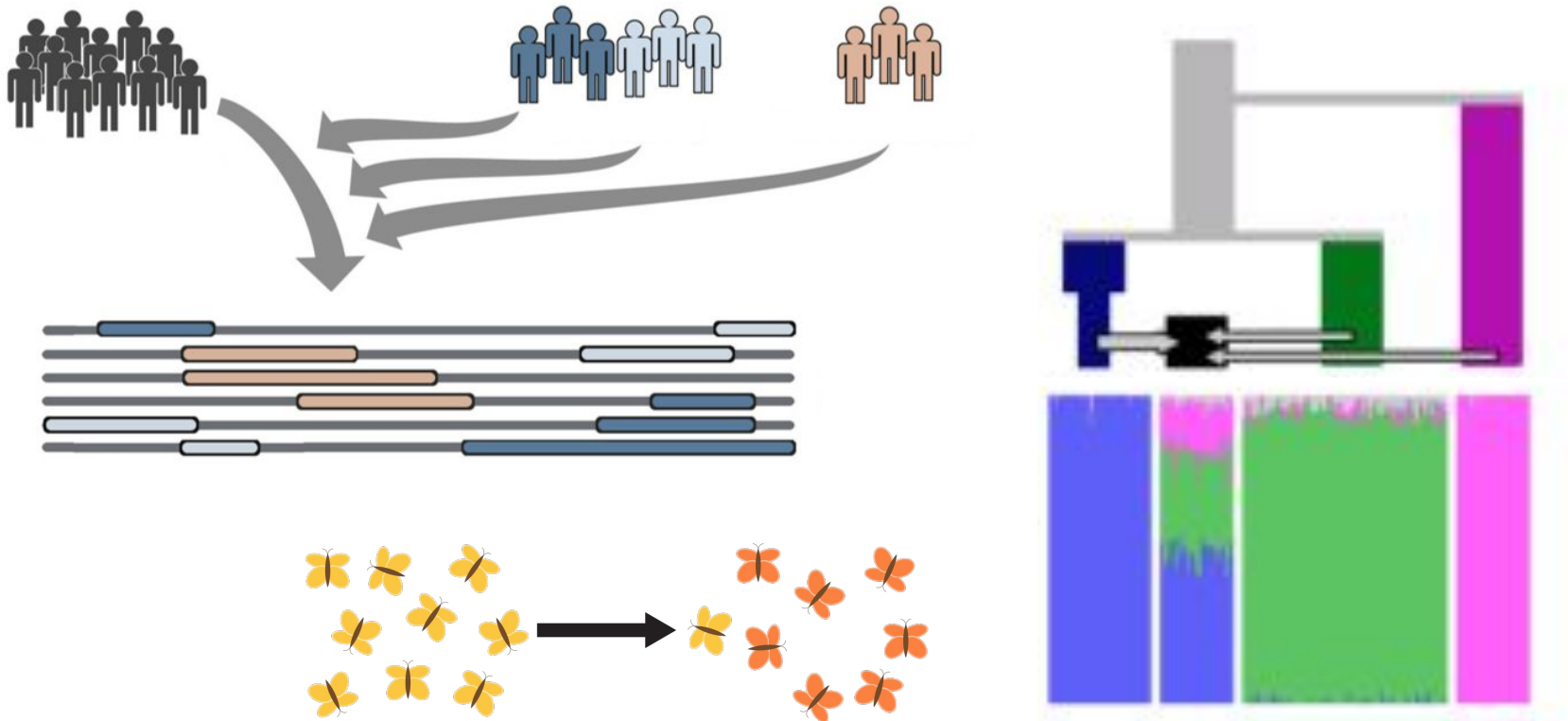# Module 4: Population genomics of admixture and introgression

Shyamalika Gopalan, PhD

AGAR Workshop, July 28th 2022

Outline

Part 1:
Population genetics and evolutionary forces (~1 hr)

Part 2:
Gene flow (~2 hrs)

** We will have two breaks **

To prepare:

Check out the github directory for this module -
https://github.com/alanrogers/agar22/tree/main/introgression

Download all necessary materials
- Exercise document
- All .R and .py scripts

Install all necessary software
- R and RStudio
- python3 (conda)
- msprime (conda)
- plink (conda)
- admixture (conda)
- rfmix (conda)

Part 1 objectives

- Understand genetic drift and its effects on patterns of neutral diversity

- Understand how demographic parameters impact drift and population differentiation

- Predict the effect of a parameter change on F_{ST} between two populations

Population genetics examines patterns of **genetic variation** within and between populations to understand how they are shaped over time by **fundamental mechanisms of evolution**

Populations evolve by four main mechanisms:

1. Mutation

2. Selection

3. Drift

4. Migration

Populations evolve by four main mechanisms:

1. Mutation - generates new variation

2. Selection

3. Drift

4. Migration

Populations evolve by four main mechanisms:

1. Mutation - generates new variation

2. Selection - directional change in frequency

3. Drift

4. Migration

Populations evolve by four main mechanisms:

1. Mutation - generates new variation

2. Selection - directional change in frequency

3. Drift - random change in frequency

4. Migration

Populations evolve by four main mechanisms:

