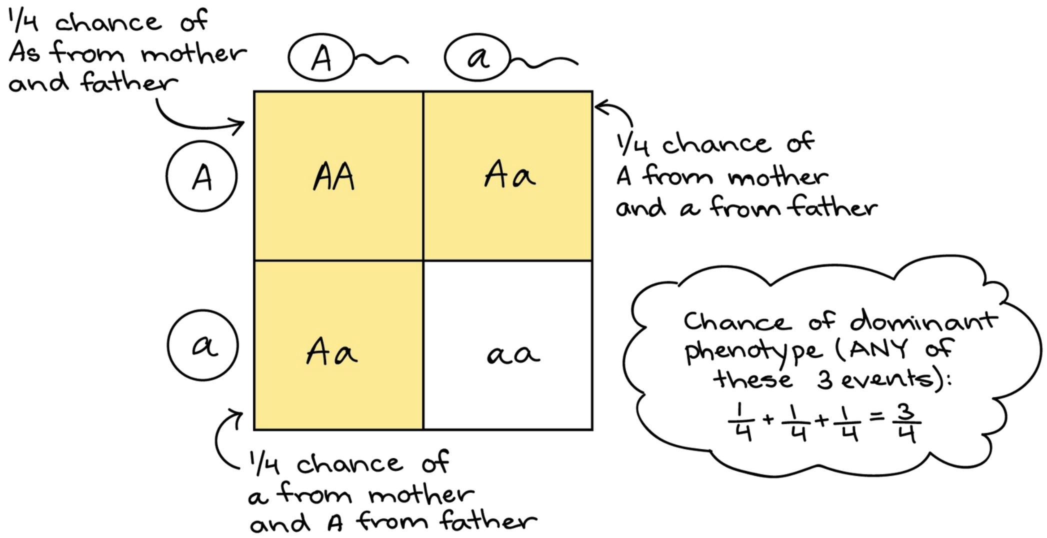


MSc in QUANTUM COMPUTING AND QUANTUM TECHNOLOGIES

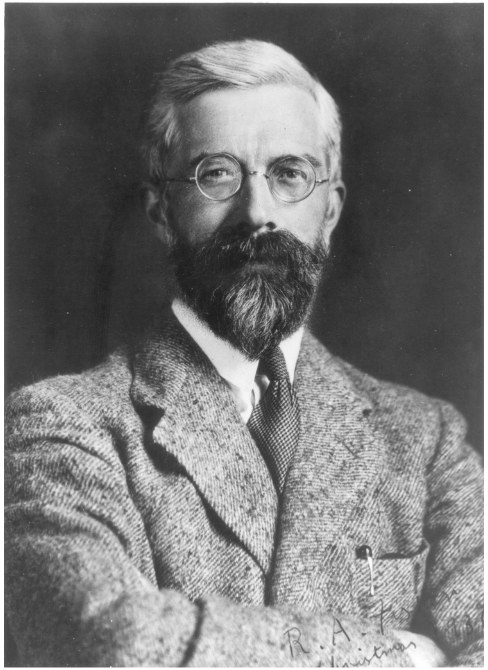
Computational Biology #3

Population Genetics & Evolution

Aristotelis C. Papageorgiou
Molecular Biology & Genetics, DUTH
apapage@mbg.duth.gr

Genetics is a computational science

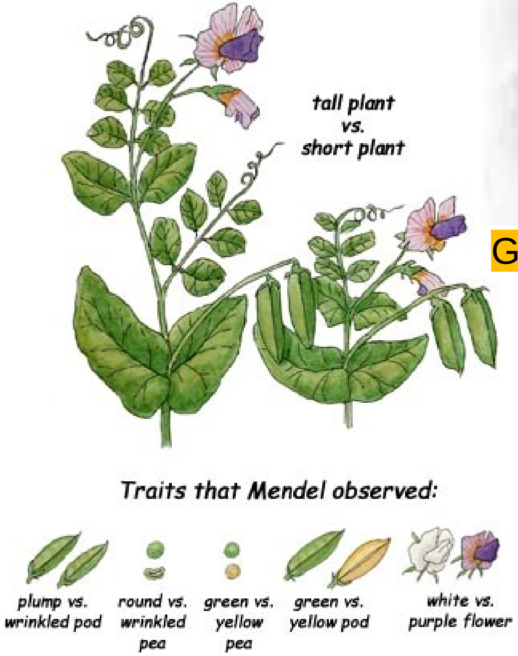
- Algebra describes the main genetic processes
- Linear algebra is used in form of **mixed models** (animal and plant breeding)
- First statisticians were geneticists (eg R.A. Fisher)
- The development of information science, computer technology and the internet have led to **bioinformatics**



R.A. Fisher (1890-1962)

Gregor Mendel

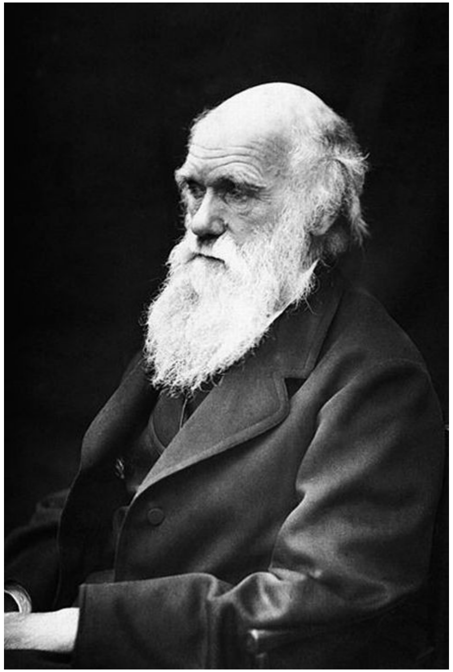
- The first biological experiment
- From his results he deducted to a theory of inheritance
- He discovered the dual nature of inheritance
- He did not know any statistics and he messed his data
- He was right!




Gregor Mendel (1822-1884)

Charles Darwin

- He described the mechanism of natural selection in evolution
- He founded a dynamic evolutionary theory
- He failed to explain heredity
- He did not use any math, or statistics or genetics
- He was (mostly) right



Charles Darwin (1809–1882)

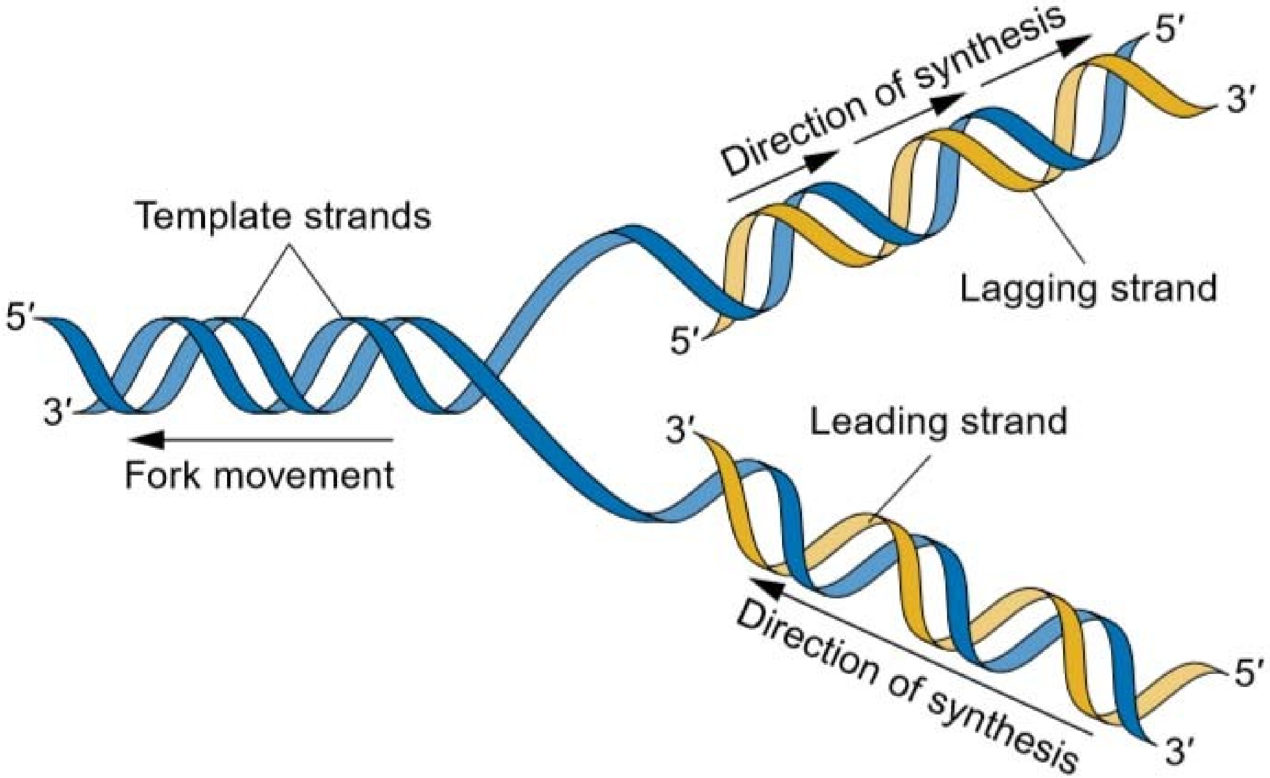
Genetic information

- Genetic information: All things passed from one generation to the next
- Genetic diversity: Differences between organisms in their genetic information
  - Within species
  - Among species
- Genetic **diversity** is widespread in nature
  - Derives from mistakes during the transfer of genetic information (mutations)
  - Is the main prerequisite for evolution to happen
- **Evolution**: Change in genetic information in a population from generation to generation

