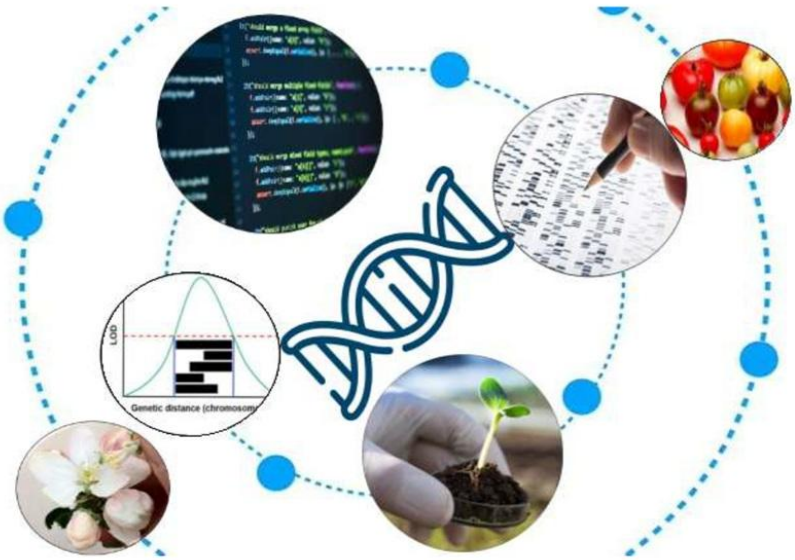
Quantitative genetics

Master 1 : Molecular Genetic and Plants Improvement.



DR. FATIMA DAHLIA
DEPARTMENT OF BIOLOGY
NATURAL AND LIFE SCIENCE FACULTY
IBN KHALDOUN UNIVERSITY OF TIARET
EMAIL: FATIMA.DAHLIA@UNIV-TIARET.DZ

May 2025

Table of content

Objectifs 6

Introduction 7

I. Prerequisites 8

1. Prerequisites test ... 8

II. Chapter 1 : Definitions and concepts 10

1. Differences between quantitative characters and qualitative characters ... 10
2. Specificities of quantitative traits. ... 11
3. Review of some statistical concepts ... 12
   - *3.1. Position statistics* ... 12
   - *3.2. Dispersion Statistics* ... 14
   - *3.3. Samples and population* ... 21
   - *3.4. Genotypes and phénotypes distribution* ... 22
4. Exercice ... 23
5. Exercice ... 23
6. Exercice ... 24

III. Chapter 2: Polygenic inheritance 25

1. General information ... 25
2. Relationships between genes ... 26
   - *2.1. Average effects (additivity)* ... 26
   - *2.2. Effect of dominance* ... 28
   - *2.3. Epistasis* ... 30
3. Exercice ... 31
4. Exercice ... 31
5. Exercice ... 31
6. Exercice ... 32
7. Exercice ... 32

IV. Chapter 3: Heritability 33

1. Kinship and Heritability ... 33
2. How to Quantify Heritability? ... 33
   - *2.1. Using Variances* ... 34
   - *2.2. Another method for estimating the heritability value* ... 38

3. The use of h^2 in breeding and improvement 38
4. Exercice 39
5. Exercice 39
6. Exercice 39
7. Exercice 40

V. Chapter 4: Heterozygosity 41

1. Heterozygosity and Allele Frequency 41
   1.1. Calculating Heterozygosity from Allele Frequencies 41
   1.2. Relationship between Heterozygosity and Genetic Homeostasis 42
2. Hardy-Weinberg laws 42
   2.1. Hardy-Weinberg equilibrium conditions 42
3. Evolution of heterozygosity over generations 43
4. Factors influencing heterozygosity 44
5. Exercice 45
6. Exercice 45
7. Exercice 45
8. Exercice 45

VI. Chapter 5: Heterosis 46

1. Hypotheses or Mechanisms 46
   1.1. Dominance 46
   1.2. Overdominance 46
2. Evolution of Heterosis Over Generations 47
   2.1. Mean of a Population 47
   2.2. Proper Evolution of Heterosis 49
3. Exercice 51
4. Exercice 51
5. Exercice 52
6. Exercice 52

VII. Chapter 6: The Value of an Individual in Crossing 53

1. Genetic Balances 53
   1.1. Internal Balance and Relational Balance 53
   1.2. Biological Reality of Balance and Representation by the Set of Genotypes 54
2. Combining Abilities 54
   2.1. Definitions 54
   2.2. Importance of Variances Related to Combining Abilities 55
3. Methods for Assessing the Values of an Individual in Crossing 55
   3.1. Open Pollination (Maternal Progeny) 56
   3.2. The top-cross test 56
   3.3. The poly-cross test 57

3.4. Hierarchical cross **57**
3.5. Diallel cross **58**

4. Exercice **60**

5. Exercice **60**

6. Exercice **60**

VIII. Chapter 7: Quantitative trait loci (QTL) 62

1. Applications of QTLs in agricultural, medical, and ecological research **62**
1.1. Agricultural Research **62**
1.2. Medical Research **62**
1.3. Ecological Research **63**
1.4. Limitations and challenges of using QTLs **63**

2. Techniques for QTL detection **63**
2.1. Linkage analysis and genetic mapping **64**
2.2. Association analysis and genome-wide association study **65**
2.3. Use of molecular markers for QTL mapping **66**
2.4. High-throughput genotyping techniques for QTL mapping **67**

3. Statistical analysis of QTLs **70**
3.1. Regression and analysis of variance models **70**
3.2. Linear Mixed Models for QTL Detection **70**
3.3. Correction for Population and Environmental Effects **71**
3.4. Assessing the Statistical Significance of QTLs **71**

4. Applications of QTLs in agricultural researchApplications of QTLs in agricultural research **72**
4.1. Marker-assisted selection for the selection of complex traits **72**
4.2. QTL mapping for resistance to diseases and pests **73**
4.3. QTL Mapping for Crop and Animal Quality **74**
4.4. Improving the Nutritional Quality of Food **74**

5. Applications of QTLs in medicine and ecology **74**
5.1. QTL mapping for human diseases **75**
5.2. Identification of QTLs involved in disease resistance in wild and domestic animals **75**
5.3. Identification of QTLs involved in adaptation to environmental changes **76**

6. Advanced Techniques in Quantitative Genetics **76**
6.1. Genetic Interaction Analysis **76**
6.2. Bayesian Regression Models **77**
6.3. Advanced Marker-Assisted Selection Methods **77**

7. Future developments in the field of QTLs **77**
7.1. Evolution of high-throughput sequencing technologies **78**
7.2. New approaches for high-throughput data analysis **78**
7.3. Prospects for the application of QTLs in marker-assisted selection and synthetic biology **78**

8. Exercice *79*
9. Exercice *79*
10. Exercice *79*
11. Exercice *80*

IX. References *81*

1. References *81*

Objectifs

Quantitative genetics is a discipline greatly involved in the field of plant breeding. Teaching the subject will allow students to distinguish quantitative traits and to know the different types of heritability and their use in plant breeding. Knowledge of the basics of quantitative genetics is also essential for the evaluation of an individual in crosses.

Introduction

Quantitative genetics is a branch of genetics that studies the variation of complex characteristics in individuals and populations, such as height, weight, growth, milk production, oil content in plants, etc. These characteristics are also called quantitative traits, because they vary continuously rather than falling into distinct categories.

Quantitative genetics studies how quantitative traits are inherited and influenced by the environment. Quantitative characteristics are often polygenic, meaning that they are influenced by multiple genes. In addition, quantitative characteristics can also be affected by the environment, such as diet, temperature, light, stress, etc.

The methods of quantitative genetics make it possible to measure the heritability of quantitative traits, that is, the proportion of the observed phenotypic variance that is due to genetic differences between individuals. Quantitative genetics methods can also be used to identify regions of the genome that are associated with variations in quantitative traits, called QTLs (Quantitative Trait Loci). The characterization of QTLs allows a better understanding of the genetic mechanisms underlying the variation in quantitative traits.

Its history dates back to the beginning of the 20th century, when scientists such as Ronald Fisher, J.B.S. Haldane, and Sewall Wright began to develop statistical methods to analyze data on complex characteristics. They proposed that quantitative traits were governed by many genes with tiny effects, and that these genes were inherited in a polygenic manner.

Over time, more advanced statistical methods have been developed to study quantitative traits, including linear regression and analysis of variance (ANOVA). These methods have led to a better understanding of how quantitative traits were inherited and how they were influenced by the environment.

In the 1980s and 1990s, the use of genetic markers made it possible to identify regions of the genome that were associated with variations in quantitative traits, called QTLs (Quantitative Trait Loci). QTLs have provided a better understanding of the genetic mechanisms underlying the variation in

quantitative traits and have paved the way for marker-assisted selection (MAS), a method of selecting individuals for specific characteristics using genetic markers.

With the advent of high-throughput DNA sequencing technologies in the early 21st century, quantitative genetics has taken another leap forward. These technologies have made it possible to identify millions of genetic variations in the genome, and to discover new genes and new regions of the genome that influence quantitative traits.

Today, quantitative genetics is used in many fields of biology, including agriculture, animal husbandry, medicine, and ecology. Recent advances in the field of quantitative genetics have opened new perspectives for understanding the complexity of quantitative traits and for their use in marker-assisted selection and synthetic biology.

I. Prerequisites

Introduction

A solid grounding in Mendelian genetics and statistics, as well as an understanding of fundamental concepts in population genetics. Prior knowledge of quantitative genetics is also useful, but not strictly necessary.

Prerequisites test

What is a gene?

In a diploid cell, a gene has:

○	One allele
○	Two alleles
○	Three alleles

An Allele could be:

☐	Dominant or recessive
☐	allelopathic
☐	Codominant
☐	Lethal
☐	Toxic

A quantitative trait is:

○	Monogenic, with discontinuous variation
○	Polygenic, with continuous variation.

Dahlia Fatima

A qualitative trait is:

○	Monogenic, with discontinuous variation
○	Polygenic, with continuous variation.

The interaction within alleles includes

☐	Dominance
☐	Additivity
☐	Epistasis
☐	Environment

The phenotype is the result of the sum of:

☐	Genotype + Dominance + Epistasis
☐	Genotype + Environment
☐	Environment + Additivity

II. Chapter 1 : Definitions and concepts

Introduction

Quantitative genetics studies differences between individuals. These differences are quantitative and not qualitative.

Any organ or any species shows possible differences. Ex: the difference in size, resistance to diseases, or yield. These individual differences can either form continuous and increasing series from one extreme to another or form distinct categories. In the first case, these are quantitative characters, and in the second case, these are qualitative characters.

1. Differences between quantitative characters and qualitative characters

Qualitative traits and quantitative traits are two types of characteristics that can be inherited and passed down from generation to generation. Here are the main differences between these two types of traits:

- **Nature of the characteristics**: Qualitative traits are characteristics that are easily observable and can be classified into distinct categories, such as eye colour, blood type, the presence or absence of a disease, etc. Quantitative traits, on the other hand, are characteristics that vary continuously and can be measured with numerical units, such as height, weight, productivity, etc.
- **Number of genes involved**: Qualitative traits are often governed by a small number of genes, often with a dominant or recessive effect, while quantitative traits are generally influenced by many genes, each having a relatively small effect.
- **Environment:** Qualitative traits are often little influenced by the environment, while quantitative traits are often strongly influenced by the environment. For example, eye colour

is generally determined by genes, while height can be influenced by many environmental factors such as nutrition, physical exercise, etc.

- **Heritability**: Qualitative traits often have high heritability, which means that a large part of the variation in these traits is due to genetic differences between individuals. In contrast, quantitative traits often have lower heritability because they are strongly influenced by the environment.
- **Statistical analyses**: for qualitative traits, we can only do an analysis of proportions; on the other hand, for quantitative traits (measurable), we can do all series of statistical analyses (means, variances, covariances, etc.).

The following table summarizes the main differences between qualitative and quantitative characters.

Quantitative traits	Qualitative traits
Continuous variability of most of traits. **Ex** : Yield, growth, height …etc.	Discontinuons variability **Ex** : Leaves colour, flowers colour …etc.
Polygenic determinisme	monogenic or oligogenic
Analysis of a series of statistical descriptors: means, variance, standard deviation Analysis of proportions	Analysis of proportions

Table 1: *Main differences between qualitative and quantitative characters.*

2. Specificities of quantitative traits.

Because of the established experimental techniques, geneticists are led to answer questions about the genetic determinism of a continuously varying character in a population. These questions, which are at the base of the study of quantitative genetics, are:

- Is the observed variation of the character in one way or another influenced by genetic variation? Is the segregation of alleles in the population responsible for the differences in the character, or is all this variation only the result of environmental variation and random fluctuation during development?
- If genetic variation exists, what are the reaction norms of the different genotypes?

- What is the importance of genetic variation in the total phenotypic variation? Are the reaction norms and the effects of the environment such that almost all the variation results from environmental differences and disturbances in development, or is genetic variation predominant?
- Do many loci or only a few affect the character? How are they distributed in the genome?
- How do the different loci interact to influence the character? Are we dealing with cases of dominance or epistasis?
- Are we in the presence of non-nuclear inheritance (maternal effect)?

In experimental organisms, it is relatively simple to establish whether a genetic influence is exerted, but the localization of these genes, even approximate, is extremely laborious.

3. Review of some statistical concepts

Statistical concepts are classified into :

- Position statistics.
- Dispersion statistics

3.1. Position statistics

position statistics gather :

- Mean.
- Median.
- Mode

a) Mean

The mean is a measure of central tendency. It represents the "typical" or "central" value of the data and is often used to describe the average performance of individuals in a population or sample. The mean is often represented by the Greek letter mu (μ) and can be calculated for continuous or discrete variables. However, it is important to note that the mean can be influenced by extreme or outlier values, which can impact its interpretation.

The calculation of the mean differs depending on whether the proportions are equal or unequal.

- *Equal proportions*: If the proportions are equal, the mean can be calculated by adding all the values of a variable and dividing the result by the total number of observations ($\mu = \Sigma xi/n$).
- *Unequal proportions*: If the proportions are unequal, the mean must be calculated by weighting each observation by its respective proportion. To do this, we multiply each observation value by its proportion, add the products, and divide the result by the sum of the proportions ($\mu = [\Sigma(xi*pi)/\Sigma pi]$. Where xi are the variables and pi the frequencies of the variables. The sum of the frequencies is equal to 1. Therefore $\mu = \Sigma(xi*pi)$.

Example 1:

Let's take the example of a study conducted in a given region to measure the average yield of wheat crops. Data were collected on wheat yields for 10 different farms, with the following results (in tons per hectare):

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Yield	6	5	5,5	7	5,5	6,5	6	6,5	7	7,5

Table 2: Average wheat crop yield of ten farms.

Mean = $\Sigma xi/n$

= (6 + 5 + 5.5 + 7 + 5.5 + 6.5 + 6 + 6.5 + 7 + 7.5) / 10

= 6.3

Example2:

Genotypes	**aa**	**Aa**	**AA**
Value (cm)	10	18	20
Frequencies	0.36	0.48	0.16

Table 3: Average sizes and genotype frequencies.

Mean = $\Sigma(xi*pi)/\Sigma pi$

= [10(0.36) + 18(0.48) + 20(0.16)] / 1

= 15.44 cm

b) Median

The median is a measure of central tendency that represents the central value of a dataset sorted in ascending or descending order. The median divides the dataset into two equal parts so that half of the

observations are less than or equal to the median and the other half are greater than or equal to the median.

The median is less sensitive to outliers than the mean. It is often used to describe the central performance of a sample or population and can be used to compare groups with different distributions.

The median is often denoted by the letter M or Md and can be calculated for continuous or discrete variables. However, it is important to note that the median may not be representative of the data distribution in some cases, such as when the distribution is skewed or highly asymmetrical.

To calculate the median, the data must be sorted in ascending or descending order.

- If the dataset has an odd number of observations, the median is the value of the central observation.
- If the dataset has an even number of observations. the median is the average of the two central observations. If the dataset has an even number of observations.

Example:

- If we have the following data: 2. 5. 8. 10. 15. the median is 8.
- If we have the following data: 3. 6. 7. 11. the median is (6 + 7) / 2 = 6.5.

c) Mode

The mode is a measure of central tendency that represents the most frequent value of a dataset. It is the value that appears most often in a dataset.

The mode is the value that appears most often in the dataset. There can be a single mode (one value that appears most often) or multiple modes (several values that appear equally often).

The mode is a measure of central tendency that can be used to describe the central performance of a sample or population. However, the mode is not always a useful measure of central tendency, as it may not be representative of the data distribution in some cases, such as when the data is uniformly distributed or when the data is highly skewed.

The mode is often denoted by the letter Mo or Mod and can be calculated for continuous or discrete variables.

Example: Suppose we have collected height measurements of wheat plants in a plot of land. The data are as follows (in cm):

50; 45; 52; 48; 54; 60; 42; 44; 47; 53; 48.

All values appear only once except the value 48; which appears twice; making it the most frequent value in the dataset. Thus; the mode of the height of wheat plants in this plot is 48 cm. This means that the height of 48 cm is the most frequent value in the dataset.

Note: It is important to note that the mean; median; and mode are different measures of central tendency and that they can be used in conjunction to gain a better understanding of the data distribution.

3 .2. Dispersion Statistics

Dispersion Statistics include:

- Variance and standard deviation.
- Error standard.
- Coefficient of variation.
- Distribution model.
- Correlations.
- Regression.

a) Variance and standard deviation

Variance and standard deviation are measures of dispersion or variation in data.

They allow us to measure how far the data points are from the mean.

Variance is a measure of the dispersion of data around the mean. It is calculated by summing the squares of the deviations between each value and the mean, then dividing this sum by the total number of values in the sample minus one. Variance is often denoted by the symbol σ^2.

$$\sigma^2 = 1/(n-1)[\Sigma(xi - \mu)^2].$$

Where:

σ^2: variance;

$\Sigma(xi - \mu)^2$: the sum of squared deviations;

xi: the value of each observation;

μ: the mean;

n: the total number of observations.

There is another formula for calculating the variance for datasets with a large number of observations ($n > 30$):

$$\sigma^2 = 1/(n-1)[\Sigma xi^2 - (\Sigma xi)^2/n]$$

σ^2: variance;

Σxi: the sum of all observations

Σxi^2: the sum of the squares of all observations

n: the total number of observations.

The standard deviation is the square root of the variance. It measures the dispersion of data around the mean in terms of the original unit of measurement. The standard deviation is often denoted by the symbol σ.

$$\sigma = \sqrt{\sigma^2}$$

b) Standard error

The standard error is a measure of the variability of sample data relative to the sample mean. It is also known as the standard deviation of the sample and is often used to estimate the accuracy of a sample mean relative to the true population mean.

Specifically, the standard error is calculated by dividing the sample standard deviation by the square root of the sample size. It represents the expected variation between the means of all possible samples from a given population of the same size.

$$\sigma\mu = \sigma/n$$

The standard error is commonly used in statistics to calculate confidence intervals and to determine whether differences between the means of two groups are significant. It is also used in regression analysis to assess the accuracy of regression coefficients.

Example: Suppose we measured the height of 20 wheat plants and wanted to estimate the standard error of the mean height for this population of plants. Here is the data:

50; 52; 53; 51; 49; 52; 48; 50; 52; 51; 50; 53; 54; 50; 52; 51; 49; 50; 49; 53

To calculate the standard error of the mean, we need the following formula:

$\sigma\mu = \sigma/\sqrt{n}$

To calculate the sample standard deviation, we can use the following formula:

$\sigma = \sqrt{\sigma^2}$

$\sigma^2 = 1/(n-1)[\Sigma(xi - \mu)^2]$

Where Xi is the is the data point, μ is the mean of the data, and n is the sample size.

First, let's calculate the mean of this data:

$\mu = (50+52+53+51+49+52+48+50+52+51+50+53+54+50+52+51+49+50+49+53)/20 = 51.1$

Now, we can calculate the sample standard deviation:

$\sigma^2 = 1/(20-1)* [(50 – 51.1)^2 + (52 – 51.1)^2 + ... + (53 – 51.1)^2] = 2.56$ $\sigma^2 = \sqrt{2.56} = 1.6$

Finally, we can calculate the standard error of the mean:

$\sigma\mu = \sigma/\sqrt{n} = 1.60/\sqrt{20} = 0.36$

The mean can therefore be written as: 51.1±0.36

c) Coefficient of variation

The coefficient of variation is a measure of the relative dispersion of a dataset. It is defined as the ratio of the standard deviation to the mean of the dataset, multiplied by 100 to obtain a percentage:

Coefficient of variation (%) =(standard deviation/mean)x 100

The coefficient of variation is used to compare the relative dispersion of two datasets with different means.

- The higher the coefficient of variation, the greater the relative dispersion of the dataset. Therefore, a high coefficient of variation may indicate that the dataset is very heterogeneous or very dispersed
- On the other hand, a low coefficient of variation indicates a low relative dispersion of the dataset, which may indicate that the data are more homogeneous or more similar to each other.

Example: Suppose we measured the nitrogen content of different soil samples from different plots and obtained the following results:

Plot 1: 0.28%

Plot 2: 0.32%

Plot 3: 0.35%

Plot 4: 0.31%

Plot 5: 0.33%

To calculate the coefficient of variation (CV) of this data, we need the following formula: $CV = (\sigma/\mu) \times 100\%$

$\mu = (0.28 + 0.32 + 0.35 + 0.31 + 0.33)/5 = 0.318$

$\boldsymbol{\sigma = \sqrt{\sigma^2}}$

Dahlia Fatima

$\sigma^2 = [1/(n-1)*(\sum(Xi - \mu)^2)]$

$\sigma^2 = ¼ [(0.28 - 0.318)^2 + (0.32 - 0.318)^2 + (0.35 - 0.318)^2 + (0.31 - 0.318)^2 + (0.33 - 0.318)^2]$

$\sigma^2 = 0.000729$

$\sigma = \sqrt{(\sigma^2)} = \sqrt{0.000729} = 0.027$

$CV = (\sigma/\mu) \times 100\%$

$CV = (0.027/0.318) \times 100\%$

$CV = 8.49\%$

The coefficient of variation of the nitrogen content for these soil samples is therefore 8.49%. This value gives us an idea of the relative variation in nitrogen content between the different soil plots, which can be useful for assessing the stability or uniformity of soil conditions in a given area.

d) Distribution model

The normal distribution is used to represent the variation of quantitative characters in a population. It is characterized by a symmetrical bell-shaped curve (Fig. 1), with the mean (central value) representing the central tendency of the population and the standard deviation representing the dispersion or variability of values around the mean. Two distributions with the same mean can differ significantly in how the measurements are grouped around the mean.

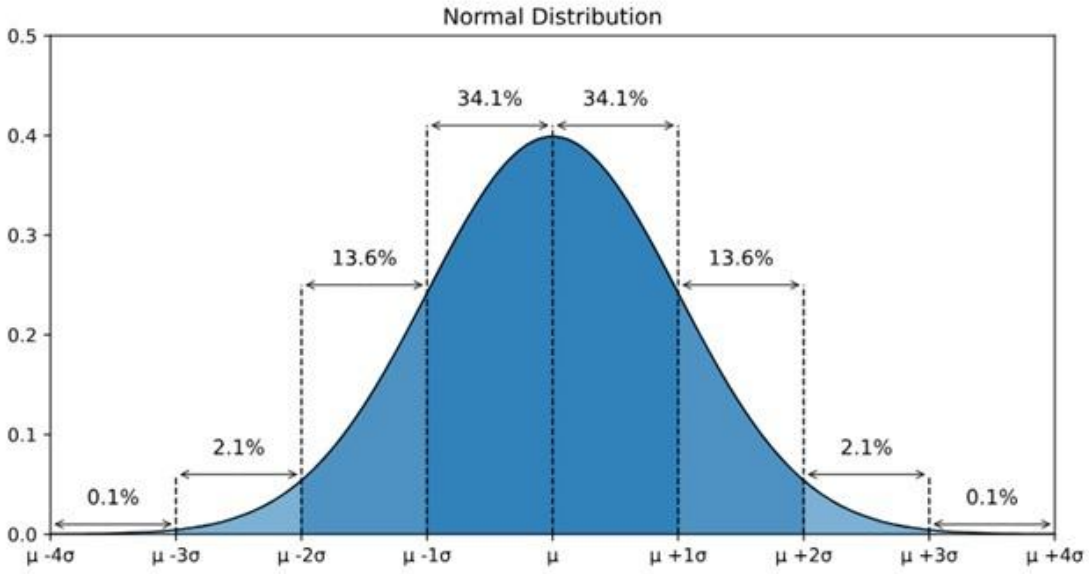


Figure 1: Normal distribution of continuous quantitative variables.

The normal distribution in quantitative genetics is based on the assumption that the effects of the many genes contributing to a given trait are independent and combine additively. According to this hypothesis, the observed variation in a trait is the result of the sum of genetic and environmental effects and therefore follows a normal distribution.

Exceptions to this configuration can be encountered. Asymmetrical (Fig. 2A) or bimodal (Fig. 2B) distributions can be encountered. An asymmetrical distribution may indicate the presence of outliers (values much higher or much lower than the other variables). A bimodal distribution may indicate that the population studied could actually be considered as a mixture of two populations, each with its own mode (for example, measuring the height of a population of male and female students).