1. Mutation - generates new variation

2. Selection - directional change in frequency

3. Drift - random change in frequency

4. Migration - introduces new variation

Mutation

- Generates **new variation** within a population

- Generally modelled by a **single fixed rate**

AGAGTGCGGTCCGGGGCGCC
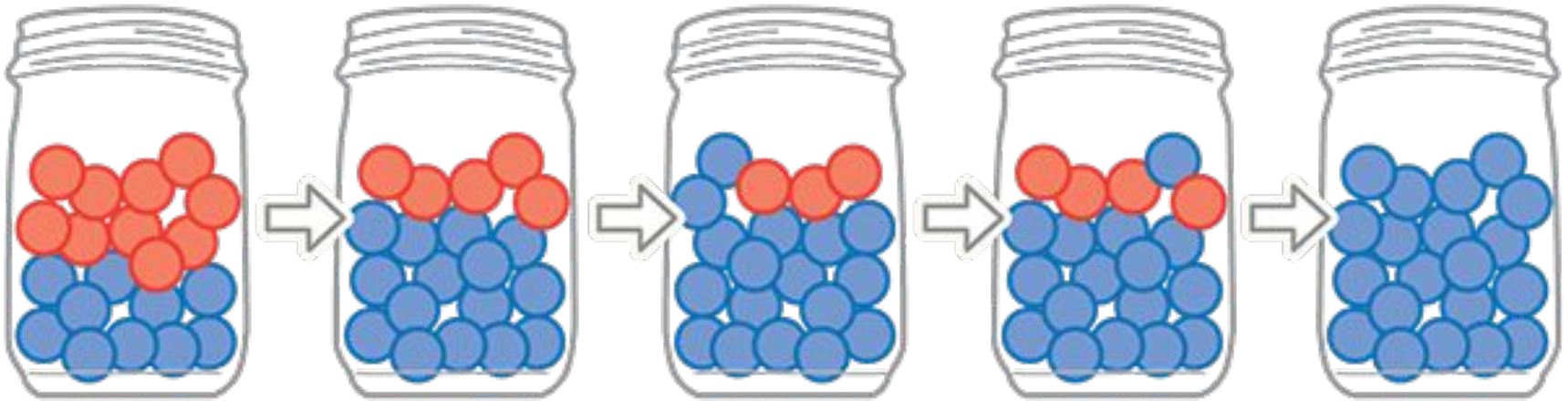
↓

AGAG**C**GCGGTCCGGGGCGCC

↓

AGAG**C**GCGGTCCGGG**A**CGCC

Selection

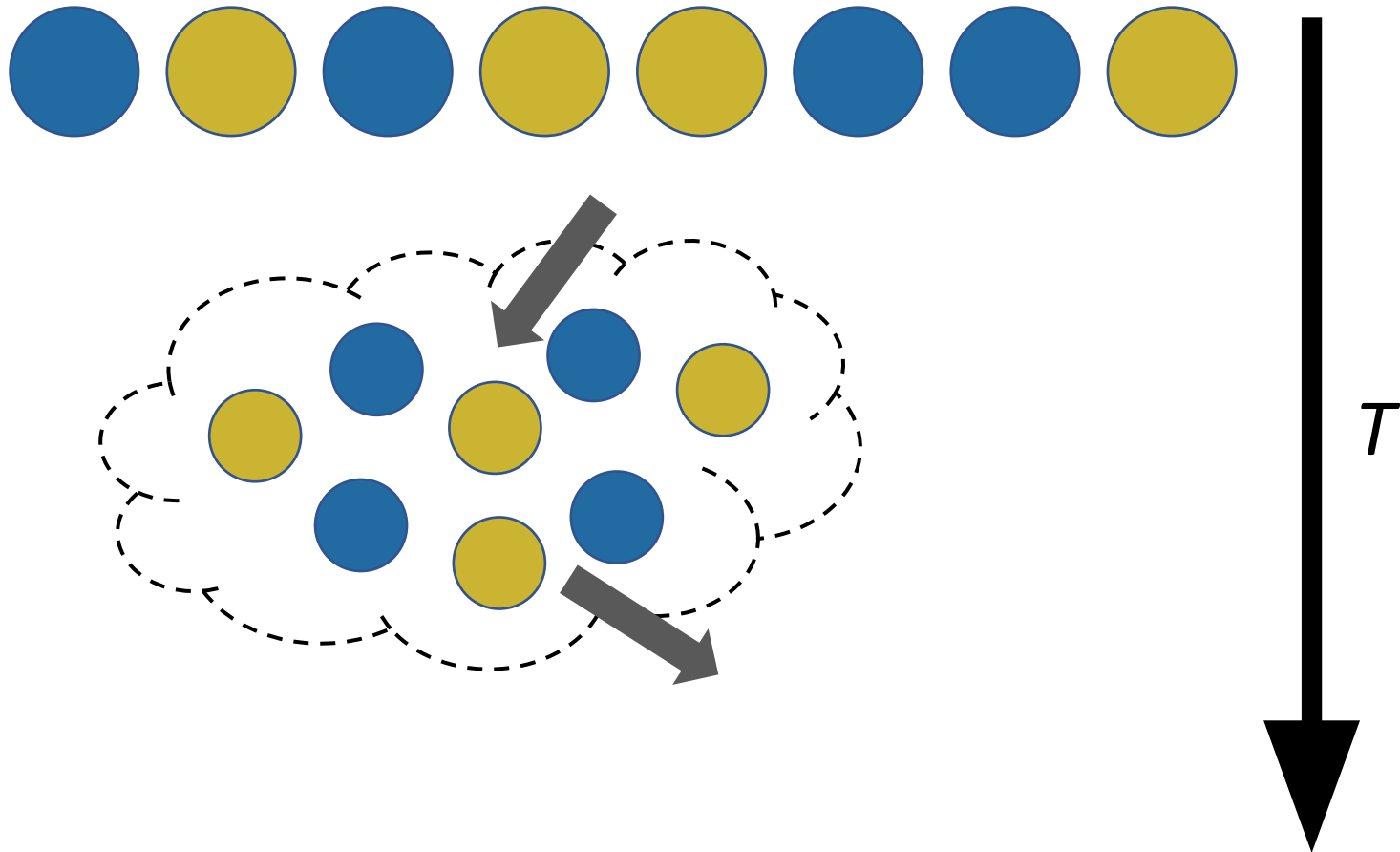- Favours **certain alleles** over others, causing them to **increase in frequency**
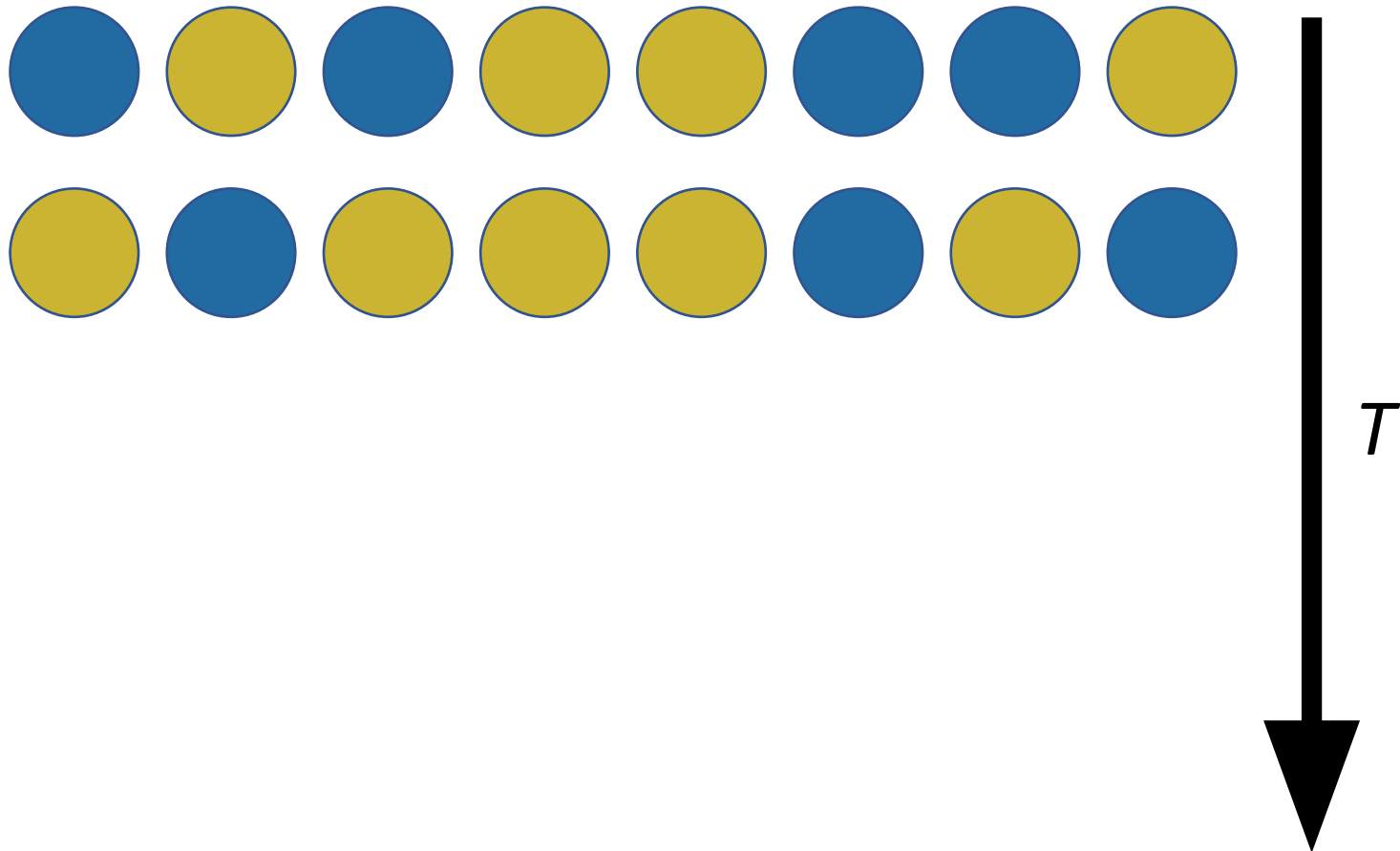
Genetic drift

