

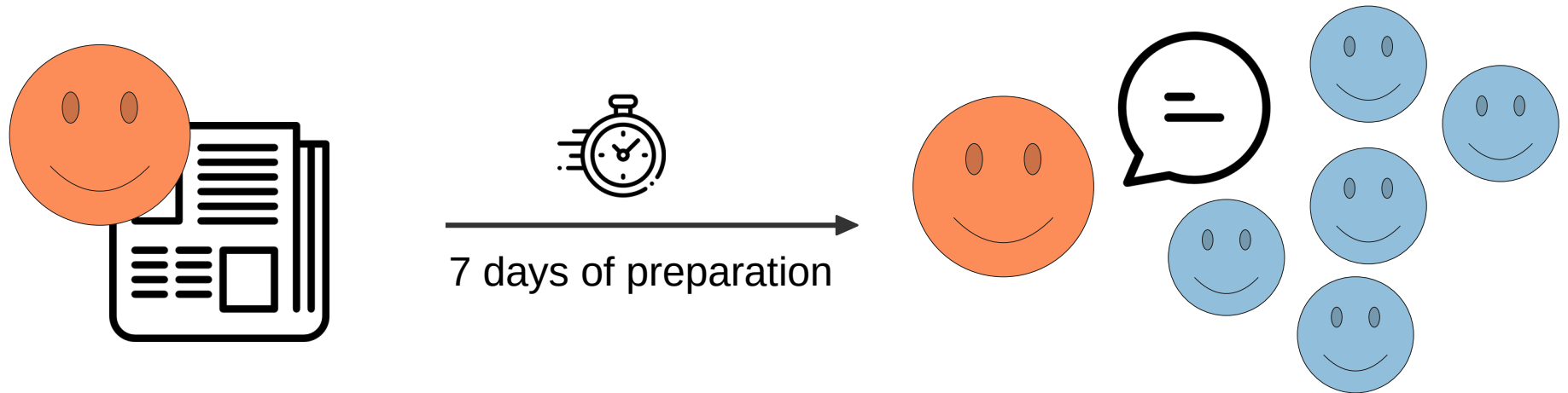
# Population genomics



[camille.roux@univ-lille.fr](mailto:camille.roux@univ-lille.fr)

## Evaluation on the module :

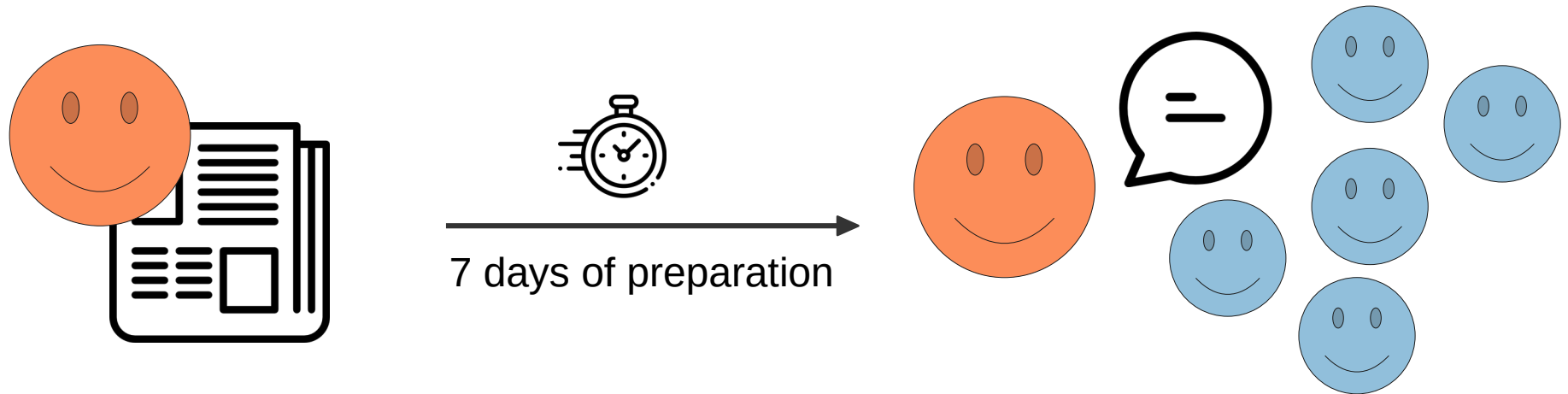
Each student will present an article at the beginning of each session



The aim of this student: **make the message of the article clear to the audience.**

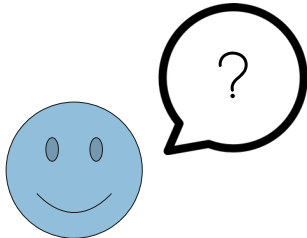
## Evaluation on the module :

Each student will present an article at the beginning of each session



The aim of this student: **make the message of the article clear to the audience.**

20 minutes of presentation ...  
... followed by questions



## Evaluation on the module :

Each student will present an article at the beginning of each session

### Clarity :

- Making clear the **context** (why o this study?): 4 points
- What is the **scientific question** ? 2 points
- In concrete terms, what **do the authors do to answer** the question? 4 points
- What are the **results** ? 4 points
- **Interpretation of** the results : 4 points
- **Take home message** from the paper : 2 points

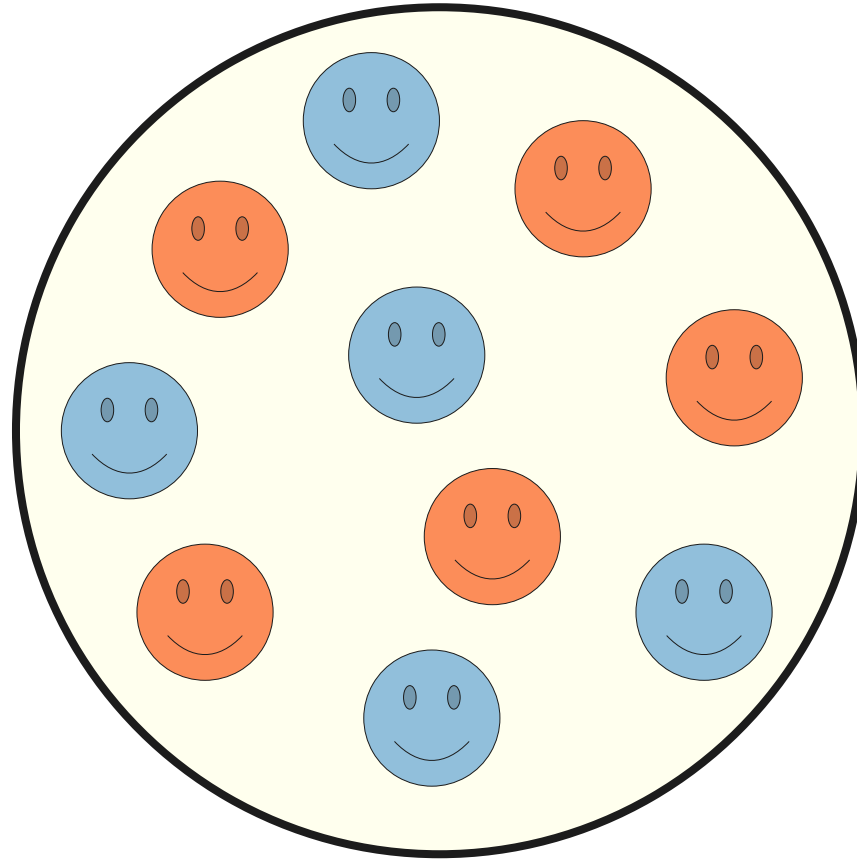
# Population genomics



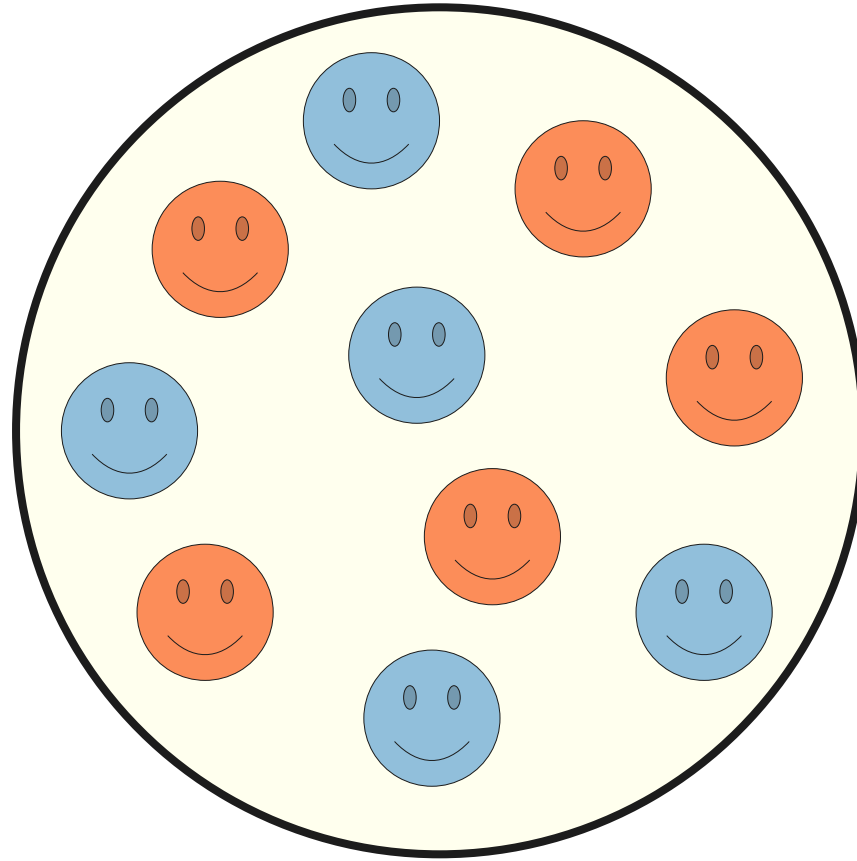
[camille.roux@univ-lille.fr](mailto:camille.roux@univ-lille.fr)

Today : an introduction to pop. geno.

# Genetic diversity

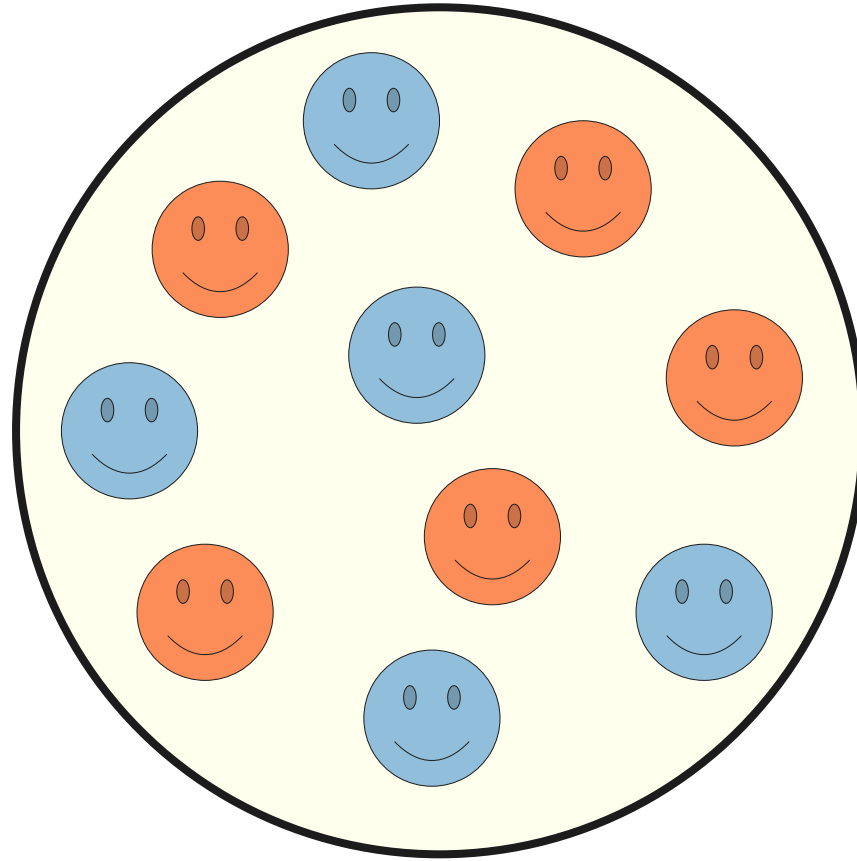


# Genetic diversity



How to quantify diversity?

# Genetic diversity

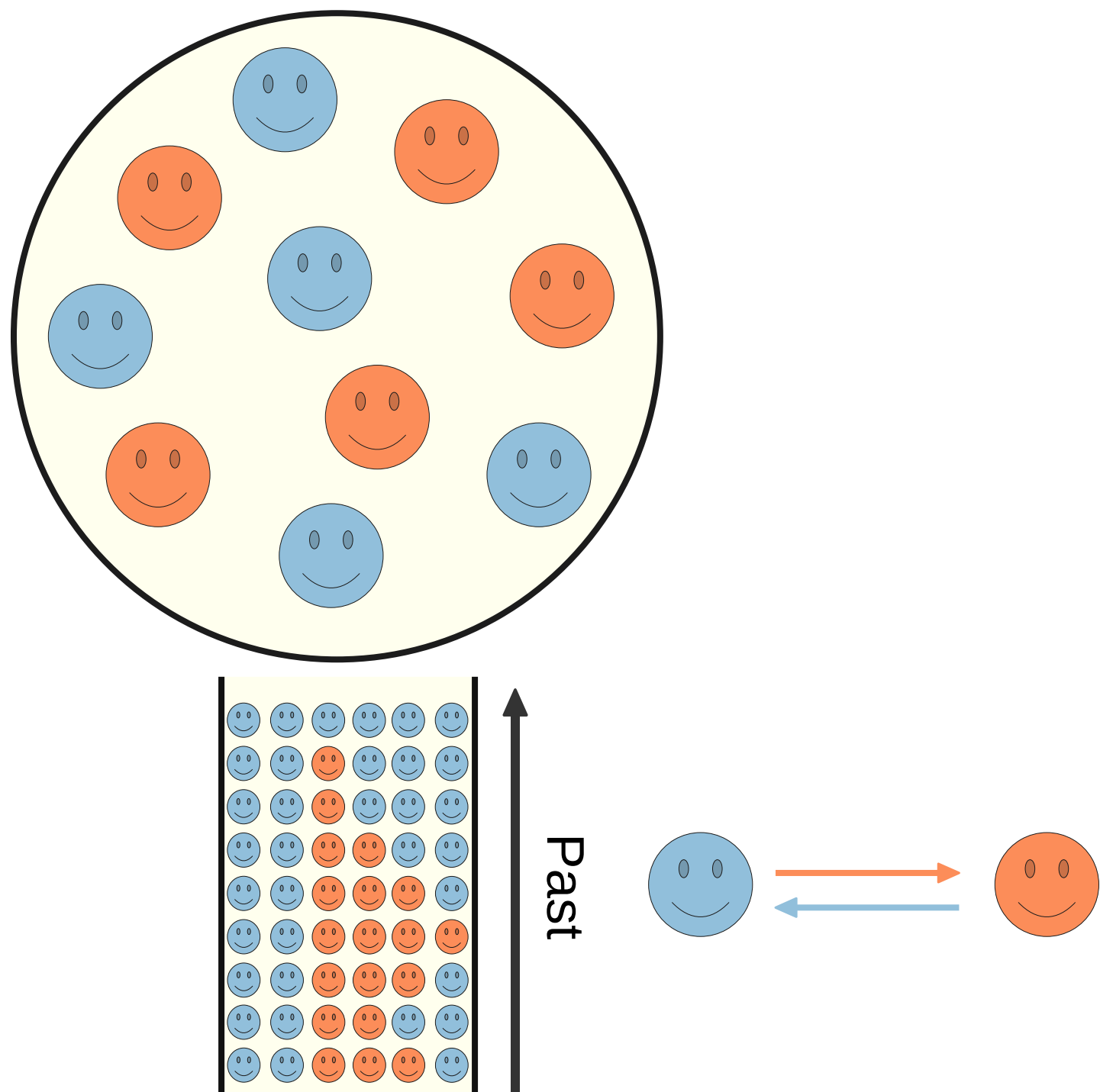


How to quantify diversity?

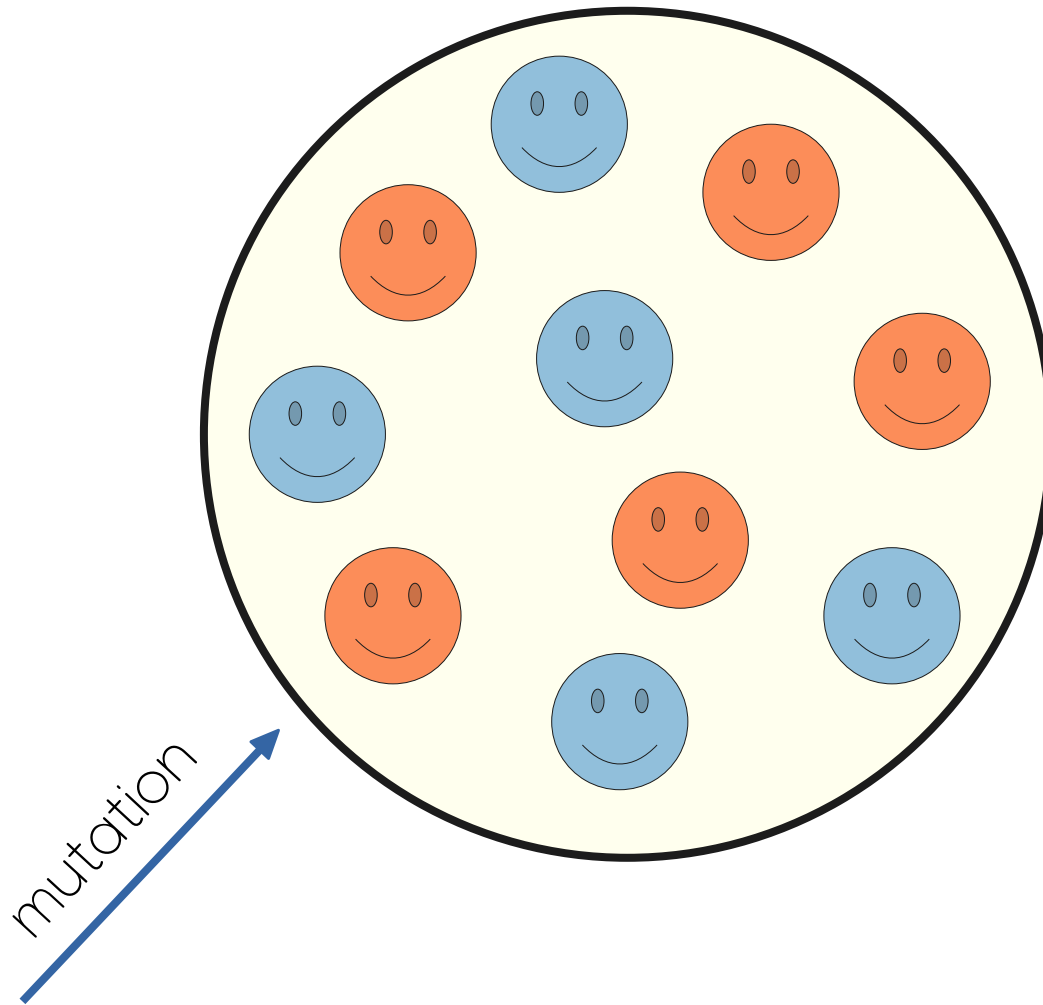
How to interpret this diversity?



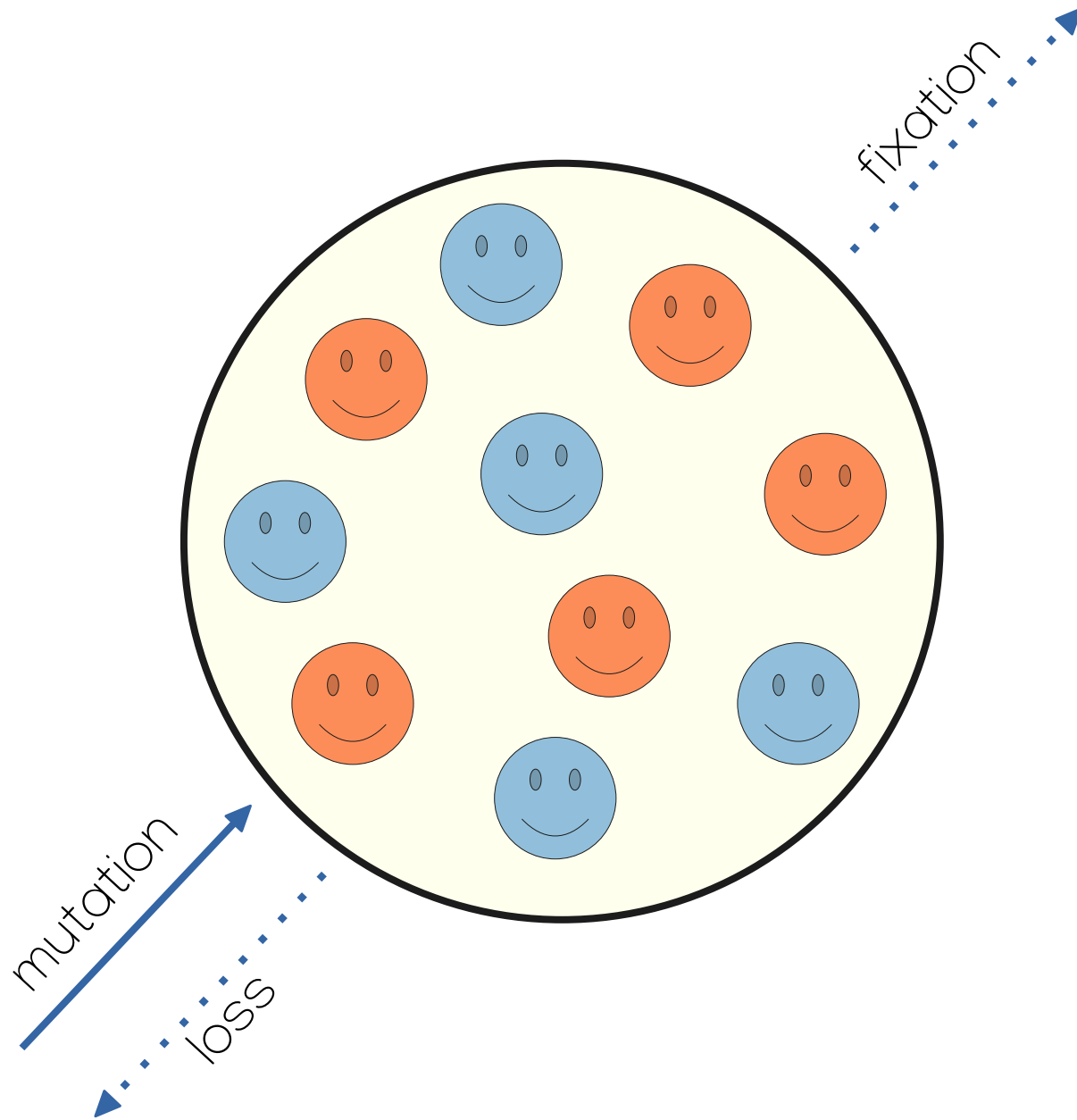
# Genetic diversity



The product  $Ne\mu$ : important determinant of genetic diversity



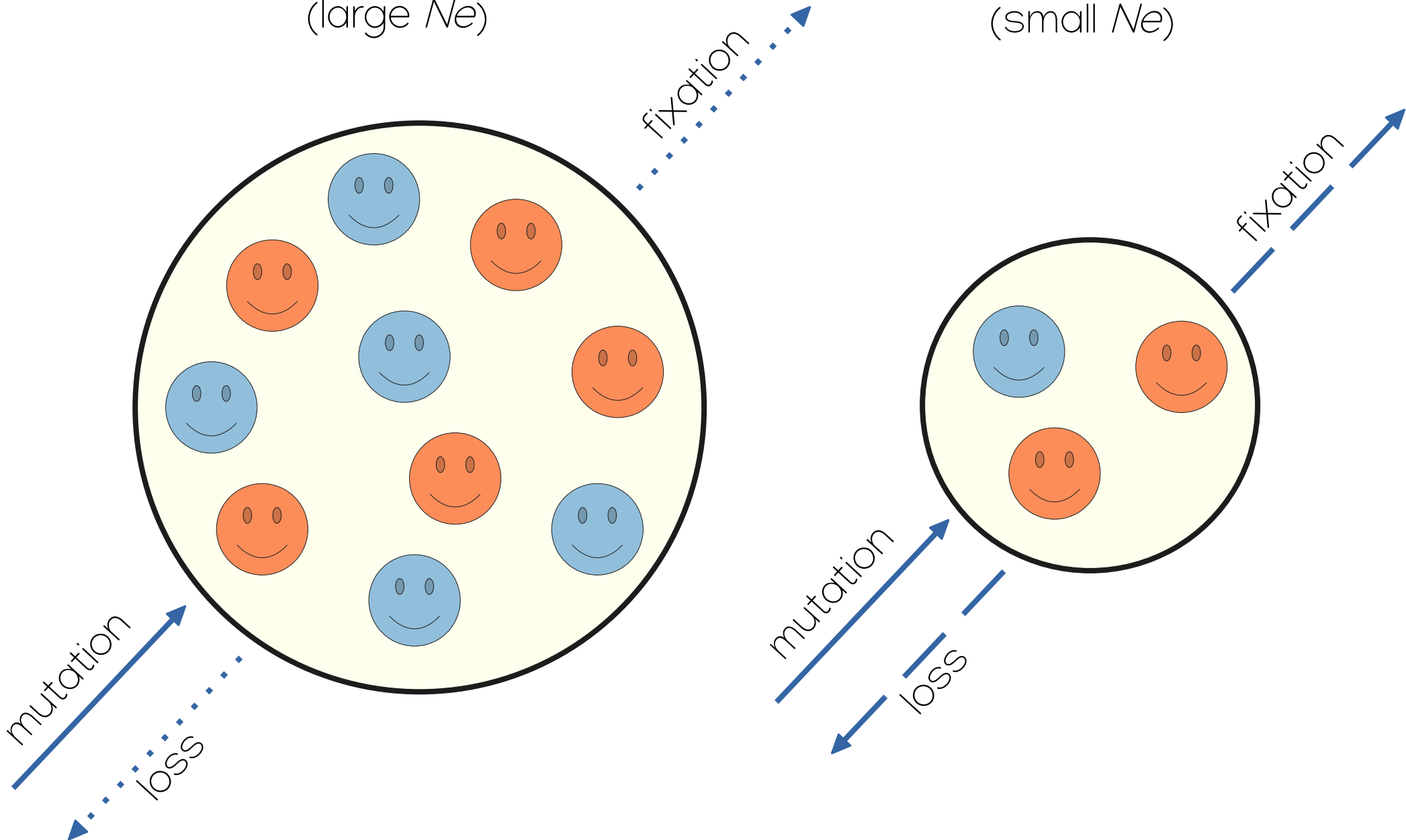
The product  $Ne\mu$ : important determinant of genetic diversity



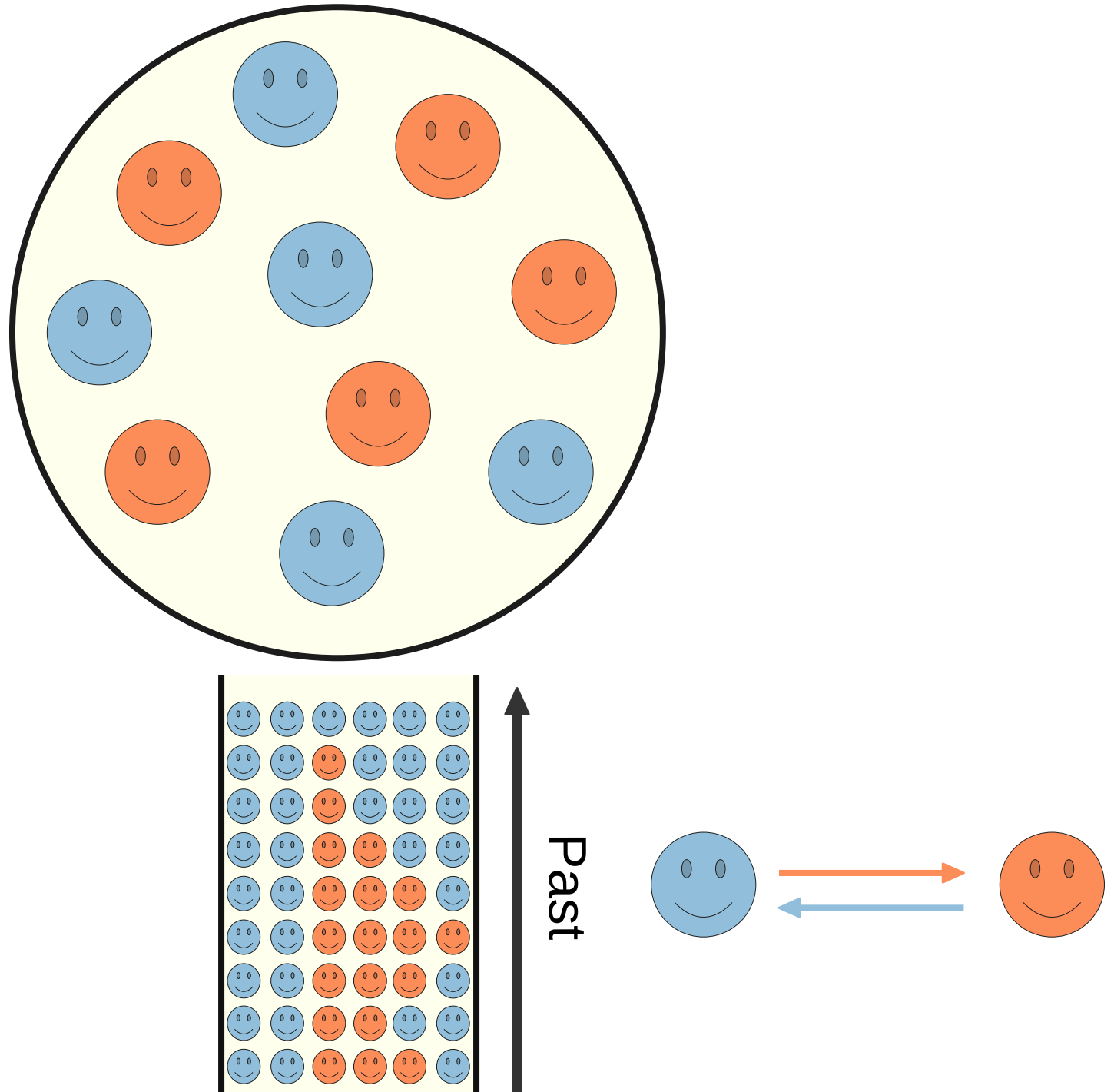
The product  $Ne\mu$ : important determinant of genetic diversity

Large population  
(large  $Ne$ )

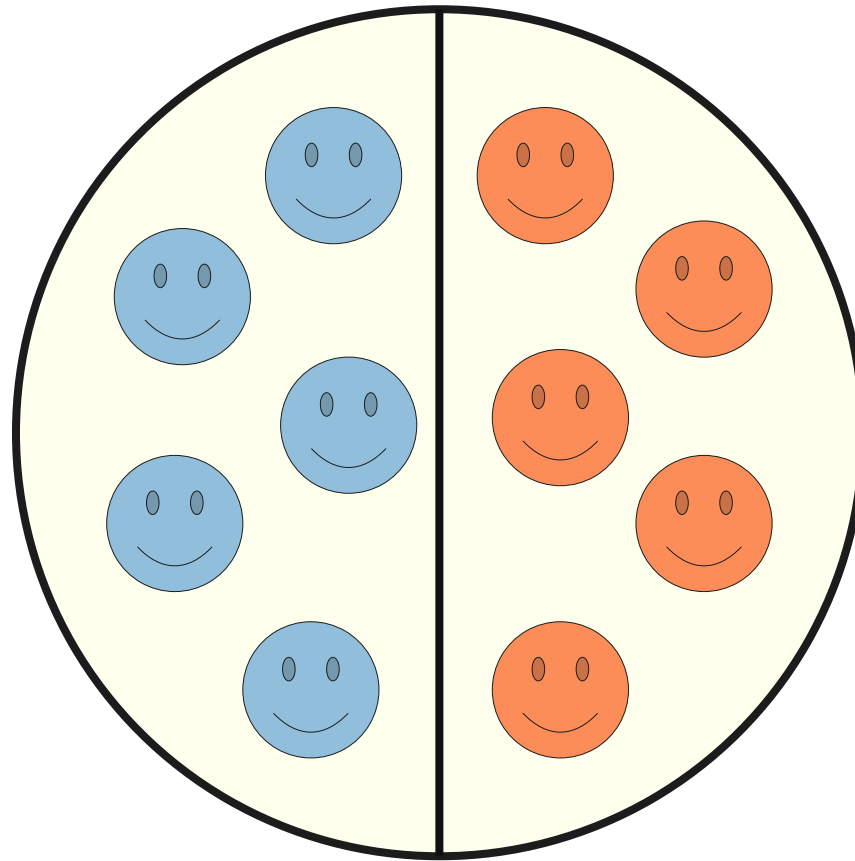
Small population  
(small  $Ne$ )



