POPULATION GENETICS COURSE

prepared by OURARI Malika

Addressed to third-year undergraduate students and first-year Master's students in the LMD system in biology, medical students

PREAMBLE

Target audience: This course packet on the genetics of natural populations is aimed at third-year undergraduate students and first-year Master's students in the LMD system in biology, medical students, and students from fields related to the study of eukaryotic natural populations.

Total course hours: 45 hours of lectures, 20 hours of tutorials.

Prerequisites: Basic knowledge of general biology (cell division, mitosis, meiosis, modes of reproduction), ecology concepts (biogeography, habitat, bioclimates), basic genetics concepts (inheritance patterns, Mendel's laws, concepts of genes, alleles, genotypes, phenotypes), molecular biology (structure and function of DNA, RNA, proteins, replication, translation, etc.), and basic concepts in statistics and probability.

Course objectives: By the end of this course, students should be able to:

- Calculate allele and genotype frequencies in natural populations in equilibrium and in populations that are out of equilibrium, considering mating patterns (panmixia, inbreeding, homogamy) and evolutionary forces (mechanisms that modify the genetic structure of populations: mutations, selection, genetic drift, migration).

- Identify the different types of selection and speciation.

- Use mathematical models to predict changes in the genetic frequencies of populations (Model genetic processes).

- Study genetic variability by exploring the causes and consequences of variability in populations, particularly in relation to adaptation, speciation, and population survival.

- Apply population genetics in medicine (antibiotic resistance), and in agronomy (resistance to insecticides, herbicides).

POPULATION GENETICS

COURSE CONTENT

1. VARIATIONS IN NATURAL POPULATIONS

1.1. Nature of variation and concept of traits: morphological, anatomical, histological, cytological, physiological, ecological, biochemical, and molecular.

1.2. Ecological and genetic components of variation: ecotype, ecophenes, ecocline.

1.3. Origin of genetic variation: mutations, recombination, polyploidy, hybridization, and introgression.

2. GENETIC STRUCTURE, EVOLUTION, AND SPECIATION

2.1. Concepts, definitions, and general characteristics of a population: size, isolation, and mode of reproduction.

2.2. Panmixia and Hardy-Weinberg equilibrium

- Theoretical data, allele frequencies, and genotype frequencies
- Deviations from panmixia : inbreeding and homogamy

2.3. Populations in disequilibrium and natural selection:

- Mutation, migration, and gene flow
- Adaptive value and selection coefficients
- Polymorphism and the alternating advantages of homozygotes and heterozygotes
- Genetic drift and different types of selection.

2.4. Speciation

- Concepts and definitions of species
- Mechanisms of reproductive isolation
- Allopatric and sympatric speciation

INTRODUCTION

Population genetics is a discipline that emerged from the synthesis of the theories of Mendel, Darwin, and early 20th-century biostatisticians (Fisher, Wright, Haldane). They introduced the first methods for analyzing (measuring) the genetic variability of natural populations. The analysis of this genetic variability within and between populations allows us to trace their evolutionary history. Population genetics explains the observed modifications in populations over time and space by studying the evolutionary forces (sexual reproduction, mutation, natural selection, migration, genetic drift, etc.). It helps to explain, and even predict, the evolution of populations.

RECALLS

Genetics is the study of the mode of inheritance of hereditary traits. Hereditary traits are governed by elements called **genes**, arranged linearly on a support: the chromosome. A gene is a sequence of DNA that ensures the synthesis of a polypeptide. Each gene on a chromosome occupies a specific location called a **locus**. Each gene is found in duplicate in diploids. Each copy constitutes an **allele**. A gene can be represented by two alleles (diallelic) or multiple alleles (polyallelic, e.g., blood groups A, B, AB, O). An allele can be either **dominant** or **recessive**. According to Mendel, the dominant allele is the one that is expressed in the first-generation hybrid, while the other is the recessive allele. For a given trait, if both alleles of the pair are identical, the individual is said to be **homozygous**. On the other hand, if the two alleles are different, the individual is said to be **heterozygous.** A gene contains coded information that ensures the synthesis of a polypeptide. All allelic forms of a gene are found at similar positions on homologous chromosomes (except in the case of translocations). Homologous chromosomes are chromosomes that carry the same loci. They are paired together in structures called bivalents during prophase I of meiosis. Each bivalent consists of one chromosome from the maternal origin and one chromosome from the paternal origin. The combination of the two alleles located face-to-face on homologous chromosomes forms a **genotype**.

Mendelian genetics (1865) studies hereditary mechanisms by observing the dissimilarities and similarities (phenotypic traits) between parents and their offspring. Molecular genetics analyzes the mechanisms that lead from the gene to the trait. It studies the biochemistry of nucleic acids (replication, transcription, translation, restriction enzymes, sequencing, etc.). On a third level, population genetics statistically studies genotype and allele frequencies in populations. A population is defined as a group of individuals of the same species

living in a geographic area small enough that all sexual partners have the same probability of encountering each other.

OBJECTIVES OF POPULATION GENETICS

An individual holds, for a short period of time, a small portion of the gene pool; a gene pool being the total collection of genes and their different alleles present in a population. An individual's contribution to the processes of adaptation, evolution, and speciation is therefore minimal compared to the entire gene pool of the population. It is within the population that genes are organized into multiple combinations (in different genotypes) through sexual reproduction (segregation, recombination, crossing over). It is also within the population that genes can undergo transformations through mutations. This genetic variability is the basis for the persistence of natural populations, enabling them to adapt to a constantly changing environment, resist parasites, and new diseases. The population thus has two characteristics that are absent in the individual: continuity over time and the ability to change. Therefore, the population is considered the fundamental unit of adaptation, evolution, and speciation (formation of new species). Evolution is defined by the geneticist Dobzhansky as any change in allele and/or genotype frequencies within populations.

The objectives of population genetics are to:

- Identify genetic variability within natural populations by estimating allele frequencies and genotype frequencies using mathematical and statistical models.

- Evaluate the variation of these frequencies between populations and over time.

- Assess the influence of reproductive regimes (panmixia, inbreeding, autogamy, homogamy) and evolutionary forces (mutations, genetic drift, natural selection, migration) on the variation of genotype and allele frequencies in space (among populations) and in time (across successive generations).

FIELDS OF APPLICATION OF POPULATION GENETICS

Population genetics is of interest in:

- Basic sciences, by explaining the mechanisms of biological evolution and tracing the evolutionary history of species.

- Ecology, by demonstrating the impact of invasive species on the environment when introduced into new habitats, in biodiversity conservation studies, in the study of GMOs (dissemination, environmental impact), and in fisheries and hunting management programs.

- Medicine, in the search for genetic defects, epidemiology, and the study of inbreeding effects.

- Agronomy, in animal and plant genetic improvement, through the use of inbreeding or homogamic crosses and artificial selection.

CHAPTER I. GENETIC VARIABILITY IN NATURAL POPULATIONS

1. Introduction

A distinctive feature of the living world is the variability of phenotypes among individuals. Within a sexually reproducing species, there are no two individuals with exactly the same phenotypic characteristics, even identical twins: each individual is unique. Some of these variations are expressed at the phenotypic level (morphology, physiology, behavior), while others can be highlighted through the use of specialized techniques (variability in proteins or DNA sequences).

2. Concept of Traits

According to Bidault (1968), a trait is the set or part of the features or properties of an individual that can be measured or qualified, allowing comparisons with the same traits or properties of another individual. Measuring or qualifying a trait allows us to indicate its values or states. For example, in the case of pollen grains from a plant, the diameter, shape, and color are traits. The trait 'diameter' has a specific value, while 'shape' is observed in a particular state.

For a long time, the only traits that attracted the attention of analysts of variability were those related to morphology, i.e., those concerning the shape and dimensions of various parts of an individual. Nowadays, the analysis of variation includes a wide range of traits of various types.

- Morphological Traits

Most morphological traits, which were previously expressed qualitatively, are increasingly being expressed quantitatively. For instance, the different parts (leaves, petals,

sepals) of organisms are most often expressed by numerical ratios rather than being described using subjective terms like "lanceolate," "narrowly lanceolate," etc.

- Anatomical and Histological Traits

In plants, among the most commonly used anatomical traits are the structures of leaves and stems, as well as those of certain epidermises and stomatal features. The stem structure (the particularity of the endodermis) distinguishes different species of *Equisetum*. In the genus *Festuca* (F. Gramineae), leaf anatomy allows for the distinction of numerous taxa at the species or subspecies level. The structure of the epidermis characterizes different tribes of grasses by the shape and arrangement of epidermal elements. Stomatal dimensions distinguish infraspecific units in *Poa annua* L..

- Cytological Traits (Karyological Traits)

These concern the number, shape, and internal structure of chromosomes, referred to as the karyotype. The morphological study is defined by the position of the centromere (terminal, subterminal, median), the centromeric index, the length of the arms, and the possible presence of secondary constrictions (satellites). Based on the variation of these traits, a karyogram is established from an average of karyotypes. A karyogram is a schematic representation of the average karyotype derived from several cells. Karyotype analysis is recognized as being highly important in explaining evolutionary phenomena and in species delimitation.

For example:

1. The subfamilies Aveneae and Hordeae in grasses have a base chromosome number of 7 and long chromosomes, while other subfamilies like Paniceae have a base number of 9 or 10 and short chromosomes.

2. The human karyotype with 46 chromosomes is obtained after rearrangement of the chimpanzee's 48-chromosome karyotype.

- Physiological and Ecological Traits

These traits translate into particular ecological adaptations related to photosynthetic capacity, resistance to cold, heat, drought, or various soil constituents. For example, *Festuca glauca* (Gramineae) includes two morphologically similar varieties: one, tetraploid, found in calcareous soils, and the other, diploid, found in siliceous soils (Bidault, 1968).

- Biochemical Traits

From a technical point of view, chemotaxonomy is relatively recent. Certain molecules such as alkaloids, proteins, enzymes, polysaccharides, and anthocyanins are used through chromatography or electrophoresis. However, it is with the progress of molecular biology that a natural classification system (reflecting the phylogenetic tree of the species) can now be established. Amino acid sequences in proteins or nucleotide sequences in DNA or RNA provide discontinuous data (site comparison), while the degree of hybridization between DNA strands from different groups, along with allele frequencies, is expressed quantitatively.

Statistical methods using computer tools allow for the simultaneous analysis of many traits (multivariate analyses), offering a clearer picture of population variability. Various data analysis methods are commonly used (Principal Component Analysis - PCA, Correspondence Factor Analysis - CFA, Hierarchical Cluster Analysis - HCA).

3. Main Types of Variation

Depending on the level of study, we distinguish between intra-population and inter-population individual variation.

3.1. Individual Variation

It affects the different parts of an individual at a given time or the same parts at different times.

- Age-related Variation: This refers to the difference between immature or larval stages and adult stages (caterpillars, butterflies). Physiologically, we observe foliar polymorphism in the genus *Plantago*.

- Seasonal Variation: Mammals in temperate and cold regions can transform their coats into winter fur. Many birds, during part of the year, have dark plumage, which is replaced by bright nuptial plumage before the mating season begins.

- Variation in the structure of the epidermis of grasses depending on the part of the considered plant or organ. Many floral traits of *Plantago coronopus* vary distinctly depending on their position along the spike.

3.2. Variation Within a Population and Between Populations

Modern biological approaches recognize the importance of intraspecific variation and aim to determine its evolutionary significance. Studying variation within and between

populations helps explain the mechanisms of speciation. The advantages of maintaining a reserve of variation are:

- The greater the number of genetic types within a population, the more likely it is to have genotypes capable of surviving seasonal changes and other temporal transformations, especially in hostile environments. For example, if there are drought-resistant genotypes within a population normally living in a wet environment, this population will have a better chance of surviving during a drought period.

- This variation also allows for a greater utilization of the environment, as it facilitates the colonization of marginal habitats.

