

GENETICS

Giacomo Castagnetti - Genomics - 2021/2022

INDEX:

- Introduction
 - Genetics
 - Genes
 - History of genetics and genomics
 - Mendel
 - Modern biology
 - Genomics era
 - Omics
- Mendel
 - Pre-mendelian theories of inheritance
 - Mendel's scientific approach
 - True breeding
 - Monohybrid crosses
 - Mendel's first law
 - Punnett square
 - Test cross
 - Dihybrid crosses
 - Mendel's second law
 - Predicting the outcome of crosses with more than two genes
 - Rediscovery of mendel's work
- Basic principles of probability
 - Basic notions
 - Conditional probability
 - Hypothesis testing
 - Observation
 - Experiment
 - Chi-square test
 - Errors
 - Degrees of freedom
 - Excel
- Chromosomes
 - DNA packaging
 - Cell division
 - Haploids, diploids and poliploids
 - Karyotype
 - Homologous chromosomes
- Cell cycle
 - Somatic cells
 - Interphase
 - Chromosomes through the phases
 - Mitosis
 - Cytokinesis
 - Nuclear spindle

- Germ cells
 - Meiosis
 - First meiotic division
 - Second meiotic division
 - Variation
- Linked inheritance
 - Genetic sexual determination
 - Morgan's work
 - Bridges' work
 - Sex differentiation in drosophila
 - Genetic sexual determination in humans
 - Sex differentiation in mammals
- Dosage compensation
 - Mosaicism
 - Counting mechanism
 - Partial inactivation
 - Epigenetic event
 - Epigenetic modifications
 - Histone acetylation and methylation
 - Non coding RNA molecules
 - DNA methylation
- Pedigree analysis
 - Garrod's work
 - Pedigrees
 - Notation
 - Autosomal recessive
 - Autosomal dominant
 - X-linked recessive
 - X-linked dominant
 - Effect of X inactivation
- Extension of Mendelian laws
 - Classical Mendelian models of inheritance
 - Basic extensions of Mendelian inheritance
 - incomplete/partial dominance
 - Codominance
 - Multiple allele of a gene
 - Recessive lethal alleles
 - Multiple phenotypes
 - Environment influences
 - Incomplete penetrance
 - Variable expressivity
 - Sex limited and influenced characters
- Gene interactions
 - Genetic analysis

- Complementation analysis
- Dihybrid crosses
 - None
 - Addition
 - Epistasis
- Linkage analysis
 - Morgan
 - Morgan's hypothesis
 - First linkage map
 - Single crossing over
 - Autosomal chromosomes
 - Physical proof of crossing over
 - Additiveness of recombination frequencies
 - Double crossing over
 - Three point mapping
 - Interference
 - Genetic map of the drosophila chromosomes
 - Genetic mapping functions
 - Physical distance
 - Recombination hotspots

INTRODUCTION

Genetics

Discipline that studies:

- The inheritance of information (traits) from a generation to the next one
- The structure and function (biological identity) of the genetic information
- The genetic variation
- The relation between a specific gene and its function (from genotype to phenotype)

Genes

Traditional genetics	<ul style="list-style-type: none">• Fundamental units of heredity, transmitted from parents to offspring• Each gene corresponds to a trait
Molecular genetics	<ul style="list-style-type: none">• Segment of DNA on a chromosome (or a union of segments which were originally separated), that encodes for a functional product (protein or non coding RNA)

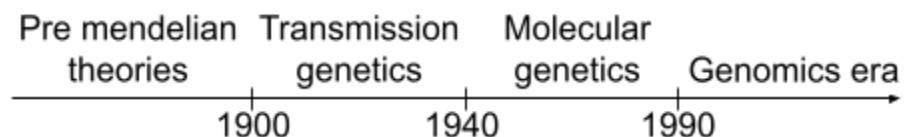
genotype=the genetic makeup of an individual (whole genes, DNA sequence)

phenotype=observable expression of genes producing observable characteristics in an individual

Genetics is important for all levels of biological organization:

levels:	molecular	cellular	organism	population
Ex: Human chromosome → 7, cystic fibrosis gene	Protein that interact with→ the cellular membrane	The protein acts like a → ionic pump	A dysfunction of the protein causes an → accumulation of mucus outside of the cell, the organism is affected by cystic fibrosis	This disease is differently spread in different populations

History of genetics and genomics



Some early theories:

- Pangenesis: genetic information from different parts of the body travels to the reproductive organs where it is transferred to the gametes
 - Preformation: the “genetic information is only provided by the sperm, it contains an homunculus (fully developed scaled human)
 - Blending inheritance: offsprings are an average of parents
-
- **Mendel (1866)**
 - Describes the fundamental law of inheritance

- Theorises the presence of genes (unchangeable units of inheritance), each couple of genes (alleles) controls a trait of the organism
- **Modern biology**
 - *Cell theory*: cells are the fundamental structural and functional unit of organisms, cells arise only from preexisting cells through mitosis and meiosis
 - *Chromosomes*: the nucleus of the cell contain the genetic information, organized in chromosomes
 - *Genes*: through the work of Morgan (fruit flies) the chromosomal theory of inheritance is accepted

Mutation: heritable change in DNA sequence, a mutant is an individual that has an altered phenotype due to an alteration in the genetic composition.

Genetic analysis: usage of *mutants* to study the inheritance, by looking *how the mutation is transmitted* through the generations it can be created a *transmission model* and discover which genes are responsible of the mutation, through molecular biology tools you the *gene* can be *isolated* and amplified (through molecular biology tools the path can be crossed in both directions, a mutation is put into a gene and it is studied what phenotype does it produces)

Model organisms: handy organisms for genetic analysis

1. Short generation time
2. Easy to maintain and grow in a lab
3. Suitable for the molecular biology tools
4. Closely resemble of other organisms or systems

Ips cells: cells that can be reprogrammed to generate cells of different and multiple types

- *DNA*: it is discovered that nucleic acids are the carrier of the genetic information (Avery, MacLeod, McCarty) and Watson & Crick (Rosalind Franklin) described the structure of dna
- *The central dogma of molecular biology*: the sequence of nucleic acids is deciphered, the stream of information goes from the dna to the proteins, practical and precise proof that a single change in the genotype can cause a change in the phenotype
- *DNA sequencing*: fred sanger discovers how to sequence DNA, the ambitious Human Genome Project starts (a single lab could sequence 200 bp/day, the human genome is constituted of 6 billion bases)

In 2001 Francis Collins, director of the Human Genome project, and Craig Venter, leader of a private genome sequencing project, both published the first half of the human genome on science and nature

● **Genomics era**

Genomics differentiates from genetics because it considers the whole “meaning” of the genome (organization, function, regulation and evolution) of an organism, not just the function of a gene.

- Genome sequencing of other organisms
- Metagenomics: DNA sequences of species that are difficult to grow in laboratories, it can be done directly were those organisms live

- Next generation of DNA sequencing: through massively-parallel sequencing the cost and the time of DNA sequencing is drastically decreased.
 (a meter of paragon for scientific discoveries is the moore's law: every two years the calculation capacity of computers doubles; the advances of dna sequencing are dramatically better than the moore's law trend)
 Nowadays to sequence DNA a few days and less than 1000 dollars are requested.