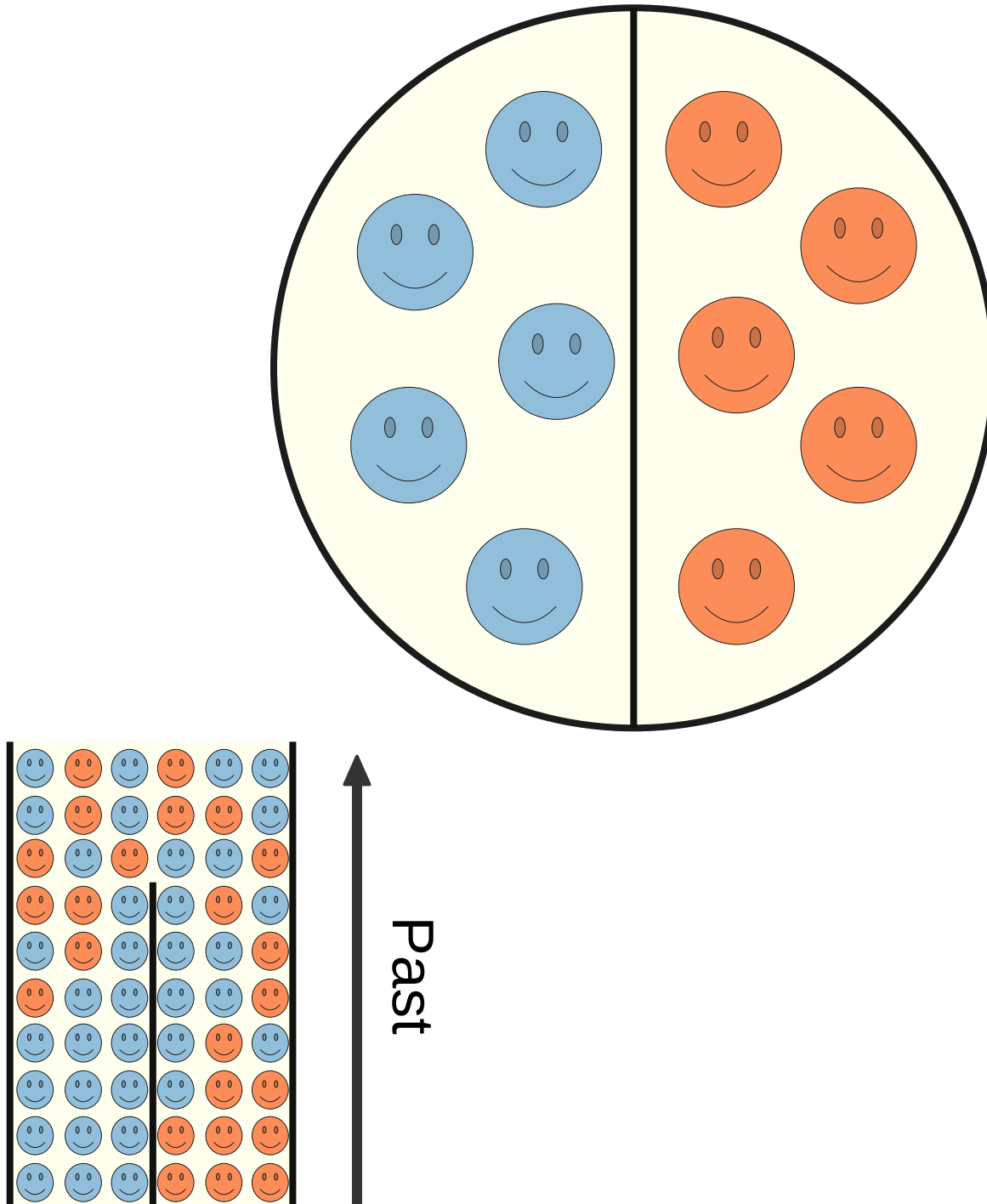
# Wright and Fisher model



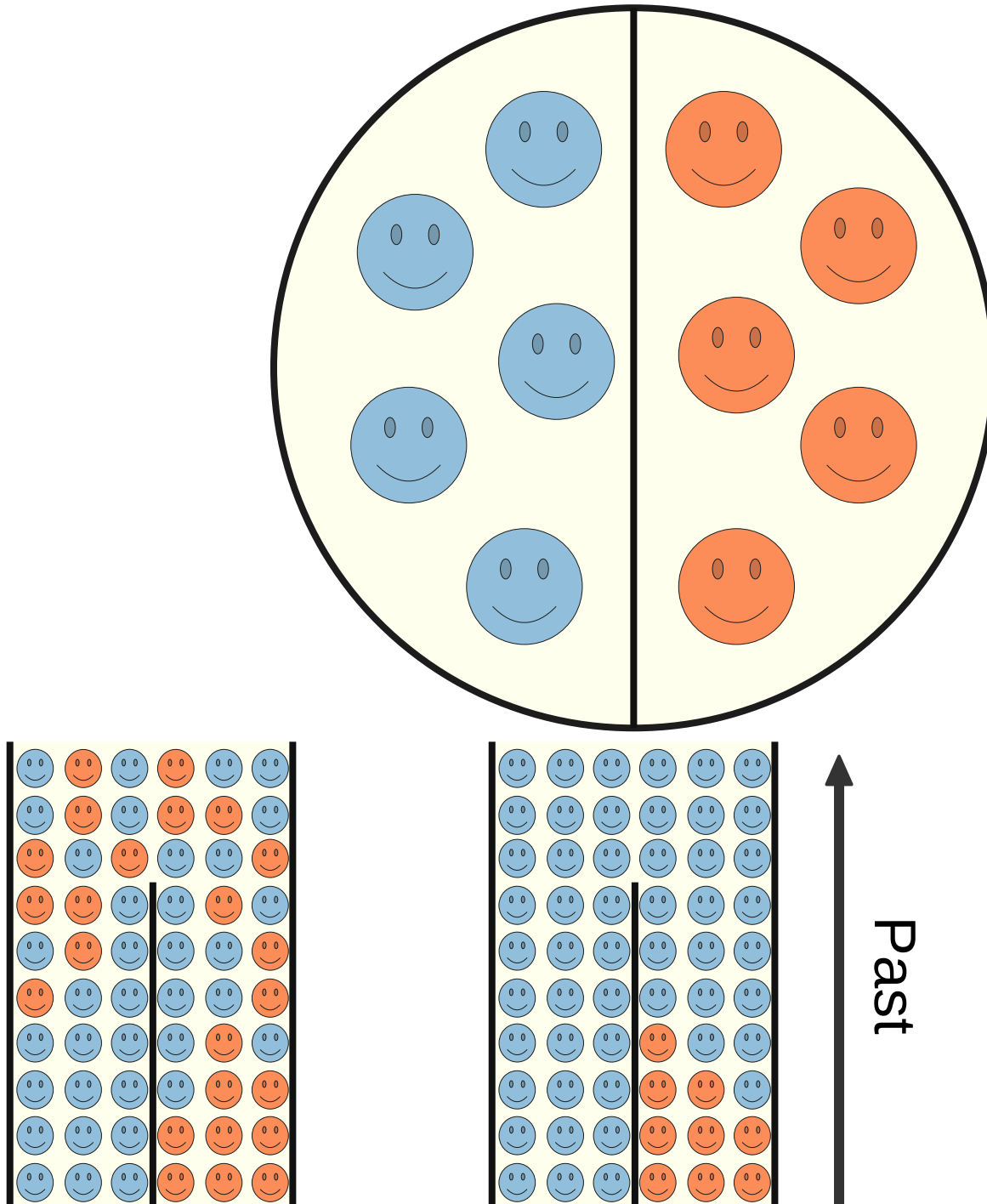
# Structured populations



# Structured populations

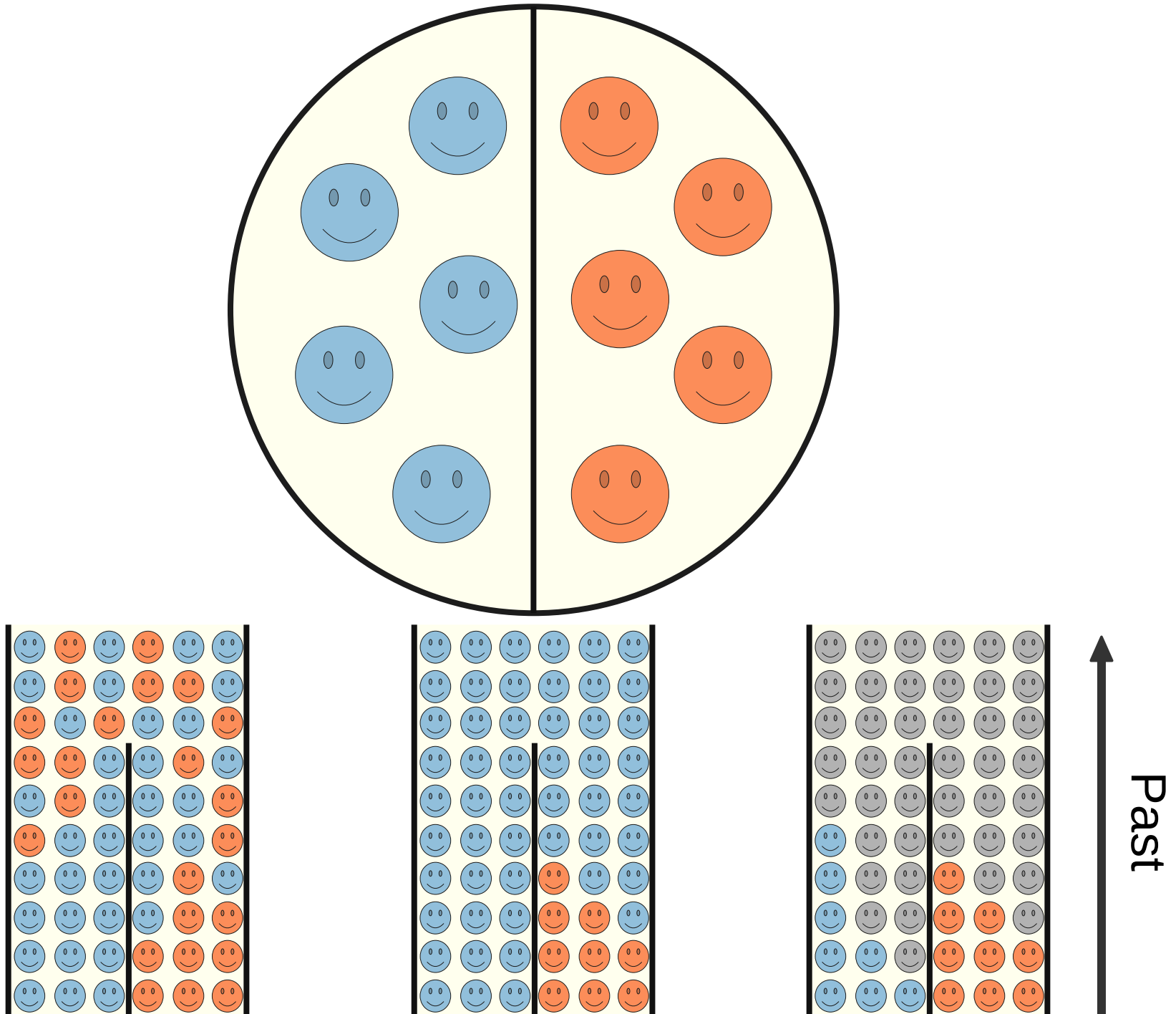


# Structured populations

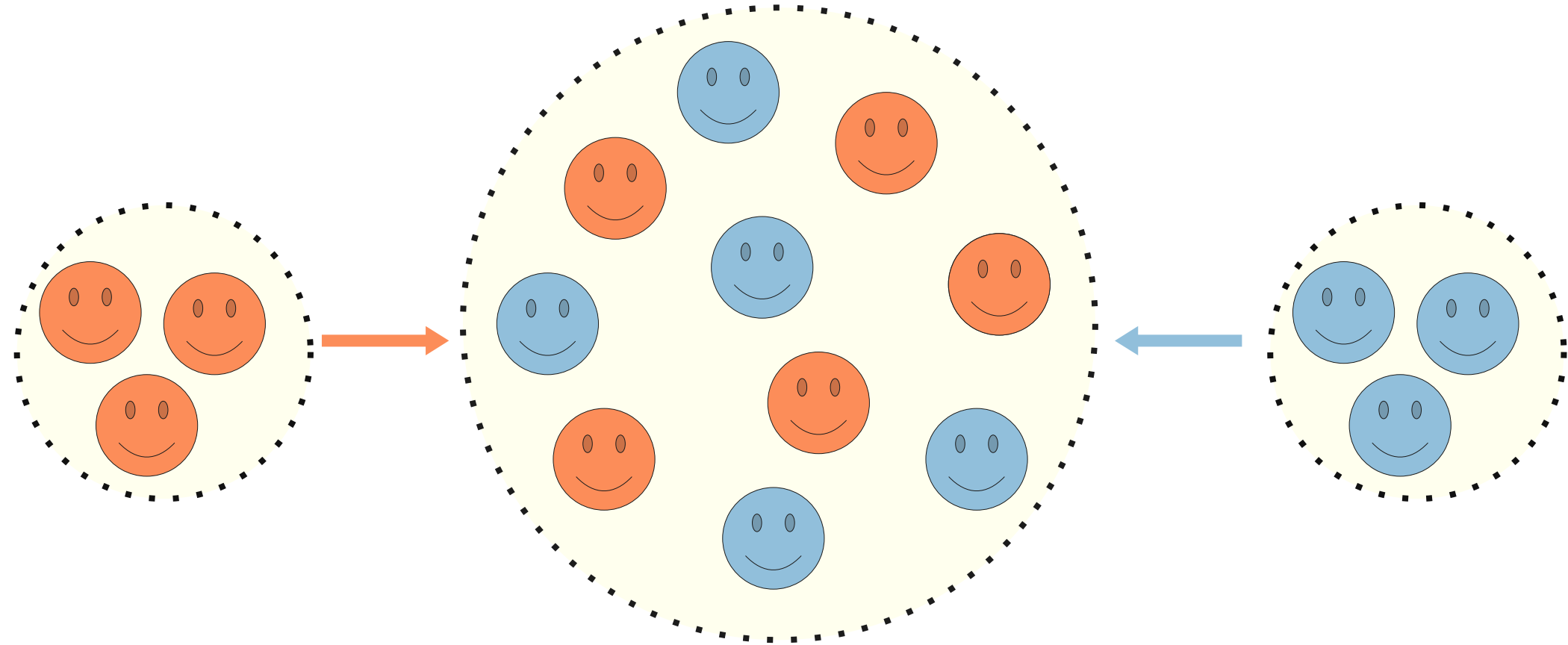




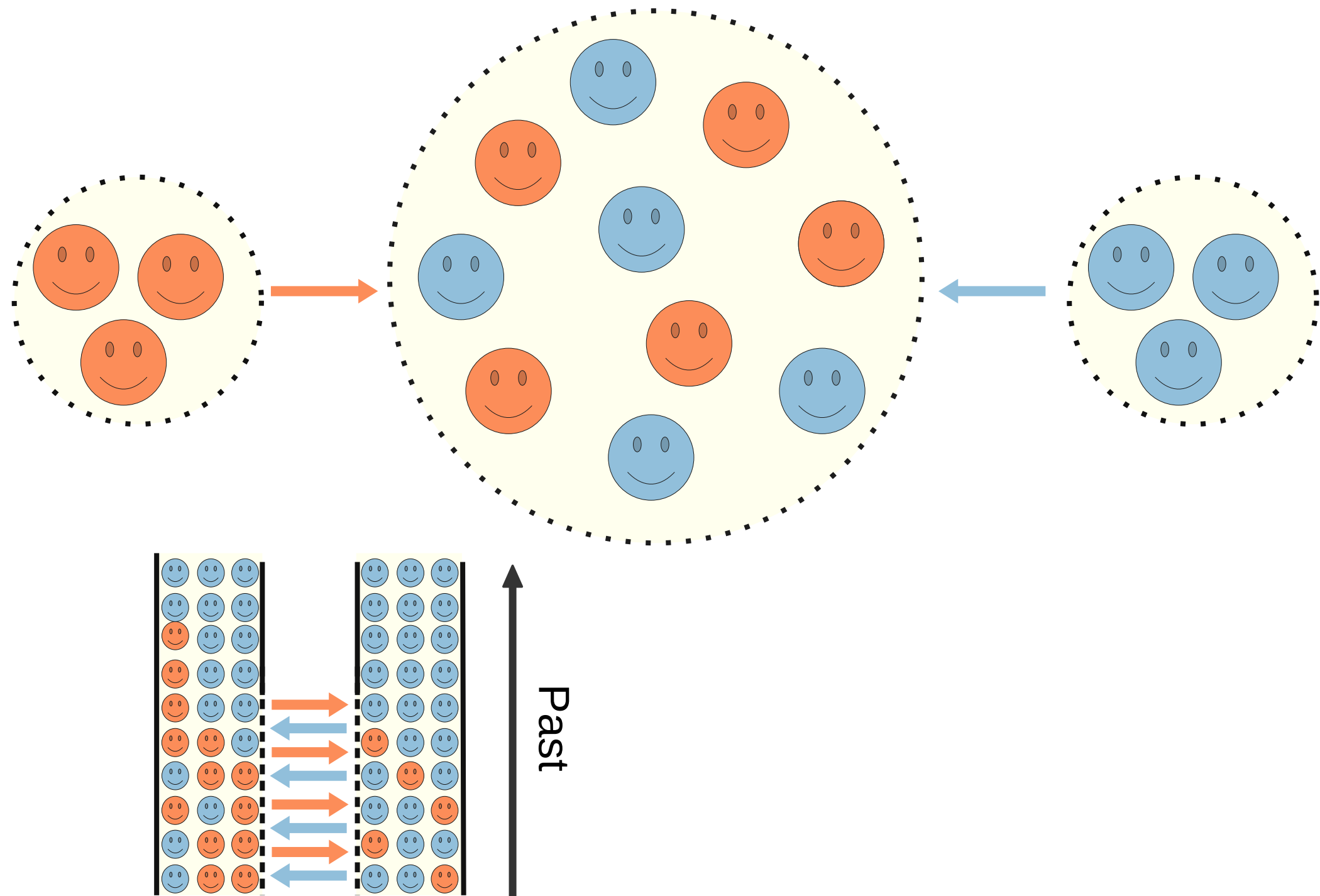
# Structured populations



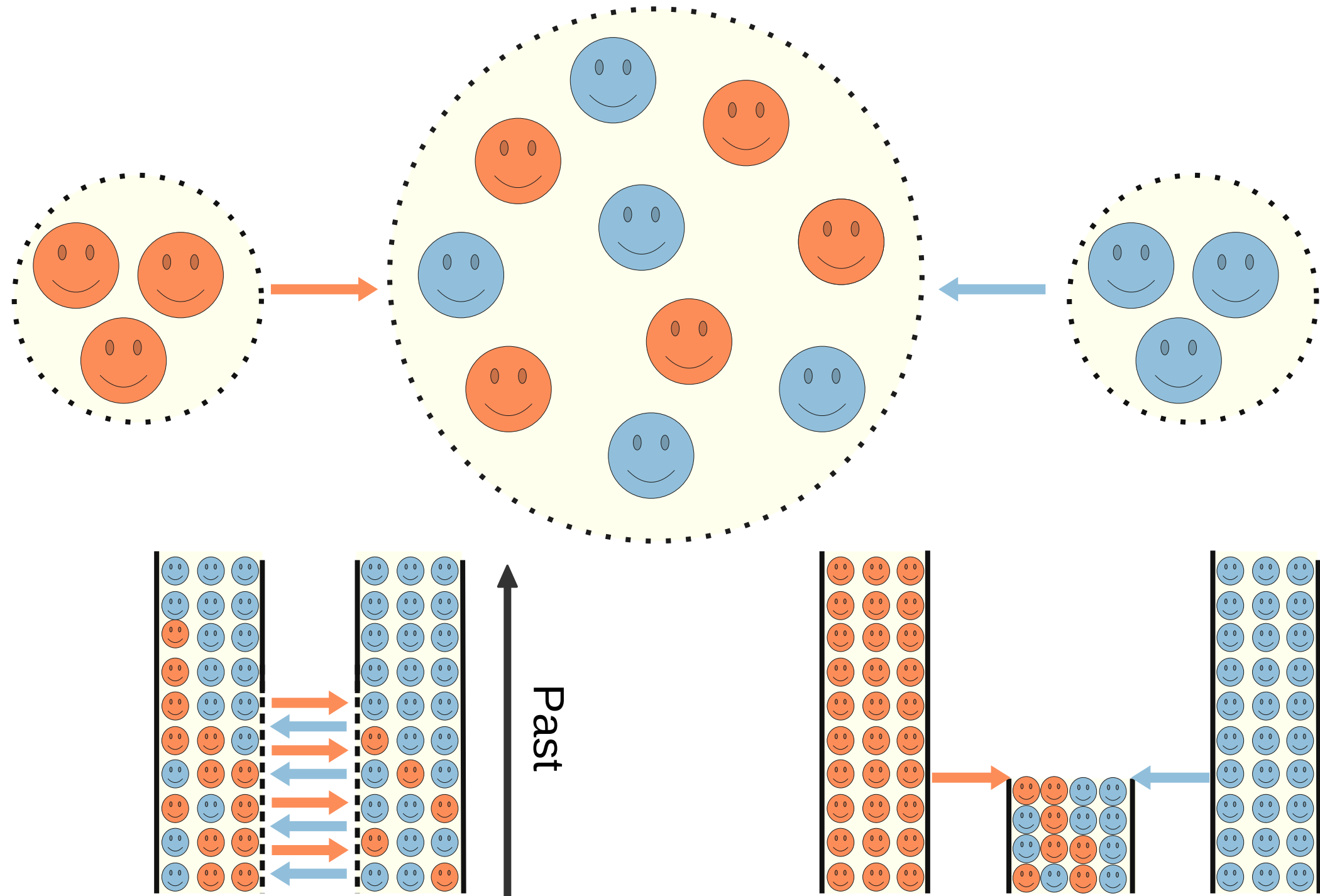
Migration



# Migration

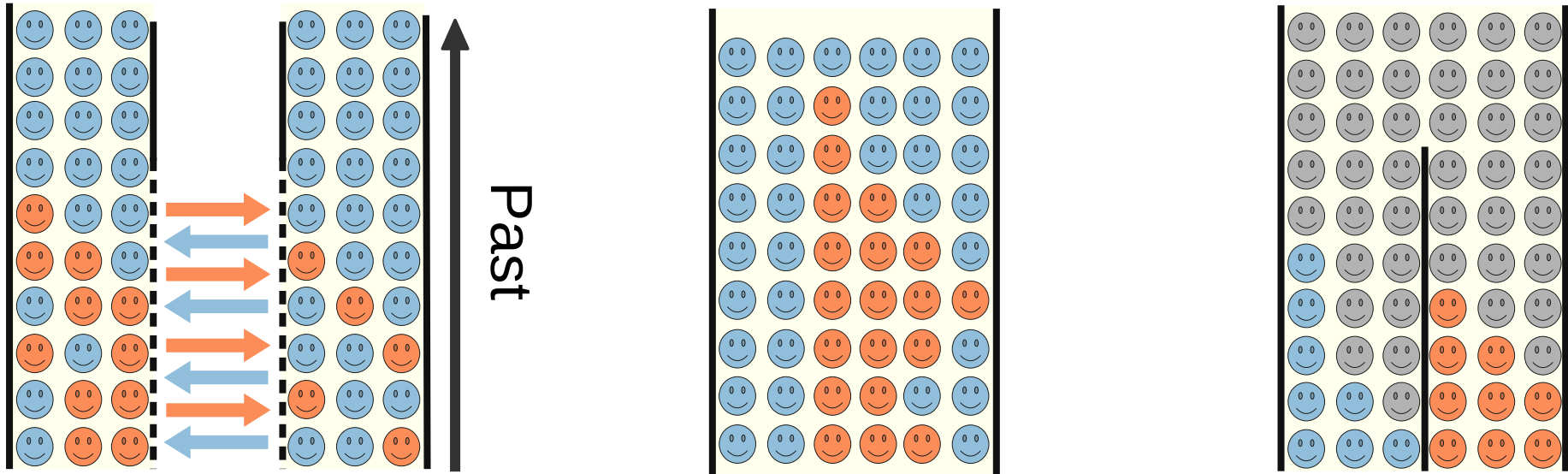


# Migration

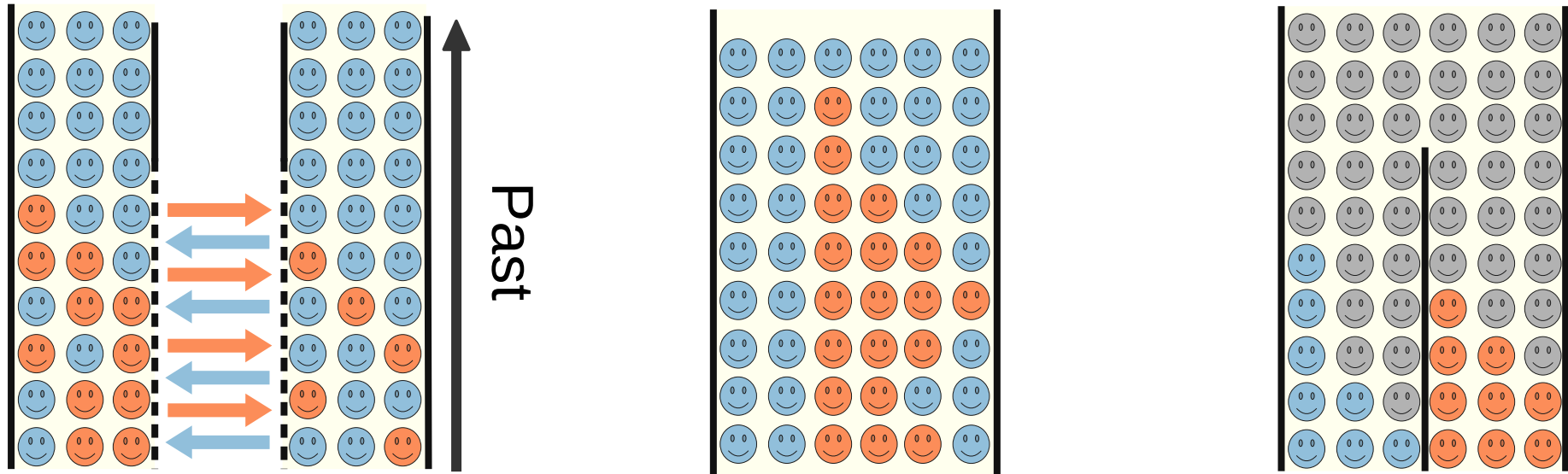




# 1. How to compare alternative demographic scenarios?

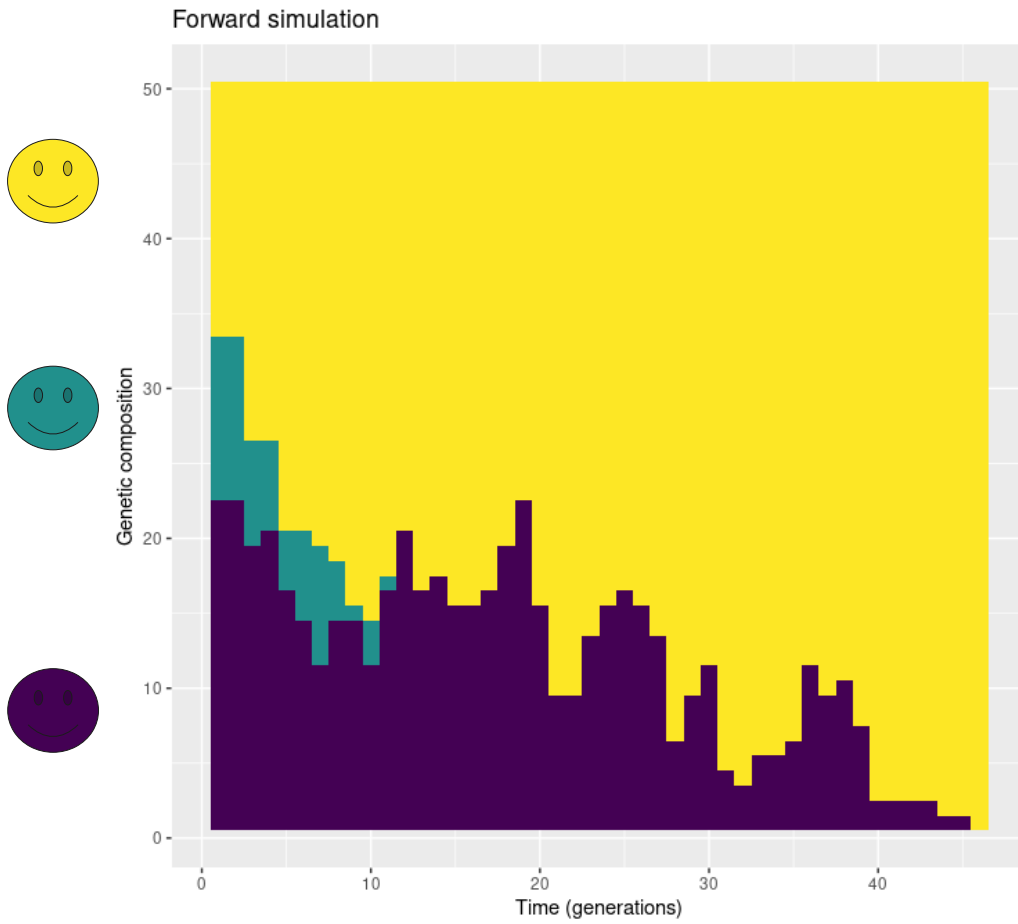


1. How to compare alternative demographic scenarios?



2. What evolutionary forces act on a given locus?

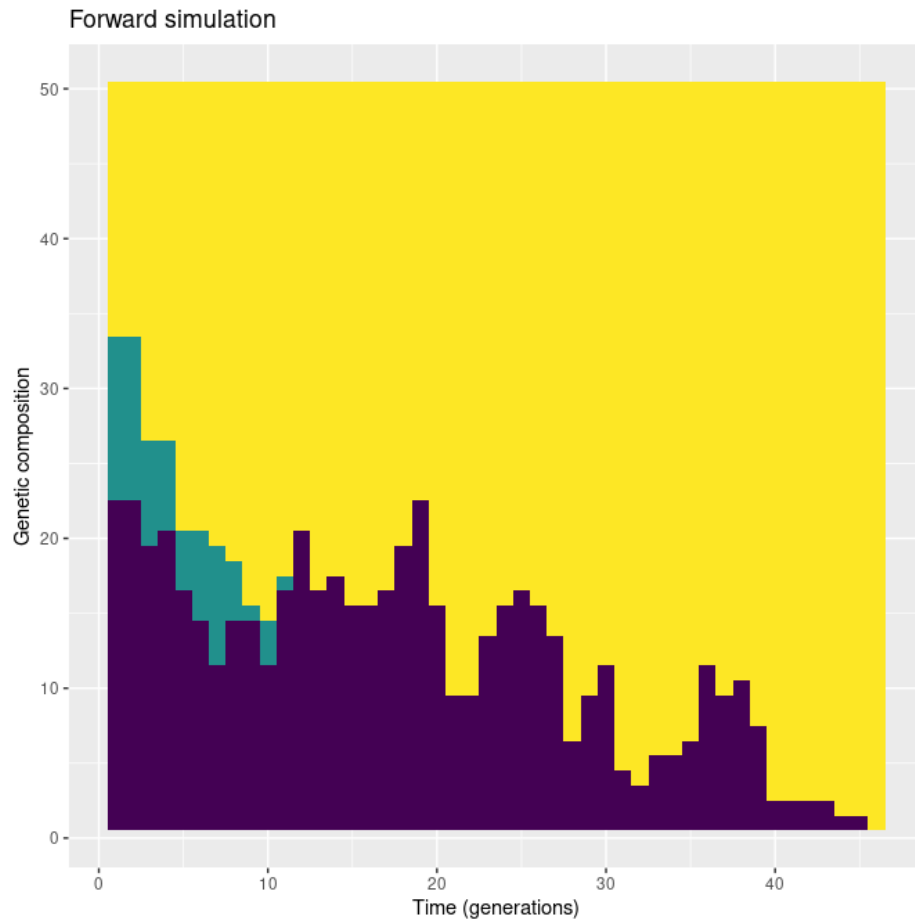
# Genetic drift



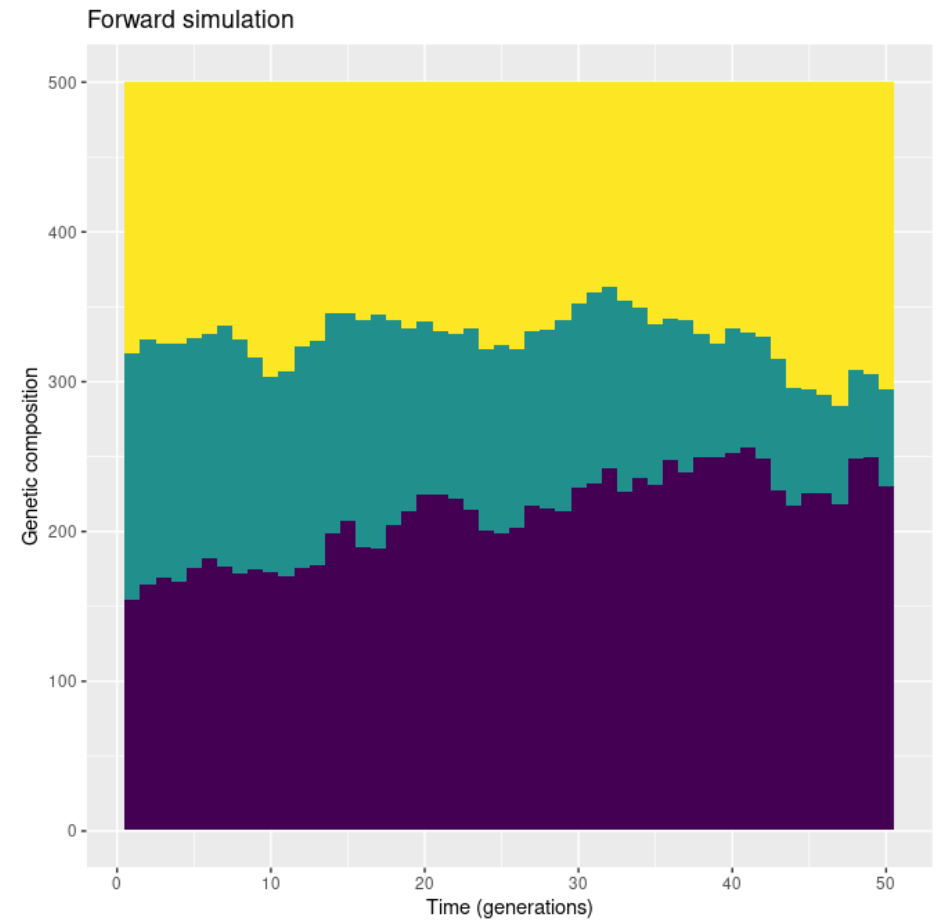
$N_e = 50$  (individuals)



# Genetic drift



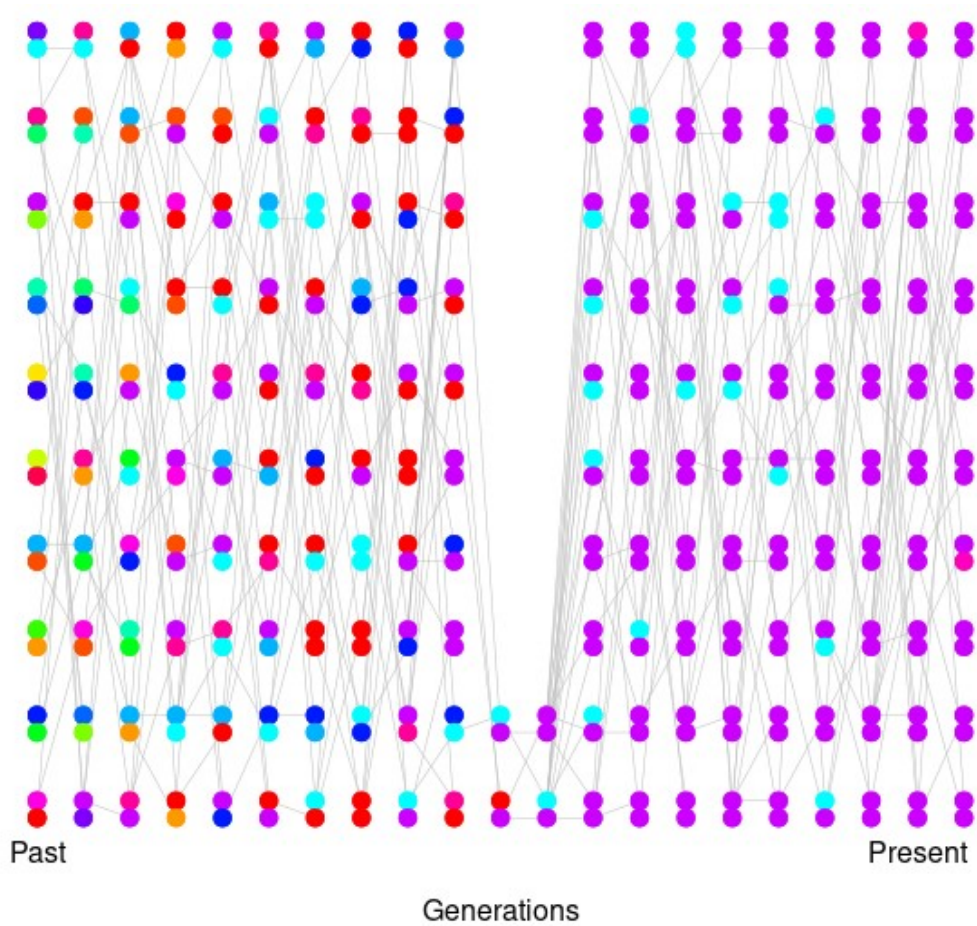
$N_e = 50$  (individuals)



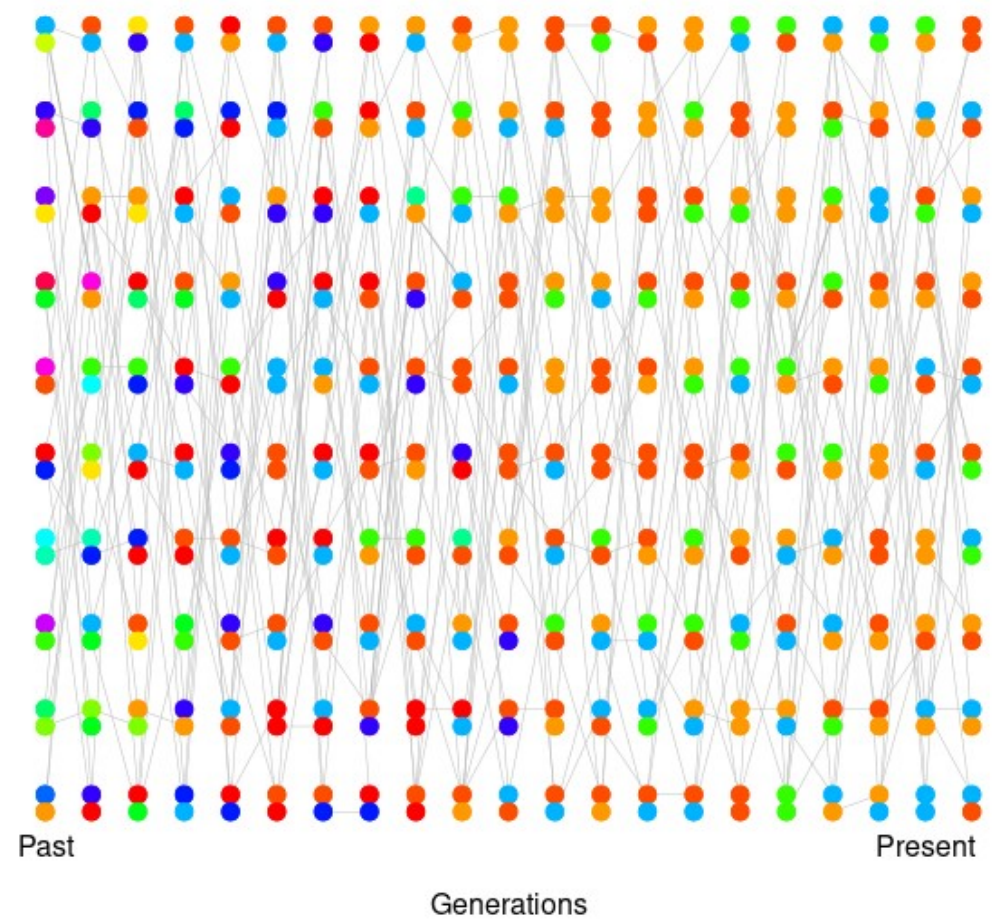
$N_e = 500$

# Genetic drift

Bottleneck

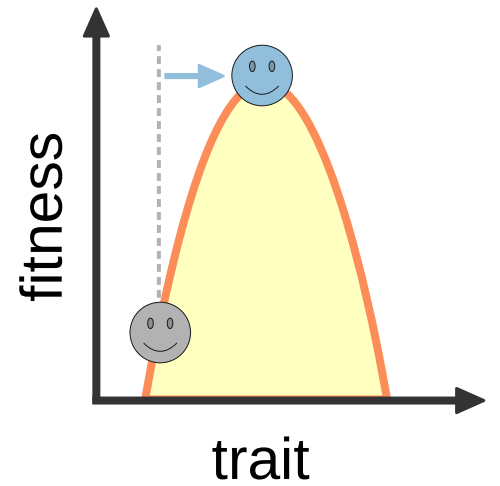


Constant size



# Natural selection

## Directional selection

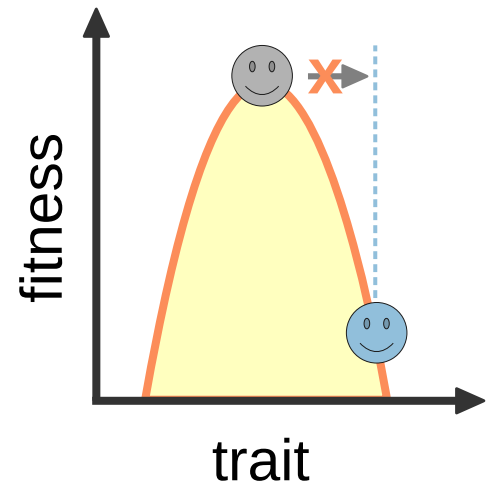


 Ancestral allele

 Derived allele

# Natural selection

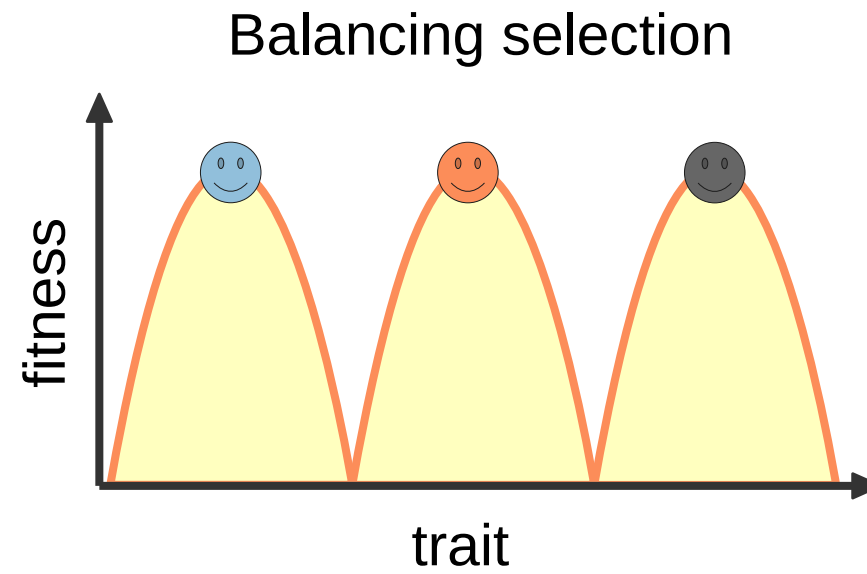
## Purifying selection



 Ancestral allele

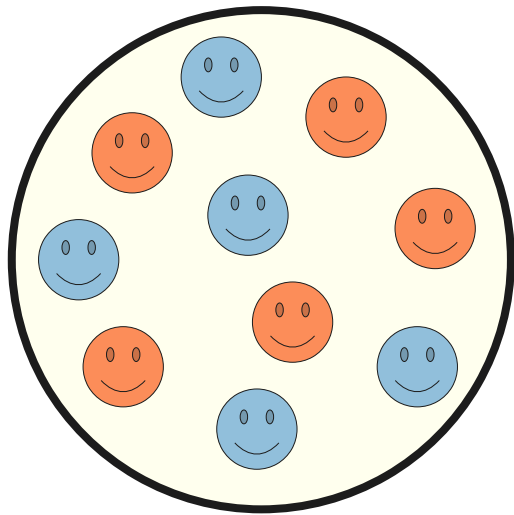
 Derived allele

# Natural selection



1. How to compare alternative demographic scenarios?
2. What evolutionary forces act at a given locus?

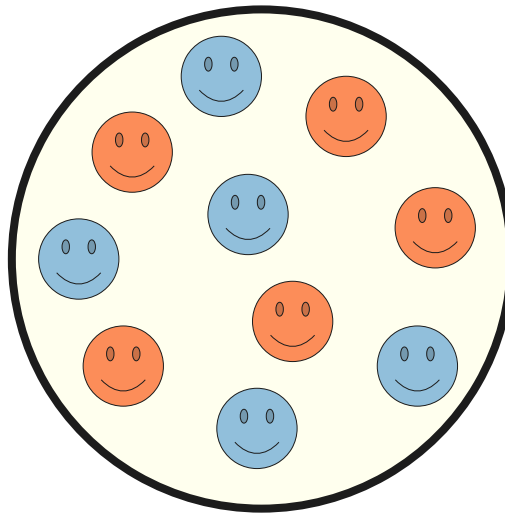
1. How to compare alternative demographic scenarios?
2. What evolutionary forces act at a given locus?



Observation

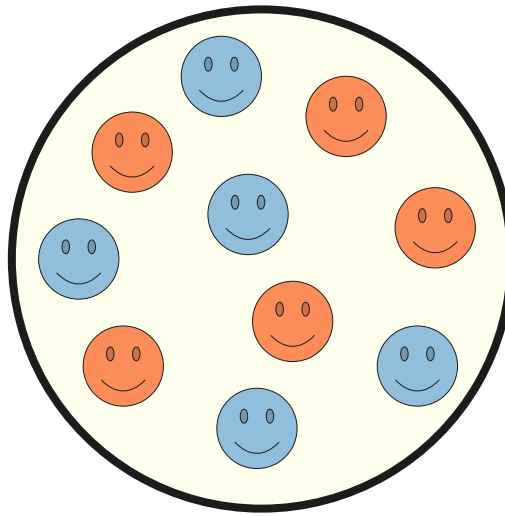
Quantifying  
genomic patterns

Model  
of  
evolution



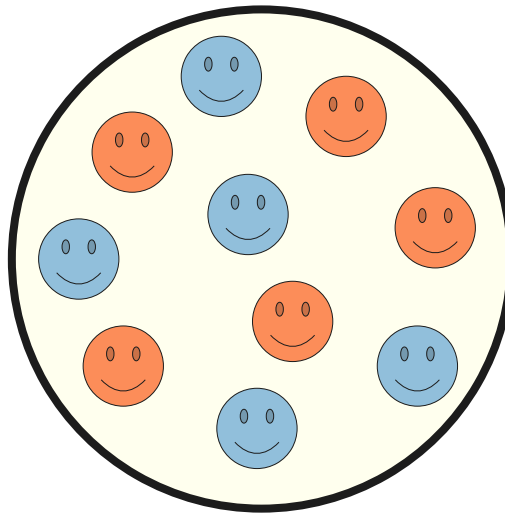
How can we measure that ?!?





How can we measure that ?!?

1. Sampling individuals from natural population(s)

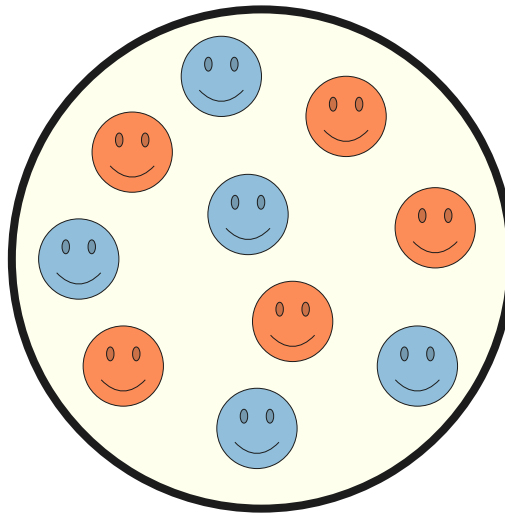


How can we measure that ?!?

1. Sampling individuals from natural population(s)



2. Sequencing them



How can we measure that ?!?

