

# Evolutionary Thinking 2022

## TA session

### week 5 – Basis of Population Genetics

Jilong Ma  
aujilongm@birc.au.dk

# Outline

---

## 1. Learning outcome of this week

Terminology (Wed)

Hardy-Weinberg Equilibrium (Wed)

Genetic Drift, Mutation (Wed, Fri)

Coalescence Theory (Fri)

– Process, Tree and Tree length, Site Frequency Spectrum

## 2. Exercises

# Terminology

- **Locus:** a specific “spot” in the genome (could be a single base-pair or an entire gene or region, depending on context)



Slides from Fernando Racimo  
“Intro to popgen”

# Terminology

- **Locus:** a specific “spot” in the genome (could be a single base-pair or an entire gene or region, depending on context)



Slides from Fernando Racimo  
“Intro to popgen”

# Terminology

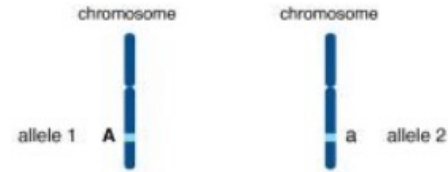
- **Locus:** a specific “spot” in the genome (could be a single base-pair or an entire gene or region, depending on context)
- **Allele:** 1 of the alternative forms of a locus that exist in a population



Slides from Fernando Racimo  
“Intro to popgen”

# Terminology

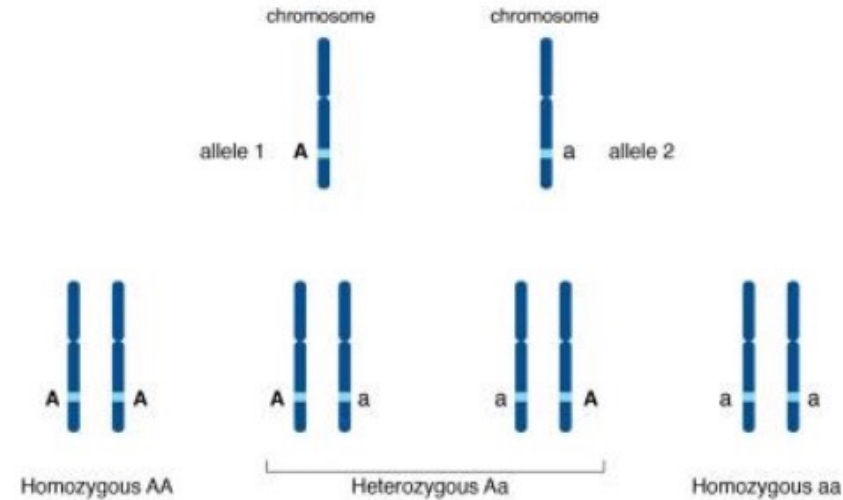
- **Locus:** a specific “spot” in the genome (could be a single base-pair or an entire gene or region, depending on context)
- **Allele:** 1 of the alternative forms of a locus that exist in a population
- **Polymorphism:** a locus with 2 or more alleles segregating in a population



Slides from Fernando Racimo  
“Intro to popgen”

# Terminology

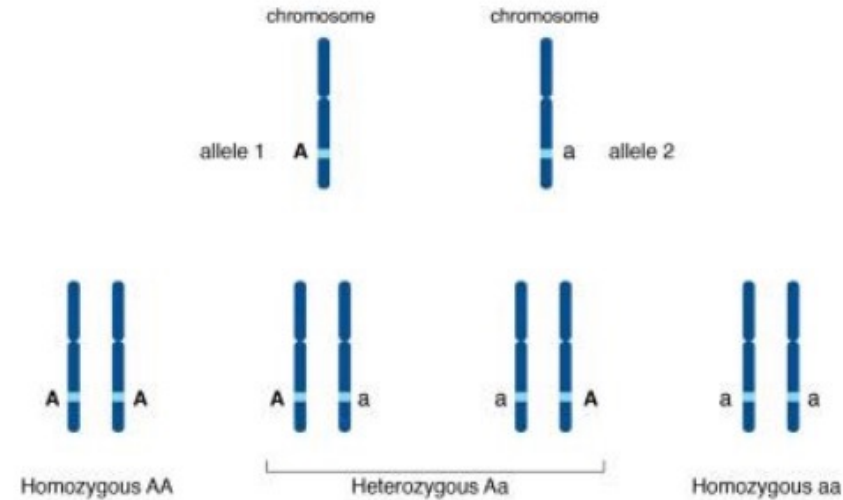
- **Locus:** a specific “spot” in the genome (could be a single base-pair or an entire gene or region, depending on context)
- **Allele:** 1 of the alternative forms of a locus that exist in a population
- **Polymorphism:** a locus with 2 or more alleles segregating in a population
- **Genotype:** the set of alleles present at a locus in a particular organism ( set size = 2 if the organism is diploid)



Slides from Fernando Racimo  
“Intro to popgen”