- **Stochastic change** in allele frequency due to the effect of sampling a **finite population**
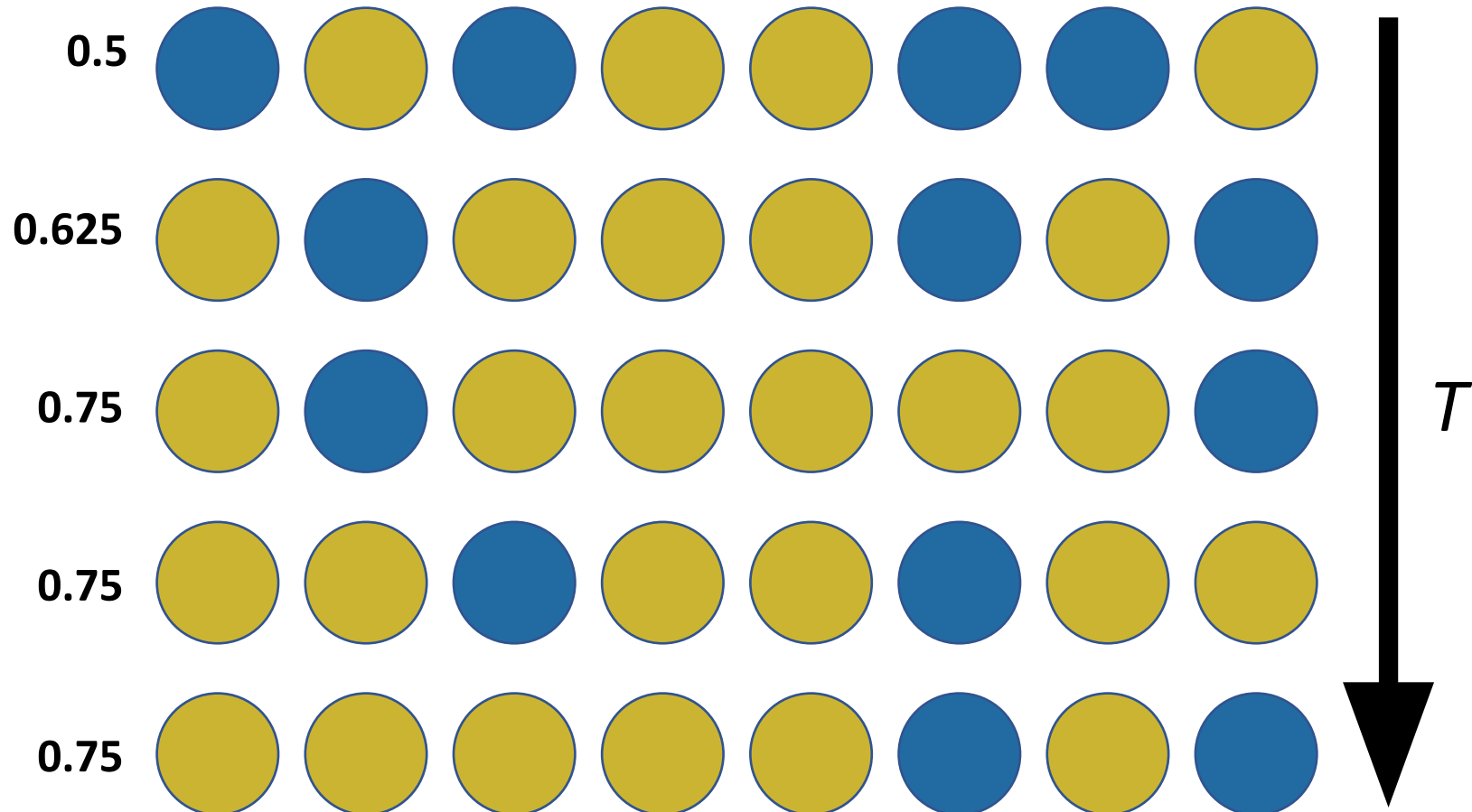
- A completely **neutral** process