The next step was to understand the genome, discovering his functions; this was achieved through:

1. Genomic variations: finding the differences between the genomes of individuals and linking them to phenotype differences.
 Only 0,1% of the genome is variable between individuals, the majority of variations are common variations, placed every time on the same bases, some of those can have phenotypic consequences and some not.
2. Coding sequences: only 1,5% of the genome encodes for proteins.
3. Non coding sequences: 10% of the genome is used for regulatory functions and it is similar between vertebrates.
4. Epigenomics marks: coding and non coding sequences of dna that cause the 3d structure to form, the epigenomics variables don't involve changes in the nucleotide sequence but they just rearrange the dna to silence or activate it.
5. Genome function: 80% of the genome has a biological function, the ENCODE project is born to keep track of everything we know about the genome, through multiple technologies and approaches in a collective effort to define the functional elements of the genome.

Genomics sits at the intersection of genetics, molecular biology and bioinformatics, it has various applications

Ex: genomic medicine

Through the study of the genome of an individual we can develop individualized prevention plans for specific diseases, this leads to improved health and reduced healthcost.

In the case of a tumor through the genome we can understand the development rate and the type of the cancer and act in consequence.

- **Omics**

Science that joins a lot of branches such as: genomics, proteomics, metabolomics, transcriptomics, metagenomics.

The today's bottleneck of data analysis is that we need to understand the meaning of some parts of the genome.

MENDEL (1822 - 1884)

Basic principles of heredity

Mendel was a monk in Brno, he was a scientist and in the monastery he was researching and teaching.

He used a totally theoretical approach, he knew nothing about cellular structure and he had no way to check visually the correctness of his hypothesis (discrete units of inheritance).

In 1866 he published the results of his experiments and laid the foundation of transmission genetics.

Pre mendelian theories of inheritance:

- Pangenesis
- Blending inheritance

Mendel's scientific approach

He tried to make complex problems simple

- Choice of **garden pea** as model organism:
 - Easy to grow
 - Lots of varieties
 - Many offsprings
- Selecting one character at a time, that showed just two distinct forms, and used true-breeding strains.
- Quantitative analysis of results, he kept track of a lot of tries.
- Scientific method, he formulated an ipotesis and made experiments to prove it.

True breeding

True breeding strains are plants that maintain their characteristics through generations.

The base of the work of mendel was the breed between his plants:

In garden peas the male germ cells are contained in pollen grains, at the top of the anthers while the eggs are at the base of the flower, in the ovary, the only access is constituted by the stigma.

The pollination occurs when the pollen reaches the stigma, usually it is a self-fertilization.

Mendel had developed a technique to cross fertilize his plants, he removed the anthers from a flower and then he took the pollen from another organism and put it into the stigma.

Mendel selected 7 perfect varieties (starting from 34), he wanted that every variety had just two contrasting phenotypes, then he kept self pollinating varieties to obtain pure lines.

Monohybrid crosses

- Parental generation: Mendel crossed 2 breeding strains that were different only for one character.
- F1 generation: one of the 2 versions of the character seem to have disappeared; Mendel made self-fertilize these plants.
- F2 generation: the disappeared character reappears in a ration of $\frac{1}{4}$ in respect to the other.

Mendel obtained the same results with every of the 7 traits that he had selected, even in **reciprocal crosses** (genes of dad and mom inverted); so he created a model based on the idea of particulate unit factors of heredity, that didn't blend.

Mendel's first law

1. Unit factors in pairs

Genetic characters are controlled by unit factors existing in pairs for each organism, each character is controlled by a gene and each gene has two or more alternative forms, called alleles

2. dominance /recessiveness

If two different alleles, responsible for the same character, are present in the same individual one of them is dominant (it has a phenotypic expression) and the other is recessive.

3. Segregation

During the formation of gametes the alleles separate, or segregate, randomly so that each gamete receives one or the other with equal possibility; at the fertilization gametes meet randomly

Ex. color alleles of pea plants:

Y =dominant, it makes the pea look yellow

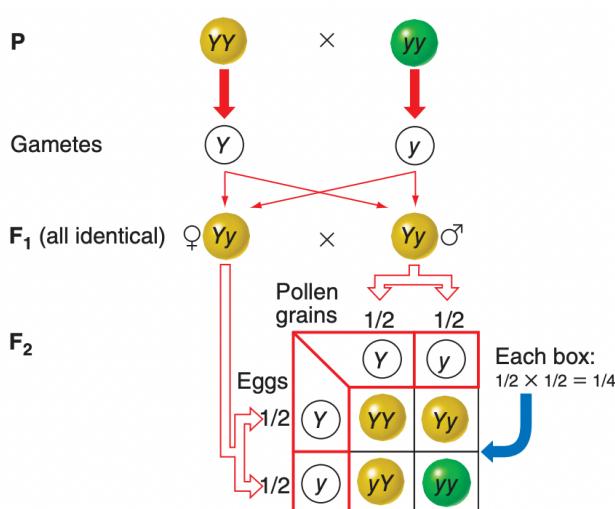
y =recessive, it makes the pea look green

Homozygote: YY (pea appears yellow and the parent transmits yellow alleles at 100%) yy (pea appears green and the parent transmits green alleles at 100%).

Heterozygote: Yy (pea appears yellow and the parent can transmit both Y and y alleles, with the same probability)

Punnett square

A simple way to visualise breeding probability is the **Punnett square**:



Probability is also computable through the **product rule**:

The probability of 2 or more independent events happening both at the same time is the product of their individual probabilities

Or through the **sum rule**:

Having 2 mutually exclusive events, the probability of at least one of them to happen is the sum of their individual probabilities.

In this case product rule can be used to find the probability of a specific genotype, to have the yy genotype we need the gametes to be both y , so the

probability of having that result is: $1/2 \times 1/2 = 1/4$.

In the case of calculating the yellow phenotype we can sum the probability of each of the yellow phenotype events, to gain the probability for at least one of those possibilities to happen:
 $\frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{4}$.

dictionary:

Gene= factors of inheritance controlling a trait

Alleles= alternative forms of a gene (R r; Y y; T t; etc)

Phenotype= appearance

Genotype= pair of alleles carried by the individual

Homozygote= two copies of the same allele (RR; yy)

Heterozygote= two alternative alleles

Dominant/Recessive= “A” is dominant on “a” if the phenotype of Aa is the same of AA. The recessive phenotype is masked by the dominant one in Aa individuals

Test cross (mendel's first law)

Now we can test if the model that mendel theorised works in other situations:

Crossing a homozygous recessive with a dominant phenotype we can have two situations:

If the dominant phenotype has a homozygous genotype the results from the cross will be only dominant phenotypes.

If the dominant phenotype is heterozygous the we will have $\frac{1}{2}$ chance to have a dominant phenotype and $\frac{1}{2}$ chance to have a recessive phenotype.