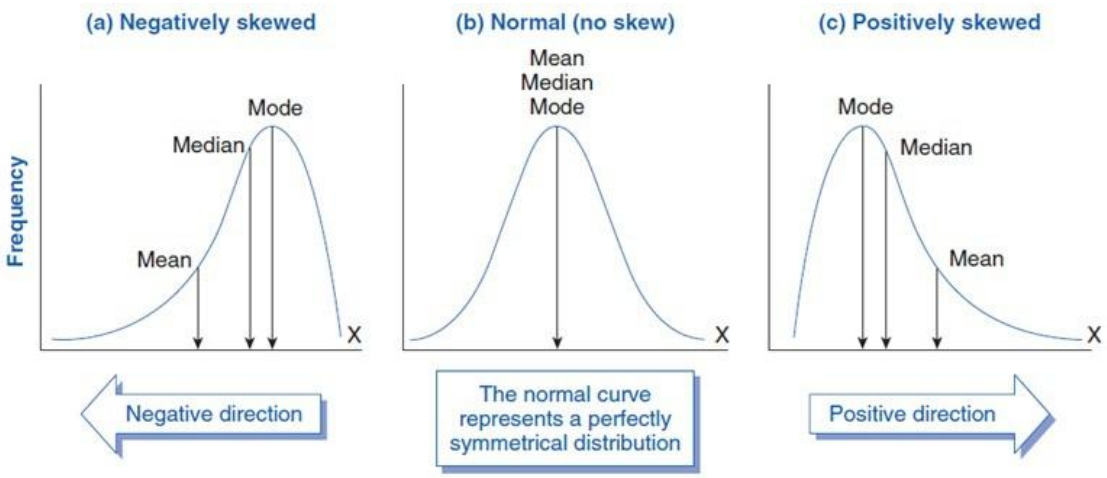


Figure 2A : Asymmetric distribution

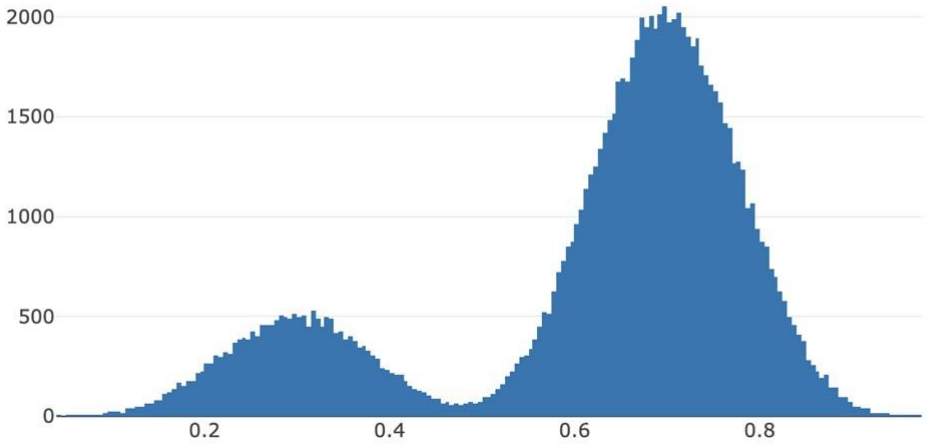


Figure 2B : Bimodal distribution

e) Correlation

Another statistical concept used in quantitative genetics is the correlation between variables. There can be very close, imprecise, or non existent relationships.

The correlation between quantitative characters, also called phenotypic correlation, measures the statistical relationship between two quantitative traits or characters in a population (Fig. 3). It allows us to determine the extent to which variations in one trait are associated with variations in another trait and whether these variations are positive (traits increase or decrease together) or negative (traits vary inversely to each other).

The correlation between quantitative characters is generally expressed by the correlation coefficient (r), which varies from -1 to +1. A value of r close to 1 indicates a strong positive correlation, meaning that the two traits tend to increase or decrease together. A value of r close to -1 indicates a strong negative correlation, meaning that the two traits tend to vary inversely to each other. Finally, a value of r close to 0 indicates a lack of linear correlation between the two traits.

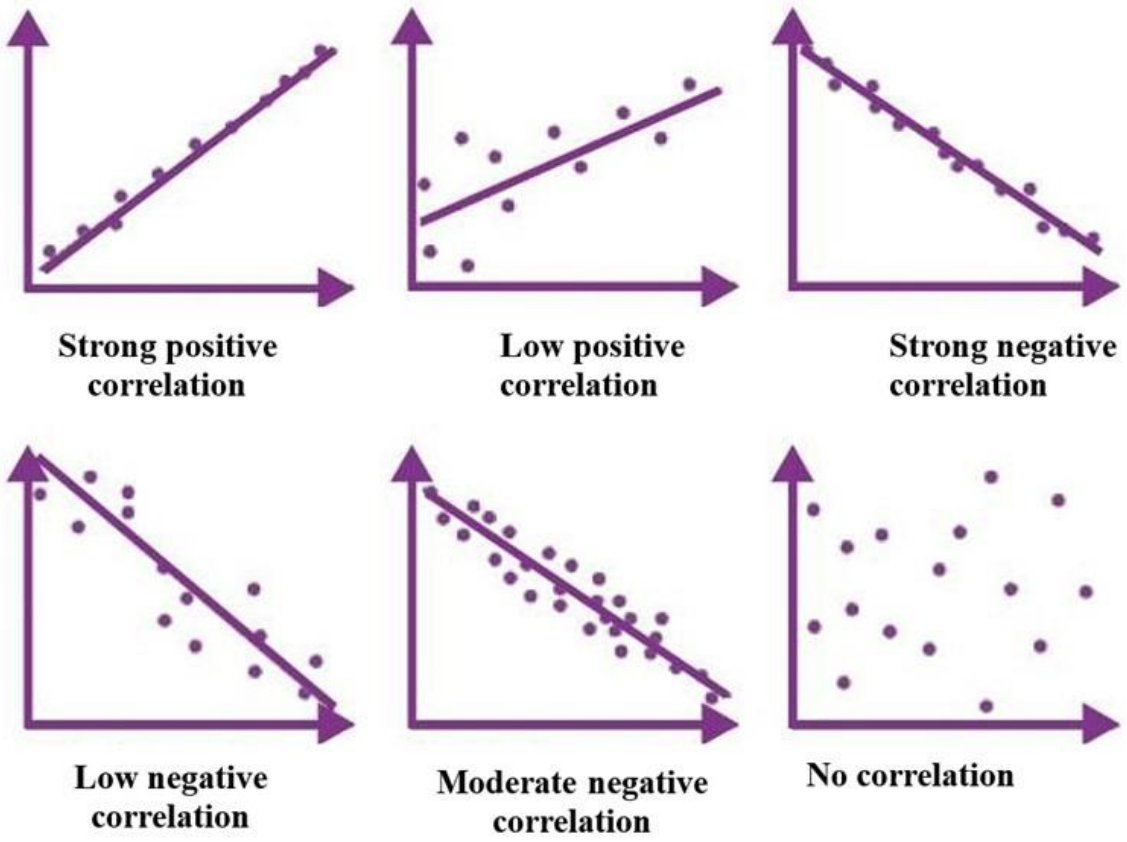


Figure 3 : Different types of correlation between quantitative characters.

The correlation coefficient "r_{xy}" is the quantity commonly used to estimate the degree of correlation between two variables, x and y. It is calculated from the deviations of each x value from the mean of the y values.

$$\mathbf{Cov_{xy} = (1/n-1)[(X_i - \mu_x)(Y_i - \mu_y)]} \quad \text{.....(1)}$$

$$\mathbf{Cov_{xy} = 1/(n-1)[\Sigma Xi.Yi - (\Sigma Xi)(\Sigma Yi)/n]} \quad \text{.....(2)}$$

$$\mathbf{r_{xy} = Cov_{xy}/\sigma x.\sigma y} \quad \text{.....(3)}$$

In formula (3), the products of the deviations are divided by the product of the standard deviations of x and y. This normalization by the standard deviations has the effect of making rxy a dimensionless number independent of the units used to measure x and y, thus defining r_{xy}.

Note: Sometimes, there is no linear relationship between two variables, but a regular non-linear relationship exists between them; the value of one variable can be perfectly predicted from the other.

Example: Parabolic relationship.

At the level of a parabola, each value of Y can be predicted from the value of X without there being a linear correlation between them (Fig. 4).

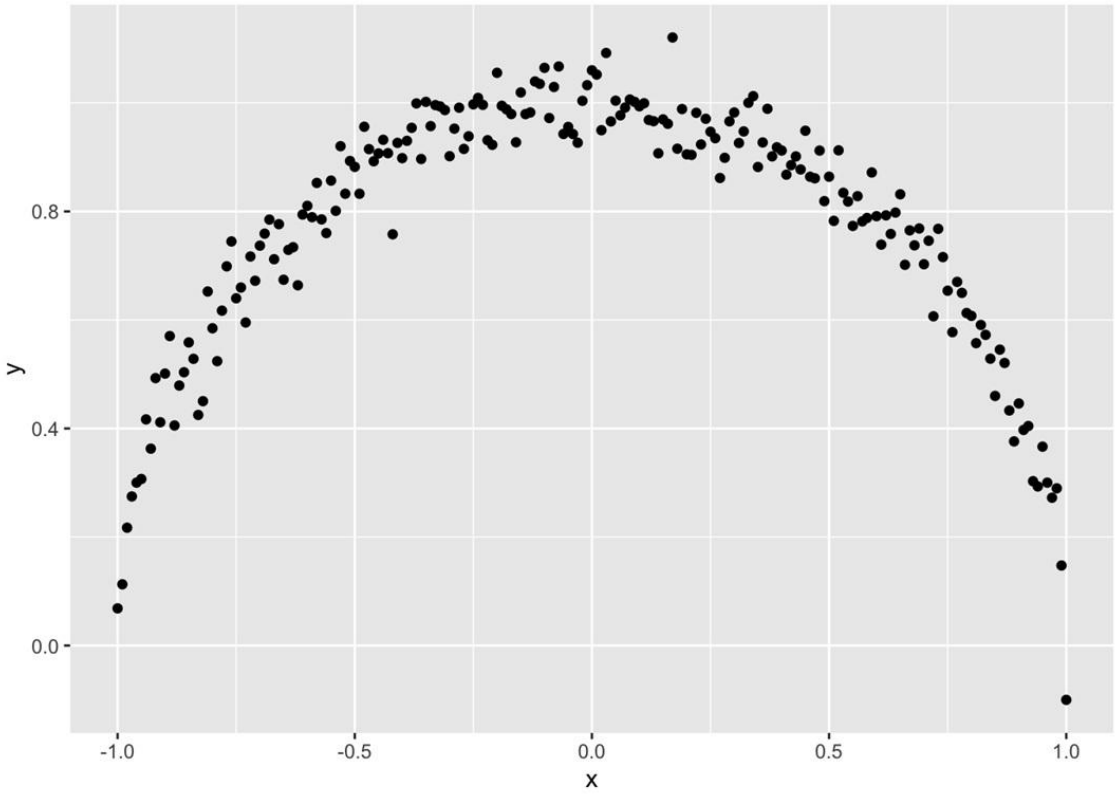


Figure 4: Parabolic correlation between two traits.

he correlation can be made between two characteristics of the same individual but also between, for example, the height of a parent (x) and that of a descendant (y) or between the heights of an older member (x) and a younger member (y) of a family. This use of correlation is the one practiced in quantitative genetics problems.

Note:

It is important to emphasize that correlation does not equal identity. Values can be very well correlated without being identical.

Example 1: The variables x (1; 2; 3; 4; 5; 6; 7; 8; 9; 10) and y (20; 22; 24; 26; 28; 30; 32; 34; 36; 38) within the pairs are perfectly correlated (r = 1), although each value of y is approximately 20 units greater than the corresponding value of x (Fig. 5). Two values are perfectly correlated if for an increase in one there is a constant increase in the other, or a constant decrease if r is negative. Here, when x increases by one unit, y increases by 2

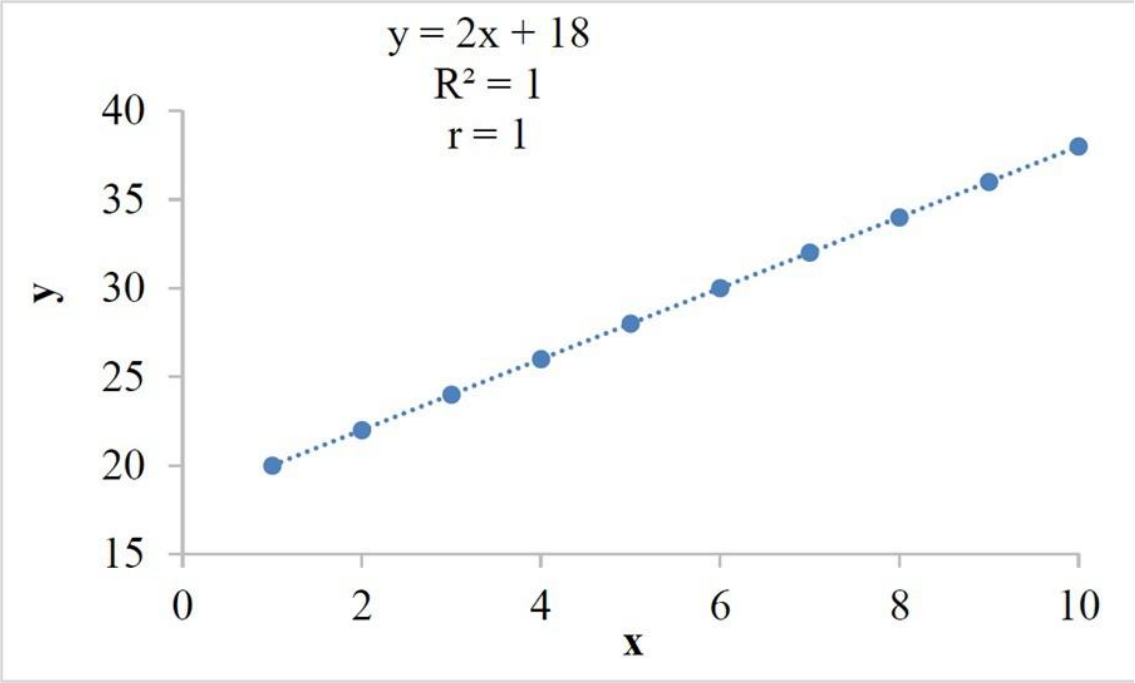


Figure 5: Perfect positive linear correlation between variables x and y.

Example 2: Parents and their offspring can be very well correlated for a characteristic such as height, but because of a change in the environment from one generation to the next, each of the children may be taller than the parent. This phenomenon appears in adoption studies where children, although correlated with their biological parents, can on average be very different from them due to a change in their social situation.

f) Regression

Measuring correlation only provides an estimate of the precision of the existing relationship between two variables. The problem is to predict the value of one variable based on the value taken by the other. If x increases by two units, by how much will y increase?

Regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. It allows us to estimate and predict the value of a dependent variable based on the values of the independent variables.

Regression can be used to analyse quantitative characters and determine the functional relationship between them. In the context of linear regression, the relationship between the dependent variable (y) and an independent variable (x) is modelled by a linear equation of the form

$$y = a + bx$$

where (a) represents the y-intercept and (b) represents the slope (regression coefficient). The regression equation allows us to predict the value of (y) based on the values of (x).

$$b = Cov_{xy}/\sigma^2 x$$

$$a = y - bx$$

Regression can also be used for more complex models, such as multiple regression, where several independent variables are included in the model. In this case, the regression equation takes the form:

$$Y = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n$$

where $x_1, x_2, ..., x_n$ represent the different independent variables and $b_1, b_2, ..., b_n$ are the corresponding regression coefficients.

The equation cannot exactly predict y for a given value of x because of the existing dispersion around the line.

3 .3. Samples and population

When studying a sample of a population, we want to estimate or know the characteristics of the overall population through the sample.

We define a population or a sample by statistical characteristics (mean, variance, standard deviation, etc.). These are certain data for a population (parameters) and random values for a sample (characteristics). The characteristics of each sample are not identical to those of the overall population but vary from sample to sample. This difference is reflected in notations with different symbols: Greek letters for the population and Latin letters for the sample (Table 4).

Example: Two alfalfa samples, extracted from the whole alfalfa population, will not have exactly the same mean and variance.

$\bar{x}$ is actually an unbiased estimate of μ. If a large number of samples are used and $\bar{x}$ calculated for each of them, then the average of the $\bar{x}$ values will be μ.

Parameters or characteristics	Population	Sample
Mean	μ	$\bar{x}$
Variance	σ^2	s^2
Standard deviation	σ	s
Proportion	π or p	f (1)

Number	N	n
Coefficient de regression	α; β	a; b
Coefficient de correlation	ρ	r

***Table 4**: Statistical notations for a population and a sample*

(1) For a sample, we will say the observed proportion or a frequency.

But s^2 is not an unbiased estimate of σ^2; it tends to be a little smaller, and the average of the s^2 values is less than σ^2. As an estimate of the whole represented by the samples, the appropriate value to use is:

$\sigma^2 = (n/n-1)s^2$

$\sigma^2 = (n/n-1).(1/n).[\Sigma(xi - \bar{x})^2]$ (when N < 30) and $\sigma^2 = (n/n-1).(1/n).[\Sigma xi^2 - (\Sigma xi)^2/n]$ (when N > 30)

$\sigma^2 = (1/n-1).[\Sigma(xi - \bar{x})^2]$ (when N < 30) and $\sigma^2 = (1/n-1).[\Sigma xi^2 - (\Sigma xi)^2/n]$ (when N > 30)

with N: number of individuals in the sample.

To estimate the standard deviation, which is simply the square root of the variance, from a sample, the following formula is used:

$\sigma = \sqrt{\sigma^2}$

$\sigma = \sqrt{[(1/n-1).(\Sigma(xi - \bar{x})^2)]}$ (if N < 30).

$\sigma = \sqrt{[(1/n-1).(\Sigma x_i^2 - \Sigma xi)^2/n]}$ (if N > 30).

This is an unbiased estimate of σ^2. The choice between the two formulas depends on the researcher's concern (the sample or the whole). All these considerations on deviations also apply to the sample covariance but not to the correlation coefficients because the factor n/n-1 will appear in both the numerator and the denominator and therefore will not play any role in the calculations. It is important to note that for data from the entire population, the variance is calculated using n instead of n-1 in the formula. The reason for using n-1 for samples is due to Bessel's correction, which takes into account the fact that the estimation of the variance from a sample is more uncertain than the estimation of the variance of the entire population.

3 .4. Genotypes and phénotypes distribution

Assuming that a plant population contains three (03) genotypes, each of them having a different effect on the growth rate. Furthermore, suppose that there is variation due to the environment from plant to plant because of heterogeneous properties of the soil on which the population grows and that

fluctuations occur during development. For each genotype, we will have a distinct phenotype distribution with a mean and standard deviation that depends on the genotype and environmental factors (Fig. 6).

If the proportions are 1/aa, 2/Aa, and 3/AA, the phenotypic distribution of the individuals in the overall population will be of type "b," resulting from the sum of the three genotypic distributions.

The mean of this overall distribution is the average of the three genotypic means. The variance of the overall distribution is produced partly by the variation due to the environment and partly by the slightly different means of the three genotypes.

Properties of the overall distribution

- Despite the existence of three distinct genotypic distributions, there is only one mode;
- ·Any individual within the height range between the two arrows (of the graph) could originate from any of the three genotypes due to their overlap, so we cannot conduct a simple Mendelian analysis to determine the genotype of an individual.

A quantitative character is defined by the fact that the average phenotypic differences between genotypes are small compared to the variation between individuals of the same genotype.

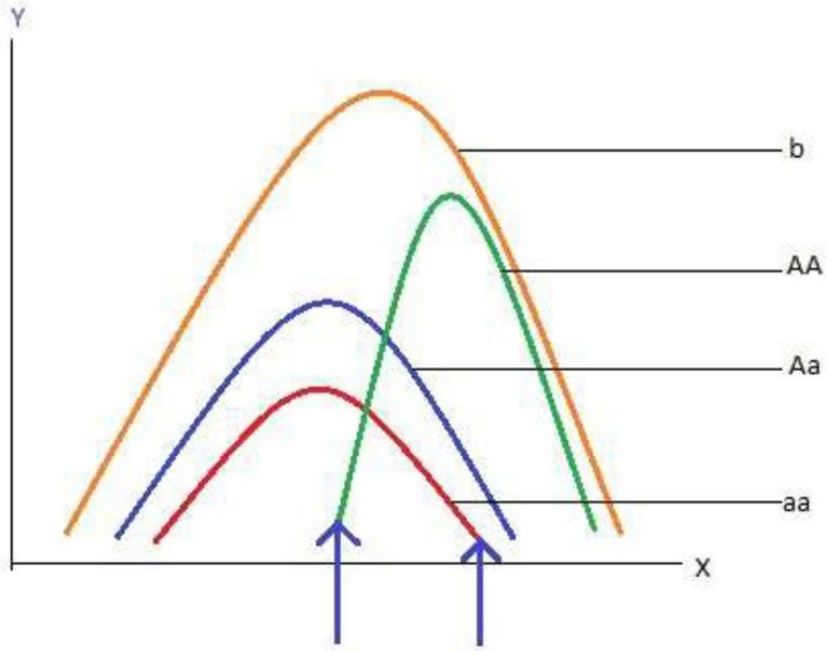


Figure 6: Genotype and phenotype distribution

4. Exercice

Quantitative genetics studies the variations of qualitative traits between individuals.

○	True
○	False

5. Exercice

Which of these traits are quantitative?

☐	Leave colour
☐	Skin colour
☐	Yield
☐	Leaves shape

6. Exercice

In the following series of values, determine the mode and the median: 12 15 13 12 21 14 13 17 27 25 13 31 11 10 19

III. Chapter 2: Polygenic inheritance

Introduction

Polygenic inheritance, also known as quantitative inheritance or polygenic heritability, refers to the influence of multiple genes on a given trait or characteristic. Unlike monogenic inheritance, where a single gene largely determines the trait, polygenic inheritance implies that many genes interact to influence the variation of a given trait.

1 . General information

Continuous characters are characters for which phenotypes vary gradually from one extreme to the other (e.g., human height). These characters vary along a continuum. But we cannot classify humans only according to whether they are "tall" or "short" (Fig. 7).

Figure 7: Extremes of height in human beings.

Dahlia Fatima

In general, continuous characters depend on more than one gene. Characters that are governed by several genes are called polygenic characters.

Many traits are determined by the contribution of multiple genes. For example, plant height, root growth, crop yield, disease resistance, nutrient content, tolerance to environmental stress, flower colour, etc., are often polygenic characteristics.

The polygenic inheritance of traits can also be influenced by environmental factors. Growing conditions, including soil, humidity, temperature, light exposure, etc., can interact with genes to influence the expression of polygenic traits in plants.

Understanding polygenic inheritance is of great importance for crop selection and improvement. By identifying the genes and alleles that contribute to desired traits, breeders can use this information to develop plant varieties with improved characteristics, such as high yield, disease resistance, etc.

Example: skin colour in humans.

Skin colour results from the interactions of several factors determined by different pairs of genes. Some genes could act on the metabolism of skin melanocytes and modify their rate of melanin production. If melanocytes produce more melanin, the skin will be darker. Other genes can determine the distribution of melanin in the thickness of the skin (in Black people, melanin is dispersed throughout the thickness of the epidermis, which is not the case in White people). Some genes could determine the relative amounts of each of the two possible types of melanin (two types of melanin are known: eumelanin, black, and pheomelanin, yellow-red). Others could affect the production of certain hormones involved in the activity of melanocytes.

Suppose, for example, that skin colour is governed by 4 different pairs of genes. Also assume that each of the genes involved exists only in two allelic forms: A and a, B and b, C and c, and D and d (the reality is probably much more complex).

Suppose further that the alleles A, B, C, and D determine factors responsible for a darker skin colour and the alleles a, b, c, and d determine factors responsible for a light colour and that there is no dominance between the alleles. An individual AA BB CC DD would have a very dark skin colour, while an individual aa bb cc dd would have a very pale colour.

It is also understood that all shades between these two extremes can be obtained depending on the possible combinations (Aa BB CC dd would give a particular colour, while aa Bb cc DD would give another).

A couple whose two parents are AaBbCcDd could have children with all possible shades of colour (from "AABBCCDD" to "aabbccdd").

2. Relationships between genes

The set of measurable characters in living beings (plants or animals) is governed by a polygenic system. These genes can have different effects on the characters studied. Indeed, the modes of action of these genes are additivity, dominance, and epistasis.

2.1. Average effects (additivity)

Average effects, also known as additivity, refer to the individual contribution of each allele (alternative form of a gene) to a given trait within the framework of polygenic inheritance.

In the context of polygenic inheritance, each gene involved in determining a particular trait can have several different alleles. Each allele can have a specific effect on the trait, and these effects add up to influence the phenotypic expression of the trait.

In a population that we assume to be panmictic (random mating), a given genotypic character is generally conditioned by several homologous structures (A1, A2, A3,…, Ai,…, and An). Each of these (n) structures has its relative frequency (P1, P2, P3,…, Pi,…, Pn). An individual (ig) of relative frequency (PiPg) in this population therefore has the genotype AiAg. Its corresponding genotypic value could be written [AiAg].

The assessment of the average value [Ai] provided by a structure (Ai) can be made from all genotypes containing (Ai). Therefore:

$$[A_i] = [P_1P_i[A_1A_i]+ P_2P_i[A_2A_i]+ P_3P_i[A_3A_i]+\ldots+ P_nP_i[A_nA_i]]/(P_1P_i+ P_1P_i+ P_1P_i+\ldots+ P_1P_i)$$

This equation simplifies by pi.

$$[A_i] = P_i[P_1[A_1A_i]+ P_2[A_2A_i]+ P_3[A_3A_i]+\ldots+ P_n[A_nA_i]]/ P_i(P_1+ P_2+ P_3+\ldots+ P_n)$$

Noting that P1 + P2 + P3 + … + Pi + … + Pn = 1

$$[Ai] = \sum Pi[AiAn]$$

When average effects are present, it means that each allele contributes independently and additively to the variation of the trait. Therefore, the sum of the effects of all alleles present in an individual largely determines the phenotypic value of the observed trait.

This expression can be applied in a simple case where the population only includes two structures, A and B, with respective frequencies p and q. This situation occurs in the panmictic multiplication of a hybrid between two homozygous lines (Tables 5 and 6).

♀ ♂	**A (p)**	**B (q)**
A (p)	AA (**p^2**)	AB (**pq**)
B (q)	AB (**pq**)	BB (**q^2**)

Table 5: panmictic multiplication of a hybrid between two homozygous lines

Hence

AA	AB	BB
p^2	**2pq**	**q^2**

Table 6: Genotypic frequencies

With: $p^2[AA] + 2pq[AB] + q^2[BB] = 1$

From which we derive the average effects of the structures:

- $[A] = p^2[AA] + pq[AB]$
- $[B] = q^2[BB] + pq[AB]$

We can consider that the three genotypes [BB], [AB] and [AA] mark graduations of types 0, 1 and 2 (Fig. 8).