The study of intra- and inter-population variability requires a suitable sampling procedure, followed by statistical methods (distribution law, mean, mode, standard deviation, correlation coefficient) to deduce the properties of the entire population. Sampling should cover the entire distribution area. The analysis must focus on homologous traits at the same phenological stage in both field samples and transplanted samples grown under the same conditions (to eliminate the effects of genetic variation).

4. Ecological and Genetic Components of Variation

4.1. Genetic Component: Concepts of Ecotypes and Ecoclines

Different environmental factors strongly influence genetic variability in populations through natural selection. The environment selects individuals that are genetically best adapted to it. Populations, if they thrive in different ecological conditions, may differentiate into several subunits, each adapted to its own environmental conditions. Turesson called these populations ‘ecotypes’, where their hereditary characteristics are the product of the interaction between genotype and environment.

Ecotypic variation is more often continuous than discontinuous, and this continuity is related to the gradual variation of the environment. This finding led Gregor et al. to adopt the concept of ‘ecocline’ to describe this continuous ecotypic variation. For example, Dobzhansky (1948) showed that the frequency of certain variations in chromosome structure increases regularly with altitude.

4.2. Ecological Component: Concept of Ecophenes

The environment can directly affect the phenotype of an individual, leading to non-heritable modifications called accommodations. The populations resulting from these modifications are known as ‘ecophenes’.

Examples:

1. Two plants, one growing in optimal conditions and the other in poor soil.

2. A mouse raised at 28°C has a tail length of 93.1 mm, but at 6°C, it decreases to 75.9 mm.

3. A man from the plains who moves to the mountains experiences an increase in the number of red blood cells: the decrease in oxygen levels due to lower atmospheric pressure is compensated by the more intense production of red blood cells, but this phenomenon disappears upon returning to the plains.

‘Homeostasis’ refers to an individual's ability to change its phenotype in such a way that it can survive and adapt to environmental variations.

5. Origin of Genetic Variation

Several factors are responsible for genetic variability: sexual reproduction (segregation, recombination), mutations, selection, etc.

5.1. Recombination and Segregation

During meiosis, homologous chromosomes from the mother and father pair up and recombine (crossing-over). Then, they randomly segregate into haploid reproductive cells (spores or gametes).

5.2. Mutations:

Mutations are sudden, hereditary changes that affect part or all of the genotype. Mutations can be classified into two types: gene mutations and chromosomal mutations.

- Gene mutations:

A gene mutation occurs when the sequence of a gene is altered. If only a small number of nucleotides are affected, this is called a point mutation. These are localized changes within a cistron (functional unit of a gene).

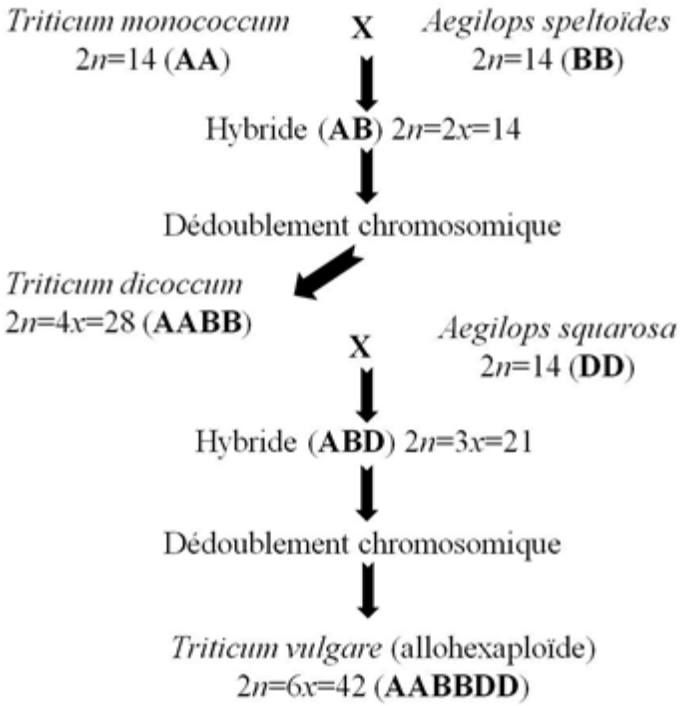
- Chromosomal mutations: Chromosomal mutations can involve structural changes to chromosomes or numerical changes in chromosomes.

* Structural changes: These include fragmentation (breakage of a chromosome), deletions (loss of a chromosome segment), duplications (addition of a chromosome segment to an original chromosome), inversions, and translocations.

* Chromosomal Numerical Changes: In many organisms, the number of chromosomes is generally constant. However, some organisms belonging to the same systematic units sometimes exhibit varying chromosome numbers. For example, in *Festuca ovina*, forms with 14, 28, 42, 49, 56, and 70 chromosomes are observed. The smallest common multiple for these numbers is 7, which represents the **base number** of the species and is symbolized by the letter '***x***'. The smallest number in the series represents twice the base number and is the **diploid number** of the species. The **haploid chromosome number** in gametes is represented by the letter '*n*'. The diploid chromosome set for the species will therefore be 2*n*=2*x*=14. Other taxa with chromosome numbers that are multiples of 7 are classified as **polyploids**. They are tetraploids when 2*n*=4*x*=28, **hexaploids** when 2*n*=6*x*=42, *octoploids* when 2*n*=8*x*=56, and so on. Polyploidy is more frequent in plants and is relatively common in certain vertebrate groups (amphibians, reptiles). In most cases, it is a pair polyploidy (4x, 6x). The existence of a triploid toad species, *Batura* (*Bufo pseudoraddei baturae*), found in the mountains of Pakistan, has been documented. This species reproduces sexually, and meiosis produces male gametes with '*x*' chromosomes after reduction of 3*x* to 2*x*, and female gametes with 2*x* chromosomes after increase from 3*x* to 4*x*. Fertilization results in individuals with 3*x* (2*x* X *x*). Chromosome numbers can also vary due to the presence or absence of one or more chromosomes, which defines **aneuploidy** in contrast to the previously mentioned **euploidy**. When a chromosome is lost, it is called **monosomy** (2*n*-1). **Polysomy** occurs when a chromosome in the normal set is present in more than two copies. **Trisomic** individuals have 2*n*+1, **tetrasomic** individuals have 2*n*+2, and so on. In plants, supernumerary chromosomes can be observed, referred to as **'B' chromosomes**, as opposed to the normal **'A' chromosomes**. Their origin is unknown, and they are smaller than 'A' chromosomes. They do not show homology with normal chromosomes and do not pair with them during meiosis. Many species also have multiple base numbers and may present several polyploid series (e.g., *Centaurea* L.*, *Crepis* L.).

- Hybridization, introgression

Hybridization is a cross between two genetically and taxonomically different individuals. It thus allows the contact of different genomes. It constitutes a source of variability for natural populations and is frequent in plants. It is generally accompanied by a reduction in fertility. For example, the cross between a donkey and a horse produces a mule, which is vigorous but sterile. In plants, in the case of sterile interspecific hybrids, the transition from the diploid form to the polyploid form allows the creation of fertile allopolyploids (polyploids after hybridization). For example, this is part of the history of modern wheat.



Triticum vulgare represents all the wheat varieties currently cultivated.

Due to their high sterility, these hybrids tend to cross with one of the parental species. This results in the production of many backcrosses. These backcrosses with either of the original types, through recombination and genetic segregation, lead to the transfer of part of the genetic material from one subspecies to another. This phenomenon is known as **hybrid introgression** or **introgressive hybridization**.

6. Detection of Variation: The Concept of Polymorphism

Population genetics primarily focuses on genetic variability, referred to as polymorphism. Polymorphism refers to variations in the nucleotide sequence of a gene's DNA within a population. A gene is considered polymorphic if there are at least two alleles with a frequency equal to or greater than 5% or 1%. Several types of polymorphism are studied in the analysis of population variation:

6.1. DNA Polymorphisms:

- **Nucleotide Polymorphism** or **Single-Nucleotide Polymorphism** (**SNP**) is the substitution of a single base pair in the genome between individuals of the same species. SNPs are the basis for differences in disease resistance (e.g., sickle cell anemia, β-thalassemia). SNPs can be found within coding regions of genes (exons), non-coding regions of genes (introns), or intergenic regions between genes. We refer to synonymous allelic forms when multiple forms of an SNP lead to the same polypeptide sequence, and to non-synonymous forms when the resulting sequences differ. SNPs allow for the identification of genotypes (e.g., recognizing individuals) or contribute to the construction of genealogical (individuals) or phylogenetic (species) trees.

- **Restriction Fragment Length Polymorphism** (**RFLP**) is characteristic of DNA molecules, allowing them to be distinguished in two different genomes (genetic fingerprints, paternity tests). After extraction and purification, the DNA is cut into restriction fragments by a restriction enzyme at a specific sequence (restriction site). The resulting restriction fragments are then separated according to their length through electrophoresis.

- **Microsatellites** are repetitions of one to six base pairs repeated '*n*' times ($5 \leq$ '*n*' ≤ 40). For example: TGTGTGTGTG….. Over 50,000 microsatellites (TG)n are present throughout the human genome. They are detected exclusively by PCR (Polymerase Chain Reaction), which amplifies a given region of DNA in vitro to obtain a sufficient quantity for detection and study.

6.2. Chromosomal Polymorphisms:

- Chromosomal Rearrangements / Chromosomal Mutations (gain, loss, rearrangement of segments)

- Change in Chromosome Number Without Structural Changes: Heteropolyploidy => Aneuploidy and Euploidy, genomic mutations (polyploidy)

6.3. Enzymatic and Protein Polymorphisms

It is through electrophoresis that these polymorphisms were identified. It is applied in the detection of severe hereditary diseases caused by the existence of genes or gene associations that are unfavorable. Electrophoresis is a useful technique for analyzing population variability, geographically differentiating them, and monitoring their evolution under the influence of various factors. Electrophoresis followed by specific revelation of enzymatic activities allows the creation of zymograms from raw extracts. Other proteins can be separated based on their electric charge or spatial conformation.

6.4. Serological Polymorphisms

The A, B, O blood groups in humans (Landsteiner, 1900) are the oldest recognized form of polymorphism. There are four blood groups (A, B, AB, and O), which differ by the presence, absence, or combination of A or B antigens on the surface of red blood cells. The second polymorphism refers to the Rhesus factors, which are present on the surface of red blood cells. If these factors are present on a person's red blood cells, they are said to be "Rh (Rhesus) positive." If these factors are absent from their red blood cells, they are said to be "Rh negative."

7. Quantification of Polymorphism

Polymorphism for a gene in a population can be measured by different indices:

- Average number of alleles per locus (A)

The average number of alleles per locus (A), also called the allele frequency or allelic richness, is defined for n_i alleles at locus 'I' and for L *loci* as:

$$A = 1/L \sum_{i=1}^{L} n_i$$

Example: For 3 *loci* numbered 1, 2, and 3, having 2, 3, and 2 alleles respectively, A = (2 + 3 + 2) / 3 = 2.33. The measurement of this parameter is particularly important for conservation strategies.

- Proportion of polymorphic loci (p)

- Allelic frequencies for polymorphic loci

- Diversity indices based on allele frequencies (e.g., Nei's index, He = 2pq)

Nei's Genetic Diversity Index. It is calculated as follows:

He = Number of polymorphic loci / Number of studied loci

~ If there is only one allele, He = 0.

~ The more alleles there are, the higher the index will be.

~ The closer the allelic frequencies are to each other, the higher the index will be.

Example 1: The allozyme polymorphism by Hamrick and Godt (1989).

Life form	P	He	
annual	30.2%	0.105%	
Perennial herbaceous plants	28%	0.096	
Perennial woody plants	50%	0.149	

According to these results, trees appear to be the most variable.

Example 2: In cheetahs (p) = 0.02 and He = 0.0004 (Bottleneck, loss of genetic variation leading to a loss of evolutionary potential).

- Fixation Index (Fis)

The Wright's Fis parameter, also known as the fixation index and previously called the coefficient of inbreeding (Wright, 1969), is calculated using the following formula:

Fis = (He - Ho) / He = 1 - (Ho / He);

Ho: observed heterozygosity; He: expected heterozygosity under Hardy-Weinberg equilibrium.

