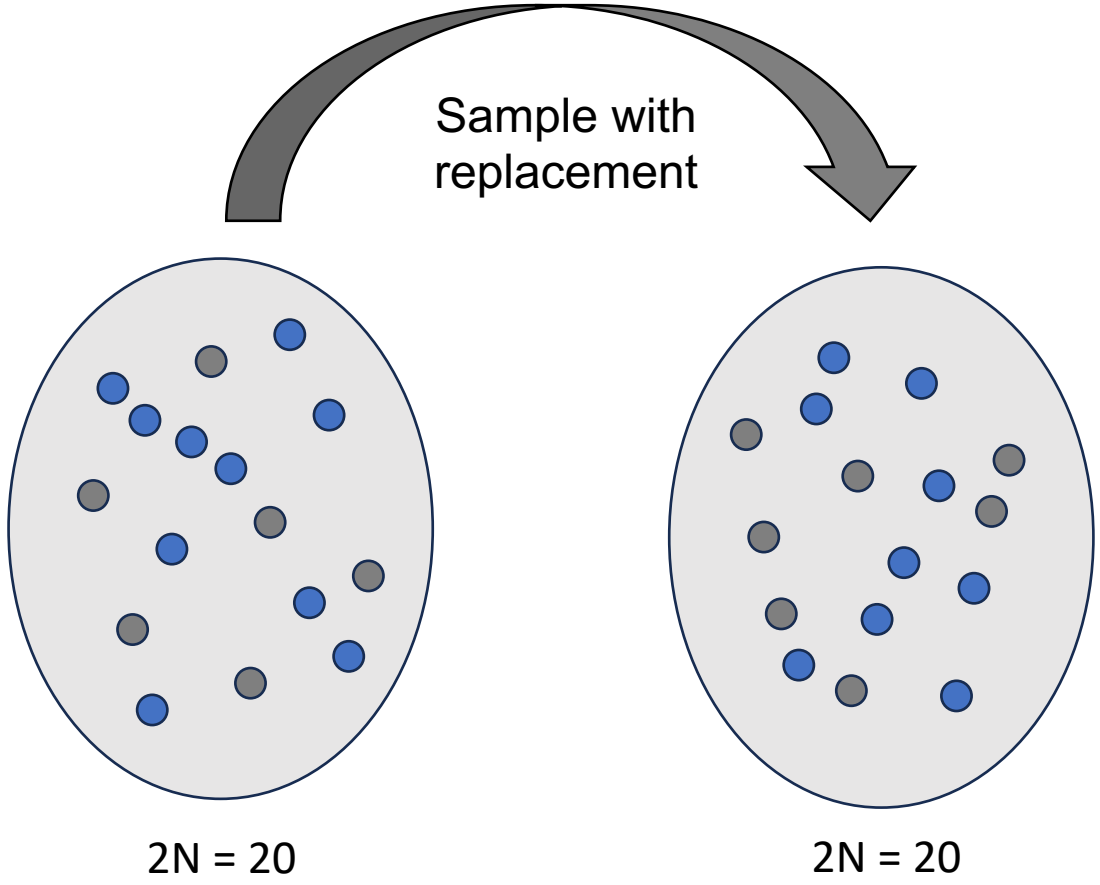
Wright-Fisher model (a version of the urn model)

- A commonly-used simplified model of population evolution
- Named after Sewell Wright and Ronald Fisher
- N hermaphroditic individuals
- Constant population size, so that the same number of chromosomes are sampled in each generation
- Alleles (gene copies) are sampled with replacement
- Discrete (non-overlapping) generations