Genetic drift

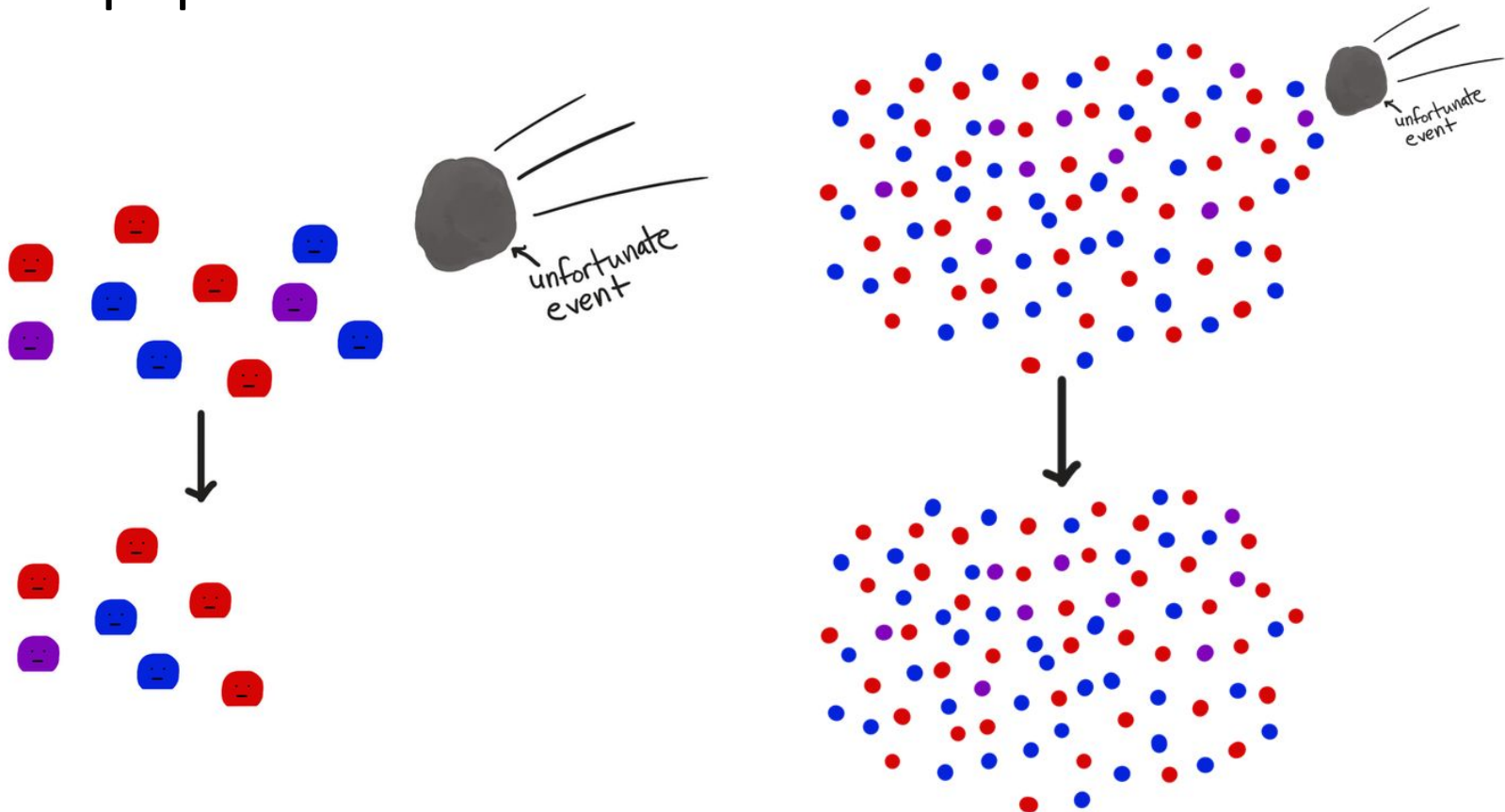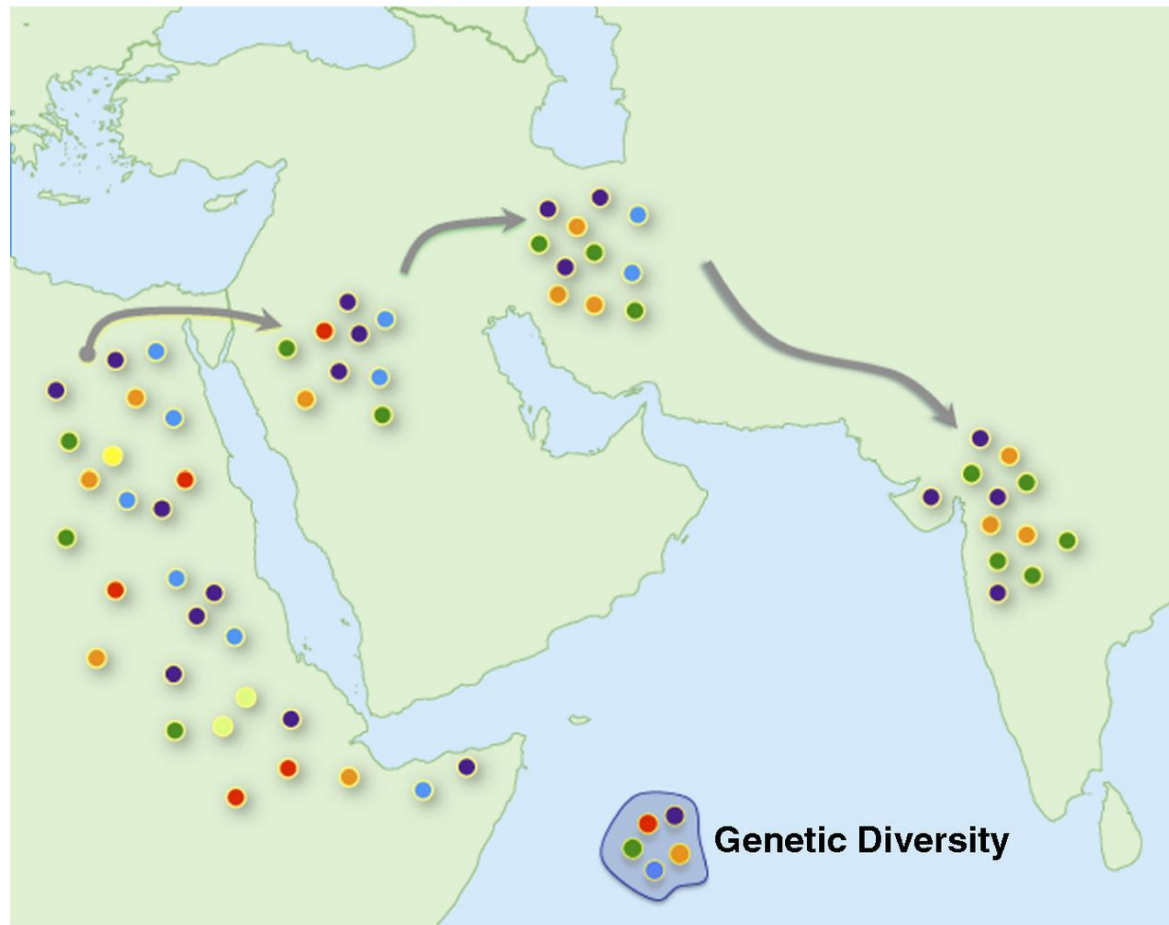# Genetic drift