Fis = 1 indicates complete fixation (in the case of self-fertilization), Fis less than 1 indicates excess heterozygosity, and Fis = 0 indicates a population in Hardy-Weinberg equilibrium.

CHAPTER II. GENETIC STRUCTURE, EVOLUTION, AND SPECIATION

1. Populations and their characteristics

1.1. Reproductive isolation

1.2. Mechanisms of isolation

1.3. Population size

1.4. Modes of reproduction

2. Panmixia and Hardy-Weinberg Equilibrium

2.1. Genotypic frequencies

2.2. Allelic frequencies

2.3. Multiple allele cases

2.4. Sex chromosome-linked genes

3. Mating Modes

3.1. Mating maintaining allelic frequencies

3.1.1. Inbreeding and self-fertilization

3.1.2. Homogamy

3.2. Mating altering allelic frequencies (e.g., dominant males, see selection chapter)

4. Populations in disequilibrium

4.1. Mutations

4.2. Natural Selection

4.3. Combined effect of mutations and selection

4.4. Migration

5. Speciation

5.1. Concepts and Definition of Species

5.2. Mechanisms of Reproductive Isolation

5.3. Mechanisms of Speciation

CHAPTER II. GENETIC STRUCTURE, EVOLUTION, AND SPECIATION

1. Populations and their characteristics

1.1. Reproductive Isolation

The various means that prevent the crossing of one species with another are called isolation mechanisms. Speciation is linked to two types of biological processes that are more or less independent and can occur successively or simultaneously. One controls the diversification of populations (mutation, recombination, segregation…), while the other establishes isolation barriers that represent the ultimate stage of the individualization of specific units (i.e., species). Diversification can occur gradually under the predominant influence of natural selection, or suddenly through major chromosomal mutations or hybridization. Therefore, we distinguish between gradual speciation and abrupt speciation. In the case of abrupt speciation, new types are immediately isolated from those that gave rise to them. Diversification and isolation are thus simultaneous. The study of populations has shown that populations evolve according to one or more of the following three structural elements:

1) Series of contiguous populations with gradual changes (cline variation);

2) Geographically separated populations from the main part of the species' range (geographic isolates);

3) Narrow zones with increasing variability, often rapidly bordered on each side by groups of stable and highly uniform populations or subspecies (hybrid zones).

- Cline Variation:

Across a series of contiguous populations, it is generally observed that changes follow a regular progression due to the gradual variation in the environment and the gene flow between adjacent populations, which tends to smooth out the differences between populations.

- Geographic Isolates:

An isolate is defined as a population or a group of populations separated by an extrinsic barrier and unable to freely exchange genes with other populations of the species. The essential

characteristic of a geographic isolate is being separated from the rest of the species by a discontinuity. However, isolation is never complete, as some gene flow always occurs (even in oceanic islands, otherwise they would not have been colonized initially). These isolates have three possible fates: they become distinct species, they disappear completely, or they re-establish contact with the main body of the species.

- Hybrid Zones:

When two populations are isolated from each other, their gene pools become independent, and their genetic compositions diverge continuously. When geographic isolation disappears and the populations re-establish contact, certain phenomena can be observed in the contact zone that indicate the degree of genetic differentiation that occurred during isolation.

1.2. Mechanisms of Isolation

Isolation mechanisms are of two types: internal barriers and external barriers.

- External Barriers: These are pre-copulatory or pre-pollination mechanisms that prevent interspecific crosses, such as:

*Seasonal Isolation: Species do not meet. For example, two species that do not flower at the same time of year.

*Ethological Isolation: (ethos = habit, customs, behavior). Species meet but do not exchange genes (no fertilization). There is a restriction on random mating.

*Mechanical Isolation**: This is very effective in plants. The stigmatic lobes have sizes and shapes that vary by species, creating barriers to pollination by insects.

*Ecological Isolation of the Sympatric Type**: Ecological niches are different.

- Internal Barriers: These are post-copulatory or post-pollination mechanisms. They reduce the success of interspecific crosses.

*Gamete Mortality: Prevents fertilization. Pollen tubes are unable to reach the ovules.

Example 1: In polyploids, pollen tubes are of large diameter and have difficulty traveling through the styles of diploid species.

Example 2: Sperm may encounter an antigenic reaction in the female reproductive tract.

*Zygote Mortality: Fertilization occurs. In the genus *Datura*, the embryo stops development when it reaches the eight-cell stage. Various mechanisms can lead to zygote mortality, such as:

~ The albumen acting as an inhibitor to embryo development.

~ The presence of a lethal gene that has no effect on individuals of the species but causes the death of the hybrid.

~ The embryo never reaches maturity due to genetic disharmony between the parental genomes.

~ The inviability of hybrids results from the interaction of certain hybrid genotypes with the presence of one of the parental species.

~ Sterility of Hybrids: There are two types of hybrid sterility (genetic and chromosomal): **Genetic sterility** occurs when sexual organs fail to undergo meiosis or due to meiotic abnormalities of genetic origin; Chromosomal sterility results from the lack of homology between the maternal and paternal chromosomes of the hybrid.

~ Inviability or Weakness of F1 descendants or subsequent generations: Sometimes isolation is only manifested in the second generation of hybrids.

1.3. Population size

Evaluating population size is a technically challenging task. It is necessary to consider variations in size over time, individual movements, and the interpopulation exchanges they entail, the duration of generations, and any overlaps that may occur. One must also take into account the actual distribution of individuals within a population in nature, as assumptions of uniformity and continuity have often been used, leading to the very insufficient notion of density. Due to these material and conceptual difficulties, the size of a population is often poorly estimated.

Example 1: A population of *Cepaea nemoralis* snails consists of approximately 5,000 to 20,000 individuals.

Example 2: Populations of moths range from 500 to 20,000. Even well-organized groups of higher vertebrates do not have a fixed size as Baboon groups (*Papio*) can range from just a few individuals to nearly 200.What matters in population genetics is the impact of these numerical fluctuations on the biology of the corresponding populations, their structures, and their genetic composition.

1.4. Modes of reproduction

There are two modes of reproduction: sexual and asexual. The biological function of sexual reproduction is to generate a wide variety of different genotypes. It allows for the recombination of genetic factors from parent individuals into numerous genetically unique zygotes. There are different types of sexual reproduction:

- **Autogamy**: This is fertilization between two gametes from the same individual. Hermaphroditism is widespread in the animal kingdom (e.g., protozoa, fish; a single individual produces both male and female gametes). This mode does not necessarily lead to inbreeding. In most hermaphroditic species, there are several mechanisms that reduce or eliminate the chances of autogamy, such as the production of only male or female gametes at a given time, one after the other (protandry or protogyny).

- **Allogamy**: This is fertilization between two gametes from different individuals. In animals, this is referred to as gonochorism (separation of sexes). The presence of a genetic incompatibility system or various floral mechanisms can promote allogamy. A genetic incompatibility system is a genetic and physiological mechanism that favors allogamy. It is characterized by the presence of S alleles. Two individuals carrying the same S allele are incompatible, preventing autogamy. Heterostyly is another incompatibility mechanism that also prevents autogamy.

- **Apomixis:** This is the formation of a seed without fertilization. Since this mode of reproduction prevents genetic recombination and segregation, it could be considered a limiting or negative factor in evolution. However, apomixis is always partial, meaning that a given plant can reproduce both sexually and apomictically at the same time (facultative apomixis). Sexual reproduction, through recombination and segregation, generates diverse genotypes, which can in turn be either sexual, apomictic, or both. Thus, certain favorable genetic recombinations can be preserved unchanged through apomixis. The primary advantage of apomixis is therefore the possibility of retaining favorable genetic combinations. It also happens that the offspring of an

apomictic plant may include individuals with one or more extra or missing chromosomes due to anomalies occurring during meiosis. These new types can only persist through asexual reproduction.

- Parthenogenesis: This is a mode of reproduction in which a new individual is produced without fertilization. It is common in plants and certain arthropods, including insects and crustaceans. Male and female gametes can develop into embryos without fertilization or meiosis. The absence of meiosis allows triploid ($3x$) and pentaploid ($5x$) individuals to reproduce without difficulty.

2. Panmixia and Hardy-Weinberg Equilibrium

We consider:

- Autosomal genes

- Genes existing in two allelic forms: Most genes exist in multiple allelic forms. Some esterases, for example, are determined by six alleles. For our study, we consider the case of a single pair of alleles.

- A panmictic population: Random mating occurs, independently of genotype and kinship.

- A very large population size: This allows us to neglect fluctuations in gene frequency.

- Stable population size: Studies of natural populations have shown that population size tends to remain relatively stable. When a population colonizes a new environment, its size increases from one generation to the next until it reaches a limit determined by factors such as food availability and space. Once this maximum is reached, the population remains in equilibrium with its environment. The population size remains practically constant from one generation to the next, meaning it is in equilibrium or stationary.

The Hardy-Weinberg law allows for the calculation of gene frequency in a stationary population. Independently discovered by Hardy (an English mathematician) and Weinberg (a German physician), this law states that in a panmictic population at equilibrium, where there is no mutation, selection, migration, and where the population is large, with all eggs reaching adulthood, all individuals being equally fertile, and all gametes participating in fertilization, the

proportion of alleles and genotypes will remain absolutely constant from one generation to the next. Regardless of the initial frequency, this state is reached within the first generation.

2.1. Genotypic Frequencies

Let there be two alleles (A and a) for the same gene in a diploid population of size N, where A is dominant over a. There are three possible genotype categories: (AA), (Aa), (aa). Let x represents the number of individuals with the (AA) genotype, y represents the number of individuals with the (Aa) genotype, and z represents the number of individuals with the (aa) genotype.

$x + y + z = N$

We aim to determine the frequency of each genotype. The genotypic frequency is the ratio of the number of individuals with that genotype to the total number of individuals in the population. The sum of the frequencies of each genotype constitutes the genotypic structure of the population.

The frequency of genotype AA = $x/(x+y+z)=x/N$ The frequency of genotype Aa = $y/(x+y+z)=y/N$ The frequency of genotype aa = $z/(x+y+z)=z/N$	**Genotypic structure of the population**

2.2. Allelic Frequencies

The same statement as before, but this time we aim to calculate the frequency of each allele. Let ‘p’ be the frequency of allele ‘A’ and ‘q’ the frequency of allele ‘a’ ($p + q = 1$).

Each ‘AA’ individual carries two ‘A’ alleles, so we have ‘2x’ ‘A’ alleles coming from the homozygous dominants in the population.

Each ‘A’ individual carries one ‘A’ allele, so we have ‘y’ ‘A’ alleles coming from the heterozygous ‘Aa’ individuals in the population.

The total number of ‘A’ alleles in the entire population is ‘2x’ (from the homozygotes) + ‘y’ alleles (from the heterozygotes).

The frequency of allele ‘A’ is given by:

‘p’ = (2x + y) / 2N = (x + ½y) / N = x / N + ½y / N

The frequency of allele ‘A’ is therefore equal to the frequency of the homozygous genotype ‘AA’ plus half the frequency of the heterozygous genotype (Aa), i.e.:

fr(A) = fr(AA) + ½ fr(Aa)

It is thus possible to calculate allelic frequencies based on the partial genotype frequencies and the total population size.

Application exercises:

Ex. 1. Consider the blood group numbers M and N in a population of 10,694 individuals.

Blood Group	M	MN	N
Number	3356	5178	2160

Calculate the frequencies of the different blood groups, phenotypic frequencies, genotypic frequencies, and allelic frequencies.

Blood group frequencies:

Fr(M): 0.3138

Fr(MN): 0.4842

Fr(N): 0.2020

The frequencies of the blood groups M, MN, and N are ‘phenotypic frequencies’. Since M and N are co-dominant (neither is dominant over the other), the phenotypes and genotypes overlap. These frequencies are also ‘genotypic frequencies’. Given that group M corresponds to the (MM) genotype, group N corresponds to the (NN) genotype, and group MN corresponds to the (MN) genotype.