Sampling under W-F

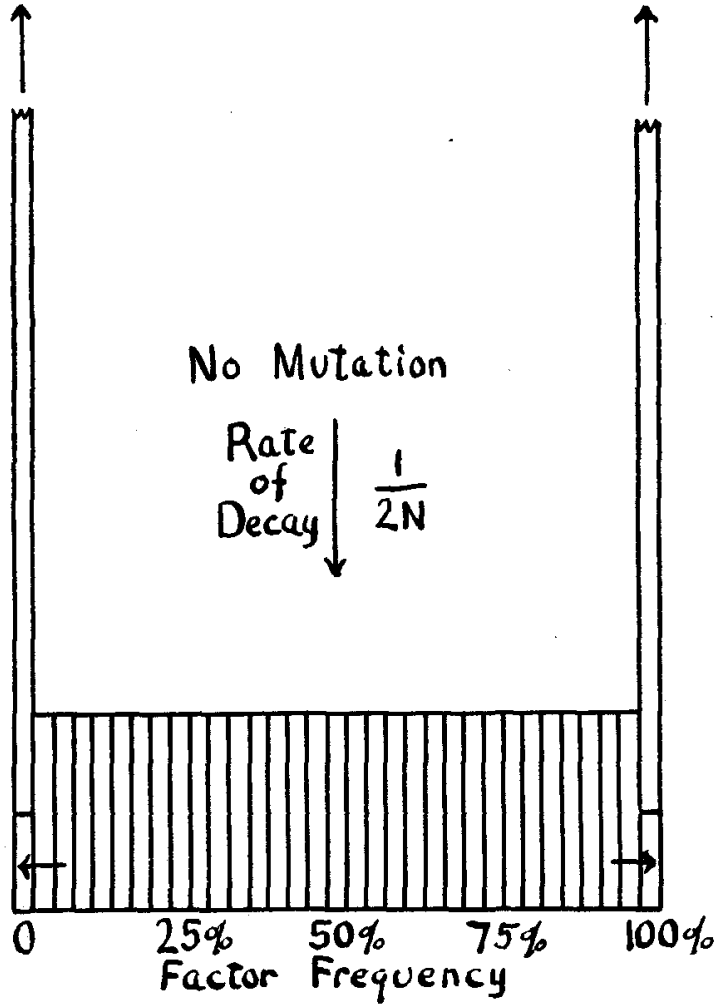


Sampling under W-F



Population size stays constant over generations

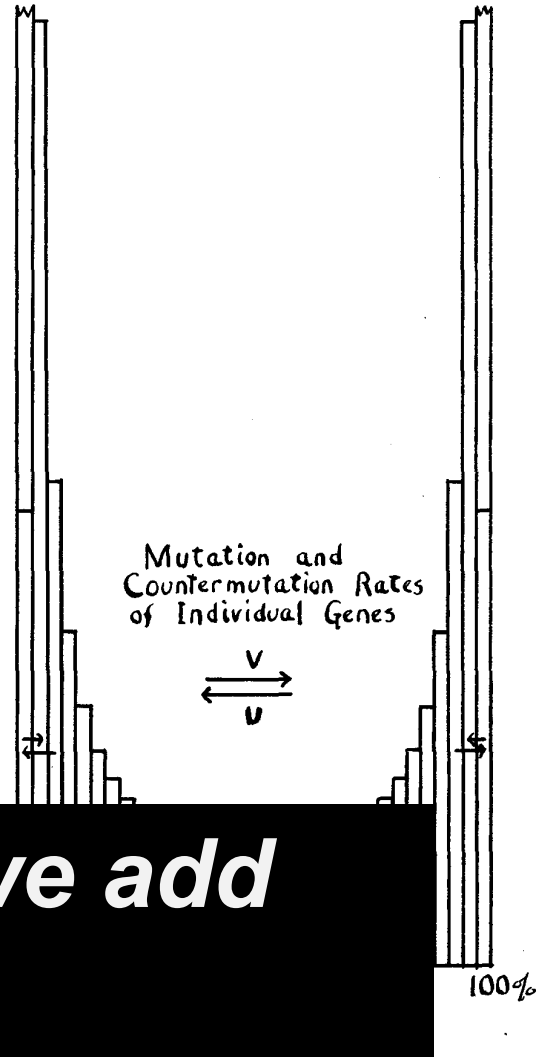
If genetic drift is constantly removing variation from populations, why do do we still find variation?



Source: Wright, 1931

If genetic drift is constantly removing variation from populations, why do do we still find variation?

Mutation adds variation to populations!



Let's see what happens when we add mutation to the model

Source: Wright, 1931

Drift versus mutation

- Genetic drift removes variation from the population at the rate of 1/2N
- Mutation adds variation to the population at the rate of $2N\mu$

How to model mutation?

A simple model:

<u>the infinite alleles model</u>

- Each mutation creates a unique allele
- Gene copies remain identical in proceeding generations only if neither mutates
- Probability of mutation is μ

How drift affects expected homozygosity

Without mutation

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}$$

With mutation

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}\right]$$

How drift affects expected homozygosity

Without mutation

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}$$

With mutation

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}\right]$$

Stated as a sentence:

The two gene copies are the same if neither mutates and they were either inherited from the same gene copy (IBD) or they had the same state.

How drift affects expected homozygosity

Without mutation

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}$$

With mutation

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}\right]$$

Stated as a sentence:

The two gene copies are the same if neither mutates and they were either inherited from the same gene copy (IBD) or they had the same state.

How drift affects expected homozygosity

Without mutation

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}$$

With mutation

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}\right]$$

Stated as a sentence:

The two gene copies are the same if neither mutates and they were either inherited from the same gene copy (IBD) or they had the same state.

How drift affects expected homozygosity

Without mutation

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}$$

With mutation

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}\right]$$

Stated as a sentence:

The two gene copies are the same if neither mutates and they were either inherited from the same gene copy (IBD) or they had the same state.

We can simplify these equations with some approximations

$$(1-\mu)^2 \approx 1-2\mu$$

$$\frac{1-2\mu}{2N} \approx \frac{1}{2N}$$

Why can we make these approximations?
How similar are they really?

Numerical example: $(1-\mu)^2 \approx 1-2\mu$

μ	$(1-\mu)^2$	$1-2\mu$
0.100000	0.810000	0.800000
0.010000	0.980100	0.980000
0.001000	0.998001	0.998000
0.000100	0.999800	0.999800
0.000010	0.999980	0.999980
0.000001	0.999998	0.999998

Numerical example: $\frac{1-2\mu}{2N} \approx \frac{1}{2N}$

μ	$2N$	$\frac{1-2\mu}{2N}$	$\frac{1}{2N}$
0.0001	10	0.0999800	0.10000
0.0001	100	0.009980	0.01000
0.0001	1000	0.0009998	0.00100
0.0001	10000	0.000100	0.00010
0.0001	100000	0.0000100	0.00001

We can simplify these equations with some approximations

Original formula:

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1-\frac{1}{2N}\right)\mathcal{G}\right]$$

After approximations:

$$\mathcal{G}' = \frac{1}{2N} + \left(1-\frac{1}{2N}\right)\mathcal{G} - 2\mu\mathcal{G}$$

At equilibrium $\mathcal{G}' = \mathcal{G}$, so :

$$\hat{\mathcal{G}} = \frac{1}{1+4N\mu} \quad \text{and} \quad \hat{\mathcal{H}} = 1-\hat{\mathcal{G}} = \frac{4N\mu}{1+4N\mu}$$

We can simplify these equations with some approximations

Original formula:

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G}\right]$$

After approximations:

$$\mathcal{G}' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\mathcal{G} - 2\mu\mathcal{G}$$

At equilibrium $\mathcal{G}' = \mathcal{G}$, so :

The probability at equilibrium that two gene copies are identical by state

$$\hat{\mathcal{G}} = \frac{1}{1+4N\mu} \quad \text{and} \quad \hat{\mathcal{H}} = 1 - \hat{\mathcal{G}} = \frac{4N\mu}{1+4N\mu}$$

We can simplify these equations with some approximations

Original formula:

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1-\frac{1}{2N}\right)\mathcal{G}\right]$$

After approximations:

$$\mathcal{G}' = \frac{1}{2N} + \left(1-\frac{1}{2N}\right)\mathcal{G} - 2\mu\mathcal{G}$$

At equilibrium $\mathcal{G}' = \mathcal{G}$, so :

$$\hat{\mathcal{G}} = \frac{1}{1+4N\mu} \quad \text{and} \quad \hat{\mathcal{H}} = 1-\hat{\mathcal{G}} = \frac{4N\mu}{1+4N\mu}$$

The probability at equilibrium that two gene copies are different by state

We can simplify these equations with some approximations

Original formula:

$$\mathcal{G}' = (1-\mu)^2\left[\frac{1}{2N} + \left(1-\frac{1}{2N}\right)\mathcal{G}\right]$$

After approximations:

$$\mathcal{G}' = \frac{1}{2N} + \left(1-\frac{1}{2N}\right)\mathcal{G} - 2\mu\mathcal{G}$$

At equilibrium $\mathcal{G}' = \mathcal{G}$, so :

$$\hat{\mathcal{G}} = \frac{1}{1+4N\mu} \quad \text{and} \quad \hat{\mathcal{H}} = 1-\hat{\mathcal{G}} = \frac{4N\mu}{1+4N\mu}$$

When $4N\mu$ is very large, $\hat{\mathcal{G}} \approx 0$ and $\hat{\mathcal{H}} \approx 1$

