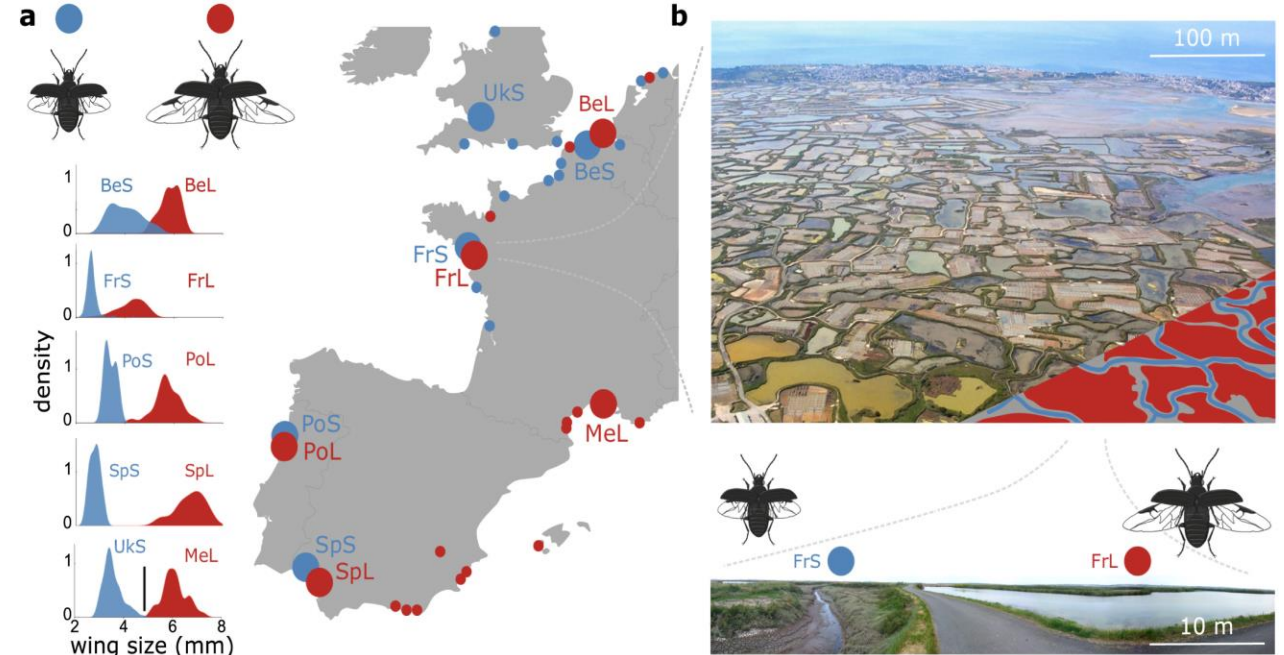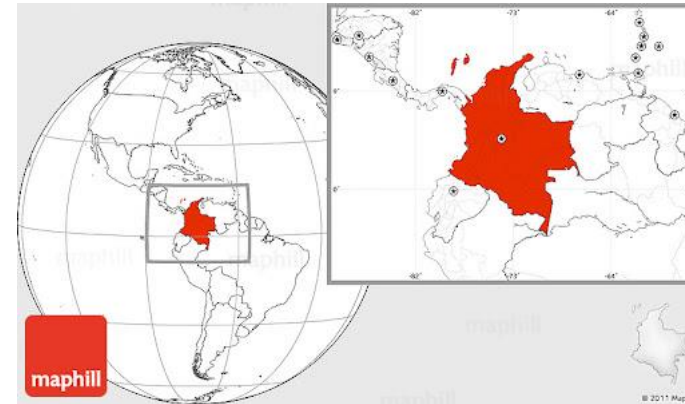# Why Populations Change: The Genetics Behind Evolution

Maria Madrid

July 31st, 2025

Who am I?

- Colombian CO

- PhD student, Eco-Evolutionary Genomics Lab

- Studying the DNA of beetles to understand their evolution 🪲

What will we learn today?

- Learn what allele frequencies are and how to calculate them

- Understand what population genetics is for

- Learn what SNPs are and how they're detected from sequencing data

- Understand and calculate F_{ST} and other metrics as measures of population differentiation

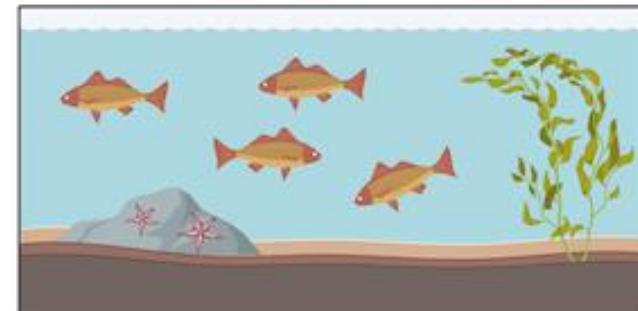# Populations, communities, and ecosystems

- **Population**: groups of individuals of the <u>same species</u> that live together and reproduce

- **Community**: groups of <u>different species</u> living together

- **Ecosystem**: <u>all</u> living organisms and non-living material in the same geographical space
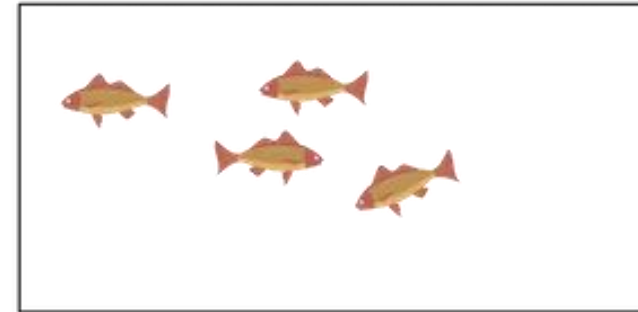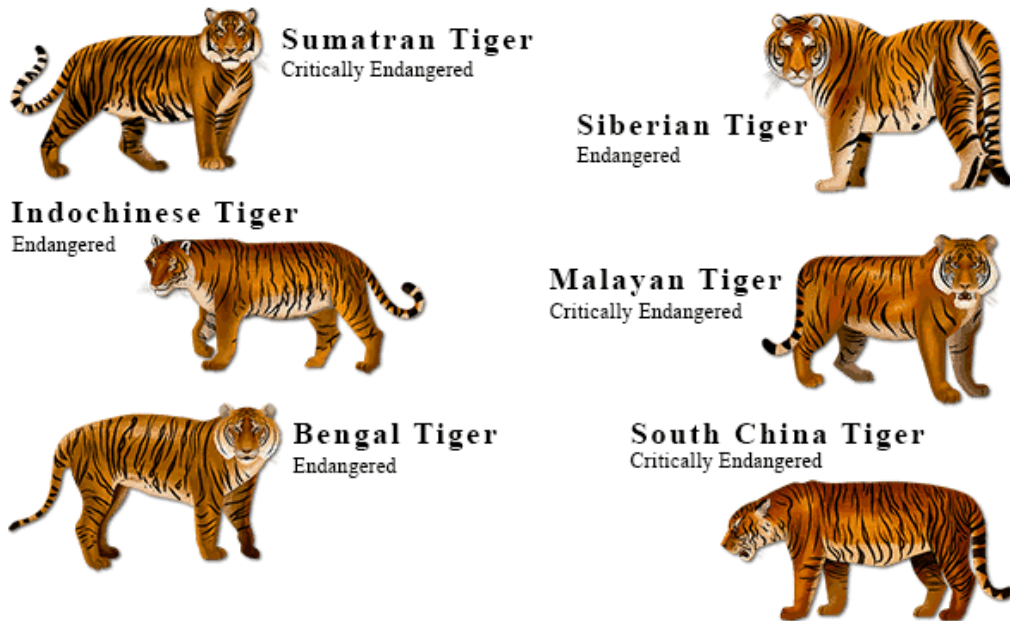

population
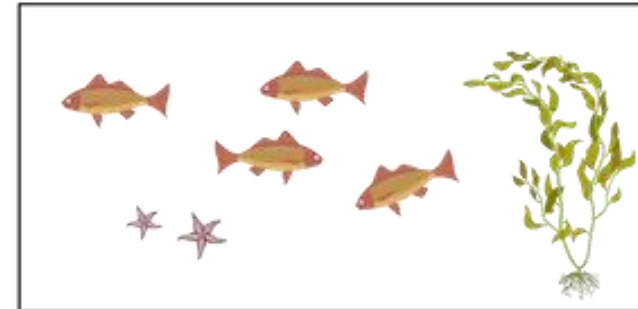
community

ecosystem

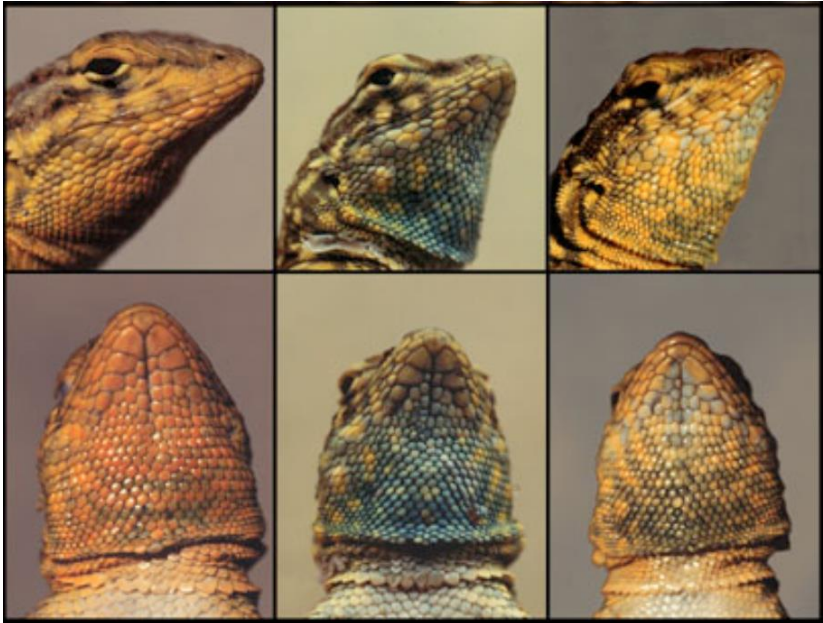# Should different populations of one species look the same?