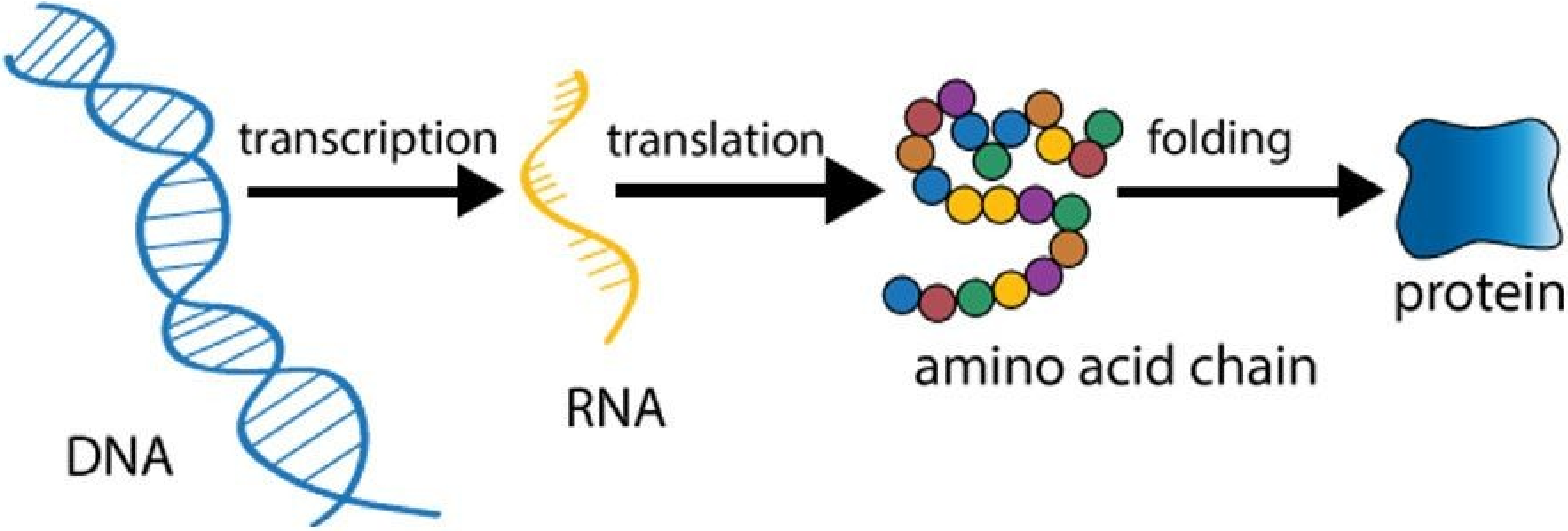
Gene and allele

- Every higher organism has a gene in **two copies**
    - Each copy comes from a different parent
- Gene: the unit of genetic information
    - A piece of DNA (fragment) containing the information (code) to make a protein
        - The protein catalyzes the reaction that defines a trait
    - The code is linear with just 4 letters (bases – nucleotides)
- DNA – replication
- DNA/RNA/protein

DNA is copied

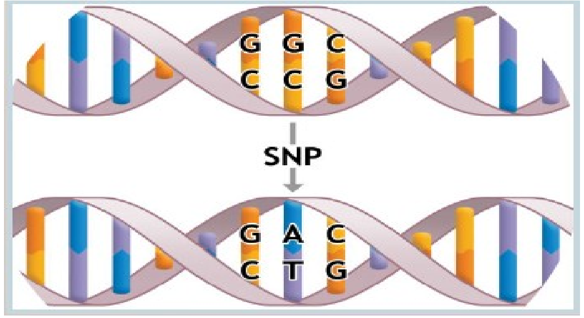


DNA is expressed

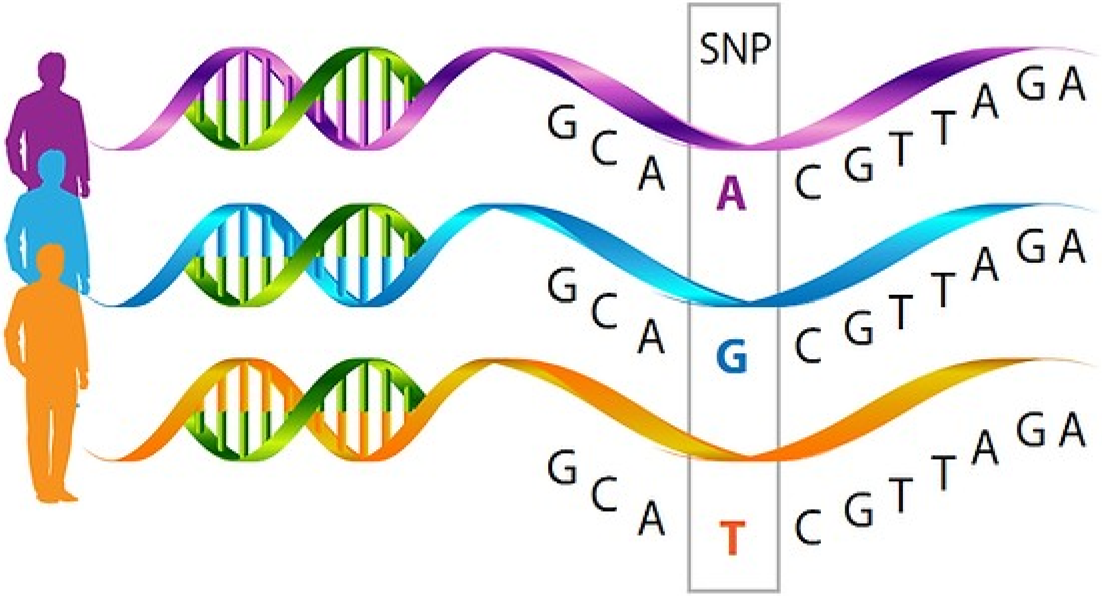


Gene and allele

- **Allele**: the alternative form of a gene
  - It does the same job, but in a different way
- Example: a gene defines fruit color in sweet pea
  - The A allele results in yellow color
  - The a allele results in green color
- The DNA sequences of the two alleles are almost the same
  - They differ in one or more bases (polymorphisms)
  - Polymorphisms come from mutations



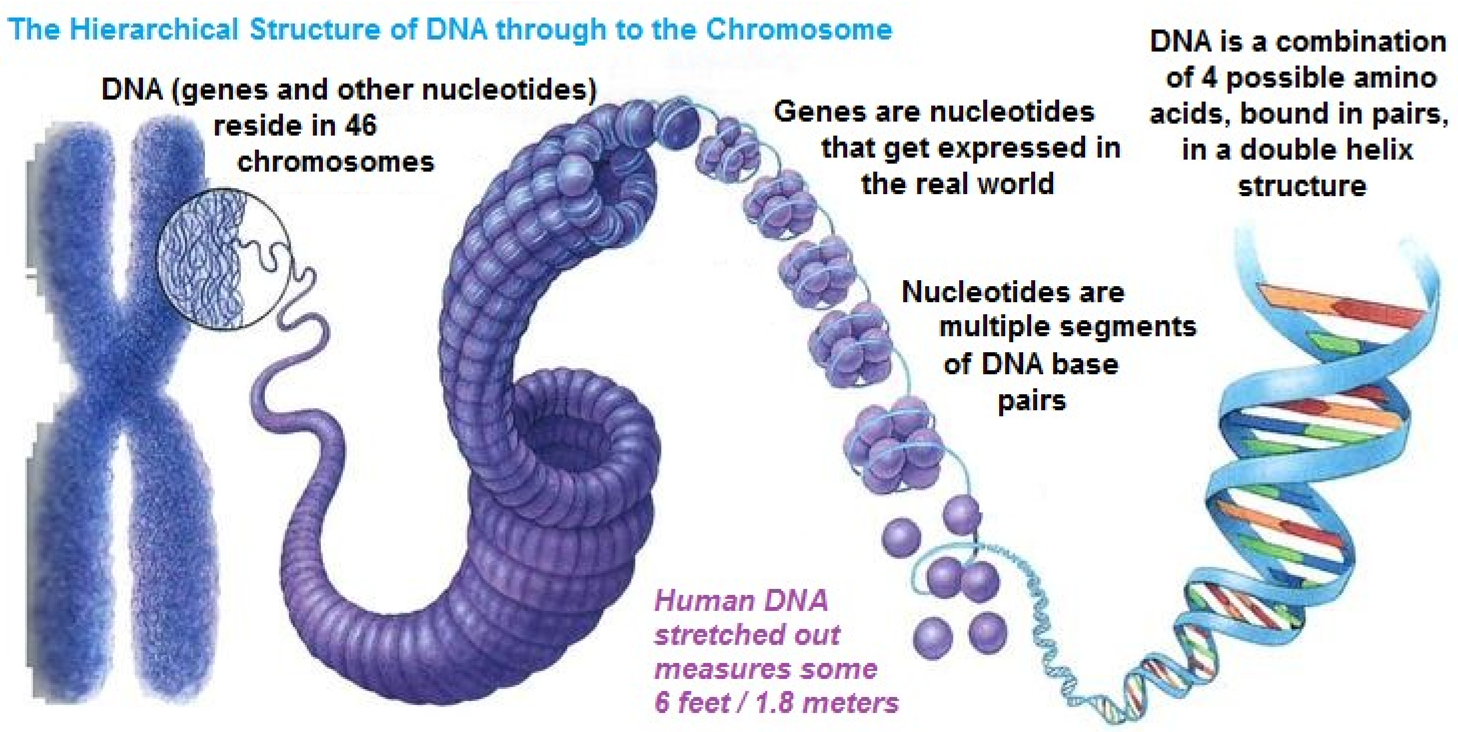
Single Nucleotide Polymorphism



Chromosomes

- Structures in the cell nucleus, found in pairs in each somatic cell of a diploid organism
- They consist of DNA and proteins and carry the genetic information during cell divisions
- They have the ability to divide and reproduce (mitosis)
- Each chromosome within a pair has one allele at each gene locus (location of a gene on a chromosome)

Genes and chromosomes



The Hierarchical Structure of DNA through to the Chromosome

DNA (genes and other nucleotides) reside in 46 chromosomes

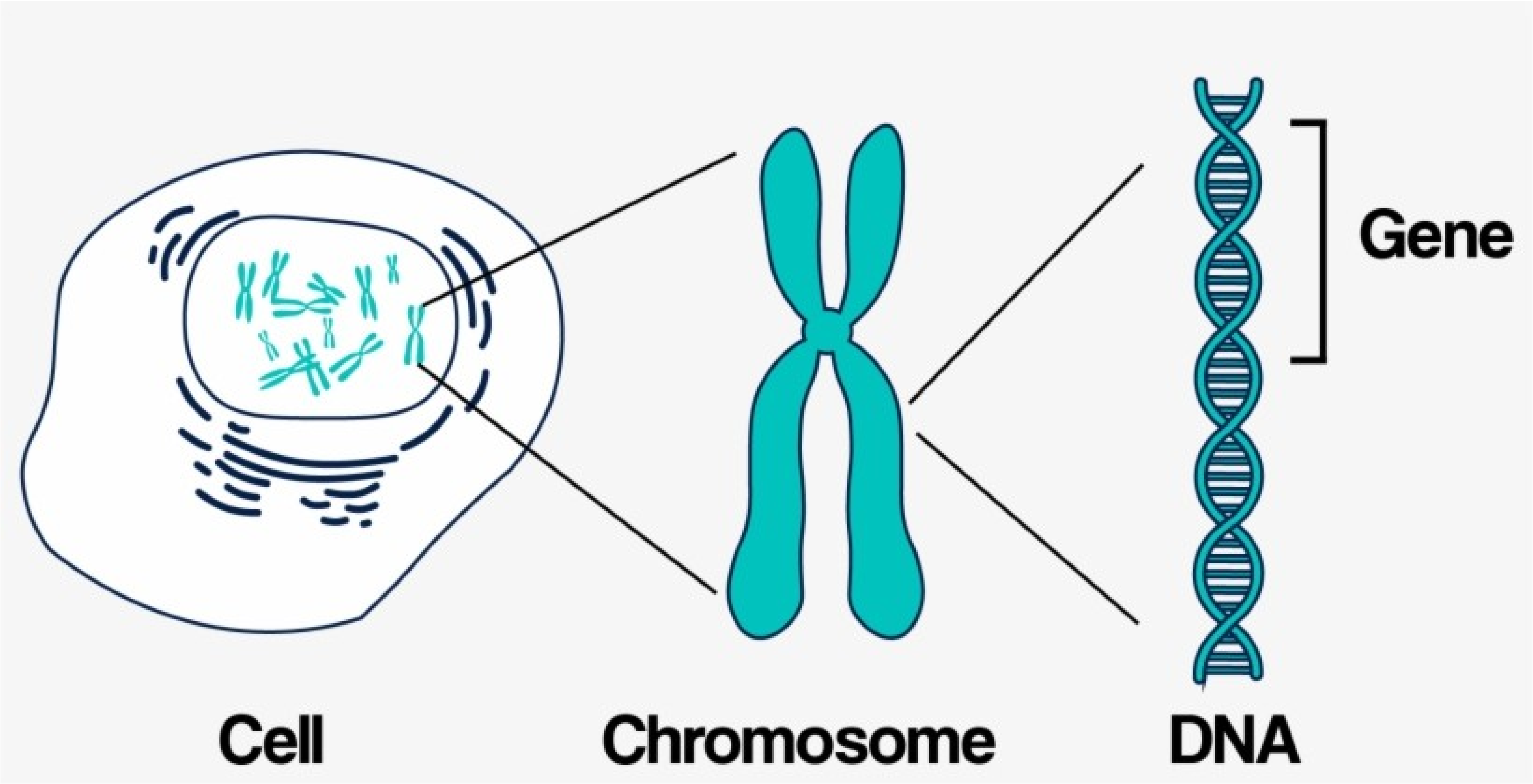
Genes are nucleotides that get expressed in the real world