T

Genetic drift

Genetic drift

- The effect of drift is **stronger** in **smaller** populations

Genetic drift

- Founder effects and bottlenecks

From Henn et al.
2012, PNAS

What is effective population size (N_e)?

- A **theoretical** population parameter

- The size that an **idealised** (i.e. Wright-Fisher) population would need to be in order to exhibit the **same qualities** as the real population

- These qualities generally derive from patterns of **neutral genetic diversity**, such as nucleotide diversity

Exercise 1

~20 min.

github.com/alanrogers/agar22/tree/main/introgression

Population divergence

- A single population becoming multiple populations - why might this occur?

- We can understand this process by **analyzing genomic data**

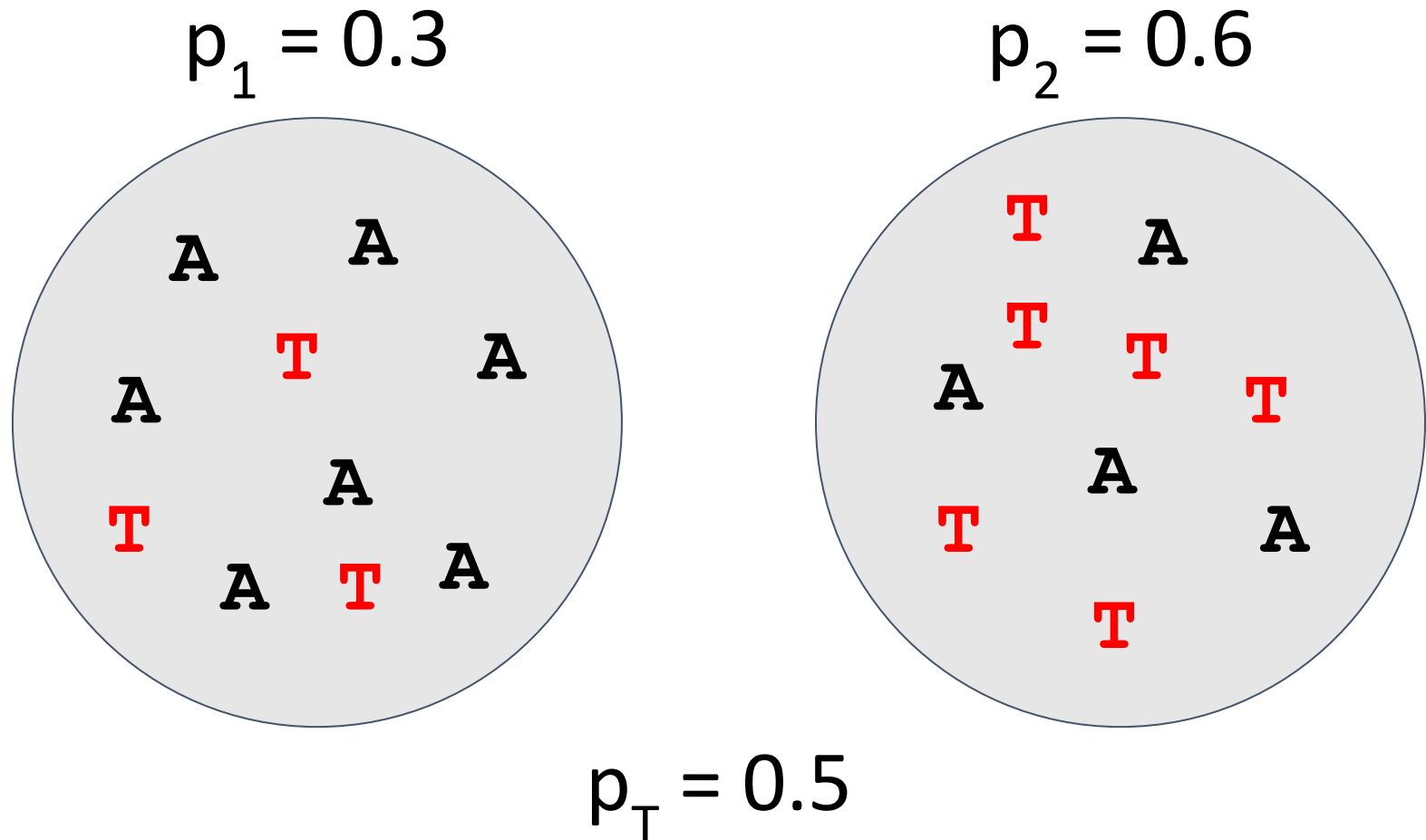- Following divergence, daughter populations begin to **evolve independently**