1. Sampling individuals from natural population(s)



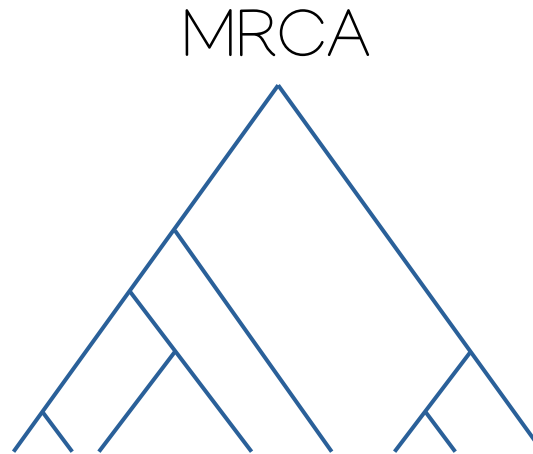
2. Sequencing them



3. Analyse the sequences

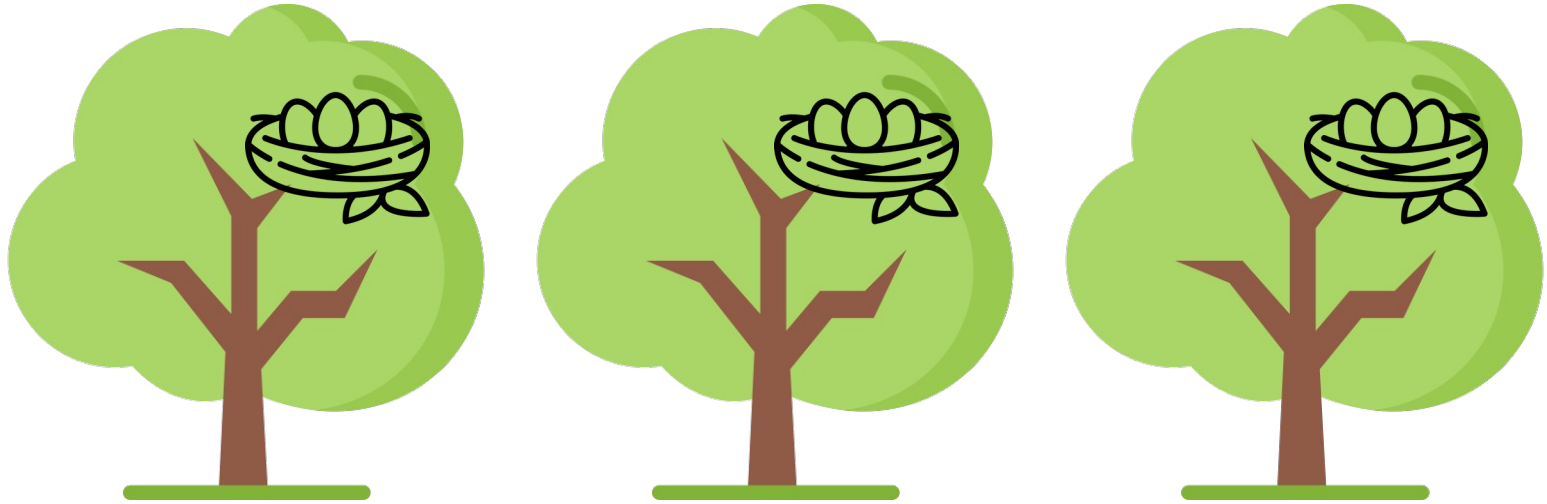
# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals

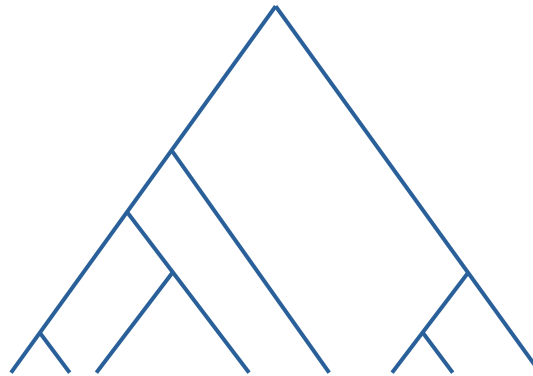


# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals

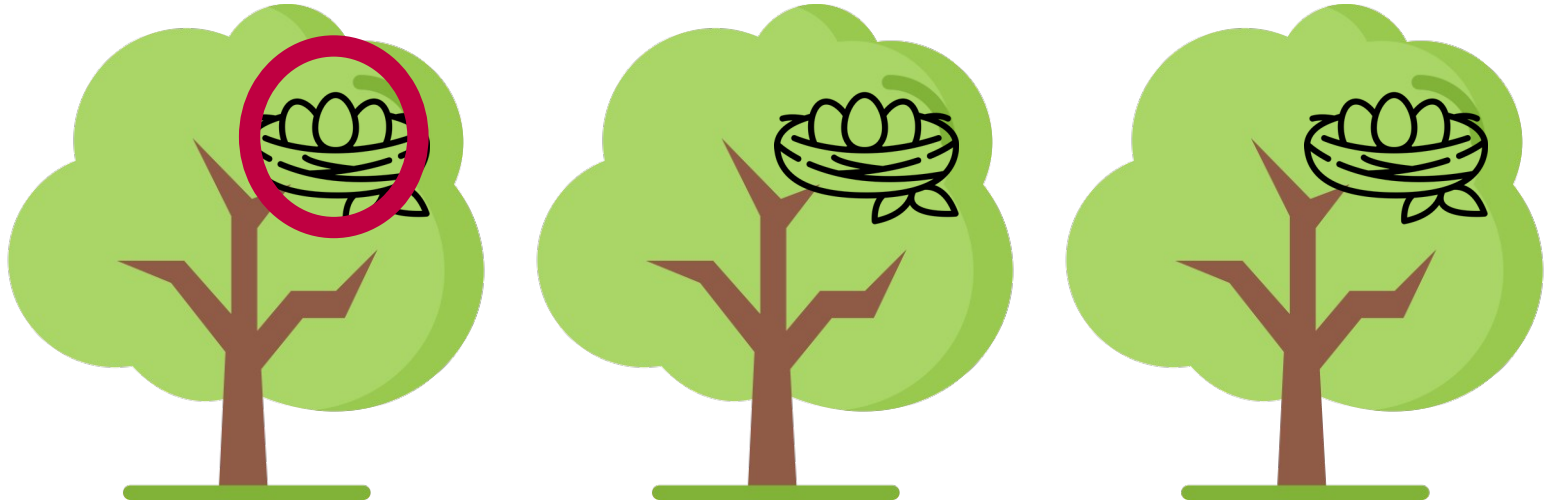


MRCA

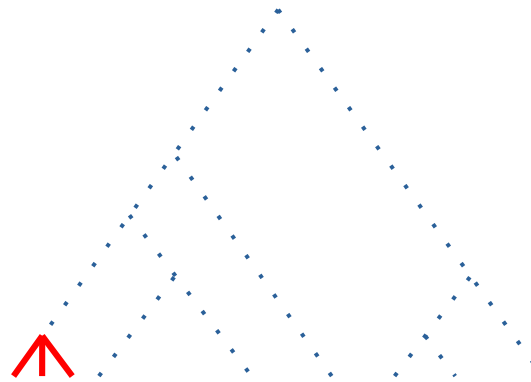


# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals

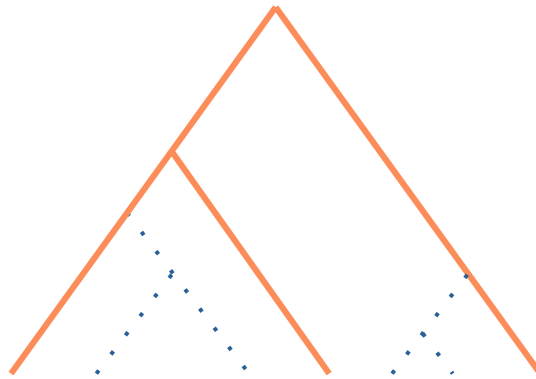
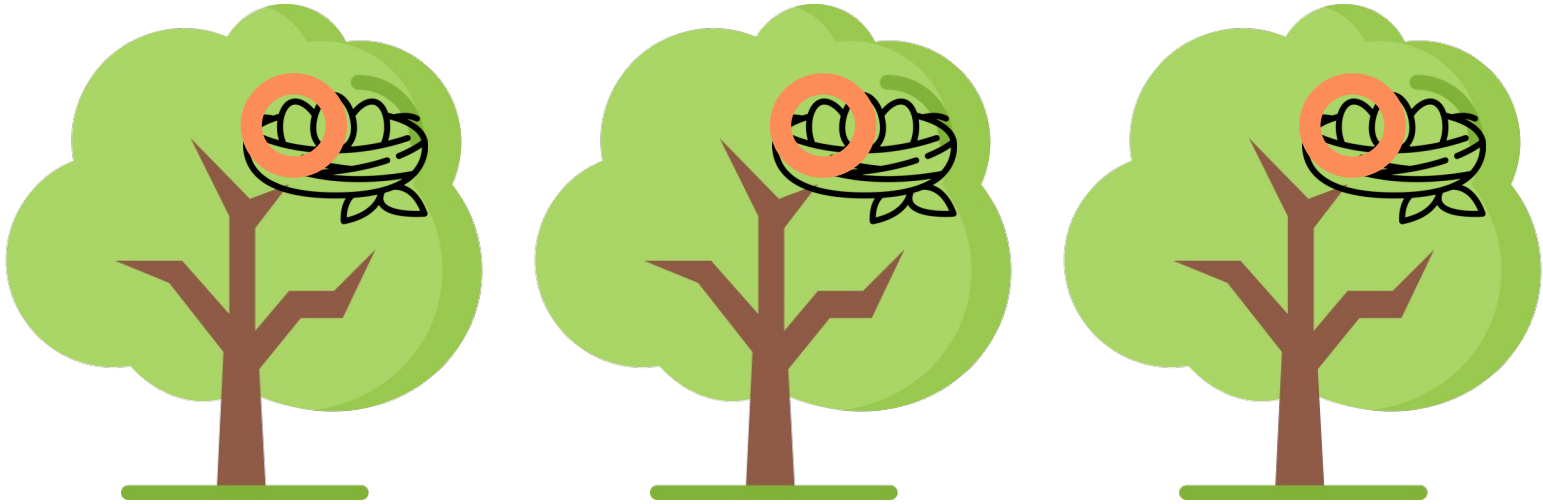


MRCA



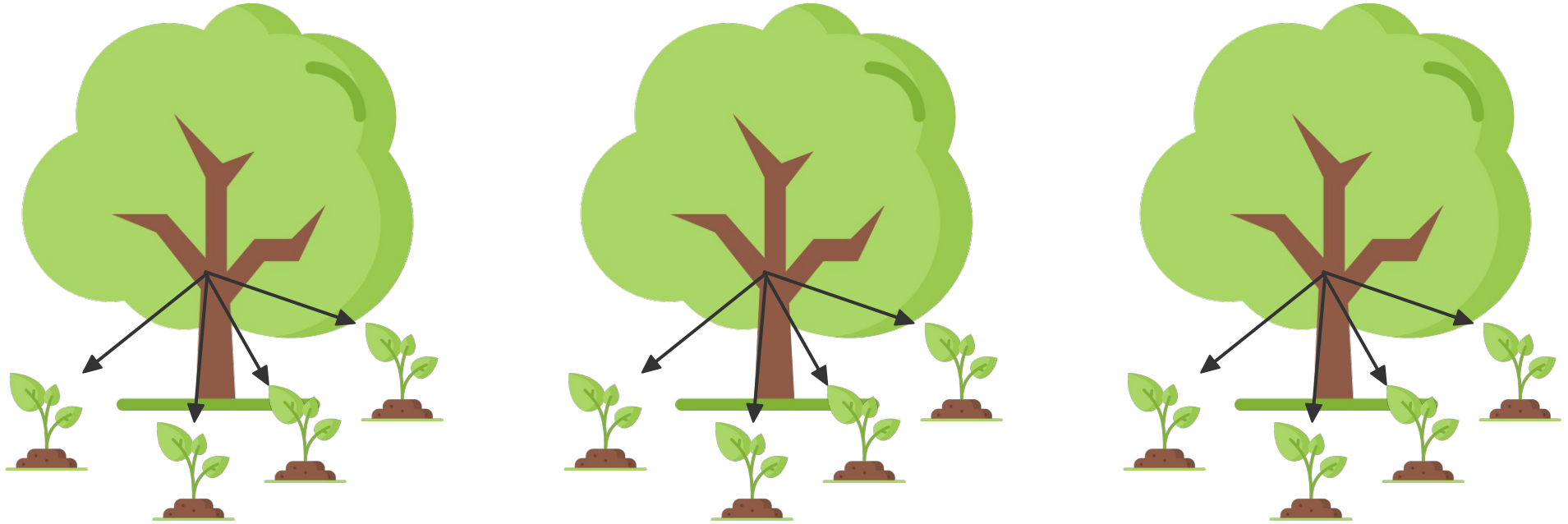
# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals



# 1. Sampling individuals from natural population(s)

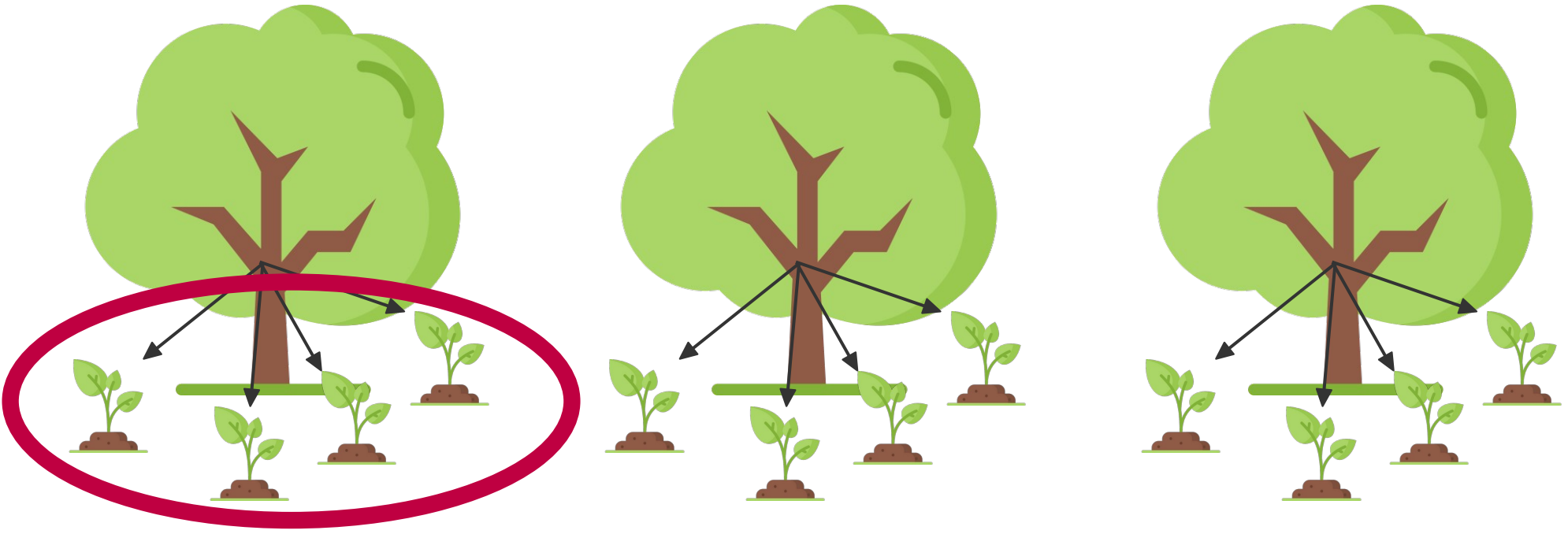
Samples must be a set of unrelated individuals





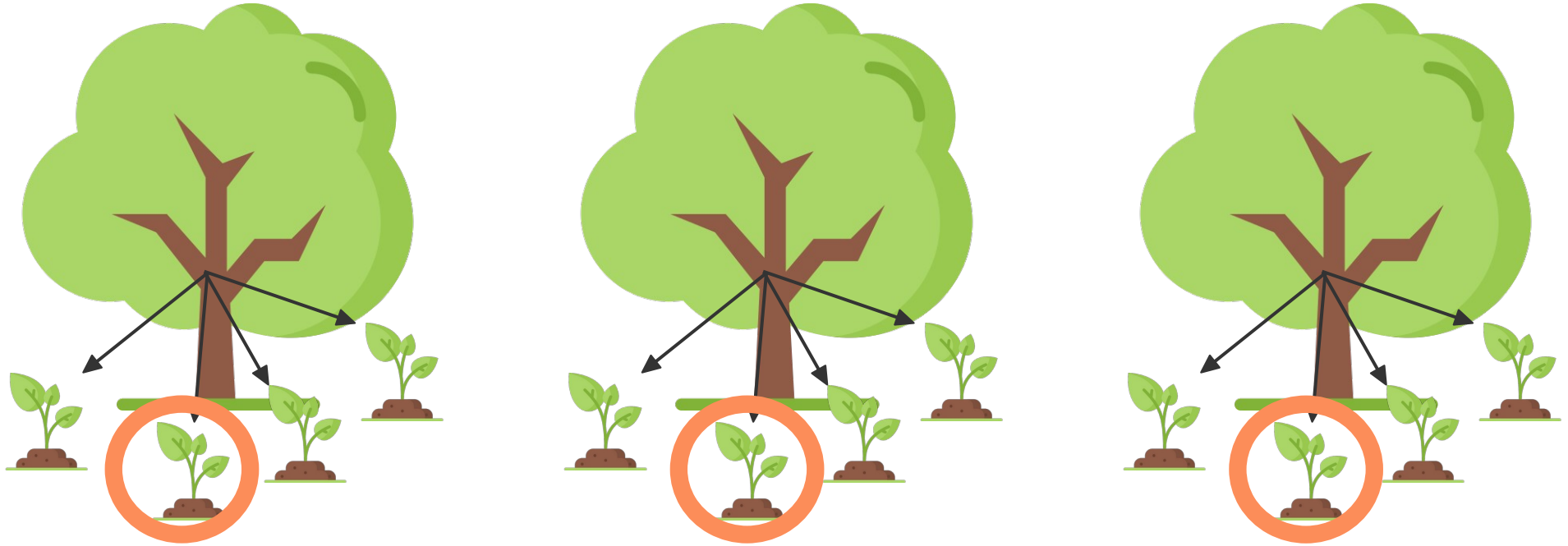
# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals



# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals



# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals

Generally : use between 10 and 50 individual chromosomes (5 to 25 diploid individuals)

# 1. Sampling individuals from natural population(s)

Samples must be a set of unrelated individuals

Generally : use between 10 and 50 individual chromosomes (5 to 25 diploid individuals)

Inference requires a set of sequences sampled from a single population

population = group of randomly mating individuals

# 1. Sampling individuals from natural population(s)

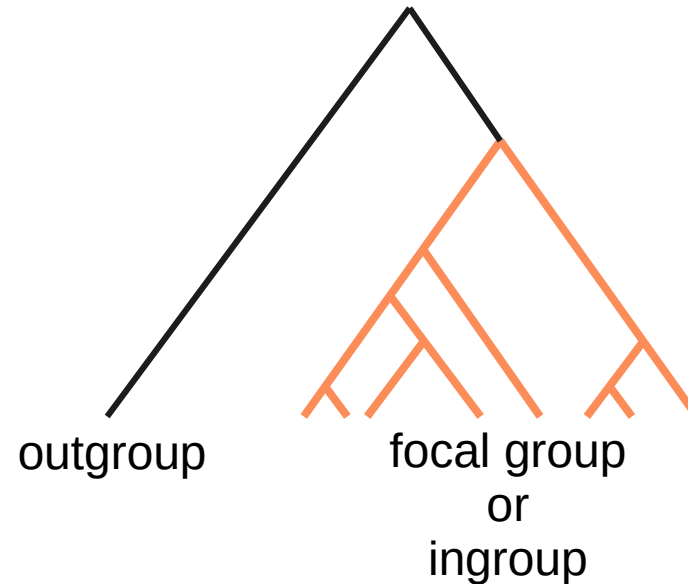
Samples must be a set of unrelated individuals

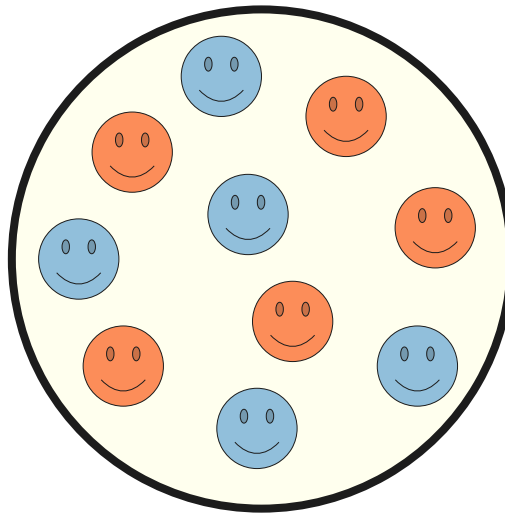
Generally : use between 10 and 50 individual chromosomes (5 to 25 diploid individuals)

Inference requires a set of sequences sampled from a single population

population = group of randomly mating individuals

For some methods : a sequenced outgroup is required



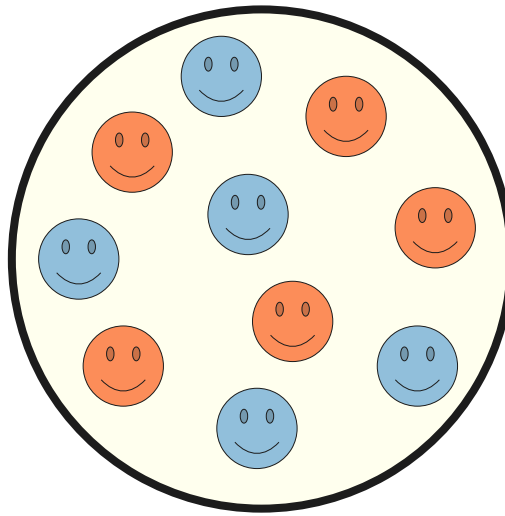


How can we measure that ?!?

1. Sampling individuals from natural population(s)



2. Sequencing them



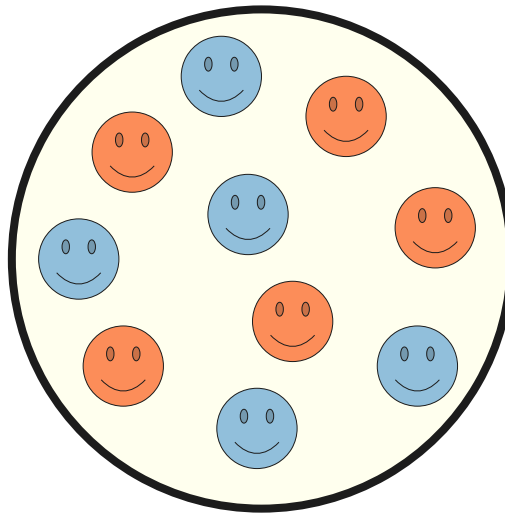
How can we measure that ?!?

1. Sampling individuals from natural population(s)



2. Sequencing them

How ?



How can we measure that ?!?

1. Sampling individuals from natural population(s)



2. Sequencing them



3. Analyse the sequences

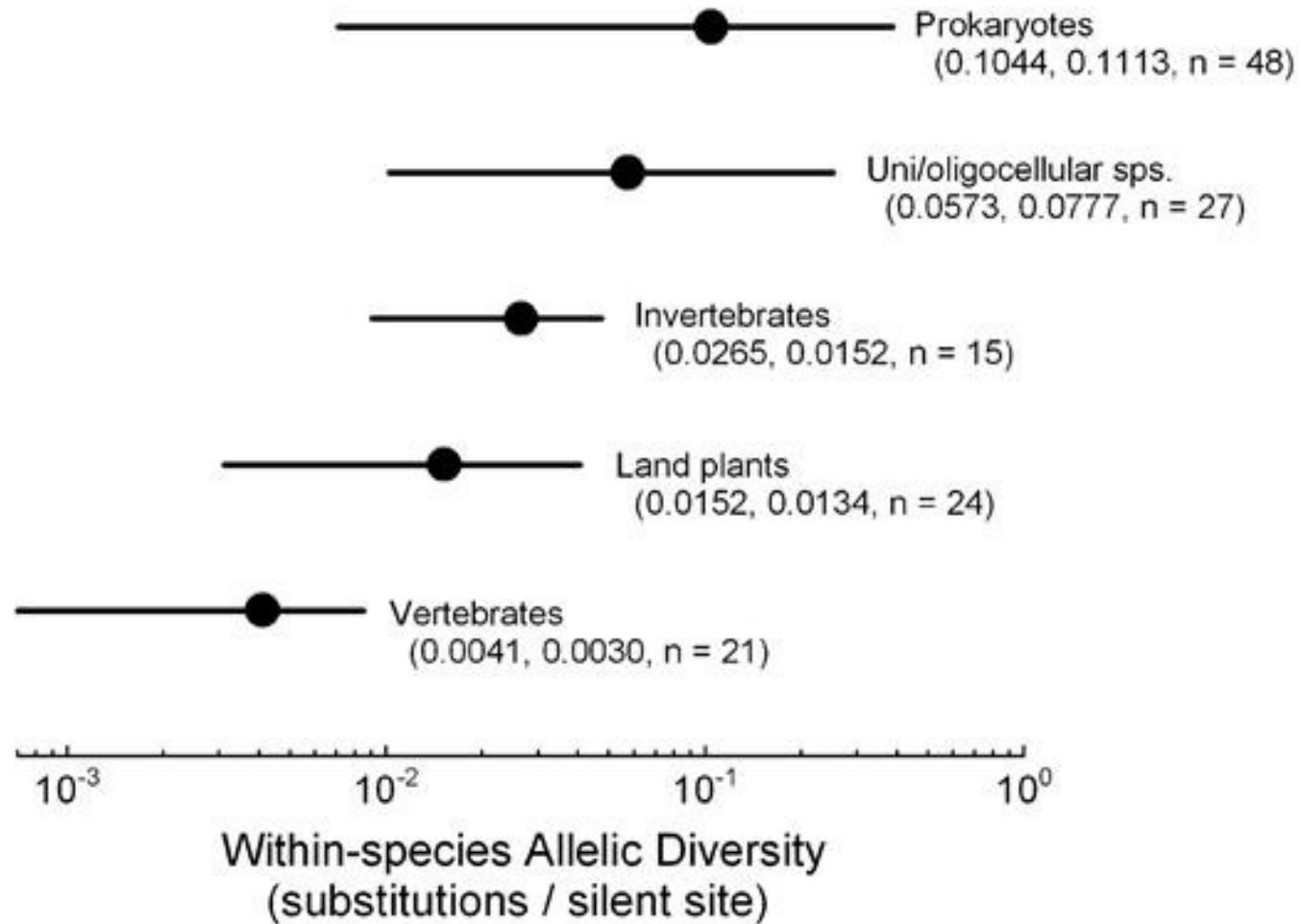


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$i_1$ :	T	T	A	C	A	A	T	C	C	G	A	T	C	G	T
$i_2$ :	T	T	A	C	G	A	T	G	C	G	C	T	C	G	T
$i_3$ :	T	C	A	C	A	A	T	G	C	G	A	T	G	G	A
$i_4$ :	T	T	A	C	G	A	T	G	C	G	C	T	C	G	T

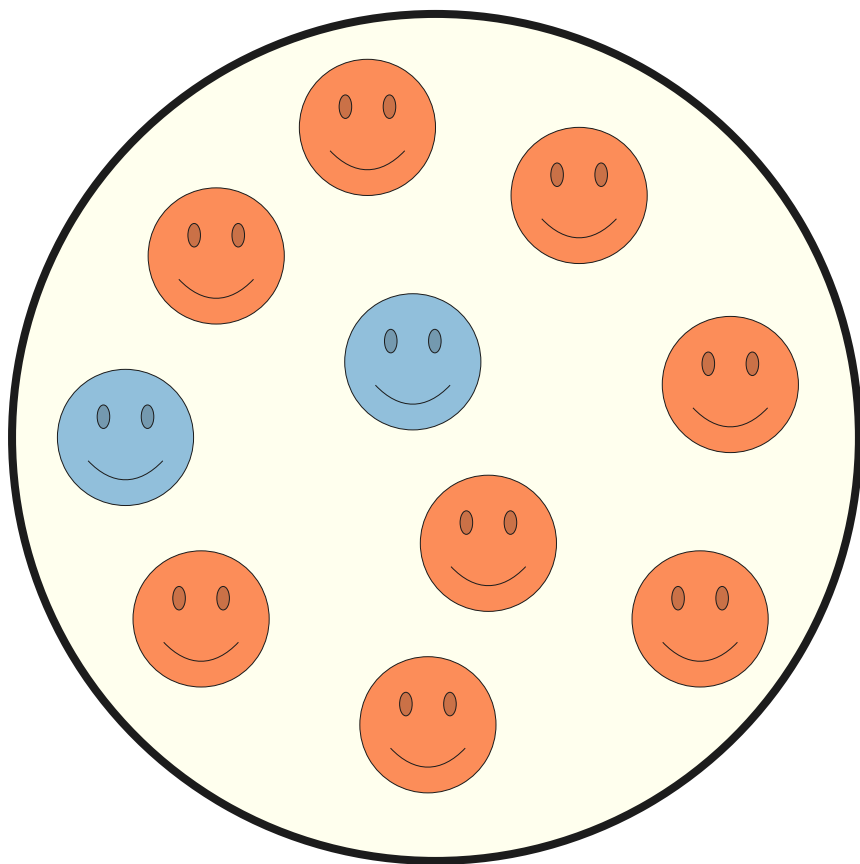
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$i_1$ :	T	<b>T</b>	A	C	<b>A</b>	A	T	<b>C</b>	C	G	<b>A</b>	T	<b>C</b>	G	<b>T</b>
$i_2$ :	T	<b>T</b>	A	C	<b>G</b>	A	T	<b>G</b>	C	G	<b>C</b>	T	<b>C</b>	G	<b>T</b>
$i_3$ :	T	<b>C</b>	A	C	<b>A</b>	A	T	<b>G</b>	C	G	<b>A</b>	T	<b>G</b>	G	<b>A</b>
$i_4$ :	T	<b>T</b>	A	C	<b>G</b>	A	T	<b>G</b>	C	G	<b>C</b>	T	<b>C</b>	G	<b>T</b>

With missing data ?

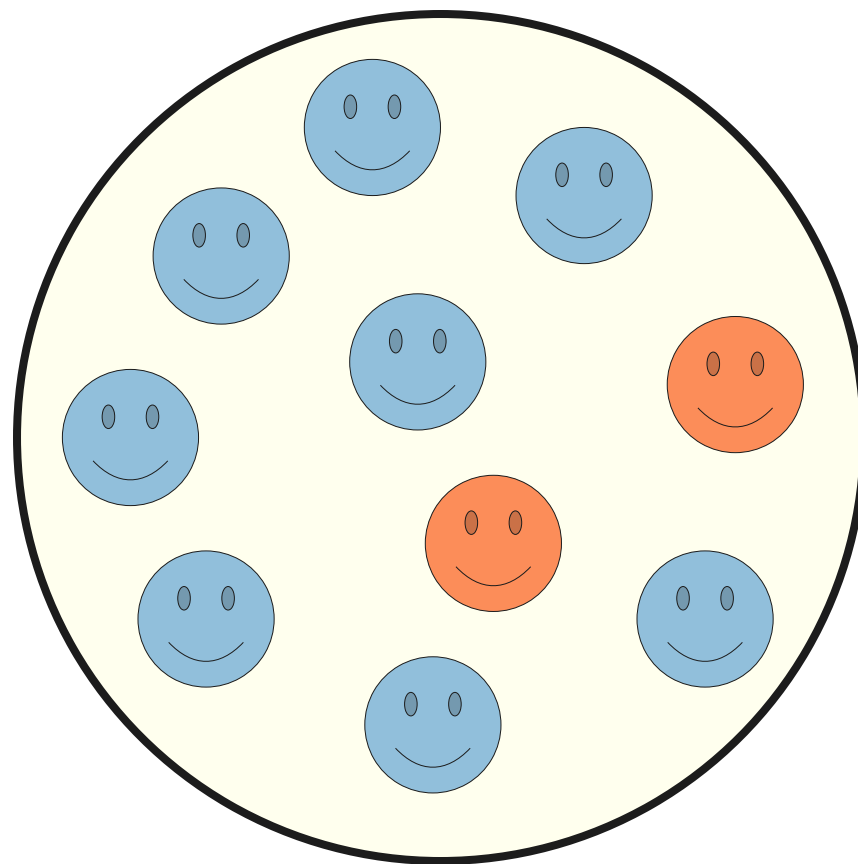
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$i_1$ :	T	T	A	C	A	A	T	C	C	G	A	T	C	G	T
$i_2$ :	-	-	A	C	G	A	T	G	C	G	C	T	C	G	T
$i_3$ :	T	C	A	C	A	A	T	G	C	G	A	T	G	G	A
$i_4$ :	T	T	A	C	G	A	T	G	-	-	C	T	C	G	T



# Population structure



Population 1

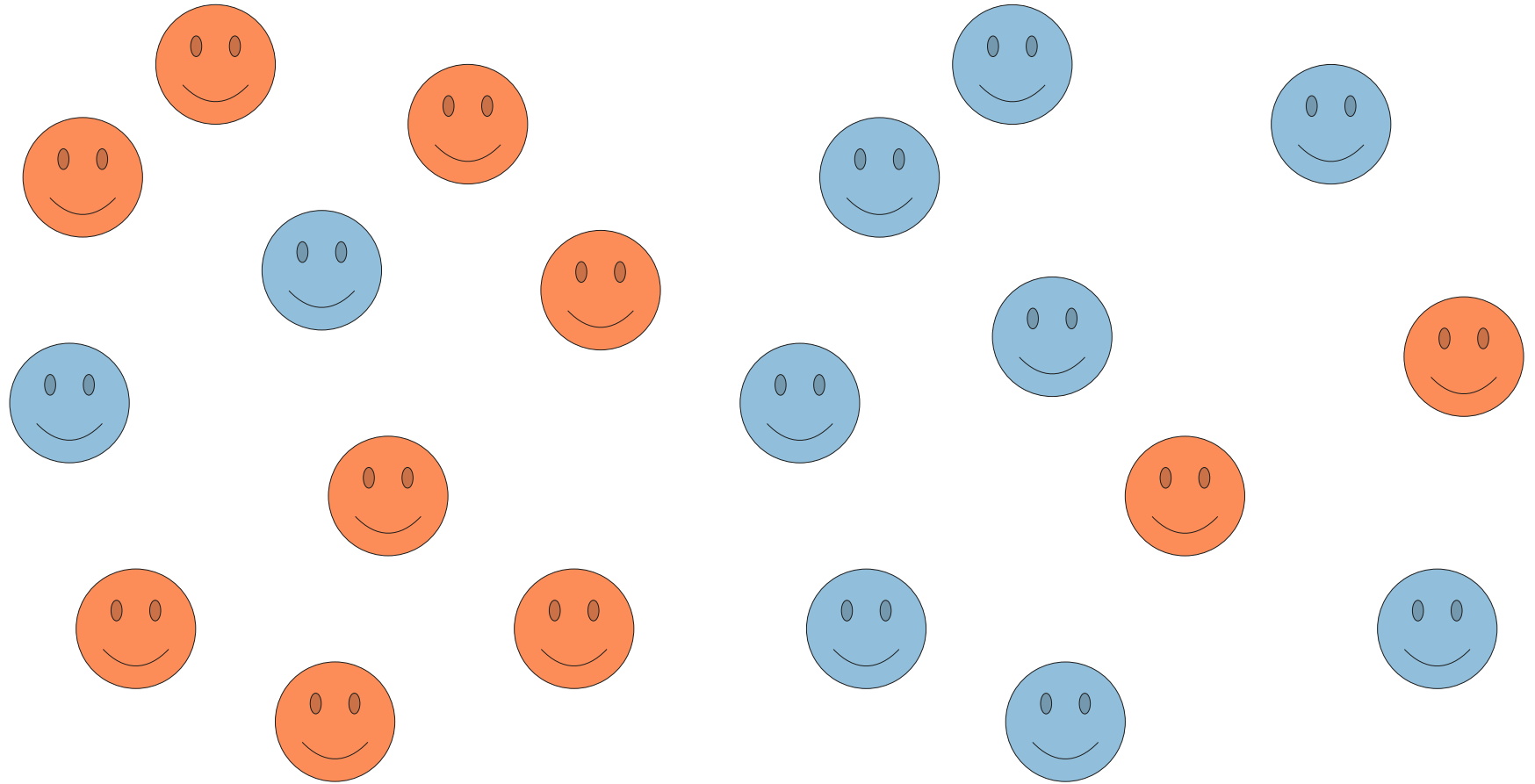


Population 2

How do we measure differences among populations ?

How do we define populations in practice?

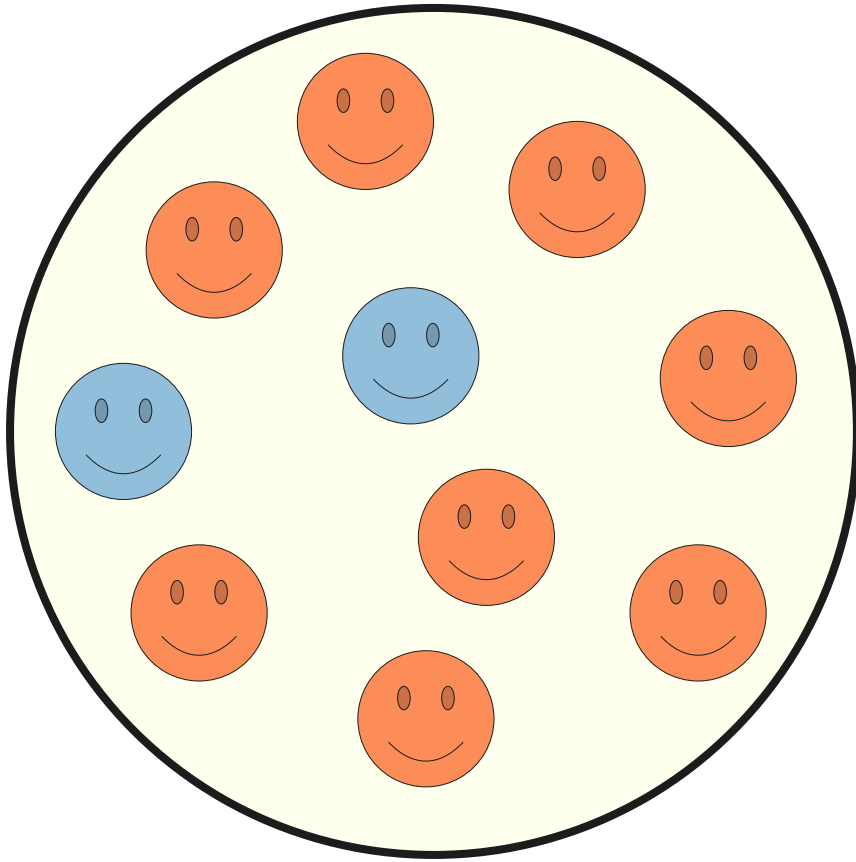
# Population structure



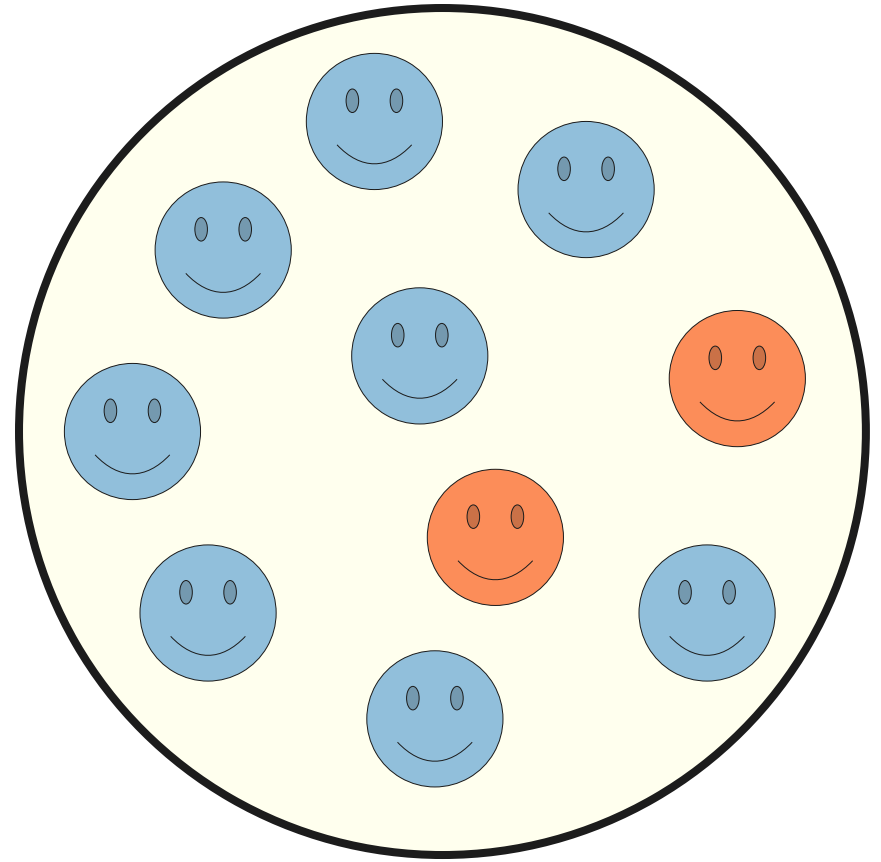
How do we measure differences among populations ?

How do we define populations in practice?