DNA is a combination of 4 possible amino acids, bound in pairs, in a double helix structure

Nucleotides are multiple segments of DNA base pairs

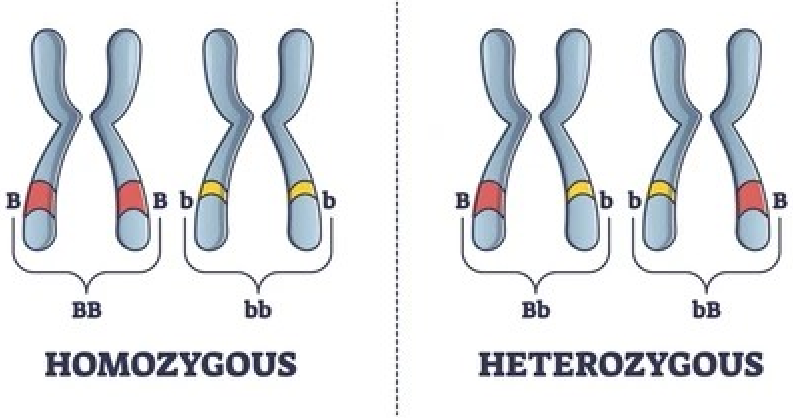
Human DNA stretched out measures some 6 feet / 1.8 meters

Genes and chromosomes



Heterozygous and homozygous

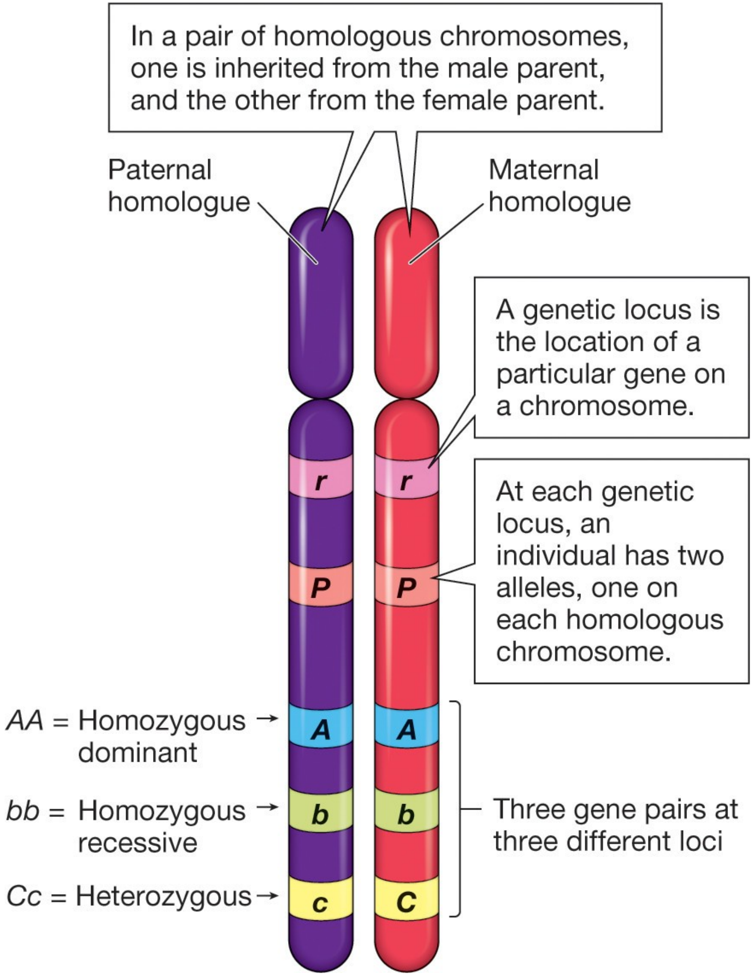
- A diploid individual that has two identical alleles at a gene locus is **homozygous**
- A diploid individual that has two different alleles at a gene locus is **heterozygous**



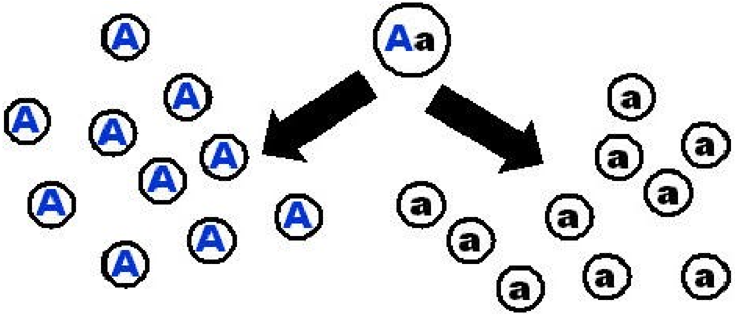
shutterstock.com · 2068666403

Homologous chromosomes

- Chromosomes come in pairs
- A diploid somatic cell contains an even number of chromosomes (23 pairs in humans = 46)
- In each homologous pair, one chromosome is of paternal and one of maternal origin

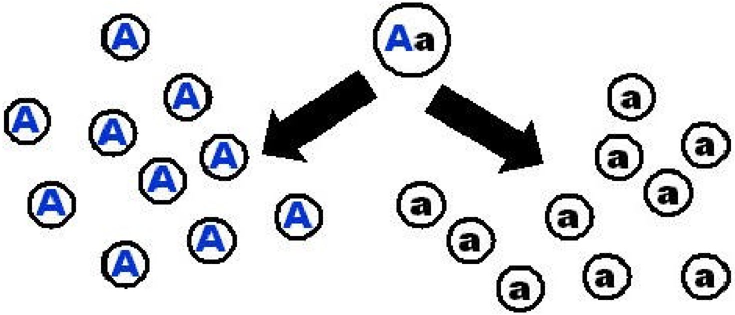


Gametes



- Homologous pairs of chromosomes divide and enter reproductive cells called **gametes**
- A gamete contains only one chromosome from each pair of homologous chromosomes
- From each parent, only one allele is inherited from each gene locus

1/2



- This is a completely random event: which chromosome from each pair of homologous chromosomes will be passed into a certain gamete
- From each parent, only one allele from each locus is inherited, with a probability of ½ for each gamete
- *Mendel's Law of Segregation*

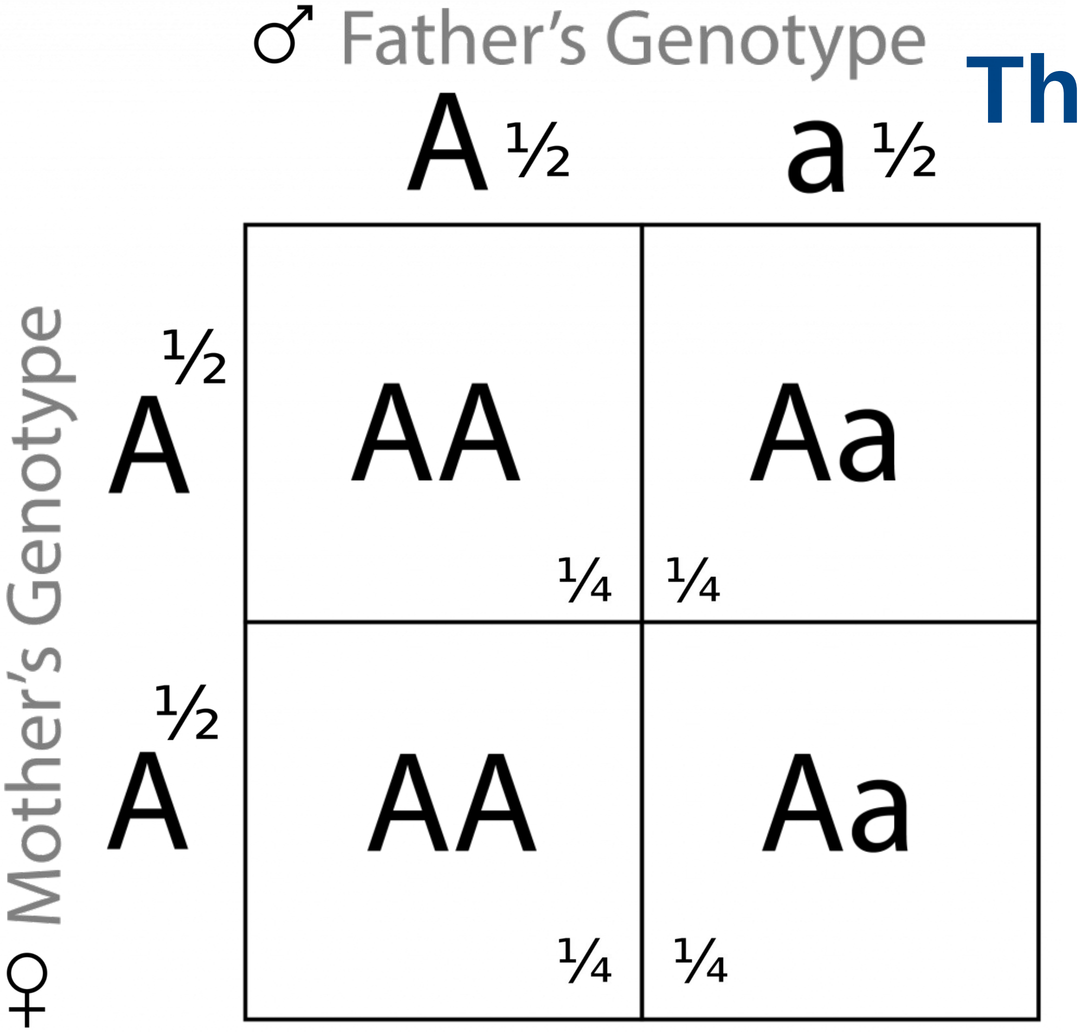
Reproduction and genetic information

- We consider the transfer of an allele to the next generation an event (with probability of occurrence ½)
- The laws of probability apply