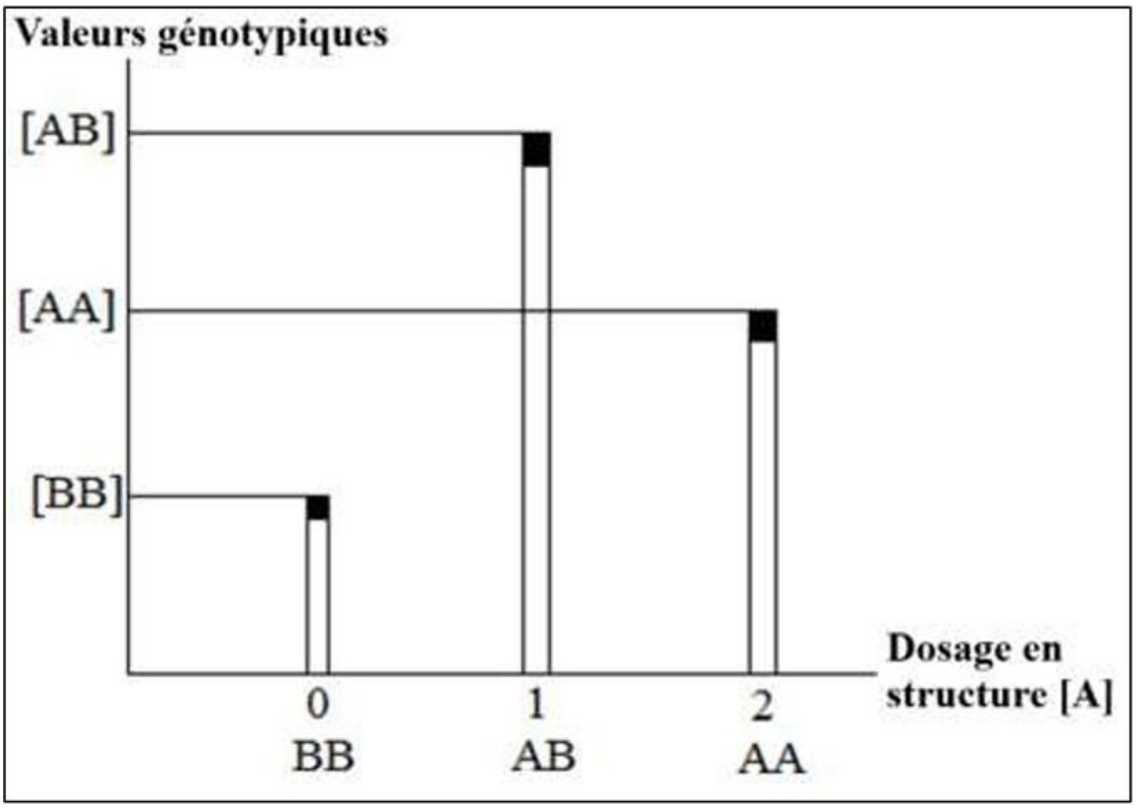


Figure 8: Genotypic values according to the dosage in structure A

If the contributions of structures A and B in the realisation of the phenotype include only additive effects [A'] and the genotype [AB], the genotype [AB] should have a phenotypic value intermediate between [AA] and [BB] (F. In this case, the three genotypes would then be aligned on a straight line (Δ') at ordinates 2B']; 2[A'] and [A'] + [B'].

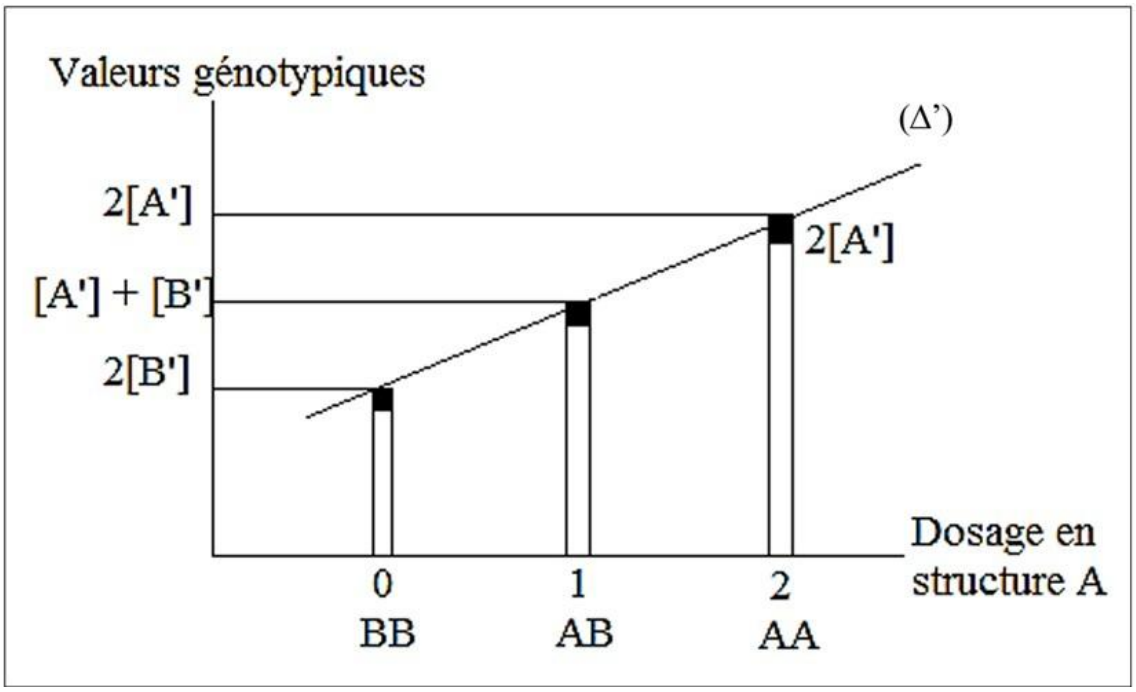


Figure 9: Representation of a case of perfect additivity.

The average effects are not always strictly additive. In practice, this case is very rarely achieved; there may be interactions between alleles, which means that the effect of one allele can be modified by the presence of other alleles. These interactions can be synergistic (combined effects greater than the addition) or antagonistic (combined effects less than the addition).

Among the multiple effects that contribute to the realisation of the actual genotypic values [BB], [AB] and [AA], the part called additivity will be that which is defined by the best-fit regression line (Δ) to the actual values; therefore, by the regression line between these genotypic values (Fig. 10).

On this line correspond respectively to the abscissa; dosages 0, 1 and 2 of structure A and are located on the ordinate points 2 [A'] + and 2 [A'], which represents the share of additivity for each phenotypic value.

Average effects are often used to quantify the contribution of each allele to the variation of a trait and to predict the phenotypic performance of individuals in the context of selection and genetic improvement.

2 .2. Effect of dominance

The effect of dominance is an important concept in genetics that describes the interaction between alleles of a given gene and their influence on the phenotypic expression of a trait. It specifically concerns cases where one allele masks or predominates over another allele in determining the trait.

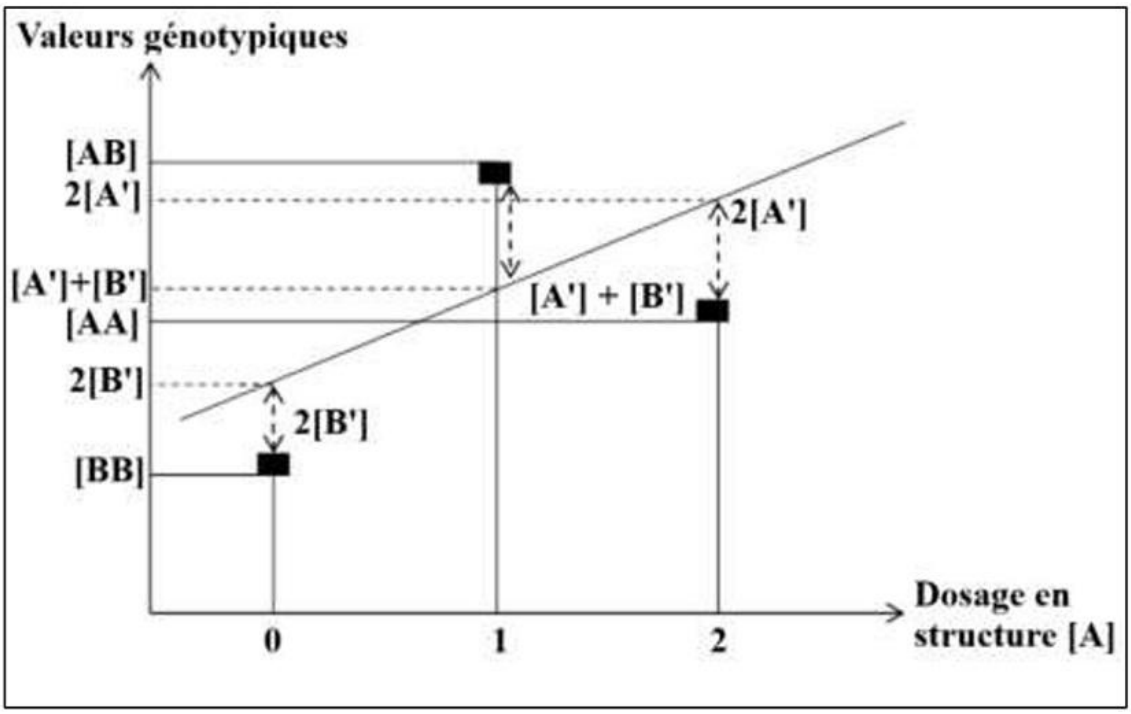


Figure 10: Regression and estimation of dominance and additivity.

Consider Figure 10, we can notice for the considered chromosomal zones that the effects of additivity only covered part of the genotypic value. The difference between these observed actual values and the additive value defines the effects of dominance. Dominance therefore appears as a sum of interaction between homologous segregates.

In a diploid plant, for a given chromosomal zone, if the heterozygote has a perfectly intermediate genotype between that of the two homozygotes, there is perfect additivity (Fig. 11).

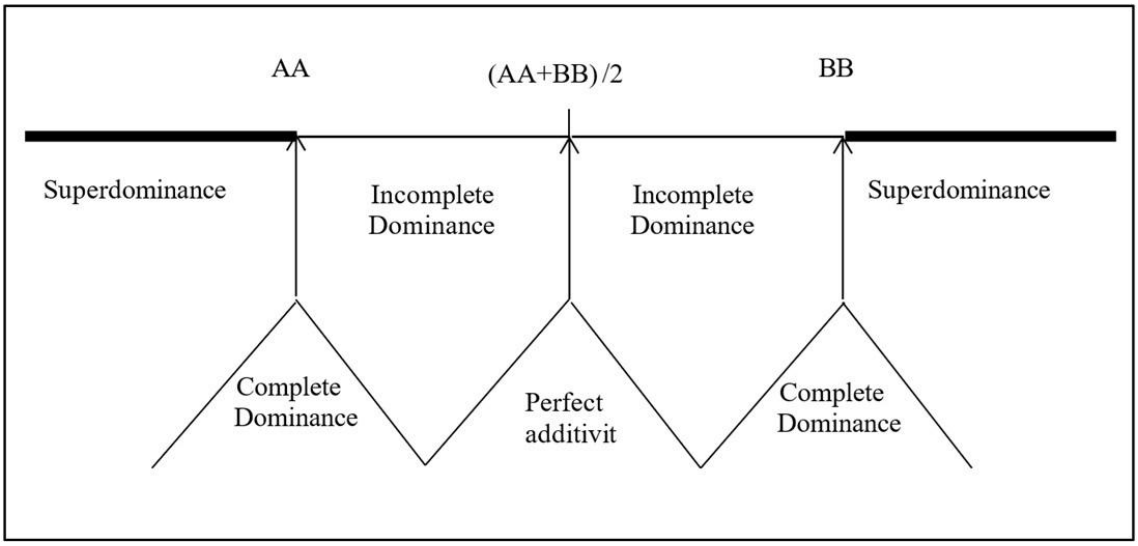


Figure 11: Dominance and definition of the heterozygote.

The other situations are defined as follows:

- [AB] = [AA] or [BB] => Complete dominance;
- [AB] < [AA] or [BB] => Incomplete dominance;
- [AB] > [AA] or [BB] => Overdominance.

Dominance and additivity are not mutually exclusive and can coexist in genetic systems. In some cases, traits can be influenced by both dominant and recessive alleles, as well as by the additive effects of multiple genes.

2 .3. Epistasis

For most complex traits, there are not only two homologous segments involved, but several zones at various points in the genome have been able to decompose the genotypic value.

Indeed; additivity interacts with dominance for homologous segments, but there are different types of interactions between non-homologous segments, the sum of which constitutes the effects of epistasis.

Epistasis is a concept in genetics that describes the interaction between genes, where the effect of one gene masks or modifies the effect of another gene in determining a phenotypic trait. Epistasis can occur when two or more genes interact to influence the same trait.

In the context of epistasis, a gene is considered epistatic if it masks or modifies the effect of another gene, called the hypostatic gene. The epistatic gene can inhibit or block the expression of a phenotypic trait, regardless of the alleles present in the hypostatic gene.

There are different types of epistasis, including dominant epistasis and recessive epistasis.

a) Dominant epistasis

In this case, a single dominant allele of the epistatic gene is sufficient to mask the expression of the hypostatic gene. Therefore, even if the hypostatic gene has recessive alleles that could normally influence the trait, they will be masked by the dominant allele of the epistatic gene.

b) Recessive epistasis

In this case, both recessive alleles of the epistatic gene must be present to mask the expression of the hypostatic gene. If at least one dominant allele is present in the epistatic gene, the effect of the hypostatic gene will be observed.

c) Epistasis can be illustrated by various examples.

For example, in the case of hair colour in mammals, the gene responsible for the production of melanin pigment can be affected by an epistatic gene that blocks pigment production, regardless of the alleles present in the hypostatic gene. Thus, even if the hypostatic gene has alleles that determine hair colour, they will not be expressed in the presence of the dominant allele of the epistatic gene.

Epistasis can complicate the prediction of phenotypes in genetics because it modifies the expected ratios of certain genetic crosses. Understanding epistatic interactions is essential for a better understanding of the genetics of complex traits.

These non-homologous interactions can occur between additive effects (AxAxA…); between dominance effects (DxDxD…); or be mixed (AxDxAxD….) (Fig. 12).

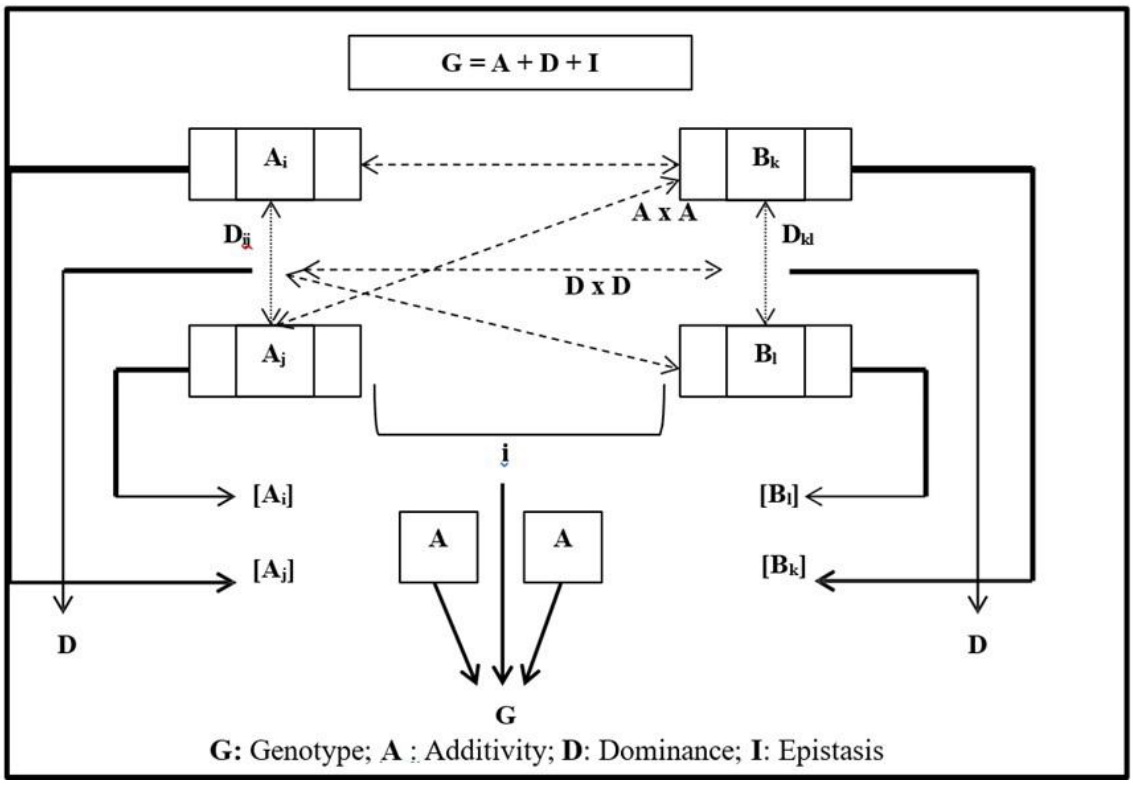


Figure 12: The various types of effects (direct and in interaction) involved in genotypic expression: case of two pairs of genes Ai; Aj; Bk and Bl.

3. Exercice

Which among these interaction, justified a polygenic inheritance ?

☐	Phenotype
☐	Epistasis
☐	Genotype
☐	Additivity

4. Exercice

Interaction between genes belonging to two non homologous chromosome is called epistasis.

○	True
○	False

5. Exercice

What is the main criteria that confirm that a trait is quantitative?

6. Exercice

A qualitative trait is also called:

7. Exercice

A continuum is synonym of:

IV. Chapter 3: Heritability

Introduction

The question of whether or not a trait is hereditary raises the question of the role played by genetic differences in the phenotypic differences existing between individuals or groups of individuals.The heritability of a trait is a measure of the proportion of the phenotypic variation of a trait in a population that is attributable to genetic differences. In other words, it is a matter of quantifying the extent to which the observed differences in a trait between individuals are due to genetic variations rather than environmental factors.

1 . Kinship and Heritability

Kinship and heritability are two concepts related to genetics and the transmission of traits from one generation to the next, but they have different meanings.

Kinship refers to the degree of genetic relationship between two individuals. Individuals who share a more recent common ancestor have a higher kinship than those who share a more distant common ancestor.

Biologically related individuals should resemble each other more than unrelated individuals. This resemblance will be reflected as a positive correlation between parents and offspring, or between individuals from the same parents.

Examples:

- Taller-than-average parents should have taller-than-average offspring.
- The more seeds a plant produces, the more its offspring should produce.

Kinship refers to the degree of genetic relationship between individuals, while heritability measures the proportion of the phenotypic variation of a trait in a population that is attributable to genetic differences.

Kinship can influence the heritability of a trait, as heritable traits are more likely to be transmitted between closely related individuals. However, kinship alone does not determine the heritability of a

trait. Heritability is a statistical measure that is determined by genetic studies and depends on the genetic variation observed in a specific population.

Traits are said to be familial if members of the same family share them regardless of the reason. These traits are only hereditary if their similarity is due to common genotypes. In organisms amenable to experimentation, there is no problem in distinguishing similarities due to the environment from those due to heredity.

Examples:

- The offspring of a disease-susceptible plant and that of a resistant plant can be grown together in the same environment to determine whether, despite the similarity of the environment, each of them resembles its own parent (genetic difference).
- The fact that parents speak a language and their children speak the same language does not mean that it is a hereditary trait, but it is a familial trait. If there is immigration, the children will speak the language of the host country.

2. How to Quantify Heritability?

If a trait shows some heritability in a population, it is then possible to quantify its degree of heritability.

2.1. Using Variances

The variation within the phenotypes of a population stems from two sources. On the one hand, the average differences between genotypes, and on the other hand, each genotype expresses phenotypic variance due to environmental variance.

The total phenotypic variance (denoted σ^2_p) can be divided into two parts:

- The variances within the mean genotypic values (σ^2_g);
- The environmental variance (σ^2_e).

$$\sigma^2_p = \sigma^2_g + \sigma^2_e$$

Broad-sense heritability (H^2) is a measure of the proportion of the phenotypic variation of a trait in a population that is attributable to genetic differences. It includes both the direct effects of genes (additive heritability) and the effects of interactions between genes (non-additive heritability), such as dominance and epistasis effects. In other words, H^2 captures the total contribution of genetic factors

to the variation of the trait, and the degree of heritability can be defined as the contribution of genetic variance to the total variance:

$$\mathbf{H^2 = \sigma^2_g/\sigma^2_p = \sigma^2_g/\sigma^2_g + \sigma^2_e}$$ **(Broad-sense heritability)**

In contrast, narrow-sense heritability (h^2) refers specifically to the proportion of the phenotypic variation of a trait that is attributable to the effects of additive genes alone. It does not take into account non-additive effects, such as dominance or epistasis. Thus, h^2 represents a more restricted estimate of the genetic contribution to the variation of the trait.

$$\mathbf{h^2 = \sigma^2_{Ad}/\sigma^2_p =}$$ **(Narrow-sense heritability)**

The distinction between H^2 and h^2 is important because it allows a better understanding of the genetic mechanisms underlying trait variation. If h^2 is significantly high, this suggests that genetic differences between individuals contribute significantly to the variation of the trait. On the other hand, if H^2 is high compared to h^2, this indicates that non-additive effects, such as dominance, play an important role in the variation of the trait.

Note

The heritability of a trait varies according to the population and the set of environments in which it develops.

Heritability is not a fixed characteristic of a trait but depends on the population within which the measurement is made as well as the set of environments in which the population evolves:

- In a population, a trait can be influenced by alleles segregating at numerous loci. In another, these loci may very well be homozygous (pure line), and in this case, the heritability of the traits will be zero since the genotypic variance (σ^2_g) = 0 with an $H^2 = 0$; such a value does not mean that genes play no role in the development of the trait, but rather that the existing variation between individuals in this population cannot be attributed to genetic variation.
- A population that develops in a very homogeneous environment will be characterised, for a given trait, by a smaller environmental variance than a population evolving in a more variable environment. The environmental variance will be zero and therefore $\sigma^2_g = \sigma^2_p$ with an $H^2 = 1$. The lack of environmental heterogeneity will express a higher heritability value, but this value does not mean that the trait is insensitive to the action of the environment.

a) In-depth analysis of variance (additive variance and environmental variance)

The subdivision of phenotypic variance can provide interesting information for breeders and growers.

Example 1: If two alleles A and a at one locus, controlling size, segregate in the environments where the population is found, the mean phenotypes (stem height) and the frequency of the three genotypes are as follows:

	aa	Aa	AA
Phenotypes	10	18	20
Frequencies	0.36	0.48	0.16

***Table 7**: Results of means and frequencies of genotypes (aa, Aa, and AA) of stem height.*

The additive variance is calculated from the square of the deviation from the mean of the average effect of A relative to the weighted average of its mean.

$\sigma^2_{Ad} = 2[fA(\bar{A} - \mu)^2 + fa(\bar{a} - \mu)^2]$

The frequency of allele a: fa = [2f(aa) + 1f(Aa)] /2 = [2(0.36) + 0.48] /2 = 0.6

with 2 because there are two a alleles in the homozygous genotype aa and 1 because there is one a allele in the heterozygous genotype Aa.

The frequency of allele A: fA = [2f(AA) + 1f(Aa)] /2 = [2(0.16) + 0.48] /2 = 0.4

The average effect of allele a is denoted ā, and the average effect of allele A is denoted Ā.

ā = [2(0.36)(10) + (0.48)(18)] / [2(0.36) + 0.48] = 13.20 cm

Ā = [2(0.16)(20) + (0.48)(18)] / [2(0.16) + 0.48] = 18.80 cm

This average difference of 5.6 cm between the effects due to alleles A and a (18.80 – 13.20 = 5.60) accounts for part of the phenotypic variance but not all of it because the heterozygote is not exactly intermediate compared to the homozygotes; there is a dominance effect.

The phenotypic mean μ = 10(0.36) + 18(0.48) + 20(0.16) = 15.44 cm

Therefore:

$\sigma^2_{Ad} = 2[0.4(18.8 – 15.44)^2 + 0.6(13.2 – 15.44)^2] = 15.05\ cm^2$

The total genetic variance associated with this locus can be divided into:

• Additive genetic variance (genetic variance due to the substitution of A by a);

• Variance due to dominance (genetic variance resulting from the partial dominance of A over a in heterozygotes).

The total genetic variance resulting from the variation among the mean phenotypes of the 3 genotypes is:

$\sigma^2_g = 0.36(10 - 15.44)^2 + 0.48(18 - 15.44)^2 + 0.16(20 - 15.44)^2$

Hence:

$\sigma^2_d = \sigma^2_g - \sigma^2_{Ad} = 17.13\ cm^2 - 15.05\ cm^2 = 2.08\ cm^2$

This subdivision of genetic variance is useful because it allows predicting the effect of selection.

Example 2: Consider the same previous example except that the alleles that control plant size at the given locus are A and A_1.

	AA	**AA1**	**A1A1**
Phenotypes	10	12	10
Frequencies	0.25	0.50	0.25

***Table 8**: Results of means and frequencies of genotypes (AA, AA_1, and A_1A_1) of stem height.*

$\sigma^2_{Ad} = 2[fA(\bar{A} - \mu)^2 + fa(\bar{a} - \mu)^2]$

The frequency of allele A: $fA = [2f(AA) + 1f(AA1)] /2 = [2(0.25) + 0.5] /2 = 0.5$

The frequency of allele A_1: $fA_1 = [2f(A_1A_1) + 1f(AA_1)] /2 = [2(0.25) + 0.5]/2 = 0.5$

$\bar{A} = [2(0.25)(10) + (0.5)(12)] / [2(0.25) + 0.5] = 11$ cm

$\bar{A}_1 = [2(0.25)(10) + (0.5)(12)] / [2(0.25) + 0.5] = 11$ cm

The phenotypic mean $\mu = 10(0.25) + 12(0.5) + 10(0.25) = 11$ cm

Therefore:

$\sigma^2_{Ad} = 2[0.5(11 - 11)^2 + 0.5(11 - 11)^2] = 0\ cm^2$

The total genetic variance resulting from the variation among the mean phenotypes of the 3 genotypes is:

$\sigma^2_g = 0.25(10 - 11)^2 + 0.5(12 - 11)^2 + 0.25(10 - 11)^2 = 1$

Hence:

$\sigma^2_d = \sigma^2_g - \sigma^2_{Ad} = 1\ cm^2 - 0\ cm^2 = 1\ cm^2$

Interpretation

There is no difference between the effects of alleles A and A1 since each of them determines an effect of 11 units; there is no additive genetic variance, although variance due to dominance is present; the largest individuals are heterozygotes.