Allelic frequencies:

'p' = fr(M) = fr(MM) + ½ fr(MN) = 3356 / 10694 + ½ (5178 / 10694) = 0.5559

'q' = fr(N) = fr(NN) + ½ fr(MN) = 2160 / 10694 + ½ (5178 / 10694) = 0.4449

Let's show that, in a population at Hardy-Weinberg equilibrium, allele frequencies and genotype frequencies remain constant from one generation to the next.

At generation 'n', let the frequency of allele 'A' be 'p' and the frequency of allele 'a' be 'q', with p + q = 1. What happens to these frequencies in generation n+1?

If alleles 'A' and 'a' are equally distributed among sperm and eggs, the individuals of generation 'n' produce 'p sperm carrying allele 'A', 'q' sperm carrying allele 'a', 'p' eggs carrying 'A', and 'q' eggs carrying 'a'.

The result of random mating, which determines the genetic composition of generation n+1, is given in the following table:

	A(p)	a (q)
A(p)	AA (p^2)	Aa (pq)
a (q)	Aa (pq)	Aa (q^2)

In the case of two alleles, there is a 'p' probability of drawing 'A', and therefore a probability of p^2 of obtaining the (AA) genotype. Similarly, there is a probability of q^2 of obtaining the (aa) genotype. To obtain the (Aa) genotype, there are two times pq possibilities.

The generation (n+1) will consist of p^2 AA zygotes, 2pq Aa zygotes, and q^2 aa zygotes.

The allele frequencies in this generation will be equal to:

fr(A) = fr(AA) + ½ fr(Aa) = p^2 + pq = p(p + q) = p,

fr (a) = fr (aa) + ½ fr (Aa) = q^2 + pq = q(p + q) = q

which are precisely the same proportions as in generation n.

It is easy to verify that the sum of the different genotypic probabilities equals 1:

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

Application to a recessive gene responsible for phenylketonuria, due to a metabolic deficiency and causing severe mental retardation in homozygous recessive individuals. This disease affects 1 in 25,000 individuals in a population. What is the frequency 'q' of this allele?

The frequency of affected individuals (aa) is $1/25{,}000 = q^2$.

The frequency of allele 'a' is $q = \sqrt{(q^2)} = \sqrt{(1/25{,}000)} = 1/159$.

$p + q = 1$, so $p = 1 - q = 158/159 = 79/80$.

The frequencies of the different genotypes are therefore:

- fr(aa) = 1/25,000 → affected individuals

- fr(Aa) = 1/80 → normal

- fr(AA) = 79/80 → normal.

It is thus observed that the frequency of heterozygotes is 300 times higher than that of homozygotes affected by phenylketonuria. In general, the rarer an allele is, the greater the proportion of heterozygotes compared to homozygotes for that allele.

Note:

The phenotype [A] due to the dominant allele (A) can correspond to two genotypes: (AA) and (Aa). The only phenotype whose genotype is certain is the one caused by the recessive allele (aa). If the population is in equilibrium, an estimate of 'q' can be obtained from q^2 (the frequency of homozygous aa individuals).

Example: If 75% of the population has phenotype [A], then 25% has phenotype [a]. If the population is in equilibrium for this locus, we can write:

q^2 = frequency of (aa), hence $q^2 = 0.25$, leading to $q = 0.5$ and $p = 0.5$.

2.3. Case of Multiple Alleles

Let there be three alleles A, a, and a' with a hierarchical dominance relationship A > a > a' and respective frequencies p, q, and r. Random crossings will produce zygotes with the following frequencies:

$(p+q+r)^2=$	p^2 (AA)	+	q^2 (aa)	+	r^2 (a'a')	+	2pq (Aa)	+	2pr (Aa')	+	2qr (aa')

Different frequencies: p^2 2pq 2pr q^2 2qr r^2

Different genotypes: (AA) (Aa) (Aa') (aa) (aa') (a'a')

Different phenotypes: [A] [a] [a']

Example: Let there be three alleles determining coat coloration in rabbits: 'C' (wild type), 'ch' (Himalayan), and 'c' (albino), following the dominance relationship C > ch > c with respective frequencies p, q, and r.

a. If a rabbit population containing wild-type, Himalayan, and albino individuals is panmictic, what will be the expected genotype frequencies in the next generation as a function of p, q, and r? Deduce the phenotypic frequencies in terms of p, q, and r.

b. A sample of rabbits contains 168 wild-type, 30 Himalayan, and 2 albino individuals. Calculate the allele frequencies p, q, and r.

c. Given the frequencies p = 0.5, q = 0.1, and r = 0.4, calculate:
- The proportions of different genotypes among wild-type rabbits.
- The frequency of wild-type rabbits.
- The frequency of Himalayan rabbits.

Answer: **a):**

Genotypic frequencies: p^2 2pq 2pr q^2 2qr r^2

Different genotypes are: (CC) (Cch) (Cc) (chch) (chc) (cc)

Different phenotypes are: [C] [ch] [c]

Phenotypic frequencies:

fr [C]= p^2+2pq+2pr, fr [ch]= q^2+2qr, fr [c]= r^2

b) :

Phenotypic frequencies:

fr [C]=168/200=0.84= p^2+2pq+2pr

fr [ch]=30/200=0.15= q^2+2qr

fr [c]=2/200=0.01= r^2

fr [c]= 2/200 = 0.01 = $r^2 \Rightarrow$ r = 0.1

(q^2+2qr+ r^2)= fr [ch]+ fr [c]= 0.15 + 0.01= 0.16=$(q+r)^2$=0.16$\Rightarrow$ q+r=0.4 $\Rightarrow$ q=0.4-0.1=0.3

q=0.3 ; since p+q+r=1 $\Rightarrow$ p = 1-q-r = 1-0.3-0.1=0.6 p=0.6

c: Frequencies of different genotypes among wild rabbits:

fr (CC)= p^2= $(0.5)^2$= 0.25

fr (Cch)= 2pq=2x0.5x0.1=0.1

fr (Cc)=2pr=2x0.5x0.4=0.4

Frequencies of wild rabbits (phenotypes): fr [C]=fr (CC)+fr (Cch) + fr (Cc)= 0.25+0.1+0.4=0.75

Himalayan rabbit frequencies (phenotypes): fr [ch]=fr (chch)+fr (chc)= q^2+ 2qr=$(0.1)^2$+ 2x0.1x0.4=0.09

2.4. Case of Sex-Linked Genes**

A gene is sex-linked when it is carried on the X chromosome. For two alleles ‘A’ and ‘a’, there are:

- Three genotypes for females: X^AX^A, X^AX^a, X^aX^a
- And two genotypes for males: X^AY, X^aY

Heterozygosity only exists in females. Males are **hemizygous** for sex-linked genes. The initial population is divided into two subpopulations: one for females and one for males, in order to calculate allele frequencies and genotype frequencies. Thus, we can calculate allele frequencies separately for females, as we did previously for autosomal genes. Similarly, we can calculate allele frequencies separately for males.

For males, the frequency of X^AY males corresponds to the frequency p of allele 'A', and the frequency of X^aY males corresponds to the frequency q of allele 'a'. **In males, the genotype frequency is equal to the allele frequency.**

Note:

The recessive phenotype is more common in males ($q♂ = 0.5$, $q^2♀ = 0.0025$).
Theoretically, when the population is in equilibrium, the allele frequencies calculated for females and males are identical. Therefore, it is sufficient to calculate the frequencies for males, due to the simplicity of the calculations, to obtain the allele frequencies for the entire population.

Often, the frequencies calculated for females and males are not identical. The frequency p of 'A' in the entire population is given by the following formula:

$$\mathbf{p=(2/3)\ p♀+(1/3)p♂}$$

Ex.1. Consider a locus A/a located on the X chromosome. In a population, we count:

90 ♂ X^AY
10 ♂ X^aY
77 ♀ X^AX^A
21♀ X^AX^a
2 ♀X^aX^a
- Is this population in equilibrium?
- Calculate the frequency p of A in the total population.

Among the ♀: p♀=(77/100)+1/2 (21/100)= 0.875 ; q♀=(2/100)+1/2 (21/100)=0.125

Among the ♂ : p♂=90/100=0.9 q♂=10/100=0.1

Let us consider the population in equilibrium, where allele frequencies in males are equal to those in females, and let the frequencies in males be p = 0.9 and q = 0.1

Nbr of ♀ X^AX^A $=p^2x100=(0.9)^2x100=$**81**
Nbr of ♀ X^AX^a $=2pqx100= 2x0.1x0.9x100=$**18**
Nbr f ♀X^aX^a $=q^2x100= (0.1)^2x100=$**1**

Let's calculate the χ_2:

$\chi_{2cal} = \Sigma(\text{ni obs-ni cal})^2/\text{ni cal} = (77-81)^2/81+(21-18)^2/18+(2-1)^2/1=1.6$

Number of degrees of freedom 3-1=2, $\chi_2$cal < $\chi_2$tab (6) => the population is in equilibrium

Frequency p of the 'A' allele in the entire population:

$$p=(2/3)p♀+(1/3)p♂=(2/3)x\ 0.875+(1/3)x0.9=0.883$$

Ex.2: In domestic cats, coat color is determined by a pair of codominant sex-linked alleles X^bX^y

	Phenotype		
	black	mosaic	yellow
♀	X^bX^b	X^bX^y	X^yX^y
♂	X^bY	--------	X^yY

Let 'p' be the frequency of 'b' and 'q' that of 'y'.

p = [2x (number of black ♀)+(number of mosaic ♀)+ (number of black ♂)]/(2 number of ♀+ number of ♂)

q = [2(number of yellow ♀)+(number of mosaic ♀)+(number of yellow ♂)]/ (2 number of ♀+ number of ♂).

3. Modes of Crossing

The Hardy-Weinberg law is essentially a theoretical law. It is only applicable if crossings occur randomly. This is an ideal case that never occurs in nature, where at least one, and generally several, of its conditions for application are absent. A number of factors always influence the choice of mates, making one type of crossing more likely than another. These non-panmictic crossings can be divided into two categories:

- Crossings that modify genotype frequencies but maintain allele frequencies.

- Crossings that modify both genotype frequencies and allele frequencies.

3.1. Crossings that Maintain Allele Frequencies

The maintenance of allele frequencies is ensured in inbreeding crossings, that is, crossings between related individuals, and in the extreme case of inbreeding, which is self-fertilization. It is also ensured in homogamic crossings, where phenotypically similar individuals tend to mate (mate choice based on size, color, ornaments, etc.).

3.1.1. Inbreeding and Self-Fertilization

Inbreeding refers to the union between individuals that share a certain degree of kinship. Inbreeding can have a social origin, as is the classic example of the dynasties of ancient Egypt: in order to preserve royal blood, marriages took place between siblings. It can also have a geographical origin, such as in isolated tribes or villages. In birds, it is quite common for pairs to form from the same brood. Another example is that of the fig wasps (blastophages) that develop and reproduce within the same fig.

Self-fertilization is widespread among flowering plants with bisexual flowers and hermaphroditic mollusks. However, self-fertilization is not an absolute rule, as special mechanisms can sometimes prevent it, such as sterile genes, heterostyly, protandry, or protogyny.

Inbreeding crossings are used to improve cultivated plants or domesticated animals. We will see why by examining the case of self-fertilization.

Let us assume that the population is originally panmictic, where we will follow the fate of the alleles ‘A’ and ‘a’, which are present at frequencies ‘p’ and ‘q’ respectively. The genotype composition of the population at the initial generation F_0 is:

$$AA : p^2 \qquad Aa : 2pq \qquad aa : q^2$$

Individuals with genotype (AA) and individuals with genotype (aa) self-fertilize, resulting in (AA) and (aa) offspring, respectively. For (Aa) individuals, the result of self-fertilization is different. In (F_1), we have:

2pq (Aa x Aa) give 2pq (¼AA + ½ Aa + ¼ aa) = pq (½AA + Aa + ½aa),

where the (AA) and (aa) will join their similar counterparts from homogamic crossings to form the third generation.