A few more terms before we continue

- Dominant character: when one dominant allele is present in a genotype, then that character is expressed
    - Usually with a capital letter (A)
- Recessive character: expressed only in the absence of the dominant allele
    - Usually lowercase (a)
- **Genotype**: the combination of two alleles at the same gene locus (eg *Aa*)
- **Phenotype**: The product of expression of the genotype

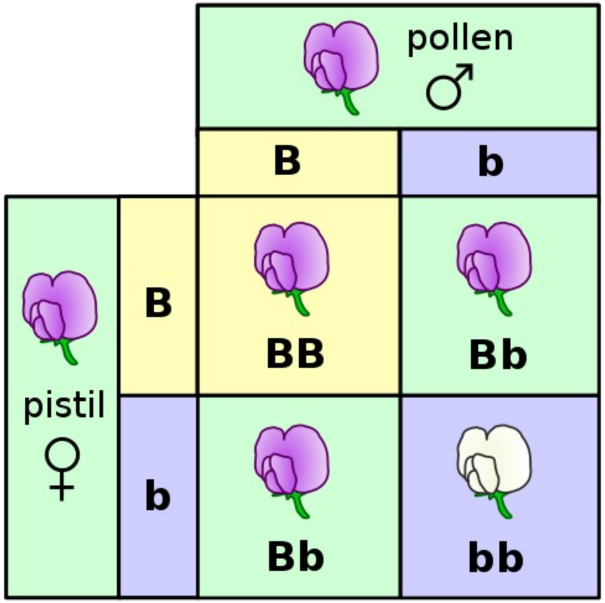
The Punnett square



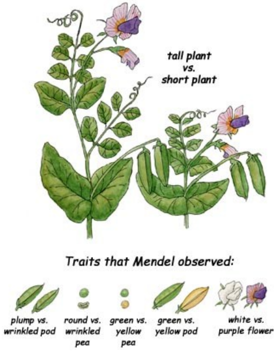
- Using the law of segregation and the laws of probability we model the reproduction and the inheritance of characters
- This is where all calculations and measurements of genetic information begin

Gregor Mendel

- He discovered the particle and binary nature of heredity
- The probability of carrying an allele to the next generation is ½
- Traits are inherited randomly
- Mendel's work remained unknown until 1900



Gregor Mendel (1822-1884)



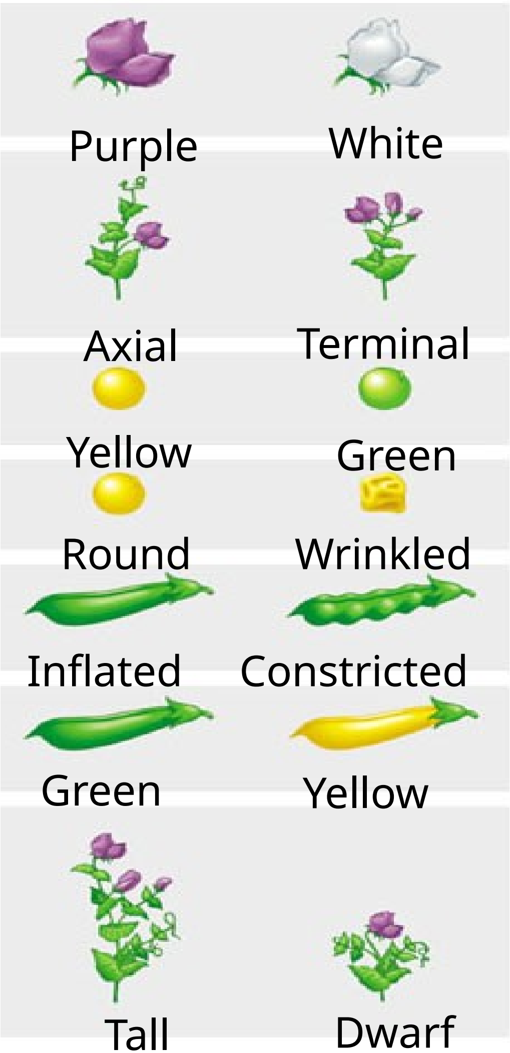
Who was Gregor Mendel?

- Austrian monk, born in 1822 to peasant parents
- Mendel was educated in a monastery and went on to study science and mathematics
- He returned to the monastery and spent the rest of his life there, eventually becoming abbot
- In the garden of the monastery, Mendel initiated a series of experiments on plant hybridization
- The results of these experiments would ultimately change our views of heredity irrevocably

Gregor Mendel (1822-1884)

Brno / Brünn

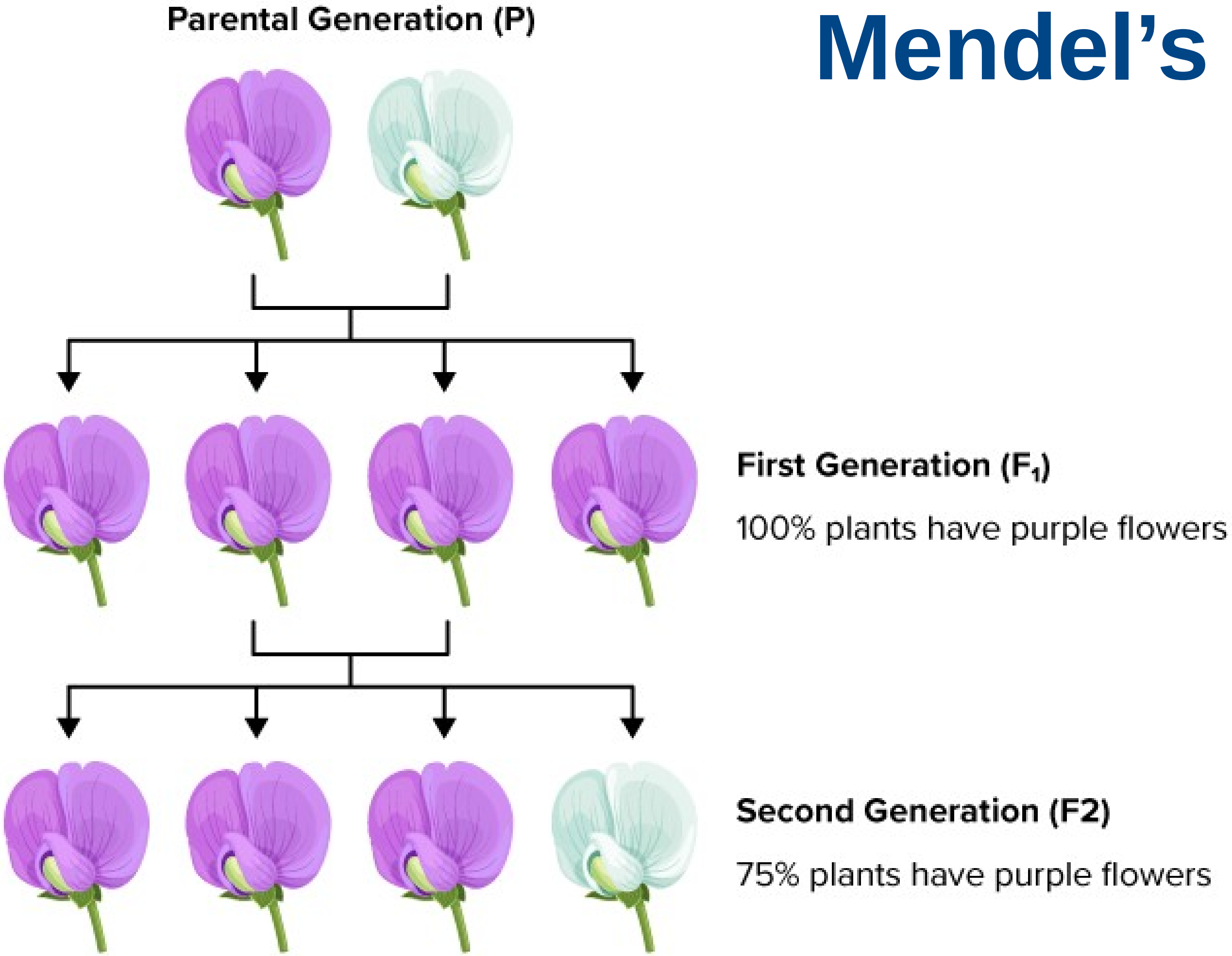
Mendel studied 7 characteristics / traits



Flower color	Purple	White
Flower position	Axial	Terminal
Seed color	Yellow	Green
Seed shape	Round	Wrinkled
Pod shape	Inflated	Constricted
Pod color	Green	Yellow
Stem length	Tall	Dwarf

Pisum sativum

Mendel’s experiment



Parental Generation (P)

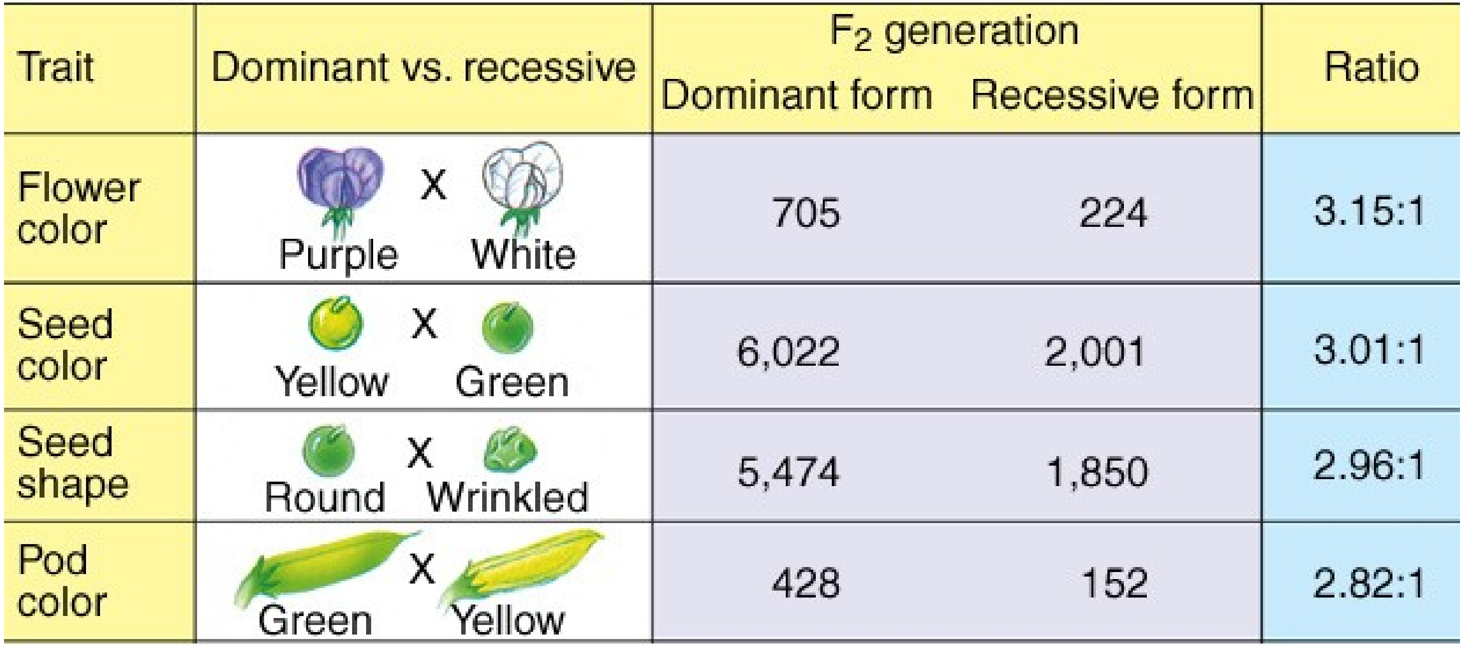
First Generation (F_1)