$\hat{\mathcal{G}}$ and $\hat{\mathcal{H}}$ are expected frequencies at equilibrium

$$\hat{\mathcal{G}} = \frac{1}{1 + 4N\mu}$$

- $\hat{\mathcal{G}}$ is the equilibrium frequency of homozygous genotypes in the population
- $\hat{\mathcal{G}}$ is the same as the inbreeding coefficient, $\hat{F}$

$$\hat{\mathcal{H}} = 1 - \hat{\mathcal{G}} = \frac{4N\mu}{1 + 4N\mu}$$

$\hat{\mathcal{H}}$ is the equilibrium frequency of heterozygous genotypes in the population

When $4N\mu$ is very large, $\hat{\mathcal{G}} \approx 0$ and $\hat{\mathcal{H}} \approx 1$

$4N\mu$ is also referred to as θ, which is a measure of diversity

$4N\mu$ is also referred to as θ, which is a measure of diversity

$$\hat{\mathcal{G}} = \frac{1}{1 + 4N\mu} = \frac{1}{1 + \theta}$$

$$\hat{\mathcal{H}} = 1 - \hat{\mathcal{G}} = \frac{4N\mu}{1 + 4N\mu} = \frac{\theta}{1 + \theta}$$

We will discuss θ and how it is estimated in population samples in future lectures

Change in heterozygosity is due to population size and mutation

$$\Delta\mathcal{H} \approx \Delta_N\mathcal{H} + \Delta_\mu\mathcal{H}$$

Change in heterozygosity is due to change due to drift and change due to mutation

$$\Delta_N\mathcal{H} = -\frac{1}{2N}\mathcal{H}$$

Change of $\mathcal{H}$ due to genetic drift

$$\Delta_\mu\mathcal{H} = 2\mu(1-\mathcal{H})$$

Change of $\mathcal{H}$ due to mutation

$$\mathcal{H}' = \mathcal{H} + (1-\mathcal{H})[1-(1-\mu)^2]$$

The probability that two alleles differ in the next generation

Change in heterozygosity is due to population size and mutation

$$\mathcal{H}' = \mathcal{H} + (1 - \mathcal{H})[1 - (1 - \mu)^2]$$

Two gene copies will differ in the next generation if

they differ in the current generation

Change in heterozygosity is due to population size and mutation

$$\mathcal{H}' = \mathcal{H} + (1 - \mathcal{H})[1 - (1 - \mu)^2]$$

Two gene copies will differ in the next generation if

they differ in the current generation

Change in heterozygosity is due to population size and mutation

$$\mathcal{H}' = \mathcal{H} + (1 - \mathcal{H})[1 - (1 - \mu)^2]$$

Two gene copies will differ in the next generation if

they differ in the current generation

or

Change in heterozygosity is due to population size and mutation

$$\mathcal{H}' = \mathcal{H} + (1 - \mathcal{H})[1 - (1 - \mu)^2]$$

Two gene copies will differ in the next generation if they differ in the current generation or they are the same in the current generation and a mutation changes the state of one copy

Relevance of heterozygosity and homozygosity

What does homozygosity and heterozygosity tell us about a population?

Many organisms self-fertilize at high rates or are not even diploid during most of their life cycle

Is heterozygosity a relevant measure for them?

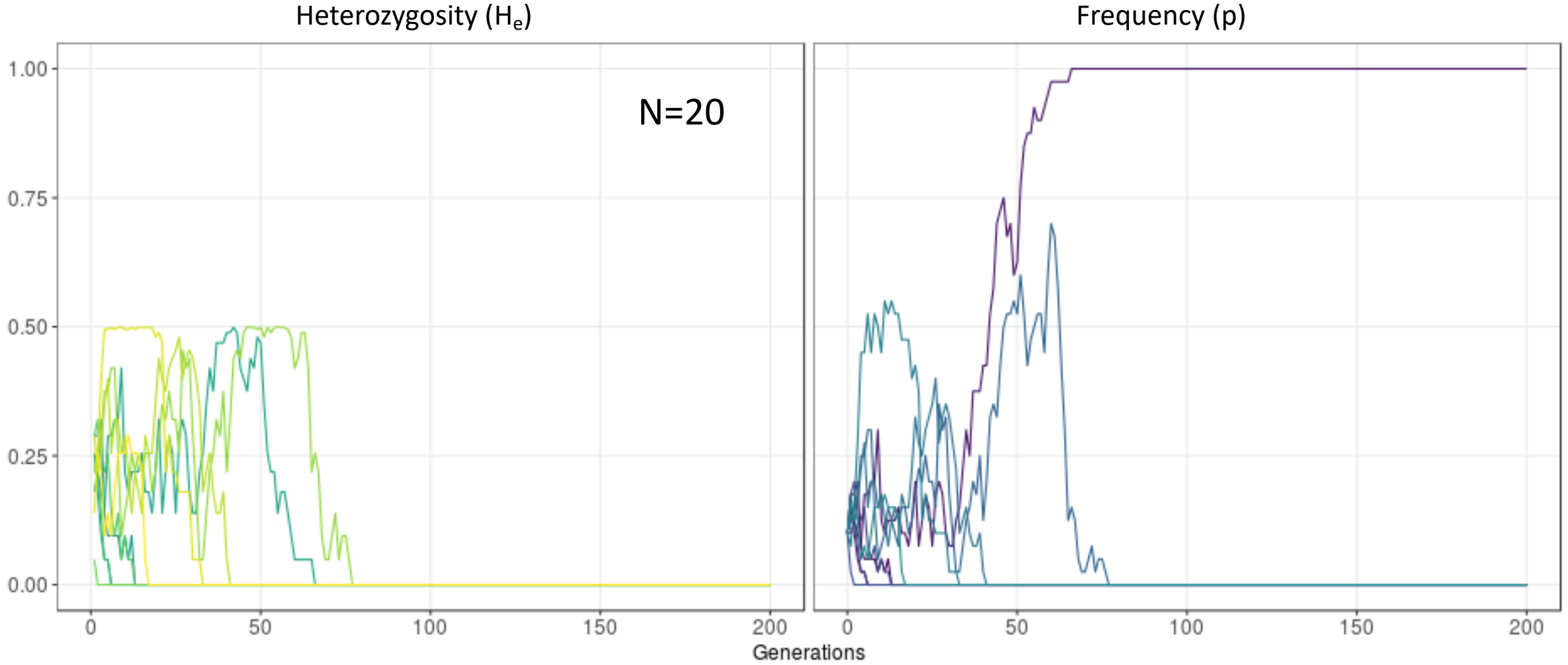
Yes!