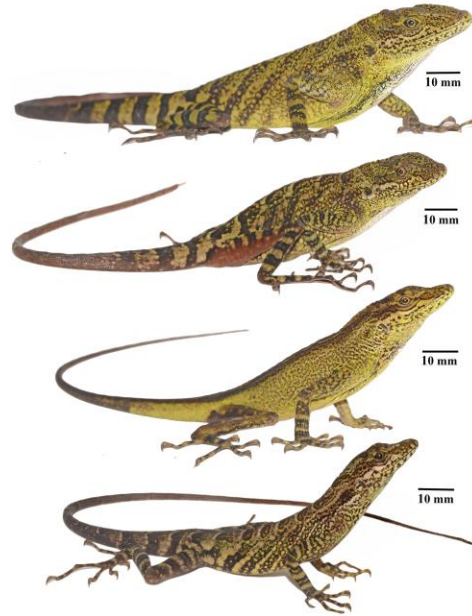Same species (*Panthera tigris*), different subpopulations

Different species of *Heliconius* butterflies

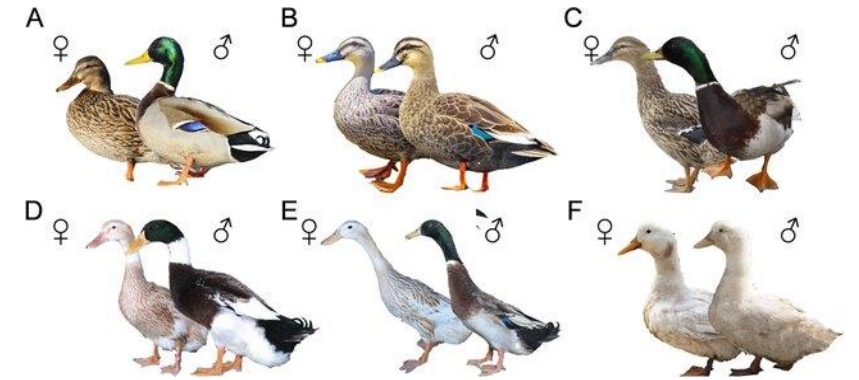# Should different populations of one species look the same?

Same species!

Different species

Same species!

Different species (different genera even!)

Should different populations of one species look the same?
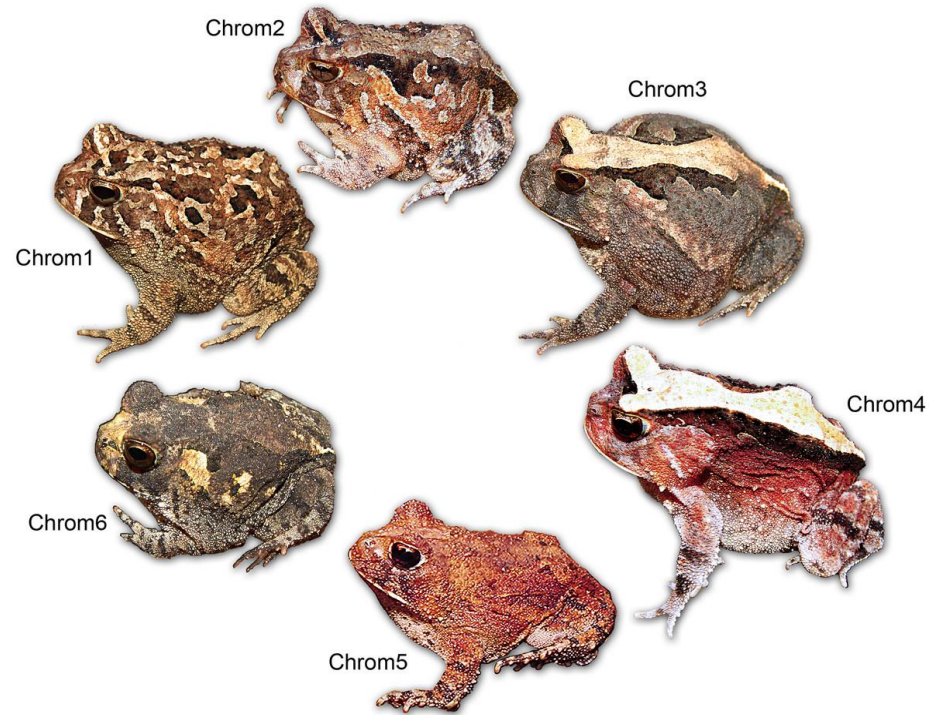
Same species!

Same species!

Why wou... so similar?

- **Convergent** ...evolve similar traits: not be... similar environmen...
- These simila... work the same, but ev...
- It's nature's ...ys
- Classic exa...
  - Wings in b...
  - Fins in dol...
- Evolution is... pecies toward simi...

Convergent evolution
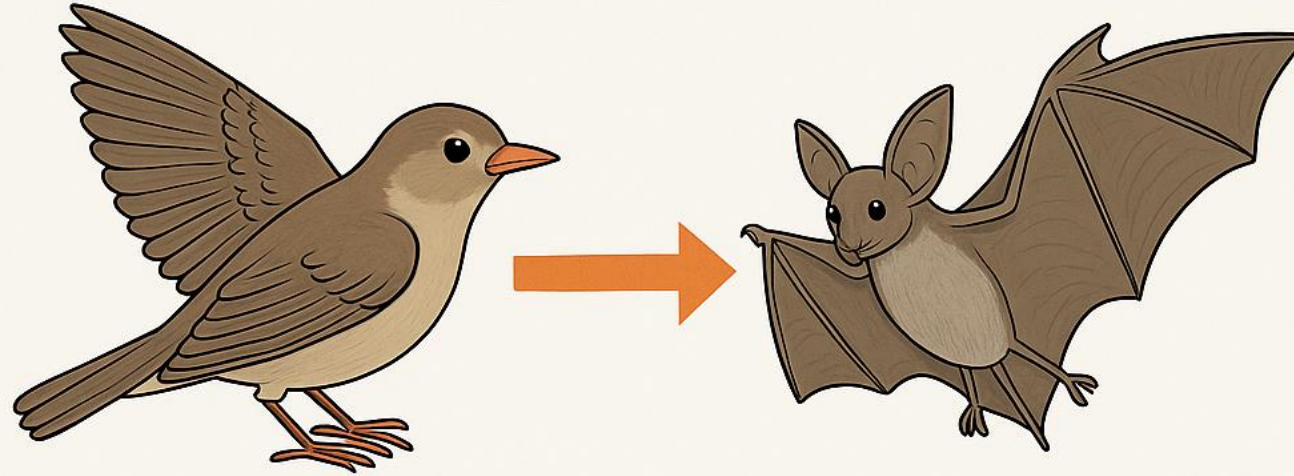
Bird wing → Bat wing
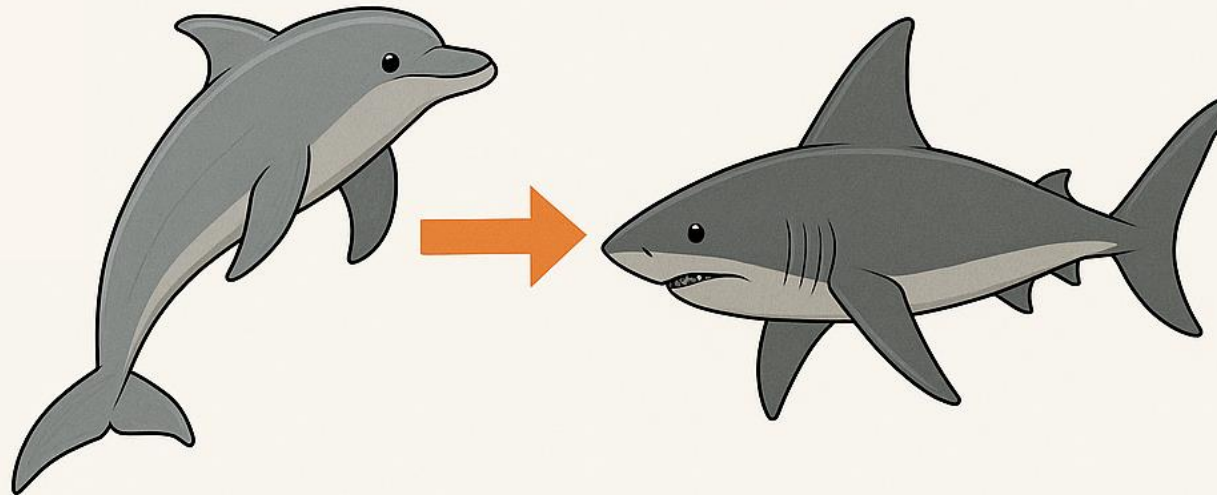
Dolphin fin → Shark fin