# The Wahlund effect



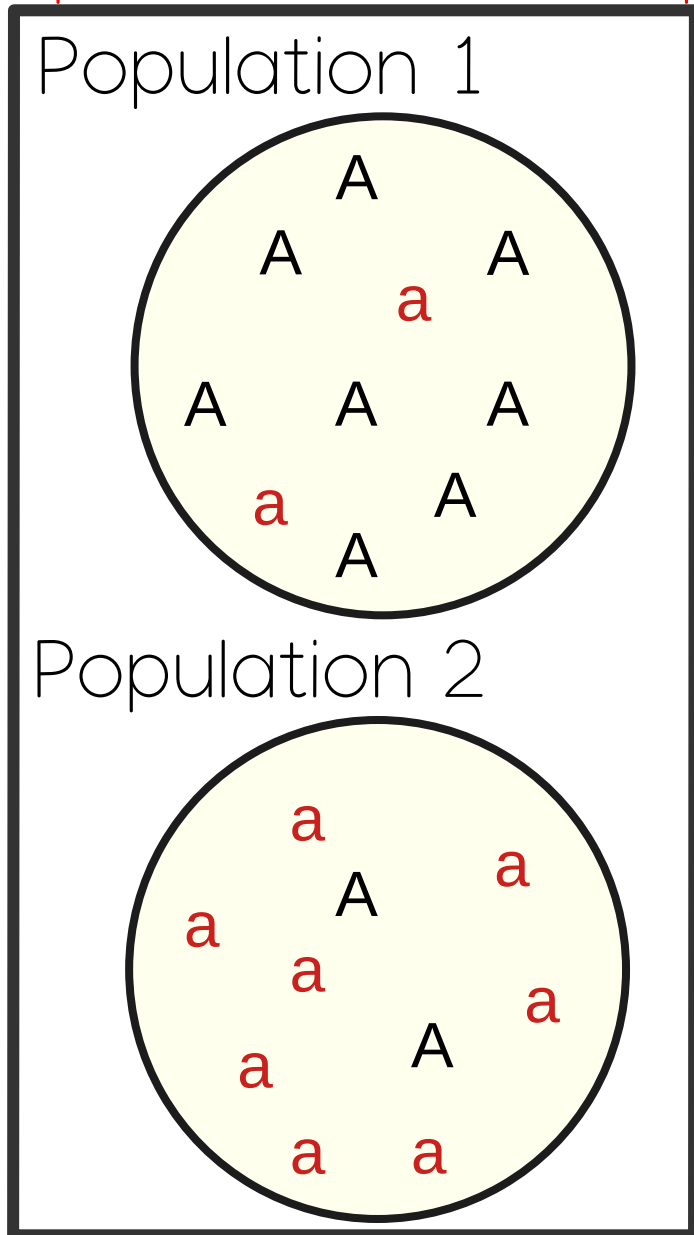
Population 1



Population 2

How different are genotypic frequencies (assuming HW) ?

The Wahlund effect  
If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies

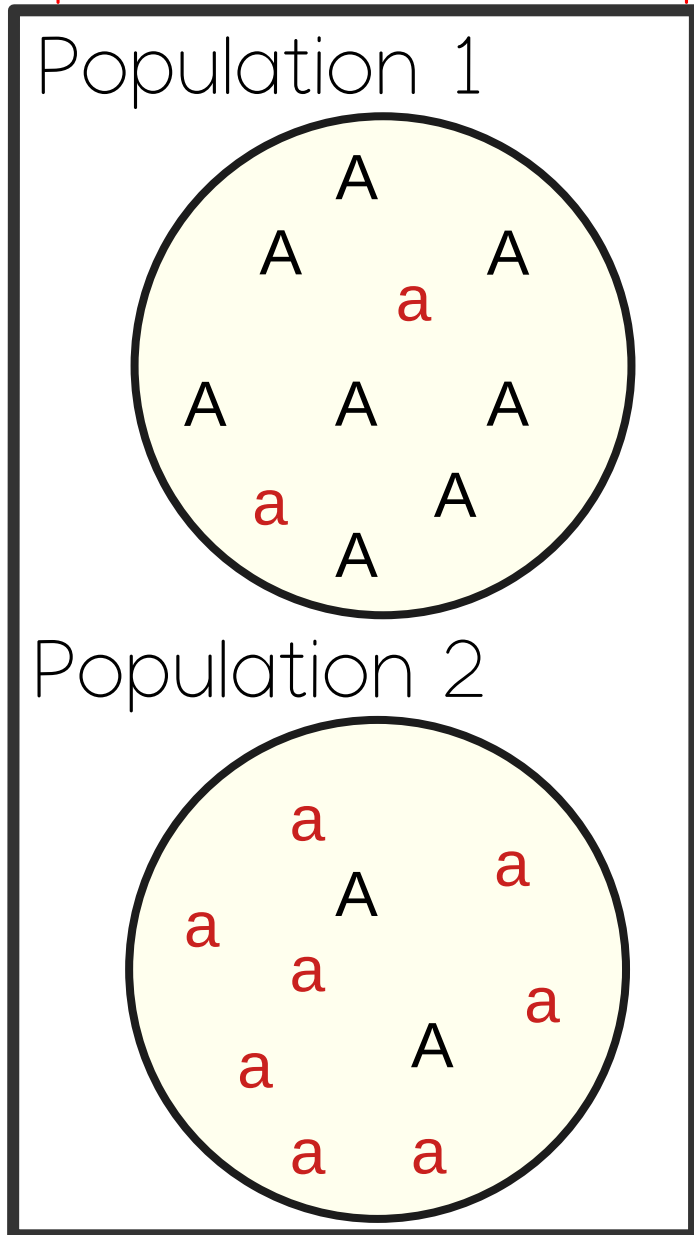


Total population



# The Wahlund effect

If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies

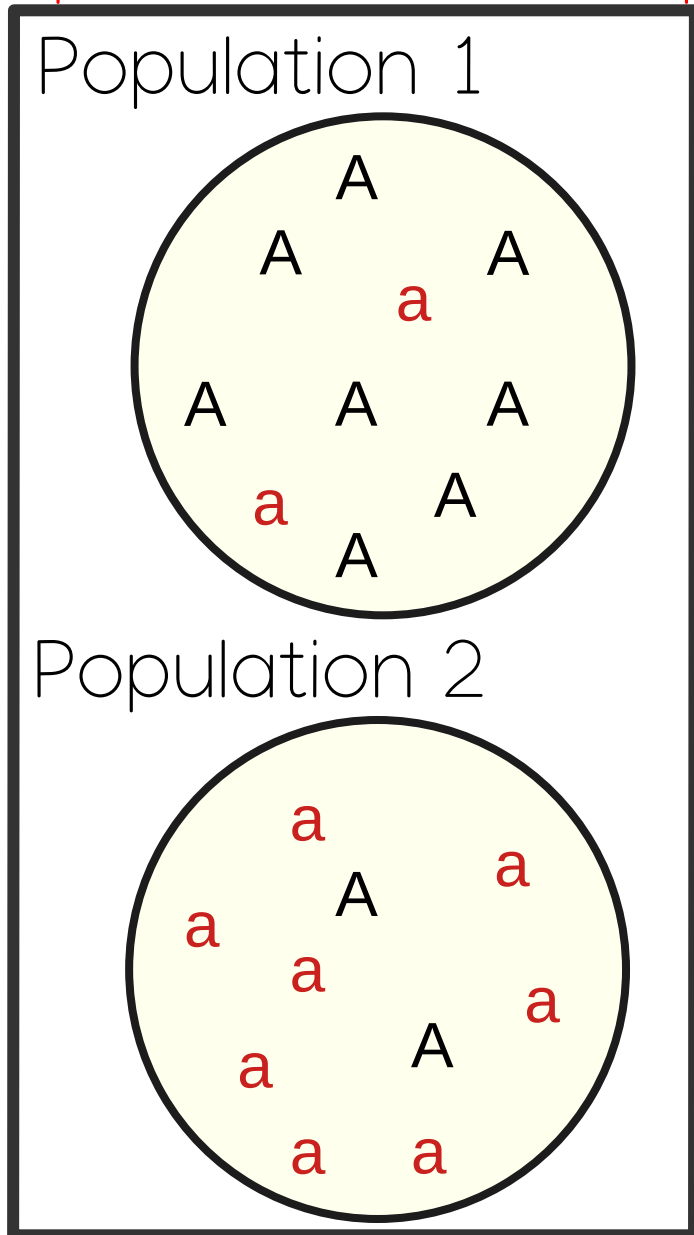


	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1			
Pop. 2			
Tot. Pop.			

Total population

# The Wahlund effect

If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies

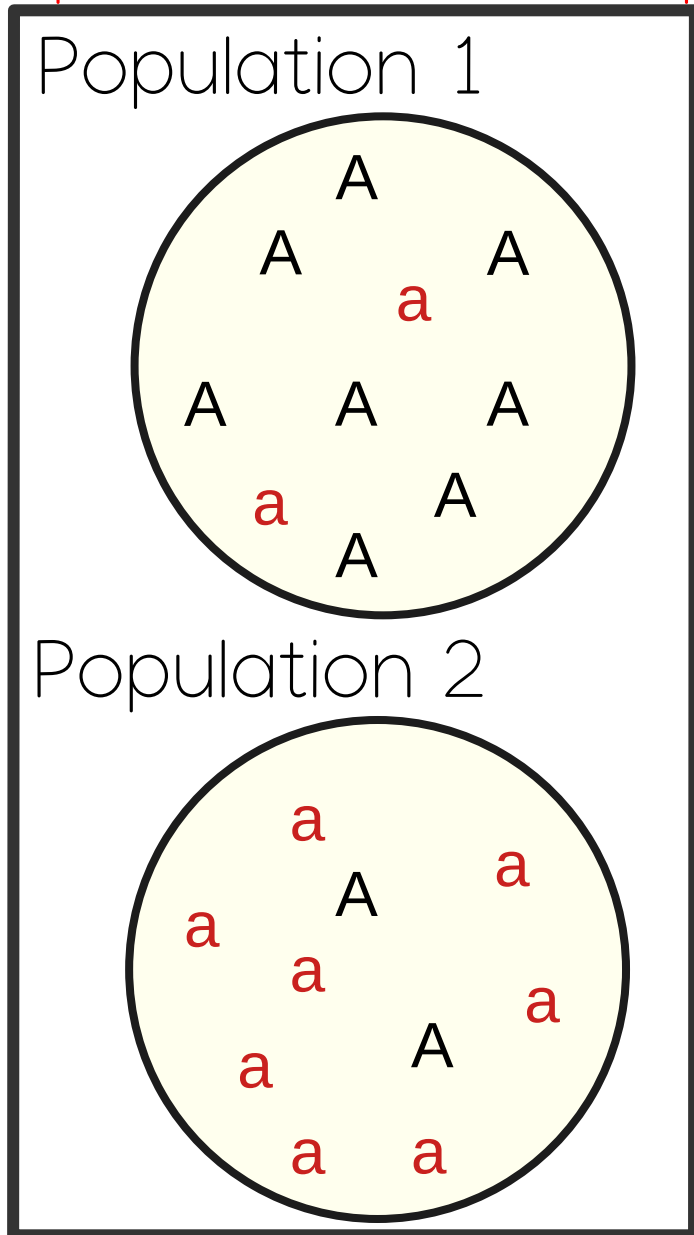


	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1	0.64	0.32	0.04
Pop. 2	0.04	0.32	0.64
Tot. Pop.	0.25	0.5	0.25

Total population

# The Wahlund effect

If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies

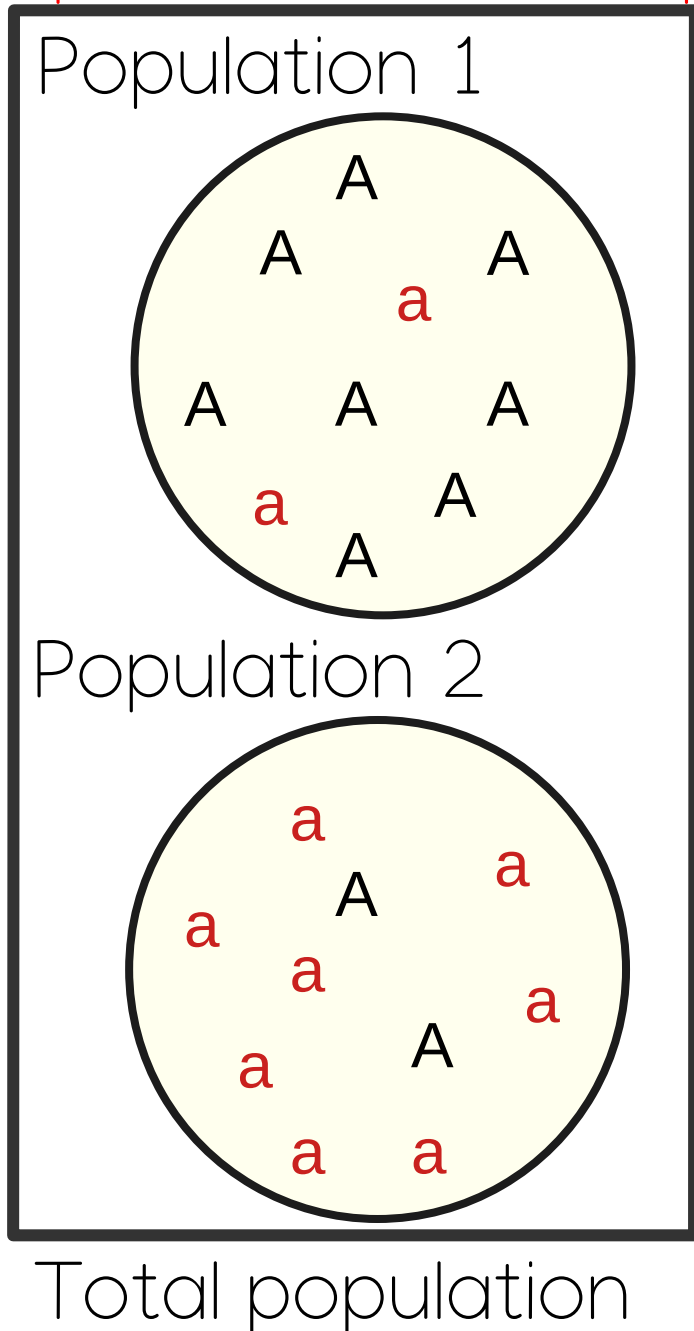


	f(AA)	f(Aa)	f(aa)
Pop. 1	0.64	0.32	0.04
Pop. 2	0.04	0.32	0.64
Tot. Pop.	0.25	0.5	0.25

50 % of individuals are expected to be heterozygous.

Total population

The Wahlund effect  
 If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies



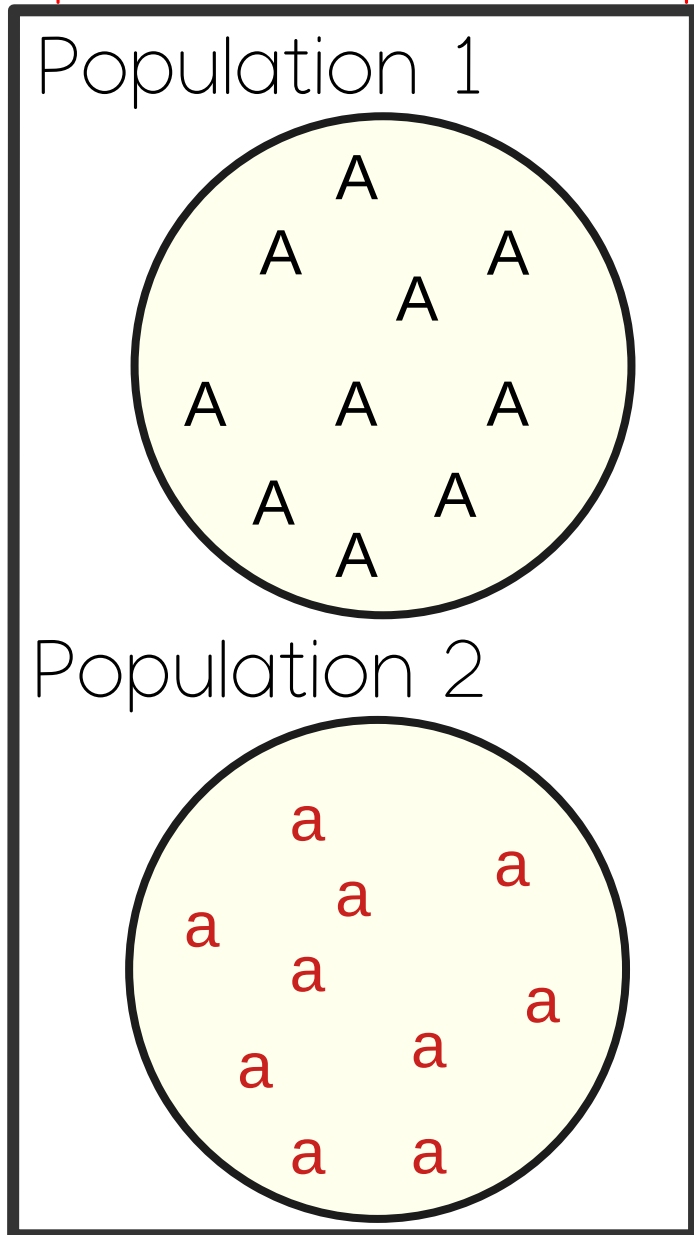
	f(AA)	f(Aa)	f(aa)
Pop. 1	0.64	0.32	0.04
Pop. 2	0.04	0.32	0.64
Tot. Pop.	0.25	0.5	0.25

50 % of individuals are expected to be heterozygous.  
 However, we observe 32 %.

$$\text{mean}(\text{pop}_1 ; \text{pop}_2) = f(\text{Aa})_{\text{Tot}} - 0.18$$

# The Wahlund effect

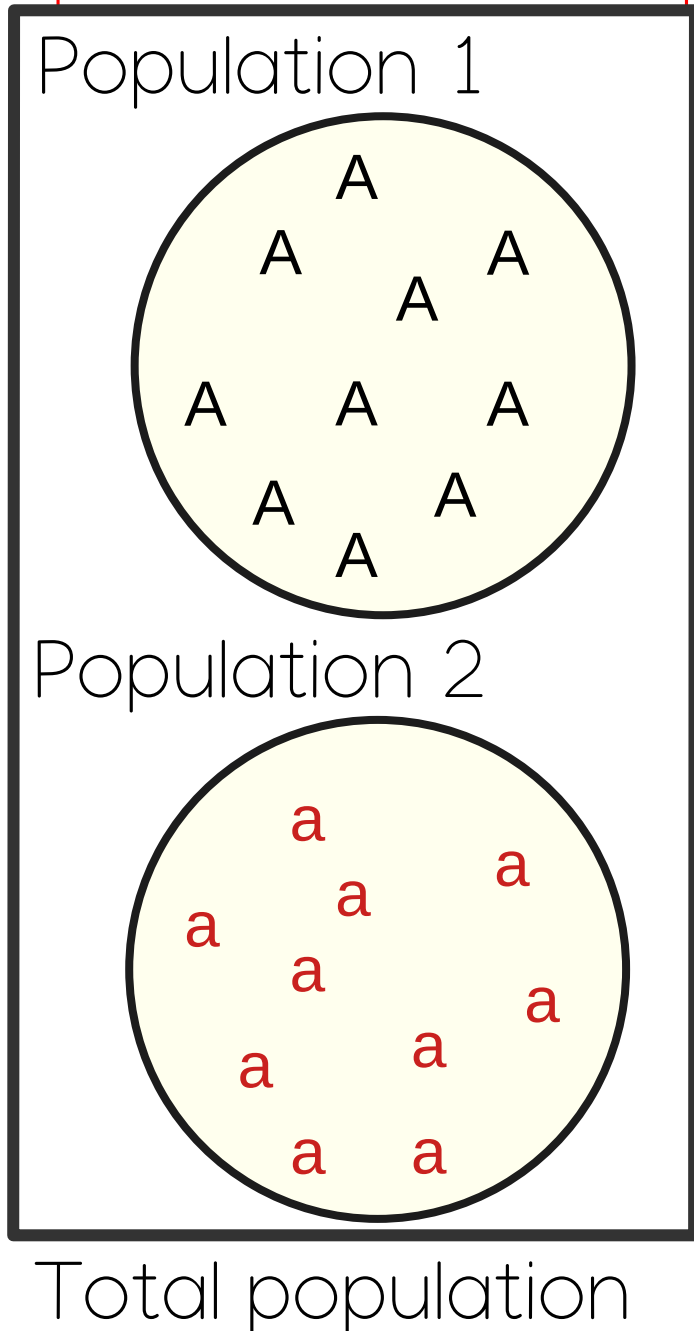
If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies



	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1			
Pop. 2			
Tot. Pop.			

Total population

The Wahlund effect  
 If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies

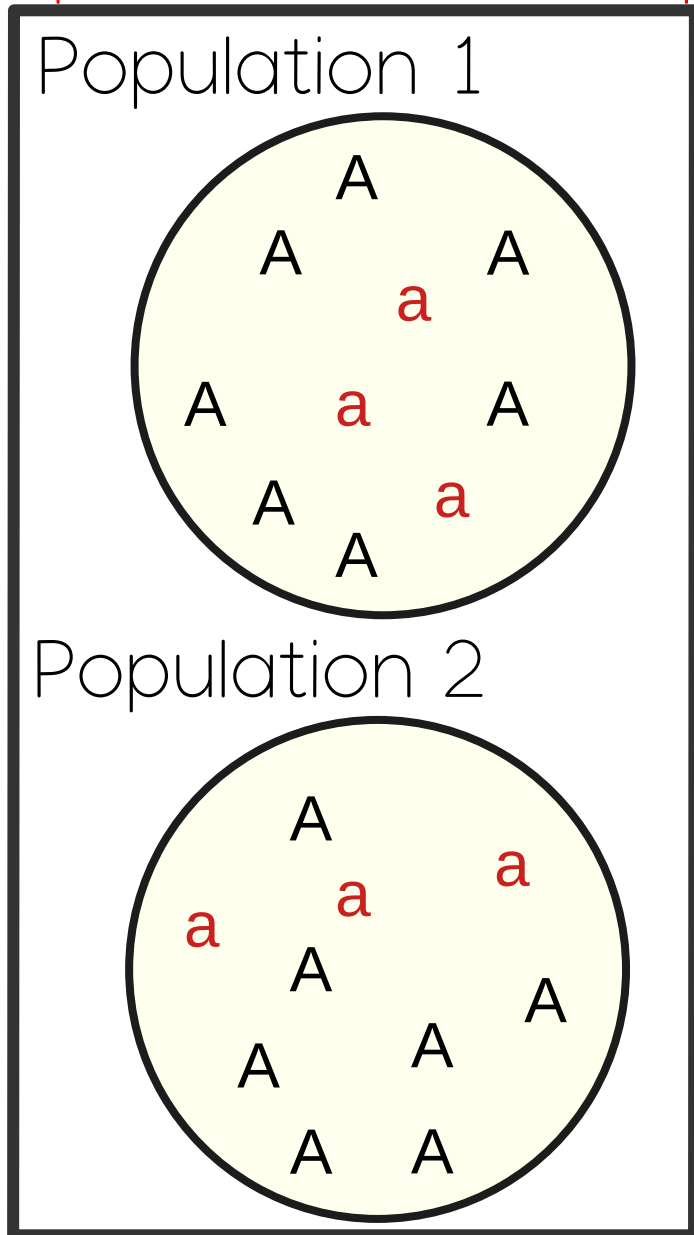


	f(AA)	f(Aa)	f(aa)
Pop. 1	1	0	0
Pop. 2	0	0	1
Tot. Pop.	0.25	0.5	0.25

50 % of individuals are expected to be heterozygous.  
 However, we observe 0%.

$$\text{mean}(\text{pop}_1 ; \text{pop}_2) = f(\text{Aa})_{\text{Tot}} - 0.5$$

The Wahlund effect  
If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies

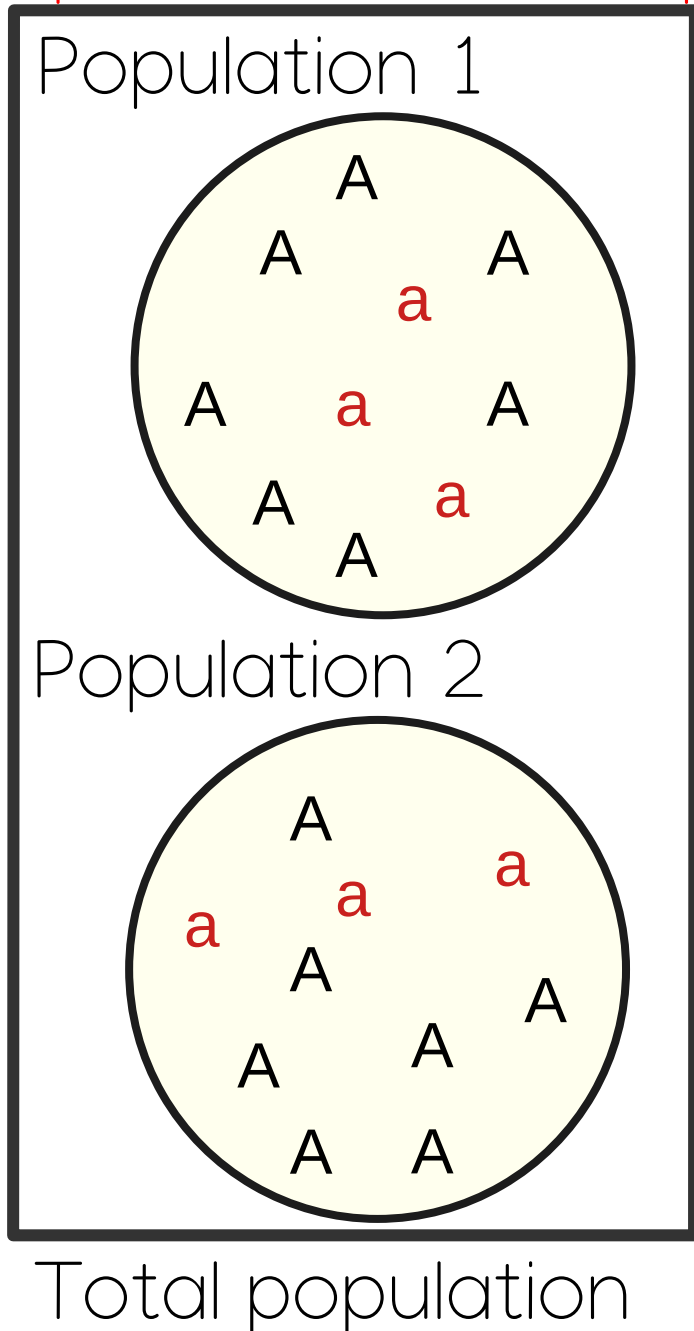


	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1			
Pop. 2			
Tot. Pop.			

Total population

# The Wahlund effect

If pop.1 and pop.2 are part from the same panmictic population → we expect similar genotypic frequencies



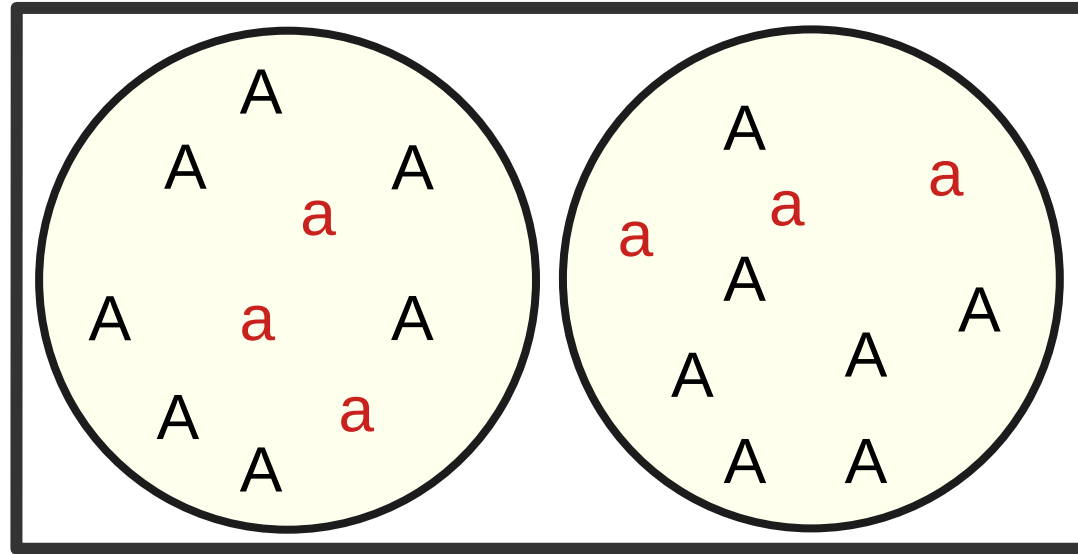
	f(AA)	f(Aa)	f(aa)
Pop. 1	0.49	0.42	0.09
Pop. 2	0.49	0.42	0.09
Tot. Pop.	0.49	0.42	0.09

42% of individuals are expected to be heterozygous.  
We effectively observe 42%.

$$\text{mean}(\text{pop}_1 ; \text{pop}_2) = f(\text{Aa})_{\text{Tot}} - 0$$

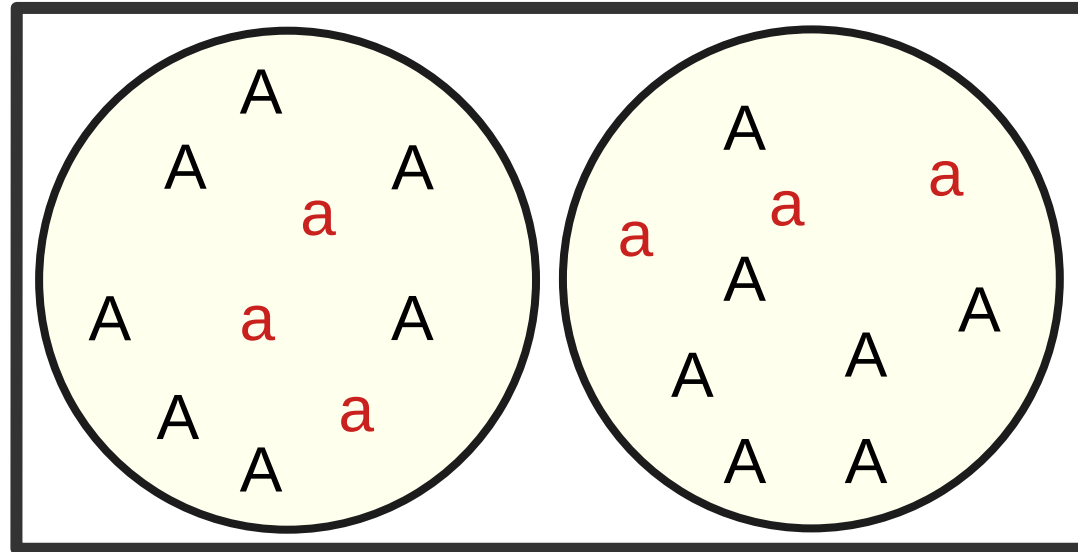


# The Wahlund effect

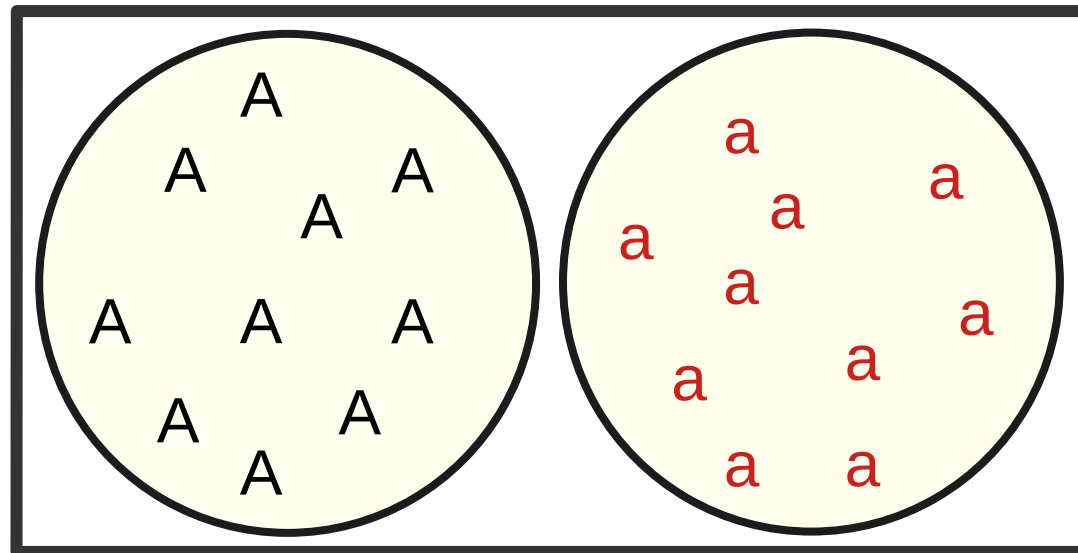


Populations with similar allele frequencies → show very small deviations from HWE.

# The Wahlund effect

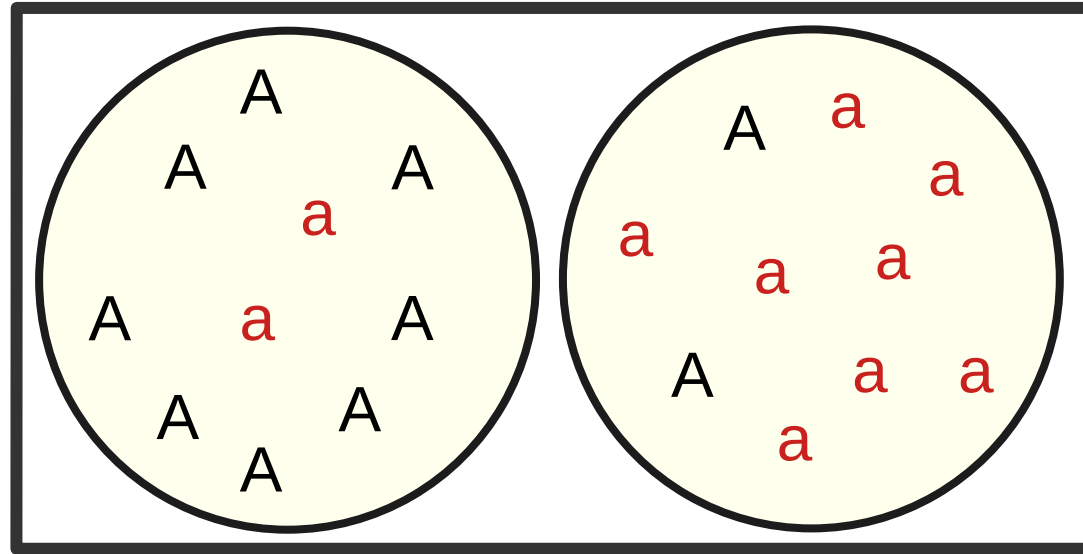


Populations with similar allele frequencies → show very small deviations from HWE.



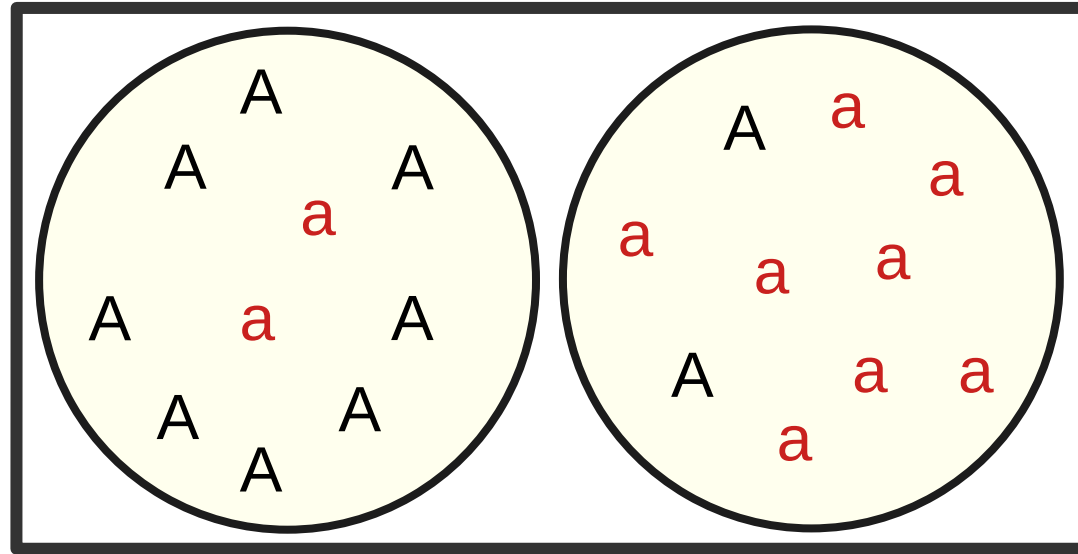
Populations with different allele frequencies → show strong deviations from HWE.

# The Wahlund effect



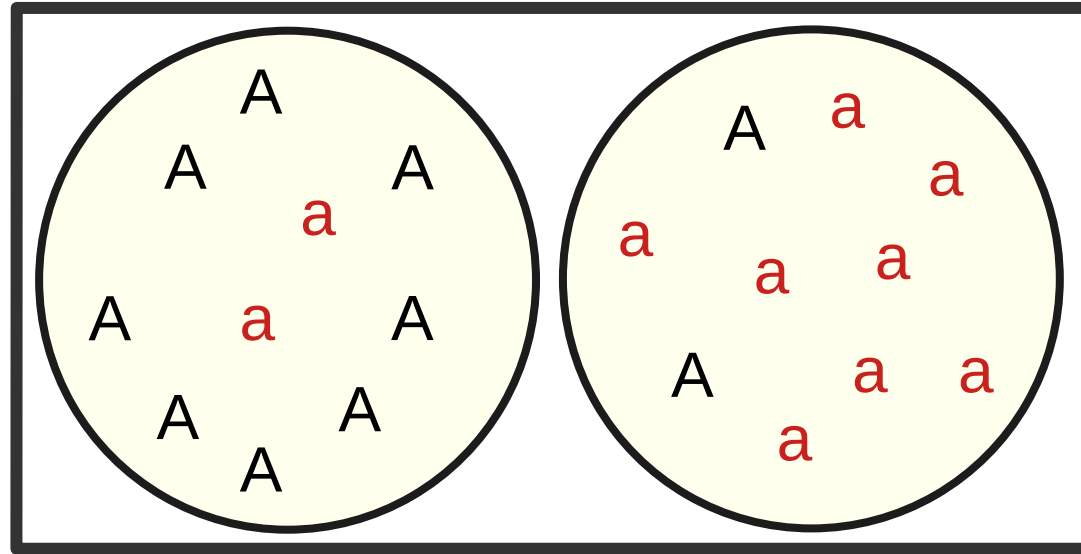
Useful measure of allele frequency differences among populations

# The Wahlund effect



Useful measure of allele frequency differences  
among populations → variance in frequencies  $\sigma^2$

# The Wahlund effect



Useful measure of allele frequency differences among populations → variance in frequencies  $\sigma^2$

$$\frac{1}{n} \cdot \sum_{i=1}^n (p_i - \bar{p})^2$$

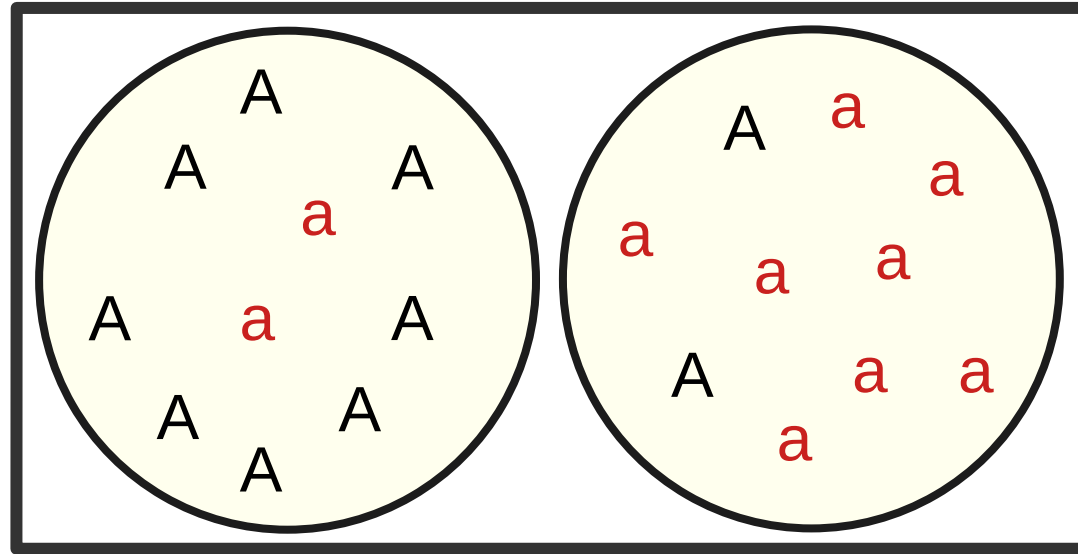
Where :

$n$  is the number of populations

$p_i$  is the frequency of allele A (or a) in population  $i$ .

$\bar{p}$  is the frequency of allele A (or a) in the whole metapopulation.

# The Wahlund effect



Useful measure of allele frequency differences among populations → variance in frequencies  $\sigma^2$

$$\frac{1}{n} \cdot \sum_{i=1}^n (p_i - \bar{p})^2$$

Where :

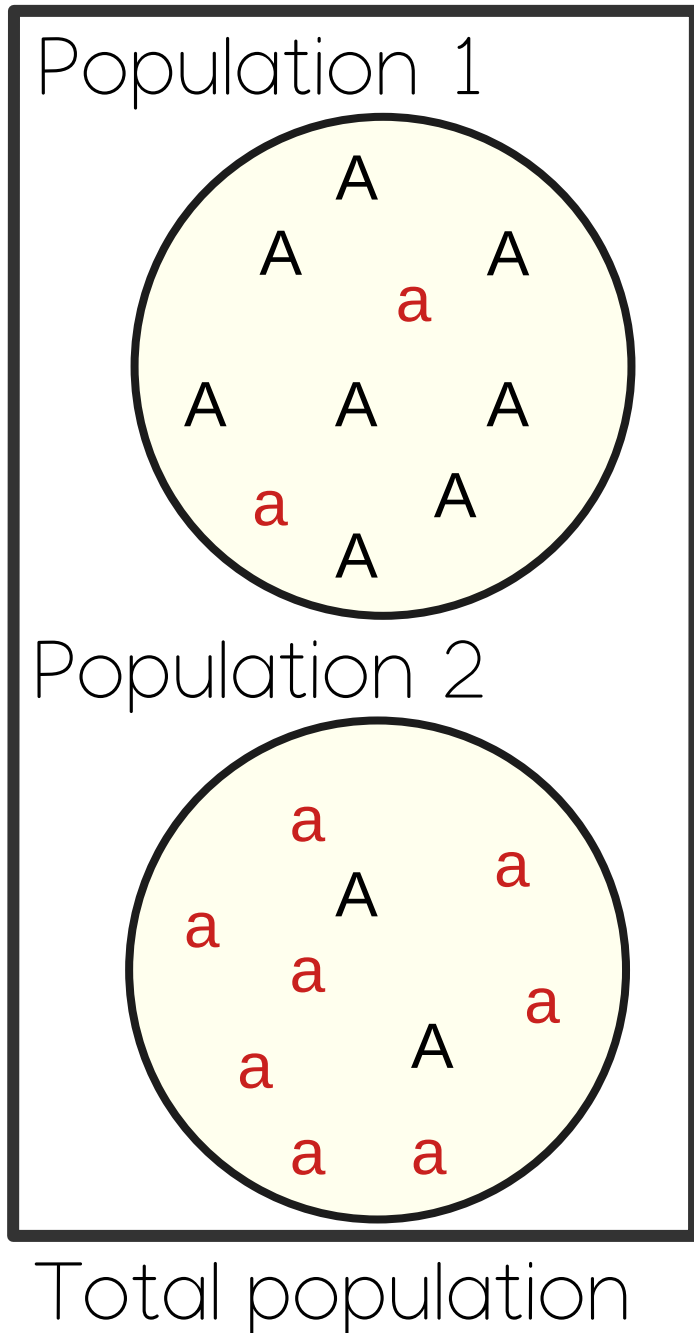
$n$  is the number of populations

$p_i$  is the frequency of allele A (or a) in population  $i$ .