- Expected heterozygosity is the ***probability*** of drawing two different gene copies
- It is a useful ***summary of variation*** at the population level even in a population that does not mate at random

More simulations, now with mutation

- https://cjbattey.shinyapps.io/driftR/
- If you want to play with this app at home, simplify the simulations by setting:
- **Mutation rate = 0.001**
- Starting allele frequency A = 0.1
- Fitness = 1 for all genotypes
- Migration rate = 0
- Population size = 20

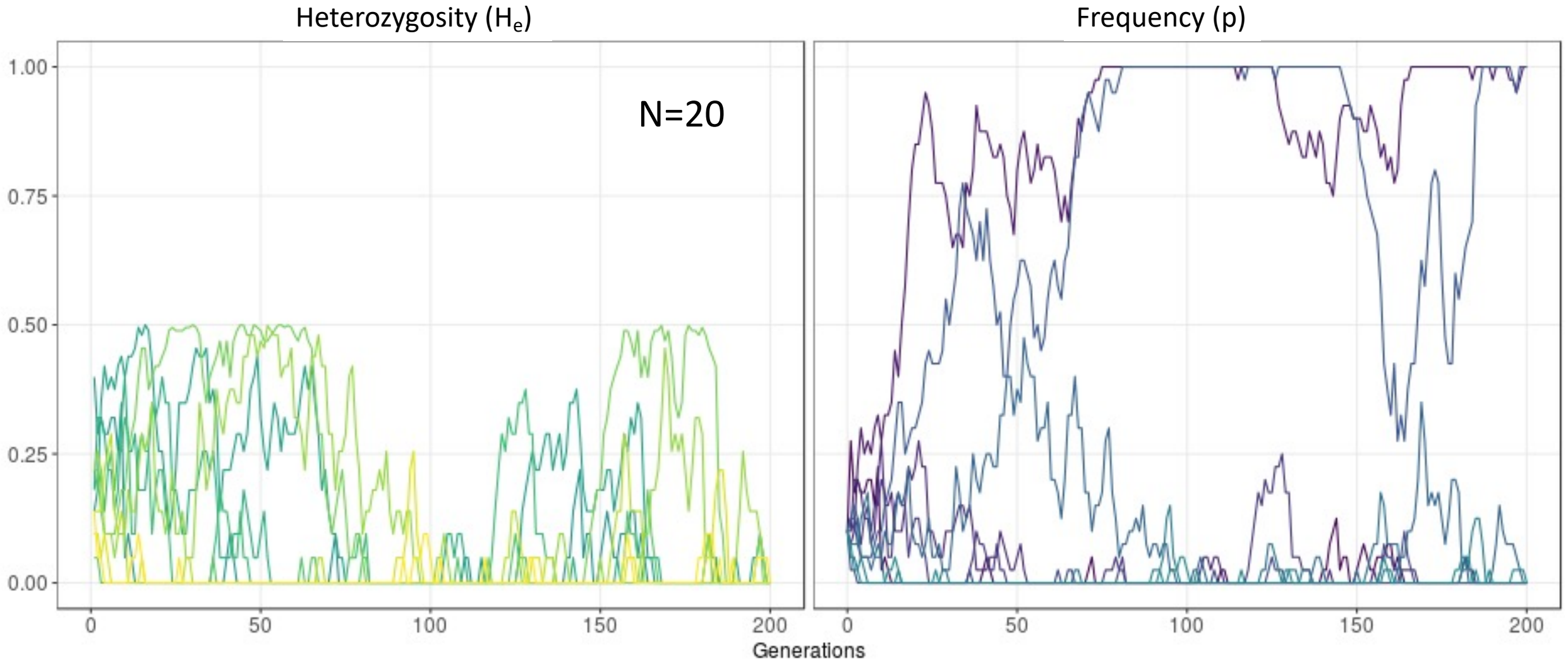
Recall that with drift alone variation is lost when the population size is very small!



Simulations, $p_{t=0}$=0.1, N=20, no mutation, 200 generations

https://cjbattey.shinyapps.io/driftR/

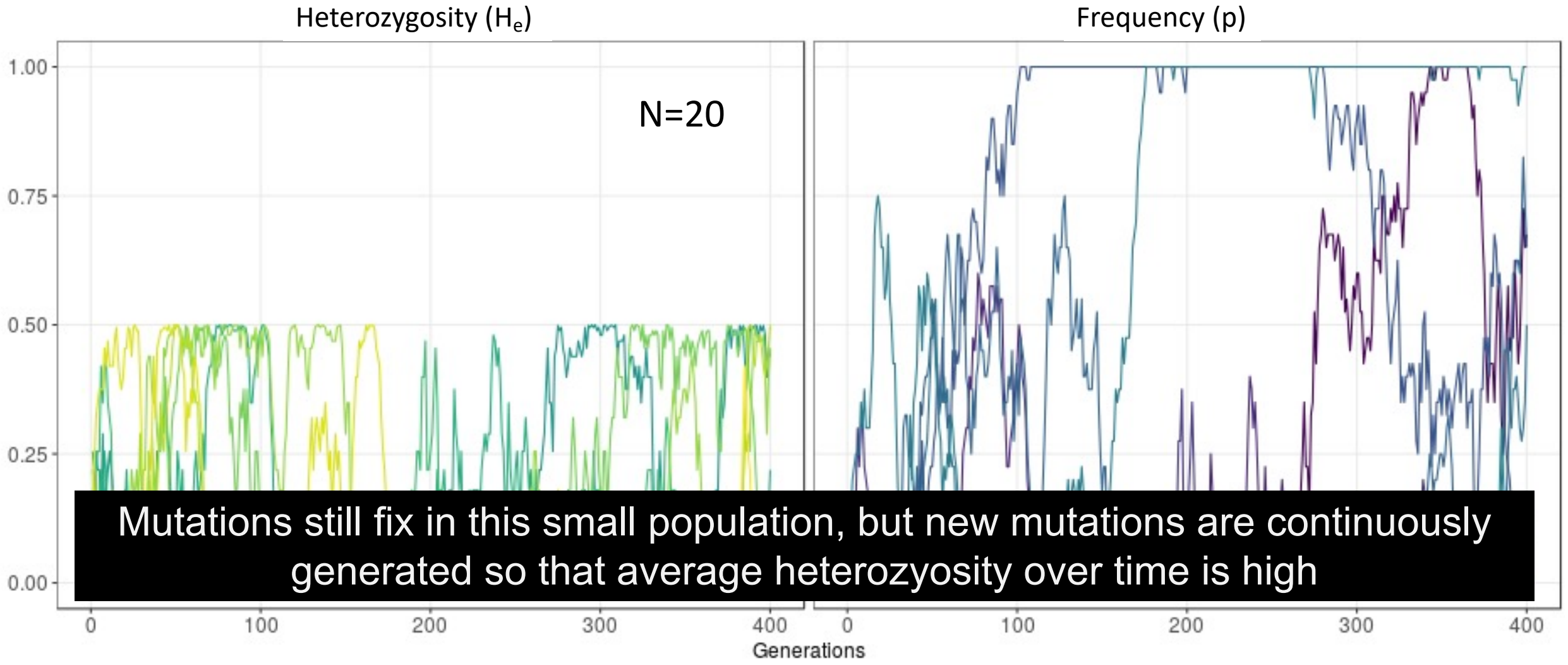
Add mutation. Now there is variation in the population!