Doing this cross between a recessive phenotype and an individual from the dominant phenotypes of F2 generation ($\frac{1}{3}$ homozygotes, $\frac{2}{3}$ heterozygotes): gives a possibility of $\frac{1}{3}$ of having a dominant phenotype + $\frac{2}{3} * \frac{1}{2}$ to have still a dominant phenotype and $\frac{2}{3} * \frac{1}{2}$ to have a recessive phenotype.

Dihybrid crosses

Mendel started crossing plants that were different for 2 traits:

- Parental generation: one individual presented 2 homozygous dominant traits and the other had the correspespective recessive alleles.
- F1 generation: all the offspring presented a dominant phenotype, mendel let the plants self-fertilize.
- F2 generation: in the offspring appeared all the characteristics, in all the possible combinations, even **combinations that weren't present in the parental generation** (the other hypothesis was that genes provenient from the same parent transmitted through a dependent assortment, so the only possibilities should have been both dominant phenotype or both recessive).

Mendel observed these ratios:

Parental combinations: 1/16 both recessive traits

9/16 both dominant traits

New combinations: 3/16 one dominant and one recessive trait

3/16 one recessive and one dominant trait.

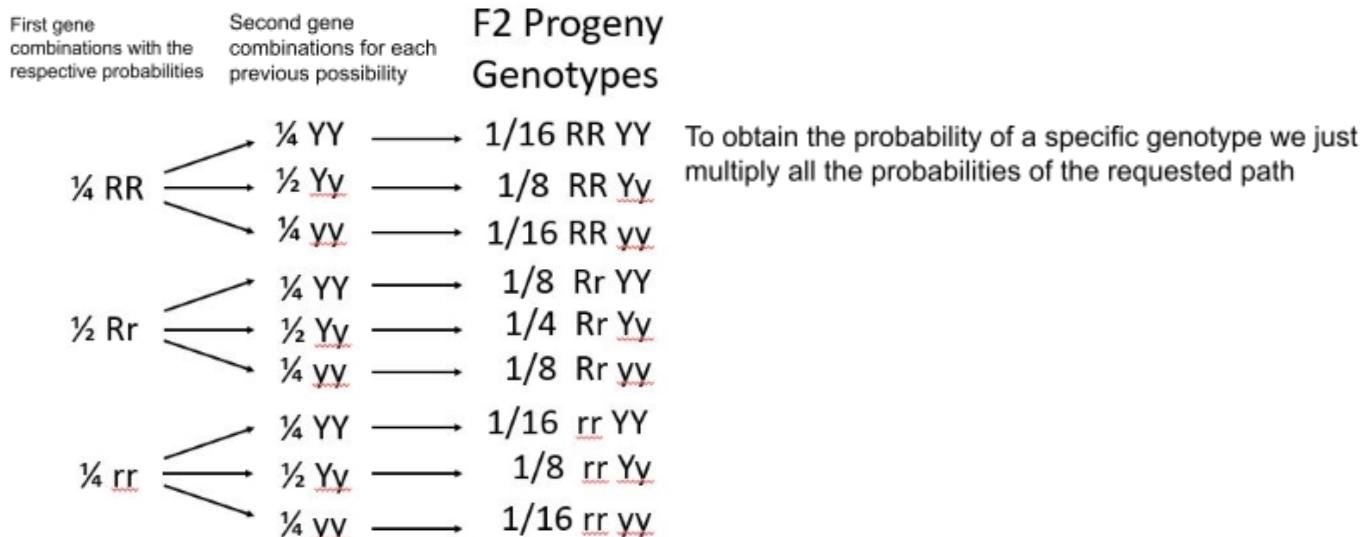
Also in this case mendel tried to start with a parental generation that had different allele combinations but the results were the same, the genes transmitted **independently**.

Through the Punnett square we can understand the meaning of the experimental ratios: every parent can transmit 4 possible combination of its alleles, each one with the same probability, so the total possibilities are $4 \times 4 = 16$, with 4 different phenotypes.

Cross of F₁ Generation

		RY	Ry	rY	ry	
		RY	RRYY	RRYy	RrYY	RrYy
round, yellow	RY	RRYY	RRYy	RrYY	RrYy	
	Ry	RRYy	RRyy	RrYy	Rryy	
	rY	RrYY	RrYy	rrYY	rrYy	
	ry	RrYy	Rryy	rrYy	rryy	

Another useful graphic display of mendel's crosses is the **branch diagram**:



Dihybrid crosses can be considered theoretically as consisting of two monohybrid crosses conducted separately.

Mendel's second law: independent assortment

Independent assortment: during gamete formation, segregating pairs of alleles assort independently of each other

(test cross again between a recessive homozygous and dominant phenotype, the results depends only on what the dominant heterozygous is transmitting: $\frac{1}{4}$ dominant heterozygous, $\frac{1}{4}$

half dominant heterozygous and half recessive, $\frac{1}{4}$ half recessive and half dominant heterozygous, $\frac{1}{4}$ recessive)

Predicting the outcome of crosses with more than two genes (assuming simple dominant-recessive relationship)

Number of heterozygous gene pairs: n

Number of different types of gametes formed: 2^n

Number of different genotypes produced in offspring: 3^n

Number of different phenotypes produced in offspring: 2^n

These relationships are based on the fortune of mendel, not all the genes are independent, in the reality genes that are near on the same chromosome are very likely to transmit together, otherwise if two genes are on different chromosomes or if they are far from each other they are not dependent one from the other (chromosome crossing over).

Rediscovery of mendel's work

In 1900 Walter Sutton and Theodor Boveri, based on the discoveries of cell division carried out by Walter Fleming (mitosis) and on the analysis of chromosomes through cell coloration and microscope, were able to independently describe the action of chromosomes during meiosis and mitosis.

Sutton proposed the chromosomal theory of inheritance: "the association of paternal and maternal chromosomes in pairs and their following separation during cell division may constitute the physical basis of the Mendelian law of heredity".

This theory was confirmed by the work of Morgan, using Drosophila.

BASIC PRINCIPLES OF PROBABILITY

Probability = mathematical **measure** of likelihood

- Empirical probability: an event is calculated by counting how many times the event occurred vs the total tries. It can give an idea about theoretical probability
- Theoretical probability: probability computed through assumptions and mathematical laws

Basic notions

P(A) = probability of outcome A, the value must lie within the range of 0 (impossible) and 1 (there is only that possibility).

P(A,B) = probability of outcome A or outcome B or both, it is called joint probability (\cup).

Sum rule: if A and B are mutually exclusive, $P(A,B) = P(A) + P(B)$.

P(AB) = probability of simultaneous outcomes A and B (\cap)

Product rule: if outcomes A and B are independent, then $P(A,B) = P(A) * P(B)$.

Conditional probability

Conditional probability = measure of the probability of an event given that another event has occurred.

$P(A|B) = \frac{P(A \cap B)}{P(B)}$ probability of A given that B has occurred

$P(A \cap B) = P(A|B) * P(B)$

if $P(A|B)$ is equal to $P(A)$ or $P(B)$ the two events are said to be **independent**.