Population divergence

- F_{ST} is a measure of divergence calculated from allele frequencies across populations

$$F_{ST} = \frac{\text{total variance - within-pop variance}}{\text{total variance}}$$

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Population divergence

$p_1 = 0.3$

$p_2 = 0.6$

$p_T = 0.5$

Population divergence

Hardy-Weinberg equilibrium for biallelic locus

$p + q = 1$

$f_{AA} + f_{Aa} + f_{aa} = 1$

$f_{AA} = p^2$

$f_{aa} = q^2$

$f_{Aa} = 2pq$

Population divergence

$p_1 = 0.3$ $p_2 = 0.6$ $p_T = 0.5$

Population divergence

$p_1 = 0.3$ $\qquad\qquad$ $p_2 = 0.6$ $\qquad\qquad$ $p_T = 0.5$

$H_1 = 2p_1q_1$

Population divergence

$p_1 = 0.3$ $p_2 = 0.6$ $p_T = 0.5$

$H_1 = 2p_1q_1$
$H_1 = 2(0.3)(0.7)$

Population divergence

$p_1 = 0.3$ $p_2 = 0.6$ $p_T = 0.5$

$H_1 = 2p_1q_1$
$H_1 = 2(0.3)(0.7)$
$H_1 = 0.42$

Population divergence

$p_1 = 0.3$ $p_2 = 0.6$ $p_T = 0.5$

$H_1 = 2p_1q_1$ $H_2 = 2p_2q_1$
$H_1 = 2(0.3)(0.7)$ $H_2 = 2(0.6)(0.4)$
$H_1 = 0.42$ $H_2 = 0.48$

Population divergence

$p_1 = 0.3$ $p_2 = 0.6$ $p_T = 0.5$

$H_1 = 2p_1q_1$ $H_2 = 2p_2q_1$
$H_1 = 2(0.3)(0.7)$ $H_2 = 2(0.6)(0.4)$
$H_1 = 0.42$ $H_2 = 0.48$

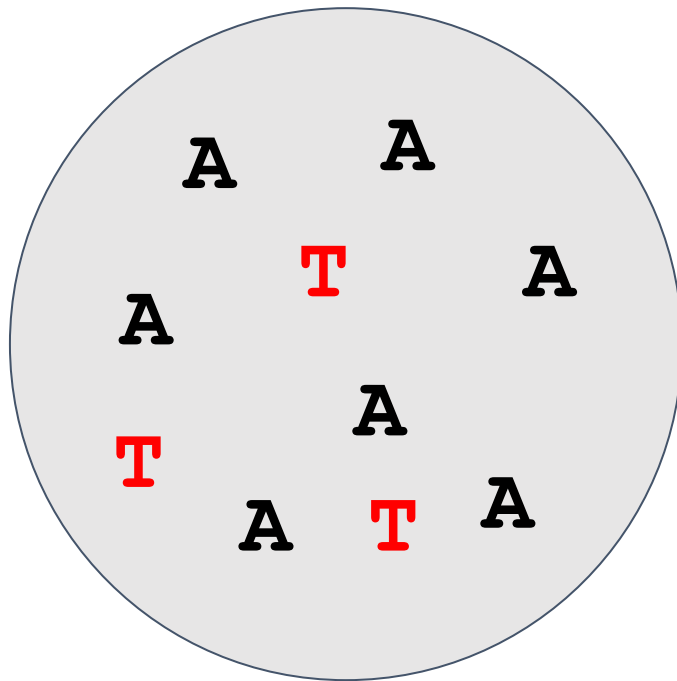$H_s = (0.42 + 0.48) / 2 = 0.45$

Population divergence

$p_1 = 0.3$ $\qquad\qquad$ $p_2 = 0.6$ $\qquad\qquad$ $p_T = 0.5$

$H_1 = 2p_1q_1$ $\qquad\qquad$ $H_2 = 2p_2q_1$
$H_1 = 2(0.3)(0.7)$ $\qquad\quad$ $H_2 = 2(0.6)(0.4)$
$H_1 = 0.42$ $\qquad\qquad\quad$ $H_2 = 0.48$