Populations share a gene pool

- The combination of all the genes (including **alleles**) present in a reproducing population or species
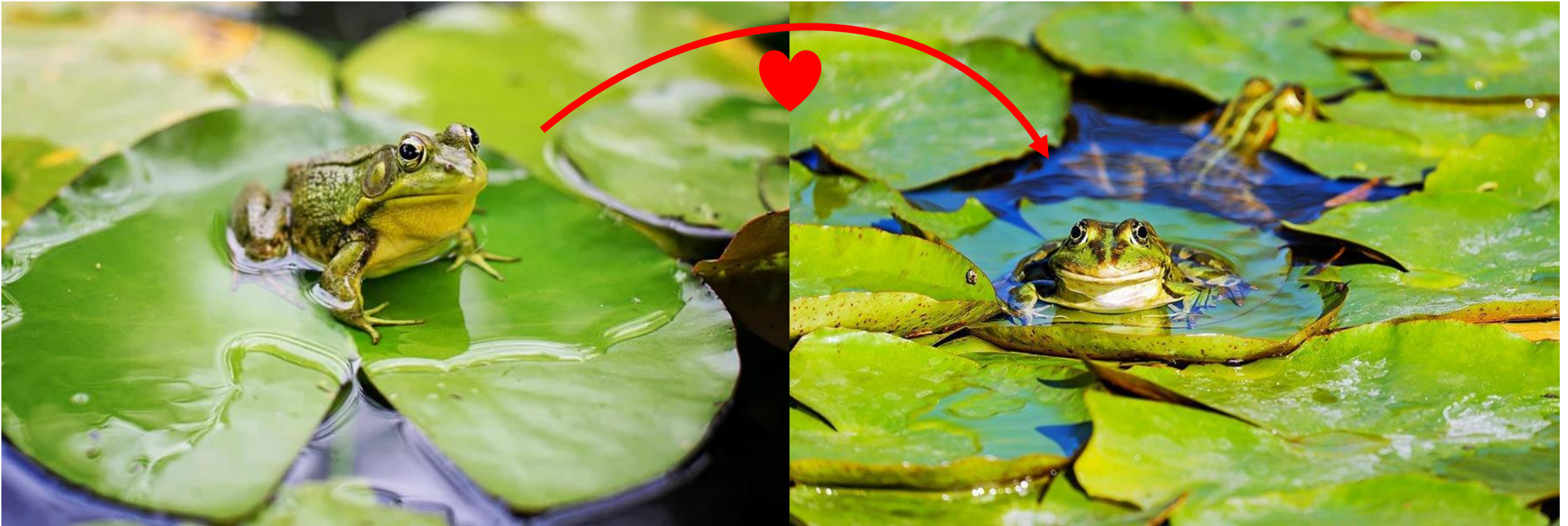
All the genetic options!

- A gene pool has all the genetic possibilities of a **population**

- Three "versions" of a gene (X_1, X_2, X_3) -> alleles!

- Since individuals are **diploid**, we assume each one has 2 copies

- We can even calculate the variation in this population

Each oval represents a single individual.

Three **alleles**—X_1, X_2, and X_3—exist at locus X in this population. Each diploid individual carries two copies of gene X, which may be the same or different alleles. No diploid individual can have more than two of the alleles.

Gene pool of population
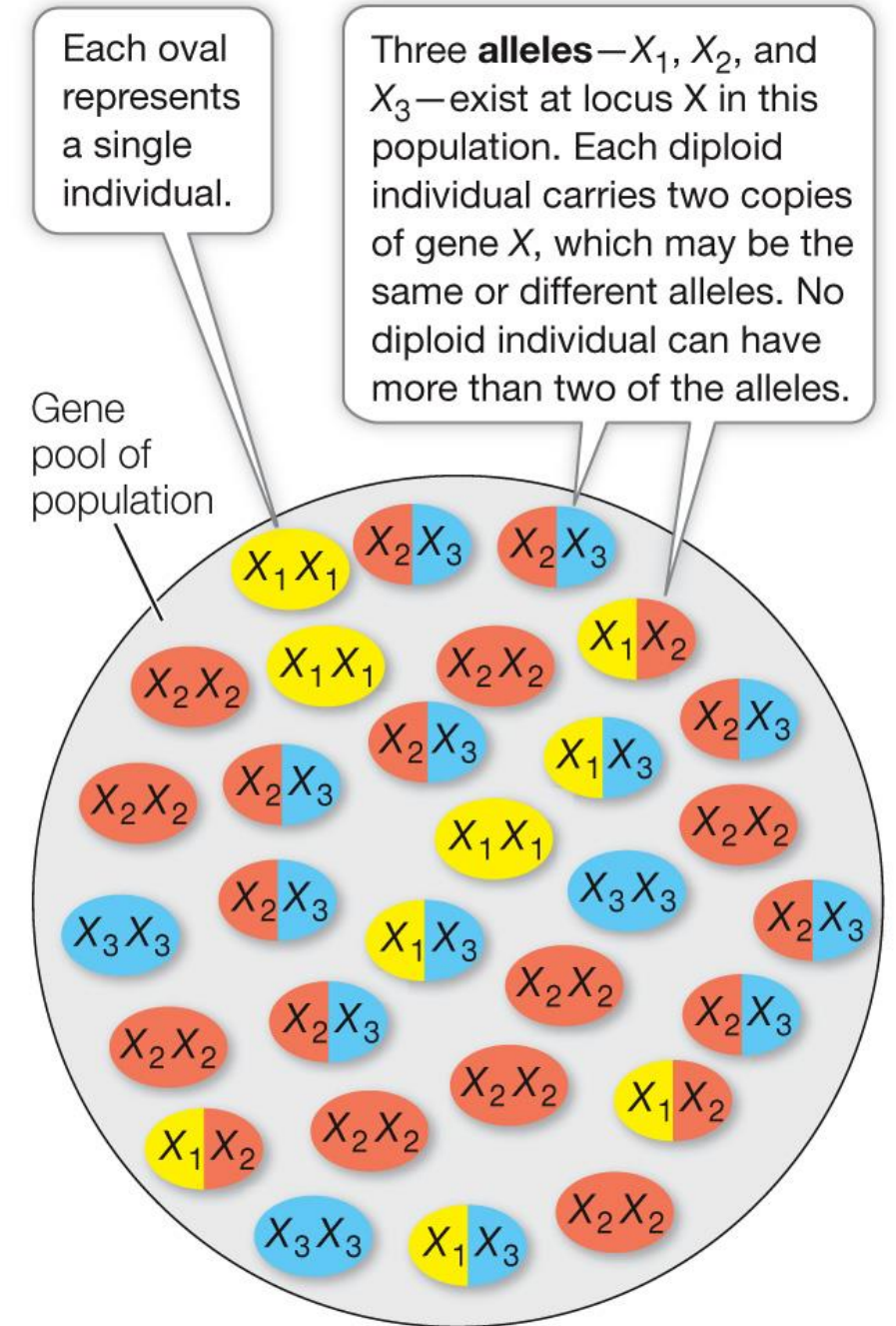
How to measure variation: allele frequencies

- In a simple, two-allele system
- We will refer to them as alleles A and B
- In one population, A + B = 1
- How many individuals do we have?
- n = 20
- How many AA individuals?
- AA = 10
- How many BB individuals?
- BB = 6
- How many AB individuals?
- AB = 4
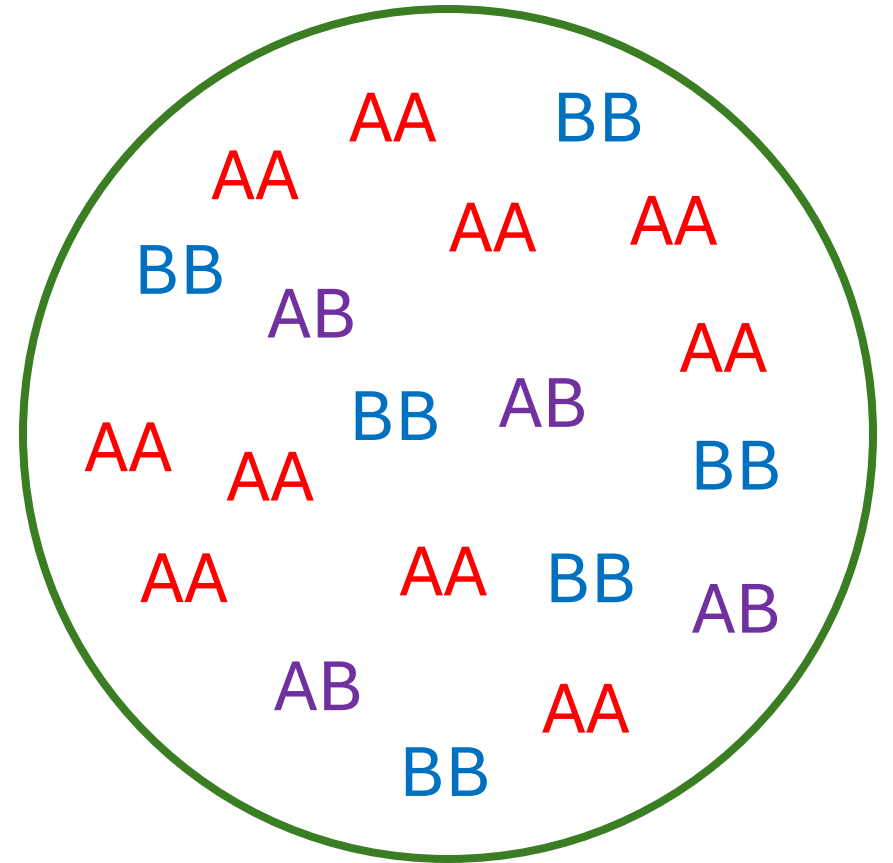
How to measure variation: allele frequencies