If B is **necessary** for A, $P(A \cap B) = P(A)$ and $P(A|B) = \frac{P(A)}{P(B)}$

Ex. In a monohybrid cross of tall plant x short plant, all F1 offspring is tall
if F1 individuals are cross-fertilized, what is the probability that a tall plant in the F2 progeny is also heterozygote?

A=heterozygote=2/4

B=tall=conditioning event=3/4

The event B may be visualized as a restriction of the sample space

$P(A \cap B) = 2/4$ in this particular case the event B is necessary to A.

$P(A|B) = \frac{P(A \cap B)}{P(B)} = 2/4 / 3/4 = 2/3$

Hypothesis testing

Hypothesis testing is a tool to accept a scientific hypothesis or reject it.

- **observation**

To create an hypothesis is necessary to start with an observation:

The ideal observation should be performed on an **entire population** (the set of all possible states of a random variable) but in most cases it is studied a **sample** of the population (subset of the population, of a finite size, that should represent the whole population) from which are **estimated the parameters** of the population.

After the observation an idea is developed, it is **created a theory (H₀)**, a formal explanation of the observed data.

- **experiment**

At this point it is necessary to carry out experiments to see whether the hypothesis works or not: to accomplish this we need hypothesis testing.

- The initial hypothesis is called **H₀**
- The experimental data is called **observed data**

We need to apply a statistical test to accept or reject the hypothesis, the test is based on the contrast between the observed and expected data and in many cases we are interested in the computation of the p-value.

Ex. mendel's dihybrid cross:

	D-D	D-d	d-D	d-d	tot.
observed	315	108	101	32	556
expected	313	104	104	35	556

Are these results close enough to the theoretical hypothesis to consider it true?

Ex. Hugo De Vries experiment

Experiment with flower color and foliage type, supposing that the transmission followed the mendelian model we have the following report:

	D-D	D-d	d-D	d-d	tot.
observed	70	23	46	19	158
expected	88.9	29.6	29.6	9.9	158

- **chi-square test**

To reach a conclusion we need the **Chi-square test (χ²)**

This permits us to estimate how frequently the observed deviation can be expected to occur just as a result of chance and not because of an error in the model.

Through the test we can calculate the Chi-value, then establish a relation with the p-value, and finally evaluate the goodness of fit between observed and expected data. The Chi-square test it is only useful with categorical data, the formula to compute the Chi-value is:

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

o is the observed value for a given category

e is the expected value for that category under H₀

We divide per e because the difference must be considered in proportion to the order of magnitude of the considered value (smaller the chi-value, smaller the difference, greater the p-value).

The p-value can be computed just through computers, we use tables that put in relation Chi-square and p-value:

Degrees of freedom

df	χ ² .995	χ ² .990	χ ² .975	χ ² .950	χ ² .900	χ ² .100	χ ² .050	χ ² .025	χ ² .010	χ ² .005	P values
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838	
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750	
6	0.676	0.872	1.237	1.625	2.204	10.645	12.592	14.449	16.812	18.548	

When p-value is high it is assumed that chance alone produced the difference, while when p-value is low it is assumed that a significant factor (not randomness) produced the deviation.

χ^2

But how can we determine a significance level that differentiate a low p-value from a high one?

- **errors**

When we are testing an hypothesis we have 4 possible situations:

		DECISION	
		Accept H0	Reject H0
H0 true	1- α	Type 1 error: α	
H0 false	Type 2 error: β	1- β	

Type one error= α error that is made when we **reject an hypothesis**, there is a small chance that the hypothesis was actually good and that we are discarding it because of a high random deviation (unluckiness).

Type two error= β error that is made when we **accept an hypothesis**, obviously there is a chance that the hypothesis doesn't really actually work in all cases, but we just tested a particular sample of the population.

The significance level of p-value that is considered as **convention** is **0.05** (5%), above this value the hypothesis is accepted: we always have to consider the 7th column of the table to check if the calculated χ^2 is higher than the one in the 7th column.

- **degrees of freedom**

To really find the p-value in the tab we need to define the degrees of freedom.

The degrees of freedom are the number of independently varying parameters of the experiment; every experiment has at least 2 random outcomes, one of those can be expressed as the total number of individual in the population or total number of tries minus the first outcome, so:

$$df = n \text{ of independent classes} - 1$$

...Ex

In our example there are 4 classes, so 3 degrees of freedom, we consider the third row of the table.

Mendel:

	D-D	D-d	d-D	d-d	tot.
observed	315	108	101	32	556
expected	313	104	104	35	556
$\frac{(o-e)^2}{e}$	0.01	0.15	0.09	0.26	$\chi^2=0.51$

On the table in correspondence to 3 degrees of freedom we can see that a 0.51 chi square gives us a 0.9 p-value that is way bigger than 0.05 so the H₀ is acceptable.
DeVries:

	D-D	D-d	d-D	d-d	tot.
observed	70	23	46	19	158
expected	88.9	29.6	29.6	9.9	158
$\frac{(o-e)^2}{e}$	4.02	1.47	9.09	8.36	$\chi^2=22.94$

Now, referring to the table, the p-value is 0.00004, it is way smaller than 0.05 so we have to reject the hypothesis (accepting type one error); there is something wrong in the experiment.

- **Excel**

Once that you have created a table with observed and expected values ($e=\text{tot} \times R$), you can obtain the p value using the function CHISQ.TEST(observed values; expected values).

CHROMOSOMES

The chromosomes have the same function in each cell: keeping the genetic information and transmitting it to the successive cell generation; but the structure of the chromosomes vary from eukaryotes to prokaryotes:

- Prokaryotes: there is just one circular chromosome situated in the nucleoid, a region of the cytoplasm, it is not surrounded by a membrane
- Eukaryotes: there are multiple chromosomes, characterized by linear portion of DNA, they are kept securely inside the nucleus

DNA packaging

To form the chromosomes the DNA is highly packed:

- Nucleosome: 8 histone protein subunits attach to the dna molecule forming loop structures.
- Chromatin: multiple nucleosomes are coiled together and stacked on top of each other forming a fiber (30 nm).
- Chromosomes: the chromatin is looped and packed through other proteins

Chromosomes are not always present, they form only when the cell is dividing, this packing mechanism is also very useful as a regulation system: when chromatin is highly packed it is impossible to express genes (this is called epigenetics and acts to differentiate the cells of different tissues or to modify cells during the development).

Cell division

- Prokaryotes: simple binary fission mechanism, the chromosome replicates to produce two identical copies, these two copies segregate from each other with one copy going to each daughter cell.
- Eukaryotes: there are two different types of cell division
 - Mitosis: useful to form 2 identical filial cells, it takes place in somatic cells
 - Meiosis: useful to form haploid gametes, it takes place in germ cells; it generates diversity, daughter cells are a complete mixture of the genome.

Haploids, diploids and poliploids

Individuals from the same species have the same number of chromosomes, the chromosomes can be arranged in pairs, each chromosome of the pair (homologous chromosomes) comes from a different parent, in this case the individual is called **diploid** (germ cells are always haploid).