# Terminology

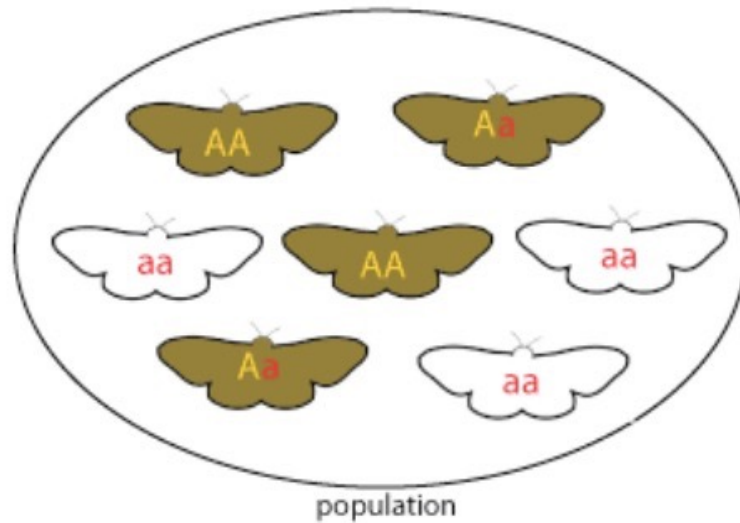
- **Locus:** a specific “spot” in the genome (could be a single base-pair or an entire gene or region, depending on context)
- **Allele:** 1 of the alternative forms of a locus that exist in a population
- **Polymorphism:** a locus with 2 or more alleles segregating in a population
- **Genotype:** the set of alleles present at a locus in a particular organism ( set size = 2 if the organism is diploid)



Slides from Fernando Racimo  
“Intro to popgen”



# Terminology



## Allele frequencies

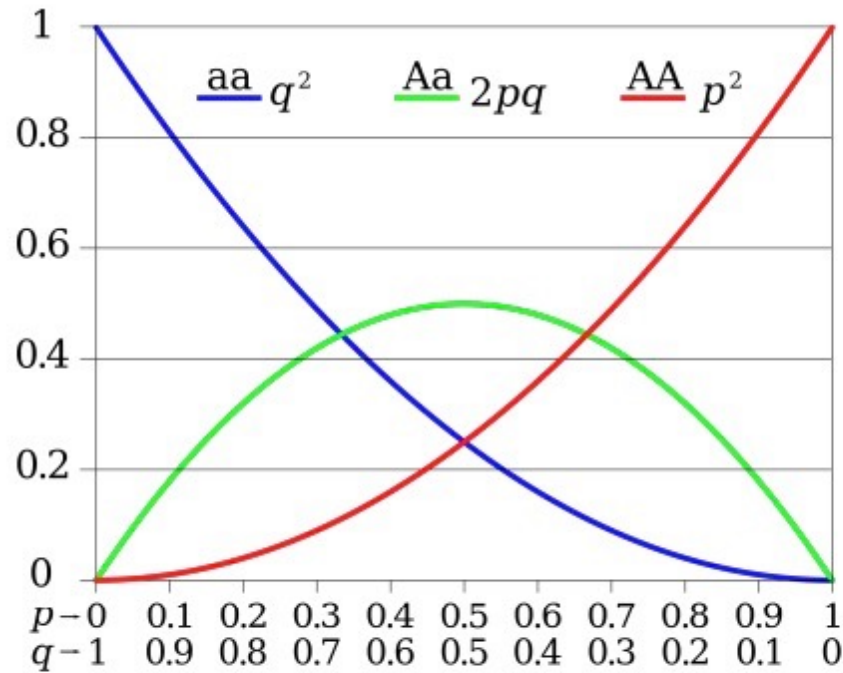
- $f(A) = 6/14$
- $f(a) = 8/14$

## Genotype frequencies

- $f(AA) = 2/7$
- $f(Aa) = 2/7$
- $f(aa) = 3/7$

Slides from Fernando Racimo  
“Intro to popgen”

# HW Equilibrium



Assumptions?

Allele frequencies

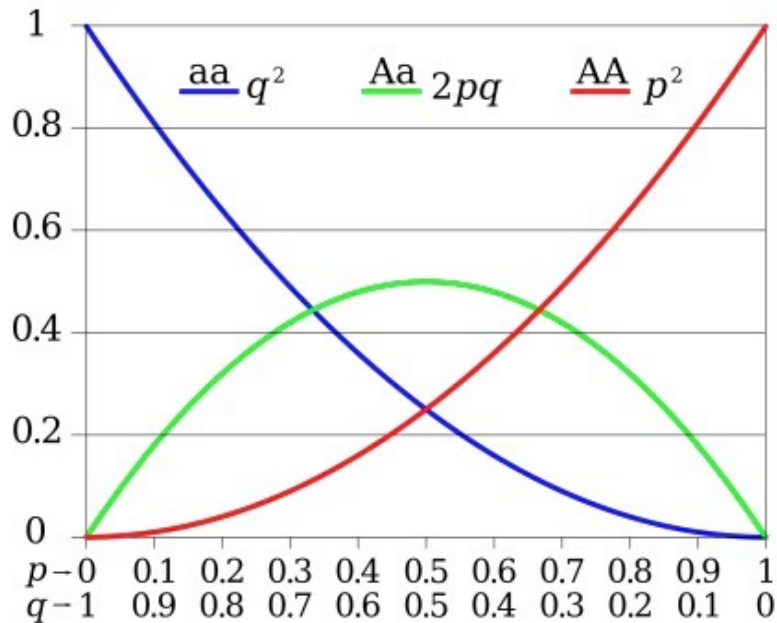


Genotype frequencies

Slides from Fernando Racimo  
“Intro to popgen”


# HW Equilibrium

Assumptions?