Thus, the number of heterozygotes decreases by half with each generation:

- 2pq at generation F_0

- pq at generation F_1

- pq/2 at generation F_2

- $pq/2^2$ at generation F_3

- $pq/2^{n}\text{-}^{1}$ at generation F_n

The number of homozygotes simultaneously increases by the corresponding amount through the addition of successive terms of a geometric progression with a ratio of '½'. Thus, at generation 'n', the increase for each type of homozygote (since there are two types) is equal to:

$\frac{1}{2}$ (pq + pq/2 + pq/4 +…….+$pq/2^{n-1}$) = pq/2 + pq/4 +……$pq/2^{n} \simeq$ pq

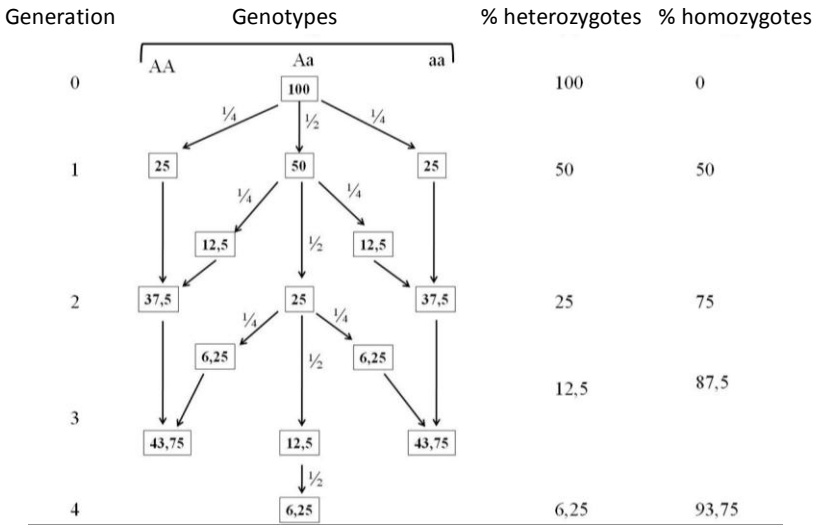
The frequency of (AA) homozygous individuals after 'n' generations is equal to:

p^2 (initial frequency) + pq (increase) = p (p + q) = p

The frequency of (aa) homozygous individuals is equal to:

pq + q^2 = q (p + q) = q

So, the genotype frequencies are equal to the allele frequencies (since there are no more heterozygotes).



The allele frequencies are therefore:

fr (A) = fr (AA) + ½ fr (Aa) = fr (A) + 0 = **p**

fr (a) = fr (aa) + ½ fr (Aa) = fr (a) + 0 = **q**

Allele frequencies thus remain constant with this type of crossing, but genotype frequencies vary. Heterozygotes are fairly quickly eliminated (after about nine generations, approximately).

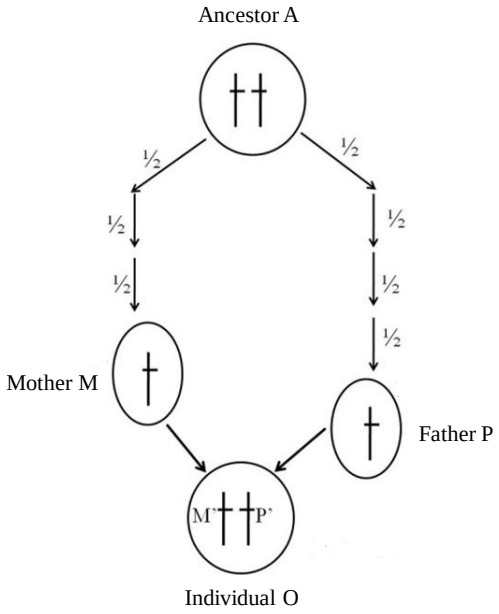
Inbreeding coefficient

The degree of inbreeding in a **population** is measured by the coefficient α, which ranges from zero (absence of homozygosity) to one (complete homozygosity) and expresses the rate of lost heterozygosity with each generation. The inbreeding coefficient α of an **individual** is the probability that two alleles it possesses at a randomly chosen locus are identical. At the individual level, this coefficient indicates the probability that the two alleles at a locus are two copies of the same allele from a common ancestor.

Calculation of the Inbreeding Coefficient

a. Case of a Common Ancestor

An individual O receives an allele P' from its father P and an allele M' from its mother M. If this mother and father share a common ancestor A, the probability that these two alleles P' and M' are identical is determined as follows:



- The probability that the allele transmitted by the father comes from A is equal to $(½)^n$, where n is the number of generations between A and P.

- The probability that the allele transmitted by the mother comes from A is equal to $(½)^{n'}$, where n' is the number of generations between A and M.

- The probability that both alleles come from A is $(½)^{(n + n')}$.

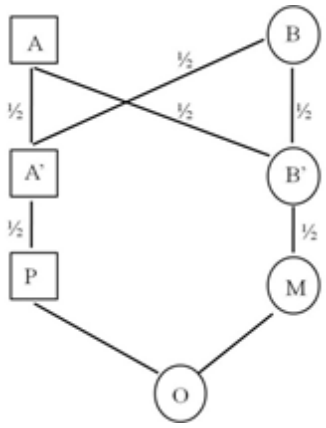
- In this case, with a probability of (½), the same allele from A was the one that gave both P' and M'. This gives a probability: $p1 = (½)^{(n + n') + 1}$.

If the ancestor A is inbred, we have a probability:

$$\mathbf{p_2 = \alpha_O = (½)^{n+n'+1} x (\alpha_A + 1)}$$

b. Case of Two Common Ancestors

Example of a marriage between two first cousins. The genealogy of such a marriage is shown in the figure below:



O must have received one of the copies she possesses from her father P. He has a one-in-two chance of having received it from A', who in turn has a one-in-two chance of having received it from A. Similarly, O must have received the other copy from her mother M, who has a one-in-two chance of having received it from B', who has a one-in-two chance of having received it from A. Therefore, the probability that both copies carried by O come from A is equal to $(½)^4$. Furthermore, there is a one-in-two chance that A passed on the same copy to both A' and B'. The probability that O possesses two identical copies of one allele or the other carried by A is: $½ \times (½)^4 = (½)^5$.

The same reasoning shows that there is also a $(½)^5$ chance that O possesses two identical copies of one of the ancestral alleles carried by B.

Thus, the calculation of the inbreeding coefficient involves adding the probabilities that the descendant possesses two identical copies from a common ancestor, in this case:

$$\boxed{\alpha_O=(½)^{n+m+1}{}_{(A)} + = (½)^{n+m+1}{}_{(B)}}$$

In general:

$$\boxed{\alpha=\Sigma(½)^{ni+mi+1}x\ (1+\alpha_i)}$$

i: common ancestor, ni: number of generations between ancestor i and the mother, mi: number of generations between ancestor i and the father, αi = inbreeding coefficient of ancestor i.

The formula (AA p^2 + Aa 2pq + aa q^2) used in the case of Hardy-Weinberg equilibrium no longer applies in the case of inbreeding. It is transformed into:

$$\mathbf{AA\ (p^2+\alpha pq) + Aa\ (2pq - 2\alpha pq) + aa\ (q^2 + \alpha pq)}$$

c. Inbreeding and Homozygosity

Inbreeding has significant consequences in the case of recessive traits. It increases the probability that the descendant will be homozygous and express the corresponding trait. In the case of inbreeding, the probability that the child expresses the defect is equal to (q^2 + αpq), whereas in the case of panmixia (Hardy-Weinberg equilibrium), it is equal to (q^2).

For example, consider the probability that the descendant of a marriage between first cousins will be phenylketonuric. In this case, (q = 1/159) and (α = 1/16), so:

q^2 + αpq= $(1/159)^2$ + [(1/16) x (158/159) x (1/159)]= **11/25000**

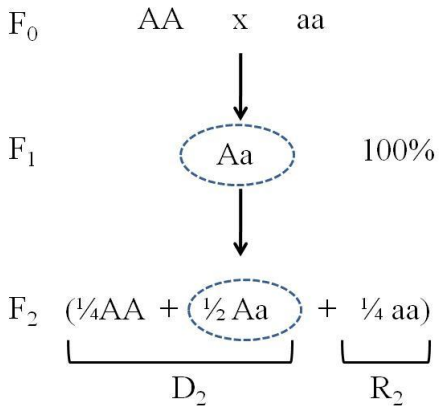
The probability is 11 times higher than what would be expected in a random marriage.

3.1.2. Homogamy

Couples may not form randomly, as we saw in panmictic populations. There may be selection based on phenotype. A crossing system is termed homogamic if there is a tendency towards resemblance (positive homogamy) or, conversely, towards dissimilarity (negative homogamy) between the partners of the same couple. Homogamy does not change allele frequencies but affects genotype frequencies. The homogamy practiced by breeders or farmers is phenotypic homogamy: when a morphological or physiological trait seems sufficiently interesting to be reproduced on a large scale, they cross individuals carrying that trait with each other. Some of these traits are controlled by dominant genes, which are therefore expressed in the heterozygous state. Others are due to recessive genes, which are only visible in the homozygous state.

Let's imagine a simple case of phenotypic homogamy. A population contains dominant alleles A and recessive alleles a, where mating occurs only between individuals with the [A] phenotype

(AA, Aa) or between individuals with the [a] phenotype (aa). Let's calculate the result of this type of cross using a panmictic F_2 population as the starting point.



In the F_2 generation, the R_2 individuals mate among themselves, resulting exclusively in (aa). The D_2 individuals mate among themselves. Evaluating the result of this second cross is straightforward by considering the allele frequencies: the frequencies of alleles A and a in the D_2 fraction of the population are 2/3 and 1/3, respectively.

$\underline{(¼AA + ½\ Aa)}$ which simplifies to AA + 2Aa=> AA + Aa + Aa
D_2

In the D_2 fraction, the frequency of allele A is fr(A) = 4 alleles A/6 alleles total, which is 2/3.

The frequency of allele a is 1/3.

The genotypes resulting from their crosses, given by the expression $[(2/3)A+(1/3)a]^2$, are found in the following proportions:

	A (2/3)	a(1/3)
A (2/3)	AA(4/9)	Aa (2/9)
a(1/3)	Aa (2/9)	aa(1/9)

Which simplifies to : (4/9)AA, (4/9) Aa et (1/9) aa

But D_2 represents only ¾ of the population, so the proportions relative to the entire F_2 population are adjusted to:

[(4/9)AA+(1/3)Aa+(1/9)aa] x (3/4) = (1/3)AA+(1/3)Aa+(1/12)aa

The F_3 is composed of:

(1/3)AA+(1/3)Aa+(1/12)aa	+	1/4aa	=	(1/3)AA + (1/3)Aa	+	(1/3)aa
from D_2		from R_2		D_3		R_3

An identical calculation can be made to calculate the F_4 from this F_3.

D_3 contains 3A alleles for every 1 a allele; it will therefore provide the following in F_4:

$[(3/4)A+(1/4)a]^2$= (9/16)AA + (6/16)Aa + (1/16)aa

And since D_3 represents 2/3 of F_3, these proportions adjusted to the total population become:

[(9/16)AA + (6/16)Aa + (1/16)aa] x (2/3)= (3/8) AA+ (¼)Aa + (1/24)aa

(3/8) AA+ (¼)Aa + (1/24)aa	+	(1/3) aa	=	3/8AA	+	1/4Aa	+3/8aa
from D_3		from R_3				

Examination of successive generations shows that the number of dominant alleles remains equal to the number of recessive alleles. The frequencies of the dominant and recessive homozygous genotypes also remain equal from one generation to the next.

$$\text{fr } (AA)_n = \text{fr } (aa)_n$$

The frequency of heterozygotes is easily obtained by recurrence:

In F_1, it is equal to 1;

In F_2, it is equal to ½;

In F_3, it is equal to 1/3;

In F_4, it is equal to ¼;

In F_n, it is equal to 1/n.