$\bar{p}$  is the frequency of allele A (or a) in the whole metapopulation.

$$\sigma^2 = (1/2) * [(0.8-0.5)^2 + (0.2-0.5)^2] = 0.09$$

# The Wahlund effect



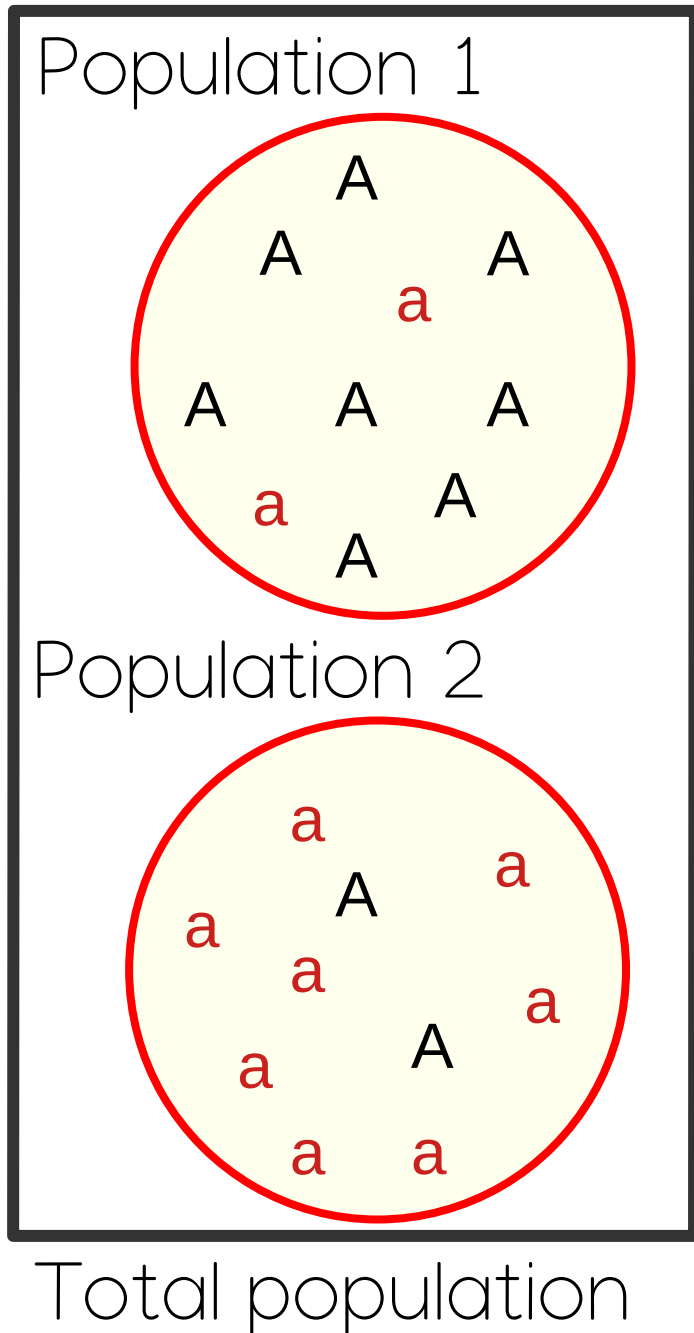
	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1	0.64	0.32	0.04
Pop. 2	0.04	0.32	0.64
Tot. Pop.	0.25	0.5	0.25

$$E[f(AA)] = \bar{p}^2 + \sigma^2$$

$$E[f(Aa)] = 2\bar{p}\bar{q} - 2\sigma^2$$

$$E[f(aa)] = \bar{q}^2 + \sigma^2$$

# The Wahlund effect



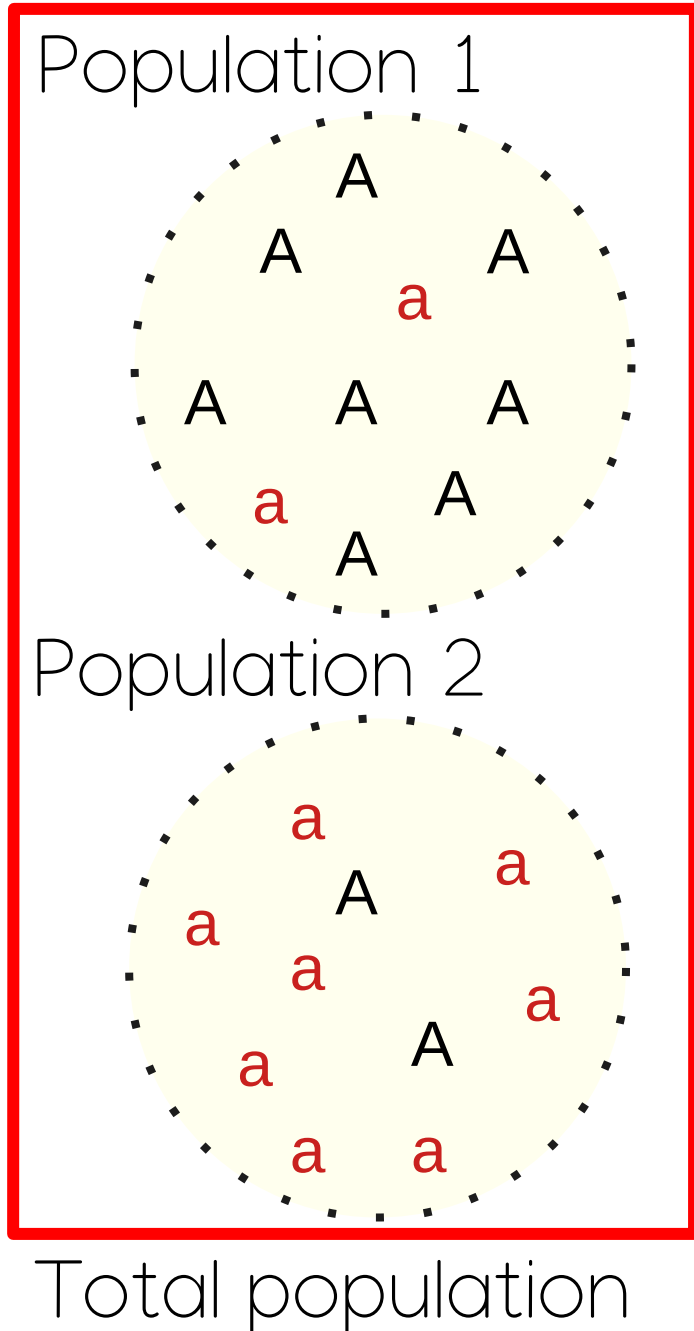
	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1	0.64	0.32	0.04
Pop. 2	0.04	0.32	0.64
Tot. Pop.	0.25	0.5	0.25

$$\begin{aligned}
 E[f(AA)] &= \bar{p}^2 + \sigma^2 \\
 E[f(Aa)] &= 2\bar{p}\bar{q} - 2\sigma^2 \\
 E[f(aa)] &= \bar{q}^2 + \sigma^2
 \end{aligned}$$

Expected genotypic frequencies within populations



# The Wahlund effect

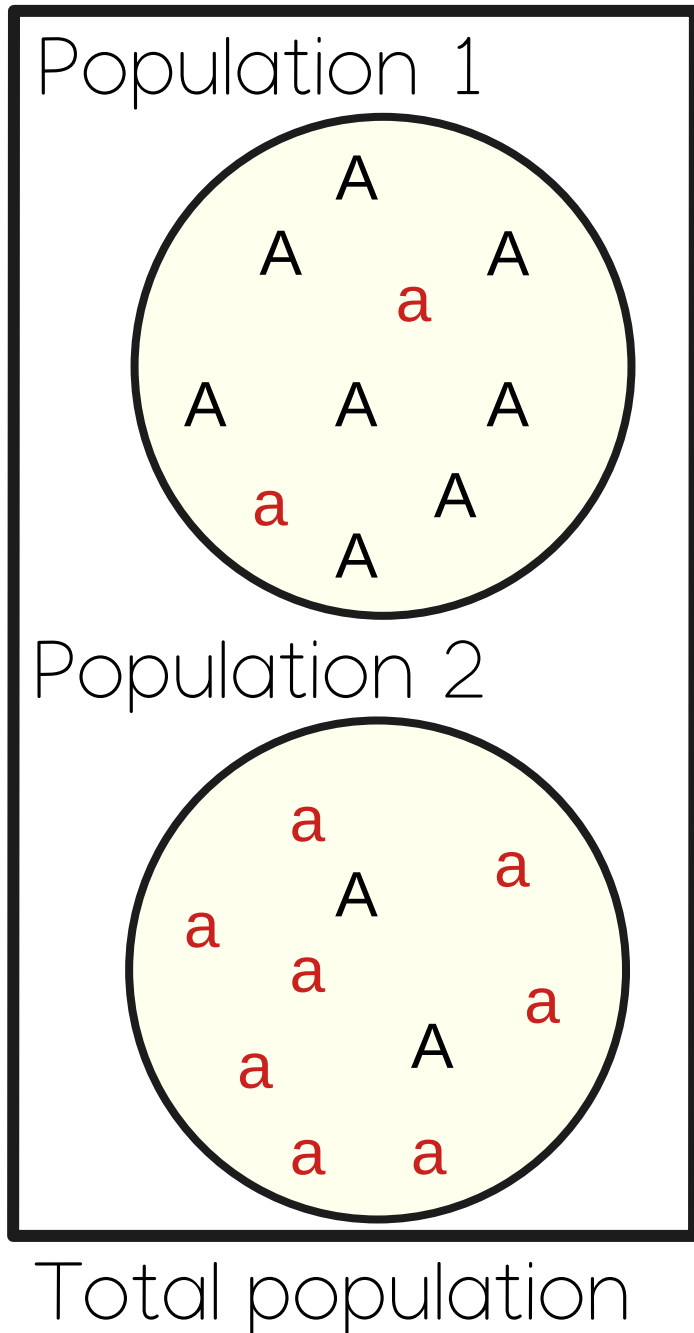


	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1	0.64	0.32	0.04
Pop. 2	0.04	0.32	0.64
Tot. Pop.	0.25	0.5	0.25

$$\begin{aligned}
 E[f(AA)] &= \bar{p}^2 + \sigma^2 \\
 E[f(Aa)] &= 2\bar{p}\bar{q} - 2\sigma^2 \\
 E[f(aa)] &= \bar{q}^2 + \sigma^2
 \end{aligned}$$

Expected genotypic frequencies in total

# The Wahlund effect

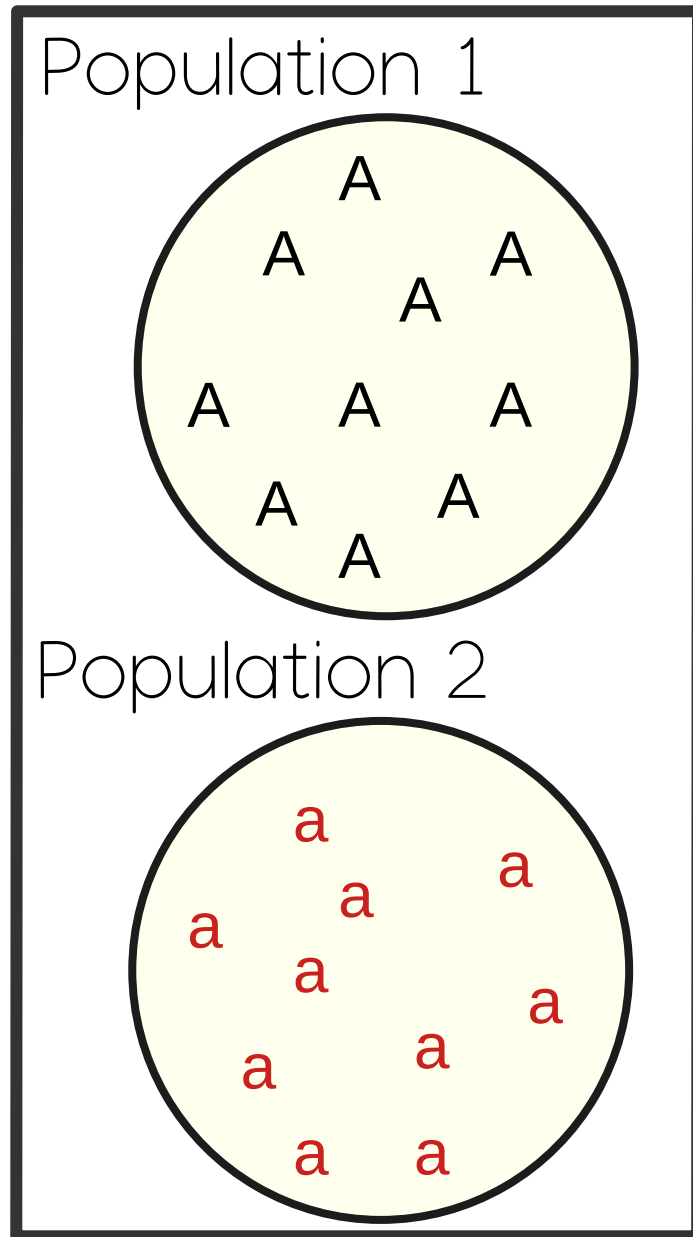


	$f(AA)$	$f(Aa)$	$f(aa)$
Pop. 1	0.64	0.32	0.04
Pop. 2	0.04	0.32	0.64
Tot. Pop.	0.25	0.5	0.25

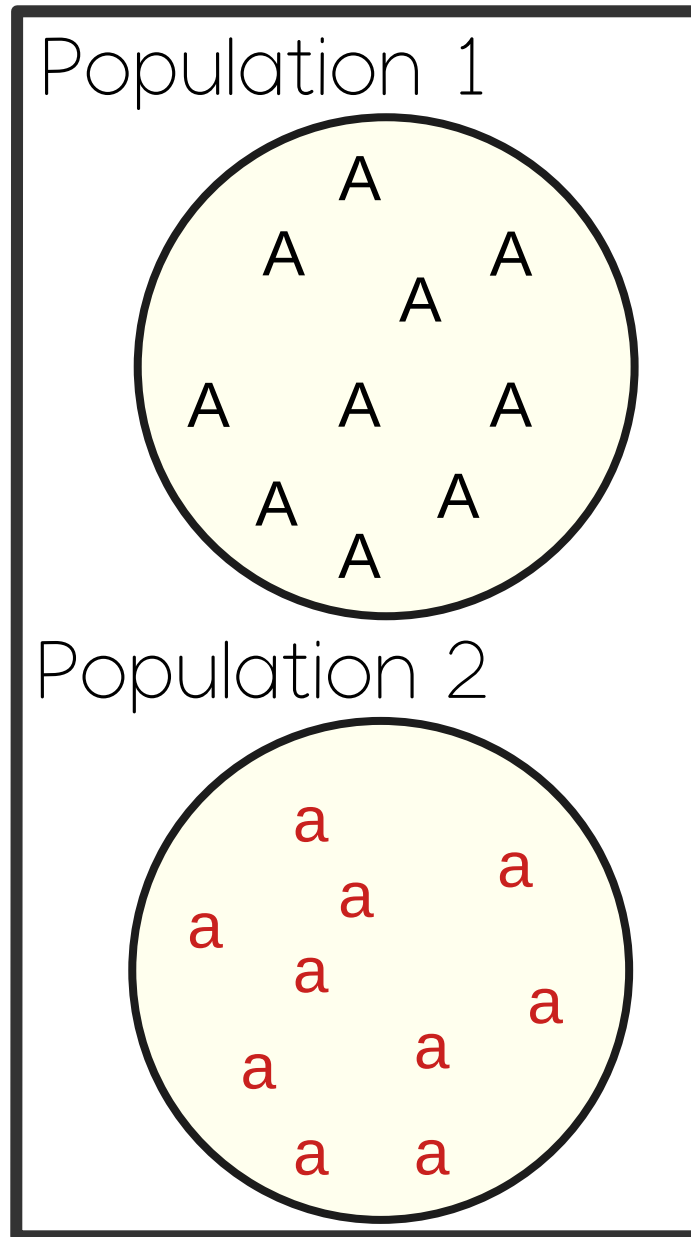
$$\begin{aligned}
 E[f(AA)] &= \bar{p}^2 + \sigma^2 \\
 E[f(Aa)] &= 2\bar{p}\bar{q} - 2\sigma^2 \\
 E[f(aa)] &= \bar{q}^2 + \sigma^2
 \end{aligned}$$

Variance in allele frequencies

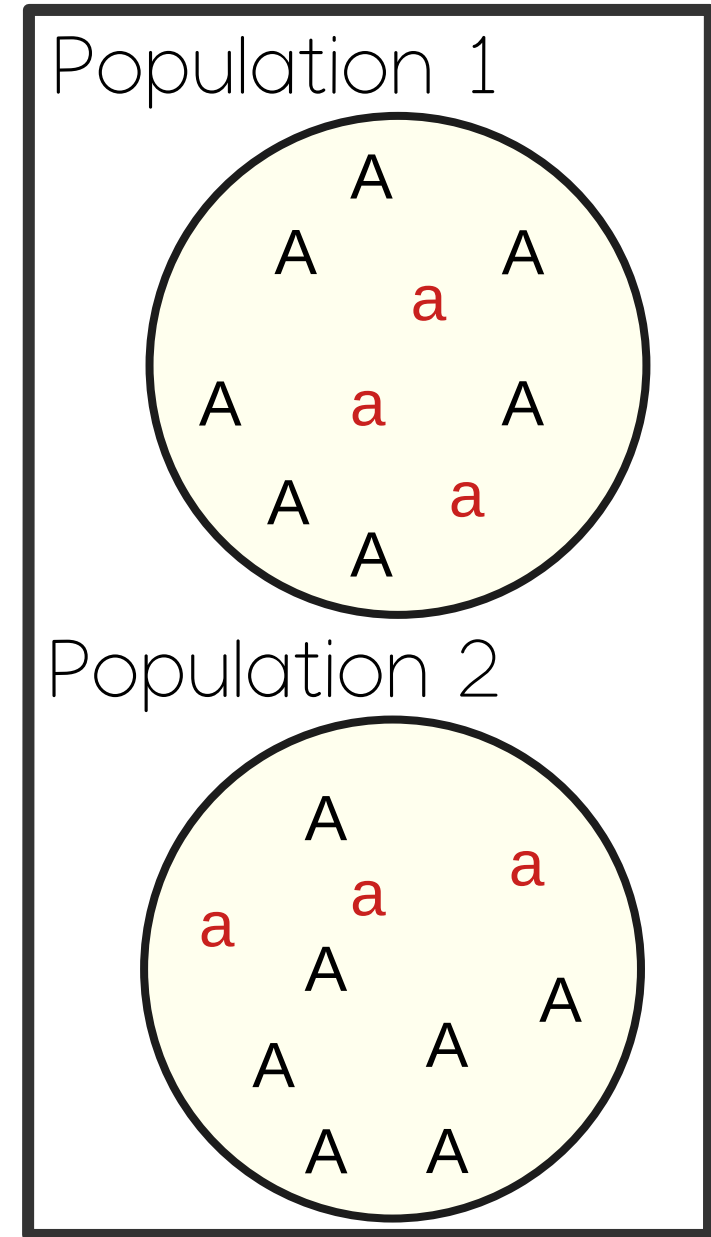
Variance in allelic frequencies ( $\sigma^2$ ) measures the genetic differentiation



Variance in allelic frequencies ( $\sigma^2$ ) measures the genetic differentiation



$\sigma^2$   
>>>

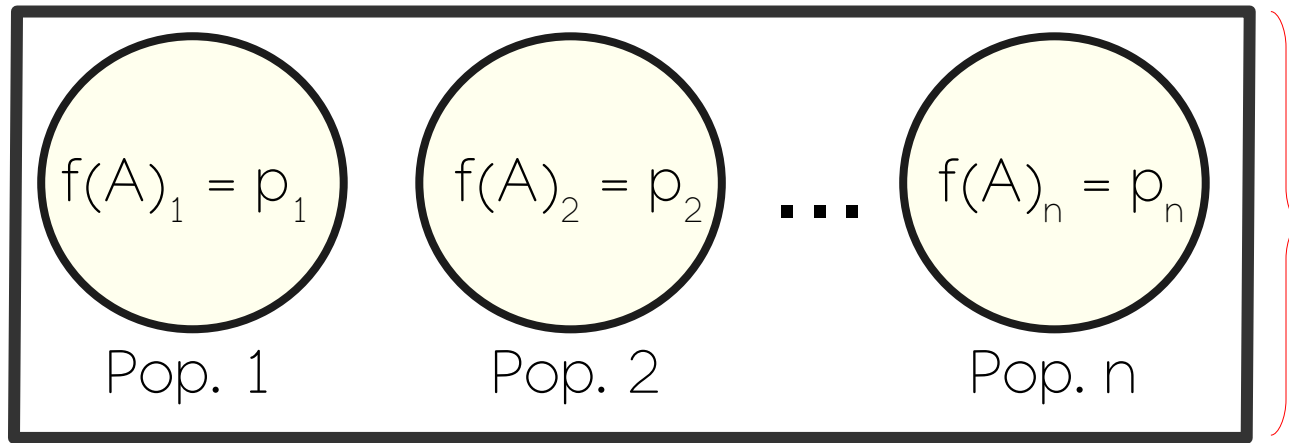


We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )

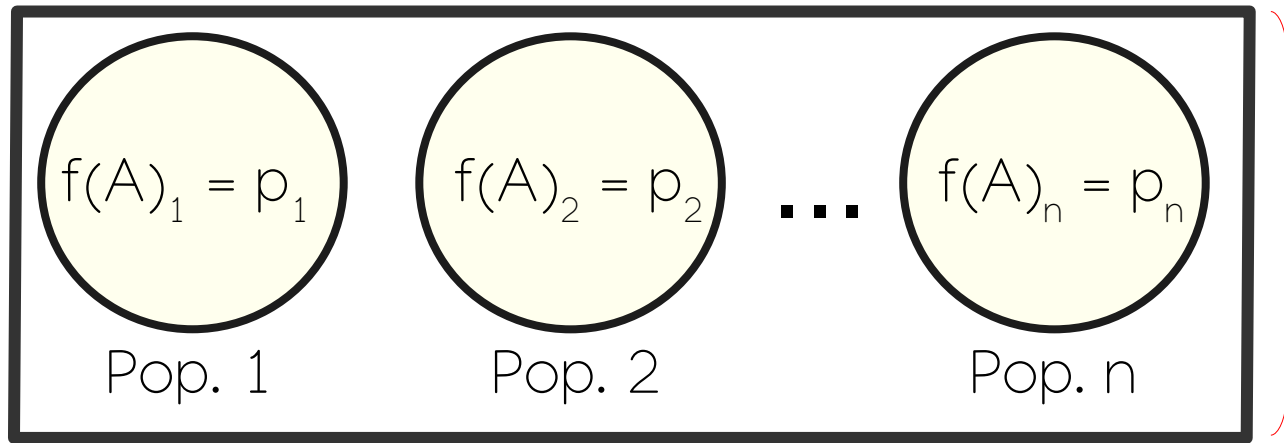


$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$

$$\bar{p} = \frac{1}{n} (p_1 + p_2 + \dots + p_n)$$

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

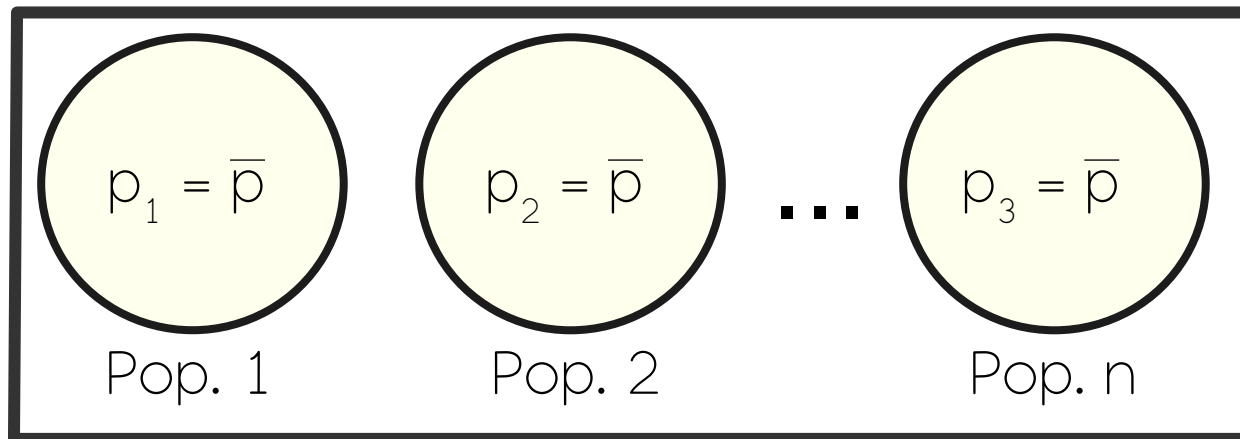
What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )



$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$

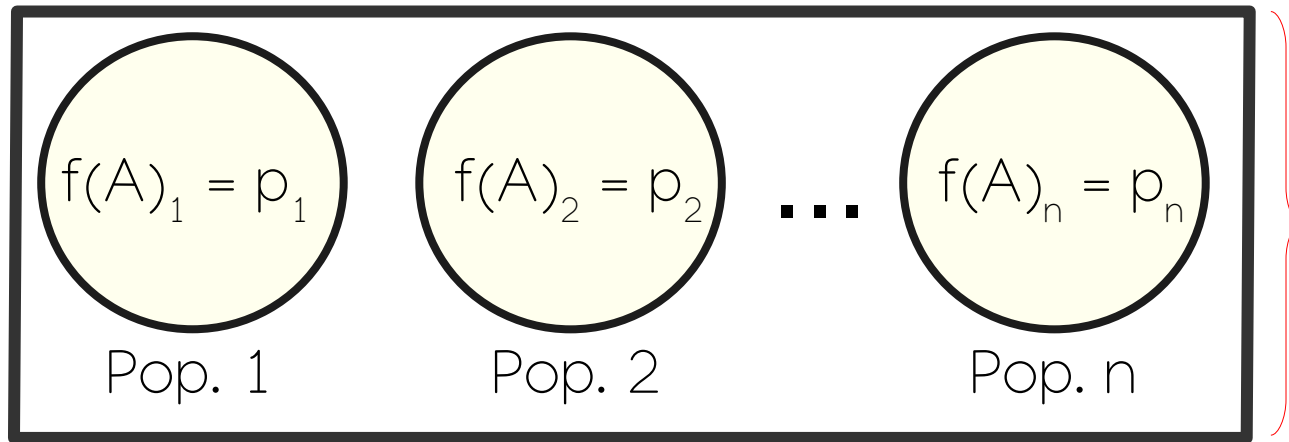
$$\bar{p} = \frac{1}{3} (p_1 + p_2 + \dots + p_n)$$

Lowest variance



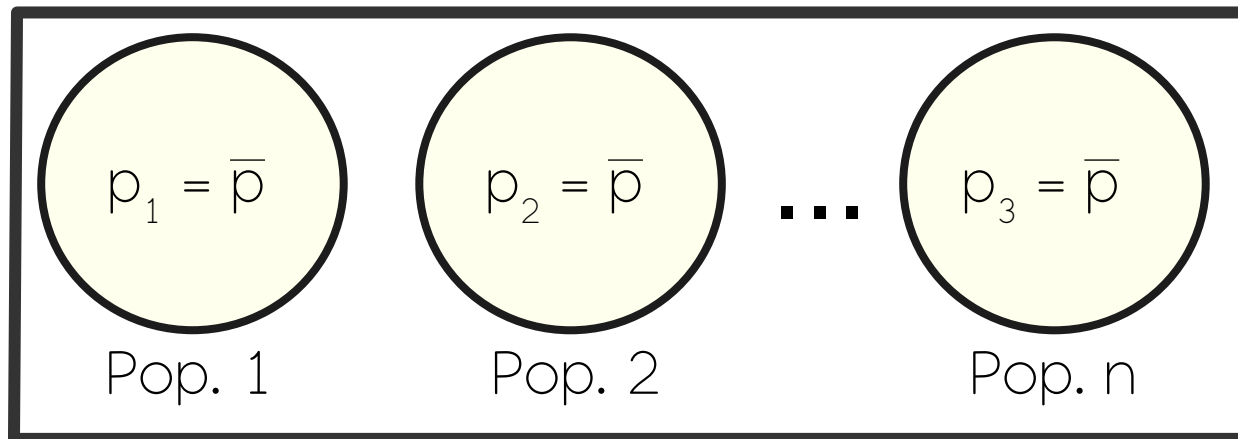
We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )



$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$
$$\bar{p} = \frac{1}{3} (p_1 + p_2 + \dots + p_n)$$

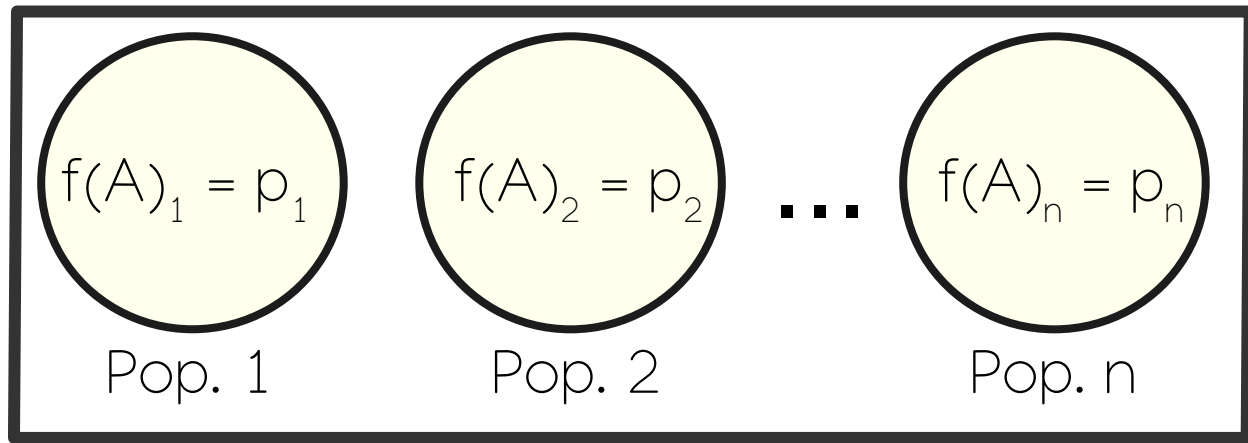
Lowest variance  $\rightarrow$  when  $p_i = \bar{p}$





We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

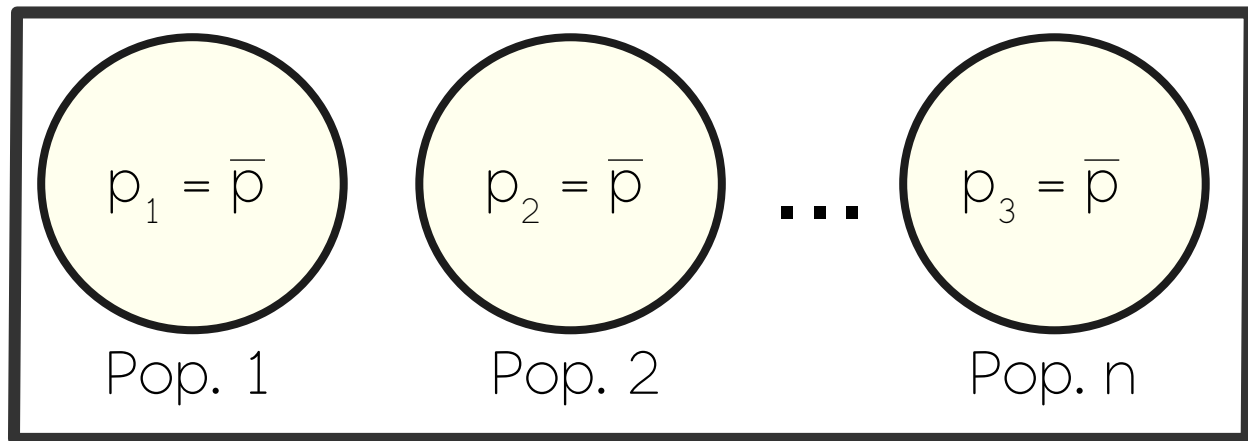
What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )



$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$

$$\bar{p} = \frac{1}{3} (p_1 + p_2 + \dots + p_n)$$

Lowest variance  $\rightarrow$  when  $p_i = \bar{p}$

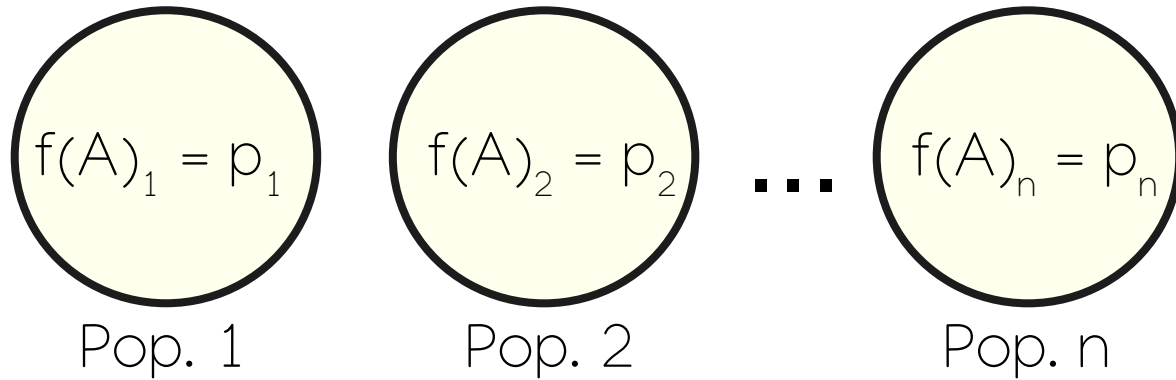


$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (p_i - \bar{p})^2$$

$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (\bar{p} - \bar{p})^2 = 0$$

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

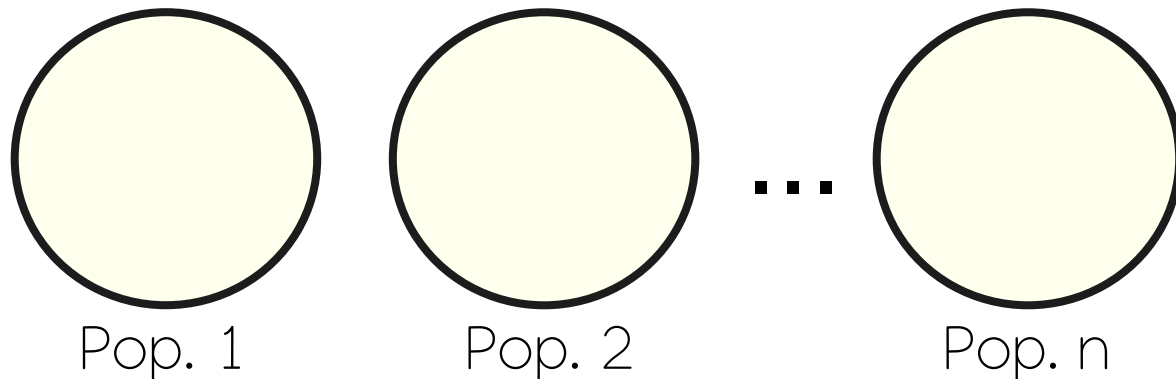
What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )



$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$

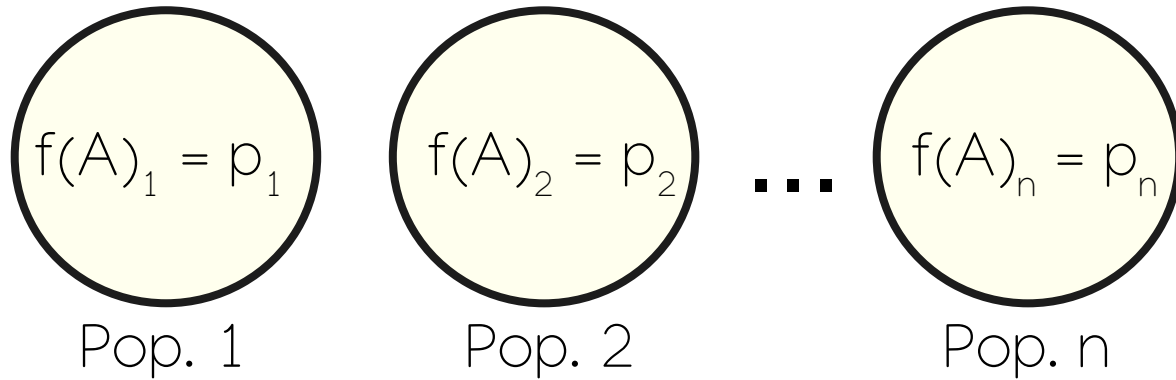
$$\bar{p} = \frac{1}{n} (p_1 + p_2 + \dots + p_n)$$

Highest variance



We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

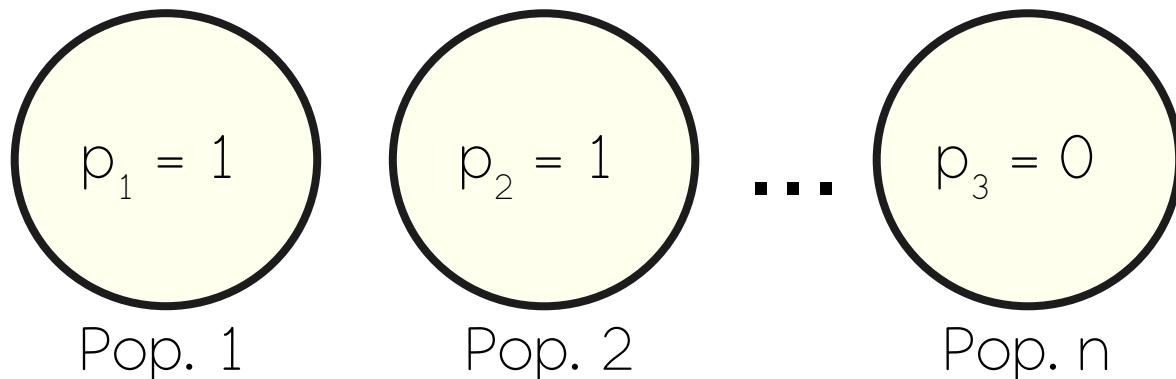
What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )



$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$

$$\bar{p} = \frac{1}{n} (p_1 + p_2 + \dots + p_n)$$

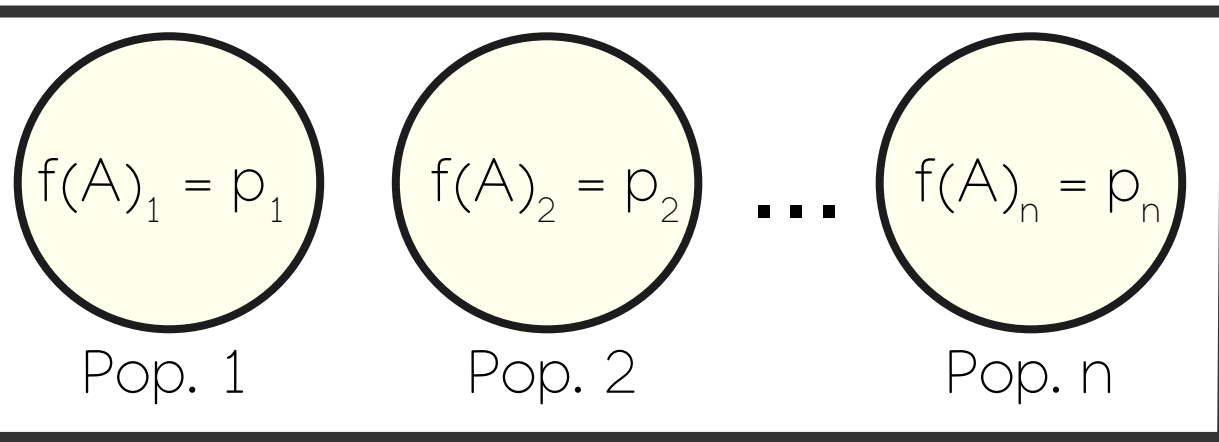
Highest variance  $\rightarrow$  when  $p_i = 0$  or  $1$



Proportion  $\bar{p}$  with  $p_i = 1$  and proportion  $\bar{q}$  with  $p_i = 0$