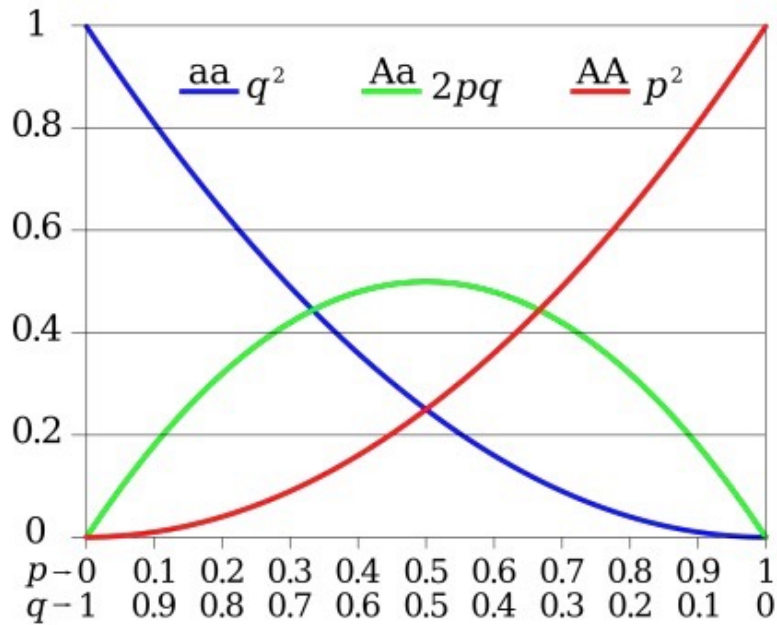
## Assuming:

- Infinite population size
- No migration
- No mutation
- No selection
- Random mating
- Non-overlapping generations

Allele frequencies  Genotype frequencies


Slides from Fernando Racimo  
“Intro to popgen”

# HW Equilibrium



## Assuming:

- Infinite population size
- No migration
- No mutation
- No selection
- Random mating
- Non-overlapping generations

Allele frequencies  Genotype frequencies

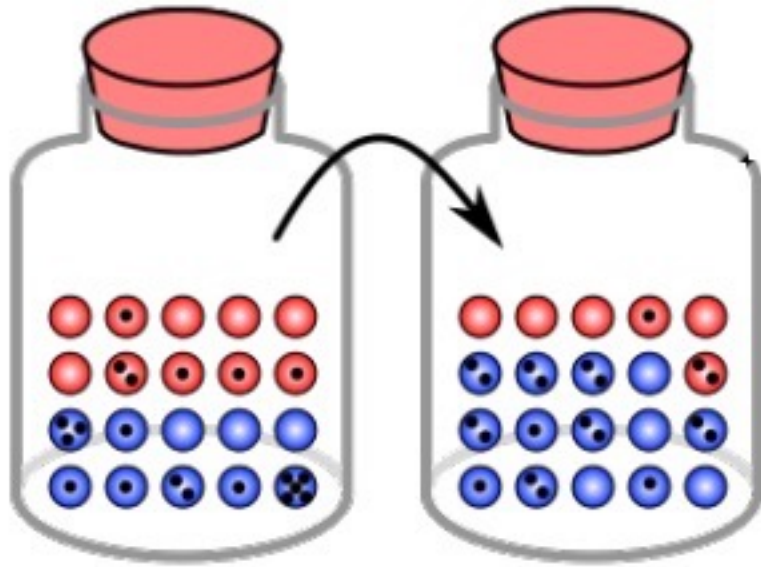
Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model



Slides from Fernando Racimo  
“Intro to popgen”