If a breeder tries to increase the size of this population by selection, the union of these heterozygotes will not be better than the original population. The selection will be ineffective (will not lead to anything; we will return to the parents).

The effect of selection depends on the part due to the additive genetic variance and not on the genetic variance in general. It follows that it is the heritability in the narrow sense and not the heritability in the broad sense that allows us to predict the response to selection.

b) Estimation of the components of genetic variance

The genetic components of variance are estimated from the covariance between related individuals (Table 9).

Related individual	**Estimated proportions for**	
	S^2Ad	S^2d
Cov (Identical twins)	1	1
Cov (Parent – child)	1/2	00
Cov (Half indreds)	1/4	00
Cov (Inbreds)	1/2	1/4

Table 9: *Relative proportions of additive variance and variance due to continuous dominance in the genetic covariance between related individuals.*

These relationships combined with the total phenotypic variance allow us to estimate heritability in the narrow sense.

Example

The covariance between parents and children contains one-half of the additive variance; therefore:

$\text{Cov(parent – child)} = ½\ \sigma^2_{Ad} \Rightarrow 2\ \text{Cov}_{(\text{parent – child})} = \sigma^2_{Ad}$

$\text{Cov(parent – child)} = ½\ \sigma^2_{Ad} \Rightarrow [2\ \text{Cov}_{(\text{parent – child})}] / \sigma^2_p = \sigma^2_{Ad} / \sigma^2_p$

$\text{Cov(parent – child)} = ½\ \sigma^2_{Ad} \Rightarrow [2\ \text{Cov}_{(\text{parent – child})}] / \sigma^2_p = h^2$

It follows that twice the parent-child correlation provides an estimate of heritability in the narrow sense.

There is another way to estimate heritability in the narrow sense. If we plot the phenotypes of the offspring against the average phenotypes of their parents (Fig. 13). It is possible to obtain the following relationships:

The regression line will pass through the mean of all parents and the mean of all offspring, which are equal since no change has occurred in the population from one generation to the next.

In addition, taller parents have taller children and shorter parents have shorter children, so the slope of the regression line is positive, but the slope is not equal to 1.

Very short parents have slightly taller children, and very tall parents have slightly shorter children than themselves. This slope of less than 1 of the regression line stems from the fact that heritability is far from perfect.

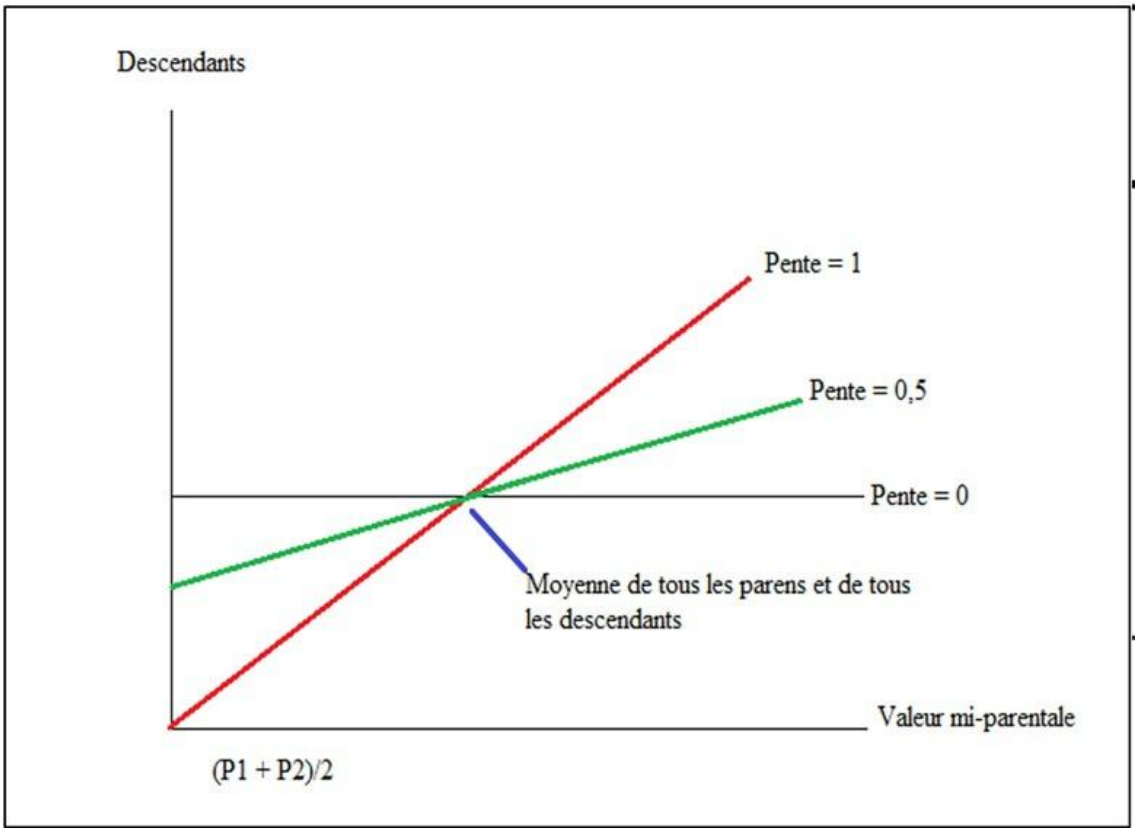


Figure 13: Regression of the measured values of the offspring relative to that of the averages of the two parents for a trait with a narrow-sense heritability of 0.5 (green line). The red line illustrates the regression slope obtained if the trait were perfectly heritable.

If a phenotype were inherited additively with perfect fidelity, the size of the offspring would be identical to the mid-parental value, and the slope of the line would be equal to 1.

On the other hand, if the offspring showed no similarity to their parents, all parents would have offspring of the same average size, and the slope of the regression line would be equal to 0.

This suggests that the slope of the regression line relating the offspring value to the mid-parental value is an estimate of the additive heritability (in the narrow sense), and in fact, this relationship is accurate.

By definition, the regression $\mathbf{b = Cov_{xy}/\sigma^2_x}$ (of the offspring value on the variance of the parents). From the table of proportions of additive and dominance variances, we will have:

Cov parent-child $= \sigma^2_{Ad}/2$

It should be noted that the mean values vary less than their components, so by setting σ^2_p as the phenotypic variance of the parents, we obtain the following:

$$2b = [2(\sigma^2_{Ad} / 2)] /\sigma^2_p = \sigma^2_{Ad} /\sigma^2_p = h^2$$

The fact that the slope corresponds to the additive heritability allows us to use it to predict the effects of artificial selection.

2 .2. Another method for estimating the heritability value

Assuming that we choose as parents of the next generation those who have, on average, two units of value above the general mean of the population from which they were chosen.

If the heritability in the narrow sense is equal to 0.5, the offspring of the selected parents from the population will be 0.5*2=1 unit above the mean of the current population since the regression coefficient predicts that an increase in y will be due to an increase of one unit of x.

We can define:

- The selection differential as the difference between the selected parents and the general mean of the selected population);
- The selection response is the difference between their offspring and the starting generation.

$$\textbf{Selection response} = h^2 * \textbf{selection differential}$$

$$h^2 = \textbf{Selection response / Selection differential}$$

This second expression gives us another way to express heritability in the narrow sense by selecting for one generation and comparing the response with the selection differential. Most often this is done for several generations, and the average response is determined.

3. The use of h^2 in breeding and improvement

The importance of h^2 is great for breeders, even if its value only applies to a particular population and a range of environments.

Example: A poultry geneticist wishing to increase the growth rate is not concerned with the genetic variance of all possible farmyard chickens distributed in all possible environments.

The question is whether a selection scheme can be designed to increase the growth rate and, if so, to what extent?

- If one group has a high genetic variance and another has a very low variance, the breeder will choose the first for selection.
- If the heritability (h^2) of the chosen group is very high, the population mean will respond quickly to the selection pressure imposed, since most of the superiority of the chosen parents will be reflected in the offspring. The higher h^2 is, the higher the parent-offspring correlation will be.
- If h^2 is low, then only a small fraction of the increased growth rate of the selected parents will be found in the next generation.
- If H^2 and h^2 are low, this implies that the importance of the variance due to the environment is large compared to the genetic variance. Can the environmental variance (S^2e) be reduced? There are several methods to decrease (S^2e): i) Modify the breeding conditions so that the variance due to the environment decreases. OR, ii) Use family selection. Rather than choosing the best individuals, the breeder produces several offspring from various pairs, and mating is decided on the basis of the average performance of the offspring. By establishing an average from the offspring, uncontrolled fluctuations due to the environment or development are eliminated, and a better estimate of the genotypic difference between pairs can be established and thus allows the best partners to be chosen as parents of the next generation.
- If h^2 is low but H^2 is high, this means that the role of the variance due to the environment is minimal. The low heritability in the narrow sense is the result of a low additive genetic variance compared to the variance due to dominance and interactions. Such a situation leads to the use of particular selection schemes based on non-additive variance. This is the controlled hybrid method widely used in corn and tomatoes. A large number of pure lines are

produced by self-fertilisation. They are then crossed with each other in all possible combinations (if economically justified), and the cross providing the best hybrid is chosen. New pure lines are produced from this superior hybrid, and crosses are made again to detect the best hybrid during this second cycle. This scheme actually selects for dominance effects because it is based on the best heterozygotes.

4. Exercice

Here is the date of ear emergence (in number of days elapsed since an arbitrary date), P1 and P2 are pure lines.

P1: σ^2_p = 12.99; n=159; μ=24.25

P2: σ^2_p = 27.61; n=148; μ=33.48

F1: σ^2_p = 18.45; n=171; μ=27.29

F2: σ^2_p = 21.20; n=552; μ=31.12

Is the heading date a quantitative character?

Estimate H² for P1, P2 and F1.

5. Exercice

We measure the grain weight per plant in maize (g). The phenotypic standard deviation in a large population is 15. The phenotypic standard deviation in a pure (inbred) line is 12.

• Calculate the broad-sense heritability of grain weight per plant in the inbred line and in the large population.

6. Exercice

The study of plant weight in a population has shown that the variance is 3.90. The covariance between half-sibs is 0.56.

• Estimate the narrow-sense heritability of plant weight in this population.

7 . Exercice

In a population of alfalfa, the average stem length is 30 cm. A set of plants with an average stem length of 35cm are isolated and crossed among themselves. The average length of the offspring is 33cm.

• From this data, calculate the narrow-sense heritability of stem length in this population.

V. Chapter 4: Heterozygosity

Introduction

Heterozygosity is a concept in genetics that refers to the presence of two different alleles for a particular gene in an individual. An allele is an alternative version of a gene that can influence a specific trait or characteristic in an organism.

Heterozygosity plays a crucial role in quantitative genetics for these reasons:

- **Genetic Diversity**: Heterozygosity is an indicator of genetic diversity within a population. A population with high heterozygosity is generally considered to be more genetically diverse. This genetic diversity is important because it offers greater variability in traits and characteristics, which can promote adaptation and survival in the face of environmental changes.
- **Eevolutionary Potential:** Heterozygosity is linked to the evolutionary potential of a population. When a population has high heterozygosity, it has a richer genetic reservoir. This means that it has a greater capacity to generate new genetic combinations through recombination during reproduction, which can lead to the emergence of new phenotypic variations and promote the evolution of the population.
- **Genetic Stability**: Heterozygosity plays a role in maintaining genetic stability within populations. When individuals are heterozygous for a given gene, it means they carry two different alleles for that gene. This increased allelic diversity can contribute to maintaining a genetic balance and preventing the accumulation of deleterious mutations.
- **Genetic Improvement**: In quantitative genetics, heterozygosity is often sought in plant and animal breeding programs. When there is high heterozygosity, it can indicate greater genetic variability and higher potential for improvement of desired traits, such as productivity, disease resistance, or product quality.
- **Study of Genetic Diseases**: Heterozygosity is also important in the study of genetic diseases. In some cases, being heterozygous for a gene can confer an advantage in terms of resistance to certain diseases. For example, in the case of recessive diseases, being heterozygous can offer protection against the disease while maintaining the genetic diversity of the population.

1 . Heterozygosity and Allele Frequency

Heterozygosity is often measured by calculating the rate of heterozygous individuals in a given population. For example, if a population has 100 individuals and 20 of them are heterozygous for a specific gene, the heterozygosity rate would be 20%.

1 .1. Calculating Heterozygosity from Allele Frequencies

In a diploid population, allele frequency can be calculated from heterozygosity and vice versa. Suppose there are two alleles for a gene, A and a, and that p represents the frequency of allele A and q the frequency of allele a in the population. Then, the frequency of homozygosity AA would be p^2, the frequency of homozygosity aa would be q^2, and the frequency of heterozygosity Aa would be $2pq$. Moreover, $p + q = 1$, because the total allele frequencies must sum to 1.

Heterozygosity can be calculated from allele frequencies using the following formula:

$$\text{Heterozygosity} = 2pq$$

where p represents the frequency of allele A and q represents the frequency of allele a in the population.

Example: Suppose that in a population, the frequency of allele A is 0.6 (p = 0.6) and the frequency of allele a is 0.4 (q = 0.4).

$$\text{Heterozygosity} = 2pq = 2(0.6)(0.4) = 0.48$$

Therefore, in this population, the heterozygosity is 0.48, which means that 48% of the individuals are heterozygous for this gene.

Note: This formula is based on certain assumptions, including Hardy-Weinberg equilibrium, which assumes the absence of evolutionary forces such as selection, genetic drift, migration, and mutation. In reality, these forces can influence the allele frequencies and heterozygosity of a population.

1 .2. Relationship between Heterozygosity and Genetic Homeostasis

Heterozygosity refers to the presence of two different alleles for a gene in an individual. It is often associated with greater genetic variability within a population. When individuals are heterozygous for a gene, it can contribute to maintaining a genetic balance and preventing the accumulation of deleterious mutations.

Heterozygosity can also promote adaptation and survival in the face of environmental changes, as it offers greater genetic diversity and higher evolutionary potential.

Genetic homeostasis, on the other hand, refers to the tendency of populations to maintain the stability of allele frequencies over time, in the absence of major evolutionary forces. Genetic homeostasis is based on the principle of Hardy-Weinberg equilibrium, which specifies that allele frequencies remain constant from one generation to the next if certain conditions are met. In this equilibrium, allele frequencies and heterozygosity can be predicted based on the initial allele frequencies.

Thus, heterozygosity can contribute to genetic homeostasis by maintaining a balance between alleles and preventing allele frequencies from deviating significantly from their initial values. If heterozygous individuals have a selective advantage over homozygotes for a given gene, this can lead to a stable equilibrium where both types of alleles are maintained in the population.

However, it is important to note that genetic homeostasis is not a permanent state and can be disrupted by evolutionary forces, such as natural selection, genetic drift, migration, and mutation. These forces can change the allele frequencies and heterozygosity of a population over time.

2. Hardy-Weinberg laws

The Hardy-Weinberg laws describe the fundamental principles of genetic equilibrium in Mendelian populations in the absence of major evolutionary forces. These laws were formulated independently by the British geneticist Godfrey Hardy and the German geneticist Wilhelm Weinberg at the beginning of the 20th century.

2.1. Hardy-Weinberg equilibrium conditions

In a population of unlimited size, not subject to selection, mutations, and/or migrations, and where unions occur randomly, gene frequencies remain constant from generation to generation, as do genotype frequencies, which can be deduced from knowledge of gene frequencies. A population is said to be in Hardy-Weinberg equilibrium when it satisfies all the conditions of validity for this law.

- **Closed population**: this condition excludes migrations and interbreeding between populations, which can modify gene frequencies;
- **Unlimited size**: gene and genotype frequencies can vary from one generation to the next simply due to chance, but these variations tend to accumulate as the population size becomes large.

- **Absence of selection**: it is assumed that the number of descendants of an individual is independent of its genotype;
- **Absence of mutation:** mutations are random events of very low probability that cause new genes to appear or existing genes to disappear in the population, thus modifying gene frequencies.
- **Panmixia**: random mating is opposed to the following systems: (i) **Homogamy**: when unions occur preferentially between individuals that resemble each other phenotypically (phenotypic homogamy) or genotypically (genotypic homogamy); (ii) **Heterogamy:** when unions occur preferentially between dissimilar individuals; (iii) **Endogamy**: when unions occur preferentially between related individuals (inbreeding); **Exogamy**: when unions occur preferentially between unrelated individuals.

Demonstration:

Consider the case of a locus with two alleles and an infinite population where the frequencies of the 3 genotypes have any values in generation N.

$A_1A_1 = P$; $A_1A_2 = 2Q$ and $A_2A_2 = R$

$P + 2Q + R = 1$

The next generation will be obtained under panmictic conditions by assuming the

3 genotypes are equally fertile (absence of selection) and not susceptible to mutation by random combination of gametes carrying the A_1 allele, whose frequency is **$P + Q = p$**, and gametes carrying the A_2 allele, whose frequency is **$Q + R = q$.**

Gametes	**A1(p)**	**A2(q)**
A1(p)	A1A1 (p^2)	**A1A2(pq)**
A2(q)	**A1A2(pq)**	A2A2 (q^2)

Table 10: Results of panmictic crossing of gametes A_1 and A_2

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

The frequency of the A_1 allele in this generation is **p^2+pq**

The frequency of the A_1 allele in this generation is $p(p + q)$ and since **$p + q = 1$** The frequency of the A_1 allele in this generation = p

The frequency of the A_2 allele in this generation is **$q^2+ pq$**

The frequency of the A_2 allele in this generation is q(p + q) and since p + q = 1 The frequency of the A_2 allele in this generation = q

Therefore, the gene frequencies in a population do not vary. Since we have the same gene frequencies, then the genotype frequencies of generation n+2 will also remain unchanged.

3. Evolution of heterozygosity over generations

If the starting individual in the first generation is homozygous for a locus, it follows that all its descendants obtained by self-fertilisation will be homozygous and identical for this locus.

If, on the other hand, the starting individual is heterozygous (Aa), self-fertilisation will produce ¼ homozygous AA, ¼ homozygous aa, and ½ heterozygous Aa. If only one individual is chosen from the next generation to ensure the propagation of the lineage, there is a 50% chance (½) that it will in turn be heterozygous; after 2 generations, 25% (¼); after 3 generations, 1/8; and at the nth generation, $H_n = 1/2H_{n-1}$ and H_n tends towards 0 (H_n is the proportion of heterozygous loci at the nth generation).

When self-fertilisation is not possible, brother-sister unions lead to the same results, albeit more slowly (Table 11).

Generations	**Heterozygosity level**	
	self-fertilisation	**Brother-sister unions**
0	1,000	1,000
1	0,500	0,750
2	0,250	0,625
3	0,125	0,500
4	0,062	0,406
10	0,000977	0,114
20	$1{,}05.10^{-6}$	0,014
n	**Hn= ½ Hn-1**	**Hn= ½ Hn-1+ ¼ Hn-2**

Table 11*: Level of heterozygosity after several generations of inbreeding in the case of two mating systems.*

4. Factors influencing heterozygosity

Heterozygosity, which refers to the presence of two different alleles for a gene in an individual, can be influenced by several factors. Here are some of the main factors that can affect heterozygosity:

- **Genetic mutation**

Genetic mutations, which are random changes in the DNA sequence, can influence heterozygosity by introducing new alleles into a population. Mutations can increase heterozygosity by creating new alleles and adding to the genetic diversity of the population.

- **Gene flow (gene migration)**

Gene flow, which is the movement of individuals or genes between populations, can influence heterozygosity. If individuals carrying different alleles migrate from one population to another, this can increase heterozygosity in the recipient population.

- **Natural selection**

Natural selection can play an important role in maintaining or reducing heterozygosity in a population. If heterozygous individuals have a selective advantage over homozygotes (what is called heterozygote advantage), this can promote heterozygosity. On the other hand, if homozygotes have a selective advantage, this can lead to a reduction in heterozygosity.

- **Genetic drift**

Genetic drift, which is the random change in allele frequencies over generations, can influence heterozygosity. In small populations, genetic drift can lead to random fluctuation of allele frequencies, which can reduce heterozygosity over time.

- **Inbreeding**

Inbreeding, which is reproduction between closely related individuals, can reduce heterozygosity. When inbred individuals reproduce, there is a greater probability that identical alleles will meet, which reduces heterozygosity and increases homozygosity.

- **Genetic segregation**

During gamete formation, alleles are distributed randomly, which can influence heterozygosity. If the alleles are distributed equitably and independently during meiosis, this can promote heterozygosity by generating a variety of allelic combinations in the gametes.

5. Exercice

If the heterozygosity level in the second generation is 0.25, estimate it in the third generation if the individuals are in self-fertilisation.

6. Exercice

If the heterozygosity level in the second generation is 0.750 and 0.625 in the third one, estimate it in the fourth generation if the individuals are in brother-sister unions.

7. Exercice

The preferential cross between individuals showing the same phenotype is called:

○	Exogamy
○	Heterogamy
○	Homogamy
○	endogamy

8. Exercice

The number of individuals in a population must be limited to consider a population as equilibrate (according to Hardy-Weinberg low)

○	True
○	False

VI. Chapter 5: Heterosis

Introduction

Heterosis has been defined, at the level of a cross between two homozygous lines, as the superiority of the F_1 hybrid compared to the better parent when moving from F_1 to F_2 by self-fertilisation or open pollination.

Heterosis can also be defined at the level of a cross between populations as the superiority of the resulting hybrid compared to the better population.

In general, heterosis is much stronger in allogamous plants than in autogamous plants. The most affected traits are complex traits.

1. Hypotheses or Mechanisms

Two main hypotheses have been formulated: the dominance hypothesis and the overdominance hypothesis.

1.1. Dominance

For a quantitative trait controlled by several genes, the fairly systematic advantage of crossing between unrelated plants compared to crossing between related plants can be due to the bringing together in a single genotype of a large number of favourable dominant genes.

Example

L_1: AABBccdd x **L_2:** aabbCCDD

Hybrid F_1: AaBbCcDd

The hybrid brings together, in a single generation, the 4 favourable genes in the same genotype. In this situation, it is possible to conceive a homozygous genotype as good as the F_1; the heterosis is said to be fixable.

Self-fertilisation of F_1 plants reveals the unfavourable recessives, hence the drop in vigour from F_1 to F_2. This hypothesis assumes that dominance is in the favourable direction.

1 .2. Over dominance

In this hypothesis, the superiority of the hybrid compared to the parents would come from the superiority of the heterozygous state at a certain number of loci. The combination of the two genes in the heterozygous state would lead to a new potential superior to that of the homozygotes.

Example 1

Let there be 2 genes with two alleles, A_1 and A_2. A_1 is effective in environment E_1, and A_2 is effective in environment E_2. E_1 and E_2 are successive environments in the life of the individual (Table 12).

Genotypes	Environments		Sum
	E1	E2	
A1A1	4	1	5
A1A2	4	4	8
A2A2	1	4	5

***Table 12**: Effects of environments on genotypes.*

In each environment, there is only dominance, but over all environments, there is overdominance; this is called marginal overdominance.

Example 2

Let there be two allelic genes with a pleiotropic effect (a gene can code for several traits) that can code for C_1 and C_2. A_1 is dominant for C_1, and A_2 is dominant for C_2 (Table 13).

Genotypes	Traits			
	T1	T2	Sum	Product
A1A1	4	1	5	4
A1A2	4	4	8	16

A2A2	1	4	5	4

***Table 13**: Effects of traits on genotypes.*

Heterozygosity, for the complex trait, therefore presents a clear advantage without there being overdominance for each of the traits. Heterosis in the case of overdominance is unfixable.

2. Evolution of Heterosis Over Generations

In this chapter, we are interested in the heterosis manifested by the crossing of two lines or two populations that have no common origin. In this case, the study of the theory of heterosis will be based on the frequencies of the genes in the two lines.

2.1. Mean of a Population

Consider the effect of a single locus with two alleles whose values for the 3 genotypes are designated by +a, b, and -a (Fig. 14).

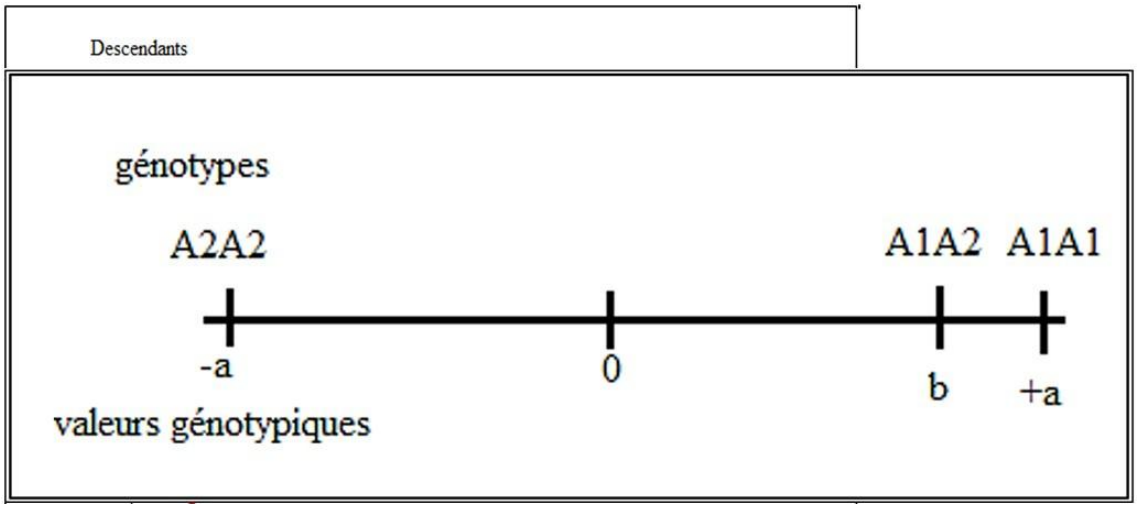


Figure 14: Graphical representation of the distribution of the three genotypes around the centre.