Simulations, $p_{t=0}$=0.1, N=20, mutation rate = 1x10^{-3}, 200 generations

https://cjbattey.shinyapps.io/driftR/

Longer time scale shows how continuous mutational input causes high average heterozygosity over time



Mutations still fix in this small population, but new mutations are continuously generated so that average heterozyosity over time is high

Simulations, $p_{t=0}$=0.1, N=20, mutation rate = $1x10^{-3}$, 400 generations

https://cjbattey.shinyapps.io/driftR/

An empirical example: genomes of wolf populations

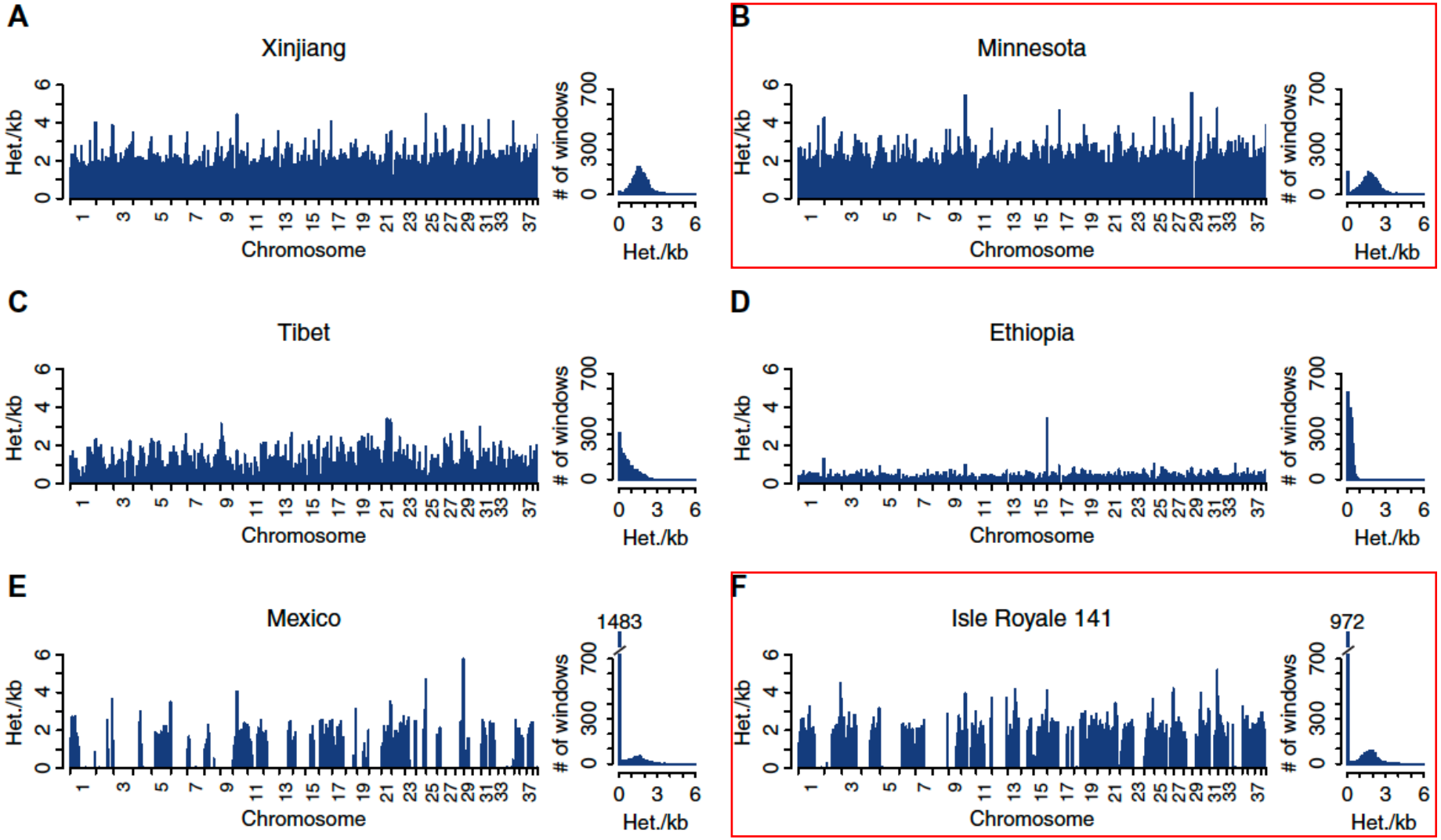
- Isle Royale, an island in Lake Superior, is home to an inbred wolf population
- In this study, the authors sequenced the genomes of Isle Royale wolves and compared them to other populations

Robinson et al., Sci. Adv. 2019; 5 : eaau0757 29 May 2019

Map of samples included in the study

Robinson et al., Sci. Adv. 2019; 5 : eaau0757 29 May 2019

Heterozygosity per kilobase is lower in small isolated populations and more variable in populations with recent inbreeding