$H_s = (0.42 + 0.48) / 2 = 0.45$
$H_T = 2(0.5)(0.5) = 0.5$

Population divergence

$p_1 = 0.3$ $p_2 = 0.6$ $p_T = 0.5$

$H_1 = 2p_1q_1$ $H_2 = 2p_2q_1$
$H_1 = 2(0.3)(0.7)$ $H_2 = 2(0.6)(0.4)$
$H_1 = 0.42$ $H_2 = 0.48$

$H_S = (0.42 + 0.48) / 2 = 0.45$
$H_T = 2(0.5)(0.5) = 0.5$

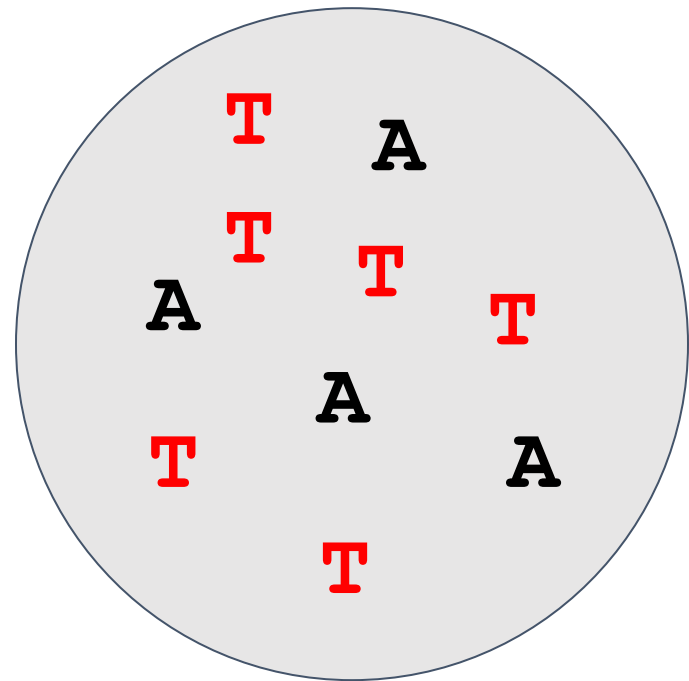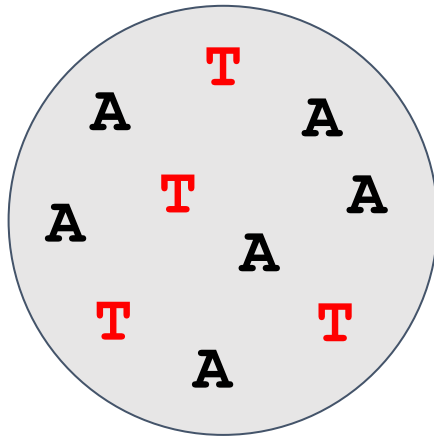$F_{ST} = (0.5 - 0.45) / 0.5 = 0.1$
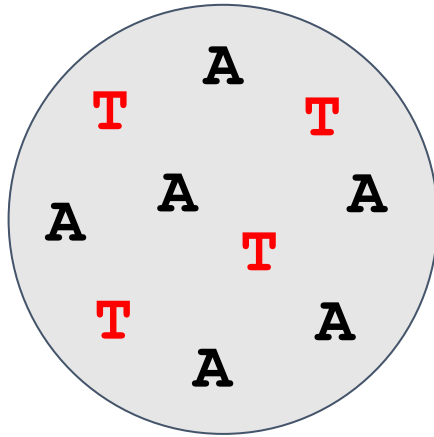
Population divergence

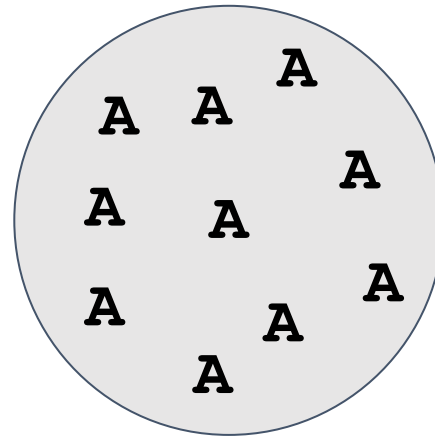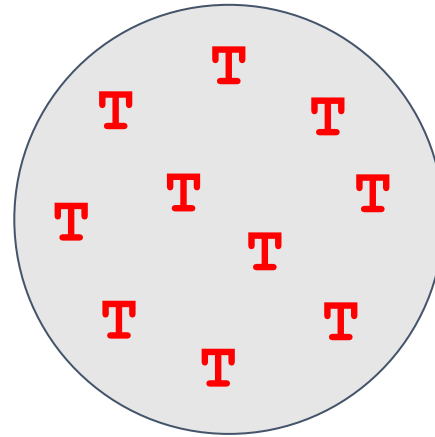$$p_1 = 0.3 \qquad p_2 = 0.6$$

$$p_T = 0.5, \; F_{ST} = 0.1$$

Population divergence

$F_{ST} = 0$

$F_{ST} = 1$

Exercise 2
~20 min.

github.com/alanrogers/agar22/tree/main/introgression