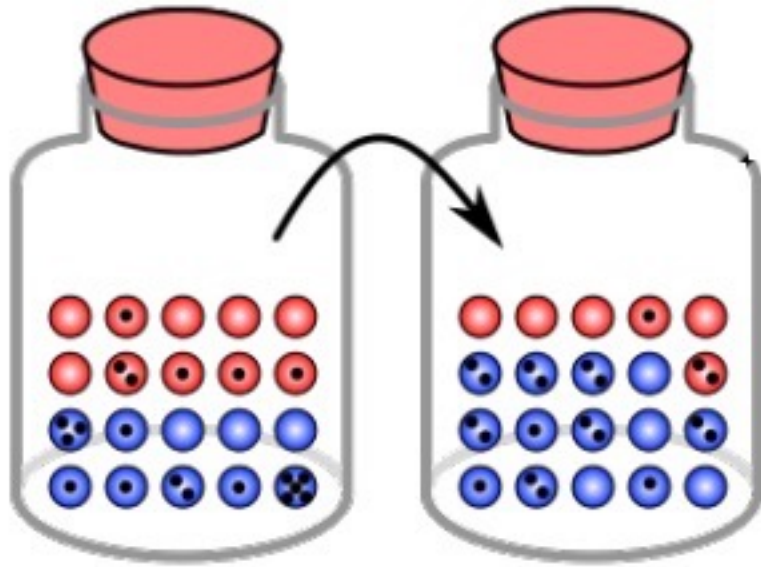
# Wright-Fisher Model



20 marbles 20 marbles

Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model

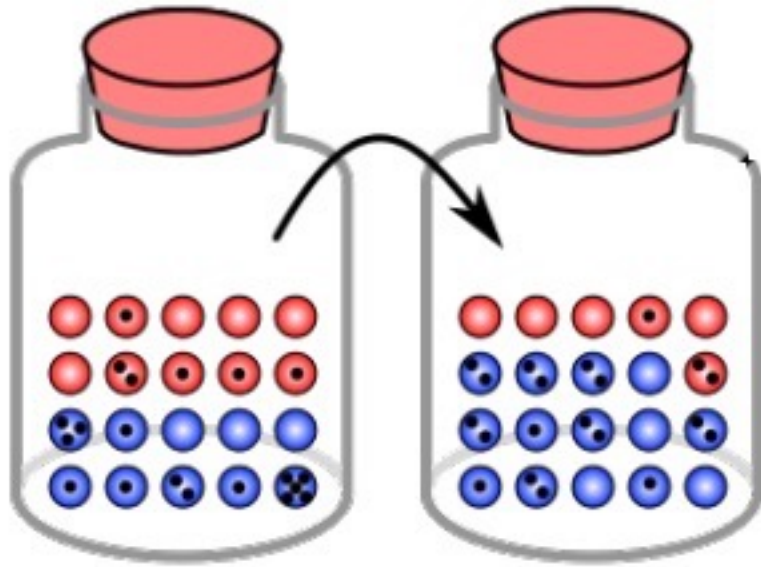


20 marbles   20 marbles

- Sample **with replacement** to fill the next jar

Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model



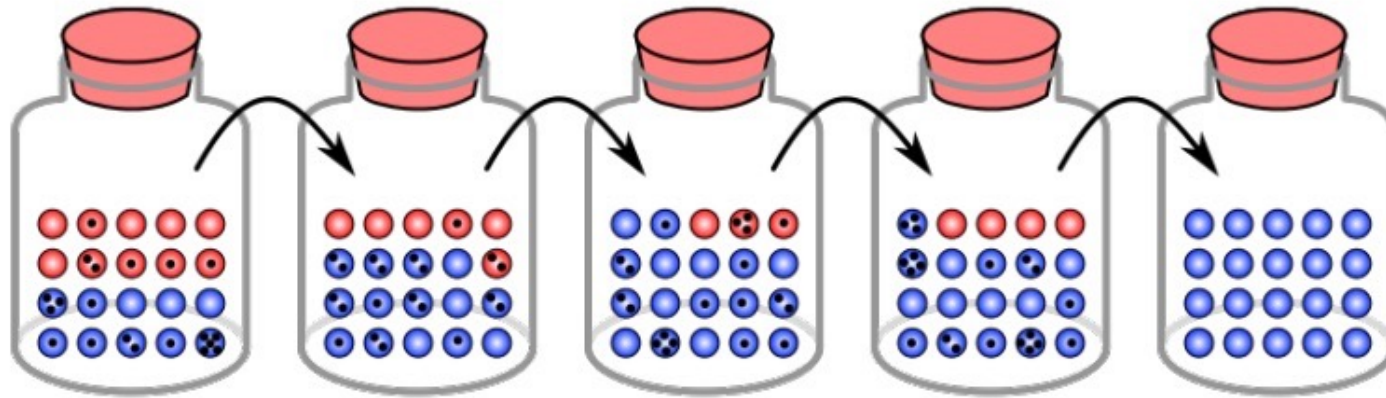
20 marbles 20 marbles

- Sample **with replacement** to fill the next jar
- The total number of marbles in each jar is the same ( $N = 20$ )

Slides from Fernando Racimo  
“Intro to popgen”



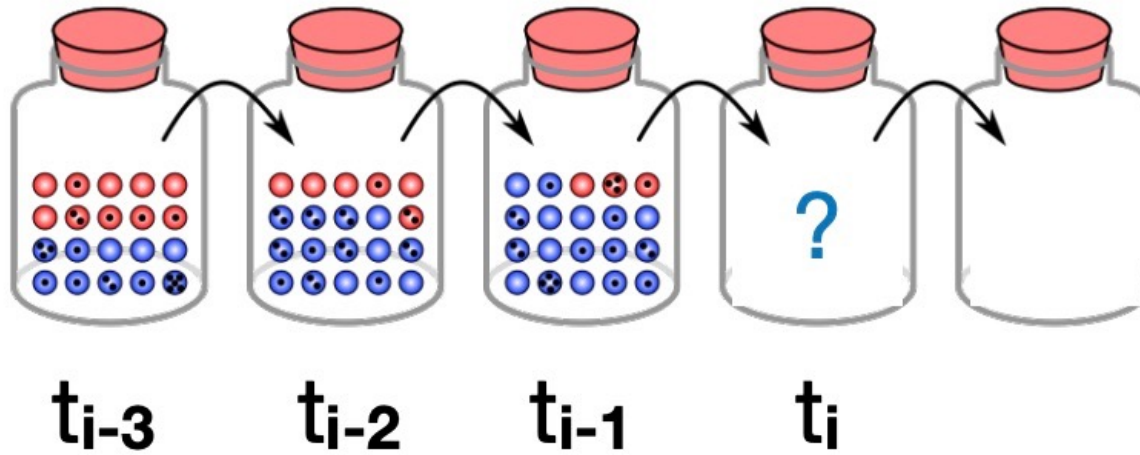
# Wright-Fisher Model



Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model

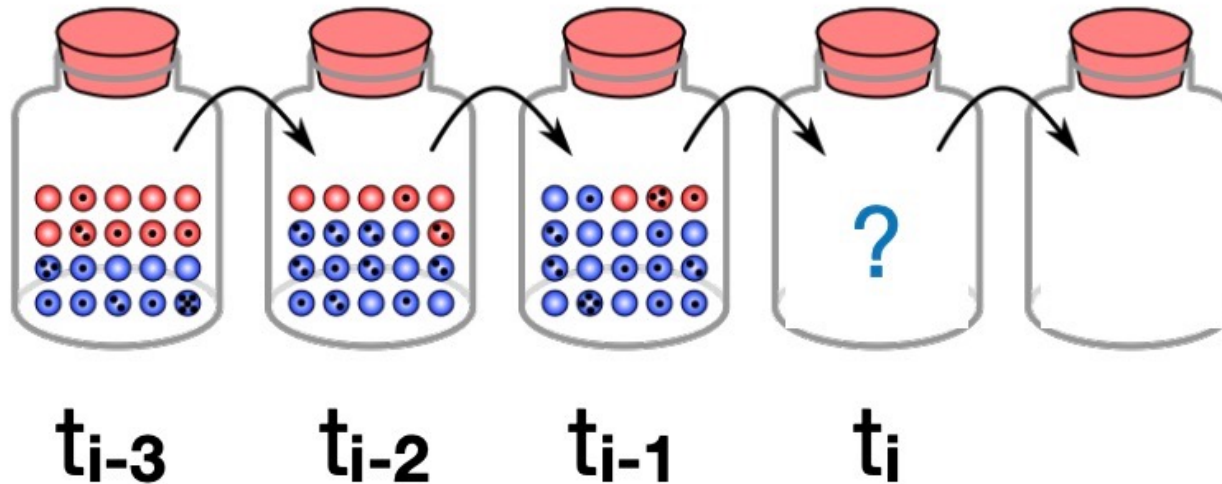
- Let's say we're interested in predicting the number of **blue** marbles at **time  $t_i$**



Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model

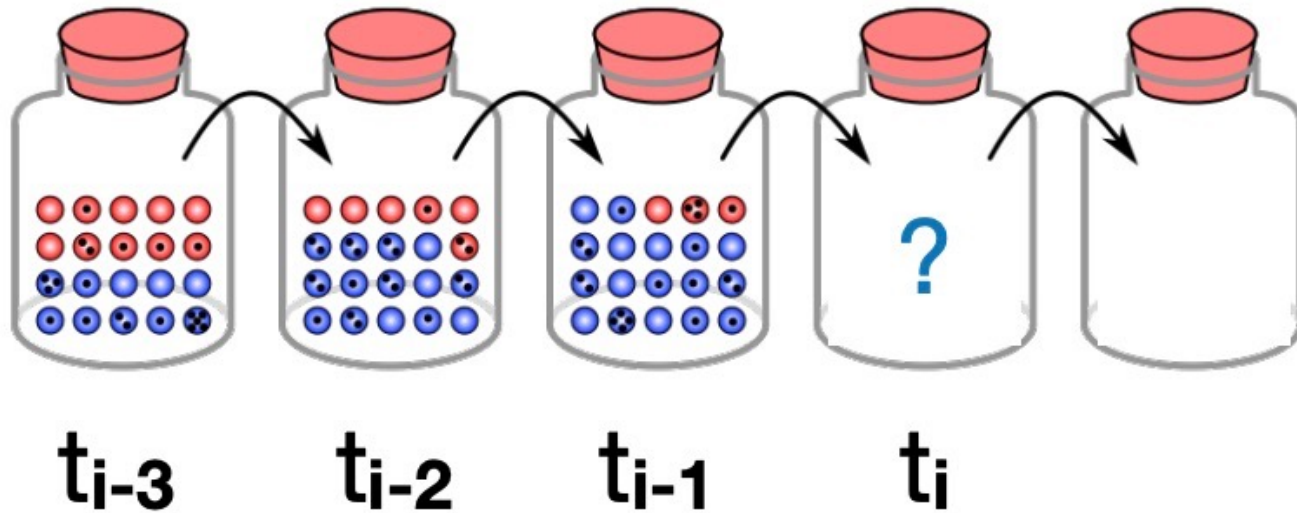
- Let  $f(t_i)$  be the frequency of blue marbles at time  $t_i$
- $P[\text{\# blue marbles} = k \mid f(t_{i-1}), f(t_{i-1}), f(t_{i-1}), \dots] = P[\text{\# blue marbles} = k \mid f(t_{i-1})]$



Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model

$$P[\text{\# blue marbles} = k \mid f(t_{i-1})] = \binom{N}{k} f(t_{i-1})^k (1 - f(t_{i-1}))^{N-k}$$



Slides from Fernando Racimo  
“Intro to popgen”