- To calculate the allele frequency for this population
- AA + AB + BB = 20
- A = 10 + 4/2
- A = 12/20 = 0.6
- B = 6 + 4/2
- B = 8/20 = 0.4
- A + B = 1
- 0.6 + 0.4 = 1
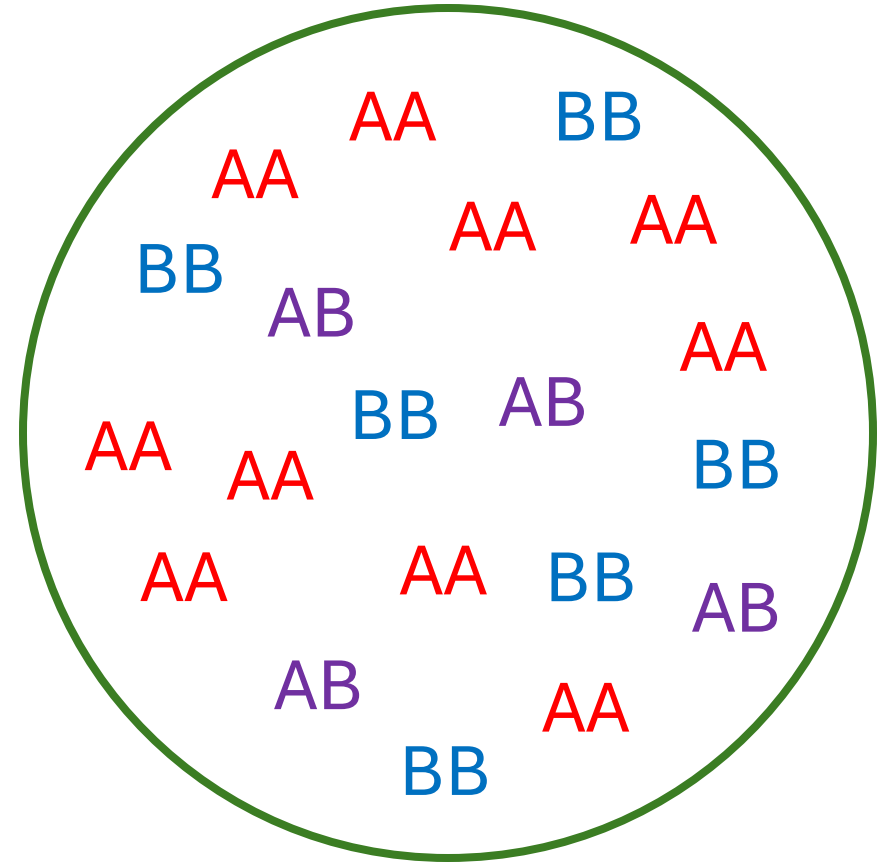
What about the first example?

- Same principle, $X_1 + X_2 + X_3 = 1$

- n = 30

- $X_1 X_1$ = 3

- $X_1 X_2$ = 3

- $X_1 X_3$ = 3

- $X_2 X_2$ = 9

- $X_2 X_3$ = 9

- $X_3 X_3$ = 3

Gene pool of population

more than two of the alleles.
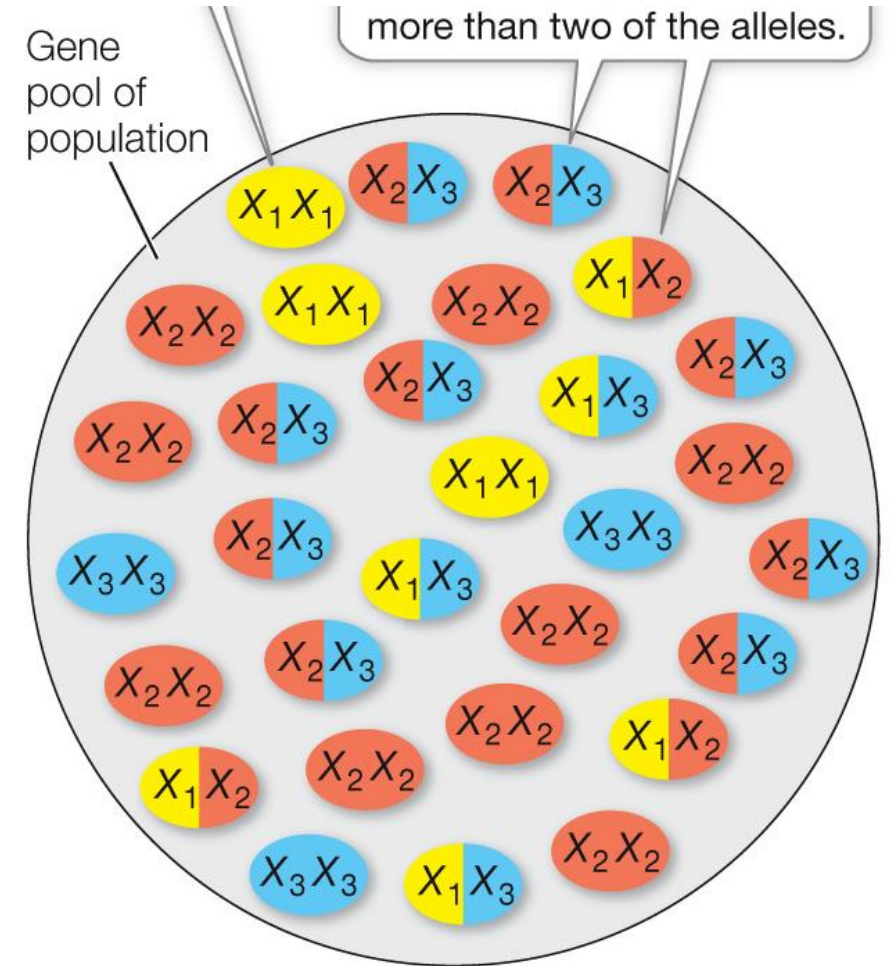
What about the first example?

- Same principle, $X_1 + X_2 + X_3 = 1$
- $X_1X_1 + X_1X_2 + X_1X_3 + X_2X_2 + X_2X_3 + X_3X_3 = 30$
- $X_1 = 3 + 3/2 + 3/2 = 6$
- $X_1 = 6/30 = 0.2$
- $X_2 = 9 + 3/2 + 9/2 = 15$
- $X_2 = 12/30 = 0.5$
- $X_3 = 3 + 3/2 + 9/2 = 9$
- $X_3 = 9/30 = 0.3$

Gene pool of population

more than two of the alleles.

X_1X_1 X_2X_3 X_2X_3 X_1X_2 X_2X_2 X_1X_1 X_2X_2 X_2X_3 X_2X_2 X_2X_3 X_2X_3 X_1X_3 X_2X_2 X_1X_1 X_3X_3 X_3X_3 X_2X_3 X_1X_3 X_2X_2 X_2X_3 X_2X_2 X_2X_3 X_1X_2 X_2X_2 X_2X_2 X_1X_2 X_3X_3 X_1X_3 X_2X_2

BEWARE! These calculations make **assumptions**!

- We assume this is a population that is not evolving, which is to say: this population is under **Hardy-Weinberg equilibrium**
- This assumes:
    - No mutations in their DNA
    - Individuals mate at random
    - There is no gene flow with other populations
    - The population size is infinitely large
    - There is no natural selection going on