100% plants have purple flowers

Second Generation (F2)

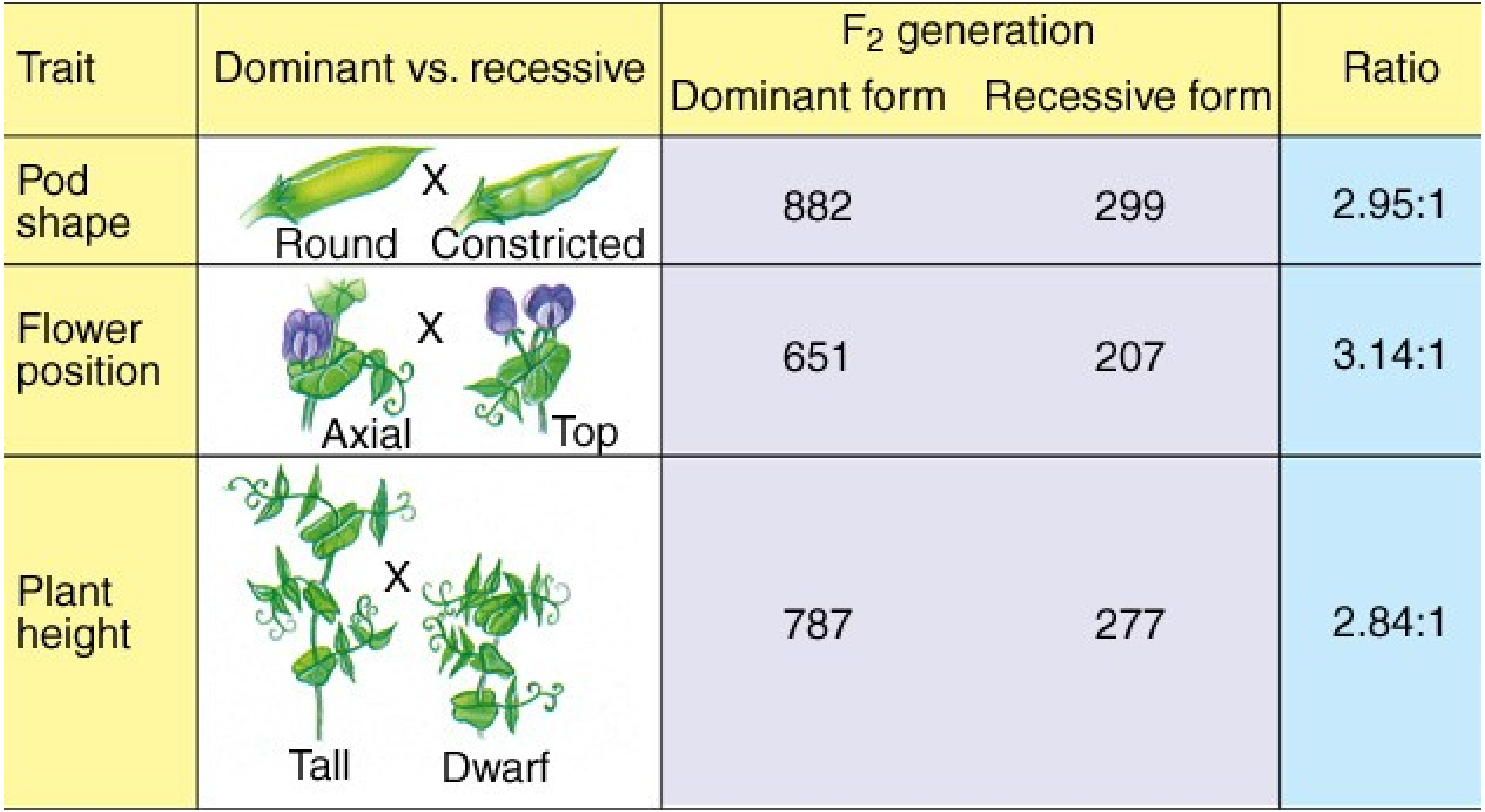
75% plants have purple flowers

Mendel’s experiment



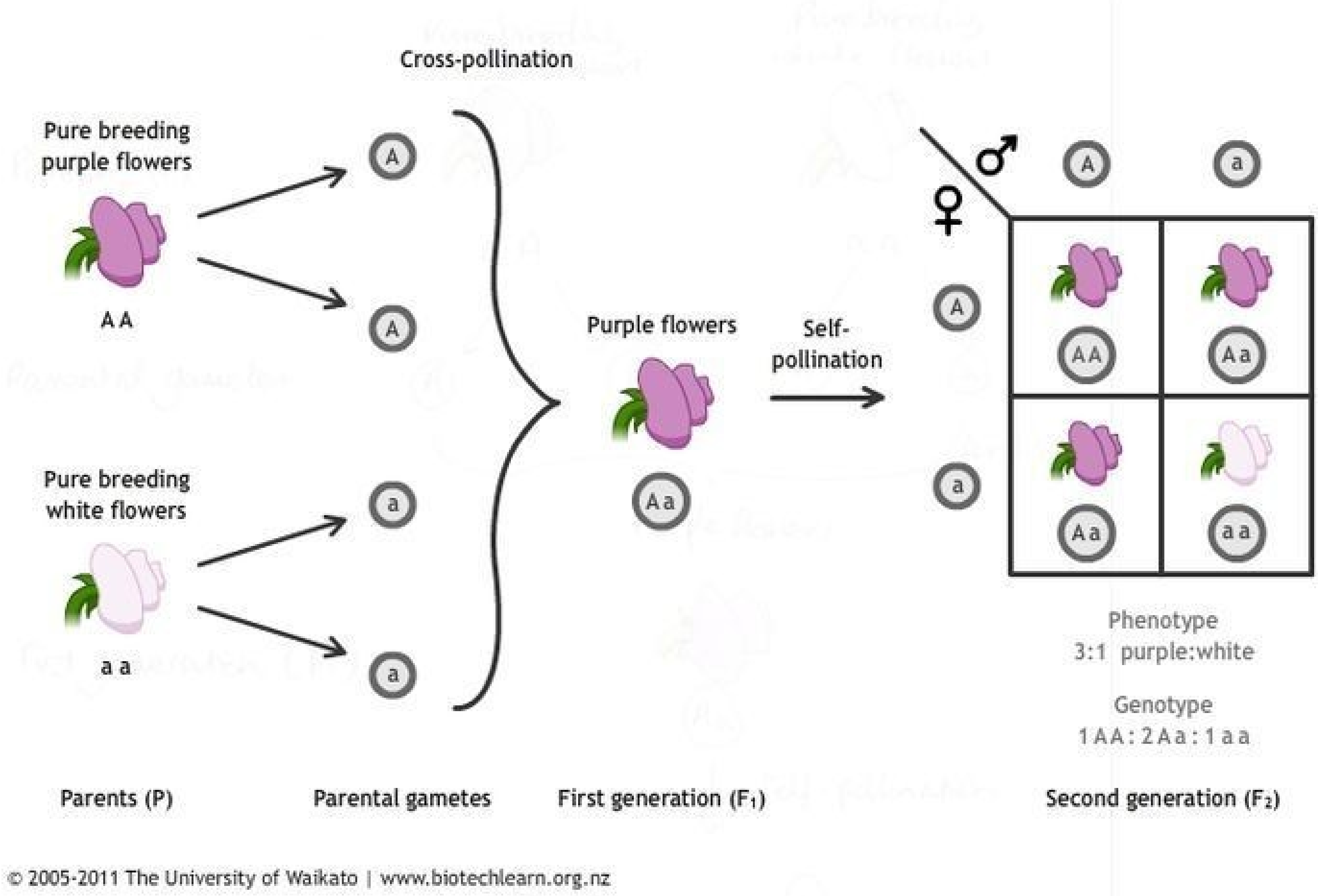
Trait	Dominant vs. recessive	F_2 generation		Ratio
		Dominant form	Recessive form	
Flower color	Purple X White	705	224	3.15:1
Seed color	Yellow X Green	6,022	2,001	3.01:1
Seed shape	Round X Wrinkled	5,474	1,850	2.96:1
Pod color	Green X Yellow	428	152	2.82:1

Mendel’s experiment



Trait	Dominant vs. recessive	F_2 generation		Ratio
		Dominant form	Recessive form	
Pod shape	Round X Constricted	882	299	2.95:1
Flower position	Axial X Top	651	207	3.14:1
Plant height	Tall X Dwarf	787	277	2.84:1

Mendel’s experiment



© 2005-2011 The University of Waikato | www.biotechlearn.org.nz

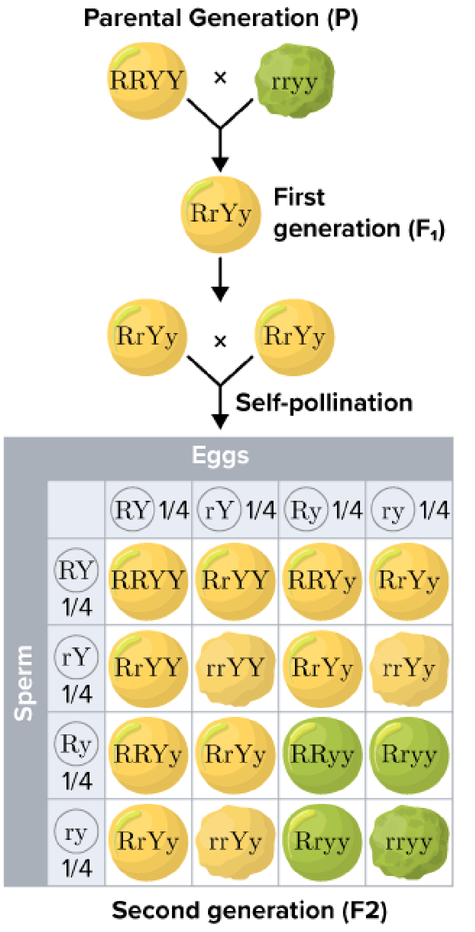
Mendel suggested that certain “elements” (genes) carry the information of a trait to the next generation

Every individual carries two copies (alleles) of a gene

Each allele is inherited with a probability of ½

Law of segregation

Mendel’s experiment



	RY 1/4	rY 1/4	Ry 1/4	ry 1/4
RY 1/4	RRYY	RrYY	RRYy	RrYy
rY 1/4	RrYY	rrYY	RrYy	rrYy
Ry 1/4	RRYy	RrYy	RRyy	Rryy
ry 1/4	RrYy	rrYy	Rryy	rryy