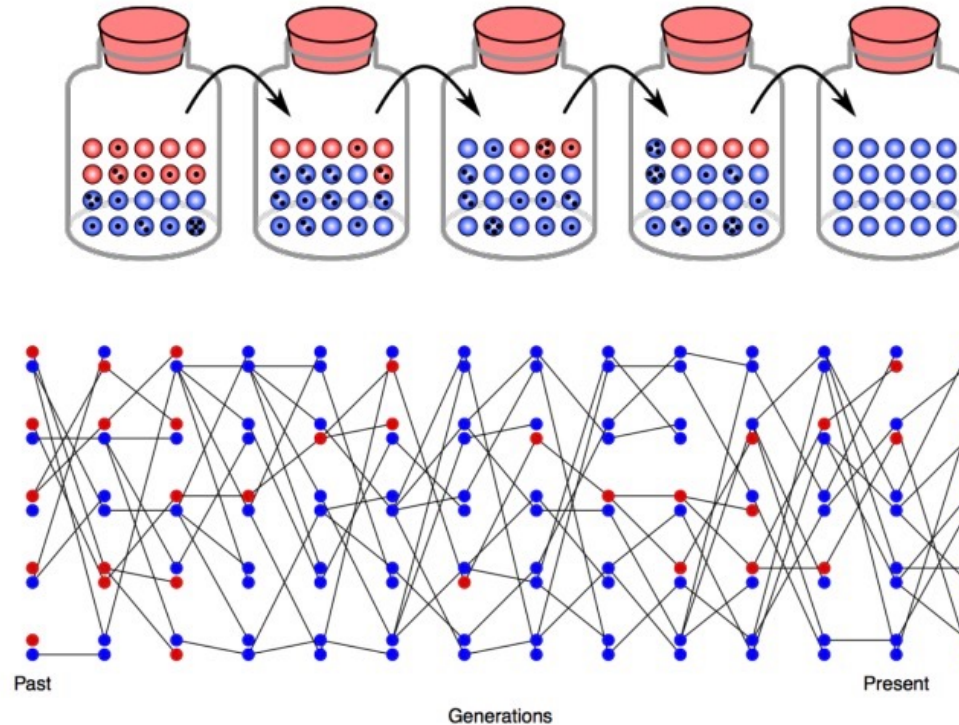
# Wright-Fisher Model

blue marbles =  
blue alleles

red marbles =  
red alleles

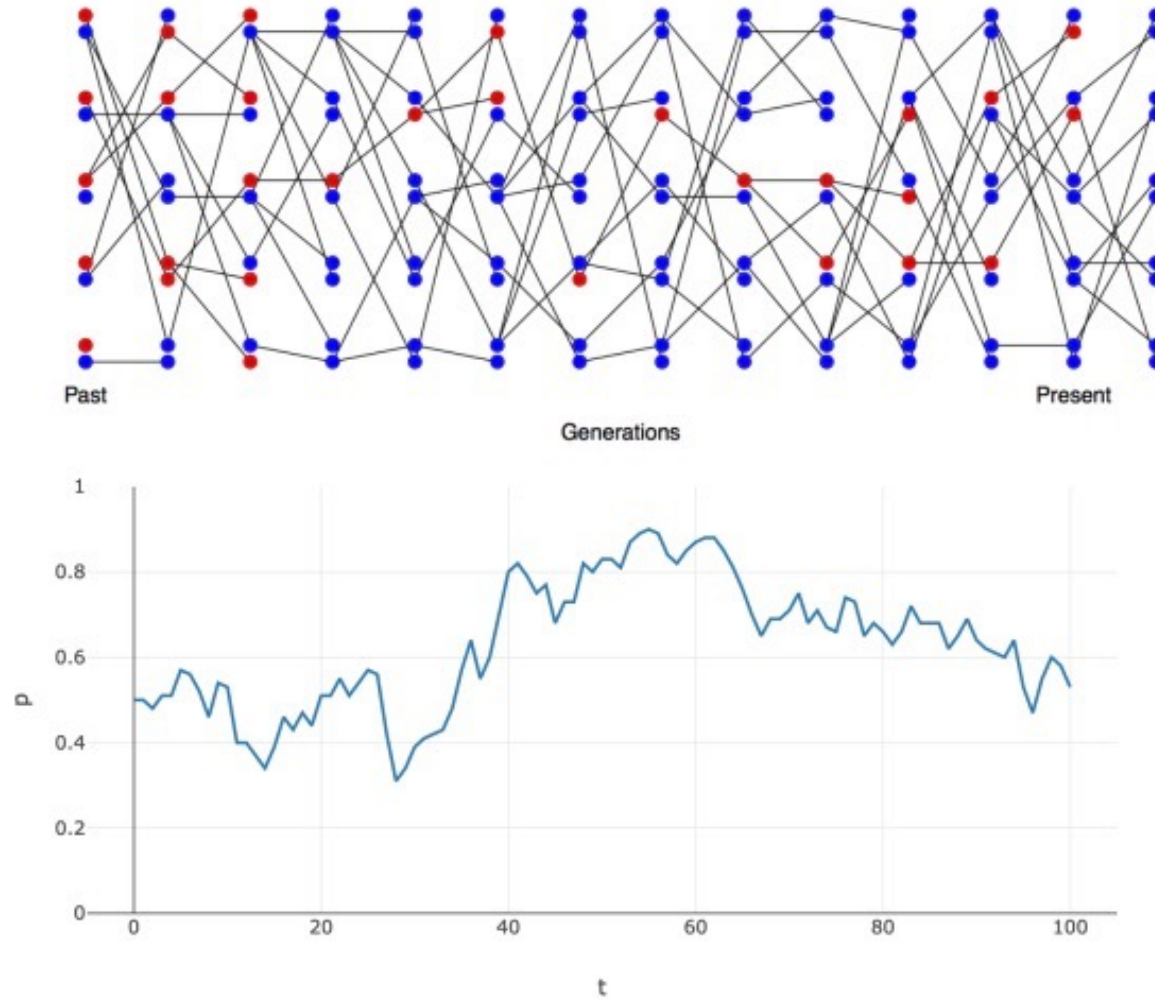
**N** total marbles =  
population size **N**

Jar  $t_i$  =  
Generation  $t_i$



Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model



Slides from Fernando Racimo  
“Intro to popgen”