$$p^2 + 2pq + q^2 = 1$$

Let's go back to this example

- Let's assume this is actually a very <u>stable</u> population of beetles!
- AA + AB + BB = 20
- A = 10 + 4/2
- A = 12/20 = 0.6
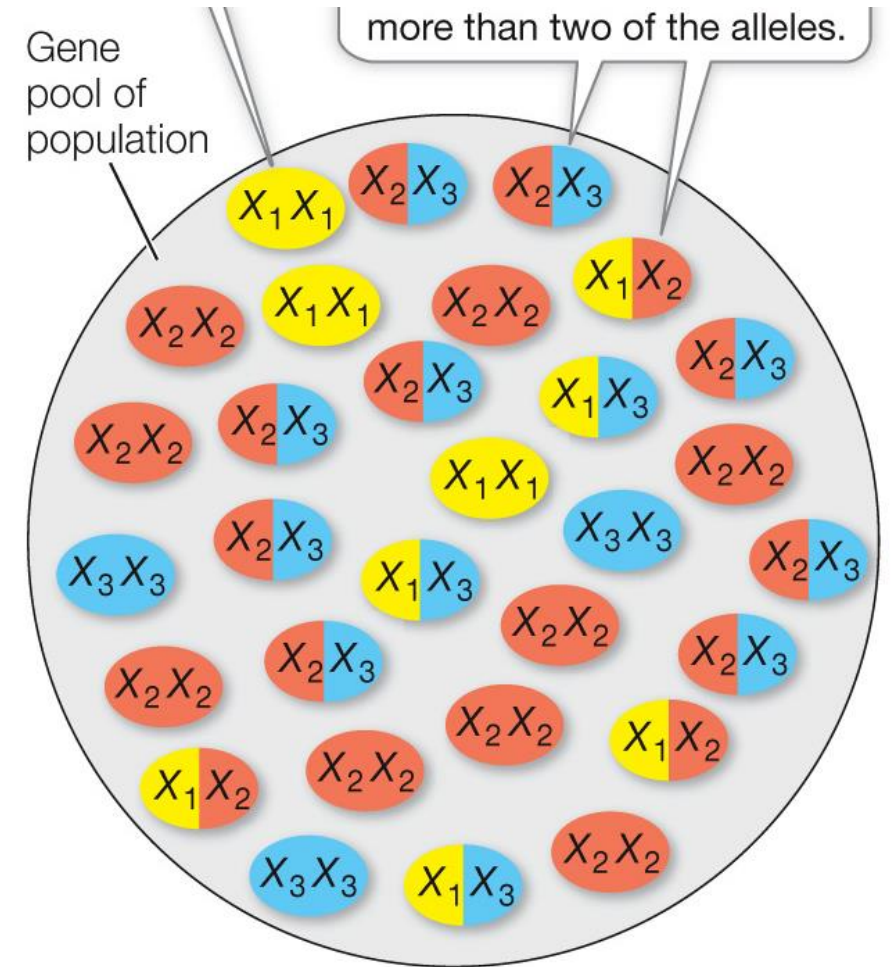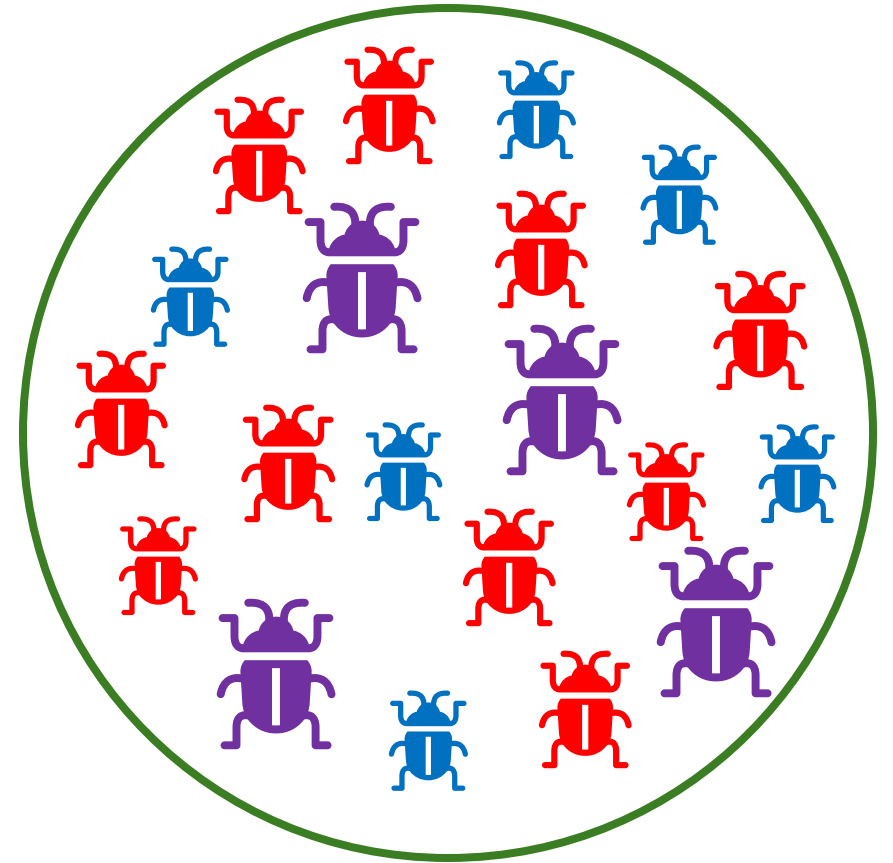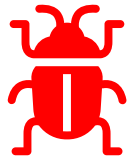- B = 6 + 4/2
- B = 8/20 = 0.4
- A + B = 1
- 0.6 + 0.4 = 1
- In this case, A = p and B = q

Hardy-Weinberg equilibrium (HWE)

AA   BB   AB

- We already know that p = 0.6 and q = 0.4
- If we plug in the numbers to our formula $p^2 + 2pq + q^2 = 1$, we can predict the **genotype frequencies** we'd expect to see if the population is <u>not evolving</u>
- $(0.6)^2 + (2)(0.6)(0.4) + (0.4)^2$
- 0.36 + 0.48 + 0.16

But why do we even care about allele and genotype frequencies?

- Because at its simplest, *evolution* is just the **change** of allele frequencies over time!

- Genotype frequencies tell us how alleles are combined in individuals

- If genotype frequencies don't match what HWE predicts, it can be a clue that:
  - **selection** is happening
  - there's population **structure**
  - there's non-random mating or migration happening

How exactly do allele frequencies change?

- There are several mechanisms:
  - Natural selection!
  - Genetic drift
  - Migration/gene flow
  - Mutations
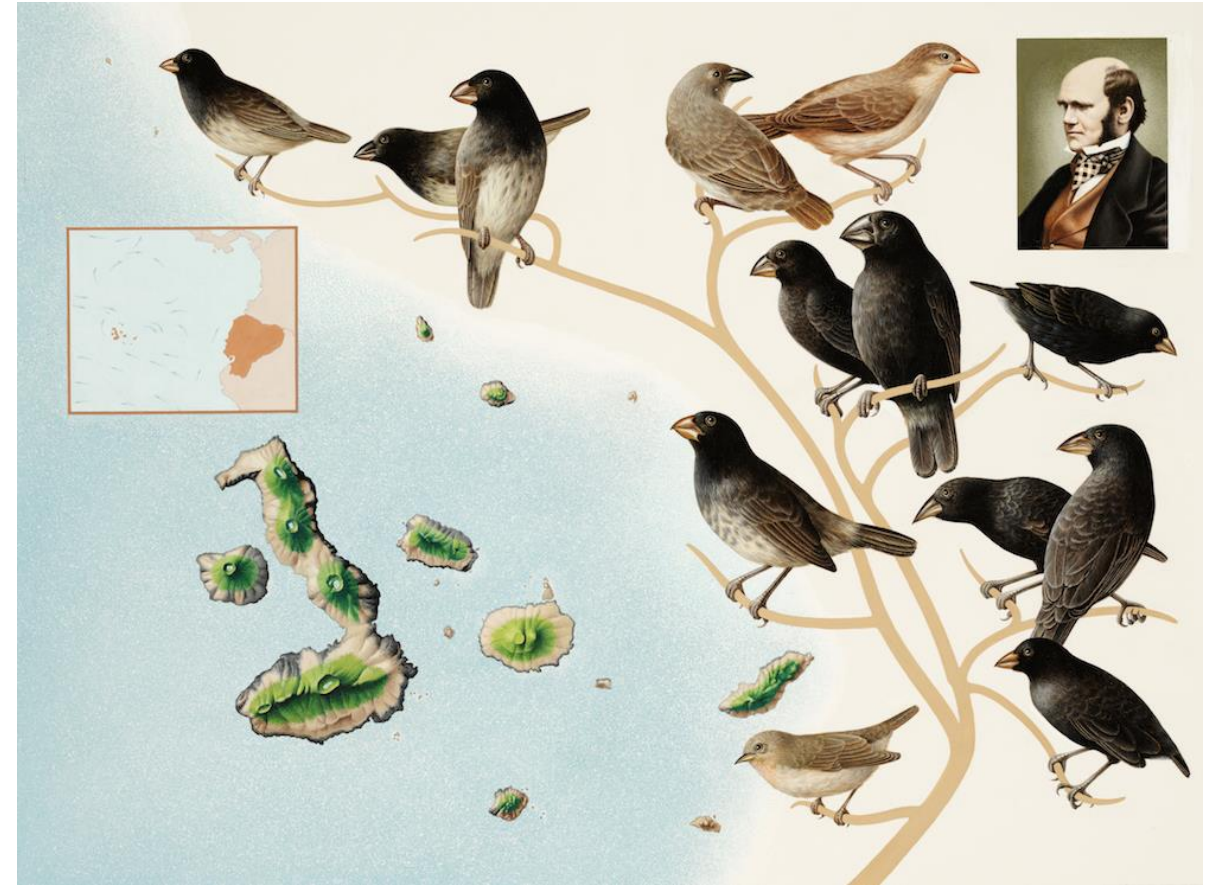
Genetic drift is random change in allele frequencies
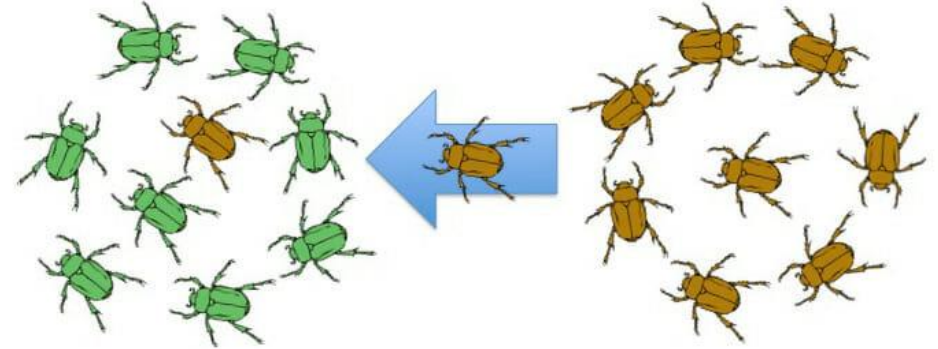
- A mechanism of evolution in which allele frequencies of a population change over generations at **random**

- It occurs in all populations, but its effects are <u>strongest in small populations</u>

- Can result in the loss or fixation of alleles

- Over time, genetic drift can reduce the genetic diversity within a population

- Not the same as **natural selection**!