The value b of the heterozygote depends on the degree of dominance:

- If there is no dominance: b = 0,
- If A_1 dominates A_2: b > 0,

- If A_2 dominates A_1: $b < 0$,
- If dominance is complete: $b = +a$ or $b = -a$.
- If there is overdominance: $b > +a$ or $b < -a$,

The degree of dominance = b/a

Example: Consider two alleles, A_1 and A_2, controlling the height of barley plant stems. The average heights for the 3 genotypes are shown in the following table:

Table 14: *Average stem heights of the three barley genotypes.*

	Genotypes		
	A1A1	**A1A2**	**A2A2**
Stem height	100 cm	85 cm	50 cm

The population mean is determined according to the procedure explained in the following table:

Table 15*: Determination of the population mean.*

Genotypes	**Frequencies**	**Values**	**Frequencies x values**
A1A1	p^2	+a	ap^2
A1A2	2pq	b	2bpq
A2A2	q^2	-a	$-aq^2$
Population mean			a(p – q) + 2bpq

Population mean = $ap^2 + 2bpq + aq^2$

Population mean = $a(p^2-q^2)$ 2bpq

Population mean = a[(p+q)(p-q)] +2bpq

Population mean = a(p – q) + 2bpq

a(p – q) is attributable to homozygotes and 2bpq is attributable to heterozygotes

Example

Assuming that the dwarfism gene A_2 is present at a frequency of 0.1 and that the homozygotes are in the case of random mating (panmixia) (Fig. 15).

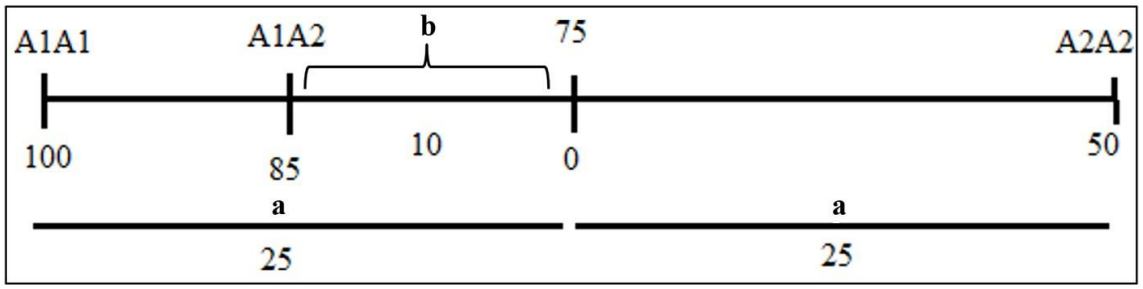


Figure 15: Graphical representation of the distribution of genotypes A1A1, A1A2, and A2A2 around their centre.

The values to be introduced into the equation are:

- Presence of the dwarfism gene q = 0.1
- Absence of the dwarfism gene p = 1 – 0.1 = 0.9

b = 10 cm ((100 + 50) / 2 = 75 b = 85 – 75 = 10)

a = 25 (the average number of units between 100 and 50 is 25, i.e., (100 – 50)/2 = 25).

The centered mean = 25(0.9 – 0.1) + 2(10)(0.9)(0.1)

The centered mean = 21.8 (this is a centered value and not the true mean)

This value of the mean is estimated from the point m, which is 75, so the real value of the population is:

Real mean = Centered mean + the value of the center

Real mean = 21.8 + 75 = 96.8 cm (this is the true mean)

2 .2. Proper Evolution of Heterosis

Consider two populations that we will call parental populations; each is the result of random mating, but they are not necessarily large. These parental populations are crossed to give an F_1 in the first-generation cross. The individuals of the F_1 are randomly crossed among themselves to give the F_2 in the second generation cross.

The value of heterosis presented by the F_1 and F_2 is measured by the difference with the value of the average parent, i.e., with respect to the average of the 2 populations.

$$\text{Heterosis F1} = mF1 - mP$$

Heterosis F2 = mF2 – mP

Where: mF1: mean value of the 1st generation, µF2: mean value of the 2nd generation, mP: the mean of parents parent

Consider the effect of a single locus with two alleles whose frequencies are p and q in one population and p' and q' in the other.

$$y = p - p' = q' - q$$

The calculation is simplified by putting the gene frequencies p' and q' in the second population in the form (p - y) and (q + y). We still denote by +a, b and -a the genotypic values, values that are assumed to be the same in both populations (epistasis not being considered).

The means of the parents Mp_1 and Mp_2 are:

$$\mathbf{Mp_1 = a\,(p - q) + 2bpq \ldots\ldots\ldots\ldots\ldots\ldots(1)}$$

$$\mathbf{Mp_2 = a\,(p - y - q\,y) + 2b(p - y)(q + y)}$$

Hence:

$Mp_2 = a\,(p - q - 2y) + 2b[pq + yp - yq - y^2]$ $Mp2 = a(p - q - 2y) + 2b[pq + y(p - q) - y^2]$

$Mp_2 = ½\,[a(p - q) + 2bpq + a\,(p - q - 2y) + 2b(pq + y(p - q) - y^2)]$

$Mp_2 = ½\,[a(p - q) + p - q - 2y) + 2bpq + 2b[(pq + y(p - q) - y^2)]$

$Mp_2 = ½\,[2a(p - q - -y) + 2bpq + 2b[(pq + y(p - q) - y^2)]$

$Mp_2 = [a(p - q - -y) + bpq + b\,[(pq + y(p - q) - y^2)]$

$Mp_2 = [a(p - q - -y) + b\,[(pq + pq) + y(p - q) - y^2)]$

$$\mathbf{Mp_2 = a\,(p - q - y) + b\,[(2pq + y(p - q) - y^2)] \ldots\ldots\ldots\ldots\ldots\ldots(2)}$$

When the two populations are crossed to produce the F1, individuals taken at random from one population are mated with individuals taken from the second (Table 16).

***Table 16**: F_1 generation obtained by random crossing of individuals from the two populations.*

Gametes de P2 et frequencies	**Gametes de P1 et frequencies**	
	A_1 (p)	**A_2 (q)**
$A_1(p - y)$	**A_1A_1** **p(p – y)**	**A_1A_2** **q(p – y)**
$A_2(q + y)$	**A_1A_2** **p(q + y)**	**A_2A_2** **q(q + y)**

Dahlia Fatima

This is equivalent to taking genes at random from the two populations; the F_1 is then constituted as follows (Table 17):

Table 17: *Allelic frequencies of the values of the individuals of the F_1*

Génotypes	A_1A_1	A_1A_2	A_2A_2
Frequencies	p(p – y)	2pq + y(p – q)	q(q + y)
Genotypic values	+a	b	-a

The genotypic means of the F_1 are then as follows:

$mF_1 = ap\ (p - y) + b[2pq + y(p - q)] + (-a)q(y + q)$

$mF_1 = a\ (p^2 - py) + b[2pq + y(p - q)] - a(q^2 + qy)$

$mF_1 = a\ [p^2 - py - q^2 - qy] + b[2pq + y(p - q)]$ And as $p^2 - q^2 = (p + q)(p - q)x$ and $q = 1$

$mF_1 = a\ [p - q - y(p + q)] + b[2pq + y(p - q)]$

$$\mathbf{mF_1 = a(p - q - y) + b[2pq + y(p - q)]\ \ldots\ldots\ldots\ldots\ldots\ldots(3)}$$

The value of heterosis expressed by the difference between the F1 and the average parent is obtained by subtracting equation 2 from equation 3; therefore, HF_1 (heterosis):

$$HF_1 = mF_1 - mp$$

$$\mathbf{HF_1 = by^2}$$

Where: b is Dominance, and y is the difference in frequencies between the 2 populations.

Thus the existence of heterosis as well as that of inbreeding depression depends on dominance. Loci without dominance (d = 0) show neither inbreeding depression nor heterosis.

The value of heterosis that manifests after crossing between 2 particular lines or 2 populations depends on the square of the difference in gene frequency (y) between the two populations. If the populations that are crossed do not differ in their gene frequencies, there will be no heterosis.

Heterosis will be maximum when one allele is fixed in one population and another allele in another population.

If we consider the combined effects of all loci for which the two parental populations differ and the genotypic values due to the separate loci combine additively, we can present the heterosis produced by the effect of all loci combined as the sum of their individual effects.

$$HF_1 = \Sigma dy^2$$

If some loci are dominant in one direction and some in another direction, their effects will tend to cancel out, and no heterosis will be observed, although there is dominance for the loci taken one by one.

The presence of heterosis in a cross is therefore, like inbreeding depression, dependent on directional dominance. The absence of heterosis is not sufficient to conclude that the loci taken individually do not exhibit dominance.

Now consider the F_2 of a particular cross of 2 parental populations (Table 18). This F_2 is obtained by randomly mating the individuals of the F_1. As a consequence of this random mating method, the genotypic frequencies of the F_2 will be the Hardy-Weinberg frequencies corresponding to the gene frequencies in the F1. The average genotypic value of the F2 can then be calculated by equation (1). The gene frequency in the F_1 is the average of the gene frequencies in the 2 parental populations.

***Table 18**: F_2 generation obtained by random crossing of F_1 individuals*

Frequencies	**Allèles**	
	A_1	**A_2**
P1	p	q
P2	p - y	q + y
Somme	2p - y	2q + y
Moyenne	(2p – y)/2	(2q + y)/2
Moyenne	p – ½ y	q + ½ y

If we replace p and q in equation (1), we have the average genotypic value of the F_2:

$MF_2 = [a(p - ½ y - (q + ½ y)] + 2b(p - ½ y)(q + ½ y)$ After simplification we will have:

$MF_2 = [a(p - ½ y - q - ½ y)] + 2b(pq + ½ py - ½ qy - ¼ y^2)$ $MF2 = [a(p - q - y)] + b(2pq + py - qy - ½ y^2)$

$MF_2 = a (p - q - y) + b [2pq + y(p - q) - ½ y^2]$

Then

$HF_2 = MF_2 - mp$

Dahlia Fatima

$$HF_2 = ½ by^2$$

$$HF_2 = ½ H F_1$$

The value of heterosis exhibited by the F_2 is only half that exhibited by the F_1.

3. Exercice

The means of stem heights for the 3 barley genotypes are : 120 cm for A1A1, 135 cm for A1A2, and 75 cm for A2A2.

Assuming that the frequency of the gigantism allele is 0.76, calculate the cantered mean and estimate the actual mean.

4. Exercice

Considering the same statement as exercise 1, Is there heterosis in this cross? If so, estimate its value.

5. Exercice

Considering the same statement as exercise 1, and if the heterozygous individuals of this population are crossed with heterozygous individuals from another population where the frequency of the A2 allele is 0.32, what will be the value of heterosis at F1 and F2?

6. Exercice

Considering the same statement as exercise 1 and 3, If the heterozygous individuals of population 2 are crossed with heterozygous individuals from a third population where the frequency of the A1 allele is 0.68, what will be the value of heterosis at F1 and F2?

VII.Chapter 6: The Value of an Individual in Crossing

Introduction

The value of an individual in crossing, also called crossing value, is a measure of the genetic contribution of an individual to the next generation when used as a parent in a breeding program. This value is based on the expected performance of its offspring for specific traits of interest.

The crossing value is used to evaluate and select the individuals that will contribute most favourably to the selection objectives. It is based on the idea that the performance of an individual in a given environment is the result of the interaction between its genes and its environment.

The question is to find the factors that determine the value of an individual as a parent. This analysis is of considerable importance in plant breeding, where every breeder knows empirically that a given genotype marks all its offspring in crossing with a given quality, while another, on the contrary, seems to contribute nothing.

This assessment is empirical and imprecise. The notions of heritability, genetic balance, and recombination aptitudes will allow us to specify the different relationships between parents and offspring.

1 . Genetic Balances

1 .1. Internal Balance and Relational Balance

The term balance here expresses an adaptation of the genetic constitution to a given environment. When the functioning of the polygenic sets is good, we say that the genotype is well balanced. On the contrary, we speak of an unbalanced structure when the set is poorly regulated, giving an impression of disharmony in development. According to Lerner, the balance of a polygenic set can be achieved in two ways (Fig. 16):

- By an internal balance playing within a genetic sequence that can lead to a balanced allelic arrangement,

- By heterozygosity at each locus that can lead to a good balance of relationship between alleles. In this case, it is a balance of relationship.

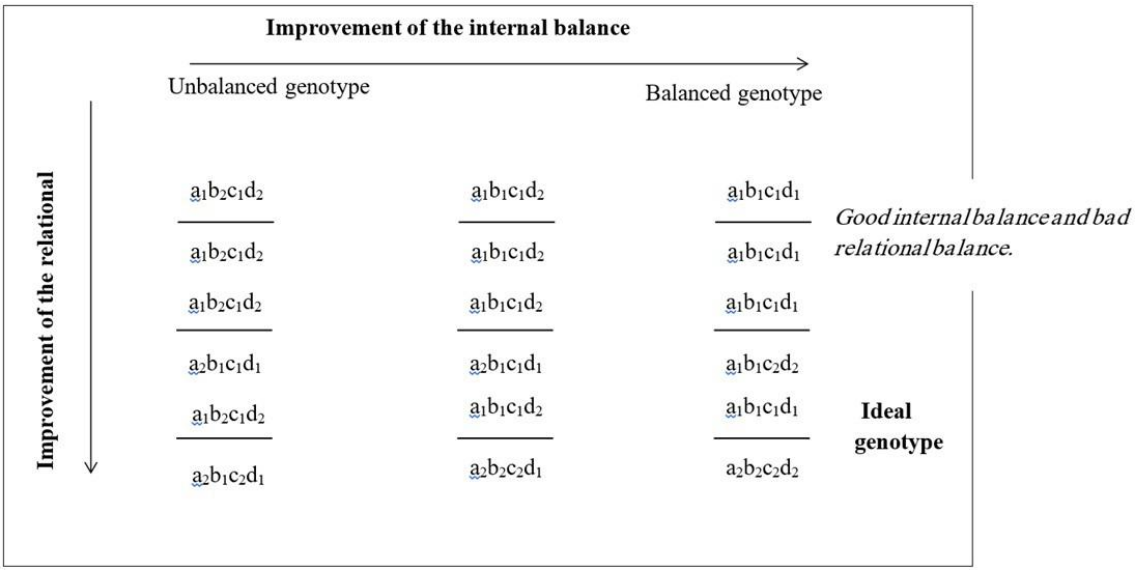


Figure 16: Different qualities of genetic balance.The letters a, b, c, and d represent the genes. These genes have a good balance of relationship when, in a locus, the two homologues have different indices. The good internal balance is represented by the set of lowercase letters all having the same indices.

1 .2. Biological Reality of Balance and Representation by the Set of Genotypes

The states of internal balance are constructed or destroyed by the play of crossing-over, which modifies the arrangement and allelic distribution along the chromosome.

The biological reality of internal balances rests strongly on the notion of linkages, that is to say, on the one hand, on chromosomal continuity and epistatic tension between the genes that follow one another along the chromosome and, on the other hand, on the existence of a continuity of expression between the juxtaposed duplicated sets.

The internal balance is determined by the quality of the interaction between homologous linkats. It is constructed or destroyed at each generation.

2. Combining Abilities

Two types of combining ability are distinguished: general combining ability and specific combining ability.

2.1. Definitions

For a given trait, the general combining ability (GCA) of a parental structure (individual, line, etc.) is estimated from the average value of the offspring when this parent is crossed with a certain number of partners. It is therefore the average of the gametic effects of an individual. It is therefore the measure of the value of the average gamete of a parent.

At the level of the entire XY progeny resulting from the crossing of two parental structures, we should therefore find the sum of the general combining abilities of the parents (XX) x (YY). In fact, the value of the XY individuals (the hybrids) shows a deviation from the predictions of additivity of the general combining abilities.

This deviation, which specifically characterises the cross (XX) x (YY), is called Specific Combining Ability (SCA). The specific combining ability is a characteristic of the zygote and not of the haploid parent.

Example: the stem height (h) of an individual resulting from a cross between 2 individuals (XX)x(YY).

$$H_{XY} = GCA_{xx} + GCA_{yy} + SCA_{XY}$$

2.2. Importance of Variances Related to Combining Abilities

The share taken by the effects of general and specific combining abilities in the differences in values between progenies is important to determine. It is the comparison of the variances related to general combining ability and specific combining ability that is essential. It determines, in fact, the strategy in the breeding program for a quality (Table 19).

Table 19: Selection strategies according to combining ability variances

Variance of general combining ability (GCA)		Variance of specific combining ability (SCA)	
Low	Little choice in parental formulas.	**Hight**	Make many crosses and then choose.
Hight	Effective choice in parental formulas.	**Low**	The choice of parents before crosses remains a priority.

3. Methods for Assessing the Values of an Individual in Crossing

Several crossing systems can be used to assess the quality of the genotype as parents. Depending on whether the species considered is self-fertilising or cross-fertilising, the techniques used are obviously not the same.

In a self-fertilising species, any systematic hybridisation requires an intervention (castration and pollination) which is often long and delicate, which limits the use of the tests described below.

3.1. Open Pollination (Maternal Progeny)

Reserved for allogamous species, this system consists of comparing the offspring of each genotype when it is freely pollinated. The comparison of the structures (homozygous or heterozygous) is done as follows (Fig. 17).

The limits of this method, whose great advantage is simplicity, are:

- The information it provides only concerns the general combining abilities and therefore essentially only concerns additive effects.
- The pollen population that fertilises one genotype is rarely the same as the one that fertilises another genotype, because the pollen of a plant is distributed around this plant according to a very marked gradient, whether the plant is anemophilous or entomophilous. Pollinating insects choose certain phenotypes that they preferentially pollinate. The adjustments of precocity between pollen maturity and stigma receptivity are very narrow in some species.

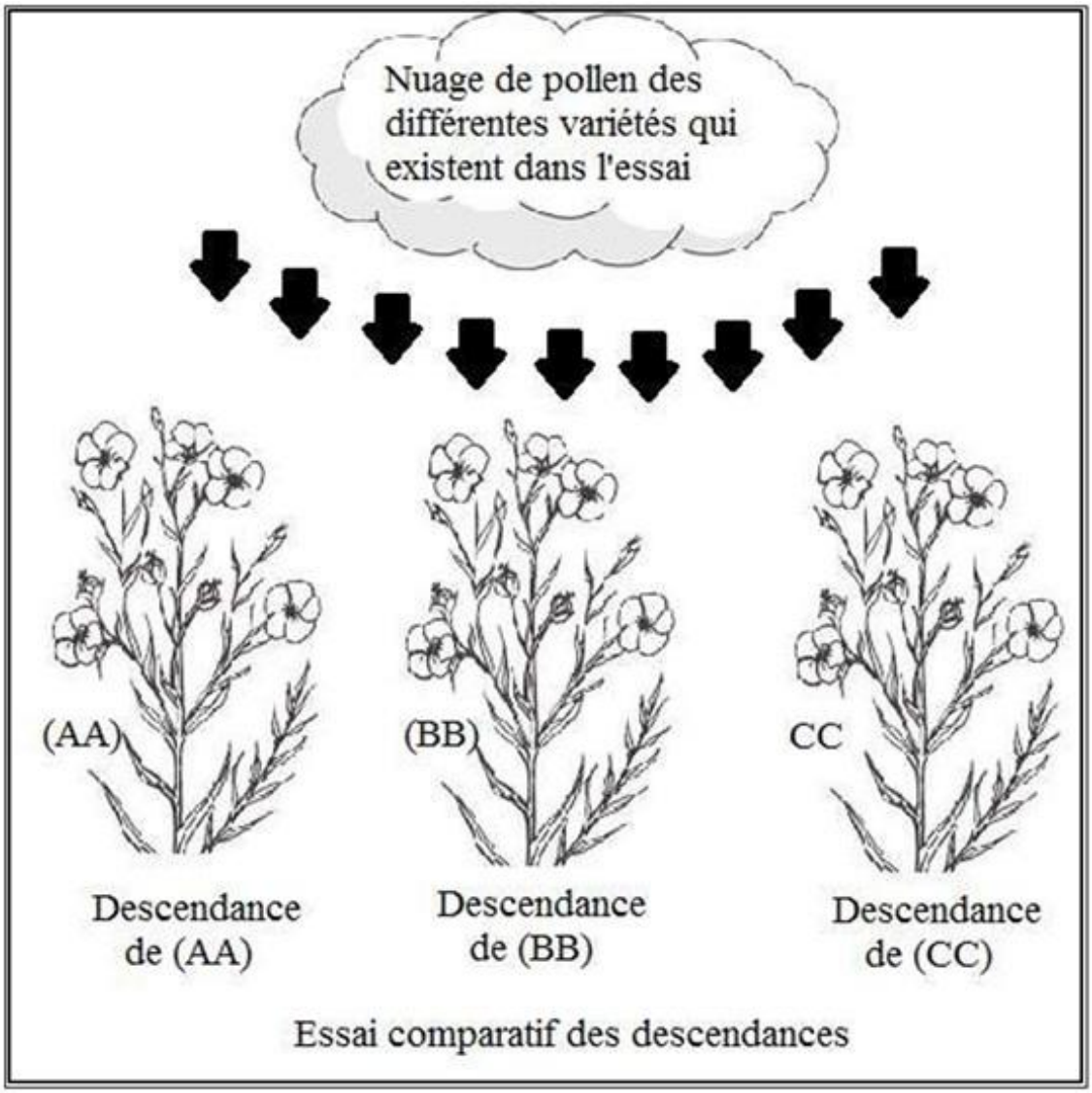


Figure 17 : Tests de descendances maternelles

3 .2. The top-cross test

The genetic structures to be studied (family, lines or clones) are pollinated by a common tester, which can be a variety, a population or a hybrid. In this case, the structures to be tested are generally represented by several individuals, which allows for better exploitation in space and time of the male gametic population. Hybridisation can be carried out in a controlled manner. The tester can be a well-defined and balanced structure which is therefore very stable.

When the tester is a broad-based genetic structure, the estimate obtained is a general combining ability. If, on the other hand, the tester has a narrow structure, the results of the top cross only reflect a specific combining ability with respect to the testers.

3 .3. The poly-cross test

In the polycross, it is the pollen population of the structures under study that constitutes the tester (Fig. 18). The polycross is the free fertilisation between (n) structures (clones, inbred lines, families, etc.) whose general combining abilities we want to compare.

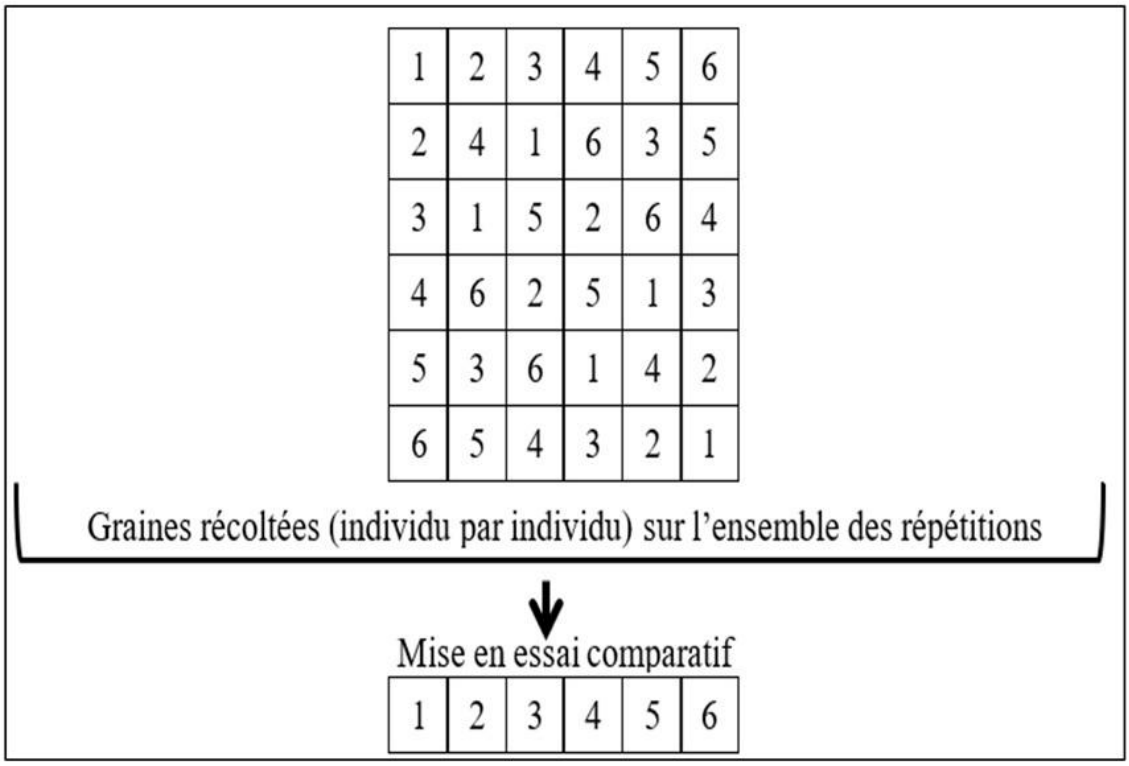


Figure 18: Systematic plan and polycross for 6 individuals.

In this cross, each female parent is artificially crossed (cross-pollination 1) with a pollen mixture from a sufficient number of male parents. To perform a progeny analysis, each female tree must be pollinated by the same pollen mixture and in the same quantity.

3 .4. Hierarchical cross

A certain number of parents P1, P2, ..., Pj are chosen at the first level, and each is crossed with different groups p1, p2, ..., pj, which may or may not have common elements and which constitute a second level (Fig. 19).