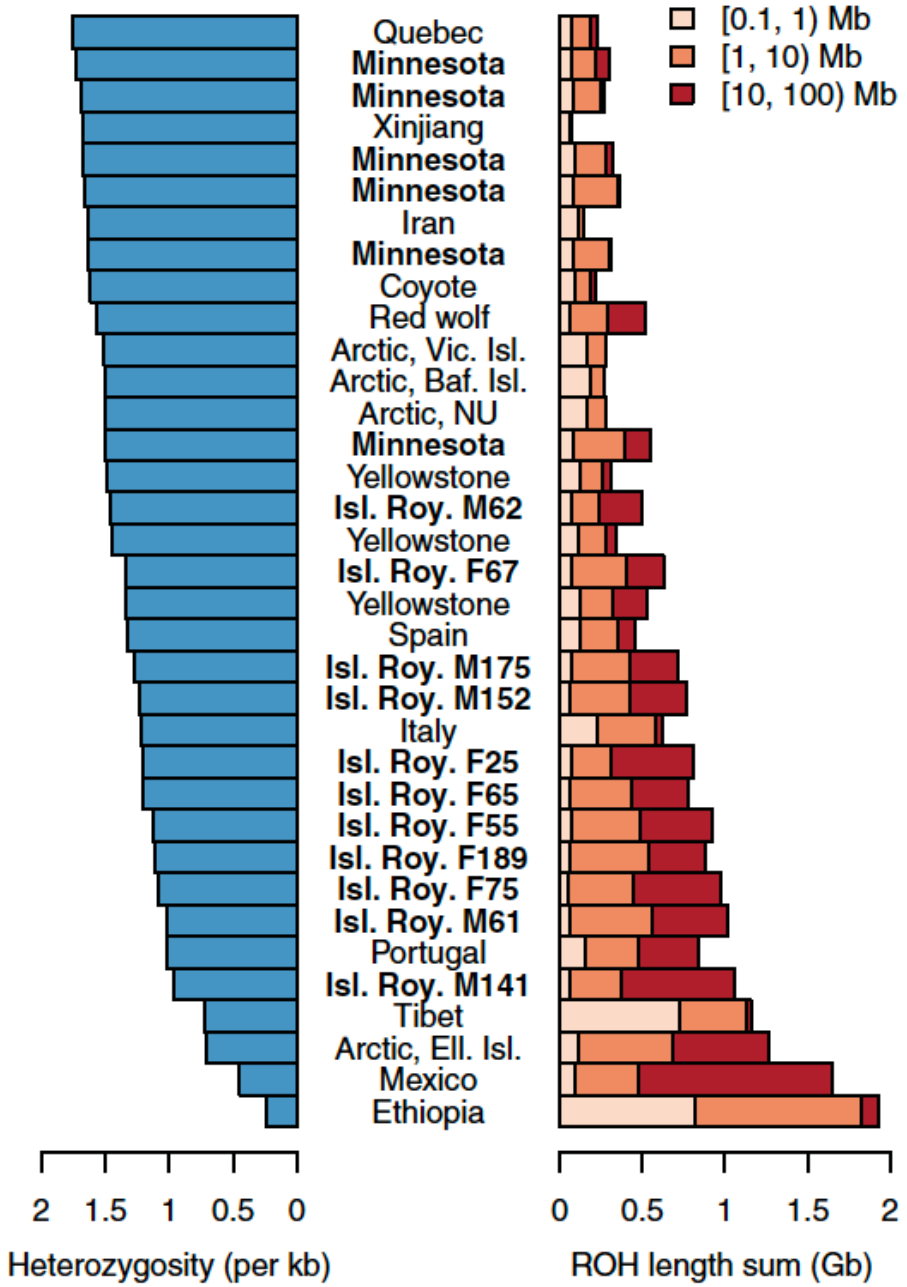


Large, outbred populations

Small, isolated populations

Populations with recent inbreeding

Robinson et al., Sci. Adv. 2019; 5 : eaau0757 29 May 2019



ROH = runs of homozygosity

Heterozygosity differs across samples

…with a tendency toward more long runs of homozygosity in samples with lower heterozygosity

Robinson et al., Sci. Adv. 2019; 5 : eaau0757 29 May 2019

Is low heterozygosity necessarily bad for a population?

- Populations with low heterozygosity have less variation for selection to act on, so they are likely to have lower evolutionary potential
- To the extent that the outbred parental population carries recessive deleterious alleles, inbreeding (reducing heterozygosity) means that those alleles will be exposed

We will discuss consequences of population size and inbreeding in terms of “genetic load” in future lectures

Variation in the predicted effects of heterozygosity across populations

N	H
1,000	0.00004
10,000	0.00040
100,000	0.00400
1,000,000	0.03700

$u = 10^{-8}$

Populations of different sizes should differ enormously in heterozygosity

Is this what is observed?

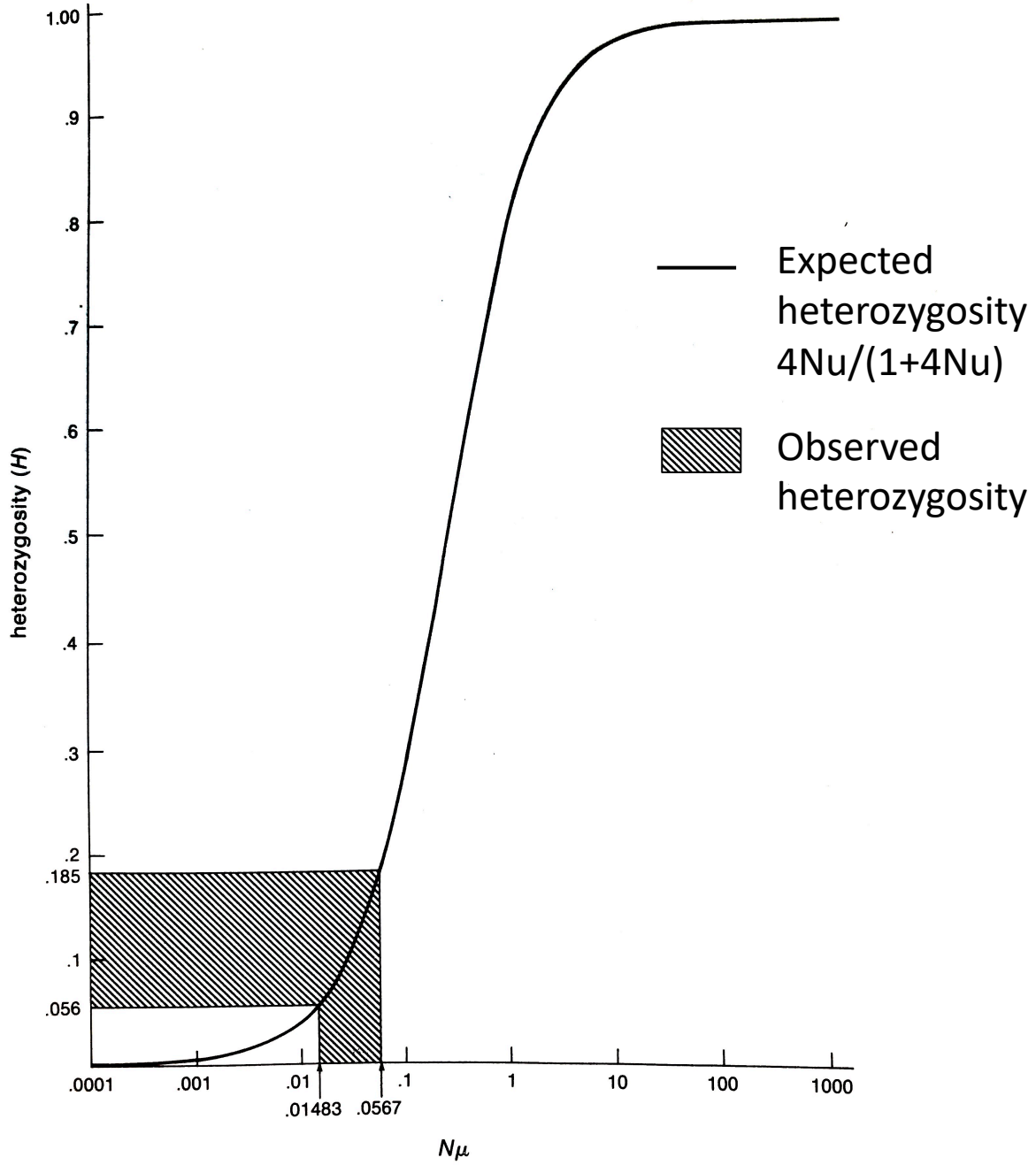
Lewontin's paradox: ***In natural populations, heterozygosity tends to be lower than expected***

- Heterozygosity should vary across populations as a function of mutation rate and population size ($2N\mu$)
- However, Lewontin observed that variation at allozymes is lower than expected

Why is heterozygosity so low?