Bottleneck and founder effects
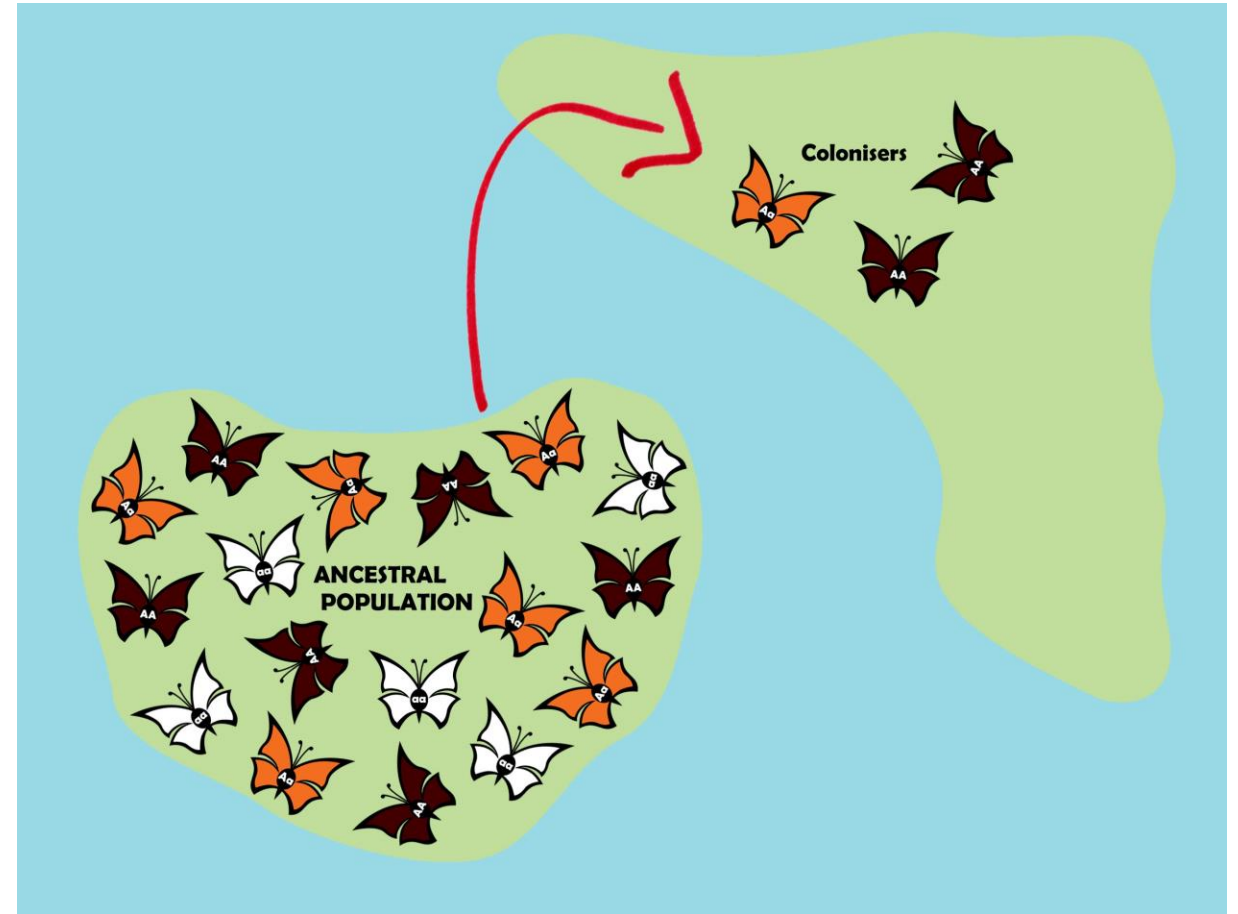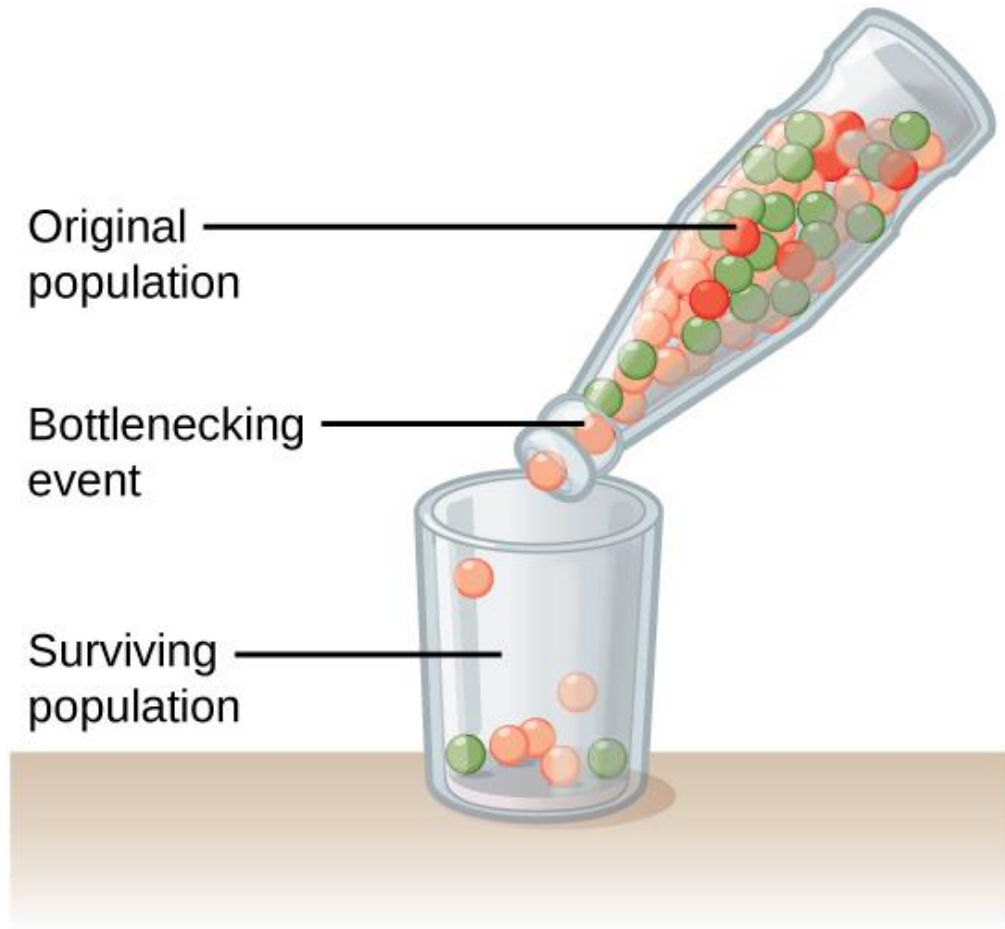
Migration between populations can result in gene flow

- The **transfer** of genetic material from one population to another

- Can take place between two **populations** of the same species

- Mediated by **reproduction** and vertical gene transfer from parent to offspring

- Well known example: humans and Neanderthals!

Neanderthal   Modern human

Lower cranium
Larger bowridge
Larger nose

Larger, broader rib cage

Shorter forearm

Broader hips
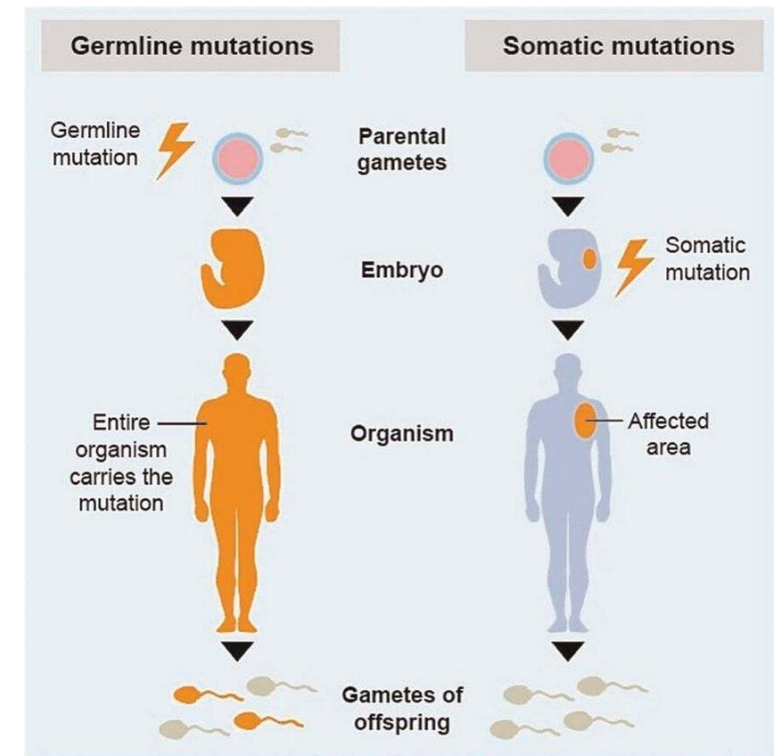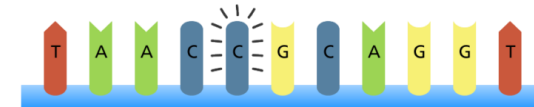
Shorter tibia

how
do we
know?

Mutations can cause permanent changes in our DNA

- **Mistakes** by the replication machinery

- **Germline** mutations (that occur in eggs and sperm) can be passed on to offspring, while **somatic** mutations (that occur in body cells) are not

- They introduce new alleles and can alter the frequencies of existing ones, contributing to **genetic diversity**
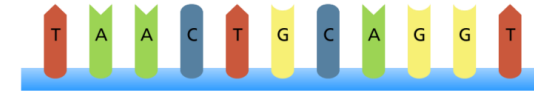
# Populations evolve, not individuals!