The majority of the eukaryotes are diploids, with the exception of plants that usually are poliploid; bacteria on the other hand are haploids and yeasts remain haploid for the majority of their life.

n=haploid number, number of pairs

2n=diploid number, number of chromosomes

Ex. humans: n=23. 2n=46 chromosomes

How to see chromosomes:

As we said before chromosomes are not always visible, when the cell is not dividing they are chromatin, we can see chromosomes under a microscope through cytogenetics.

Cytogenetics is a set of techniques that allows to observe chromosomes at the microscope: We have to get a cell population and add a chemical to simulate mitosis, after 2-3 days we stop the mitosis in metaphase through another chemical; then the sample needs to be centrifuged and fixed on a microscope slide; at this point we add stain that enhance chromosomes and give rise to bending patterns on the chromosomes.

Karyotype

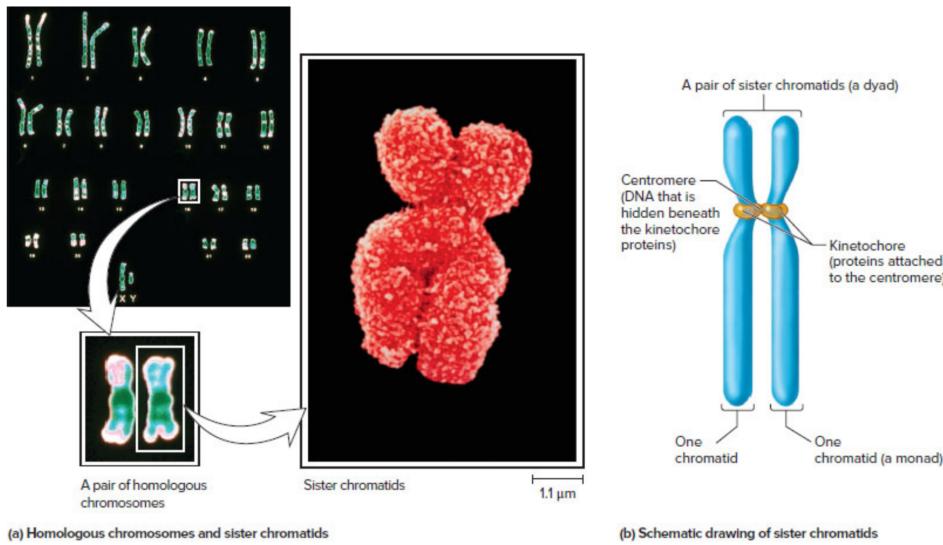
Through the microscope we can see a **disordinate** set of chromosomes, **bended** in certain specific points.

Those chromosomes can be displayed in homologous pairs to obtain the karyotype.

Homologous chromosomes

The homologous chromosomes are copies of the same chromosome, one from the father and one from the mother.

In the metaphase each homologous chromosome is formed of two identical sister chromatids, held together by a centromere (highly repetitive DNA sequence that permits the attach of proteins that form the kinetochore).



Homologs are not identical chromosomes, they can differ for some allele, while non homologs carry completely unrelated sets of genes.

Each pair of chromosomes have a different shape and size, the position of the centromere can differ a lot (at the centre of or at the end of a chromosome) but also the length of each chromosome, only in the case of the xy pair the homologs have different length.

CELL CYCLE

Somatic cells

The life of a cell is characterized by 2 important events: the mitosis (M) and the duplication of the DNA that precedes the mitosis (S).

Interphase

Between the **M** phase and the **S** phase there are the **G phases**:

- After the mitosis there is the **G1** phase, in which the cells prepare for the division and reach the **G1/S** checkpoint to assure that everything is at the right place before the start of the replication; some specific cells (as the brain cells) doesn't experience the G1 phase but enter in the **G0** phase, they won't divide anymore.
- After the DNA replication there is the **G2** phase in which the cells pass through another checkpoint to confirm that's the right timing to divide.

In some cases the checkpoints lose their control function and the cell starts dividing in an uncontrolled way, sometimes this leads to the development of a tumor.

Chromosomes through the phases:

- **G1:** chromosomes have a single double helix (no sister chromatids), chromosomes=2n, DNA=2c.
 - **G2:** sister chromatids have been produced, chromosomes=2n, DNA=4c.
- The ensemble of these phases constitute the interphase, after the interphase the mitosis starts.
- Mitosis: during the mitosis the sister chromatids splits (after the metaphase), chromosomes=4n, DNA=4c, but right away the cell get splitted and the cells reach the G1 condition, chromosomes=2n, DNA=2c

Mitosis

- Prophase: chromosomes condense and become visible
- Prometaphase: chromosomes are linked to the spindle fibers.
- Metaphase: chromosomes are lined up at the equatorial plane of the cell
- Anaphase: sister chromatids are pulled to the opposite poles by microtubules
- Telophase: chromatids are separated to the opposite poles, nuclear membrane reforms and cell divide.

In plants: occurs the formation of a new cell plate.

Cytokinesis: separation of the membrane to form 2 cells, first the cell is compressed by a contractile ring that divides the cell in nearly equal halves.

By now the organelles in the cell have been replicated, and are now divided between the two halves of the cell.

Nuclear spindle (metaphase and anaphase)

The nuclear spindle is a structure that is useful to move and to separate the chromosomes through mitosis and meiosis.

It is composed by two **centrosomes** from which start some long fibers called **microtubules**, the microtubules are composed of tubulin, proteins that can assemble and disassemble.

Those fibers grow, they reach the **kinetochore** and then chromosomes are aligned through opposite forces applied by the fibers that are pushing from opposite sides then microtubules start to disassemble (from the chromosome extremity), becoming shorter they pull apart the sister chromatids.

The sister chromatids were completely joined from cohesin, and by other proteins that protect cohesin until it is time to separate them.

The fibers of the nuclear spindle are torn down before the start of the telophase, in which the freed chromosomes are surrounded by a new nuclear membrane.

Germ cells

Meiosis

Meiosis produces haploid gametes that have mixed chromosomes in respect to the parental cell.

S phase is identical to mitosis, before starting meiosis we have pairs of homologous chromosomes.

1. First meiotic division: reductional

- Prophase: paternal and maternal homologous chromosomes join each other (synapsis phase) forming the synaptonemal complex along the whole length of the chromosome, they form a tetrad consisting of 4 chromatids.
Along the synapses are formed recombination nodules that will become the chiasmata of the **crossing over**: regions that delimit the portion of DNA exchanged between non-sister chromatids during recombination; after the crossing over the synaptonemal complex dissolves and the homologous chromosomes start separating, they remain in touch just in correspondence of the chiasma.
- Metaphase I: tetrads are attached to the nuclear spindle, each homologous kinetochore is reached by an opposite pole of the spindle, the tetrads get lined up along the metaphase plate
- Anaphase I: the centromere does not divide, each centromere is pulled apart, dissolving the chiasmata: the sister chromatids remain in touch while the homologous chromosomes are separated and moved to the opposite poles.
- Telophase I: the nuclear envelope reforms and the resultant cells have half of the number of chromosomes, each consisting of 2 sister chromatids.
- Interkinesis: similar to the interphase but no chromosomal duplication takes place

2. Second meiotic division: equational

- Prophase II: chromosomes condense, centrioles move toward the poles and the nuclear envelope is broken down
- Metaphase II: chromosomes align at the metaphase plate, sister chromatids attach to spindle fibers from opposite poles
- Anaphase II: chromosomes divide and sister chromatids move to opposite poles
- Telophase II: chromosomes begin to uncoil and the nuclear envelopes are reformed

- Cytokinesis: four new haploid cells are formed

Variation

Meiosis is a huge variation generator, it operates through the independent assortment of chromosome and through the crossing over:

Depending on how the **chromosomes align in metaphase 1** (homologous combinations) the possible resulting daughter cells are 2^n .