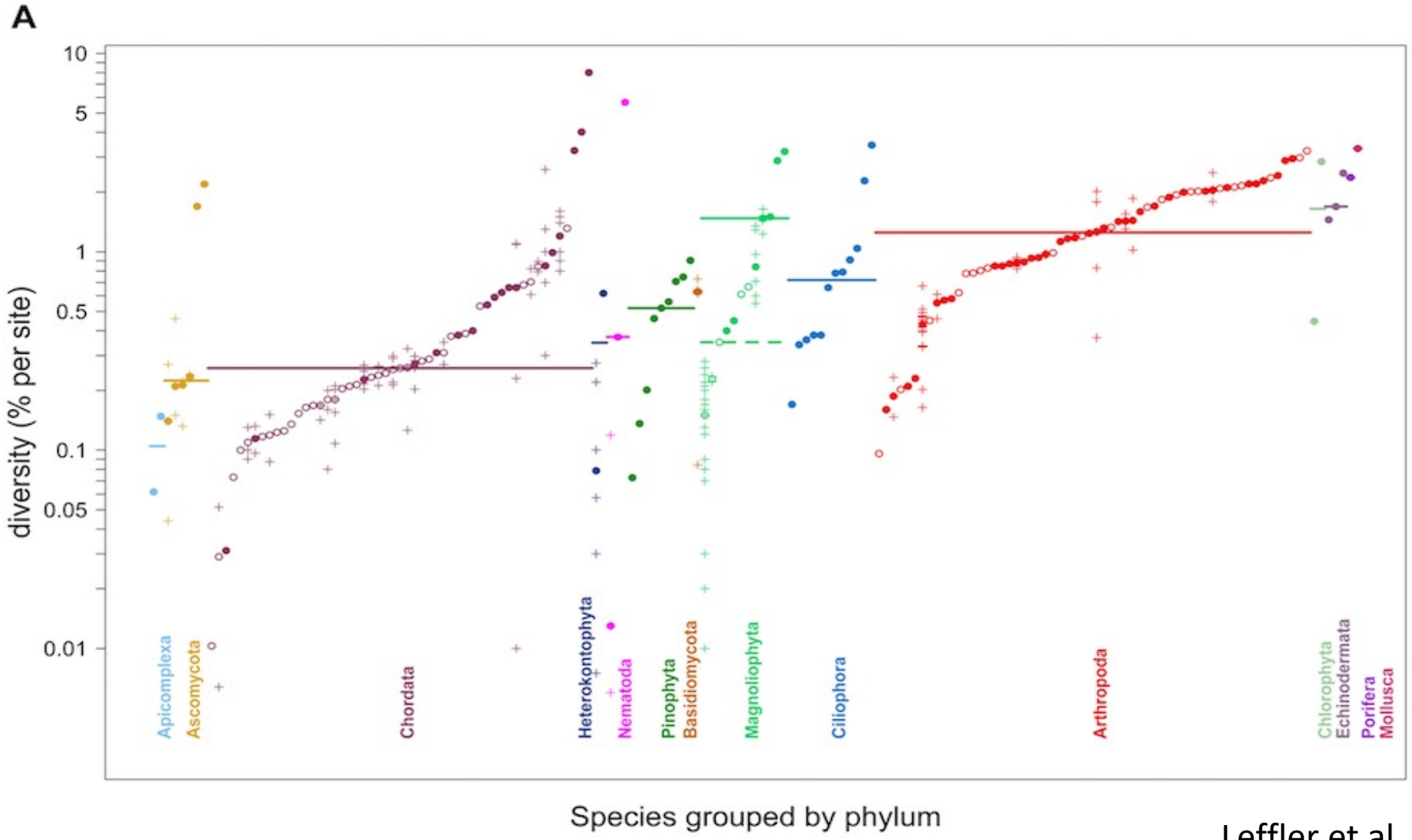
“Lewontin’s paradox”:

DNA variation differs among species far less than expected based on their census population sizes