Mendel performed crosses observing two traits at the same time

The F2 generation had ratios, such as: 9:3:3:1

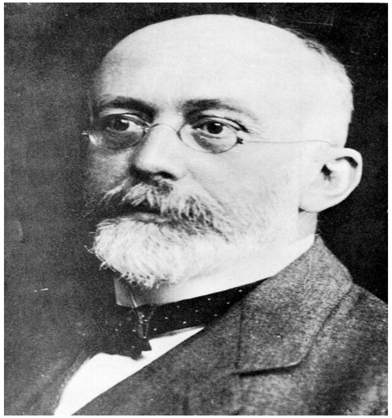
Genes are inherited independently and the probability of inheritance of each gamete with a certain combination of alleles from different genes equals the product of the individual probabilities of each allele

Genetic information is mixed independently in each gamete (at least for genes on different chromosomes): recombination

Law of independent assortment

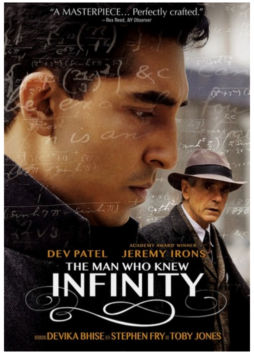
Hardy & Weinberg

- They applied Mendel's law of segregation to describe allele and genotype frequencies in **a population**
- They found that the frequencies of alleles and genotypes from generation to generation do not change
  - *Under certain conditions*



Wilhelm Weinberg
(1862–1937)

Godfrey Harold Hardy
(1877–1947)



The population

- A set of similar interreproducing organisms
    - A species
    - An area
    - Reproduction
- The population as a whole has a **gene pool** with a specific genetic structure

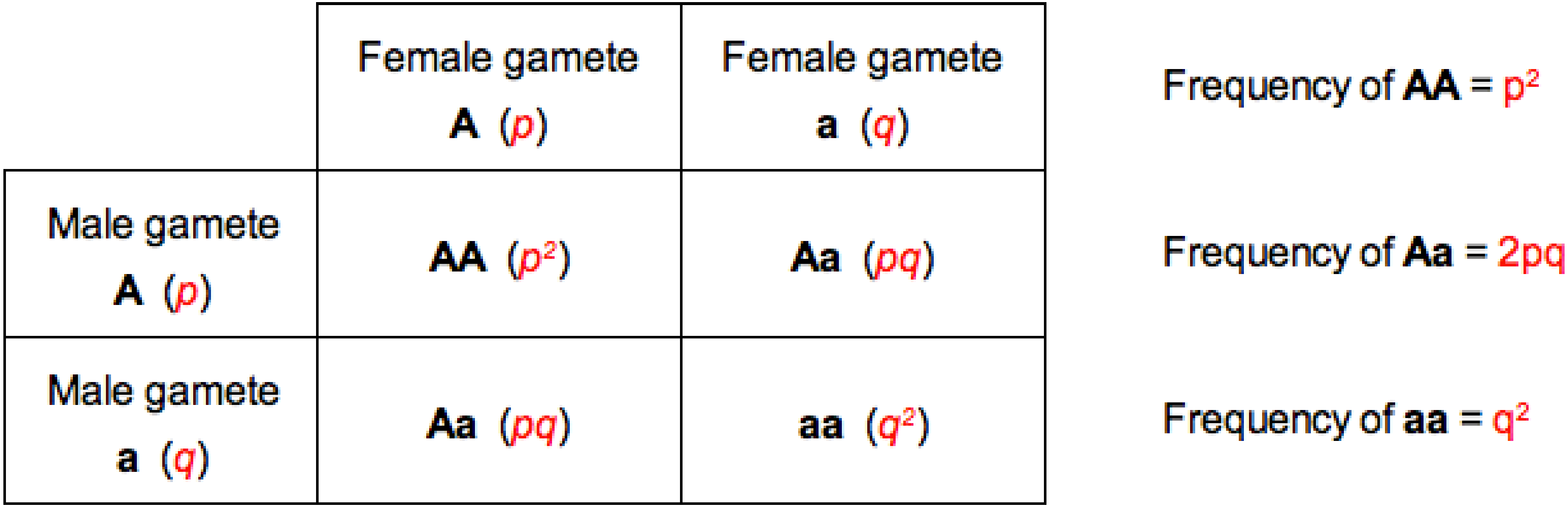
Genotype and allele frequencies

- From the results of genetic analyzes we get a record of the genotypes of a sample for a population, for a specific gene
- Let be a gene locus with two alleles, A and a
    - p: the frequency of the A allele
    - q: the frequency of allele a
- The three possible genotypes are AA, Aa and aa
    - P: the frequency of AA genotype
    - H: the frequency of genotype Aa
    - Q: the frequency of genotype aa

Mendelian population

- Is there a way to calculate the frequency of offspring genotypes from the frequencies of parental genotypes in a population?
    - Yes, if reproduction is **random**
- A population in which the matings between adult individuals are random is called a **Mendelian population**

Genotype frequencies in the next generation



	Female gamete **A** (p)	Female gamete **a** (q)
Male gamete **A** (p)	**AA** (p^2)	**Aa** (pq)
Male gamete **a** (q)	**Aa** (pq)	**aa** (q^2)

Frequency of **AA** = p^2

Frequency of **Aa** = $2pq$

Frequency of **aa** = q^2

$P=p^2$
$Q=q^2$
$H=2pq$

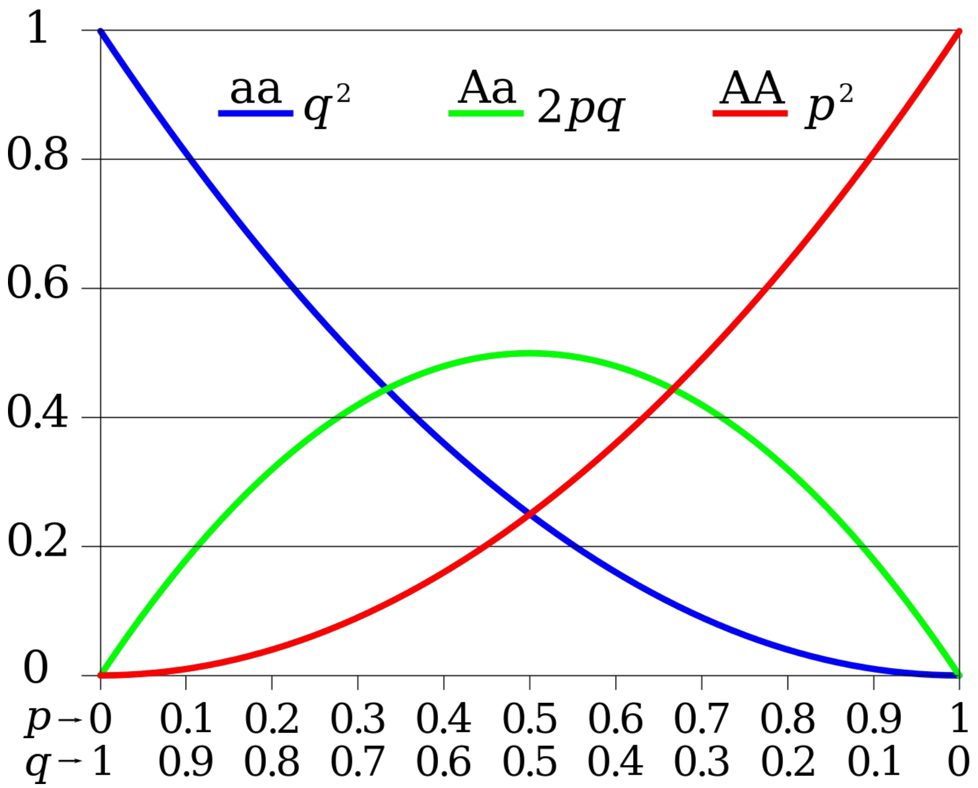
	Γύρη	
Ωάρια	A_1 p	A_2 q
A_1 p	A1A1 p^2	A1A2 pq
A_2 q	A1A2 pq	A2A2 q^2

HW equations

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

$$(p+q)^2 = 1$$

- Under random reproduction, alleles A and a with frequencies p and q are randomly combined in pairs



Example

- Let's calculate the frequencies of the genotypes in the next generation (P', Q', H')
  - p=0.6 q=0.4
- Then let's calculate the allele frequencies in the next generation (p', q')
- Are there changes in the allele frequences from one generation to the next?

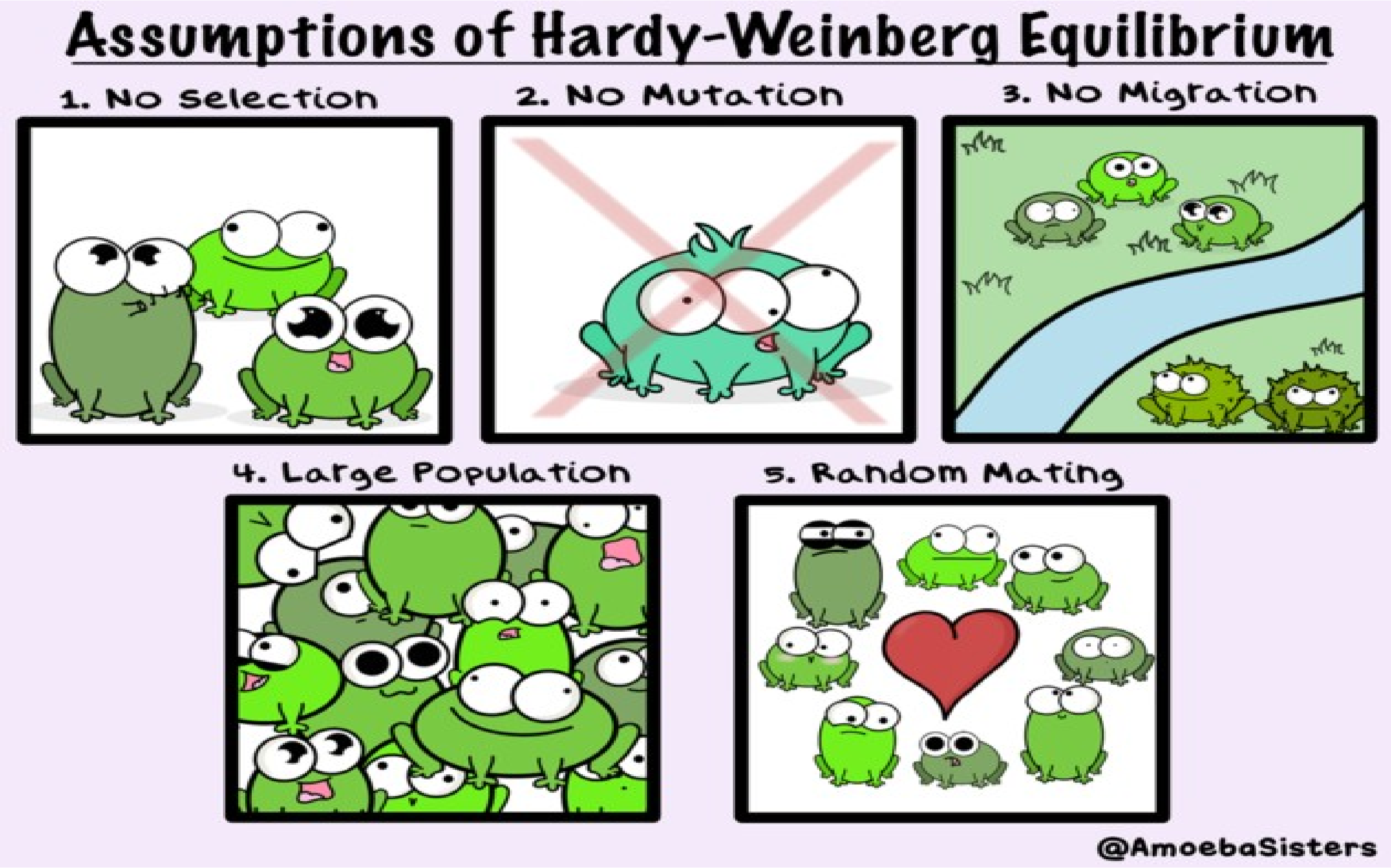
Hardy-Weinberg Law

- The most basic relation in population genetics
- When we have random reproduction, we can calculate the genotype frequencies of a population as a binomial function of the allele frequencies of the previous generation
- Under random reproduction, allele and genotype frequencies do not change from generation to generation
  - Populations do not change genetically…
- **Equilibrium population**