And we can write:

$$\mathbf{fr\ (Aa)_n = 1/n\ fr\ (AA)_n = fr(aa)_n = [1 - (1/n)] \times 1/2 = (n-1)/2n}$$

A reduction to 1/1000 of the frequency of heterozygotes therefore requires 1000 generations, whereas only 9 generations were needed in the case of absolute inbreeding. It is thus faster and easier to select an interesting variety by using inbreeding (self-fertilization in the ideal case) of individuals carrying the desired trait.

It is important to note that this breeding method has drawbacks due to the loss of genetic polymorphism it causes.

4. Populations in Disequilibrium

4.1. Mutations

Mutation is the fundamental cause of genetic variability. The term 'mutation' refers to any genetic change (inversion, translocation, polyploidy). The effect of mutation on the population structure differs depending on whether it is a rare event or an event that repeats from generation to generation at a certain rate. A rare mutation does not confer any particular advantage and cannot produce significant changes in a population, and is likely to disappear unless genetic drift (discussed later) causes it to persist at a high frequency purely by chance. Unlike the case of rare mutations, a repetitive mutation is regular as well as its frequency. The normal gene may disappear, and only the mutated gene will remain.

In general, mutation is reversible, meaning that it occurs in two directions, usually with different frequencies.

Let's denote 'μ' as the probability of mutation from the dominant allele 'A' to the recessive allele 'a', and 'v' as the probability of mutation from the recessive allele 'a' to the dominant allele 'A'.

$$A\ (p) \underset{v}{\overset{\mu}{\rightleftharpoons}} a\ (q)$$

If the frequency of allele 'A' is p in generation 'n', we can predict its frequency (p') in generation (n+1). The fraction of 'a' alleles mutated into 'A' (vq) will be added to p, while the fraction of 'A' alleles mutated into a through the reverse mutation (μp) will be subtracted.

$$\mathbf{p'=p+vq-\mu p=(1-\mu)p+vq}$$

In general, Δp represents the change in the frequency of allele 'A' in each generation.

$$\mathbf{\Delta p = p'-p=p+vq-\mu p-p=vq-\mu p}$$

$$\mathbf{\Delta p = vq-\mu p}$$

An equilibrium state will be reached when Δp = 0, meaning when the number of mutations from 'A' to 'a' exactly compensates for the number of mutations from 'a' to 'A'.

$$\mathbf{\Delta p = vq-\mu p \Rightarrow vq=\mu p \Rightarrow v(1-p)=\mu p \Rightarrow v-vp-\mu p=0 \Rightarrow v-p(v+\mu)=0}$$

$$\mathbf{\bar{p}=v/(\mu+v) \qquad \bar{q}=\mu/(\mu+v)}$$

In conclusion, the allele frequencies at equilibrium depend only on the mutation frequencies. Knowing the values of μ and v, it is very easy to calculate $\mathbf{\bar{p}}$ **et** $\mathbf{\bar{q}}$

Case where mutation is irreversible

Suppose an allele 'A' has a frequency of p_0 and μ is the mutation rate from 'A' to 'a'. After the mutation, the new frequency of 'A' is $p_1 = p_0(1 - \mu)$. Therefore, the frequency of 'A' will decrease from generation to generation.

By recurrence:

$p_1 = p_0 - \mu\, p_0 = p_0 (1-\mu)$

$p_2 = p_1 - \mu\, p_1 = p_1 (1-\mu) = p_0 (1-\mu) (1-\mu) = p_0 (1-\mu)^2$

After n generations: $p_n = p_0 (1- \mu)^n$.

Application exercise

How many generations will it take to reduce the frequency p = 0,9 to p_n=0,3 if $\mu=10^6$?

$p_n=p_0(1- \mu)^n => p_n/p_0 = (1- \mu)^n => \log(p_n/p_0) = \log (1- \mu)^n => \log p_n - \log p_0 = n \log (1- \mu) => n= (\log p_n - \log p_0)/\log (1- \mu)$

$n= (\log p_n - \log p_0)/\log (1- \mu)$=1098623 générations !!!

4.2. Selection

The Hardy-Weinberg law assumes that all genotypes are equivalent and participate equally in the formation of the next generation. In natural populations, different genotypes do not have the same viability or fertility. Natural selection refers to the phenomenon where different genotypes are not equally viable or fertile. As a result, certain genotypes are removed from the population, either completely or partially. Selection also occurs at the gamete level. In this case, it always leads, in the short or long term, to the elimination of the considered allele. Natural selection is the driving force of evolution. It is the process that leads to better adaptation of organisms to their somewhat different environments. If the conditions are different, natural selection will favor genetic divergence between populations. If the conditions are similar, natural selection will prevent populations from genetically diverging. Selection acts on phenotypes, not on genotypes. The adaptive value (or selective value) is defined as the ratio of the observed number in a generation to the theoretical number expected by applying the Hardy-Weinberg law.

Example: When 100 eggs of wild-type Drosophila give rise to 100 adults, the adaptive value of the considered phenotype is equal to 100/100 = 1. However, when 100 "Bar" eggs of the same species only produce 70 adults, the selective value of "Bar" is 0.70.

Several factors can influence this selective value: fertility, longevity, resistance to starvation, climatic factors, developmental speed, etc. The different components of the selective value do not have an absolute value. They vary with environmental conditions.

Let's consider a population containing alleles 'A' and 'a' with frequencies 'p' and 'q', respectively. In the absence of gametic selection, the different gametes would give rise to p^2 eggs (AA), 2pq eggs (Aa), and q^2 eggs (aa). However, not all of these eggs will develop into adults, as some will die due to their inability to withstand harsh environmental conditions. There will thus be selection of resistant genotypes; if we denote δ_1, δ_2, and δ_3 as the global selective values of each genotype:

The $\mathbf{p^2}$ eggs (AA) will produce $\mathbf{\delta_1 p^2}$ adults (AA).

The **2pq** eggs (Aa) will produce $\mathbf{2\delta_2 pq}$ adults (Aa).

The $\mathbf{q^2}$ eggs (aa) will produce $\mathbf{\delta_3 q^2}$ adults (aa).

The frequency of allele 'A' produced by these adults—assuming again that there is no gametic selection—will therefore be:

$$\mathbf{p_1 = (\delta_1 p^2 + \delta_2 pq)/(\delta_1 p^2 + 2\delta_2 pq + \delta_3 q^2) = p\ [(\delta_1 p + \delta_2 q)/(\delta_1 p^2 + 2\delta_2 pq + \delta_3 q^2)]}$$

with: $\delta_1 p^2 + 2\delta_2 pq + \delta_3 q^2 = w$= The average selective value of the population

the frequency of allele a will be:

$$\mathbf{q_1 = (\delta_3 q^2 + \delta_2 pq)/(\delta_1 p^2 + 2\delta_2 pq + \delta_3 q^2) = q\ [(\delta_2 p + \delta_3 q)/(\delta_1 p^2 + 2\delta_2 pq + \delta_3 q^2)]}$$

The change in frequency of allele **A** between two generations is therefore equal to:

$$\boxed{\mathbf{\Delta p = p_1 - p = [(\delta_1 p^2 + \delta_2 pq)/w] - p = pq\,[\,p(\delta 1 - \delta 2) + q(\delta 2 - \delta 3)]/w}}$$

In this formula for Δp, the population is in equilibrium if p = 0, p = 1, or if $p = (\delta_3 - \delta_2) / (\delta_1 - 2\delta_2 + \delta_3)$.

As the number of generations increases indefinitely, the population evolves towards one of these three equilibrium states.

- If δ_2 is less than or equal to the largest of the two values δ_1 and δ_3.

$p \rightarrow 0$, if $\boldsymbol{\delta_1 < \delta_3}$

$p \rightarrow 1$, if $\boldsymbol{\delta_1 > \delta_3}$

- If '$\delta_2 \geq$ the largest of the two values δ_1 and δ_3, then $p \rightarrow (\delta_3 - \delta_2) / (\delta_1 - 2\delta_2 + \delta_3)$.

The effect of selection is therefore either the elimination of polymorphism, if one of the homozygous genotypes is favored over the heterozygous genotype, or the maintenance of polymorphism, if the heterozygous genotype is the most favored.

Δp allows us to predict the evolution of allele frequencies based on the relative values of δ_1, δ_2, and δ_3. In the formula for Δp, the denominator is always positive. Every time p and q are different from 1 (i.e., the two alleles coexist in the population), the sign of Δp depends solely on the sign of the numerator M:
- If $M > 0 \rightarrow \Delta p > 0 \rightarrow$ the frequency of allele A increases.
- If $M < 0 \rightarrow \Delta p < 0 \rightarrow$ the frequency of allele A decreases.
- Finally, if M = 0, Δp = 0, and the population is in equilibrium.

We will now examine the three possible situations:

1. One homozygote is advantaged while the other is disadvantaged.
2. The heterozygote is advantaged.
3. The heterozygote is disadvantaged.

Case where one homozygote is advantaged and the other is disadvantaged: (The adaptive value of the heterozygote is intermediate).

This situation can be represented by the relationship:

- $\delta_1 \geq \delta_2 \geq \delta_3$ or $\delta_1 \leq \delta_2 \leq \delta_3$

If we plot the value of M as a function of p, we observe that this value is always positive. This means that Δp is always positive.

As a result, the frequency of allele A increases in each generation until it reaches 1. This leads to the elimination of allele a. **The population will lose its genetic diversity since it will consist only of homozygous AA individuals. In this case, we say that the population has fixed allele A. The only equilibrium state in this scenario is when p = 1.**

This situation is a classic example of **directional selection**, where one homozygous genotype is favored over the others, leading to the fixation of a beneficial allele.

Example: Resistance to DDT

The evolution of resistance to DDT (DichloroDiphénylTrichloroéthane) in insect populations illustrates this concept. Initially, insect populations were susceptible to DDT, and the allele conferring resistance (R) was either absent or at very low frequency. The introduction of DDT imposed strong selective pressure, favoring individuals carrying the resistance allele (RR or even Rr if partially dominant). Over generations, the frequency of the resistance allele increased rapidly because resistant individuals survived and reproduced, while susceptible individuals (rr) were eliminated. By 1957, resistance to DDT had spread worldwide, rendering the pesticide ineffective. Similar cases of pesticide and antibiotic resistance in various organisms follow the same pattern, demonstrating how selection reduces genetic diversity by eliminating less favorable alleles and fixing the advantageous ones.

Genotypes	(RR)	(Rs)	(ss)
Phenotypes	[R]	[R]	[s]

δ_1=1	[R]	(RR)	resistant
δ_2=1	[R]	(Rs)	resistant
δ_3=0	[s]	(ss)	sensitive

We will consider a special case of this situation where a dominant allele is genetically lethal. Genetic lethality means the total absence of descendants. It can result from the individual's death before reaching reproductive age or from their sterility. If lethality is due to a dominant allele, we have the relationship $\delta_1 \geq \delta_2 = \delta_3 = 0$. In this case, we observe that $\Delta p = q$. Indeed, no individual carrying the disadvantageous allele has descendants, and this allele is entirely eliminated in each generation. A situation similar to this case in humans is the condition known as retinoblastoma. This disorder is caused by a dominant allele, which leads to a retinal tumor that is typically fatal in early childhood. However, with surgery (such as eye removal), the child's life can be saved, and the transmission of the allele can occur.

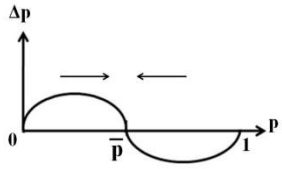
This selection, which favors extreme phenotypes, is called **directional selection**. It occurs when the environment gradually changes in a particular direction (e.g., from a hot bioclimate to a cold one). It shifts the population in one direction.

1. $\delta_1 \leq \delta_2 \leq \delta_3$: Recessive homozygotes are eliminated, and the population mean will gradually shift toward the dominant phenotype.

2. $\delta_1 \geq \delta_2 \geq \delta_3$: Dominant homozygote are eliminated, and the population mean will shift toward the recessive phenotype.