- Individuals can change, but <u>only populations evolve</u>
- Evolution is the change allele frequency within a *population's* gene pool over time
- It's impossible for a single individual to evolve in the sense of changing its genetic makeup in a way that would be passed on to its offspring
- Evolution happens across generations and affects the entire population
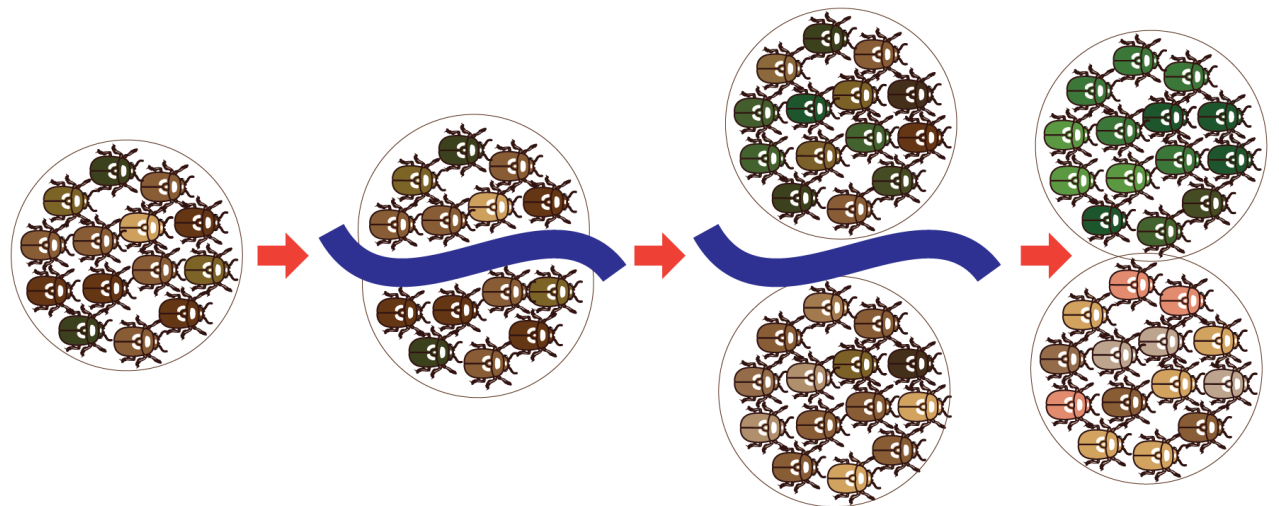
Now, let's zoom in to the DNA itself!

- Single nucleotide polymorphisms (**SNPs**) are the most important mistakes that lead to evolution
- Some of these mistakes can be irrelevant, some can be life-changing!
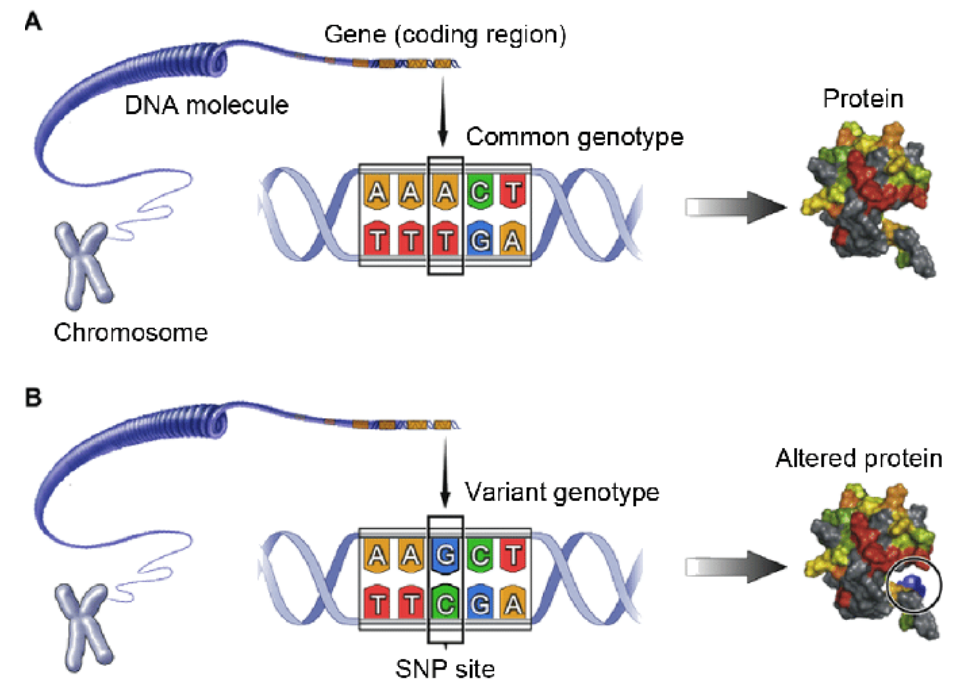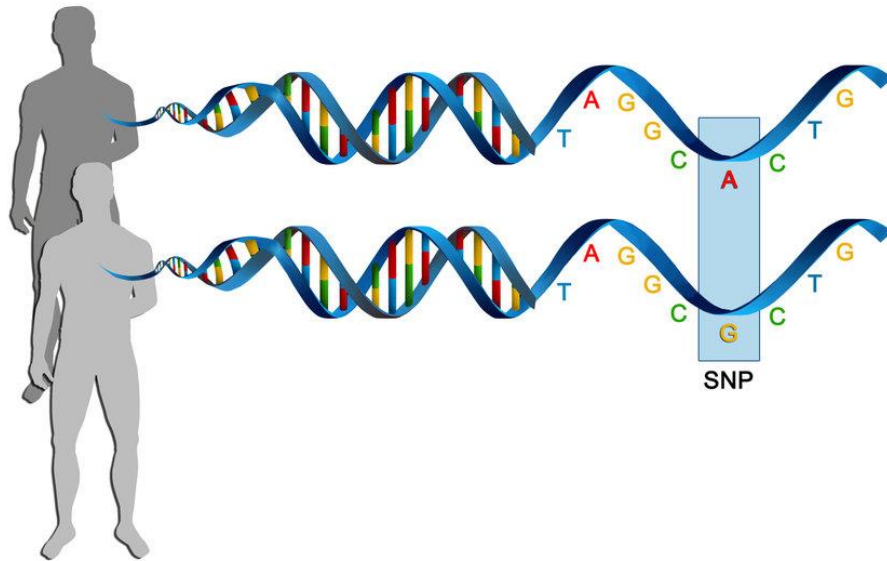
What makes SNPs so useful in genetics?

- Found in the entire genome, heritable, and often neutral

- Abundance, stability, and ability to act as **genetic markers**

- They can be unique or very common, and can be used to study disease, ancestry, or even drug response
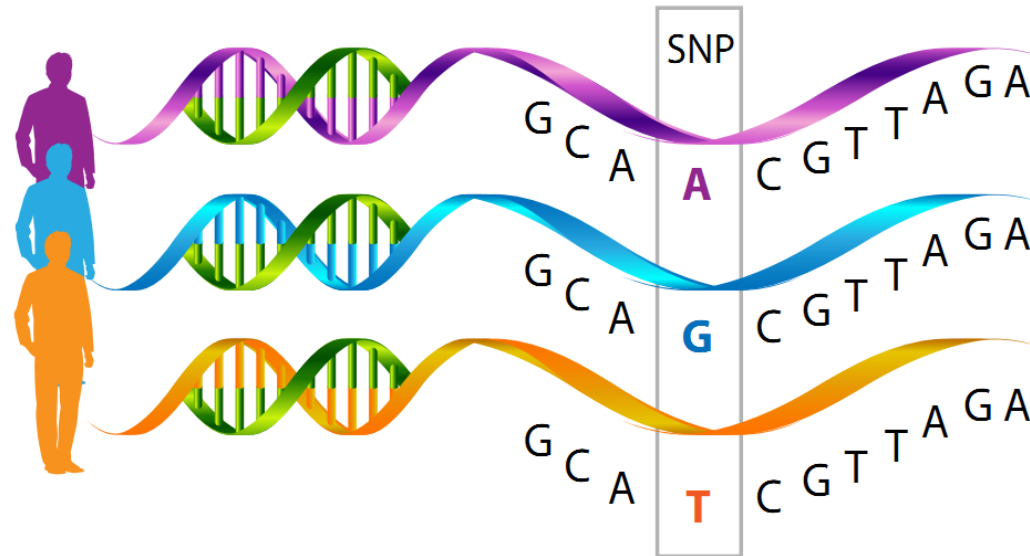
From ger

- Using SNPs ... nes at the same time, ... gene (not all!)

- A **genome** ... n. It consists of all the se ...