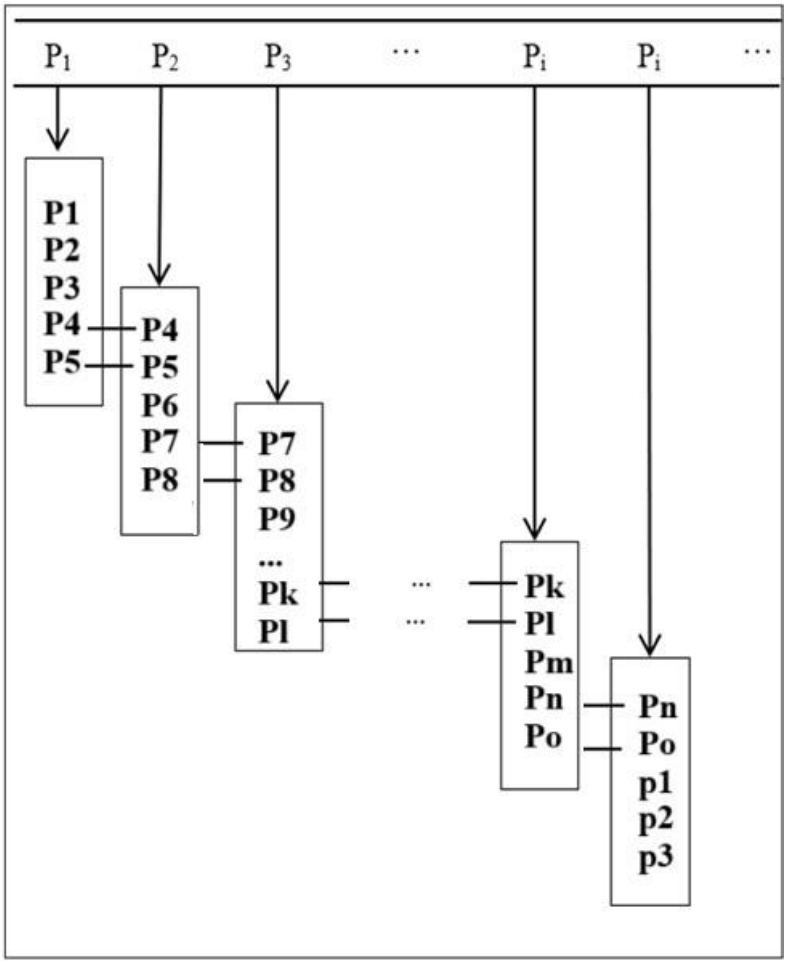


Figure 19: Hierarchical cross

3.5. Diallel cross

In all these tests, only the general combining ability (GCA) was accessible. Diallel crosses, on the other hand, allow for a more thorough interpretation in terms of combining ability.

The diallel cross applies to both self-pollinating and cross-pollinating species. It is a set of directed hybridisations between structures to be studied, systematically including a whole series of combinations (the seeds from each male parent being individualised on each female parent).

We call a diallel cross a system of systematic crosses combining

(n) parents. The set of operations can be written in the form of a table or diallel matrix (Fig. 20).

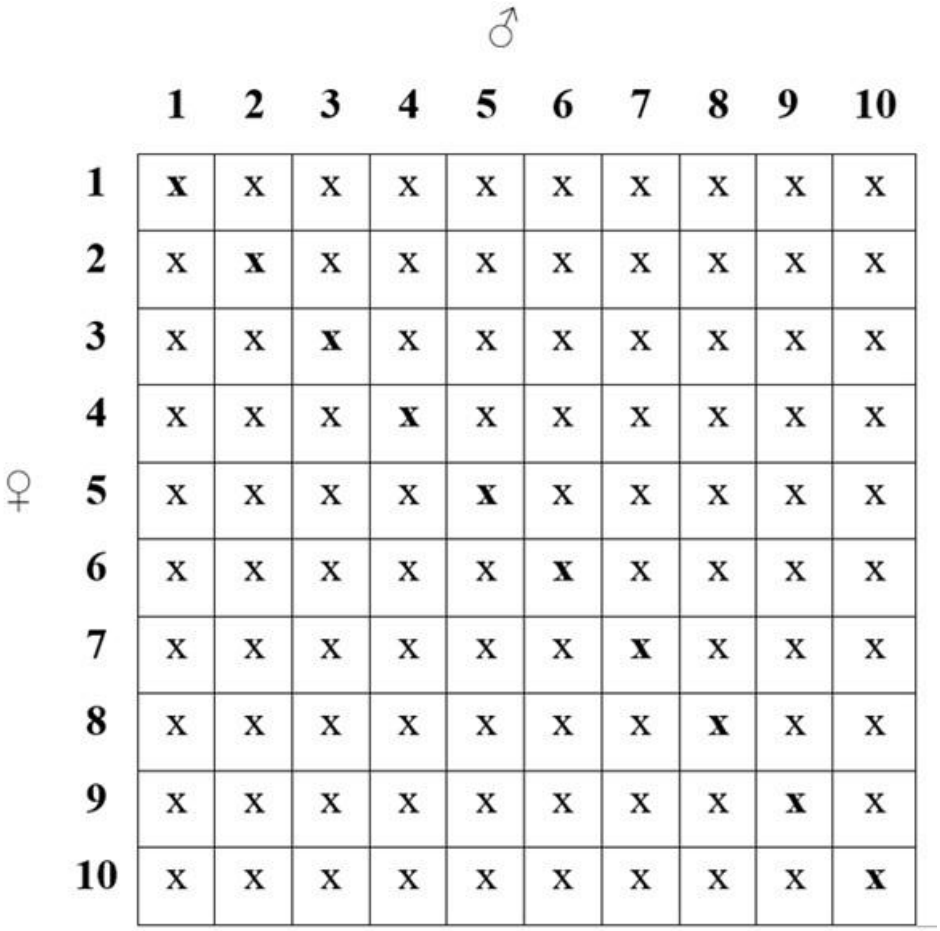


Figure 20: Complete diallel cross. All parents are crossed with each other parent in all possible combinations, including self-crossing.

The complete diallel of 10 parents therefore requires: 10 x 10 = 100 crosses

- The diallel can be complete and include the square of the pairwise combinations. The number of crosses = $(n)^2$, with (n) = number of parents or structures to be studied,
- The diallel can cover n(n – 1) hybrids. These are the inter-parent crosses (self-fertilizations are excluded).
- The diallel can include p(p+1)/2 types. A single direction of crossing, including self-fertilisation, it is also called the half-diallel (Fig. 21).
- The diallel can include p(p-1)/2 types. A single direction of crossing without self-fertilisation (Fig. 21).

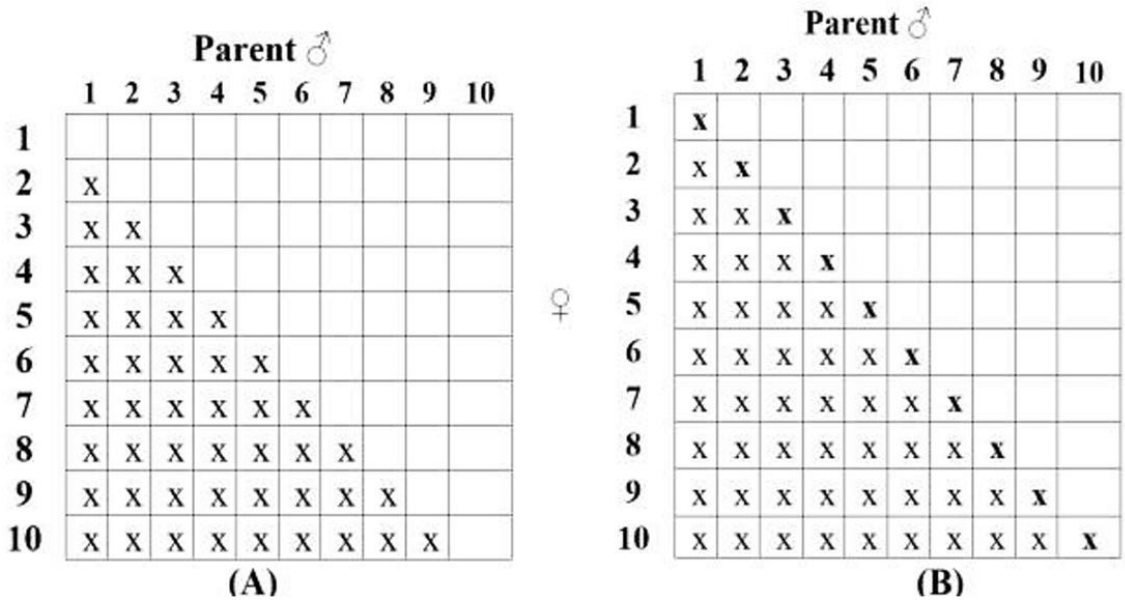


Figure 21: Half-diallel cross: (A) half-diallel cross not including self-fertilisations; (B) diallel cross including self-fertilisations.

4. Exercice

the main factor determining the value of an individual for a cross are:

☐	Additivity
☐	Heritability
☐	Combining ability
☐	Heterosis
☐	Genetic balance
☐	Polygenic inheritance

5. Exercice

Combining ability is the mean of the gametic effects

○	True
○	False

6. Exercice

The best methods for assessing the values of an individual in crossing is:

○	Top cross
○	Polycross
○	Diallele

VIII. Chapter 7: Quantitative trait loci (QTL)

Introduction

Quantitative Trait Loci (QTLs) are loci of quantitative traits. They are responsible for the control of quantitative effect traits, that is to say, having a continuous variation in their value (yield, size, etc.). QTLs are regions of the genome that are associated with the quantitative variation of complex traits, such as size, weight, milk production, disease resistance, nutrient content, meat quality, etc. QTLs are often sought using linkage or association analysis methods, which compare the genotypes and phenotypes of plant or animal populations.

QTLs are important for understanding the genetics of complex traits because they can help identify the genes that contribute to the quantitative variation of traits. Indeed, QTLs are genetic markers that allow the location of regions of the genome that are linked to a given trait. QTLs can thus help identify the genes that control a trait, as well as the interactions between genes and the environment.

QTLs can be detected using molecular markers such as SNPs (Single Nucleotide Polymorphisms) and microsatellites. Molecular markers are used to map QTLs by comparing the genotypes and phenotypes of plant or animal populations.

QTLs are used in many applications in agricultural, medical, and ecological research. For example, QTLs can be used for marker-assisted selection. QTLs can also be used to study the genetics of human and animal diseases, as well as to study the adaptation of species to changing environments.

1. Applications of QTLs in agricultural, medical, and ecological research

QTLs (Quantitative Trait Loci) have many applications in agricultural, medical, and ecological research. Here are some examples of QTL applications:

1.1. Agricultural Research

- **Marker-assisted selection**

QTLs can be used for marker-assisted selection, which allows the selection of plant or animal varieties with improved characteristics, such as disease resistance, product quality, and productivity. Marker-assisted selection reduces the time and costs associated with conventional selection based on phenotypic characteristics.

- **QTL mapping for disease and pest resistance**

QTLs can be used to map the genes involved in disease and pest resistance, which can help develop resistant varieties and reduce the use of pesticides.

- **Improving the nutritional quality of food**

QTLs can be used to map the genes that control nutrient content in plants, which can help develop varieties with increased content of nutrients such as vitamins and minerals.

1.2. Medical Research

- **QTL mapping for human diseases**

QTLs can be used to map the regions of the genome associated with complex diseases such as diabetes, Alzheimer's disease, and cancer. This can help understand the genetics of these diseases and develop new treatments.

- **Identification of QTLs involved in disease resistance in wild and domestic animals**

QTLs can be used to identify the genes that contribute to disease resistance in animals, which can help develop more resistant varieties and reduce the use of antibiotics.

- **Identification of QTLs involved in adaptation to environmental changes**

QTLs can be used to understand how species adapt to environmental changes such as climate change, which can help predict how species will react in the future.

1 .3. Ecological Research

- **Studying the adaptation of species to changing environments**

QTLs can be used to understand how species adapt to changing environments, which can help predict how species will react in the future.

- **Studying biodiversity**

QTLs can be used to understand the genetic diversity in wild populations, which can help protect biodiversity and preserve endangered species.

- **Studying the genetics of interactions between species**

QTLs can be used to understand how interactions between species are governed by genetics, which can help understand ecosystems and ecological processes.

1 .4. Limitations and challenges of using QTLs

Although QTLs (Quantitative Trait Loci) are a powerful method for studying complex characteristics, their use presents certain limitations and challenges:

- **High cost**: The identification and mapping of QTLs require significant resources, particularly in terms of financial costs, time, and qualified personnel.
- **Data complexity**: The genetic data generated during the study of QTLs is often very complex and difficult to interpret, requiring expertise in statistics and bioinformatics.
- **Environmental variability**: Phenotypic characteristics can be influenced by environmental factors such as temperature, humidity, and soil nutrients, which can make QTL mapping more difficult.
- **Linkage effects and complex genetic interactions**: QTLs are often linked to other QTLs and other genes, which can make it difficult to identify the specific effect of each QTL on the characteristic studied. In addition, interactions between genes can be complex and difficult to model.
- **Ethics**: The use of QTLs in marker-assisted selection raises ethical questions about the social and environmental implications of the genetic manipulation of plants and animals.
- **Intellectual property**: The use of QTLs in marker-assisted selection also raises intellectual property issues, as companies may claim ownership of the genetic markers used for selection.

2. Techniques for QTL detection

the most important techniques allowing the detection of ATLs are.

2.1. Linkage analysis and genetic mapping

The detection of QTL (Quantitative Trait Loci) can be performed using two main techniques: linkage analysis and genetic mapping.

a) Linkage analysis

Linkage analysis is a technique that seeks to establish a relationship between a quantitative trait and a specific region of the genome. It generally involves the comparison of genetic markers with phenotypic characteristics in populations of inbred lines or families of animals.

The technique consists of establishing a correlation between the presence of genetic markers and the presence of a specific trait. If a marker is present in individuals who exhibit the trait, then the marker is considered to be linked to the trait. In other words, the frequency of a specific marker is compared to the frequency of the phenotypic characteristic being studied, which allows determining if these two characteristics are correlated.

b) Genetic mapping

Genetic mapping is a technique that uses genetic markers to map the position of QTLs on the genome. This technique uses genetic markers to map the positions of genes on chromosomes. These markers can be simple DNA sequences or DNA variants such as SNPs (Single Nucleotide Polymorphisms).

Genetic markers are mapped using genetic maps. These maps are created by examining the frequency of recombinations that occur during meiosis, during the formation of gametes. By using these maps, QTLs can be mapped in relation to known genetic markers.

The combination of linkage analysis and genetic mapping allows for the precise location of QTLs and the identification of genetic markers that are closely linked to these QTLs. These markers can then be used for marker-assisted selection, which allows for the selection of individuals with desired characteristics using genetic markers rather than phenotypic characteristics.

Example:

The location of QTLs involved in the variation of wheat grain hardness (Fig. 22).

One of the factors affecting the baking quality of cultivated wheat is the hardness of the grains. The establishment of genetic maps on wheat has allowed the identification of several QTLs involved in the variation of grain hardness. One of the major QTLs is located on chromosome 5. The associated infographic represents a genetic map of a part of chromosome 5. The abscissa shows the distances between the RFLP markers along this chromosome arm. The size of the bars is associated with a test for the presence of the QTL: the largest bar corresponds to the most statistically probable position of the QTL. Thus, a grain hardness QTL was found near the Xmta9 marker. This is a marker of the QTL.

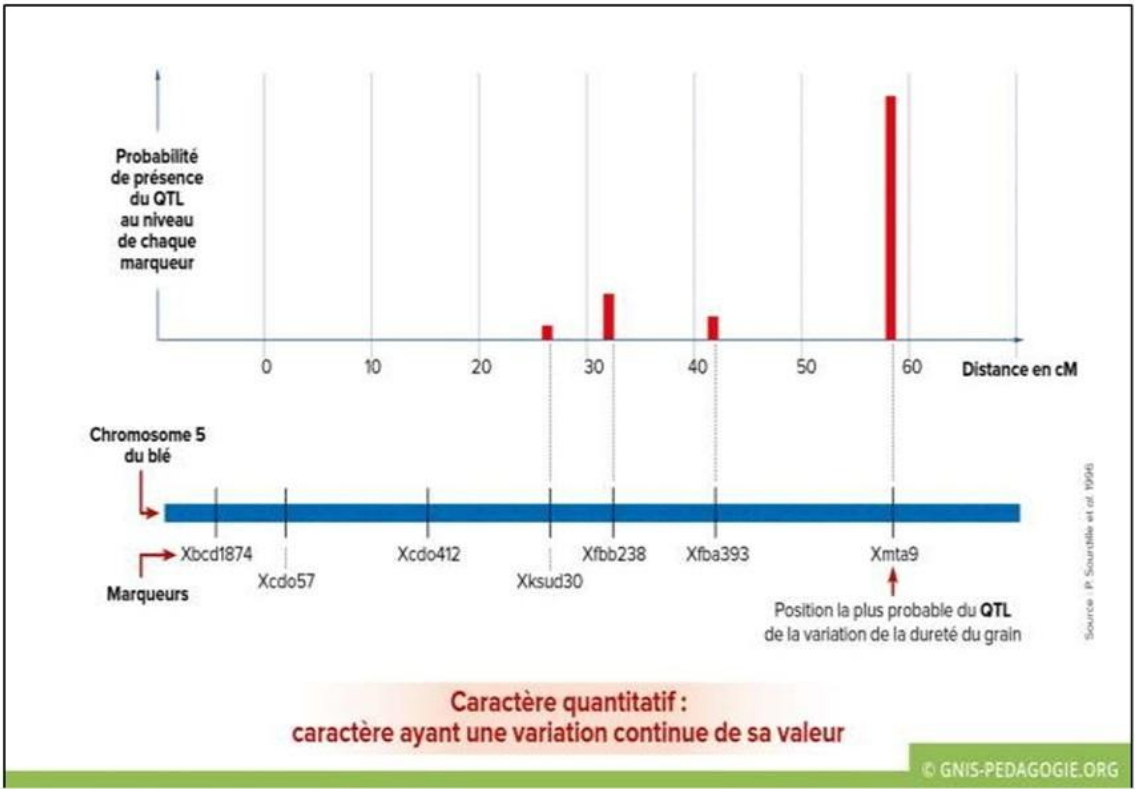


Figure 22: The location of QTLs involved in the variation of wheat grain hardness

2 .2. Association analysis and genome-wide association study

In addition to linkage analysis and genetic mapping, two other techniques for the detection of QTL (Quantitative Trait Loci) are widely used: association analysis and genome-wide association study.

a) Association analysis

Association analysis is a technique that uses genetic markers to study the relationships between quantitative traits and genetic variations. It differs from linkage analysis in that it uses a population of unrelated cases and controls to study the relationships between genetic markers and quantitative

traits. Genetic markers are compared to the presence or absence of the characteristic being studied to determine if there is a significant association between them.

Association analysis is useful for the identification of QTLs because it allows the identification of genetic variations associated with quantitative traits in large populations. However, this technique is limited by the need to use genetic markers that are in strong LD (linkage disequilibrium) with the QTLs, which can make it difficult to identify the actual genetic variations that are responsible for the QTLs.

b) Genome-wide association study

The genome-wide association study (GWAS) is a technique that uses genetic markers covering the entire genome to identify genetic variations associated with quantitative traits. This technique is similar to association analysis but uses a large number of genetic markers to maximise genome coverage.

GWAS have allowed the identification of QTLs associated with many quantitative traits in various species, including humans, plants, and animals. This technique also allows the identification of genes and molecular pathways involved in quantitative traits, which can provide important information on the underlying molecular mechanisms.

However, the genome-wide association study is also limited by the high cost of genotyping analyses and the need to use large populations to identify significant associations with sufficient statistical power. In addition, this technique does not allow for precise mapping of QTLs and determining the interaction effects between different QTLs.

2 .3. Use of molecular markers for QTL mapping

The use of molecular markers for QTL (Quantitative Trait Loci) mapping is a widely used technique for the detection of QTL in animal and plant species. This technique is based on the construction of a genetic map that links QTLs to molecular markers.

The construction of a genetic map is done in two main steps: genotyping and mapping. Genotyping consists of determining the genotypes for molecular markers in a population of recombinant lines. Mapping then consists of placing the molecular markers on a genetic map and linking them to the QTLs.

There are two types of genetic maps: marker-based genetic maps (or linkage maps) and physical maps (or position maps). Linkage maps are constructed from molecular markers that are closely linked to

QTLs, while physical maps are constructed from molecular markers that are associated with known physical positions in the genome (Fig. 23).

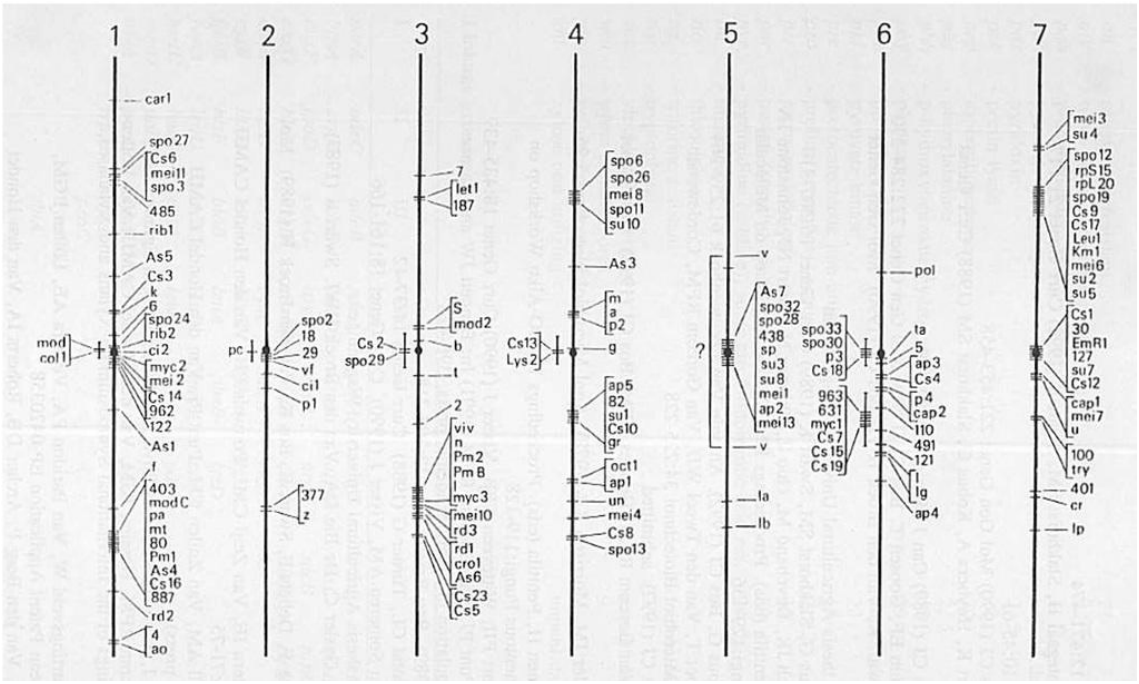


Figure 23: Example of a genetic map: Genetic map of Podospora anserina.

Once the genetic map is established, QTLs can be mapped using statistical methods such as regression analysis and analysis of variance. These methods allow the determination of the contribution of each QTL to the variation of the quantitative trait studied.

The use of molecular markers for QTL mapping has several advantages over classical methods based on phenotypic selection. First, this technique allows QTLs to be located precisely on the genetic map, which facilitates their molecular characterisation and cloning. In addition, it allows the mapping of QTLs that are difficult to detect by classical phenotypic selection methods, such as minor effect QTLs.

However, the use of molecular markers for QTL mapping also has limitations, including the need to construct an accurate genetic map and the complexity of interactions between QTLs, which can be difficult to detect. In addition, this technique does not allow the identification of the molecular mechanisms underlying QTLs, which requires further studies.

2.4. High-throughput genotyping techniques for QTL mapping

High-throughput genotyping techniques are increasingly used tools for QTL (Quantitative Trait Loci) mapping. These techniques allow the simultaneous genotyping of thousands of molecular markers, which allows the construction of genetic maps with high resolution and the detection of QTLs with increased accuracy.

The two main high-throughput genotyping techniques used for QTL mapping are next-generation sequencing (NGS) and DNA microarrays.

a) Next-generation sequencing (NGS)

It allows rapid sequencing of large amounts of DNA at an affordable cost. This technique can be used for genotyping molecular markers such as SNPs (Single Nucleotide Polymorphisms) and SSRs (Simple Sequence Repeats). The sequencing data can then be analysed to detect QTLs using statistical methods such as regression analysis.

The commercialisation of NGS technologies since 2005 has revolutionised the dimension of genetic analyses in recent years by a major change in the scale of sequencing capabilities.

Having quickly found numerous applications in the field of research, particularly for the identification of new genes involved in monogenic diseases, NGS has gradually been validated for applications in genetic diagnostics.

NGS relies on the massive generation of sequence data obtained by successive cycles of nucleotide incorporation and thus the emission of signals which are then converted into sequence information. Different technologies currently exist, notably based on parallel sequencing of millions of DNA molecules, with an ever-increasing increase in sequencing capacity associated with a progressive decrease in costs, and new approaches are under development (in particular the direct sequencing of single DNA molecules). Schematically, the NGS process consists of multiple steps of data generation and analysis, taking into account sequencing quality criteria (in particular the analysis of "coverage" and "read depth" of the sequence of interest), which are presented synthetically in Figure 24 in association with commonly used terms.

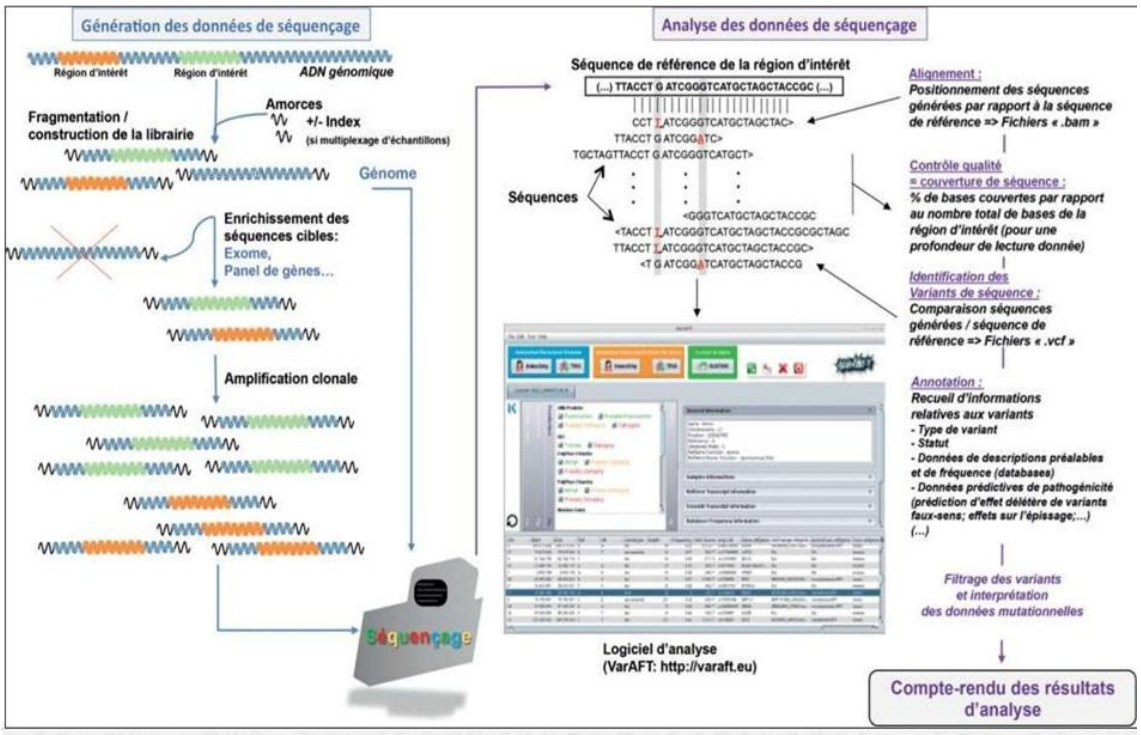


Figure 24: Principle of the next-generation sequencing (NGS) method

b) DNA microarrays

DNA microarrays allow the simultaneous genotyping of thousands of molecular markers, such as SNPs and SSRs, at an affordable cost. These chips are available for many animal and plant species, making them widely used for QTL mapping. DNA microarray data can be analysed to detect QTLs using statistical methods such as genome-wide association analysis.