Case where the heterozygote is advantaged (the adaptive value of the heterozygotes is higher than that of the homozygotes, $\delta 3 < \delta 1 < \delta 2$). In this case, there is an equilibrium value $\bar{p}$ for which p is different from zero and different from 1. This equilibrium is stable.

If $p < \bar{p}$, $\Delta p > 0$ => p increases in each generation.
If $p > \bar{p}$, $\Delta p < 0$ => p decreases in each generation.



Regardless of the value of p, selection tends to modify it in the direction of the equilibrium value $\bar{p}$.

Sickle cell anemia (thalassemia) is a classic example of this situation. This disease is determined by a semi-dominant allele. There is a modification of the amino acids in the β-chain of hemoglobin.

Voici la traduction :

AA → normal (Hb^SHb^S)

aa → lethal anemia (Hb^sHb^s)

The genotype Aa (Hb^SHb^s) presents red blood cells that change shape when the oxygen pressure decreases. They take on a crescent, sickle shape. Studies of human populations have shown that the frequency of the unfavorable allele is very low in Northern European populations (homozygotes are selected against), but becomes relatively high in certain populations of sub-Saharan Africa, India, and the Mediterranean region. Furthermore, there is a strong correlation between the frequency of this allele and the geographical distribution of malaria. It has been shown that heterozygotes are less susceptible to malaria than normal homozygotes (AA). Thus, in regions where malaria is prevalent, heterozygotes resist the disease, while homozygotes (Hb^SHb^S) and (Hb^sHb^s) are not resistant. There is a selective advantage for heterozygotes, an advantage that ensures the preservation of the harmful allele Hb^s.

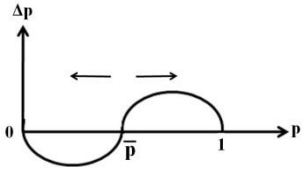
Among the African American population in the USA, the frequency of the allele responsible for sickle cell anemia is ten times lower (q = 0.04). Since malaria does not exist in this region, heterozygotes have lost their selective advantage. The selective advantage strongly depends on environmental conditions. Whether malaria is present or not, heterozygotes are either advantaged or slightly disadvantaged compared to normal individuals. This type of selection maintains the lethal allele in the population. It is **stabilizing selection**. It operates when the environment changes little or not at all. It eliminates poorly adapted mutants and opposes their spread. Its effect is to maintain a constant genetic population because it favors average or normal phenotypes and eliminates extreme individuals. It promotes **balanced polymorphism**.

Case where the heterozygote is disadvantaged ($\delta 3 > \delta 1 > \delta 2$).

In this case, a change in the sign of the numerator M, and therefore in Δp, is observed as a function of p. There is an equilibrium state where $\Delta p = 0$, and p and q are different from 1. This **equilibrium is unstable**.

$p < \bar{p} \Rightarrow \Delta p = p' - p < 0 \Rightarrow$ p decreases

$p > \bar{p} \Rightarrow \Delta p = p' - p > 0 \Rightarrow$ p increases



As soon as the value of p deviates from the equilibrium point, the selection pressure will increasingly drive it further away, leading to the fixation of one of the two alleles and a reduction in genetic variability.

This situation can be applied to the incompatibility systems of the ABO and Rhesus blood groups.

The Rhesus factor is determined by two alleles: (A), responsible for synthesizing a surface antigen on red blood cells (Rh^+), and (a), which represents the absence of this antigen. The 'A' allele is dominant, while the 'a' allele is recessive. The genotypes (AA) and (Aa) correspond to the [Rh^+] phenotype. Individuals with the [Rh^-] phenotype do not have antibodies against Rhesus antigens but can synthesize them upon exposure to [Rh^+] red blood cells.

Thus, a woman with the genotype (aa), meaning an [Rh^-] phenotype, who has a child with a man of genotype (Aa) (Rh^+ phenotype) may give birth to an (Aa) child with an [Rh^+] phenotype during the first pregnancy without complications. However, during childbirth, a small quantity of [Rh^+] antigens from the fetus can enter the mother's bloodstream, stimulating the production of antibodies. In a second pregnancy, these antibodies increase and can cause hemolytic disease of the newborn, often leading to fatal anemia. Today, such complications can be prevented.

This is a mechanism of selection against heterozygotes. However, no fixation of one of the two alleles has been observed. This type of selection is called **disruptive selection**. It splits a homogeneous population into several different adaptive types. It helps maintain variation within the population. The establishment and survival of these different types require a heterogeneous environment, meaning one composed of multiple ecological niches.

Natural selection is therefore a powerful evolutionary force. The effectiveness of selection depends on the environment, as it is the environment that determines the difference between appropriate and non-adapted genotypes. It is also the environment that directs selection toward the most suitable form:

~ It is **stabilizing** when the environment remains unchanged.

~ If the environment changes, it becomes **directional**.

~ Finally, if the environment offers a variety of ecological niches, it can become **disruptive**.

Selection Coefficient

This coefficient indicates the intensity of selection acting on a genotype. It is the inverse of the fitness or adaptive value and is expressed by the equation $S = 1 - \delta$. While the fitness value

represents the degree of advantage of a genotype, the selection coefficient expresses the degree of disadvantage.

If:

	AA	Aa	aa
δ	**1**	**0.85**	**0.60**
S	0	0.15	0.40

Let's examine various possible cases for a locus A/a.

First case: The 'A' allele is dominant, the 'a' allele is recessive, and selection is directed against the recessive allele.

	AA	Aa	aa
Adaptive value (δ)	**1**	**1**	**1**
Frequency	p^2	2pq	q^2
δ after selection	1	1	1-S
Fréquency after selection	p^2	2pq	q^2 (1-S)

After selection, the frequency q_1 is:

$q_1 = [q^2 (1-S) + pq] / [p^2 + 2pq + q^2 (1-S)] = [q^2 (1-S) + pq] / [1- Sq^2]$

The change in frequency $\Delta q = = q_1-q = [-Sq2 (1-q)] / [1-Sq^2]$

The change in frequency (Δq) depends on the selection coefficient (S) and the initial frequency of the allele 'a' (q). This fact is observed in all the cases examined.

Second case: Selection against the dominant allele

	AA	Aa	aa
Adaptive value (δ)	**1**	**1**	**1**
Frequency	p^2	2pq	q^2
δ after selection	1-S	1-S	1
Frequency after selection	$(1-S)p^2$	(1-S)2pq	q^2

Proceeding as before, we obtain:

$\mathbf{\Delta q} = q_1 - q = [Sp^2 (1-q)] / [(1-S) (1-p^2)]$

The effectiveness of selection is determined by Δq. If Δq is large, selection is effective. Δq depends on two parameters: the initial frequency of the a allele (q) and the value of the selection coefficient (S).

Application:

We will examine this effectiveness in three populations with two different values of the selection coefficient (S).

- Population I: q = 0.9
- Population II: q = 0.5
- Population III: q = 0.1

In the first case, S = 0.2, and in the second case, S = 0.8, with S applied to recessives in the case of dominance.

Δq = q_1-q=[-Sq2 (1-q)] / [1-Sq^2]

First case: S=0.2 pop I, **Δq= -0.0193**

Pop II, **Δq= -0.025**

Pop III, **Δq=-0.0018**

Second case: S=0.8 pop I, **Δq= -0.184**

Pop II, **Δq= -0.125**

Pop III, **Δq=-0.007**

Remarks:

> When S is low (first case), selection is more effective when the allele frequency is moderate. Δq is largest when q = 0.5.

> When the selection coefficient (S) is large, selection is very effective when the allele frequency (q) is high, but it quickly becomes much weaker. This type of selection is one that can be applied to a population that is being improved, for example, by removing recessives. However, very quickly, this operation will have only a very limited benefit.

Number of generations required to achieve a given frequency change:

Development for improving an animal or plant population. Suppose we apply a selection coefficient (S) to the recessives. Initially, the population contains $p^2 + 2pq + q^2$ genotypes, with q as the initial frequency of the 'a' allele. After complete selection against the recessives:

$q_1 = pq/ (p^2 + 2pq) = q/(1+q)$

$q_2= p_1q_1/ (p_1^2 + 2p_1q_1) = q_1/(1+q_1)= [q/(1+q)]/ 1+ [q/(1+q)]= q/(1+2q)$

By recurrence: q_n =q/(1+nq)

The same formula can be used to determine the number of generations required to go from a frequency q to a frequency q_n : **n= 1/q_n-1/q**

Example:

If the frequency of a recessive allele is 0.001 in a population, how many generations will be required to reach a frequency of 0.0005 if S = 1?

n = 1 / 0.0005 - 1 / 0.001 = 1000 generations

Note:

Since recessive alleles often occur at frequencies around 10^{-5}, it becomes practically impossible to completely eliminate a recessive allele from a population.

4.3. Combined Effect of Mutations and Selection

In most cases, a mutant allele is subject to a selection coefficient. As a result, the wild-type allele remains in the population. Since mutation and selection have opposite effects, it is evident that an equilibrium must be established.

We will consider three cases:

First case: A → a, with S applied to the recessives, where 'A' mutates to 'a' at a rate (μ) and 'a' mutates back to 'A' at a rate (v). When equilibrium is reached,

$$\mathbf{\Delta q = q_1 - q =} \mu p - vq = [-Sq2\ (1-q)\ ] / [1-Sq^2]=0$$

$$\Rightarrow \mathbf{q = \sqrt{(\mu/S)}}$$

Second case: A = a (codominance, S applied to homozygotes (aa)).

$$\mu p - vq= [(-1/2\ Sq)\ (1-q)\ ] /(1-Sq)$$

After approximation, we obtain: $q = \mu / S$

Third case: A > a, S applied to [A].

At equilibrium: $\mathbf{\mu p = vq = (Sq^2p)/}\ [\mathbf{1-S(1 - q^2)}\ ]$

$$\mathbf{\bar{p}=2v/S}$$

4.4. Migration

In a population, individuals may migrate. They have two possibilities: either they form a new population on their own or they integrate into another population. The influence of migration on allele frequencies depends on the rate of immigration and the differences in allele frequencies between the source population and the immigrants.

In the first case, migrants take a portion of the gene pool with them. As a result, the allele frequencies in this group are no longer identical to those of the original population, leading to

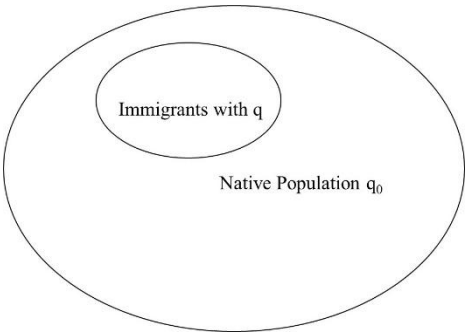
genetic drift (the unpredictable variation in frequencies due to small population size). Through successive events of genetic drift, the gene pool increasingly diverges, especially when mutation and selection act simultaneously. Over time, individuals from a population derived from the original population may develop such distinct genomes that they can no longer interbreed with individuals from the parent population. This leads to the formation of a new species—a speciation event.

Immigration Rate and Allele Frequencies

When migrants integrate into another population, the allele frequency of that population will change. This variation depends on the immigration rate and the difference between the allele frequencies of the immigrant population and those of the native (host) population.

Let's consider a population that receives a proportion **'m'** of new immigrants in each generation. In the native population, m = number of immigrants / total population size, meaning that the remaining proportion of the host population is 1 - m.

Let '**q**' be the frequency of a certain allele among the immigrants and $\mathbf{q_0}$ its frequency among the native population.



After immigration, the frequency of the considered allele in the new population is:

$$q_1 = mq + (1-m)\, q_0 = m\,(q-q_0) + q_0$$

$$\boxed{\Delta q = q_1 - q_0 = m\,(q - q_0)}$$

This immigration model corresponds to the colonization of an island. If the immigration rate remains constant from generation to generation, the island population loses its individuality and becomes comparable to that of the mainland.