Lewontin, The Genetic Basis of Evolutionary Change, 1974

Diversity varies widely across species



Leffler et al., Plos Biology, 2012

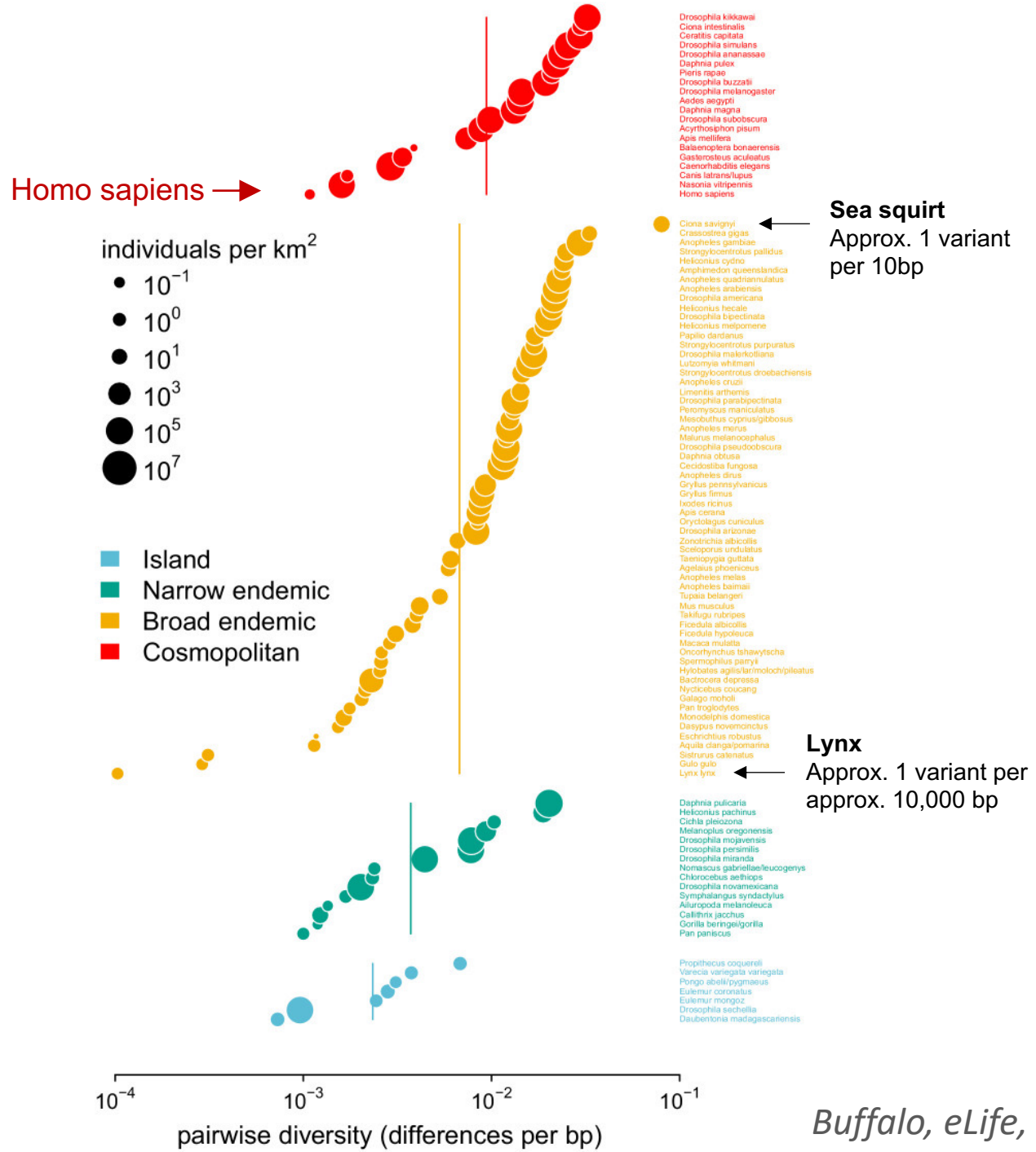
Does variation in census size (N) explain variation in diversity?

Buffalo, eLife, 2021

Not completely.

Variation in N explain only some of the variation in diversity

What other explanations could there be?



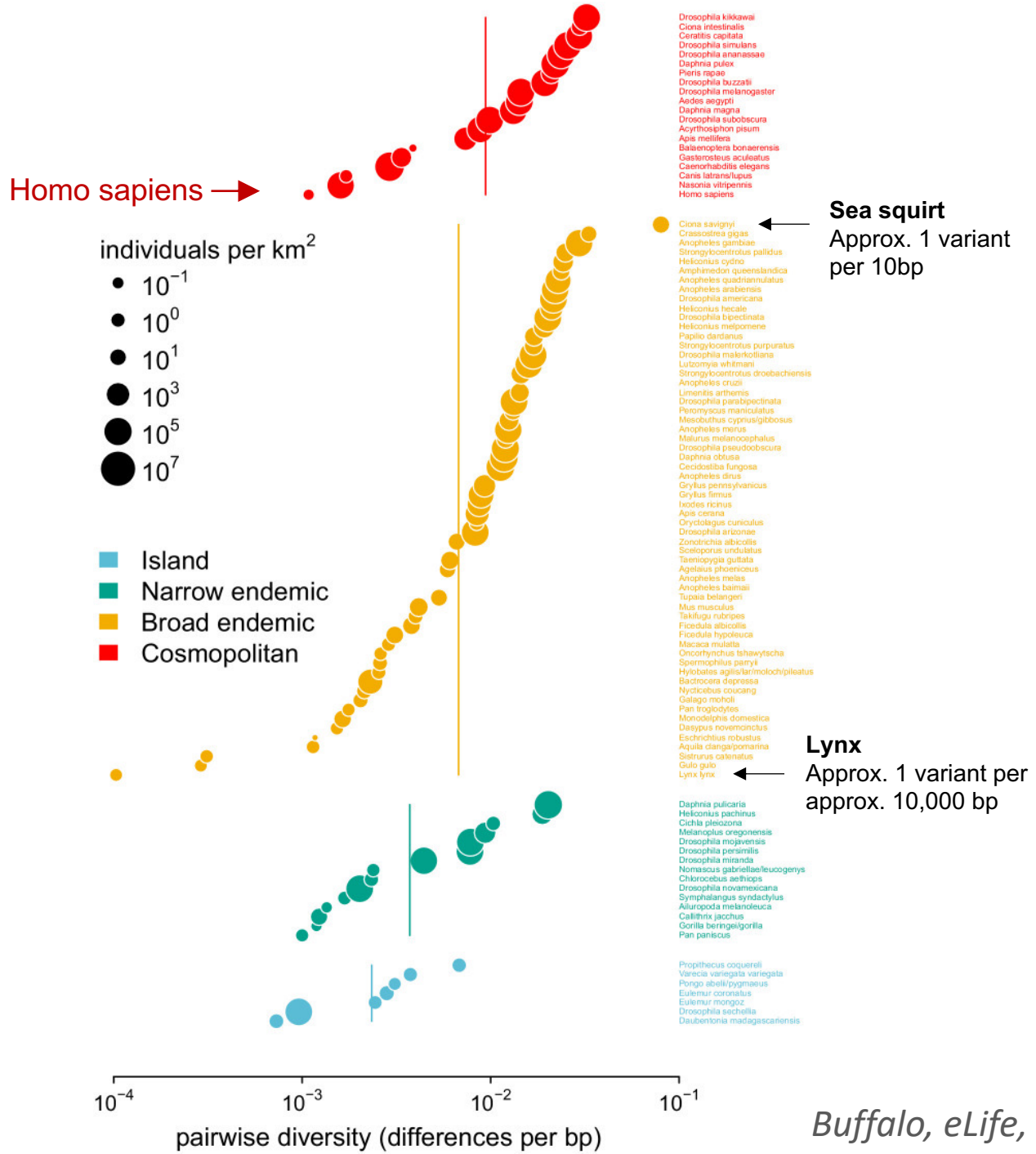
Buffalo, eLife, 2021

Not completely.

Variation in N explain only some of the variation in diversity

What other explanations could there be?

- Population bottlenecks
- Selection



Buffalo, eLife, 2021

Summary

- Genetic drift removes variation from the population at a rate of of $\frac{1}{2N}$
- Mutation adds variation into the population at a rate of 2Nμ
- Measured levels of variation in real populations are often lower than expected based on population size (N)
- What other forces might remove variation from populations?