The principle of a DNA microarray lies in the recognition, i.e., hybridisation, between two complementary single-stranded DNA molecules (Fig. 25). The sample (DNA or RNA), fluorescently labelled, is brought into contact with the chip carrying several thousand probes which are DNA fragments or oligonucleotides of known sequence. After washing the non-specifically bound material, the signal is quantified at each probe. Its value will depend on the concentration of labelled molecules complementary to the probe in the sample and the degree of complementarity (the percentage of identity) with the probe.

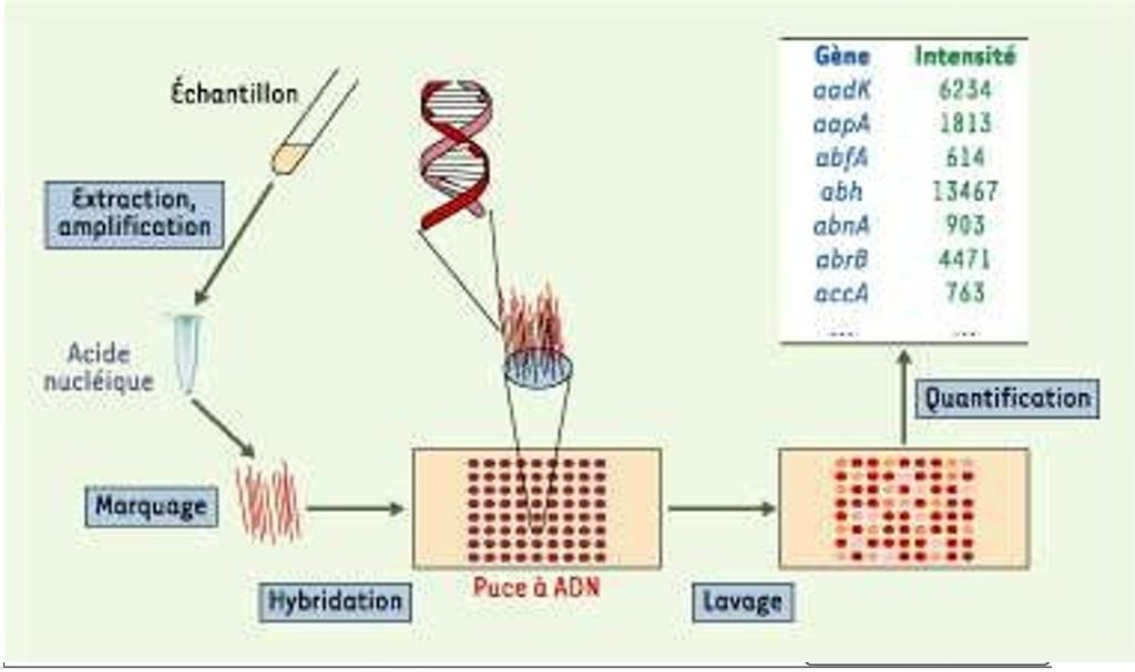


Figure 25: Nucleic acid analysis by DNA microarray.

Currently, the most common use of DNA microarrays concerns the quantification of messenger RNAs from a sample (transcriptome) in order to compare the transcription profiles of two samples obtained under different growth conditions.

In the context of bacterial identification and typing, the use of DNA chips is different: these chips allow the detection of a DNA sequence in a mixture, the identification of a sequence polymorphism (SNP) and the re-sequencing of a DNA fragment (Fig. 26). DNA chips serve as:

- **Detection tools**

DNA chips are used to detect a PCR product, replacing gel migration by validating the specificity of the amplification by its complementarity with the probe. For the analysis of a single sample, this procedure is not very competitive compared with real-time PCR or capillary electrophoresis. On the other hand, in the case of a complex sample, DNA chips enable a large number of fragments to be analysed simultaneously. Microarrays that can identify the bacteria present in a sample on the basis of 16S ribosomal RNA sequence analysis are being developed by several institutions (Fig. 26A). Total DNA is extracted from a sample and all DNAs encoding 16S RNAs are amplified, using universal primers, and hybridized on chips carrying specific probes for the 16S RNA of the species sought. This tool is developed for bioterrorism surveillance with a set of probes specific to potentially used pathogens.

Dahlia Fatima

- **For re-sequencing**

Knowledge of the complete sequence of a genome is the ultimate level of typing, but, in the current context, sequencing each isolate is not feasible. Oligonucleotide chips are therefore developed to obtain partial information on bacterial genomes (Fig. 26B). Chips are used for MLST typing of Staphylococcus aureus. The seven loci analyzed are amplified by PCR and hybridized to the chip instead of being sequenced one by one. Improving the sensitivity of the technique should allow the use of total genomic DNA to detect a set of polymorphic positions distributed throughout the genome. It would then be possible to have a global view of the genome and to point to particular mutations, such as those leading to antibiotic resistance.

- **For genomic characterization**

DNA chips are also used to characterize the variable part of a clone's genome. These "biodiversity" chips carry probes corresponding to genes that are not present in all isolates of a species (Fig. 26C). They are established from the comparison of the sequences of several genomes. By a single hybridization experiment, these chips make it possible to establish a true fingerprint corresponding to the genes present or absent in a clone. Thus, unlike most typing methods, DNA chips provide functional information on the nature of the genes that differentiate two isolates. These results can be compared to phenotypic data on the strains, particularly in relation to their virulence.

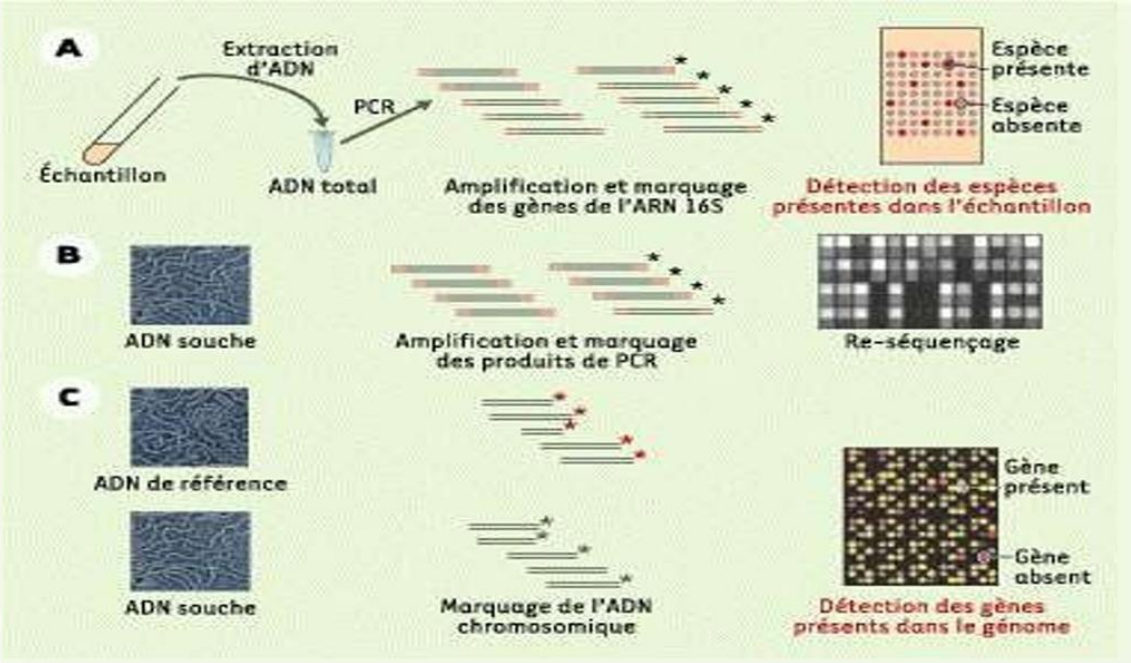


Figure 26: Three applications of DNA microarrays in microbiological analysis.(A: Analysis of a mixed bacterial population, B: Affymetrix re-sequencing chip, C: Detection of specific chromosomal regions of an isolate).

Dahlia Fatima

c) Advantages and limitations of high-throughput genotyping techniques for QTL mapping

High-throughput genotyping techniques for QTL mapping have several advantages over conventional techniques based on limited molecular markers. First, these techniques allow for more complete genome coverage, which makes it possible to detect QTLs that would be difficult to detect with limited molecular markers. In addition, these techniques allow for high-resolution analysis of genome structure and genetic variations, which facilitates the identification of QTLs associated with complex traits.

However, high-throughput genotyping techniques for QTL mapping also have limitations, including the need for significant computing resources for data analysis and the need to verify results with independent validation techniques. In addition, these techniques do not allow for the identification of the molecular mechanisms underlying QTLs, which requires further study.

3. Statistical analysis of QTLs

3.1. Regression and analysis of variance models

Statistical analysis of QTLs (Quantitative Trait Loci) relies on the use of regression and analysis of variance (ANOVA) models to identify regions of the genome that are associated with phenotypic variations.

Regression models are widely used for QTL analysis. These models allow analysis of the relationship between a phenotypic trait and the molecular markers that are associated with that trait. The most commonly used regression model is the simple linear model, which can be written as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where: Y is the phenotypic trait, X is the genotype of the molecular marker, β_0 and β_1 are the regression coefficients and ε is the residual error.

Analysis of variance (ANOVA) is another commonly used method for QTL analysis. ANOVA allows the decomposition of total phenotypic variation into components of variation associated with genetic and environmental effects. This method can be applied to data from a biparental cross, in which individuals are generated by crossing two different parental lines.

ANOVA for QTLs can be performed using a single or multiple QTL model. The single QTL model can be written as follows:

$$Y_{ij} = \mu + G_i + E_j + \varepsilon_{ij}$$

Where: Y_{ij} is the phenotypic value of individual i of group j, μ is the overall mean, G_i is the genetic effect of individual i, E_j is the environmental effect of group j and ε_{ij} is the residual error.

The multiple QTL model can be written as follows:

$$Y_{ij} = \mu + \sum \beta_k Z_{ik} + E_j + \varepsilon_{ij}$$

Where: Z_{ik} is the value of molecular marker k in individual i, βk is the effect of molecular marker k and the symbol ∑ indicates the sum over all molecular markers. This model allows for the detection of multiple QTLs simultaneously.

Regression and ANOVA models for QTLs have advantages and disadvantages. Regression models are simple and easy to use, but they do not account for environmental effects and do not allow for the detection of multiple QTLs simultaneously. ANOVA models allow for environmental effects and the detection of multiple QTLs simultaneously, but they require a large number of samples and can be more complex to implement.

The statistical analysis of QTLs relies on the use of regression models and ANOVA to identify regions of the genome that are associated with phenotypic variations. The choice of model depends on the research question and the available data.

3 .2. Linear Mixed Models for QTL Detection

Linear mixed models are an advanced statistical analysis method for the detection of QTLs (Quantitative Trait Loci). These models are particularly useful for taking into account environmental effects, kinship effects, and correlations between molecular markers. Linear mixed models combine a fixed component, which describes the main effects of molecular markers, and a random component, which describes the random effects of the environment and kinship.

The linear mixed model can be written as follows:

$$Y = X\beta + Zu + e$$

Where: Y is the vector of phenotypic values, X is the design matrix for the fixed effects of molecular markers, β is the vector of coefficients for the fixed effects, Z is the design matrix for the random effects of kinship and environment, u is the vector of random effects, and e is the vector of residual errors.

Linear mixed models allow for the estimation of the genetic effects of molecular markers while taking into account environmental and kinship effects. These models are particularly useful for data from crosses of related lines, in which individuals share a variable proportion of their DNA.

A commonly used method for QTL analysis using linear mixed models is the EMMA (Efficient Mixed-Model Association) method. This method allows for the consideration of correlations between molecular markers and environmental effects, which can improve the accuracy of QTL detection. It is also effective for high-throughput genotyping data, which can include thousands or millions of molecular markers.

Linear mixed models are an advanced method for the statistical analysis of QTLs. These models allow for the consideration of environmental effects, kinship effects, and correlations between molecular markers, which can improve the accuracy of QTL detection. The EMMA method is a commonly used method for QTL analysis using linear mixed models.

3 .3. Correction for Population and Environmental Effects

When detecting QTLs (Quantitative Trait Loci), it is important to consider environmental and population effects to avoid false positives and improve the accuracy of detection. Environmental effects can include variations due to growing conditions, climatic variations, and differences in agricultural practices, while population effects can be due to population structure or kinship.

Correction for population effects can be achieved using different approaches, such as adjusting for population structure using subsampling methods, adding covariates to statistical analysis models, or using linear mixed models that account for kinship effects.

Correction for environmental effects can also be achieved using different approaches. For example, plot blocks can be defined to control environmental variations and minimize the impact of measurement errors, or environmental covariates can be added to statistical analysis models.

There are also methods to correct for both environmental and population effects. One such method is the Quality Control (QC) method, which uses genetic markers to detect experimental errors, genotyping errors, and mislabeled individuals, in order to eliminate noisy data and improve the accuracy of the analysis.

Correcting for population and environmental effects is essential for accurate QTL detection. Approaches for correcting for population and environmental effects include adjusting for population structure, adding covariates to statistical analysis models, using linear mixed models, and using

quality control methods. Combining these approaches can improve the accuracy of the analysis and allow for a better understanding of the mechanisms underlying the quantitative traits studied.

3.4. Assessing the Statistical Significance of QTLs

Once QTLs (Quantitative Trait Loci) have been detected, it is important to determine their statistical significance to assess their biological relevance. Different approaches are used to assess the statistical significance of QTLs.

One of the most common approaches is to use significance thresholds based on statistical tests such as the Likelihood Ratio Test (LRT) or the Wald test. These tests allow for the comparison of the goodness of fit between a model with the QTL and a model without the QTL, and to determine whether the difference in fit is significant.

Significance thresholds are often established using simulations, using simulated data to estimate the null distribution of the test statistic under the assumption that there is no QTL. This approach allows for controlling the false positive rate and establishing appropriate significance thresholds for each study.

Another approach to assess the statistical significance of QTLs is to use confidence intervals (CIs) to estimate the precision of the QTL's position and effect. CIs allow for determining whether the estimated positions and effects are precise enough to be considered significant.

Finally, it is also possible to use permutation approaches to assess the statistical significance of QTLs. This approach consists of permuting the genotypes of individuals while maintaining the phenotypes and positions of genetic markers, in order to estimate the null distribution of the test statistic. QTLs detected in the real data are then compared to the null distribution to determine their statistical significance.

The evaluation of the statistical significance of QTLs is essential to determine their biological relevance. Different approaches are used to assess the statistical significance of QTLs, including the use of significance thresholds based on statistical tests, the use of confidence intervals, and the use of permutation approaches. The combined use of these approaches can improve the reliability of QTL detection and allow a better understanding of the mechanisms underlying the quantitative traits studied.

4. Applications of QTLs in agricultural researchApplications of QTLs in agricultural research

4.1. Marker-assisted selection for the selection of complex traits

QTLs (Quantitative Trait Loci) have many applications in agricultural research, particularly for marker-assisted selection (MAS). This technique allows the identification and selection of individuals carrying molecular markers associated with traits of interest, such as disease resistance, grain quality, or growth.

The use of molecular markers allows rapid identification of individuals carrying markers associated with traits of interest, without having to perform costly and time-consuming phenotyping tests. This technique also allows the selection of individuals carrying markers for complex traits, which are generally difficult to select by conventional methods.

Marker-assisted selection is used in many crops to improve crop quality and yield. For example, in rice cultivation, QTLs have been identified for disease resistance, salinity tolerance, and grain quality. By using molecular markers associated with these QTLs, breeders can quickly identify individuals carrying markers for these traits, which accelerates the selection process.

In livestock breeding, marker-assisted selection is used to improve production characteristics, such as growth, meat quality, and disease resistance. QTLs have been identified for these traits in many animal species, such as cattle, pigs, and chickens. By using molecular markers associated with these QTLs, breeders can quickly identify animals carrying markers for these traits, which accelerates the selection process and improves livestock productivity and profitability.

Example of Bt gene introgression in corn

A series of backcrosses is performed between an elite line and a genetically transformed line. The latter is characterized by a single insertion of the Bt gene (resistance transgene to the corn borer) on chromosome 1.

During backcrosses, individuals carrying the Bt gene can be selected, and having recombined the smallest fragment of the donor line around the Bt gene. Indeed, thanks to molecular markers, individuals are selected for markers close to the gene, the genotype of the elite line (Fig. 27).

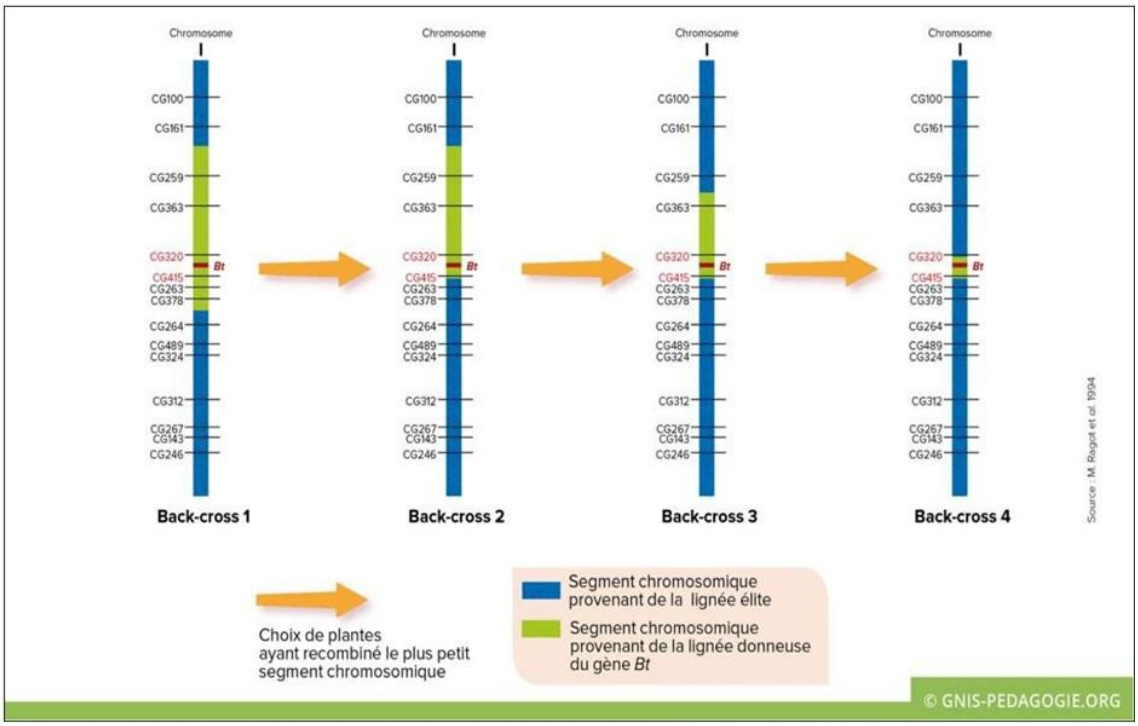


Figure 27: Molecular Marker-Assisted Selection, Example of Bt gene introgression in corn

In addition, it is also possible to accelerate the return to the elite parent thanks to molecular markers distributed throughout the genome. At each backcross, individuals with the most fragments from the recurrent elite parent will be chosen.

In the fourth generation, a near-isogenic line of the elite line is obtained, that is, identical to the starting elite line, but having integrated the Bt gene. There is therefore a gain in time, and therefore in efficiency.

4 .2. QTL mapping for resistance to diseases and pests

QTL (Quantitative Trait Loci) mapping for resistance to diseases and pests is an important application of QTLs in agricultural research. This technique allows the identification of genomic regions associated with resistance to diseases and pests, which allows the development of plant varieties resistant to these biotic stress factors.

Resistance to diseases and pests is an important characteristic for agricultural production, as it reduces crop losses and improves crop quality. Diseases and pests can cause significant damage to crops, which can lead to a decrease in productivity and profitability. QTL mapping for resistance to diseases

and pests allows the identification of genomic regions that control these characteristics, which allows the development of plant varieties resistant to these stress factors.

For example, QTLs have been identified for resistance to wheat leaf rust and corn mosaic. These QTLs have been used to develop wheat and corn varieties resistant to these diseases, which has reduced crop losses and improved crop quality. Similarly, QTLs have been identified for resistance to root rot in tomatoes and peppers, which has allowed the development of varieties resistant to this disease.

By using QTL mapping, it is also possible to understand the molecular mechanisms underlying resistance to diseases and pests. For example, QTLs have been identified for resistance to wheat leaf rust, which are associated with disease resistance genes. This knowledge can be used to develop disease and pest control strategies that specifically target these resistance genes.

QTL mapping for resistance to diseases and pests is an important application of QTLs in agricultural research. This technique allows the identification of genomic regions associated with resistance to diseases and pests, which allows the development of plant varieties resistant to these biotic stress factors. This improves crop productivity and profitability, while reducing crop losses and improving crop quality.

4 .3. QTL Mapping for Crop and Animal Quality

QTL (Quantitative Trait Loci) mapping is also widely used in agricultural research to improve the quality of crops and animals. QTLs associated with quality characteristics such as protein content, lipid content, sugar content, fruit size, and meat quality have been identified in various species.

Using QTL mapping, scientists can identify the genes that control quality characteristics and the regions of the genome that are associated with these genes. This information can be used to improve crop varieties and animal breeds to enhance their quality and economic value.

For example, QTLs have been identified for protein content in soybeans, which has led to the development of high-protein soybean varieties for use in animal and human food. Similarly, QTLs have been identified for oil content in corn, which has led to the development of high-oil corn varieties for use in the food industry.

Regarding animals, QTLs have been identified for beef quality, including tenderness, juiciness, and flavor. These QTLs have been used to develop cattle breeds that produce higher quality meat for the premium food market.

In addition, QTL mapping has also been used to improve the quality of fruits and vegetables. QTLs have been identified for the size, colour, and flavour of fruits and vegetables, which has led to the development of higher quality fruit and vegetable varieties for consumers.

QTL mapping for crop and animal quality is an important application of QTLs in agricultural research. This technique allows for the identification of genomic regions associated with quality characteristics, enabling the development of crop varieties and animal breeds that meet the needs of consumers and the food industry. This improves the economic value of crops and animals while meeting consumer needs for food quality.

4 .4. Improving the Nutritional Quality of Food

QTL (Quantitative Trait Loci) mapping is also used in agricultural research to improve the nutritional quality of food. By identifying the regions of the genome that are associated with nutrient content, scientists can develop crop varieties that have a high content of essential nutrients for human health.

For example, QTLs have been identified for the omega-3 fatty acid content in salmon, which has led to the development of salmon varieties with high omega-3 fatty acid content to meet consumer demand for healthier foods. Similarly, QTLs have been identified for vitamin C content in citrus fruits, which has led to the development of citrus varieties with high vitamin C content to meet consumer demand for nutrient-rich foods.

In addition, QTL mapping has also been used to improve the content of essential nutrients such as iron and zinc in staple crops, such as rice, wheat, and corn. This is particularly important for regions of the world where malnutrition is prevalent and where staple crops constitute a large part of the daily diet.

QTL mapping for improving the nutritional quality of food is an important application of QTLs in agricultural research. This technique allows for the identification of genomic regions associated with nutrient content, enabling the development of crop varieties that have a high content of essential nutrients for human health. This contributes to improving the quality of food and combating malnutrition in regions of the world where it is widespread.

5. Applications of QTLs in medicine and ecology

5.1. QTL mapping for human diseases

QTL (Quantitative Trait Loci) mapping is a technique used in genetics to identify regions of the genome that are associated with quantitative traits, such as height, weight, or blood pressure. This technique is also used in medicine to identify regions of the genome that are associated with human diseases.

QTL mapping is particularly useful for complex diseases, such as diabetes, cardiovascular diseases, and autoimmune diseases, which are influenced by many genetic and environmental factors. By identifying the regions of the genome that are associated with these diseases, scientists can better understand the genetic mechanisms that underlie these diseases and develop new, more targeted treatments.

QTL mapping is generally performed by analysing genetic data from families affected by a specific disease. Scientists examine the genetic differences between family members with the disease and family members who do not have the disease to identify the regions of the genome that are associated with the disease.

Once the regions of the genome associated with the disease are identified, scientists can conduct further studies to understand the genetic mechanisms underlying the disease. This may include identifying the specific genes that are involved in the disease, as well as genetic mutations that increase the risk of developing the disease.

Using this information, scientists can develop genetic tests to predict the risk of developing the disease, as well as more targeted gene therapies and drugs. QTL mapping is therefore an important technique for understanding and treating complex diseases in humans.

5.2. Identification of QTLs involved in disease resistance in wild and domestic animals

QTL mapping is also used in animal research to identify regions of the genome that are associated with traits of interest, such as disease resistance in wild and domestic animals.

Disease resistance is an important trait in agriculture, as it can have a significant impact on animal health and production. QTL mapping can be used to identify regions of the genome that are associated

with disease resistance in animals, which can help breeders select animals that are more resistant to disease.

QTL mapping for disease resistance in animals can be performed in different ways. In some cases, genetic data are collected from wild or domestic animals that are naturally resistant to a particular disease. The genetic differences between these resistant animals and those that are susceptible to the disease are then analysed to identify the regions of the genome that are associated with resistance.

In other cases, crosses between resistant and susceptible animals can be used to produce a population of offspring that exhibit genetic variability for disease resistance. The genetic data from this population can then be used to identify the regions of the genome that are associated with resistance.