Adding the **crossing over** the possibilities multiply exponentially, creating new and unique combinations of genes.

LINKED INHERITANCE

The major discoveries about linked inheritance were carried out by Thomas Hunt Morgan, through his experiments he managed to prove the chromosomal theory of inheritance.

Morgan used the *drosophila melanogaster* (fruit fly) as a model organism, it had:

- Short life cycle
- Large progeny
- Easy to rear and cross in the lab
- Only 4 pairs of chromosomes
- Male and female organisms were easily distinguished
- Polytene chromosomes in the salivary glands, large chromosomes that derive from a non-segregation of a cell that has duplicated its DNA.

Genetic sexual determination in *drosophila*

The first discovery of Morgan was that male and female individuals had different chromosomes. Each species has a particular type of genetic differencing between male and female individuals, the differences in DNA provide the different features of each sex, in the case of *drosophila*, **female** individuals are provided with a pair of homologous chromosomes (**homogametic sex**) called XX while **males** have the heterologous couple XY (**heterogametic sex**); female flies have a more rounded abdomen and are provided with sexual combs.

Humans have the same genetic mechanism for sex determination as *drosophila* while, in other species males are characterized just by one X chromosome, while females have two of them.

Morgan's work

Morgan was just trying to confirm Mendel's model when he discovered **mutant gene** in the eye colour of flies: normally every fly has a gene that makes the eyes look red (that's the wild type allele, denoted with +), while white-eyed flies have a different gene (mutant allele w).

First cross:

	female ♀	male ♂
P1:	red eyes	white eyes
	↓	
F1:	$\frac{1}{2}$ red eyes	$\frac{1}{2}$ red eyes

	↓	
F2:	$\frac{1}{2}$ red eyes	$\frac{1}{4}$ red eyes $\frac{1}{4}$ white eyes

We can deduct that white is a recessive character but in F2 the proportions don't respect the mendelian model, the character appears only in males.

Second cross:

	female ♀	male ♂
P1:	white eyes	red eyes
	↓	
F1:	$\frac{1}{2}$ red eyes	$\frac{1}{2}$ white eyes
	↓	
F2:	$\frac{1}{4}$ red eyes $\frac{1}{4}$ white eyes	$\frac{1}{4}$ red eyes $\frac{1}{4}$ white eyes

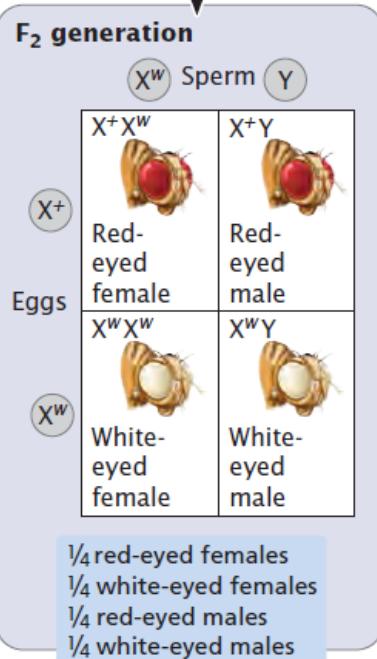
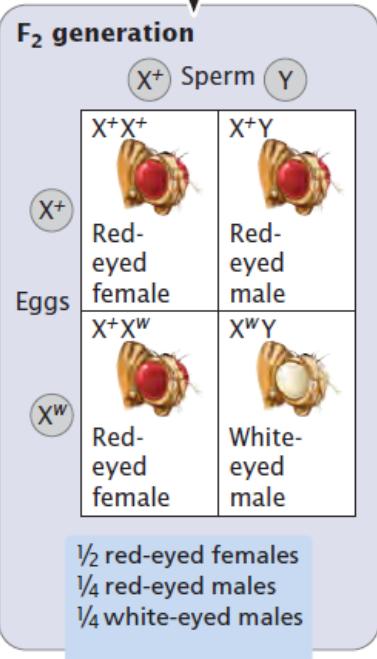
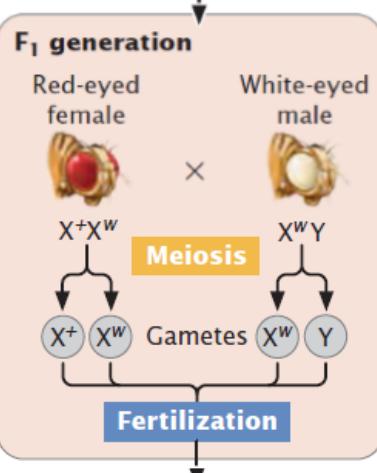
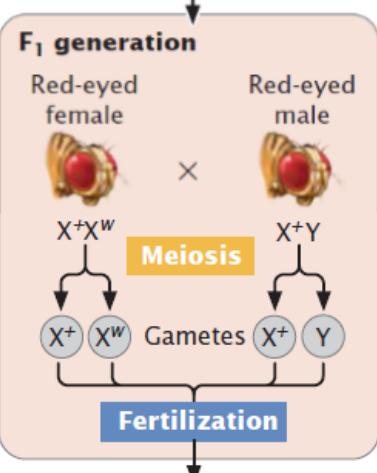
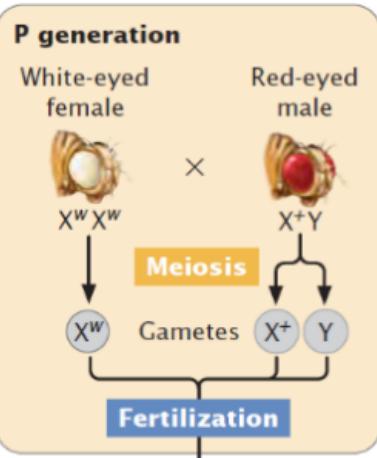
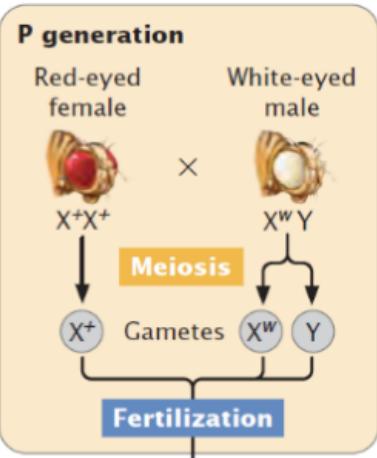
The hypothesis that Morgan elaborated is that the gene affecting the eye color is on the X chromosome, so that females can be homozygous or heterozygous and males are hemizygous, the presence of the gene on the single chromosome determines his appearance in the phenotype.