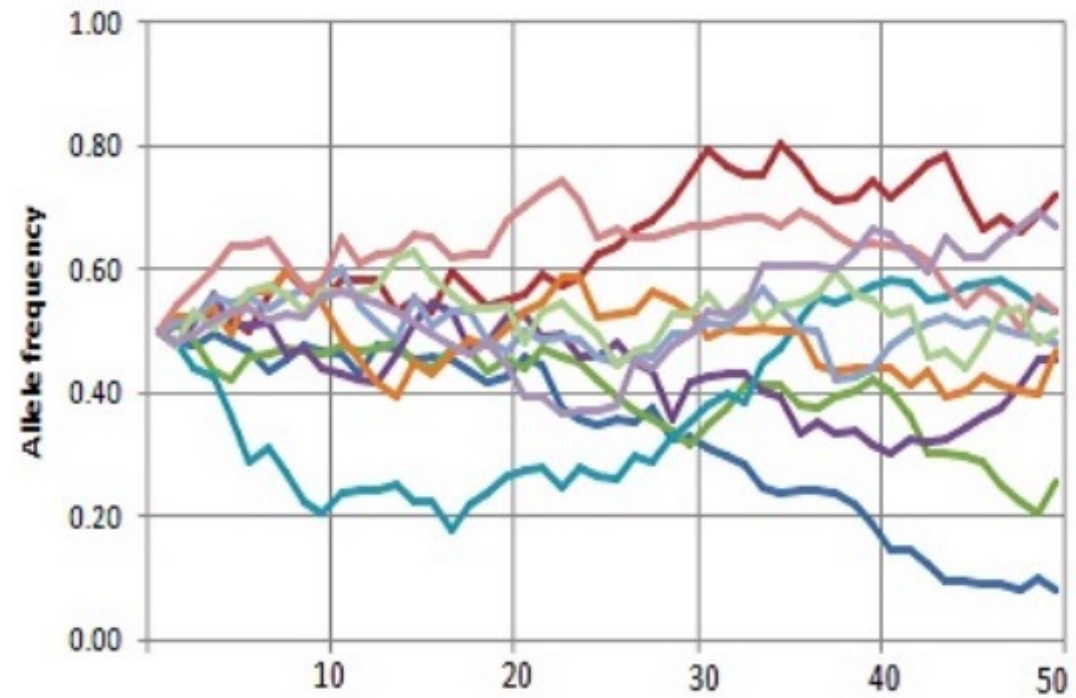
# Assumptions of the WF Model

- Constant population size
  - Individuals reproduce asexually and randomly
  - No migration
  - No selection
  - No mutation
  - Non-overlapping generations
- 
- **N** if we are studying a population of N haploid organisms.
  - **2N** if we are studying a population of N diploid organisms (which we treat as if they were equivalent to a haploid population twice its size)

What assumption of the Hardy-Weinberg model are we now relaxing?

Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model

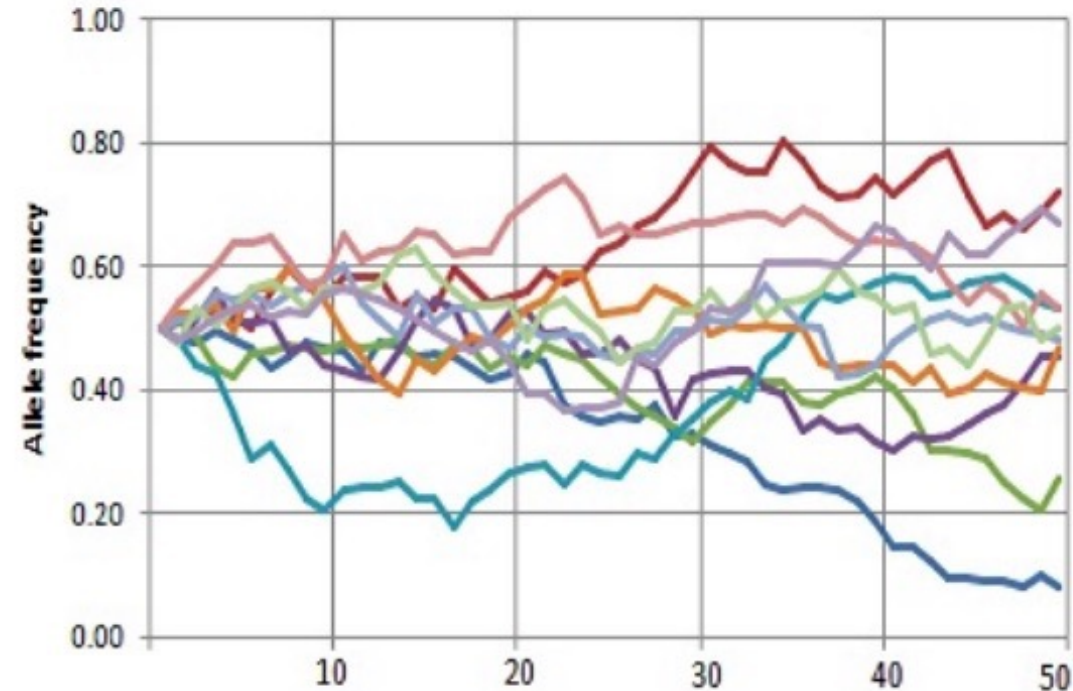


Slides from Fernando Racimo  
“Intro to popgen”



# Wright-Fisher Model and Genetic Drift

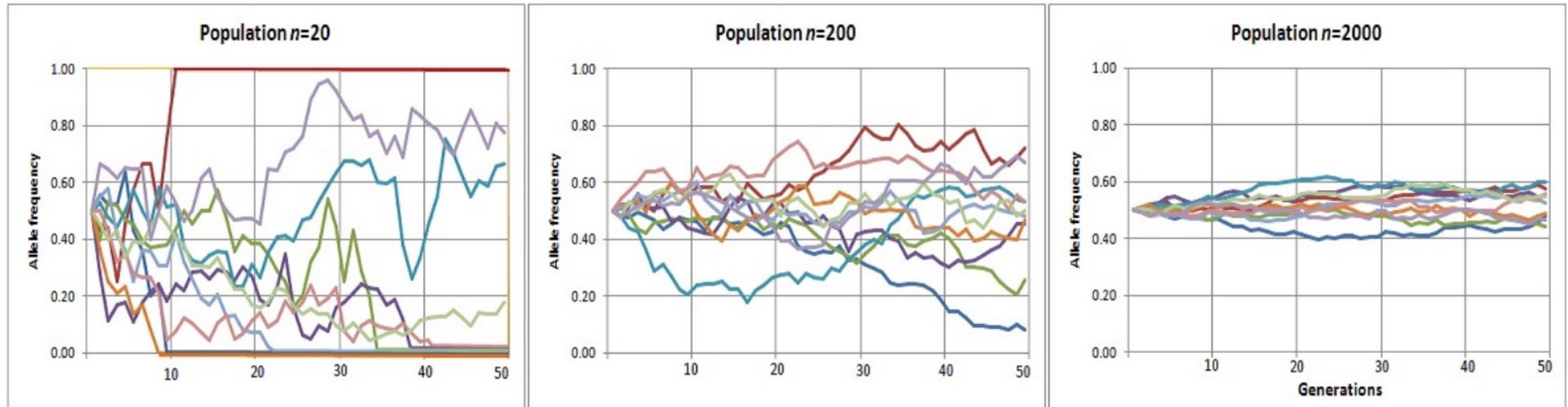
- The Wright-Fisher model is a stochastic process
- If you start from the same allele frequency, you might end up at a different frequency **purely by chance**.
- **Genetic drift** is the change in allele frequencies over time due to **random sampling**.
- No allele has any special advantage over the others.



Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model and Genetic Drift

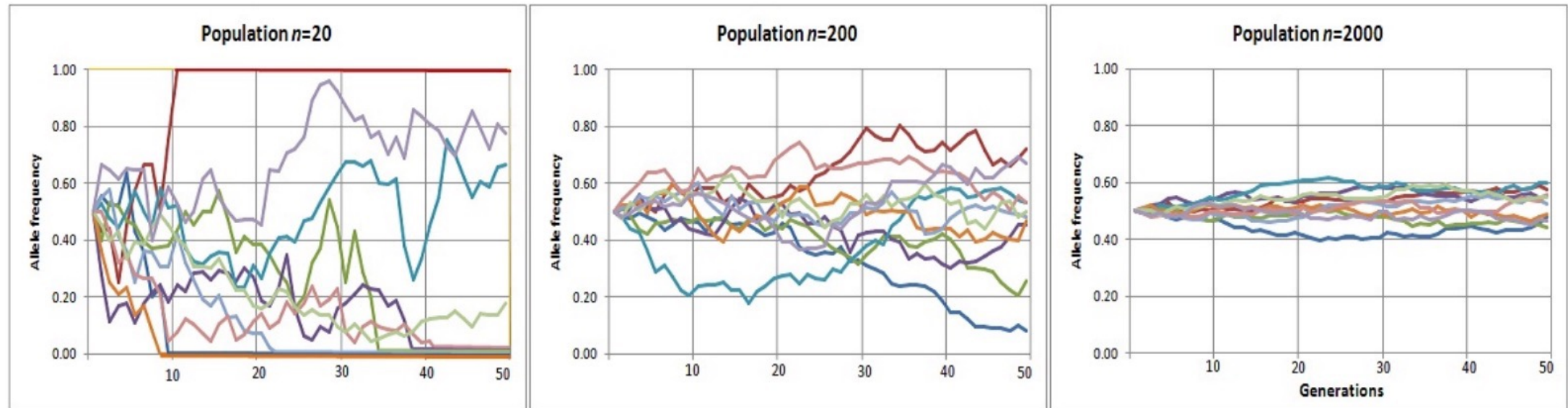
## Comments on the figure



Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model and Genetic Drift

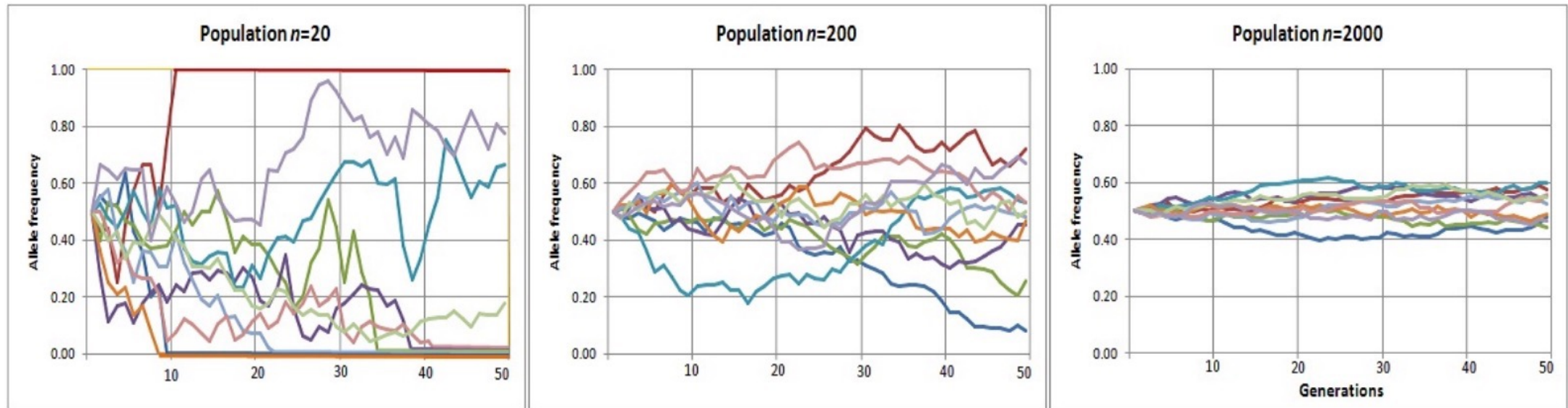
The expected allele frequency is the same at each generation, and equal to the starting frequency  $p$ .



Slides from Fernando Racimo  
“Intro to popgen”

# Wright-Fisher Model and Genetic Drift

- When an allele reaches a frequency of 1, we say it has **fixed**.
- When an allele reaches a frequency of 0, we say it has **gone extinct**.
- **Assuming no recurrent mutation**, alleles that fix or go extinct remain so forever
- All alleles must eventually fix or go extinct, given enough time.



Slides from Fernando Racimo  
“Intro to popgen”

# Exercises

---

Chapter 1:

1.1-1.4

Chapter 2:

2.1-2.3

2.7-2.9

Chapter 3:

3.1-3.8



AARHUS  
UNIVERSITY