Once the QTLs associated with disease resistance are identified, breeders can use this information to select disease-resistant animals for breeding. This can help improve disease resistance in animal populations, which can have benefits for animal health and production.

QTL mapping is an important technique in animal research to identify regions of the genome that are associated with traits of interest, such as disease resistance. This information can be used to select resistant animals for breeding, which can help improve disease resistance in animal populations.

5.3. Identification of QTLs involved in adaptation to environmental changes

QTL mapping is also used to identify regions of the genome that are involved in adaptation to environmental changes in plants and animals. Environmental changes such as climate change, pollution, and habitat degradation can have a significant impact on the survival and reproduction of species.

A species' ability to adapt to these changes depends on its genetic variability and its ability to evolve rapidly. QTL mapping can be used to identify regions of the genome that are associated with adaptive traits, such as tolerance to extreme temperatures, disease resistance, the ability to use different types of nutrients, and the ability to adapt to different habitats.

QTL mapping for adaptation to environmental changes can be performed using populations of species that exhibit high genetic variability and phenotypic diversity. The genetic and phenotypic data from these populations can then be analyzed to identify the regions of the genome that are associated with adaptive traits.

Once the QTLs associated with adaptation are identified, this information can be used to select individuals that exhibit desirable adaptive traits for breeding. This can help improve the adaptation

of species populations to environmental changes, which can have a positive impact on their long-term survival and reproduction.

QTL mapping is an important technique for identifying regions of the genome that are involved in adaptation to environmental changes in plants and animals. This information can be used to select individuals with desirable adaptive traits for breeding, which can help improve the adaptation of species populations to environmental changes.

6. Advanced Techniques in Quantitative Genetics

6.1. Genetic Interaction Analysis

Genetic interaction analysis is an advanced technique in quantitative genetics that seeks to identify the combined effects of multiple genes on a given trait. This technique is particularly important for understanding the complex processes that underlie quantitative traits and can be used to improve the accuracy of phenotype prediction.

Genetic interaction can be defined as a situation where the effect of one gene on a trait depends on the allele present at another locus. For example, two genes that have a positive effect on milk production may have a negative effect when both are present in the same line. This interaction can be caused by molecular mechanisms such as modulation of gene expression or interactions between proteins.

There are several methods for genetic interaction analysis, including additive models, multiplicative models, and non-additive models. Additive models consider that each allele contributes independently to the expression of the trait. Multiplicative models take into account the combined effects of the alleles present on the trait. Non-additive models consider that the effects of alleles depend on their interaction with other alleles.

Genetic interaction analysis can be performed using populations of species that exhibit high genetic variability and phenotypic diversity. The genetic and phenotypic data from these populations can then be analyzed to identify genetic interactions that are associated with quantitative traits.

The results of genetic interaction analysis can be used to understand the molecular mechanisms that underlie quantitative traits and to improve the accuracy of phenotype prediction. For example, genetic interactions can be used to predict the response of a population to specific treatments, such as the administration of drugs or dietary supplements.

Genetic interaction analysis is an advanced technique in quantitative genetics that allows the identification of the combined effects of multiple genes on a given trait. This technique is particularly important for understanding the complex processes that underlie quantitative traits and can be used to improve the accuracy of phenotype prediction.

6 .2. Bayesian Regression Models

Bayesian regression models are an advanced statistical analysis technique that is increasingly used in quantitative genetics for QTL analysis. This approach allows for taking into account the uncertainty in the estimates of parameters and genetic effects, as well as the hierarchical structure of the data.

The Bayesian regression model assumes that the parameters are distributed according to a prior probability distribution and that the posterior distribution is obtained by combining this information with the observed data. The use of Bayesian inference thus allows updating the probability distributions of the parameters as new data are collected.

This approach is particularly useful for the analysis of complex data, such as gene expression data, where the data dimension is high and the genetic effects are often small. Bayesian regression models also allow for taking into account the effect of many genes at the same time, as well as the effect of interactions between these genes.

However, this approach can be more complex to implement and require significant computing resources. In addition, the results can be sensitive to the choices of prior distributions and model parameters. Therefore, careful interpretation of the results is essential to avoid misinterpretations.

6 .3. Advanced Marker-Assisted Selection Methods

Marker-assisted selection (MAS) methods are advanced techniques in quantitative genetics that allow the selection of individuals with the best genetic characteristics for a trait of interest, based on specific molecular markers.

Traditional selection involves evaluating the phenotypic performance of individuals, which is often costly, time-consuming, and can be inaccurate. Marker-assisted selection overcomes these obstacles by using molecular markers to predict the genetic value of individuals for a trait of interest, without having to directly evaluate the phenotype.

Advanced marker-assisted selection methods, such as genomic selection or genome-wide selection, use genomic information from the entire genome rather than focusing on a small number of markers.

These methods are based on the use of powerful statistical tools to predict the genetic value of individuals using all available genomic information.

Genomic selection uses molecular markers to predict the genetic value of individuals, while genome-wide selection uses the entire genome to predict the genetic value. These methods have been widely used in plant and animal breeding, where they have significantly accelerated the selection process and achieved significant genetic gains.

However, it is important to note that the use of advanced marker-assisted selection requires a thorough understanding of the genetics of the trait of interest, a large amount of genomic data, and expertise in bioinformatics and statistics. Therefore, these methods are not always accessible to small producers or breeders who do not have sufficient resources.

7 . Future developments in the field of QTLs

7 .1. Evolution of high-throughput sequencing technologies

QTLs (quantitative trait loci) are regions of the genome that are associated with quantitative phenotypic variations for traits of interest. Technological advances in the field of high-throughput sequencing have revolutionized the way QTLs are identified and characterized, and have led to exciting new future developments in this field.

One of the key trends in the field of QTLs is the use of high-throughput sequencing technologies to map QTLs at the whole-genome level. This approach allows for the identification of QTLs that cannot be detected by traditional mapping methods, which can improve the accuracy of selection of individuals in plant and animal breeding programs. In addition, high-throughput sequencing also allows for the identification of rare genetic variants that could have significant phenotypic effects.

Another important trend in the field of QTLs is the use of high-throughput sequencing methods to understand the molecular mechanisms underlying phenotypic variations. Sequencing data allows genetic variants to be linked to changes in protein structure or function, which can help to understand the metabolic pathways and regulatory networks involved in phenotypic variation.

Finally, advances in the field of high-throughput sequencing have also led to the emergence of new approaches for QTL analysis, such as the use of gene regulatory networks to understand the interactions between genes and metabolic pathways involved in phenotypic variation.

Overall, future developments in the field of QTLs will likely be linked to the continued evolution of high-throughput sequencing technologies, as well as a better understanding of the molecular mechanisms underlying phenotypic variations. These advances should help accelerate the development of new plant varieties and new animal breeds with improved and more resilient characteristics.

7 .2. New approaches for high-throughput data analysis

Future developments in the field of QTLs will also be linked to the evolution of new approaches for high-throughput data analysis. High-throughput sequencing technologies produce massive amounts of genomic data, which requires advanced tools for data analysis and interpretation.

One approach under development is the integration of multi-omics data, which involves combining genomic data with other types of omics data, such as transcriptomic, proteomic, and metabolomic data. This approach allows for understanding the molecular mechanisms underlying phenotypic variation across different biological scales, ranging from the DNA level to the protein and metabolic levels.

Another important approach is the use of machine learning for the analysis of high-throughput genomic data. This approach allows for the identification of complex patterns in the data and the prediction of relationships between genes and phenotypic traits. Machine learning can also be used for the identification of rare genetic variants and for the classification of genetic subpopulations.

Finally, new approaches for high-throughput data analysis also include tools for interactive data visualization and for the creation of integrated databases that allow for storing and querying genomic and phenotypic data on a large scale.

Future developments in the field of QTLs will be largely dependent on the evolution of new approaches for high-throughput data analysis. These advances should contribute to a better understanding of the molecular mechanisms underlying phenotypic variations and to the identification of new traits of interest for plant and animal improvement.

7 .3. Prospects for the application of QTLs in marker-assisted selection and synthetic biology

QTLs have important implications in the field of marker-assisted selection, which is a method of selecting individuals for traits of interest using genetic markers associated with those traits. Advances in QTL mapping and characterization, as well as in high-throughput data analysis, have improved the

accuracy of marker-assisted selection. In addition, the use of synthetic biology techniques for genome modification offers possibilities for the creation of new artificial QTLs.

In the field of marker-assisted selection, future developments should focus on the identification of high-resolution genetic markers, as well as on the implementation of marker-assisted selection methods based on more sophisticated statistical models. These methods should allow for optimizing selection for complex and multigenic traits.

Regarding synthetic biology, future developments could allow for the creation of artificial QTLs for traits that do not naturally exist in the population of interest. Synthetic biology techniques can be used to synthesize genes and regulatory networks to control specific metabolic pathways, which could allow for modifying metabolite production and the response to biotic and abiotic stresses.

Finally, QTLs can also be used in the design of artificial biological systems, such as modular gene regulatory networks for the production of molecules for medical and industrial purposes. QTLs can also be used to design biosensor systems for the detection of specific molecules in the environment.

The prospects for the application of QTLs in marker-assisted selection and synthetic biology are promising. Future developments in these areas should allow for expanding the range of phenotypic traits available for selection and the creation of artificial biological systems. However, these advances also raise important ethical and regulatory issues, which will need to be considered to ensure responsible use of these technologies.

8. Exercice

The nature of QTLs is

○	Qualitative
○	Quantitative

9. Exercice

Detection of QTLs is achieved using:

☐	Molecular markers
☐	Phenotypic markers
☐	Biochemical marker

10. Exercice

QTLs are only used in agriculture

○	True
○	False

11. Exercice

No need to statistical analysis in QTLs detection

○	True
○	False

IX. References

1. Aguet, F., Alasoo, K., Li, Y.I., Battle, A., Im H.K., Montgomery, S.B. Lappalainen, T. (2023).
2. Molecular quantitative trait loci. *Nature Reviews Methods Primers*, 3 : 4.
3. Azouzi, B. (2006). L'outil statistique en expérimentation. Ed. Office des publications universitaires, Alger. 163 p.
4. Baril, C., Bergonzini J.C. (1994). De la Génétique des Populations à la Génétique Quantitative.
5. Ed. CIRAD, Forêt. 75 p.
6. Barton, N. H., Turelli, M. (1989). Evolutionary Quantitative Genetics: How Little Do We Know? *Annual Review of Genetics*, 23(1): 337–370.
7. Barton, N.H., Keightley, P.D. (2002). Multifactorial genetics: Understanding quantitative genetic variation. Nature Reviews Genetics, 3: 11–21.
8. Becker, W.A. (1967). Program in Genetics: Manual of procedures in quantitative genetics. Ed.
9. Pullman, Washington State University. 130 p.
10. Broman, K. W., Speed, T. P. (2002). A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4): 641–656.
11. Brun, J.M. (1992). Les bases de la génétique quantitative : Définition et mesure des paramètres du croisement. In. Eléments de Génétique Quantitative et application aux populations animales. Ed. INRA, Production animale. Pp : 101-105.
12. Bulmer, M. G. (1985). The mathematical theory of quantitative genetics. Ed. Clarendon Press, Oxford, UK. 255 p.
13. Caballero, A. (2020). Quantitative Genetics. Ed. Cambridge University Press. 325 p.
14. Charlesworth, B. (1990). Optimization Models, Quantitative Genetics, And Mutation.
15. *Evolution*, 44(3): 520–538.
16. Charlesworth, D., Charlesworth, B. (1995). Quantitative Genetics in Plants: The Effect of the Breeding System on Genetic Variability. *Evolution*, 49(5): 911–920.
17. Choo, T. M., Reinbergs, E., & Jui, P. Y. (1988). Comparison of F2 and F∞ diallel analyses in barley. Genome, 30(6), 865–869.

18. Churchill, G.A., Doerge, R.W. (1998). Mapping Quantitative Trait Loci in Experimental Populations. In. Paterson, A.H. (eds.), Molecular Dissection of Complex Traits. CRC Press. 11 p.
19. Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(1-2): 169–196.
20. Cristobal, M.S., Robert-Granié, C., Foulley, J.L. (2002). Héteroscédasticité et modèles linéaires mixtes : théorie et applications en génétique quantitative. *Journal de la société française de statistique*, 143 (1-2) : 155-165.
21. Cruz, C. D. (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum, Agronomy*, 35(3): 271-276.
22. Cyplik, A., Bocianowski J. (2022). Analytical and numerical comparisons of two methods of estimation of additive×additive×additive interaction of QTL effects. *Journal of Applied Genetics*, 63: 213–221.
23. De Jong, G. (1990). Quantitative Genetics of reaction norms. Journal of Evolutionary Biology, 3(5-6), 447–468.
24. Doebeli, M. (1996). Quantitative Genetics and Population Dynamics. Evolution, 50(2), 532–546.
25. Doerge, R.W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics*, 3: 43–52.
26. Doerge, R.W., Zeng Z.B., Weir B.S. (1997). Statistical Issues in the Search for Genes Affecting Quantitative Traits in Experimental Populations. Statistical Science, 12 (3): 195-219.
27. Doumith, M., Cazalet, C., Simoes, N., Frangeul, L., Jacquet, C., Kunst, F., Martin, P., Cossart, P., Glaser, P., Buchrieser, C. (2004). New Aspects Regarding Evolution and Virulence of Listeria monocytogenes Revealed by Comparative Genomics and DNA Arrays. *Infection and Immunity*, 72(2), 1072–1083.
28. Ducrocq, V. (1992). Les bases de la génétique quantitative : Du modèle génétique au modèle statistique. In. Eléments de Génétique Quantitative et application aux populations animales. Ed. INRA, Production animale. Pp : 75-81.
29. Erickson, D. (2005). Quantitative trait loci: Mapping the future of QTL's. *Heredity*, 95: 417–418.

30. ERNER, L.M., 1954. Genetic Homeostasis. Ed. Oliver and Boyd, Edinburgh.134 p. Falconer, D.S. (1983). Problems on quantitative genetics. Ed. Longman, London, UK. 104 p.
31. Falissard, B. (2005). Comprendre et utiliser les statistiques dans la science de la vie. Ed.
32. Masson, Paris. 372 p.
33. Frankham, R. (1999). Quantitative genetics in conservation biology. Genetical Research, 74(3), 237–244.
34. Frost, J., 2020. Hypothesis Testing: An intuitive guide for making data driven decisions. Ed.
35. Jim Frost. 371 p.
36. Gallais, A. (1990). Théorie de la sélection en amélioration des plantes. Ed. Masson, Paris. 588 p.
37. Gallais, A. (2009). Hétérosis et variétés hybrides en amélioration des plantes. Ed. Quae, France. 356 p.
38. Garcia-Oliveira, A. L., Tan, L., Fu, Y., Sun, C. (2009). Genetic Identification of Quantitative Trait Loci for Contents of Mineral Nutrients in Rice Grain. *Journal of Integrative Plant Biology*, 51(1), 84–92.
39. Gibson, G., Weir, B. (2005). The quantitative genetics of transcription. *Trends in Genetics*, 21(11), 616–623.
40. Glaser, P. (2005). Les puces à ADN vont-elles révolutionner l'identification des bactéries *?*
41. *Médecine/sciences,* 21(5), 539–544.
42. Goffinet, B., Gerber, S. (2000). Quantitative Trait Loci: A Meta-analysis. Genetics, 155(1): 463–473.
43. Gomulkiewicz, R., Kirkpatrick, M. (1992). Quantitative Genetics and the Evolution of Reaction Norms. *Evolution*, 46(2): 390–411.
44. Gorokhova, S., Biancalana, V., Lévy, N., Laporte, J., Bartoli, M., Krahn, M. (2015). Clinical massively parallel sequencing for the diagnosis of myopathies. *Revue Neurologique*, 171: 558–571.
45. Grant, I., Beversdorf, W.D. (1985). Heterosis and combining ability estimates in spring-planted oilseed rape (*Brassica napus* L.). *Canadian Journal of Genetics and Cytology*, 27(4), 472–478.
46. Hackett, C. A., Weller, J. I. (1995). Genetic Mapping of Quantitative Trait Loci for Traits with Ordinal Distributions. Biometrics, 51(4): 1252.
47. Hallauer, A.R., Carena, M.J., Miaranda Filho, J.B. (2010). Quantitative Genetics in Maize Breeding. Ed. Springer Science & Business Media, Lowa State University. 662 p.

48. Hill, W. G. (2009). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537): 73–85.
49. *https://www.semae-pedagogie.org/*[1]
50. Jacquard, A. (1970). Génétique des populations et raisonnement probabiliste. *Journal de la société statistique de Paris*, 111 : 223-229.
51. Jacquard, A. (1970). Structure génétique des populations. Ed. Masson, France. 388 p.
52. Jiang, C., Zeng Z.B. (1995). Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. *Genetics*, 140: 11 11 – 1127.
53. Johannes, F., Colot, V., Jansen, R.C. (2008). Epigenome dynamics: a quantitative genetics perspective. *Nature Reviews Genetics*, 9: 883–890.
54. Kao, C.H., Zeng, Z.B., Teasdale R.D. (1999). Multiple Interval Mapping for Quantitative Trait Loci. *Genetics*, 152: 1203–1216.
55. Kearsey, M.J. Farquhar A.G.L. (1998). QTL analysis in plants, where are we now? *Heredity*, 80:137–142.
56. Knapp, S. J., Bridges, W. C., & Birkes, D. (1990). Mapping quantitative trait loci using molecular marker linkage maps. Theoretical and Applied Genetics, 79(5).
57. Krahn, M., Arveiler, B. (2016). Le séquençage de nouvelle génération : principe, applications en diagnostic et perspectives. In : Livre national d'enseignement - Génétique Médicale
58. - DFGSM2/3. Ed. Issy-les-Moulineaux : Elsevier-Masson.
59. Kumar, J., Gupta, D. S., Gupta, S., Dubey, S., Gupta, P., Kumar, S. (2017). Quantitative trait loci from identification to exploitation for crop improvement. *Plant Cell Reports*, 36(8): 1187–1213.
60. Lamkey K.R., Edwards J.W. (1999). Quantitative Genetics of Heterosis. In. Coors, J.G., Pandey, S., Lamkey, K.R., Edwards, J.W. (eds), The Genetics and Exploitation of Heterosis in Crops. *ACSESS Publications*. Pp: 31- 48.
61. Le Roy, P. (1992). Les bases de la génétique quantitative : Les méthodes de mise en évidence des gènes majeurs. In. Eléments de Génétique Quantitative et application aux populations animales. Ed. INRA, Production animale. Pp : 93-99.
62. Lefort-Buson, M., Dattée, Y., Lavoisier, O. (1982). Genetic study of some agronomic characters in winter oilseed rape (*Brassica napus* L.) II. - Genetic parameters. *Agronomie*, 2 (4): 323-332.

1 https://www.semae-pedagogie.org/

63. Lellouch, J., Lazar, P. (1974). Méthodes statistiques en expérimentation biologique. Ed.
64. Flammarion, France. 283 p.
65. Louis, O. (2002). Éléments de génétique quantitative. Ed. Quae, France. 184 p.
66. Malécot, G., Malecot, G. (1955). La Génétique de population : Principes et applications.
67. *Population* (French Edition), 10(2) : 239.
68. Mallar, J. (1992). Les bases de la génétique quantitative : Population et Variabilité. In. Eléments de Génétique Quantitative et application aux populations animales. Ed. INRA, Production animale. Pp : 55-60.
69. Mallol, C., Varro, L. (). Conditions d'existence d'une loi de Hardy-Weiberg périodique dans des populations sélectivement neutres. In. Costa, R., Grishkov, A., Guzzo, H.V., Peresi,
70. L.A. (eds). Nonassociative algebra and its application. Marcel Dekker Inc., New York. Pp: 439-446.
71. McIntyre, L.M., Coffman, C.J., Doerge, R.W. (2001). Detection and localization of a single binary trait locus in experimental populations. Genetics Research, 78(01): 79-92.
72. Miles, C., Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nature Education*, 1(1): 208.
73. Moll, R. H., Stuber, C. W. (1974). Quantitative Genetics—Empirical Results Relevant to Plant Breeding. *Advances in Agronomy*: 277–313.
74. Morgenthaler, S. (2008). Génétique statistique. Ed. Springer-Verlag France. 156 p.
75. Motulsky, H. (2018). Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking.
76. Ed. Oxford University Press. 605 p.
77. Motulsky, H. (2019). Biostatistique : Une approche intuitive. Ed. De Boeck Supérieur s.a., France. 577 p.
78. Mullin, T. J., Park, Y. S. (1992). Estimating genetic gains from alternative breeding strategies for clonal forestry. *Canadian Journal of Forest Research*, 22(1), 14–23.
79. Namkoong, G. (1979). Introduction to Quantitative Genetics in Forestry. *Technical Bulletin*, 1588. 342 p.
80. Posthuma, D., Beem, A. L., de Geus, E. J. C., van Baal, G. C. M., von Hjelmborg, J. B., Iachine, I., Boomsma, D. I. (2003). Theory and Practice in Quantitative Genetics. *Twin Research*, 6(05), 361–376.
81. Poujardieu, B., Mallard, J. (1992). Les bases de la génétique quantitative : Les méthodes d'estimation del'héritabilité et des corrélations génétiques. In. Eléments de Génétique

Quantitative et application aux populations animales. Ed. INRA, Production animale. Pp : 87-92.

82. Roff, D. A. (2007). A Centennial Celebration For Quantitative Genetics. *Evolution*, 61(5): 1017–1032.
83. Roff, D.A. (1997). Evolutionary Quantitative Genetics. Ed. Springer Science & Business Media. 492 p.
84. Rousson, V. (2013). Statistique appliquée aux sciences de la vie. Springer-Verlag France. 327 p.
85. Rouyrre, N., Gilles, B. (1995). Génétique quantitative : aspects généraux et application à l'amélioration des plantes. Ed. CIRAD-Forêt, France. 31 p.
86. Rumbaugh M.D., Caddel J.L., Rowe D.E. (1988). Breeding and Quantitative Genetics. In. Hanson, A.A., Barnes, D.K., Hill, R.R., Rumbaugh, M.D., Caddel, J.L., Rowe, D.E., Alfalfa and Alfalfa Improvement. *Agronomy Monograph*, 29: 777-808.
87. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., Davis, R.W. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences*, 93(20): 10614–10619.
88. Seaton, G., Haley, C. S., Knott, S. A., Kearsey, M., Visscher, P. M. (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics*, 18(2): 339–340.
89. Sehgal, D., Singh, R., Rajpal, V. R. (2016). Quantitative Trait Loci Mapping in Plants: Concepts and Approaches. *Sustainable Development and Biodiversity*, 31–59.
90. Shi, X.M. (2020). eQTL Analysis: Methods and Protocols. Ed. eQTL Analysis Methods and Protocols. Ed. Humana Press, Spronger Protocols, USA. 252 p.
91. Southern, E. M. (1996). DNA chips: analysing sequence by hybridization to oligonucleotides on a large scale. *Trends in Genetics*, 12(3), 110–115.
92. Stuber, C.W., Edwards, M. D., Wendel, J.F. (1987). Molecular Marker-Facilitated Investigations of Quantitative Trait Loci in Maize. II. Factors Influencing Yield and its Component Traits1. *Crop Science*, 27(4) : 639.
93. Tourab, D. (2006). Initiation à la biostatistique. Ed. Amazon Fulfillment, Poland. 220p.
94. Trad, S., Allignet, J., Frangeul, L., Davi, M., Vergassola, M., Couve, E., Morvan, A., Kechrid, A., Buchrieser, K., Glaser, P., El Solh, N. (2004). DNA Macroarray for Identification and Typing of Staphylococcus aureus Isolates. *Journal of Clinical Microbiology*, 42(5), 2054–2064.

95. Troesch, A., Nguyen, H., Miyada, C.G., Desvarenne, S., Gingeras, T.R., Kaplan, P.M., Cros, P., Mabilat, C. (1999). Mycobacterium Species Identification and Rifampin Resistance Testing with High-Density DNA Probe Arrays. *Journal of Clinical Microbiology*, 37(1), 49–55.
96. Van Buijtenen, J.P. (2001). Genomics and quantitative genetics. *Canadian Journal of Forest Research*, 31(4), 617–622.
97. Van Leeuwen, W.B., Jay, C., Snijders, S., Durin, N., Lacroix, B., Verbrugh, H.A., Enright, M.C., Troesch, A., van Belkum, A. (2003). Multilocus Sequence Typing of *Staphylococcus aureus* with DNA Array Technology. *Journal of Clinical Microbiology*, 41(7): 3323–3326.
98. Vilain, M. (1999). Méthodes expérimentales en agronomie : Pratique et analyse. Ed. Technique et Documentation, Paris. 337 p.
99. Wada, Y., Akita, T., Awata, T., Furukawa, T., Sugai, N., Ishii, K., Ito, Y., Kobayashi, E., Kusumoto, H., Matsumoto, T., Mikawa, S., Miyake, M., Murase, A., Shimanuki, S., Sugiyama, T., Uchida, Y., Yanai, S., Yasue, H. (2000). Quantitative trait loci (QTL) analysis in a Meishan x Gottingen cross population. *Animal Genetics*, 31(6), 376–384.
100. Walsh, B. (2001). Quantitative Genetics. *Encyclopedia of Life Sciences*: 1-7.
101. William, R. Meredith, J. (1984). Quantitative Genetics. In. Kohel, R.J., Lewis, C.F., Meredith,
102. W.R. (eds), Cotton, *Agronomy Monograph*, 24: 131-150.
103. Wricke, G., Weber, E. (1986). Quantitative Genetics and Selection in Plant Breeding. Ed.
104. Walter de Gruyter, Berlin. 407 p.
105. Xu, Y., Li, P., Yang, Z., Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *The Crop Journal*, 5(2): 175–184.
106. Yeri, S.B., Kumari, V., Sharma, R., Punia S.S. (2022). Quantitative Trait Locus (QTL) Mapping in Crop Improvement. In. Kumar, N. (eds), Biotechnology and Crop Improvement. CRC Press. 10 p.
107. Yun, S. J., Gyenis, L., Hayes, P. M., Matus, I., Smith, K. P., Steffenson, B. J., &
108. Muehlbauer, G. J. (2005). Quantitative Trait Loci for Multiple Disease Resistance in Wild Barley. Crop Science, 45(6), 2563.
109. Zeng, Z.B. (1994). Precision Mapping of Quantitative Trait Loci. *Genetics*, 136: 1457-1468.

Informations légales