The possible genotypes are:

- Female with red eyes: X^+X^+ , X^+X^w
- Female with white eyes: X^wX^w
- Male with red eyes: X^+Y
- Male with white eyes: X^wY

At the time there still was no proof of chromosomal theory, here we see that a specific trait is transmitted by a specific chromosome

↓
Genes are located on chromosomes



Bridges' work

Bridges carried out Morgan's experiments on larger samples, he crossed white eyed female with red eyed male, obtaining half of red eyed females and half of white eyed males.

The difference between Morgan's results is that Bridges obtained about 0.25% of red eyed males which were sterile and 0.25% of fertile, white eyed females.

Looking at the chromosomes of the exceptional phenotypes Bridges found:

- $X^w X^w Y \rightarrow$ white eyed females
- $X^+ O \rightarrow$ infertile red eyed males

Bridges theorized that the sexual determination in drosophila is carried out just by chromosome X, in particular we can deduct the sex of an individual through the **ratio** between X chromosomes and the sets of autosomes (A)

- $X:A = 1$ female
- $X:A=0.5$ male
- $0.5 < X:A < 1$ intersex
- $X:A \geq 1$ abnormal development
- $X:A \leq 0.5$ abnormal development

Y chromosome is non involved in sex determination but it is **required for male fertility**.

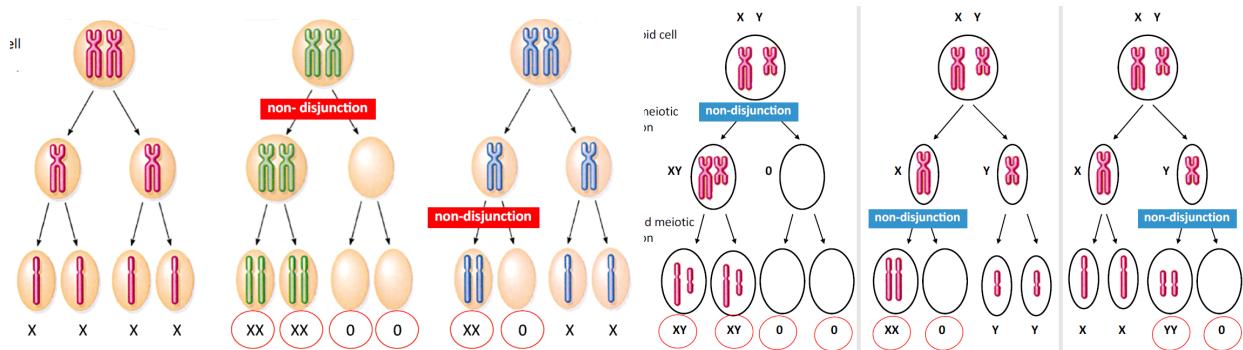
XX	female
XY	male
XO	sterile male
XXY	female
XXX	abnormal development
XXYY	abnormal development

The cases that Bridges registered were caused by an error in meiosis for which X chromosomes don't separate correctly, forming the gametes XX and OO:

		female
male	$X^w X^w$	OO
X^+	$X^+ X^w X^w$	$X^+ O$
Y	$X^w X^w Y$	Y O

The only survival genotypes are $X^w X^w Y$ (female with white eyes) and $X^+ O$ (infertile male with red eyes).

The **non-disjunction in meiosis** can occur in both male and female, at the first or second meiosis, forming all the possible abnormal gametes; only a few of them bring to an observable phenotype in a living organism.



Bridges's hypothesis was confirmed by observing the chromosome constitutions of the "exceptional" flies on the microscope, the "odd" pattern of inheritance of an X-linked trait was correlated to the "odd" chromosome constitution, proving that a specific phenotype was associated with a specific complement of chromosomes.

This was considered as the final confirmation of the chromosome theory of inheritance.

Sex differentiation in drosophila

The phenotypic differences between male and females in every species are determined by the different expression of genes: the little chromosomal difference generates a huge cascade of regulatory events that generates multiple differences, the proteins that get synthesized act on the alternative splicing, allowing only the expression of specific genes.

In the case of Drosophila the differentiation depends only on the **quantity of X chromosome** in the individual: a high dosage of X genes results in the expression of **Sxl protein** that promotes female specific **alternative splicing** of important regulatory factors that have the purpose to silence male genes and to express female characters.

In **male** flies the little amount of X chromosomes cause a **normal splicing** that includes a stop codon, female regulatory factors are not produced while the cascade of male regulators and proteins starts.

The splicing consists in the removal of introns (non synthesizing portions of DNA) from DNA sequences that get transcribed, alternative splicing consists in the possibility to leave introns and remove exons, changing the resulting synthesized proteins.

Genetic sexual determination in humans

In humans the presence of Y is compulsory for the development of a male, the differentiation between male and females starts from a gene on the Y chromosome which generates a cascade of regulatory events.

Sex chromosomes aneuploidies:

XX	female
XY	male
XO	Turner (sterile "female")
XXY	Klinefelter (sterile "male")

XXX

Triplo X (female)

X and Y chromosomes have originated from a pair of autosomes that lost most of the homology, they have different size and gene content.

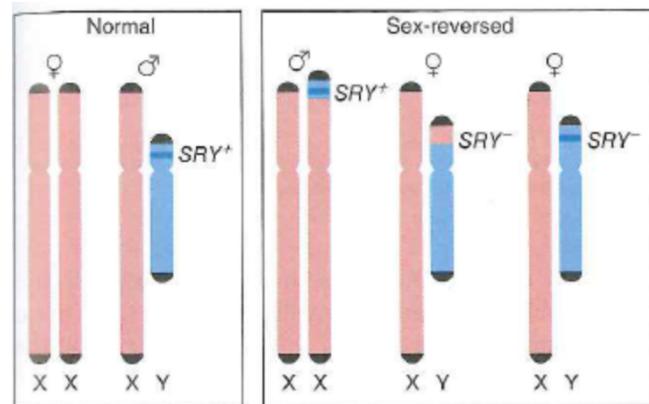
X and Y only have a small homologous region, the **pseudoautosomal regions** are located at the extremity of every sister chromatid, these regions are essential for the chromosome pairing in male meiosis, and these are the only regions in which it is possible to have crossing over. The main contents of Y chromosome are: SRY, male fertility genes and some essential genes shared with X chromosome; the majority of Y chromosome is heterochromatin.

Sex differentiation in mammals

In mammals, during early development the primitive gonads are undifferentiated (bipotential gonads), in males they will become testicles while in females they will develop into the ovaries. SRY is the critical sex determining gene on the Y chromosome, it leads to a cascade of regulatory events that form the testis through the AMH factor and that prevent the formation of female organs.

In absence of SRY gene the female genes don't occur in any problem during their expression. Mice experiment

Transgene DNA containing SRY gene is microinjected into the nucleus of a fertilised female mouse oocyte and implanted in a pregnant mouse; the result is a male mouse that has a female genotype and which is sterile. This is the sex reversal phenomenon, if any organism has a couple of sexual chromosomes that present the SRY gene it will appear as a male, while if it has any type of chromosome but it doesn't present the SRY gene it will appear as female.