$P=p^2$

$Q=q^2$

$H=2pq$


Assumptions of Hardy-Weinberg Equilibrium
1. No Selection
2. No Mutation
3. No Migration
4. Large Population
5. Random Mating
@AmoebaSisters

Evolution

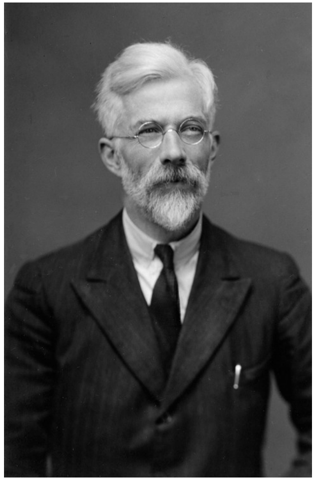
- When a population does not change genetically, then it is an equilibrium population and HW law applies
  - Such a population does not evolve
- **Evolution** applies to populations and is a dynamic process
- A population evolves when its allele frequencies change from generation to generation
- Evolution occurs when the prerequisites for an equilibrium populations are violated
  - Evolution forces are deviations from HW conditions

Evolutionary forces

- What will happen to a population if:
  - Some genotypes are more likely to have more offspring than others?
    - Natural selection
  - Population size is finite?
    - Genetic drift
  - Reproduction is not random?
    - Inbreeding and assortative mating
  - Mutations do occur?
    - New alleles
  - There is migration?
    - Gene flow

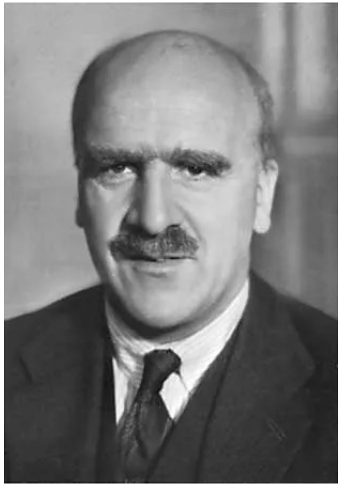
Population genetics

- Explaining heredity in a mathematical way
- Description of evolutionary forces through Mendel
- The basis of genetics and evolution



RA Fisher (1890-1962)

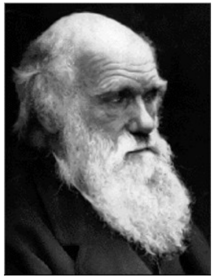
RA Write (1890-1962)



JBS Haldane (1892-1964)

But things are not that simple...

- Most traits are controlled by multiple genes that in many cases interact with each other
- Most features are also controlled by the environment
- **Quantitative characteristics**
    - **P = G + E**
- These characteristics follow the normal distribution and we study the mean and standard deviation of measurements
- **Quantitative genetics**



How selection works

We have selection when individuals with different genotypes living under specific environmental conditions systematically produce different numbers of offspring

- This is the statistical measure of the difference in the ability to survive or reproduce, between individuals in a population who differ in some critical trait
    - When the trait is hereditary, then there is **selection**
- This statistical measure is called **fitness** (or adaptive value) of the genotype
- Changing gene frequencies through selection is a systematic, and often slow, process, and the direction as well as the amount of change can be predicted
- Selection leads a population to **adaptation** (increase in mean fitness)

Basic selection model

- We consider a theoretical population, where all HW conditions apply except for differences in fitness of genotypes
    - w11, w22 and w12
- Values between 1 and 0
- We calculate the allele frequencies in the next generation

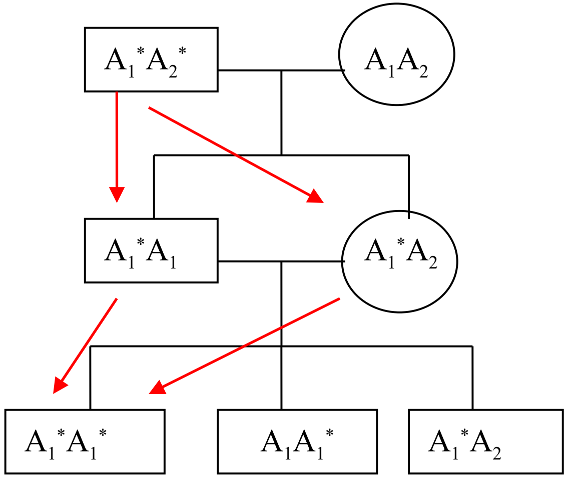
$$q_1 = \frac{1}{2}\left(\frac{2p_0q_0w_{12}}{\overline{w}}\right) + \frac{q_0^2w_{22}}{\overline{w}}$$

$$q_1 = \frac{p_0q_0w_{12} + q_0^2w_{22}}{\overline{w}}$$

Peppered moths on birch trees

Inbreeding: when two alleles are identical by descent

- An **inbreeding coefficient** *f* is the probability that two alleles found in the same genotype are identical by descent
  - They are both copies of an allele from a common ancestor
- **Inbreeding** is a deviation from random mating
  - When the probability of mating among relatives is higher than random

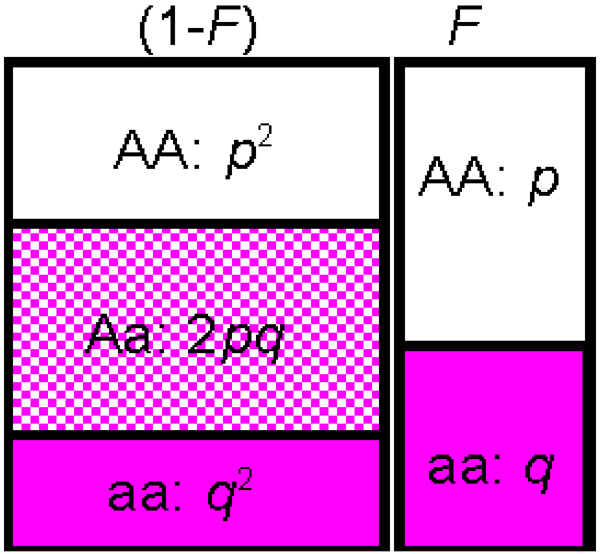


Inbreeding influences genotype frequencies

$$P = p^2 + fpq$$

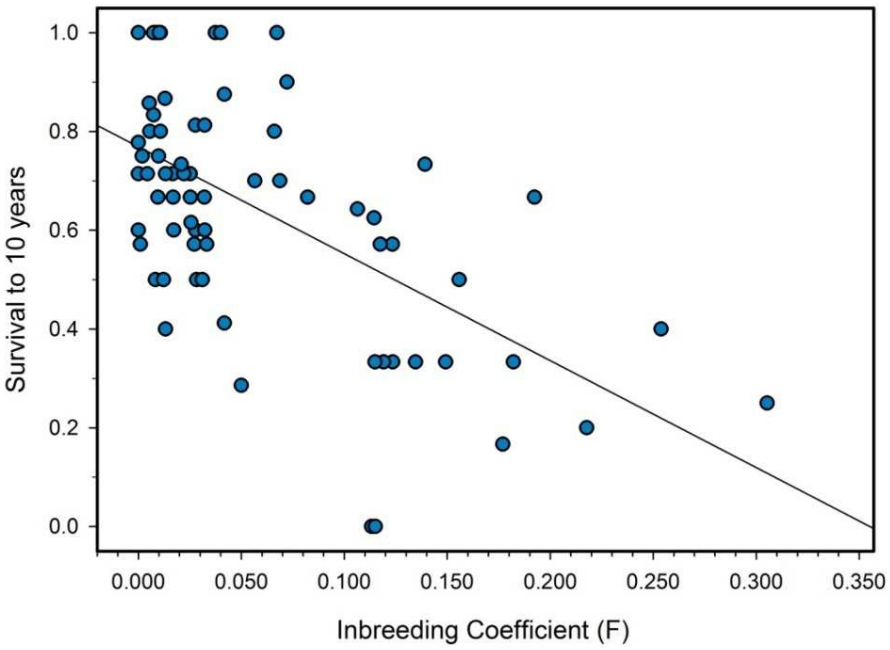
$$H = 2pq - 2fpq$$

$$Q = q^2 + fpq$$



Inbreeding, heterozygosity and inbreeding depression

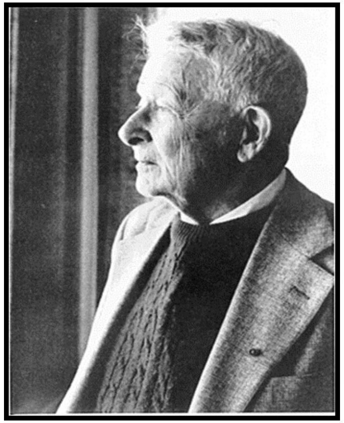
- Inbreeding increases the frequency of homozygous genotypes and decreases that of heterozygous genotypes, in a population
- For this reason, in inbred populations, there is a greater chance of hereditary diseases
- Allele frequencies do not change



Survival and inbreeding coefficient (*F*) of offspring from 71 marriages from the Habsburg royal dynasty