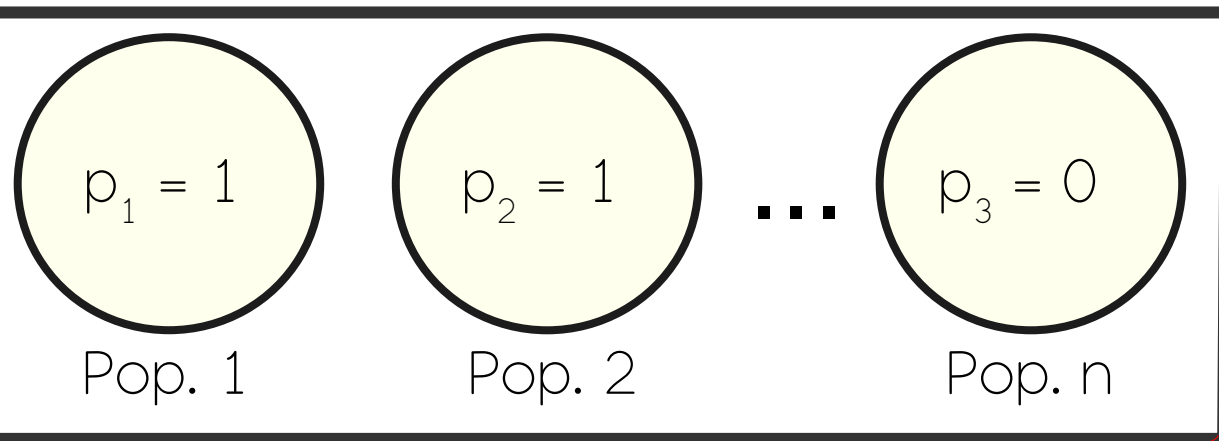
We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )



$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$
$$\bar{p} = \frac{1}{3} (p_1 + p_2 + \dots + p_n)$$

Highest variance  $\rightarrow$  when  $p_i = 0$  or  $1$

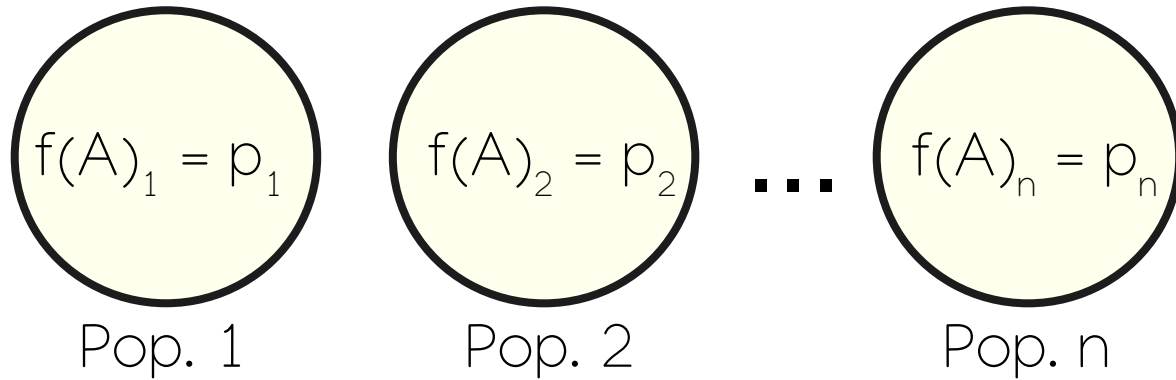


$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (p_i - \bar{p})^2$$
$$\sigma^2 = \bar{p} (1 - \bar{p})^2 + \bar{q} (0 - \bar{p})^2$$

Proportion  $\bar{p}$  with  $p_i = 1$  and proportion  $\bar{q}$  with  $p_i = 0$

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

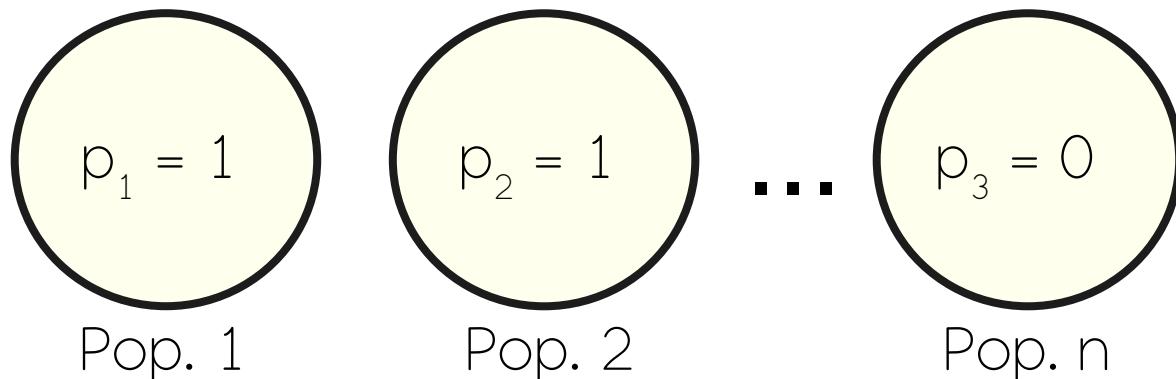
What are the limits of  $\sigma^2$  ?  
(the min. and max. values for  $\sigma^2$ )



$$\bar{p} = \frac{1}{n} \cdot \sum_{i=1}^n p_i$$

$$\bar{p} = \frac{1}{3} (p_1 + p_2 + \dots + p_n)$$

Highest variance  $\rightarrow$  when  $p_i = 0$  or  $1$



$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (p_i - \bar{p})^2$$

$$\sigma^2 = \bar{p} (1 - \bar{p})^2 + \bar{q} (0 - \bar{p})^2$$

$$\sigma^2 = \bar{p} \bar{q}$$

Proportion  $\bar{p}$  with  $p_i = 1$  and proportion  $\bar{q}$  with  $p_i = 0$

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\overline{pq}$  (max)

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\overline{pq}$  (max)



Depend on allele frequencies

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\overline{pq}$  (max)



Depend on allele frequencies



Makes comparisons between loci complicated

Locus 1 :  $\sigma^2 = 0.09$

Locus 2 :  $\sigma^2 = 0.12$

Differentiation at locus 2  
appears higher than locus 1



We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\bar{p}q$  (max)

↓  
Depend on allele frequencies

↓  
Makes comparisons between loci complicated

Locus 1 :  $\sigma^2 = 0.09$

If  $\bar{p} = 0.1$

Then  $\sigma^2$  in  $[0 ; 0.09]$

Locus 2 :  $\sigma^2 = 0.12$

Differentiation at locus 2  
appears higher than locus 1

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\bar{p}q$  (max)

Depend on allele frequencies

Makes comparisons between loci complicated

Locus 1 :  $\sigma^2 = 0.09$

If  $\bar{p} = 0.1$

Then  $\sigma^2$  in  $[0 ; 0.09]$

Locus 2 :  $\sigma^2 = 0.12$

If  $\bar{p} = 0.5$

Then  $\sigma^2$  in  $[0 ; 0.25]$

Differentiation at locus 2  
appears higher than locus 1

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\bar{p}q$  (max)

Depend on allele frequencies

Makes comparisons between loci complicated

Locus 1 :  $\sigma^2 = 0.09$

If  $\bar{p} = 0.1$

Then  $\sigma^2$  in  $[0 ; 0.09]$

Locus 2 :  $\sigma^2 = 0.12$

If  $\bar{p} = 0.5$

Then  $\sigma^2$  in  $[0 ; 0.25]$

Differentiation at locus 2  
appears higher than locus 1

We standardize  $\sigma^2$  by the maximum variance possible

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\bar{p}\bar{q}$  (max)

Depend on allele frequencies

Makes comparisons between loci complicated

Locus 1 :  $\sigma^2 = 0.09$

If  $\bar{p} = 0.1$

Then  $\sigma^2$  in  $[0 ; 0.09]$

Locus 2 :  $\sigma^2 = 0.12$

If  $\bar{p} = 0.5$

Then  $\sigma^2$  in  $[0 ; 0.25]$

Differentiation at locus 2  
appears higher than locus 1

We standardize  $\sigma^2$  by the maximum variance possible

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

We can measure the variance in allelic frequencies ( $\sigma^2$ ) to quantify the genetic differentiation

Values for  $\sigma^2$  are lying from 0 (min) to  $\bar{p}\bar{q}$  (max)

Depend on allele frequencies

Makes comparisons between loci complicated

Locus 1 :  $\sigma^2 = 0.09$

If  $\bar{p} = 0.1$

Then  $\sigma^2$  in  $[0 ; 0.09]$

Locus 2 :  $\sigma^2 = 0.12$

If  $\bar{p} = 0.5$

Then  $\sigma^2$  in  $[0 ; 0.25]$

Differentiation at locus 2  
appears higher than locus 1

We standardize  $\sigma^2$  by the maximum variance possible

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

Locus 1 :  $F_{ST} = 1$

Locus 2 :  $F_{ST} = 0.48$

Locus 1 actually more differentiated than 2

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = \frac{\sigma^2}{\bar{p} \bar{q}}$$

$$\sigma^2 = F_{ST} \cdot \bar{p} \bar{q}$$

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = \frac{\sigma^2}{\bar{p} \bar{q}}$$

$$\sigma^2 = F_{ST} \cdot \bar{p} \bar{q}$$

$$E[f(Aa)] = 2 \bar{p} \bar{q} - 2 \sigma^2$$



Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

$$\sigma^2 = F_{ST} \cdot \bar{p}\bar{q}$$

$$E[f(Aa)] = 2\bar{p}\bar{q} - 2\sigma^2$$

$$E[f(Aa)] = 2\bar{p}\bar{q} - 2F_{ST} \cdot \bar{p}\bar{q}$$

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

$$\sigma^2 = F_{ST} \cdot \bar{p}\bar{q}$$

$$E[f(Aa)] = 2\bar{p}\bar{q} - 2\sigma^2$$

$$E[f(Aa)] = 2\bar{p}\bar{q} - 2F_{ST} \cdot \bar{p}\bar{q}$$

$$E[f(Aa)] = 2\bar{p}\bar{q}(1 - F_{ST})$$

$$1 - F_{ST} = \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

$$\sigma^2 = F_{ST} \cdot \bar{p}\bar{q}$$

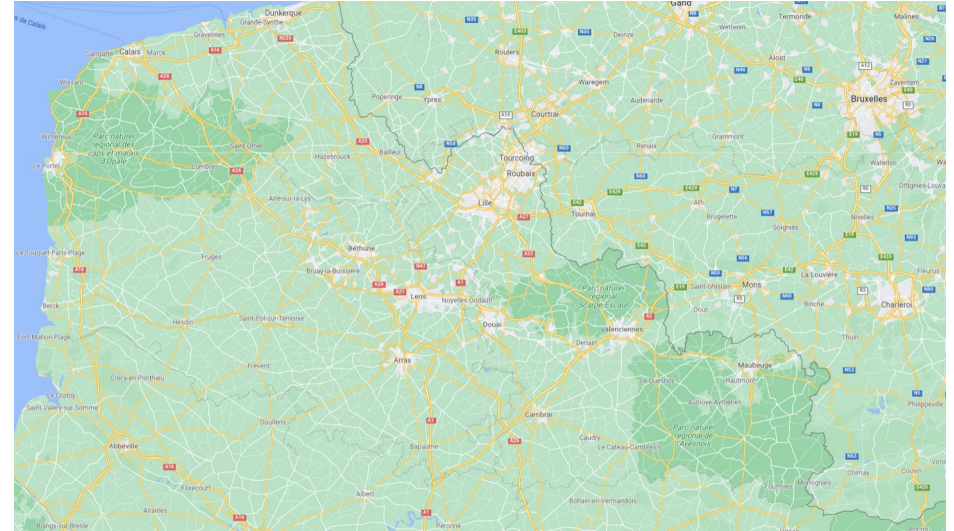
$$E[f(Aa)] = 2\bar{p}\bar{q} - 2\sigma^2$$

$$E[f(Aa)] = 2\bar{p}\bar{q} - 2F_{ST} \cdot \bar{p}\bar{q}$$

$$E[f(Aa)] = 2\bar{p}\bar{q}(1 - F_{ST})$$

$$1 - F_{ST} = \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$



# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = \frac{\sigma^2}{\bar{p}\bar{q}}$$

$$\sigma^2 = F_{ST} \cdot \bar{p}\bar{q}$$

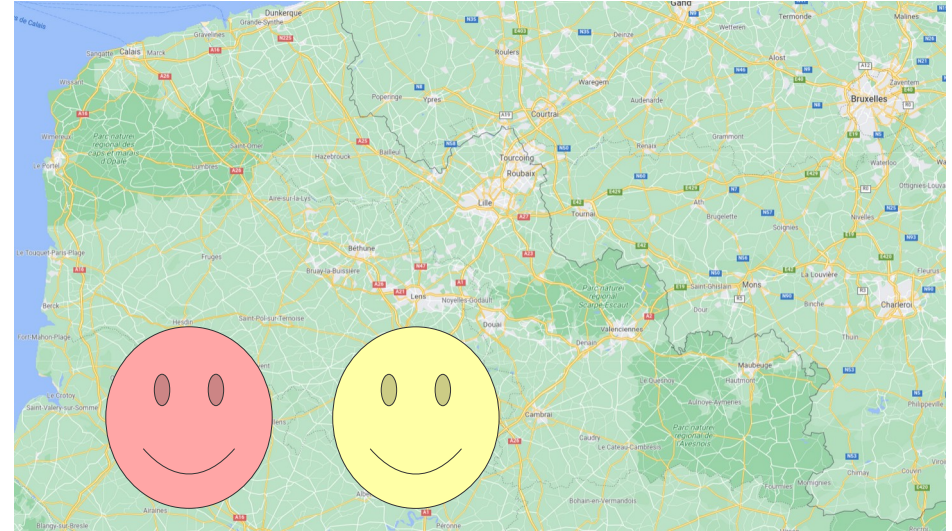
$$E[f(Aa)] = 2\bar{p}\bar{q} - 2\sigma^2$$

$$E[f(Aa)] = 2\bar{p}\bar{q} - 2F_{ST} \cdot \bar{p}\bar{q}$$

$$E[f(Aa)] = 2\bar{p}\bar{q}(1 - F_{ST})$$

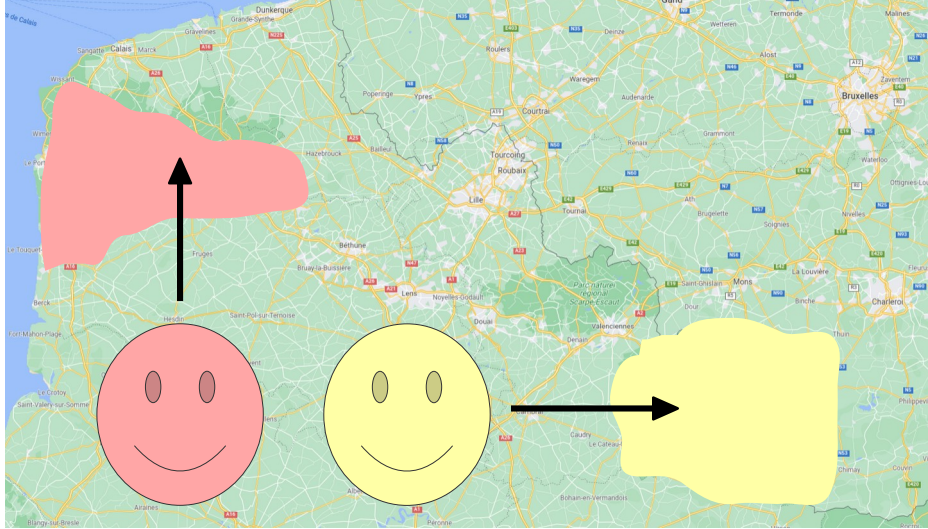
$$1 - F_{ST} = \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$



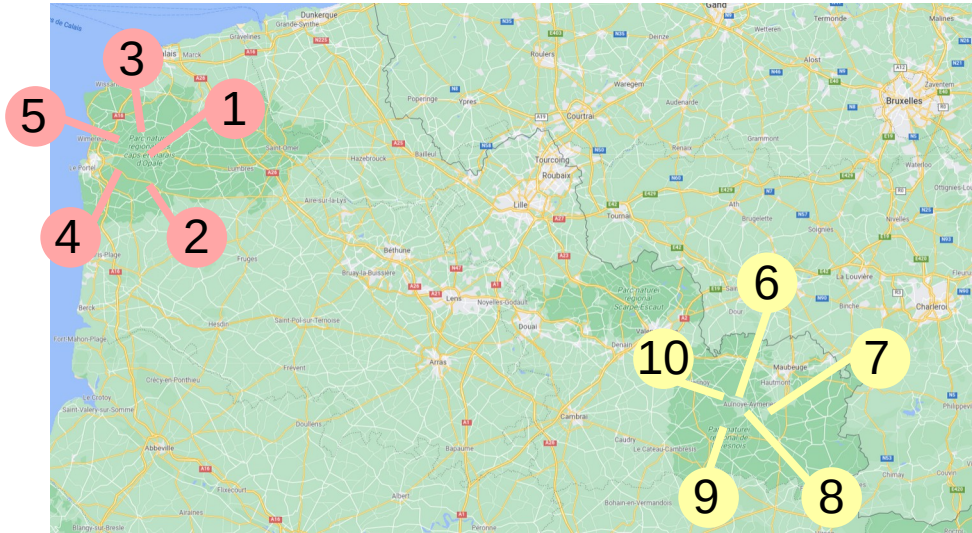
# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$



# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

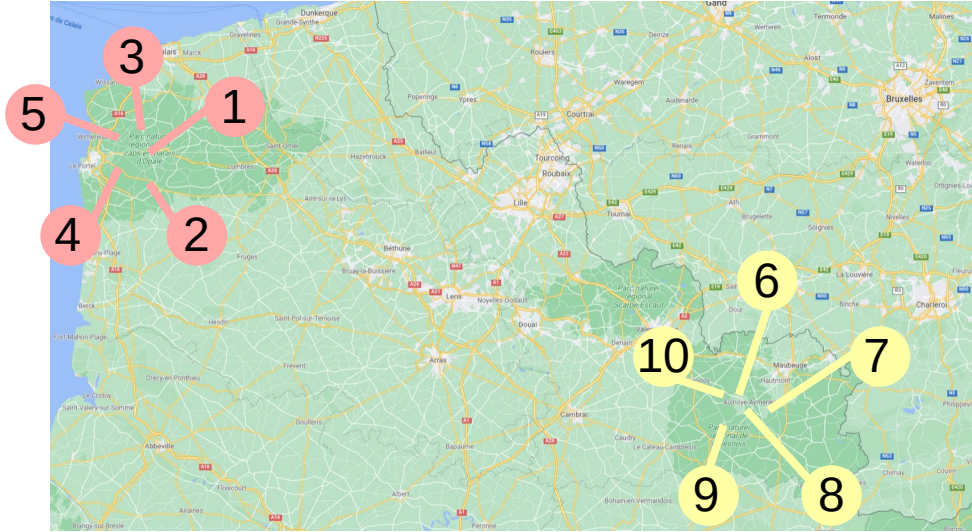


Sampling



# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$



Sampling  
Sequencing

10 sequenced  
copies of one  
gene

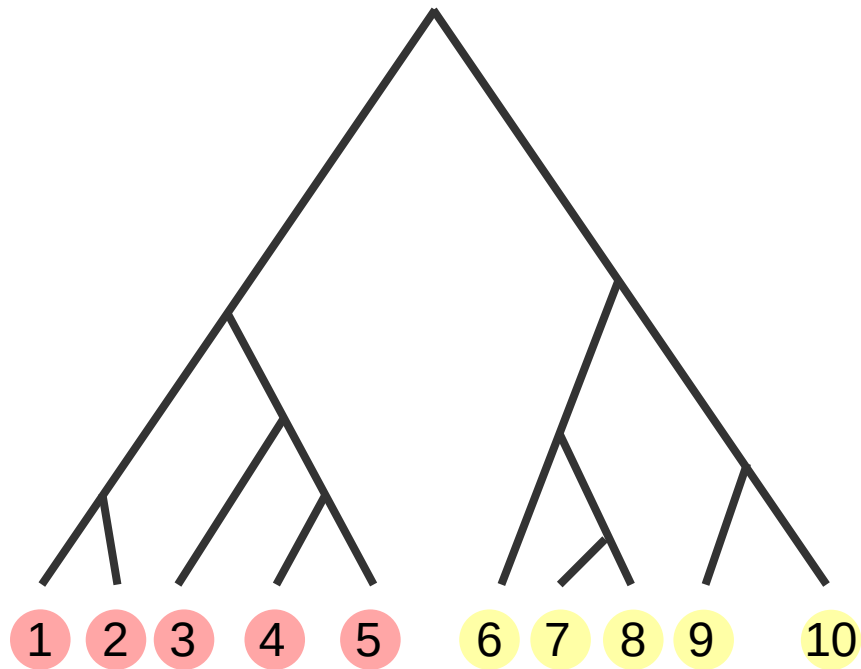
- 1 \_\_\_\_\_
- 2 \_\_\_\_\_
- 3 \_\_\_\_\_
- 4 \_\_\_\_\_
- 5 \_\_\_\_\_
- 6 \_\_\_\_\_
- 7 \_\_\_\_\_
- 8 \_\_\_\_\_
- 9 \_\_\_\_\_
- 10 \_\_\_\_\_

length =  $L$  nucleotides

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

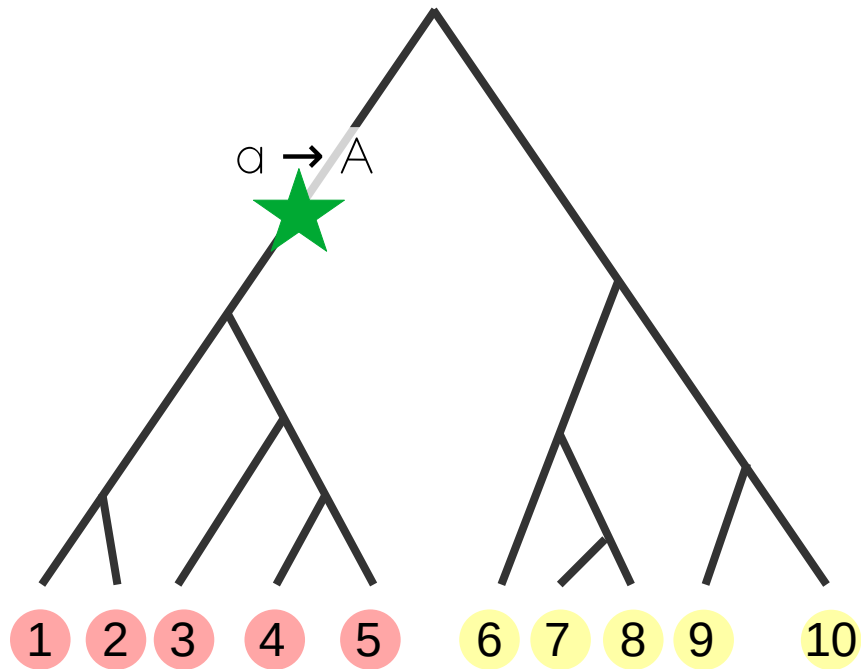




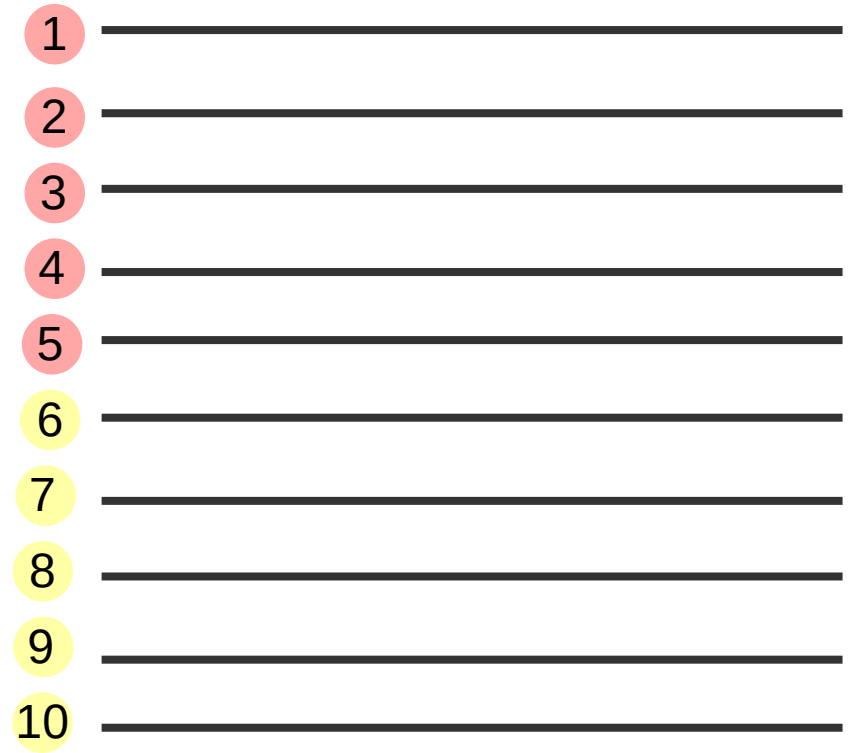
# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



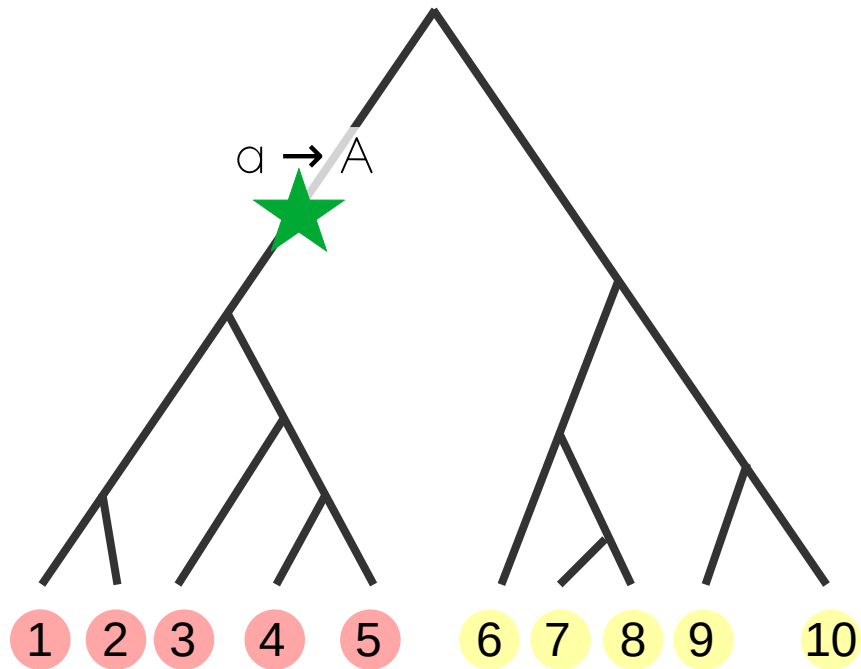
Sequenced dataset



# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



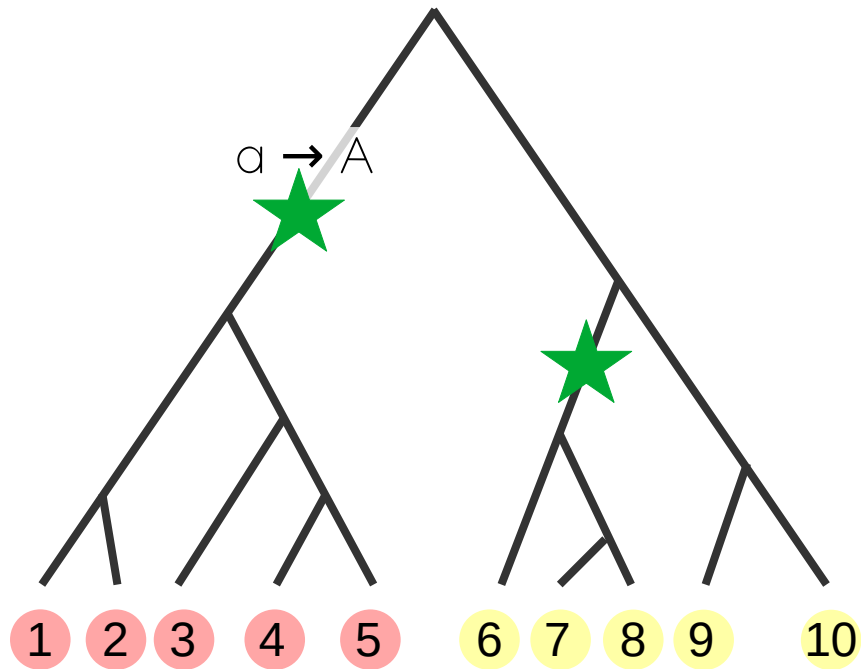
Sequenced dataset

1	— A	_____
2	— A	_____
3	— A	_____
4	— A	_____
5	— A	_____
6	— a	_____
7	— a	_____
8	— a	_____
9	— a	_____
10	— a	_____

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



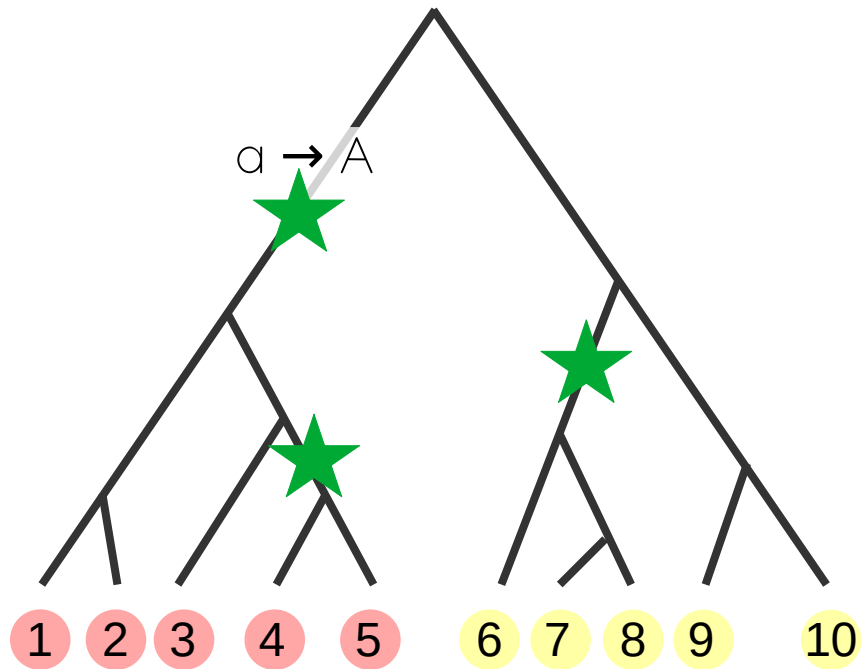
Sequenced dataset

1	-	A	-	a	_____
2	-	A	-	a	_____
3	-	A	-	a	_____
4	-	A	-	a	_____
5	-	A	-	a	_____
6	-	a	-	A	_____
7	-	a	-	A	_____
8	-	a	-	A	_____
9	-	a	-	a	_____
10	-	a	-	a	_____

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



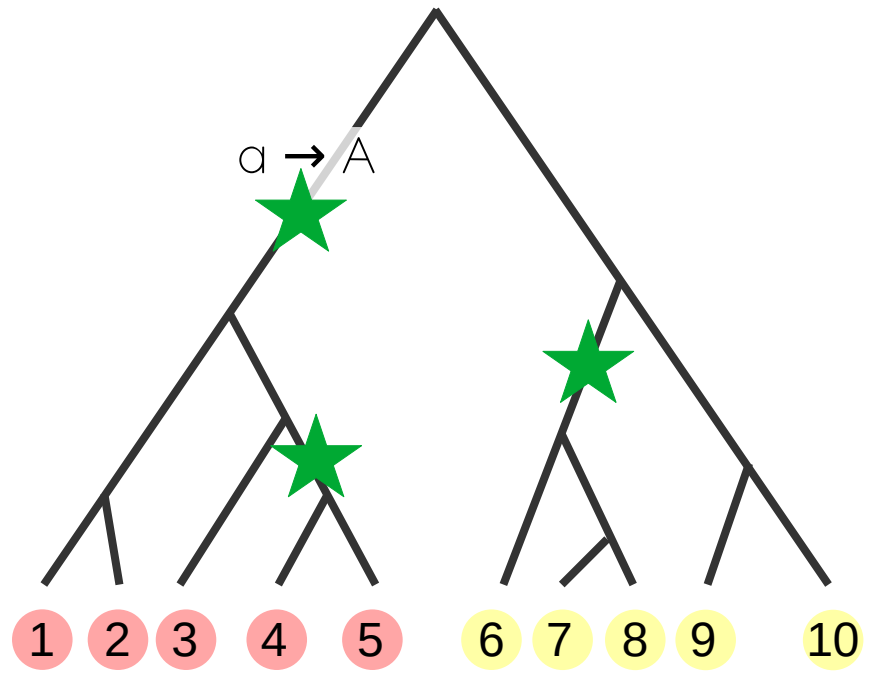
Sequenced dataset

1	-	A	-	a	_____
2	-	A	-	a	_____
3	-	A	-	a	_____
4	-	A	-	a	_____
5	-	A	-	a	_____
6	-	a	-	A	_____
7	-	a	-	A	_____
8	-	a	-	A	_____
9	-	a	-	a	_____
10	-	a	-	a	_____

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2 \bar{p} \bar{q}}$$

Real (unknown)  
coalescent tree



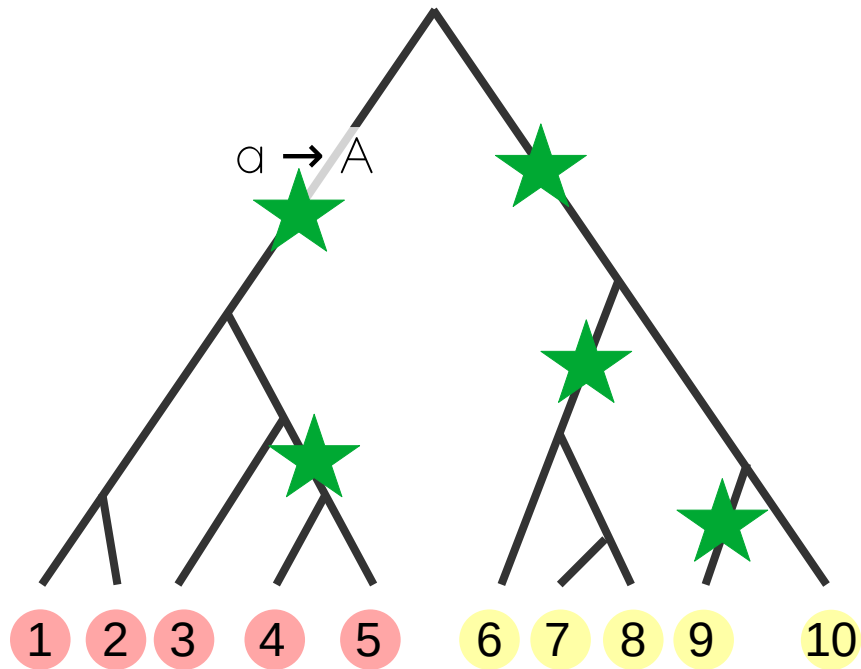
Sequenced dataset

1	-	A	-	a	-	a	-	-	-	-
2	-	A	-	a	-	a	-	-	-	-
3	-	A	-	a	-	a	-	-	-	-
4	-	A	-	a	-	A	-	-	-	-
5	-	A	-	a	-	A	-	-	-	-
6	-	a	-	A	-	a	-	-	-	-
7	-	a	-	A	-	a	-	-	-	-
8	-	a	-	A	-	a	-	-	-	-
9	-	a	-	a	-	a	-	-	-	-
10	-	a	-	a	-	a	-	-	-	-

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



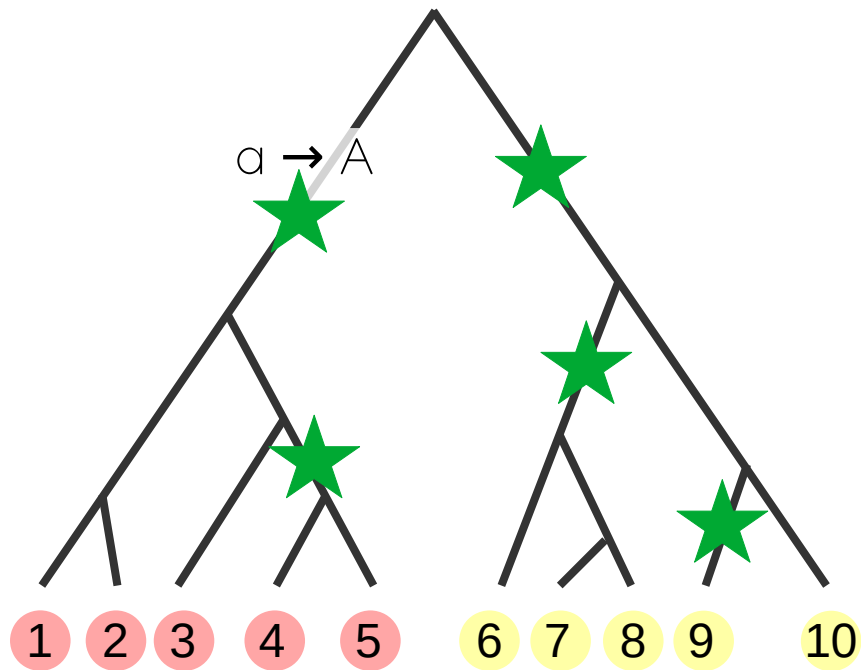
Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

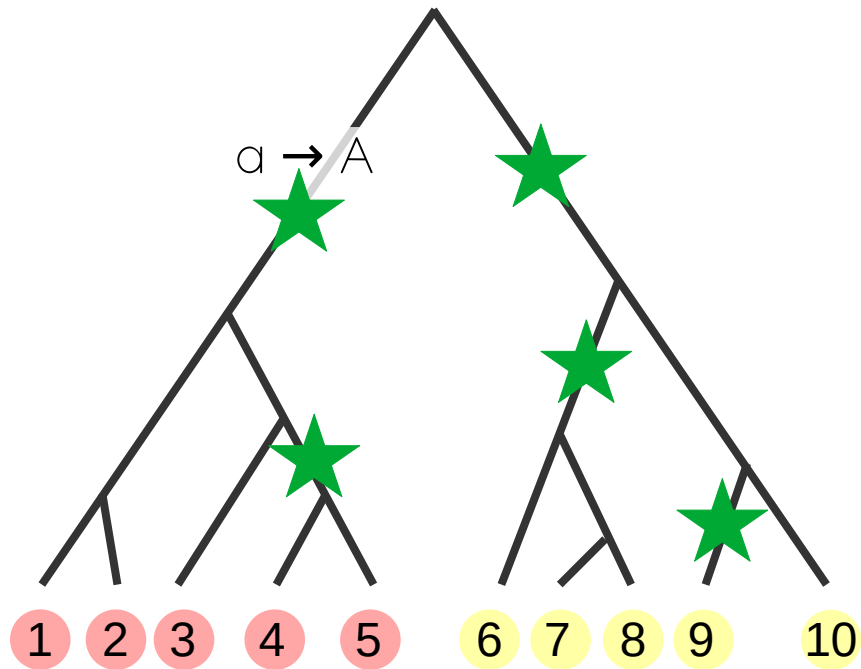
1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

$E[f(Aa)]$  = average heterozygosity among populations

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

$E[f(Aa)]$  = average heterozygosity among populations

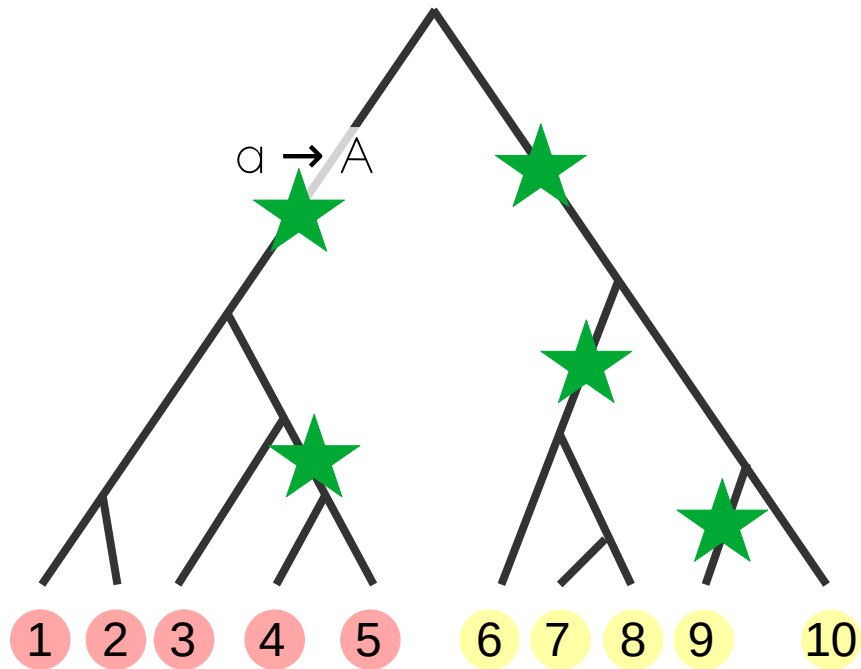
$$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i$$



# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

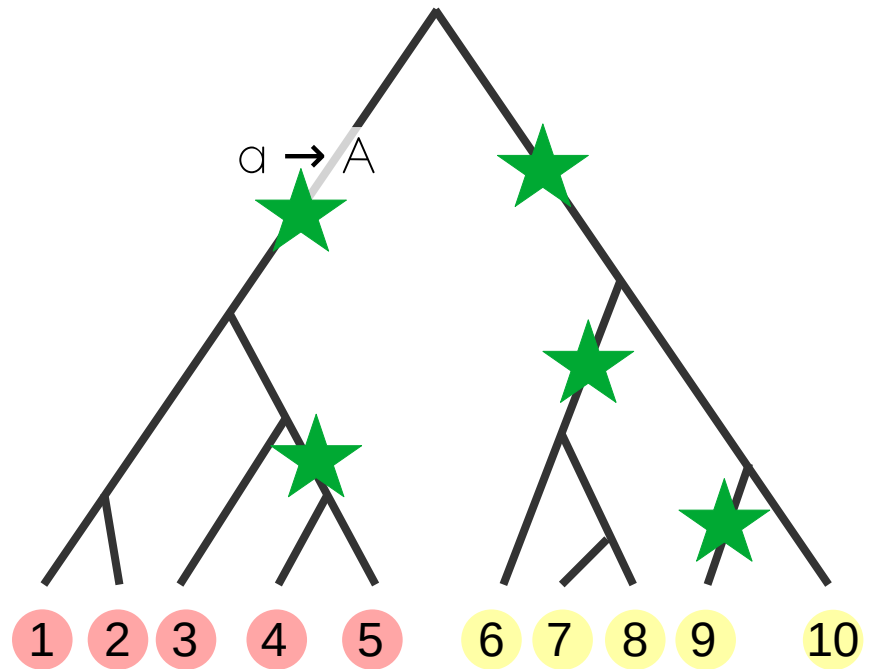
$E[f(Aa)]$  = average heterozygosity among populations

$$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2p_1q_1 + 2p_2q_2)$$

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2 \bar{p} \bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

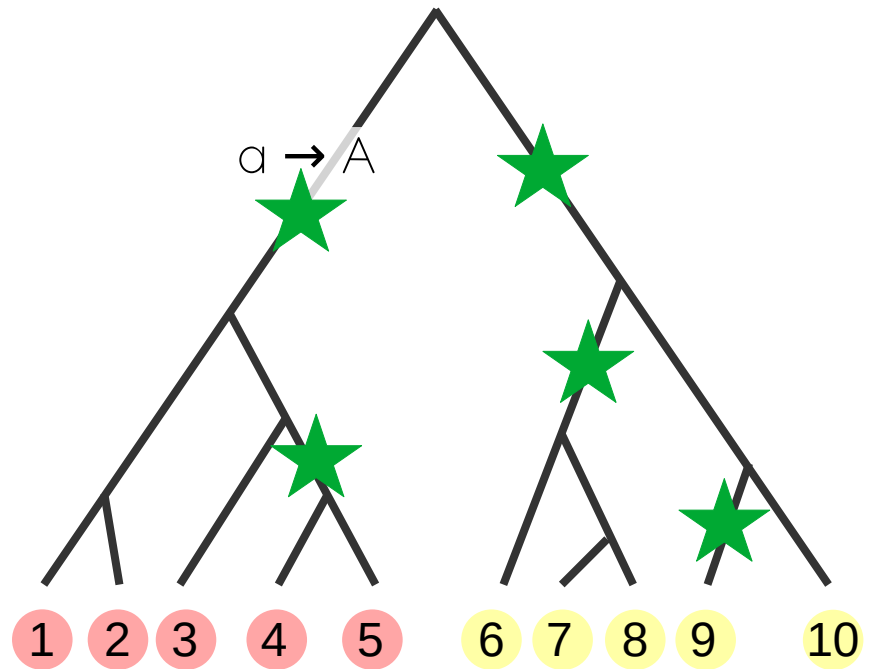
$E[f(Aa)]$  = average heterozygosity among populations

$$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2 p_1 q_1 + 2 p_2 q_2) \longrightarrow \quad ? \quad ? \quad ? \quad ? \quad ?$$

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2 \bar{p} \bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

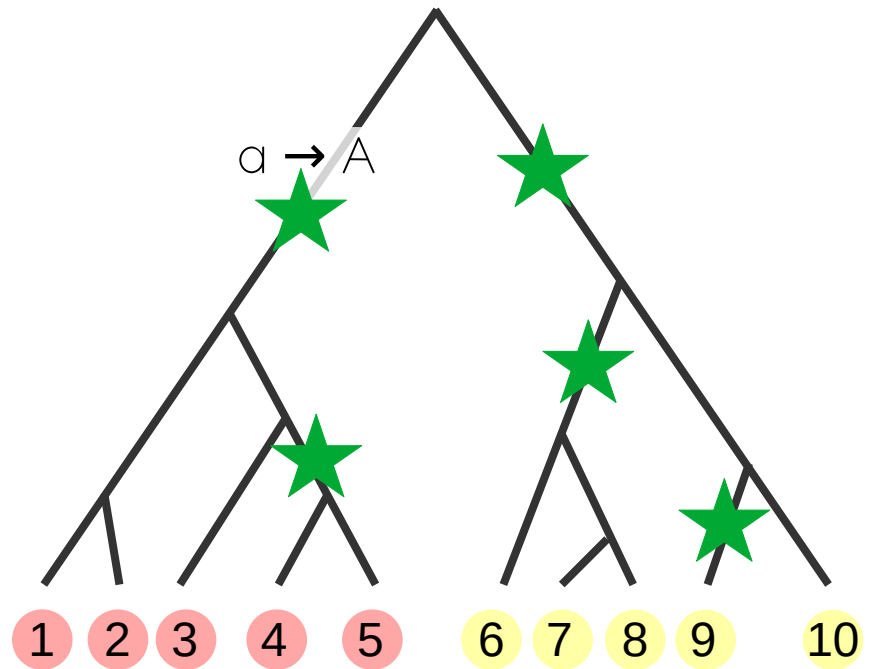
$E[f(Aa)]$  = average heterozygosity among populations

$$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2 p_1 q_1 + 2 p_2 q_2) \longrightarrow 0 \quad 0.24 \quad 0.24 \quad 0.16 \quad 0$$

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2 \bar{p} \bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

$E[f(Aa)]$  = average heterozygosity among populations

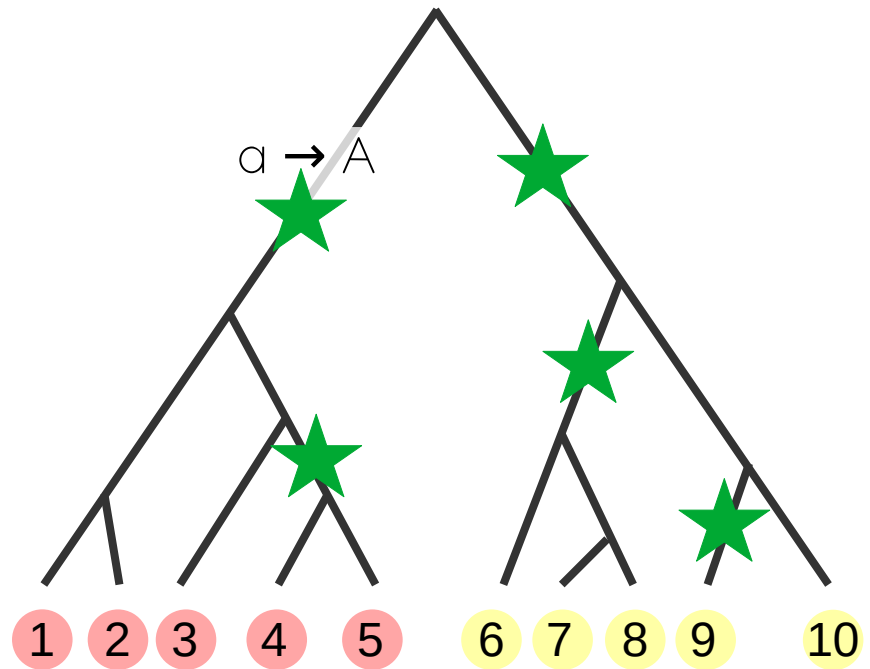
$$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2 p_1 q_1 + 2 p_2 q_2)$$

	0	0.24	0.24	0.16	0
$2 \bar{p} \bar{q}$					

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2 \bar{p} \bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

$E[f(Aa)]$  = average heterozygosity among populations

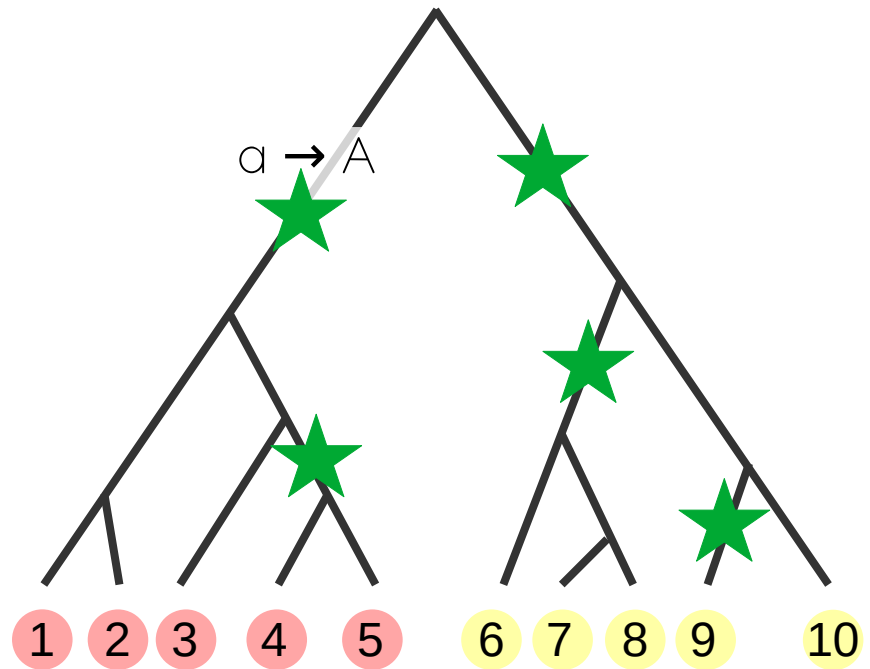
$$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2 p_1 q_1 + 2 p_2 q_2)$$

$2 \bar{p} \bar{q}$	→	0	0.24	0.24	0.16	0
	→	0.5	0.42	0.32	0.18	0.5

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2 \bar{p} \bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

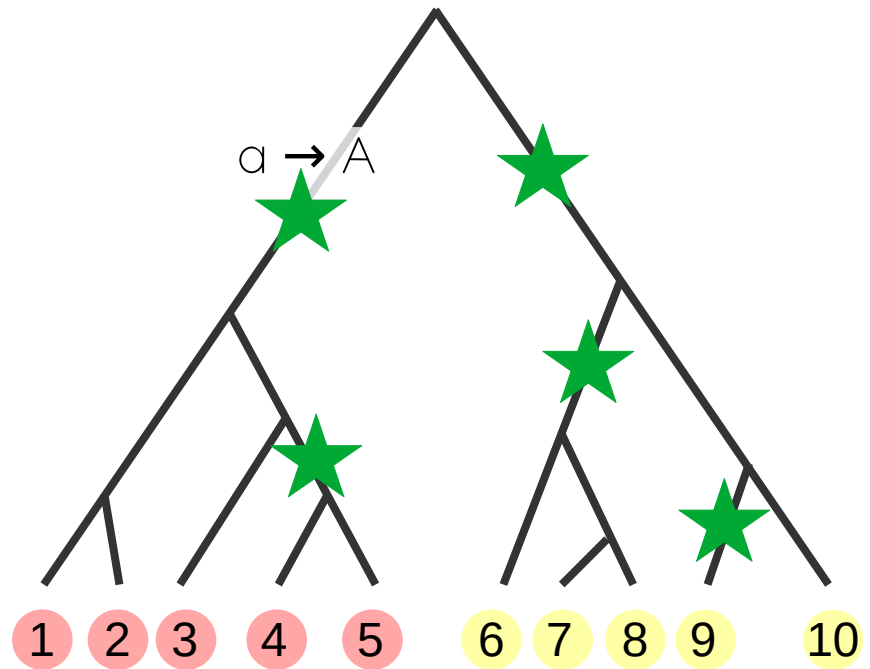
$E[f(Aa)]$  = average heterozygosity among populations

$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2 p_1 q_1 + 2 p_2 q_2)$	→	0	0.24	0.24	0.16	0
$2 \bar{p} \bar{q}$	→	0.5	0.42	0.32	0.18	0.5
$F_{ST}$	→					

# Measuring $F_{st}$ with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2\bar{p}\bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

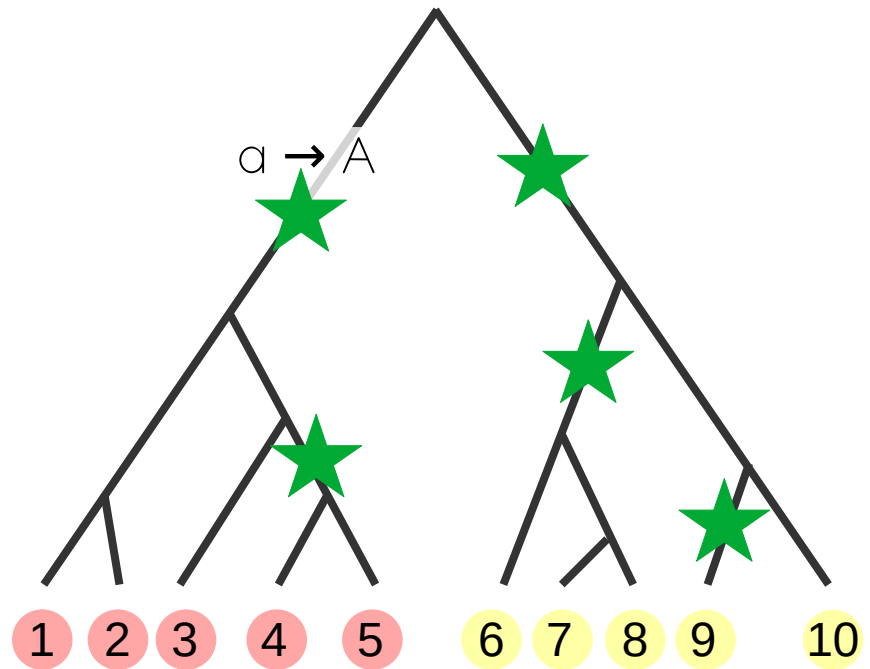
$E[f(Aa)]$  = average heterozygosity among populations

$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2p_1q_1 + 2p_2q_2)$	→	0	0.24	0.24	0.16	0
$2\bar{p}\bar{q}$	→	0.5	0.42	0.32	0.18	0.5
$F_{ST}$	→	1	0.43	0.25	0.11	1

Measuring  $F_{st}$  with DNA sequences

$$F_{ST} = 1 - \frac{E[f(Aa)]}{2 \bar{p} \bar{q}}$$

Real (unknown)  
coalescent tree



Sequenced dataset

1	-	A	-	a	-	a	-	a	-	a	-
2	-	A	-	a	-	a	-	a	-	a	-
3	-	A	-	a	-	a	-	a	-	a	-
4	-	A	-	a	-	A	-	a	-	a	-
5	-	A	-	a	-	A	-	a	-	a	-
6	-	a	-	A	-	a	-	a	-	A	-
7	-	a	-	A	-	a	-	a	-	A	-
8	-	a	-	A	-	a	-	a	-	A	-
9	-	a	-	a	-	a	-	A	-	A	-
10	-	a	-	a	-	a	-	a	-	A	-

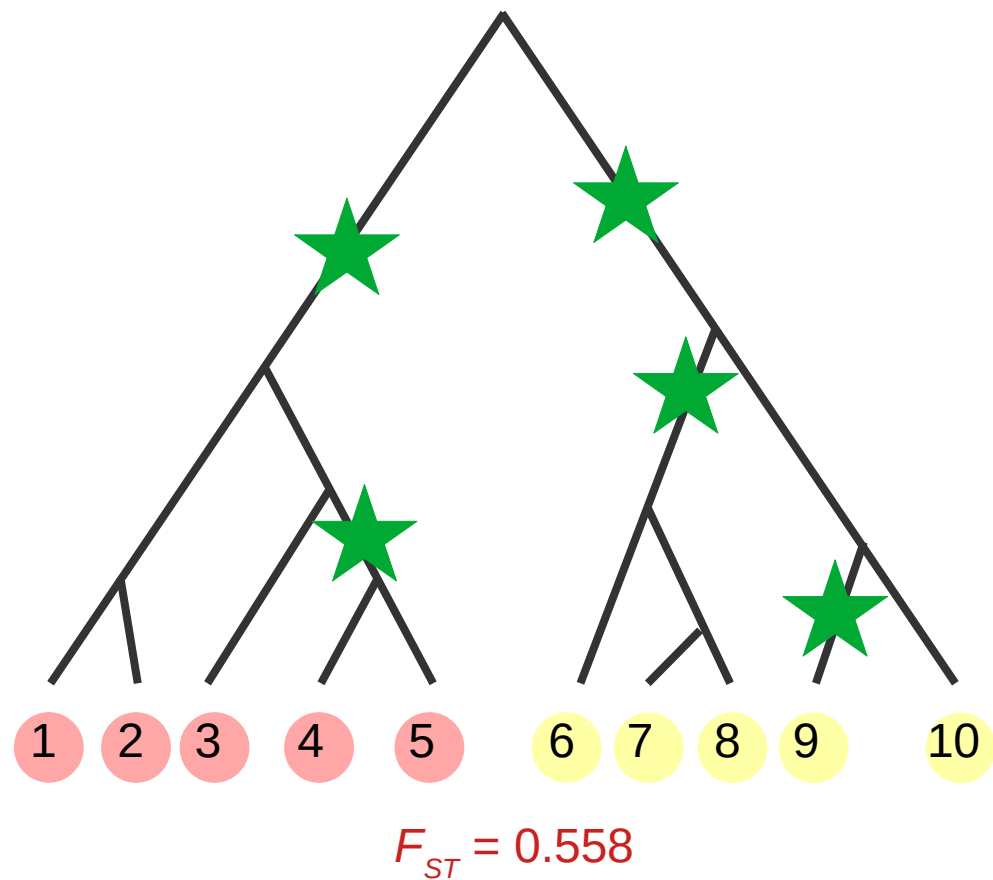
$E[f(Aa)]$  = average heterozygosity among populations

$E[f(Aa)] = \frac{1}{n} \sum_i^n f(Aa)_i = \frac{1}{2} (2 p_1 q_1 + 2 p_2 q_2)$	→	0	0.24	0.24	0.16	0
$2 \bar{p} \bar{q}$	→	0.5	0.42	0.32	0.18	0.5
$F_{ST}$	→	1	0.43	0.25	0.11	1

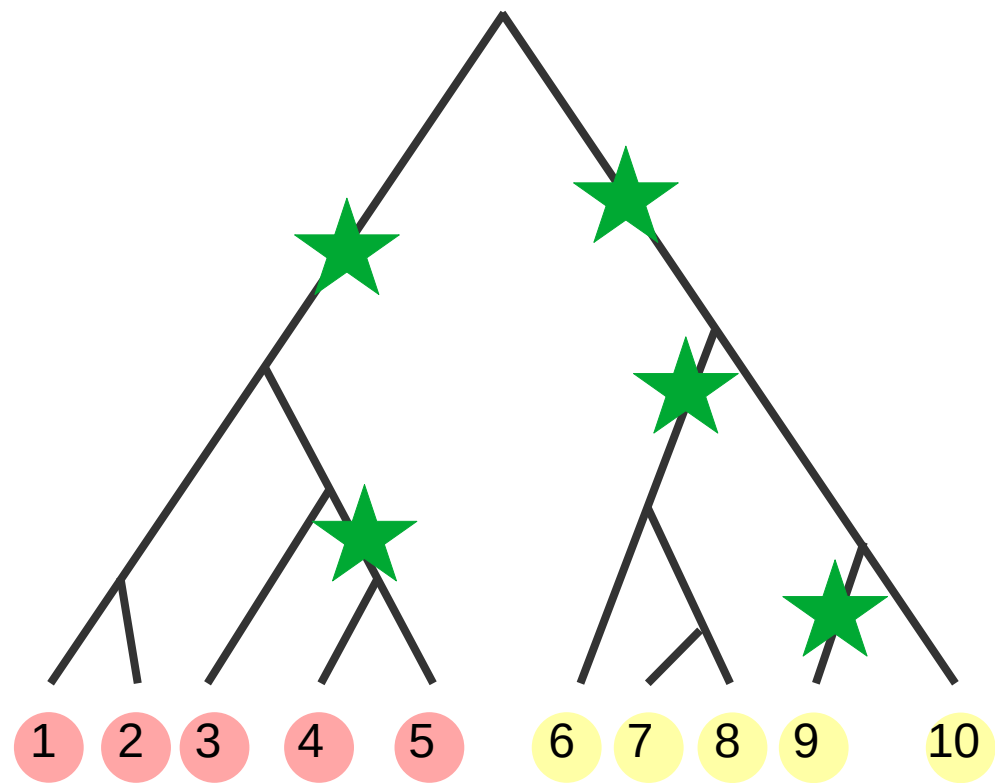
$F_{ST} = 0.558$



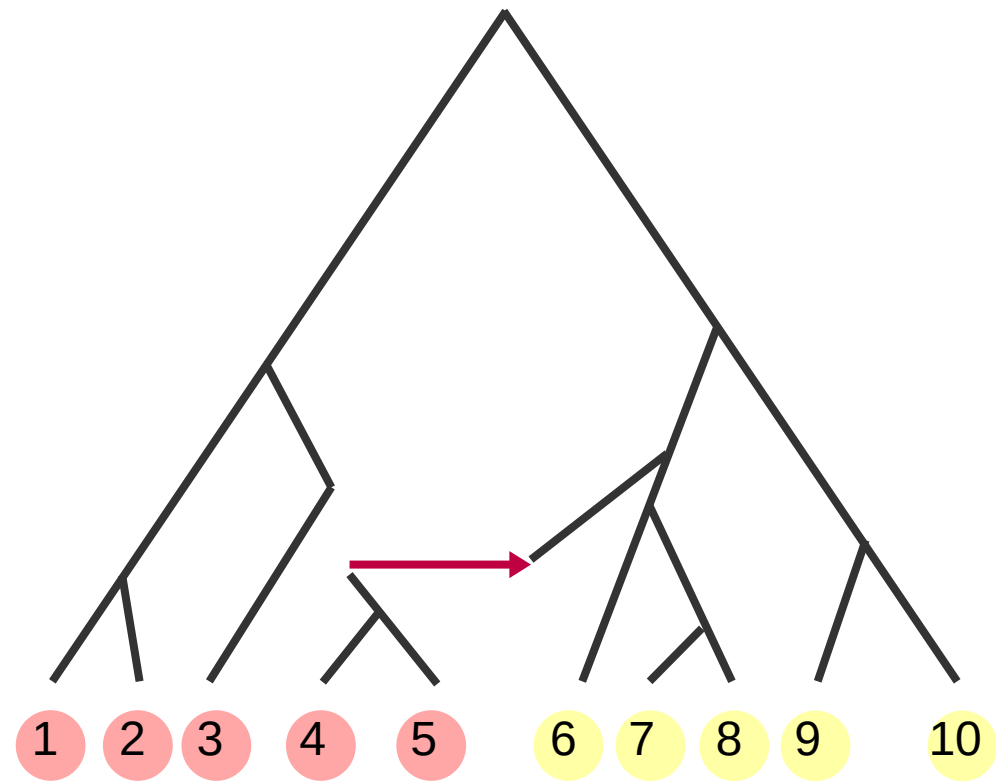
# The effect of migration on $F_{ST}$



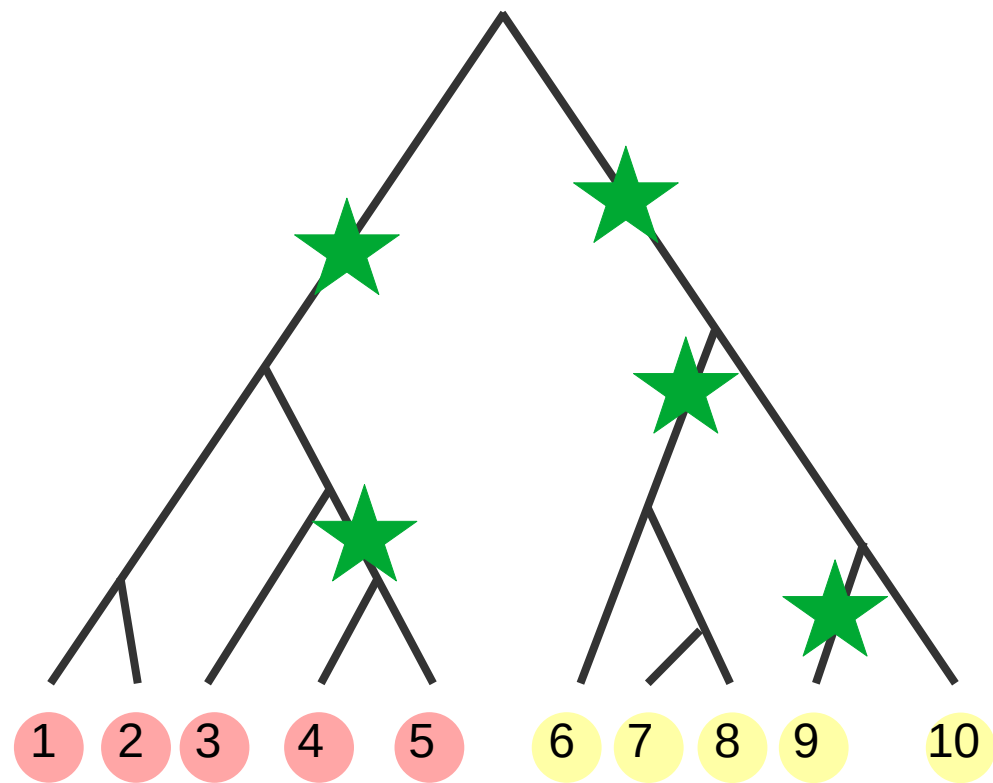
# The effect of migration on $F_{ST}$



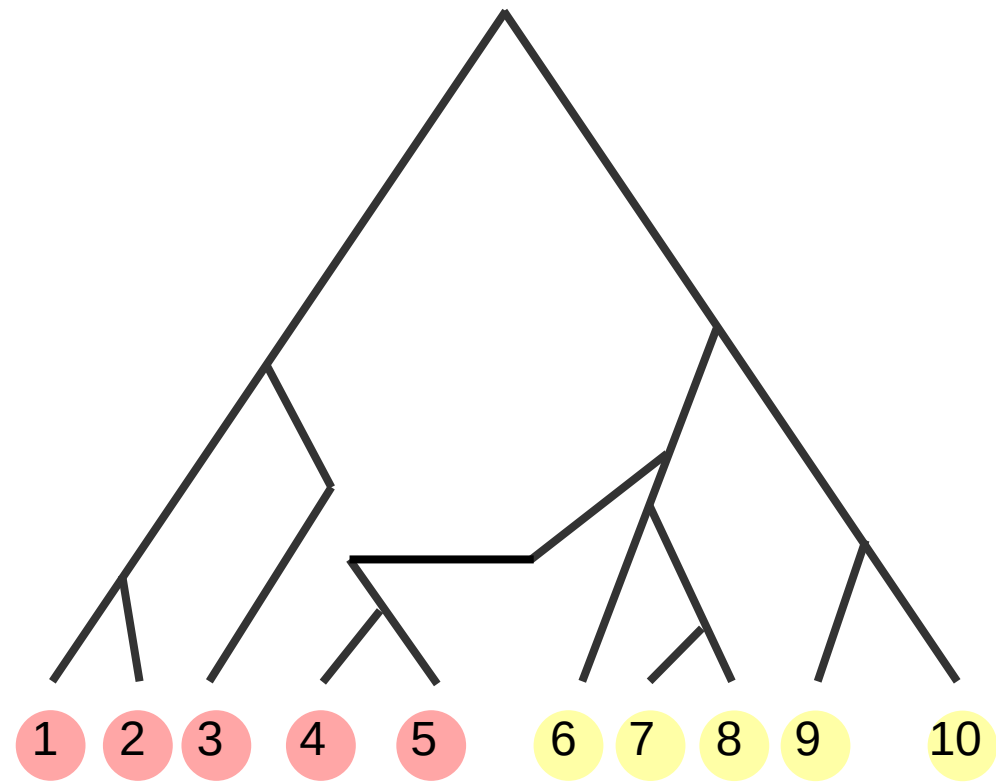
$$F_{ST} = 0.558$$



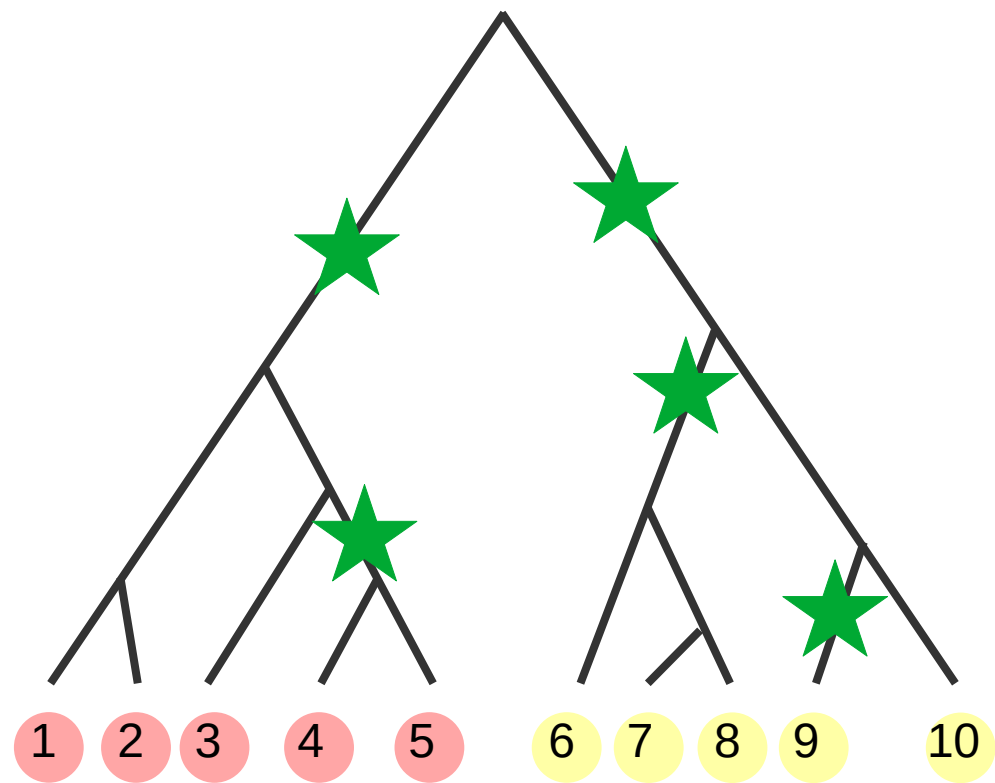
# The effect of migration on $F_{ST}$



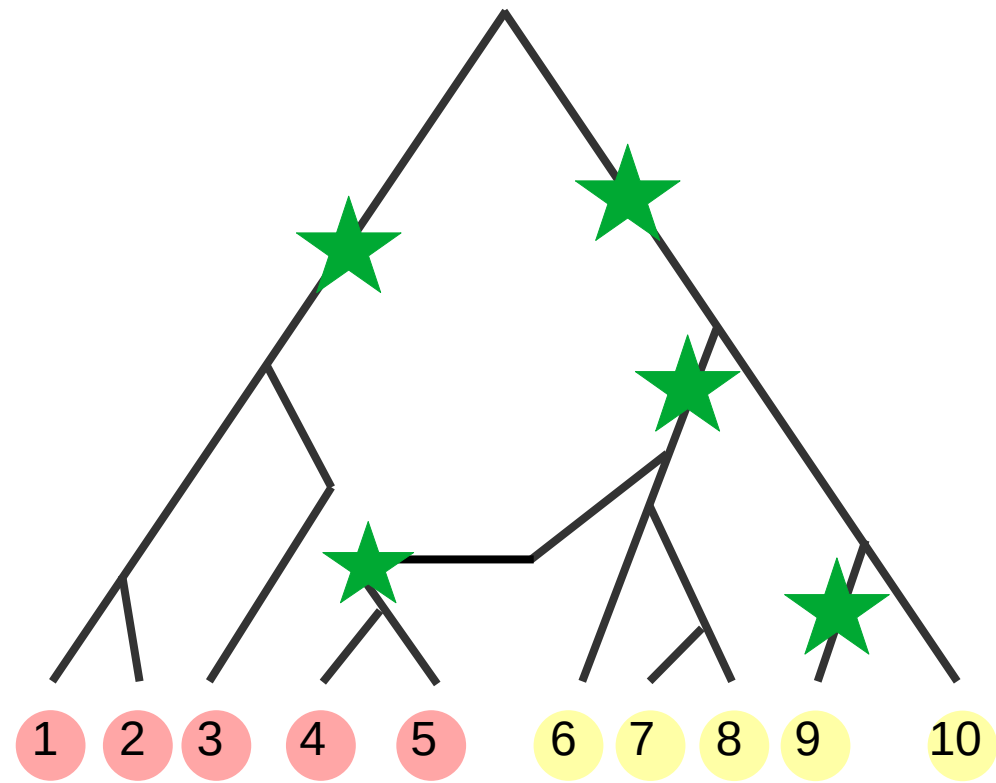
$$F_{ST} = 0.558$$



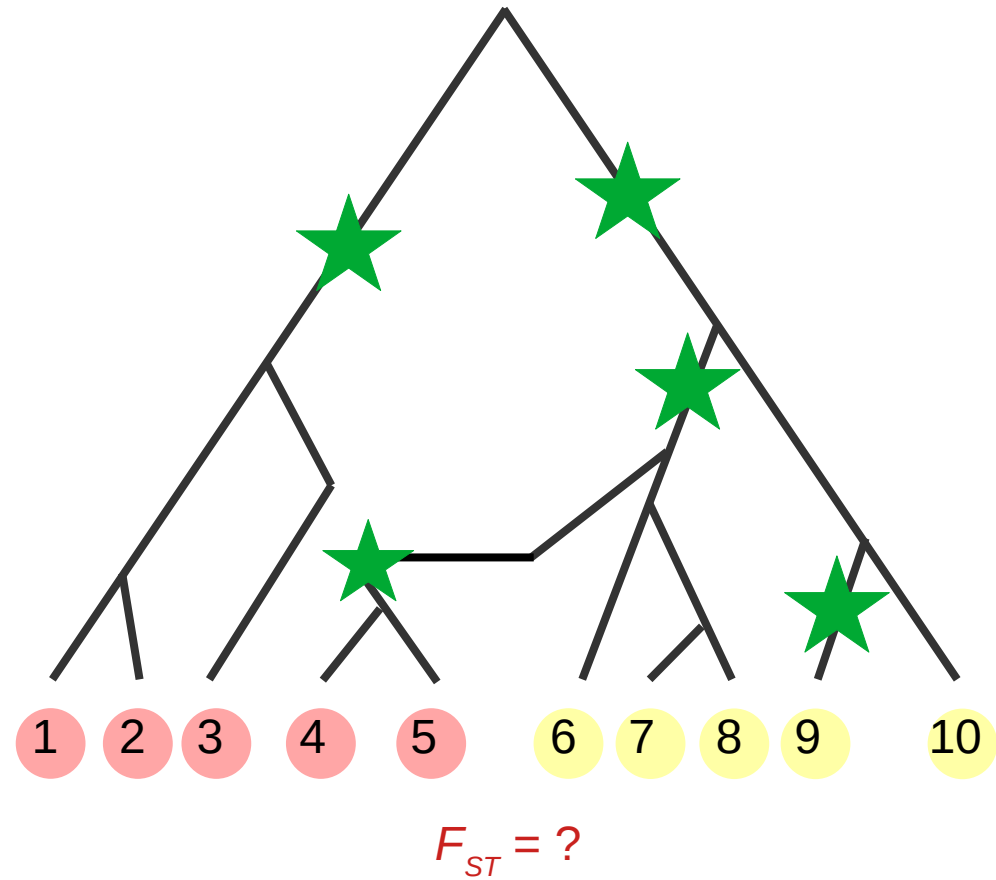
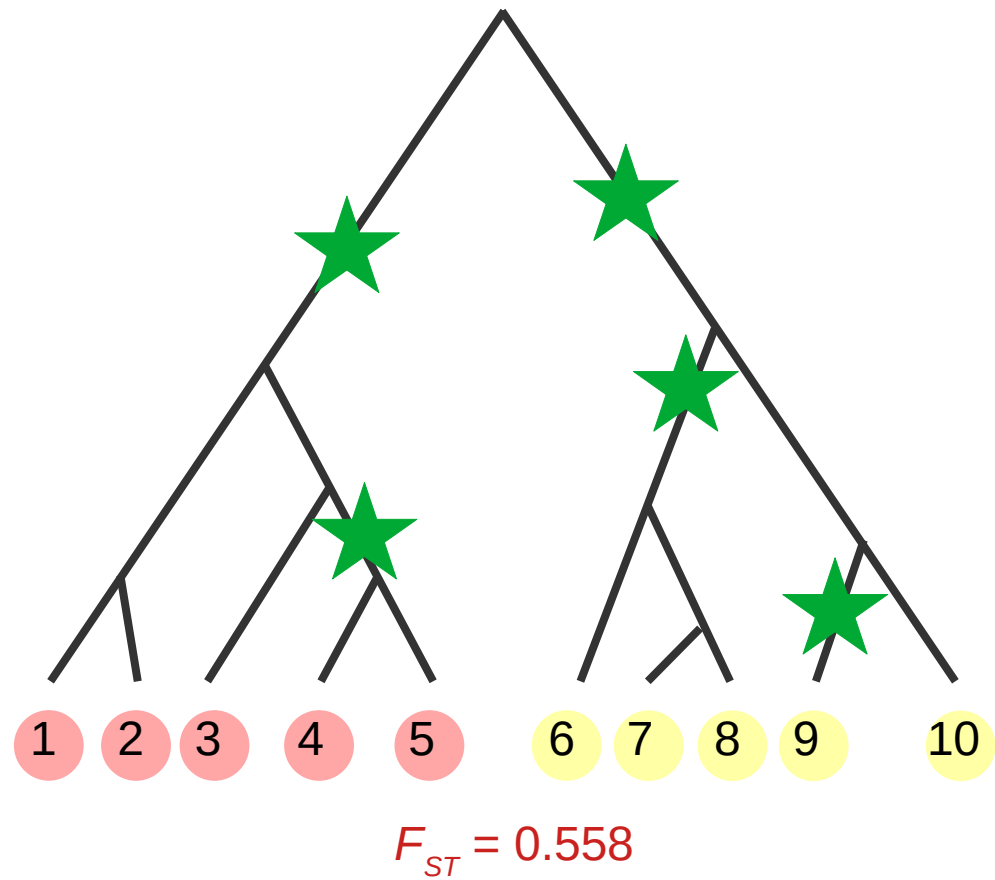
# The effect of migration on $F_{ST}$



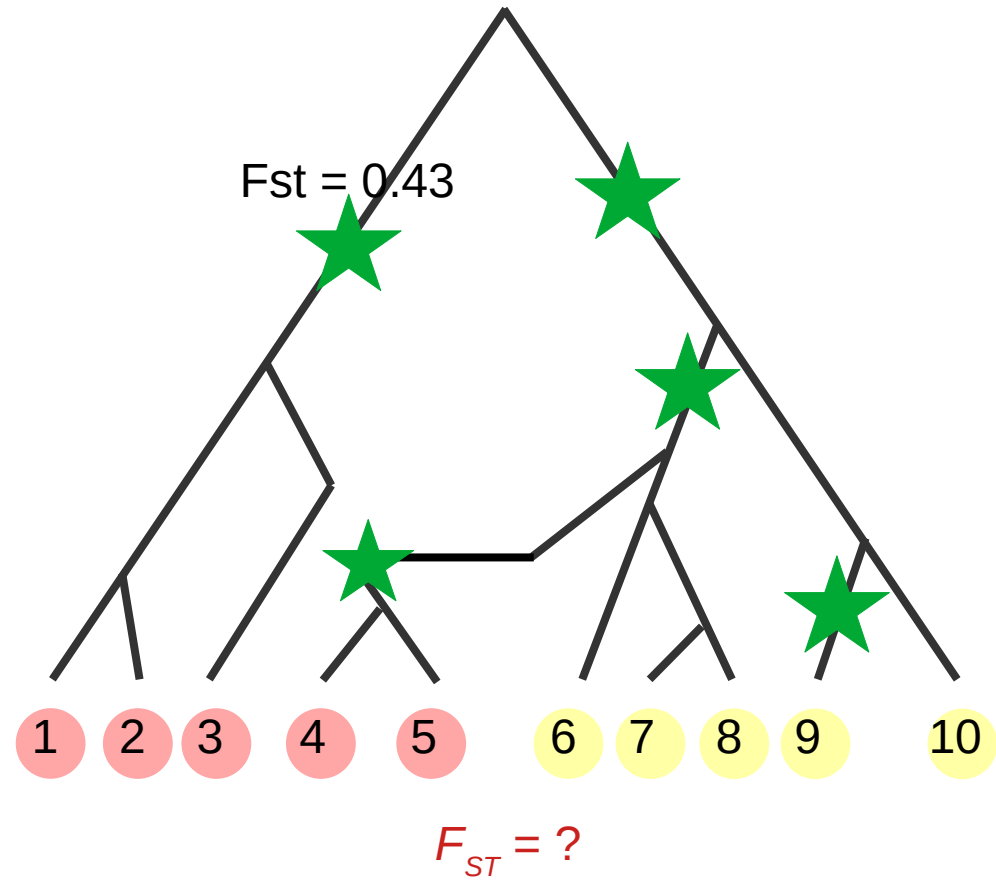
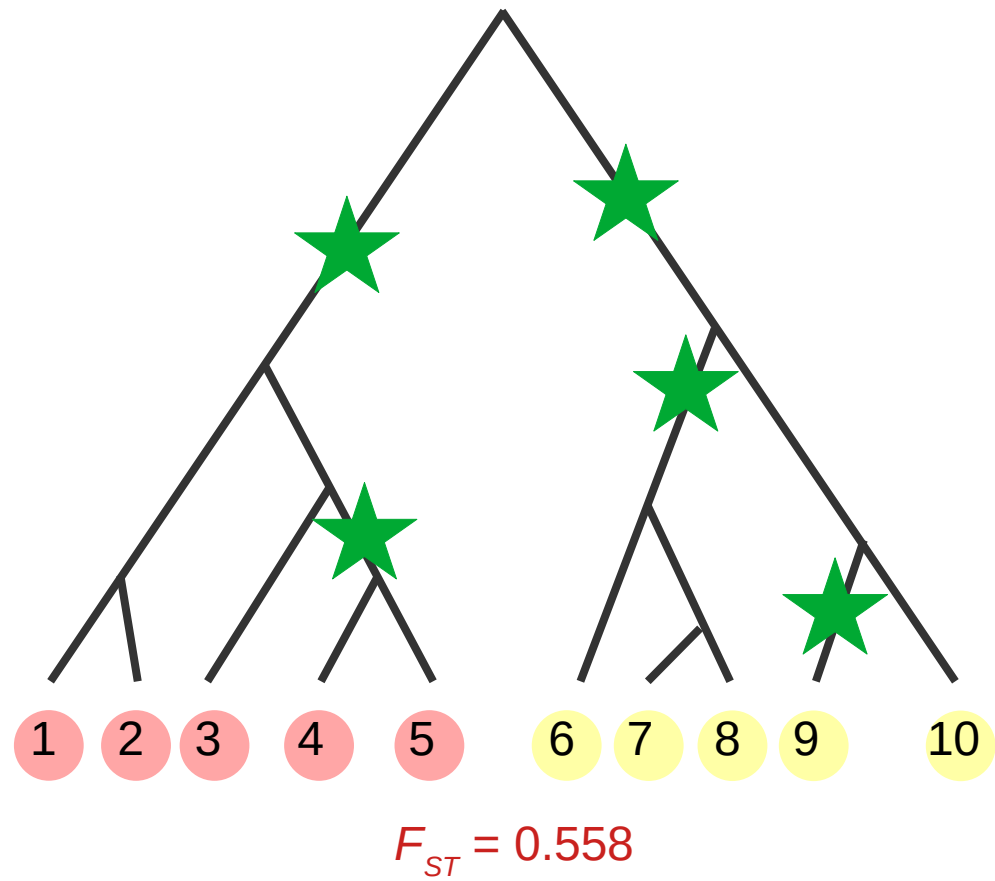
$$F_{ST} = 0.558$$