- Genes help ... nderstand evolution. T ...
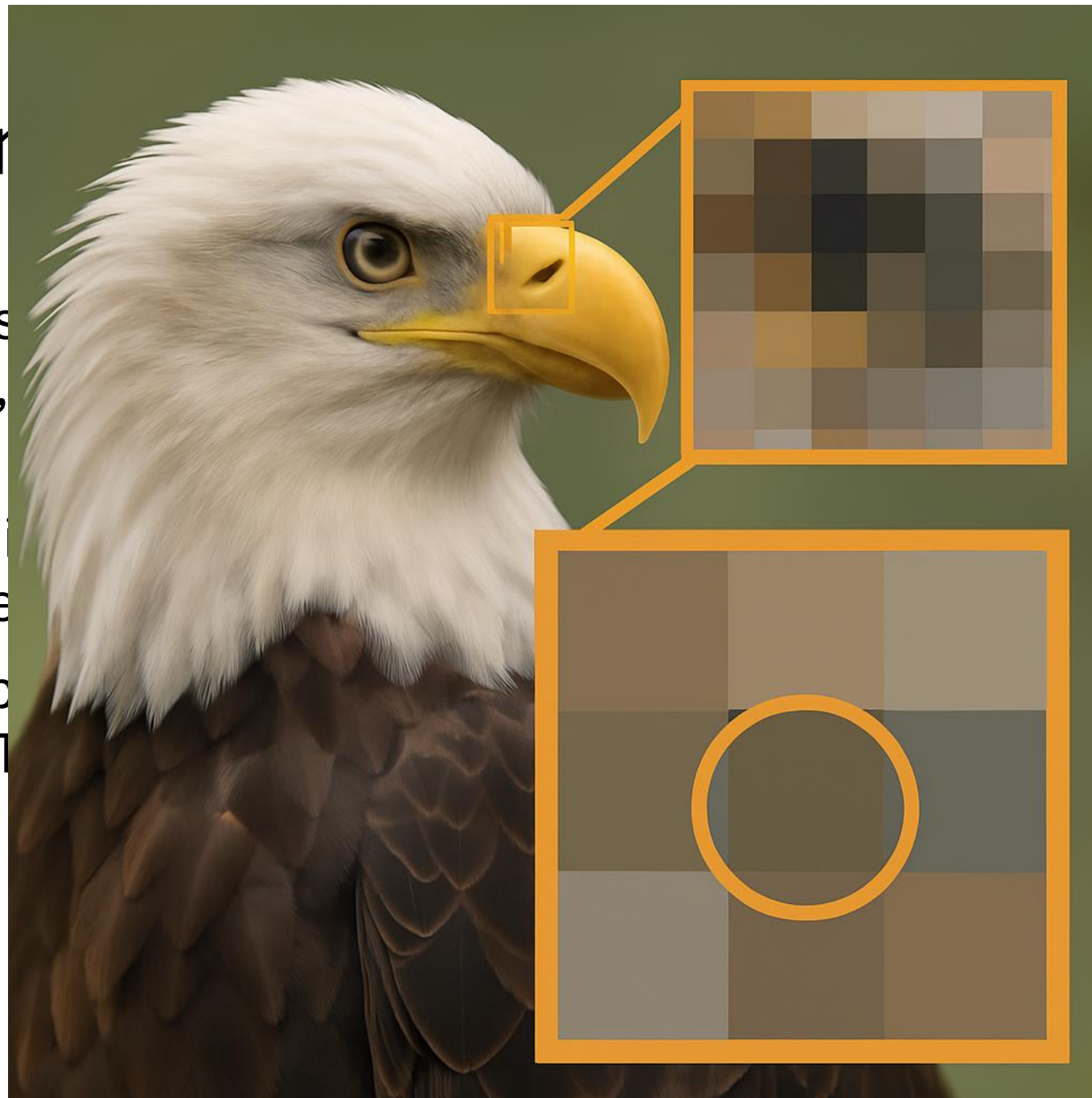
How do you compare two populations?

- To <u>compare</u> two populations, researchers use allele frequencies and genetic distance

- **Genetic distance** is a measure of the genetic differences between two populations

- If two populations have very different **allele frequencies**, it suggests they have been **evolving separately**

- One way to measure this difference is $F_{ST} \rightarrow$ tells us how much genetic variation is shared or is unique between populations

f-statistics and F_{ST}
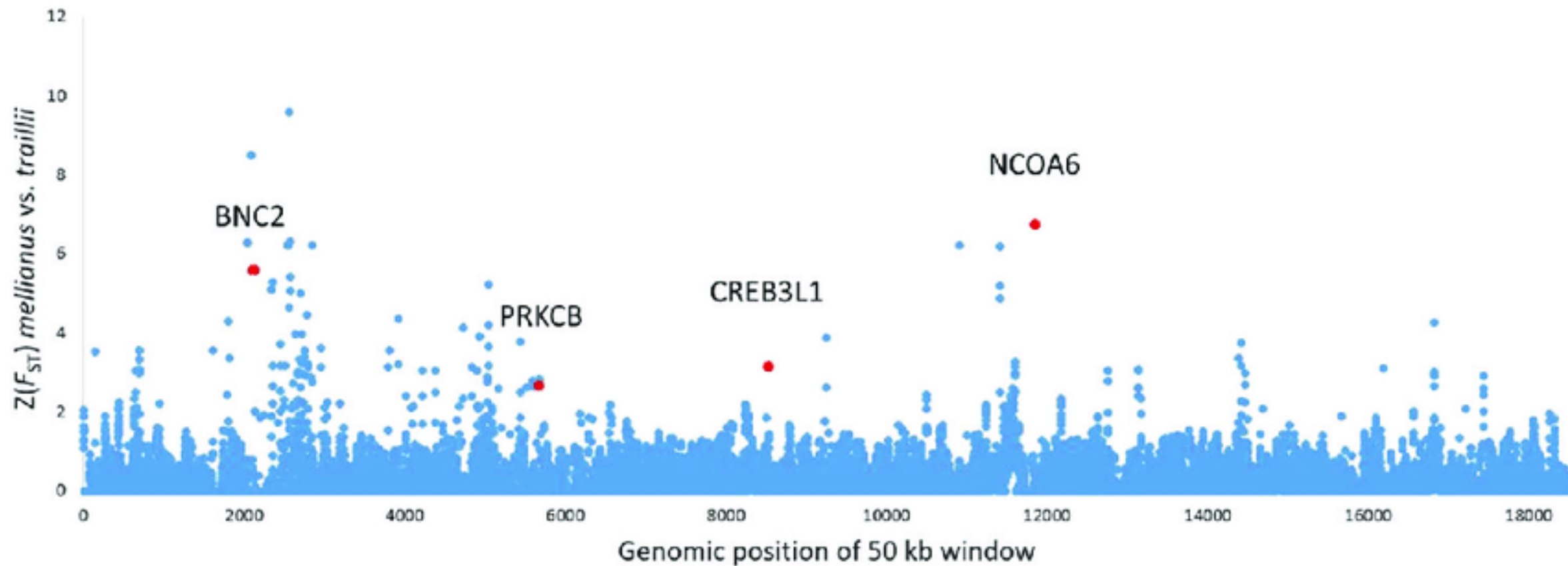
How do you calculate F_{ST}?

- A **low F_{ST}** (close to zero) means populations are genetically **similar** (they're sharing most of their gene variants)

- **A high F_{ST}** (close to one) means populations are more **different** (they've likely been evolving separately)

- To calculate it, we count how common different alleles are in each population, using SNPs as input!

- F_{ST} is a number that tells us how different two populations are at the genetic level, based on how alleles are distributed

Other important pop-gen statistics

- **π (pi)** measures **genetic diversity** within a population: it tells us how different individuals are from each other at the DNA level

- A **higher π** means more variation within the population, which can reflect larger population sizes or past gene flow

- **D_{xy}** measures the **absolute genetic difference between two populations**: how many base pairs are different on average

- While F_{ST} is a **relative** measure, D_{xy} is **absolute**, and helps identify long-term separation or regions with strong divergence
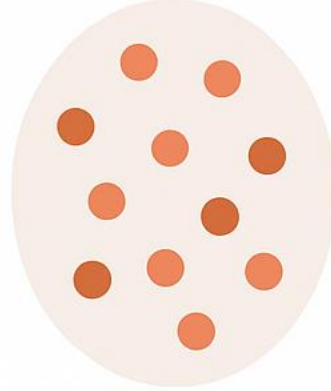
F_{ST}

π

D_{xy}
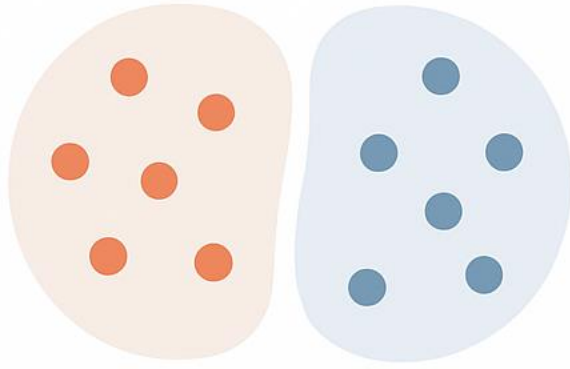
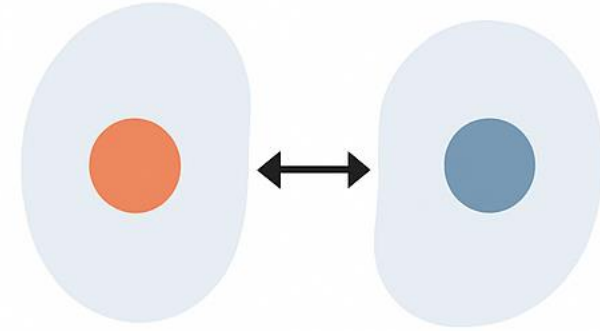**relatedness**

Measures how genetically similar two populations are

genetic diversity

Measures variation within a population

absolute difference

Measures how many genetic changes separate two populations

Let's do it ourselves!

- Two *Daphnia* populations: one is adapted to predation by fish, the other isn't

- What are the genetics behind this?

- Can we mathematically calculate the genetic difference between these two populations? Yes, we can!

- Looking through the entire genome will take a long time! Instead, we will zoom in to **one** of the ten **chromosomes**

- We will work using **R** and **RStudio**

R

- R is a **programming language** made for working with data

- Researchers use R to analyze, visualize, and understand data, especially in biology

- It's great for making graphs, plots, and statistics from real-world data

- In this course, we'll use R to look at genetic data and explore patterns of evolution

Let's jump in!

- Follow this link: https://github.com/mariamadrid19/SoS-daphnia/wiki