Genetic drift

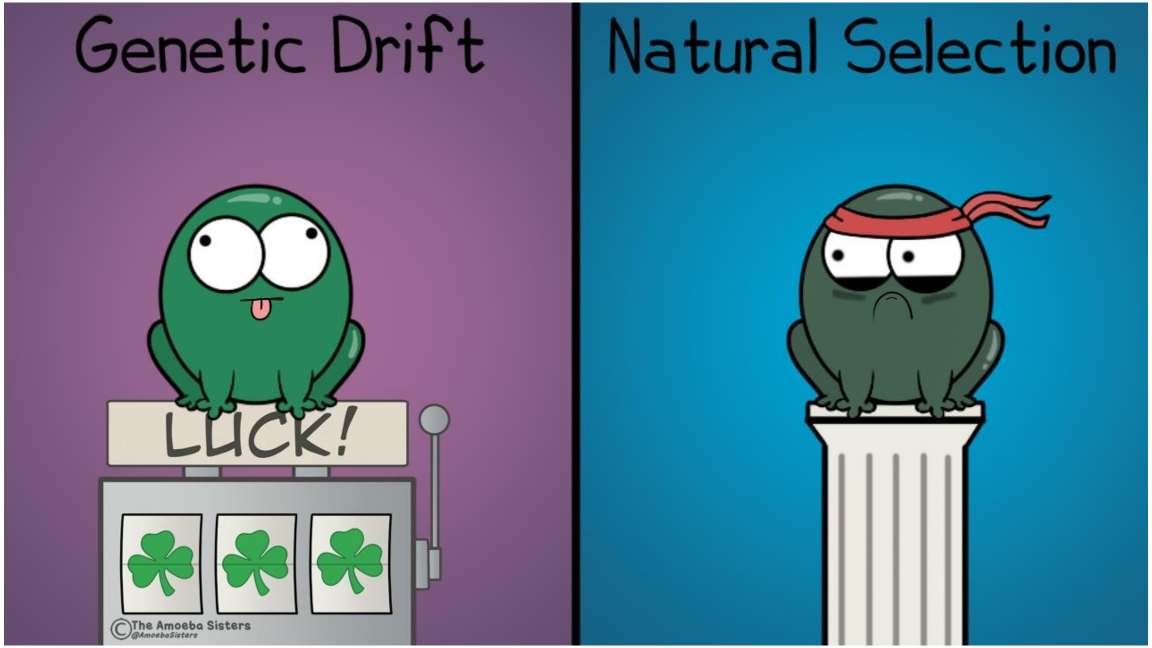
- Random variations in allele frequencies due to sampling in finite populations, from generation to generation
  - It happens because the gametes that eventually participate in reproduction are a subset (sample) of the total produced



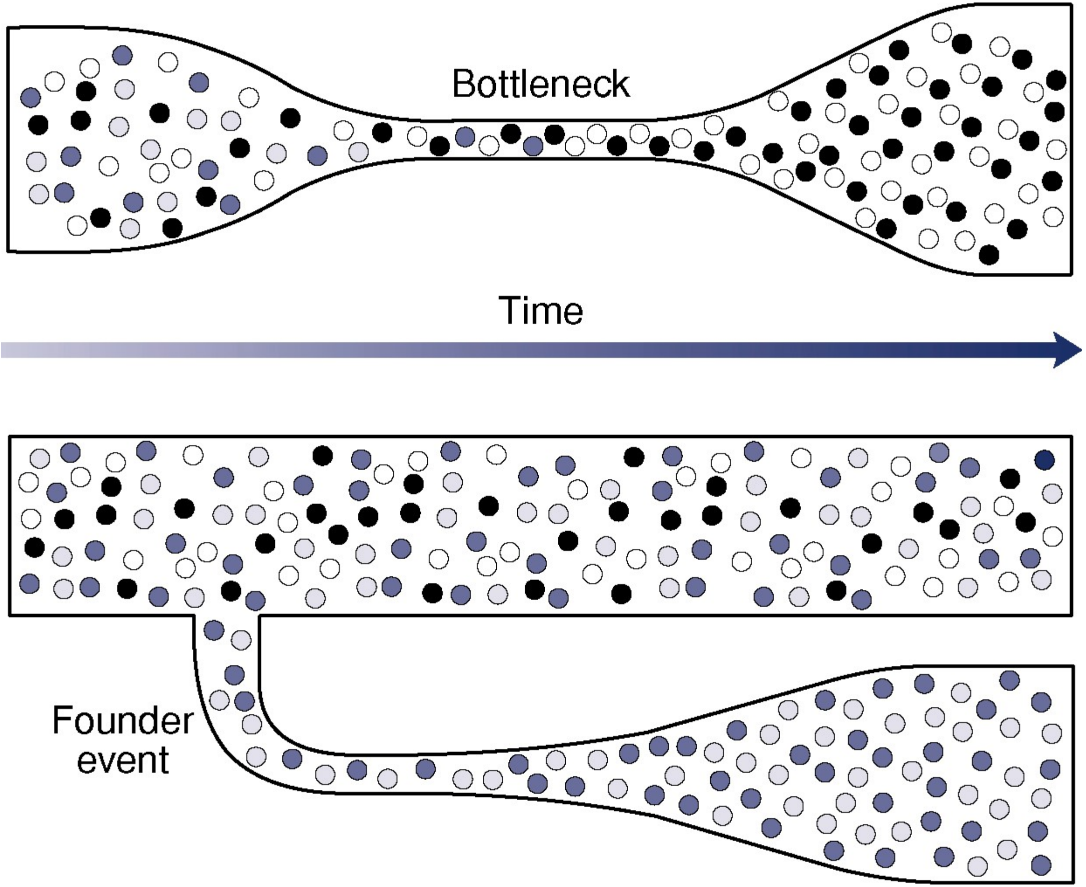
Sewall Wright

One allele wins (and the rest are gone)

- The smaller the population the stronger the effect of genetic drift on allele frequencies
- In a population, the initial frequency of an allele is significant
  - A rare allele is more likely to be lost
  - A frequent allele is more likely to be fixed



Genetic bottleneck and founder effect



A population may derive from a small sample of another larger population

Because the older population was destroyed and a small number of survivors remained (**bottleneck**)

Because a small subgroup of the population migrated away and established a colony (**founder effect**)

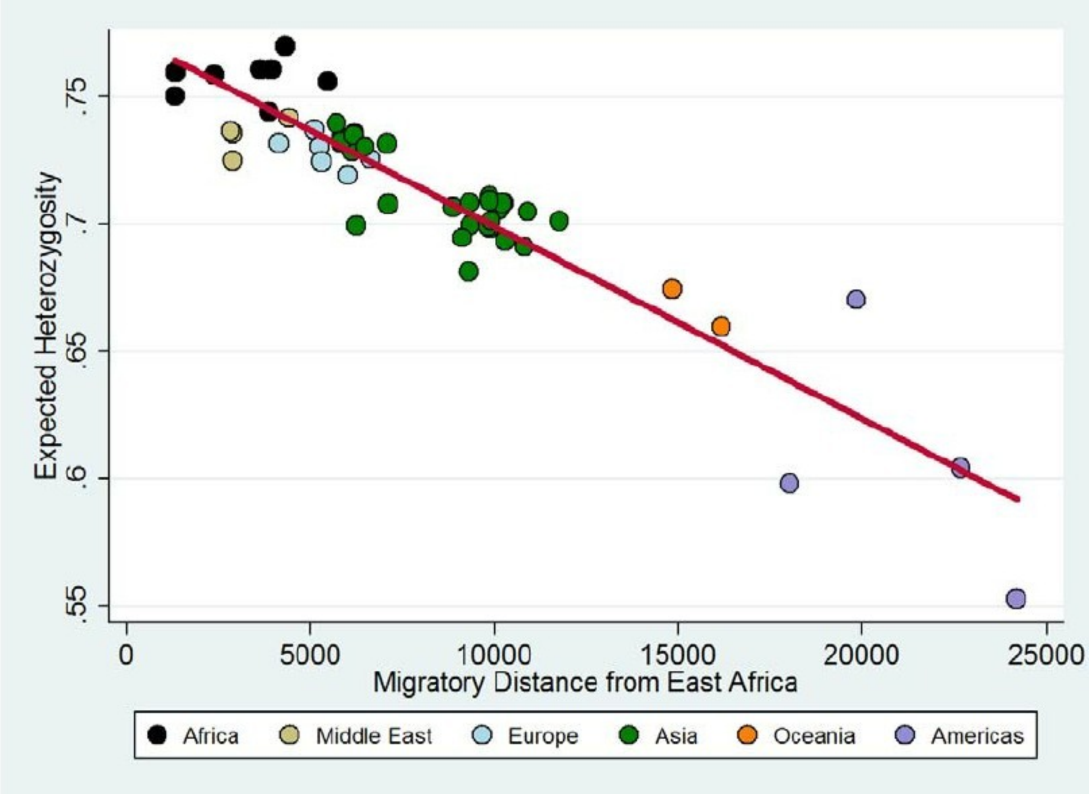
Such populations demonstrate low genetic diversity

The case of the elephant seal

- The elephant seal (*Mirounga angustirostris*) today reaches 160,000 animals on the west coast of America
- In 1890 there were only 1000 people on an island in Mexico
- Genetic research has shown very little diversity in the species

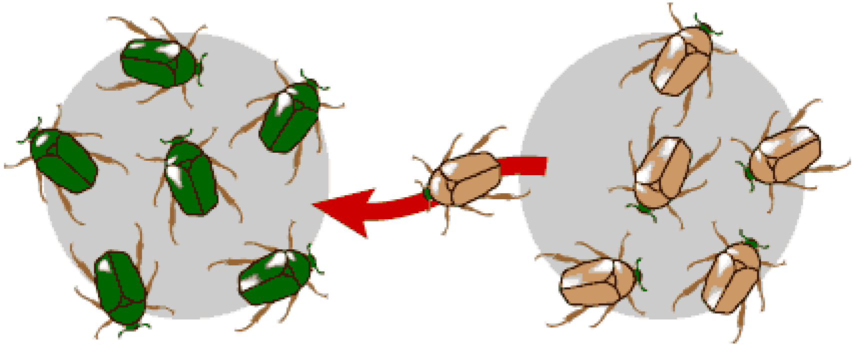
The confirmation of “out of Africa”

- The “out of Africa” scenario suggests that humans migrated out of Africa some 100,000 years ago and formed modern human populations
- Genetic diversity of modern human populations is lower at the furthest destination (Americas) and highest in the place of origin (Africa)



Gene flow

- Also called migration
- Movement of individuals or gametes between populations (groups) resulting in an exchange of genetic material
- Migration changes allele frequencies in natural populations and largely determines their genetic variation
- Gene flow works against genetic drift
  - It increases genetic diversity within populations
  - It reduces differentiation between populations



Mutation

- Spontaneous change of nucleotides, chromosomes or the number of genes
    - New alleles
    - Primary source of diversity
- A small percentage of them create **new alleles**, which ultimately improve the adaptability of a population to its environment
    - Mutation rate of 10^{-6} to 10^{-8} per generation
    - Even at these slow rates many new alleles can be created due to the large genome size and species populations
        - In humans: 500-1000 genes with new mutations in each generation
- Mutation alone does not significantly change gene allele frequencies from generation to generation

The fate of a new mutation

- A mutation creates a new allele
    - Harmful
        - Disappears or remains in a recessive form in heterozygotes, at a low frequency
    - Neutral
        - It remains in the population
    - Profitable
        - Its frequency is increasing due to natural selection
- A new allele can be lost for random reasons
    - Genetic drift
- A deleterious or neutral allele can become beneficial in a new environment

Thank you