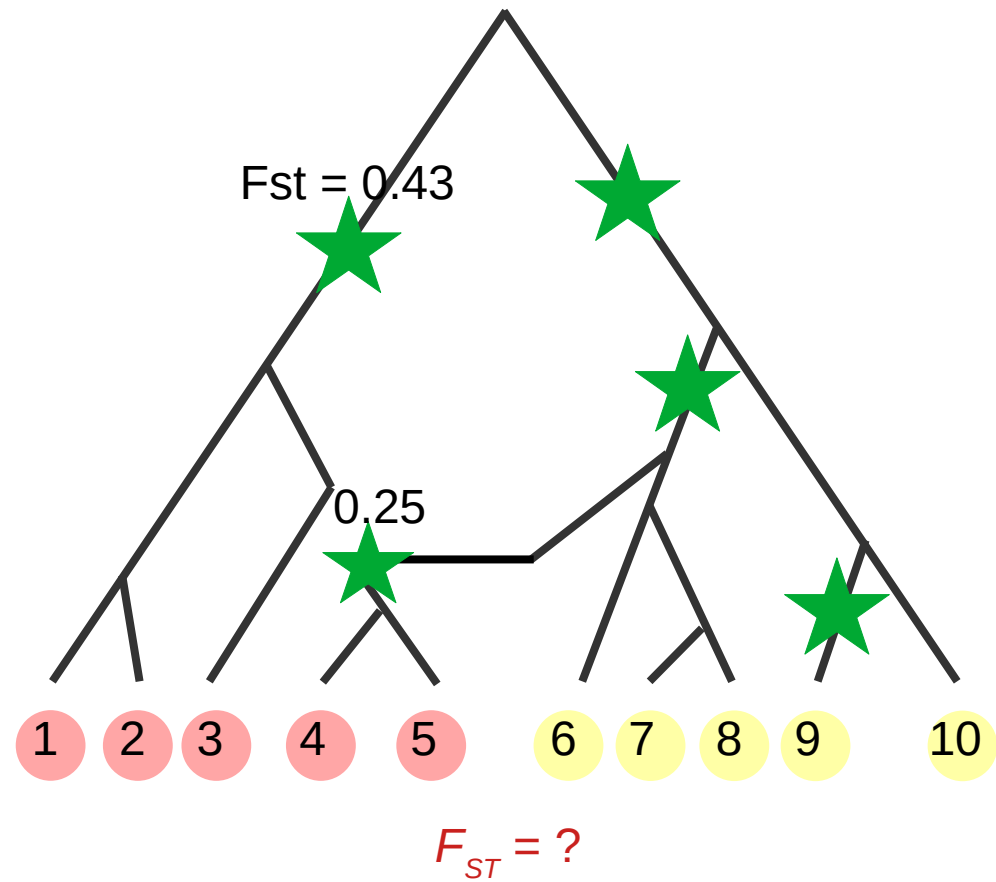
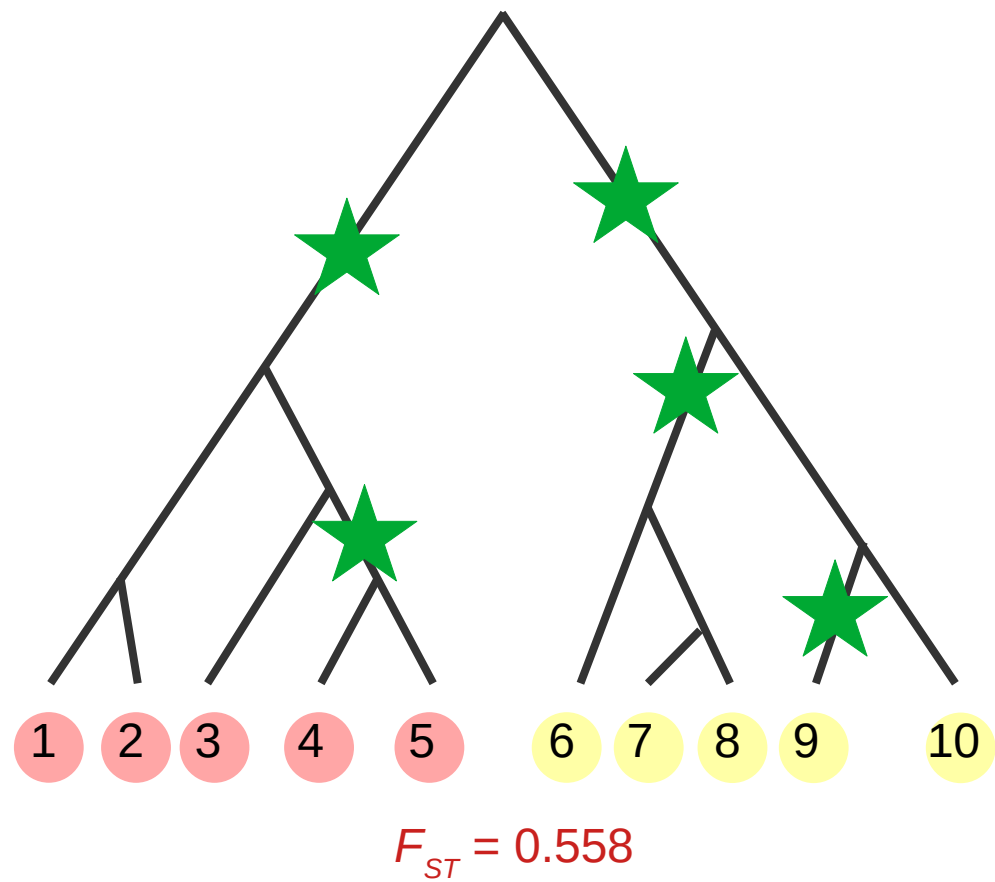
# The effect of migration on $F_{ST}$



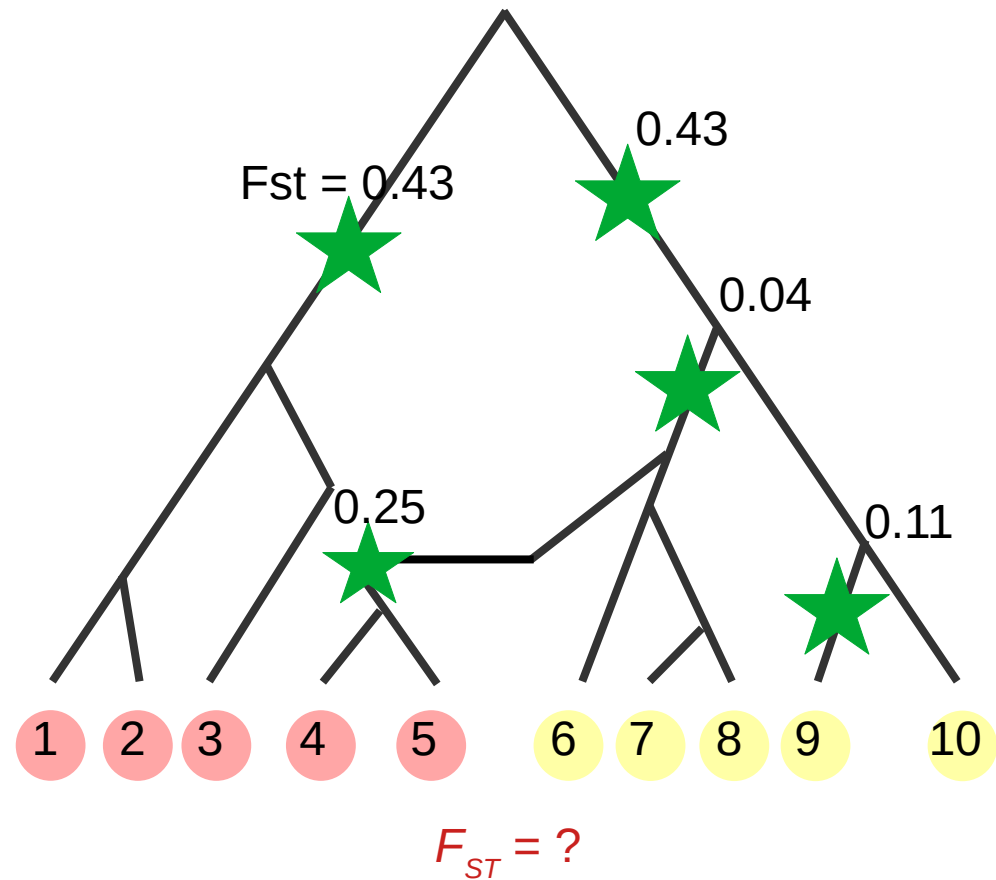
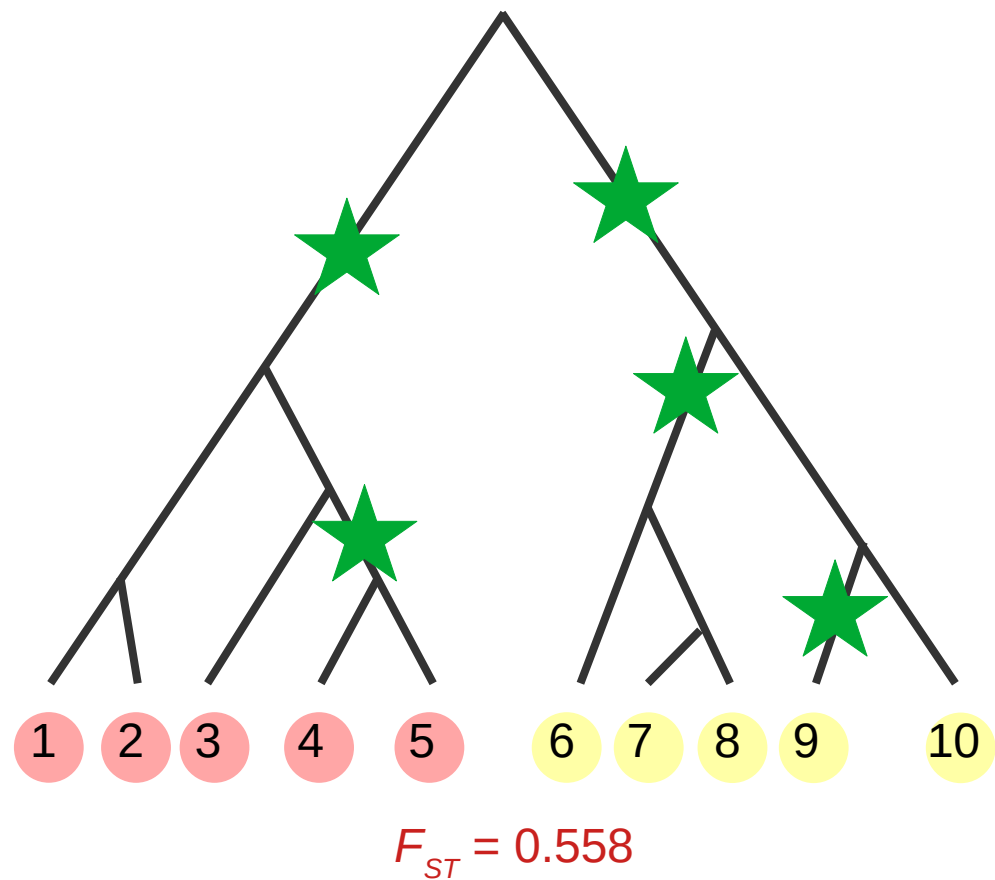
# The effect of migration on $F_{ST}$



# The effect of migration on $F_{ST}$

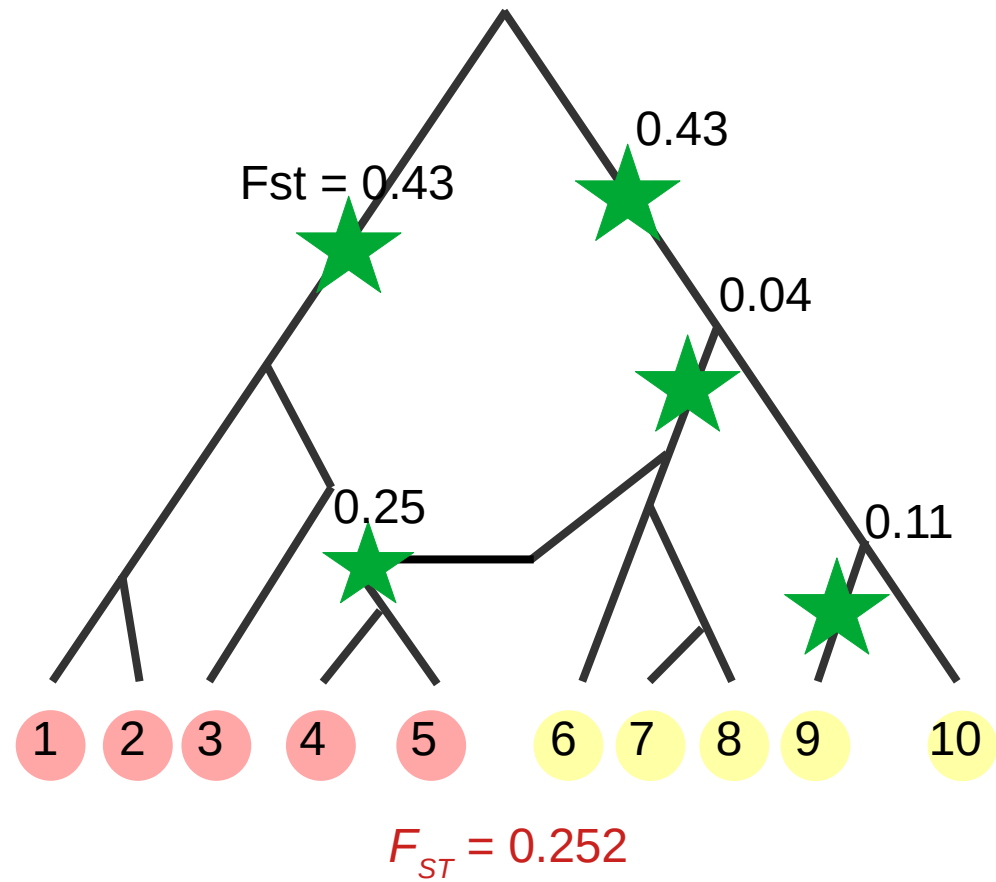
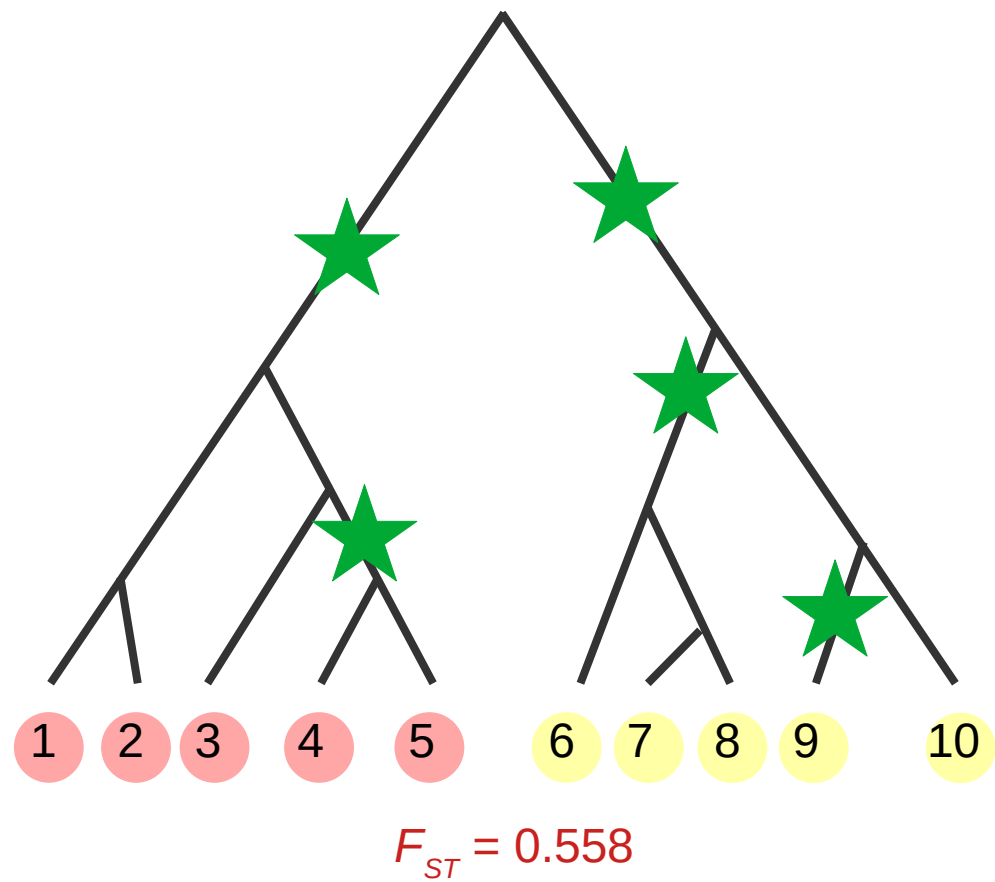


# The effect of migration on $F_{ST}$

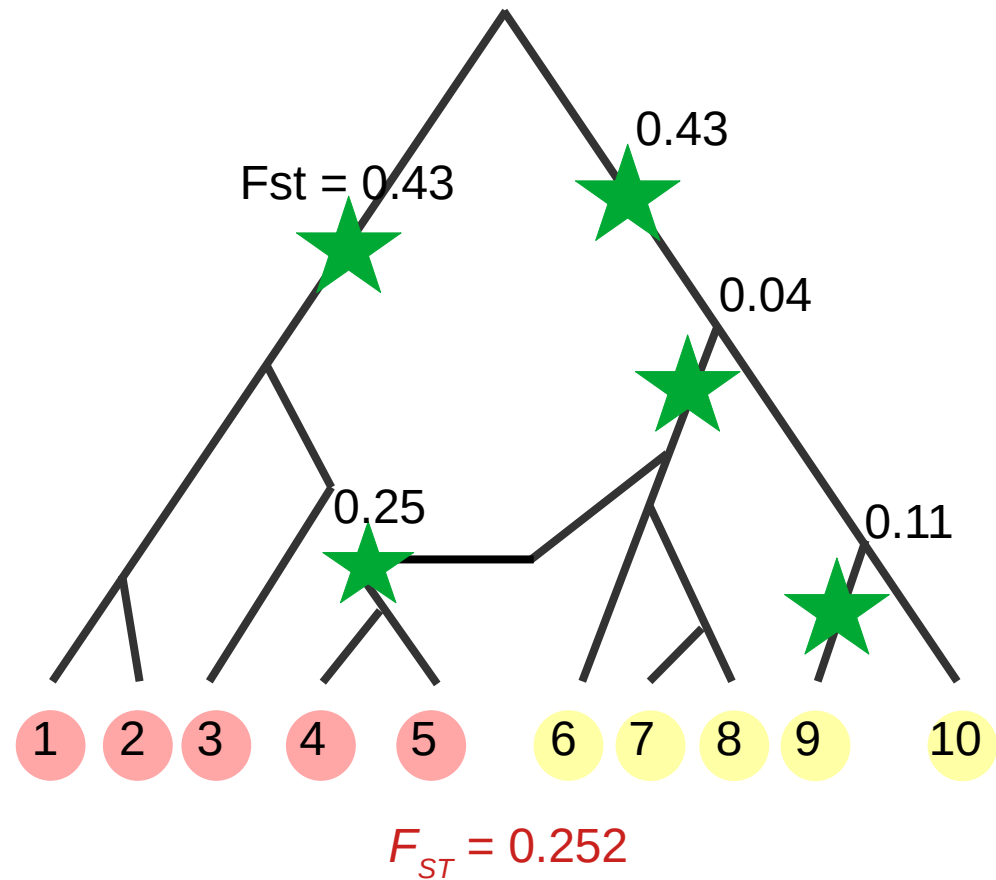
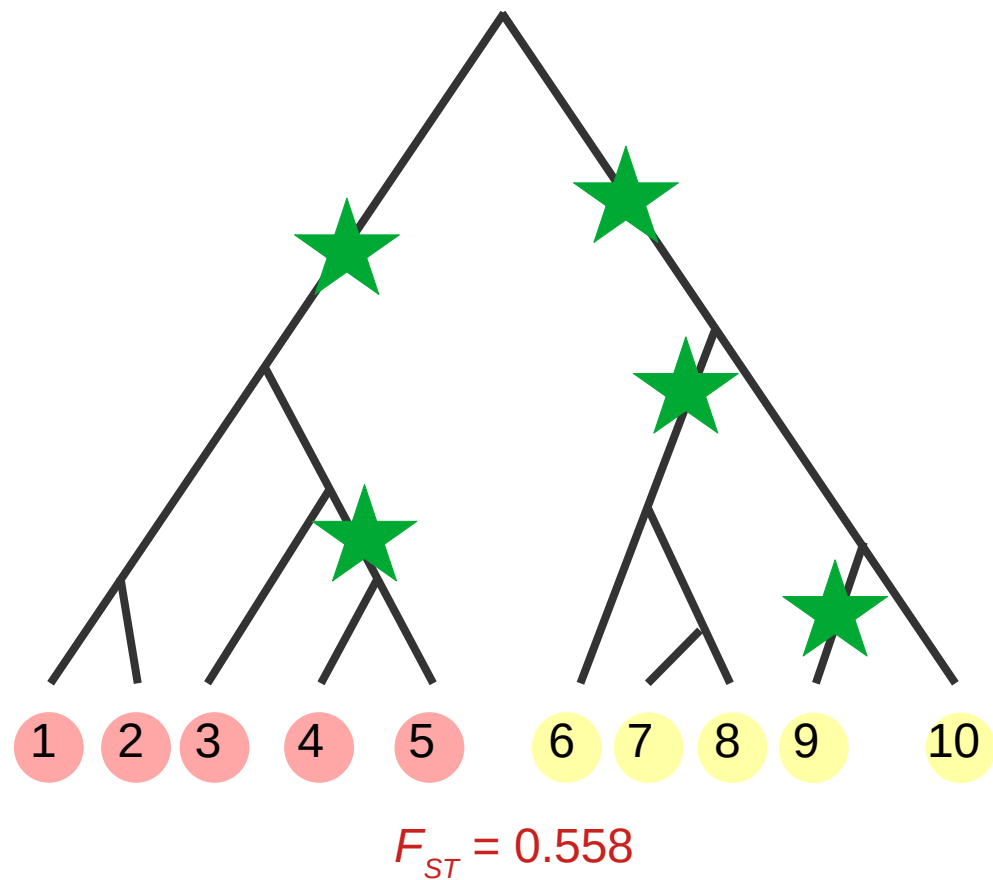




# The effect of migration on $F_{ST}$



# The effect of migration on $F_{ST}$

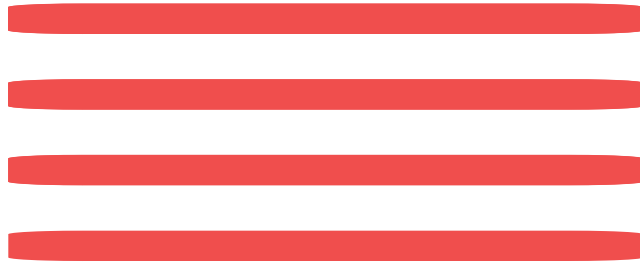


Migration ↗  $F_{ST}$  ↘

# Migration + recombination + genetic drift

Modern human pop.

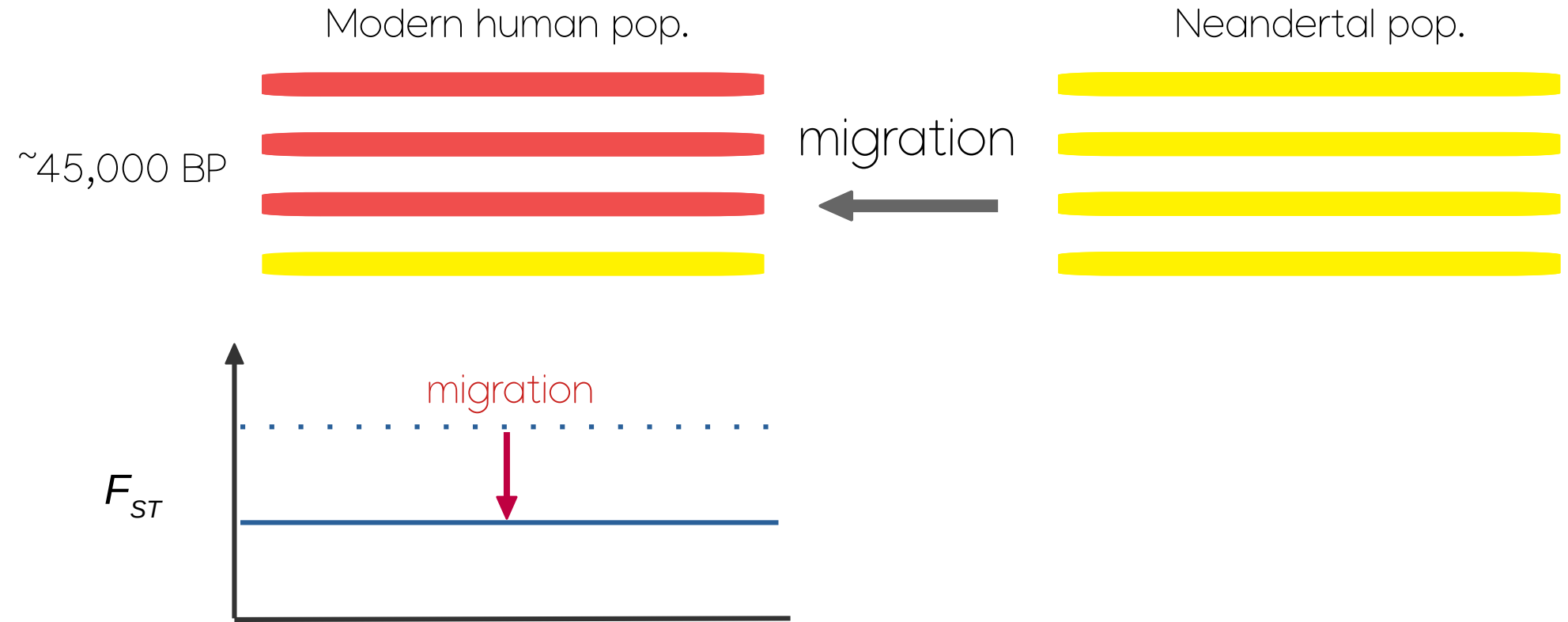
Neandertal pop.



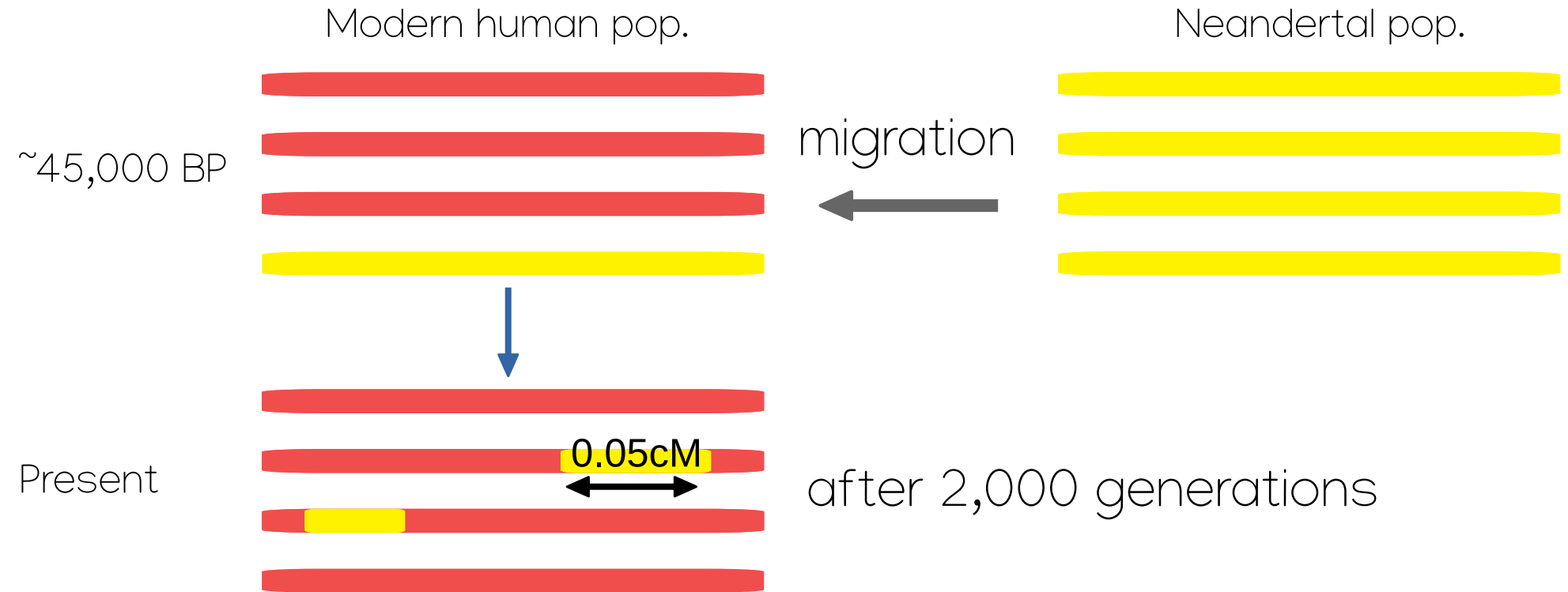
~50,000 BP



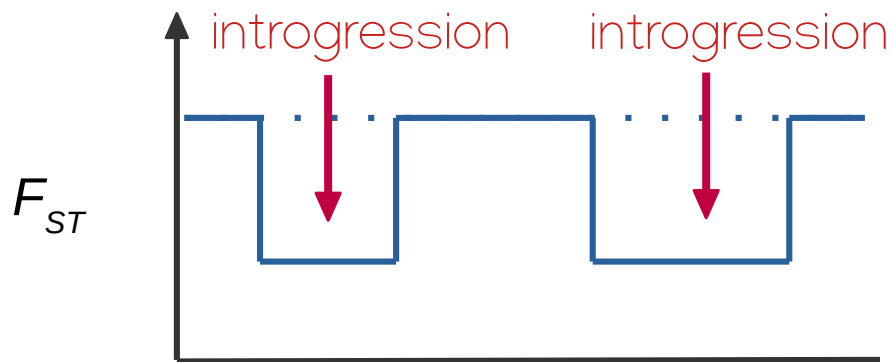
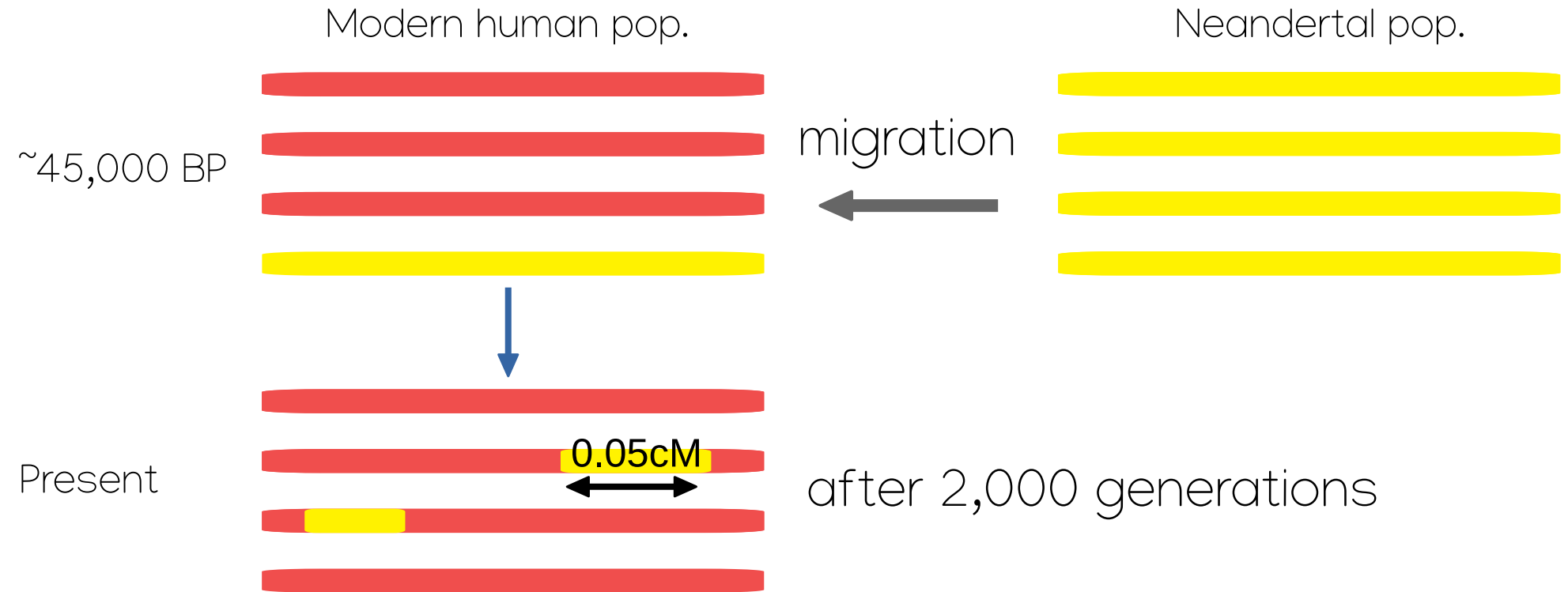
# Migration + recombination + genetic drift



# Migration + recombination + genetic drift



# Migration + recombination + genetic drift









# The Wahlund effect

```
q1=1-p1
q2=1-p2
q3=1-p3

p = (p1+p2+p3)/3
q = (q1+q2+q3)/3

fAa_1 = 2*p1*q1
fAa_2 = 2*p2*q2
fAa_3 = 2*p3*q3

fAa = 2 * p * q

sigma_sqr = (1/3) * ( (p1-p)**2 + (p2-p)**2 + (p3-p)**2 )

res = round(c(fAa_1, fAa_2, fAa_3, fAa, sigma_sqr, sigma_sqr/(p*q), (fAa_1+fAa_2+fAa_3)/3 + 2*sigma_sqr), 2)

names(res) = c('fAa (pop 1)', 'fAa (pop 2)', 'fAa (pop 3)', 'fAa (tot. pop)', 'sigma_sqr', 'Fst', 'mean(fAa) + 2 sigma_sqr')

tmp = paste("\n#pop. 1\nAA\tAa\taa\n", p1**2, "\t", 2*p1*q1, "\t", q1**2, "\n", sep="")
cat(tmp)

tmp = paste("\n#pop. 2\nAA\tAa\taa\n", p2**2, "\t", 2*p2*q2, "\t", q2**2, "\n", sep="")
cat(tmp)

tmp = paste("\n#pop. 3\nAA\tAa\taa\n", p3**2, "\t", 2*p3*q3, "\t", q3**2, "\n", sep="")
cat(tmp)

tmp = paste("\n#pop. total\nAA\tAa\taa\n", p**2, "\t", 2*p*q, "\t", q**2, "\n", sep="")
cat(tmp)
return(t(t(res)))
```





## **Comparative population genomics in animals uncovers the determinants of genetic diversity**

J. Romiguier<sup>1,2</sup>, P. Gayral<sup>1,3</sup>, M. Ballenghien<sup>1</sup>, A. Bernard<sup>1</sup>, V. Cahais<sup>1</sup>, A. Chenuil<sup>4</sup>, Y. Chiari<sup>5</sup>, R. Darnat<sup>1</sup>, L. Duret<sup>6</sup>, N. Faivre<sup>1</sup>, E. Loire<sup>1</sup>, J. M. Lourenco<sup>1</sup>, B. Nabholz<sup>1</sup>, C. Roux<sup>1,2</sup>, G. Tsagkogeorga<sup>1,7</sup>, A. A.-T. Weber<sup>4</sup>, L. A. Weinert<sup>1,8</sup>, K. Belkhir<sup>1</sup>, N. Bierne<sup>1</sup>, S. Glémin<sup>1</sup> & N. Galtier<sup>1</sup>

## RESEARCH ARTICLE

# Natural Selection Constrains Neutral Diversity across A Wide Range of Species

**Russell B. Corbett-Detig<sup>1,2</sup>, Daniel L. Hartl<sup>1</sup>, Timothy B. Sackton<sup>1\*</sup>**

**1** Department of Organismic and Evolutionary Biology, Harvard University, Cambridge Massachusetts, United States of America, **2** Department of Integrative Biology, University of California, Berkeley, Berkeley, California, United States of America

\* [tsackton@oeb.harvard.edu](mailto:tsackton@oeb.harvard.edu)

# Statistical evaluation of alternative models of human evolution

**Nelson J. R. Fagundes<sup>†‡§</sup>, Nicolas Ray<sup>§</sup>, Mark Beaumont<sup>¶</sup>, Samuel Neuenschwander<sup>§||</sup>, Francisco M. Salzano<sup>†++</sup>, Sandro L. Bonatto<sup>†,++</sup>, and Laurent Excoffier<sup>§++</sup>**

<sup>†</sup>Laboratório de Biologia Genômica e Molecular, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), 90619-900 Porto Alegre, RS, Brazil; <sup>‡</sup>Departamento de Genética, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil; <sup>§</sup>Computational and Molecular Population Genetics (CMPG), Zoological Institute, University of Bern, CH-3012 Bern, Switzerland; <sup>¶</sup>School of Animal and Microbial Sciences, University of Reading, Reading RG6 6AJ, United Kingdom; and <sup>||</sup>Department of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland

Contributed by Francisco M. Salzano, August 31, 2007 (sent for review August 1, 2007)

## RESEARCH ARTICLE

# Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence

**Camille Roux<sup>1,2,3\*</sup>, Christelle Fraïsse<sup>1,2,4</sup>, Jonathan Romiguier<sup>1,2,3</sup>, Yoann Anciaux<sup>1,2</sup>, Nicolas Galtier<sup>1,2</sup>, Nicolas Bierne<sup>1,2</sup>**

**1** Université Montpellier, Montpellier, France, **2** CNRS Institut des Sciences de l'Évolution, CNRS-UM-IRD-EPHE, Montpellier, France, **3** Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland, **4** Institute of Science and Technology, Klosterneuburg, Austria

\* [camille.roux.1983@gmail.com](mailto:camille.roux.1983@gmail.com)



[nature](#) > [nature ecology & evolution](#) > [articles](#) > article

[Published: 13 January 2017](#)

## Experimental test and refutation of a classic case of molecular adaptation in *Drosophila melanogaster*

[Mohammad A. Siddiq](#), [David W. Loehlin](#), [Kristi L. Montooth](#) & [Joseph W. Thornton](#) 

[Nature Ecology & Evolution](#) **1**, Article number: 0025 (2017) | [Cite this article](#)

**5223** Accesses | **16** Citations | **181** Altmetric | [Metrics](#)



RESEARCH ARTICLE

# Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis

**Nicolas Galtier\***

Institut des Sciences de l'Evolution UMR5554, Université Montpellier–CNRS–IRD–EPHE, Montpellier, France

\* [nicolas.galtier@univ-montp2.fr](mailto:nicolas.galtier@univ-montp2.fr)