By comparing the frequencies $q_1 ... q_n$ of the host population to that of the immigrants q, we have:

$$\mathbf{q_1 - q} = m(q - q_0) + q_0 - q = m(q - q_0) - (q - q_0)$$

$$= (q - q_0)(m - 1) = (q_0 - q)(1 - m)$$

$$q_2 - q = (q_1 - q)(1 - m) = (q_0 - q)(1 - m)(1 - m) = (q_0 - q)(1 - m)^2$$

$$\boxed{\mathbf{q_n - q = (q_0 - q)(1 - m)^n}}$$

When $q_n = q$, the island loses its individuality.

Application:

Among African Americans, the frequency of the R_0 allele of the Rhesus factor is 0.45. Among Africans, it is 0.63, and among Europeans, it is 0.03.

- Calculate the theoretical frequency of R_0 in African Americans, given that m = 0.03 over 10 generations.

- What is the proportion of African genes in African Americans?

- What is the proportion of genes from Europeans?

Let's use the following formula:

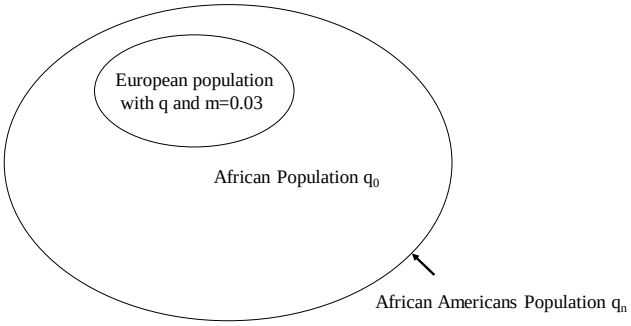
$\mathbf{q_n - q = (q_0 - q)(1 - m)^n}$, where**:**

- Africans → q_0 = frequency in the natives,

- African Americans → q_n

- Europeans → q = frequency in the immigrants. These are the European genes that have migrated into the Black population.

$\mathbf{q_n = (q_0 - q)(1 - m)^n + q} = (0.63 - 0.03)(1 - 0.03)^{10} + 0.03 = 0.47$



The proportion of genes of African origin in African Americans is:

$$\mathbf{(1-m)^{10}} = (0.93)^{10} = 0.73 = 73\%$$

The proportion of genes coming from Europeans is equal to the difference, which is 27%.

5. Speciation

5.1. Concepts and Definition of a Species

The species is the basic unit in the Linnaean hierarchy of taxonomy (the science that studies the norms and rules of systematics). Taxonomists are constantly confronted with the difficulties of identifying and naming species. In general, the issue is whether the category 'species' can be defined objectively. Historically, three general conceptions of species have successively prevailed in the scientific community:

- **The Typological Concept of Species**:

This concept dates back to the time of philosophers like Plato, Aristotle, and also to Linnaeus and his successors. The typological concept of species corresponds to a description made on a small number of individuals collected from the same locality. According to this concept, individuals of the same species are considered relatively uniform and conform to the described type.

According to this concept, the diversity observed in the universe reflects the existence of a limited number of 'universal phenomena' or types. Individuals of the same species do not have any particular relationship with one another. Thus, for the essentialist, morphological similarity is the criterion for defining a species. This is known as the "**morphological concept of species**." Morphological characteristics provide valid facts for determining species status. However, it is

completely different to use the degree of morphological differentiation as the essential criterion for determining the status of a species.

- The Nominalist Concept of Species

Nominalists deny the existence of "real" universal phenomena. For them, only individuals exist, while species are abstractions created by humans. According to them, species were invented to collectively refer to large numbers of individuals. This nominalist concept of species was popular in France in the eighteenth century and still has adherents today.

- The Biological Concept of Species

This concept emphasizes that:

~ The members of a species form a reproductive community. Individuals of a species behave towards other members of the same species as potential mates and seek one another for reproduction.

~ The species is also an ecological unit in which, regardless of the individuals that make it up, interacts as a whole with other species sharing its environment.

~ The species is a genetic unit consisting of an interconnected gene pool, whereas an individual is merely a temporary vessel carrying a small fraction of the gene pool's content for a short period of time.

The definition of species resulting from this concept is as follows: a species is a group of natural populations capable of interbreeding and reproductively isolated from other similar groups. A species is a protected gene pool. A species has its own mechanisms (isolation mechanisms) to protect itself from harmful gene flow coming from other gene pools.

5.2. Mechanisms of Reproductive Isolation

These are mechanisms that suppress or, at most, only allow very limited genetic exchange between the members of different species. Such barriers result in maintaining species integrity. These mechanisms (barriers) are of two types: internal barriers and external barriers.

- External Barriers: These correspond to mechanisms that prevent interspecific crossbreeding.

~ Geographic Isolation: The simple separation in space leads to the suppression of exchanges between species.

~ Habitat Isolation or Ecological Isolation: This arises from the fact that individuals belonging to two different species have distinct ecological preferences. A good example is that of two neighboring species of violet: *Viola arvensis* and *V. tricolor*, which can experimentally produce fertile hybrids but remain distinct in nature due to their ecological preferences; *V. arvensis* being calcicolous and *V. tricolor* being silicolous.

~ Seasonal Isolation: This occurs when the reproduction periods of two species do not coincide during the year. A classic example is that of two neighboring species of lettuce (*Lactuca canadensis* and *L. graminiflora*) that produce viable and fertile artificial hybrids but do not bloom at the same time of year. *L. graminiflora* has an early spring bloom, while the other blooms in summer.

~ Ethological Isolation: This refers to behavioral isolation (*ethos* = habit, custom). There is a restriction on random mating. It is very common among animals with developed mental faculties. It is based on the interaction between stimuli coming from both sexes' partners; these stimuli can be visual, auditory, olfactory, or tactile. An example is two species of frogs: *Microhyla olivacea* and *M. carolinensis*. Their mating calls are very distinct. There is practically no mechanism other than the vocalization that allows the separation of groups. Other stimuli, such as chemical signals, can also cause isolation mechanisms, with each signal being more or less specific.

~ Mechanical Isolation: This is very effective in plants. The stigmatic lobes have sizes and shapes that vary between species, creating barriers to pollination by insects. In some flowers, fertilization can fail because the pollen from a species with a short style cannot develop a pollen tube long enough in species with long styles, or because this growth is slowed or prevented.

- Internal Barriers: These are mechanisms that reduce the success of certain interspecific crossbreeding.

~ Prevention of Fertilization: This refers to the inviability of gametes from one species when they come into contact with the genital tract or style of another species. In many *Drosophila* species, the sperm of a foreign species induces an antigenic reaction in the genital

tract of the receiving female. In polyploids, pollen tubes are of large diameter and have difficulty traveling through the style of diploid species.

~ Zygote Mortality: In this case, fertilization occurs. In the genus *Rana* (frogs), hybrid development stops at different critical stages during embryonic development depending on the cross. In the genus *Datura* (Solanaceae), the embryo stops developing once it reaches the 8-cell stage. The main causes of hybrid mortality are varied. In some cases, the endosperm acts as an inhibitor of embryonic development. There may be a lethal gene that has no effect on the individuals of the species but causes the death of the hybrid. The embryo never reaches maturity due to a genetic disharmony between the parental genomes. Hybrid inviability results from the interaction of certain hybrid genotypes with the cytoplasm of one of the parental species.

~ Sterility of Hybrids: The most famous example of hybrid sterility in the first generation is the mule. This animal is particularly hardy but sterile. Hybrid sterility is a general phenomenon. There can also be a barrier to genetic exchanges in subsequent generations (e.g., in the genus *Gossypium*). There are two types of sterility: **genic** and **chromosomal**. Genic sterility is characterized by the failure of reproductive organs to undergo meiosis or the presence of meiotic abnormalities. Chromosomal sterility results from the non-homology between the maternal and paternal chromosomes of the hybrid.

5.2. Mechanisms of speciation

There are multiple mechanisms that can prevent gene exchange between certain individuals of the same species, leading to the splitting of a genetically balanced pool (the species) into two groups that are unable to exchange genetic material, effectively becoming two different species. Speciation, therefore, is the division of a phylogenetic lineage. It is a process of species multiplication. According to Mayer (1942), several modes of speciation are possible, including phyletic evolution (geological scale), fusion, instantaneous multiplication (via chromosomal restructuring, polyploidy), and gradual multiplication (sympatric, allopatric), where populations diverge progressively until they become distinct species.

- **Allopatric or geographical speciation**: According to Mayer (1942), a new species forms when a population is geographically isolated from other populations of its parent species and, during the isolation period, acquires traits that provide or ensure reproductive isolation (marginal isolates, colonization, differentiation of geographical races).

- **Sympatric speciation**: According to this theory, two species can become distinct from one another without going through a stage of geographical isolation. The division of the gene pool is caused by ecological factors. A strong specificity regarding food sources would significantly increase the effectiveness of reproductive isolation mechanisms. One example is the case of the apple maggot fly (*Rhagoletis pomonella*). Initially fixed on hawthorn and possibly a few other rosaceous plants, these insects later invaded orchards. Over the next 100 years, a shift in the sexual maturation period of apple maggot populations compared to hawthorn populations was observed.

Bibliography

1. **Beaudry, J.R. 1985.** Génétique générale [texte imprimé] / J.R. Beaudry. - Paris : Maloine, 1985. - 501 p. : fig. ; 23 cm. Index p. 479-497 Bibliogr. p. 499-501. - ISBN 2224011113. **576.5/28, 576.5/28.2**
2. **Binder, Eugène, 1972.** La génétique des populations [texte imprimé] / Eugène Binder. - 3e éd. - Paris : PUF., 1972. - 126 p. : fig., graph. ; 18 cm. - (Que sais-je ? / Anne-Laure Angoulvent). Index p. 125 Bibliogr. p. 126. **576.5/21**
3. **Futuyma D. J. 1998.** Evolutionary biology. 3d Ed. Sinauer Associates, Inc. Massachussetts. 763p
4. **Griffiths, Anthony J.F., 2001.** Analyse génétique moderne [texte imprimé] / Anthony J.F Griffiths; Ed. Aragnol Denise Charmot Dominique. - Paris : De Boeck, 2001. - 676 p. : fig. ; 28 cm. Bibliogr. Index. Glossaire. - ISBN 2744501115. **576.5/55, 576.5/55.2**
5. **Jacquard, Albert, 1977.** Concepts en génétique des populations [texte imprimé] / Albert Jacquard. - Paris : Masson, 1977. - 230 p. : couv.ill.en coul ; 23 cm. - (Biologie évolutive ; 4)). Index p.127.128. - ISBN 2225471576. **576.5/61, 576.5/61..3**
6. **Lucotte, Gérard, 1983.** Génétique des populations [texte imprimé] / Gérard Lucotte, Préf. Jacques Ruffié. - Paris : nter-ed., 1983. - 200 p. : fig. en coul. ; 25 cm. bibliogr.p197-198 index p.199-200. - ISBN 2729600558. **576.5/62, 576.5/62.2**
7. **Petit, Claudine,** Evolution génétique des populations et évolution moléculaire [texte imprimé] / Claudine Petit, Emile Zuckerkandi. - Paris : Hermann, s.d. - 277 p. : fig., tabl. ; 22 cm. - (Méthodes). bibliogr.p.272-273, index p.274.278. - ISBN 2705658297. **576.5/60, 576.5/60.10**
8. **Felsenstein, Joseph 2005**. Theoretical Evolutionary Genetics. University of Washington, Seattle, Washington. Ebook.
9. **Génermont, J. 1979.** Les mécanismes de l'évolution. Bordas, Paris. 232p. **8441/02**
10. **Génermont, J.** Recueil d'exercices de génétique des populations [texte imprimé] : Avec rappels des définitions et notions fondamentales / J. Genermont, Préf. G. Tessier. - [s.l] : [s.n], [s.d]. – 141 p. : fig. ; 24 cm. **576.5/59, 576.5/59.2**
11. **Hartl, Daniel L. 1987.** Génétique des populations. (Traduction de Nicomas Borot). Ed. Médecine-Sciences, Flammarion, Paris. 306p.