Exceptions to this statement are the "differences of sexual development" (DSD), but they derive from secondary regulatory factors that present a dysfunction.

DOSAGE COMPENSATION

The chromosome-based sex determination creates a different gene dosage between male and females: males will have 1 X chromosome and females will have 2 X chromosomes.

So the cells have developed a dosage compensation system.

The sexual chromosomes are the only ones that can benefit from this system, all other kinds of misdosage (trisomy) are not tolerated, except for chromosome 21 that brings down syndrome. X chromosome has about 700 genes while Y chromosome has 50 genes, to balance the genetic information between females (700*2 genes) and males (700+50 genes) one of the two X genes in females gets inactivated.

(in drosophila there is a higher expression of the male X chromosome, and in other organisms there could be a lower expression of the couple of female chromosomes)

Barr body

This mechanism was discovered by Murray Barr, by looking at a female cells that were intercurring interphase he could see darker regions that were not present in male cells, he supposed that they were chromatin bodies, from which derive the name **Barr body**. Onho formulated an hypothesis according to which the Barr body is the condensed form of an inactivated X chromosome.

Mosaicism

The inactivation of a X chromosome obviously influences the expression of the genes that were present on that chromosome.

According to Mary Lyon's hypothesis the random **inactivation** of a X chromosome **occurs** at the blastocyst stage during **embryonic development**; once established, the inactivated state is transmitted through mitosis through cell generations, forming **clonal patches** (cell with equal inactivation).

The adult female individual results in a patchwork (mosaic).

This characteristic is easely seeable on the outside of female mammals, only one allele is expressed in each cell so the proteins produced from each cell randomly vary, that's what makes the colored coat of cats.

The confirmation of this theory comes from putting mammal cells in a culture, the resulting cloned cells always present just one of the two alleles that were present in the starting cells.

Counting mechanism

Some mammalian cells can have an abnormal number of X chromosomes, in some of these cases the cell is able to **vary its number of Barr bodies**, in the case that an individual has just one X chromosome (male or female) there will not be any Barr body, in the case that the cell has more than two X chromosomes there will be $nX-1$ Barr bodies.

In most of these cases the organisms are affected by syndromes that don't act on the survival of the individual but on its fertility.

We don't exactly know how this mechanism works.

Partial inactivation

We can experimentally see that some parts of the X chromosome remain unpacked:

- **Par regions:** these regions have to remain unpacked because they are present also on the Y chromosome, so for each gene appertaining to these regions there are two alleles in both males and females.
- **“Random” genes spread along the X chromosome:** some genes spread along the X chromosome are expressed, this generate an obvious difference from the males (males don't have the possibility to express alleles for genes in the central zones of the X chromosome) and also between the females individual, because these activation are not equal for each organism.

We do not know how and why this mechanism works.

Epigenetic event

Epigenetic modifications are changes in the gene expression, due to environmental or regulatory influences, that do not derive from a change in the DNA sequence, but these changes are still heritable between cell generations, through mitosis.

In the offspring epigenetic modifications are reestablished, starting from scratch (we can observe epigenetic changes that get transmitted between parental and filial generations, there is an active discussion about this topic; this could mean that modification *induced by the environment* could be passed through generations).

In the majority of cases epigenetic changes are set during the embryonic development, to differentiate the cells between different tissues.

↓

Though epigenetic changes cells acquire their identity

Chromatin structure:

- **Heterochromatin:** DNA double helix is wrapped around the histones forming nucleosomes, and then the nucleosomes are tightly packed together.
In this state genes are not expressed.
(ex. Chromosomes during cell division, Barr body, Satellite DNA: compose repetitive regions as centromeres, they have a structural function)
- **Euchromatin:** DNA is wrapped around the histones but there is space that allows the enzymes to act on the genetic information to translate or transcribe it.
Sometimes the sequences that are useful to start these processes (promoters) are hidden around the histones, so specific factors (chromatin remodelling complexes) recognize and free those sequences.
(ex. DNA during the interphase)

Epigenetic modifications

- **Histone acetylation and methylation**

Histones are composed of 4 types of proteins that form the central part around which the DNA is stacked, each histone has several tails that are used to keep the DNA in place, these tails have a protein composition and they have specific amino acid sites that undergo modification to pack or unpack the genetic information.

- **Acetylation:** specific enzymes, HAT, are responsible to add acetyl groups to the histone tails in order to loosen their structure and open the chromatin.
Deacetylation, on the other hand, is catalysed by HDACs and it closes the chromatin.
- **methylation:** the addition of a methyl group doesn't have a specific effect such as acetylation but it closes or opens the histone tails depending on the amino acid modified.

The histone modifications influence each other, the acetylation of a histone can produce acetylation in nearby histones, for this reason the DNA is divided into areas of acetylated and deacetylated histones called **domains**.

Barrier insulator: protein complex that divides different domains, it blocks the spread of heterochromatin factors.

- **Non coding RNA molecules**

The chromatin of one X chromosome is condensed through the action of a non coding RNA molecule.

This mechanism starts from the **X inactivation centre**, a portion of the chromosome found on the long arm, close to the centromere.

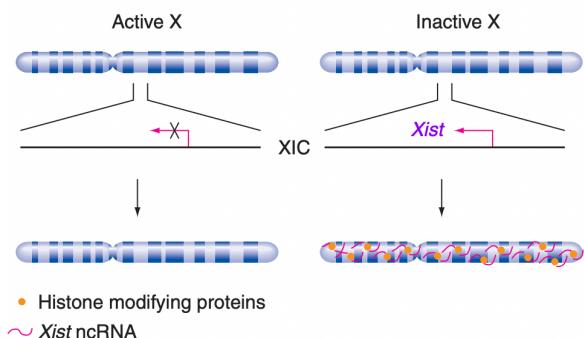
It was discovered by studying several cases in which the X chromosome was rearranged, the inactivation centre is exchanged between the chromosome X that is supposed to remain active and an autosome; in this way it remains only one chromosome X with an inactivation centre, it will be the only one to produce non-coding RNA.

The inactivation centre contains one critical gene for X inactivation: XIST.

During embryo development X chromosomes are active, at some point one of them starts the inactivation by the transcription of XIST gene.

It is not known how the two homologs undergo different paths but one of them becomes inactive through the association of the RNA to the whole chromosome from which it was transcribed.

The binded RNA recruits some histone modification proteins that progressively change the active status of DNA.

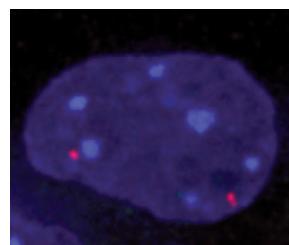


Experimental proof: fluorescent in situ hybridization

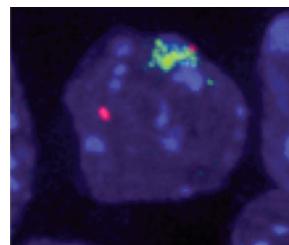
Technique used in molecular biology experiments.

It uses labeled sequences of DNA or RNA, those are complementary to the Xist RNA sequence and to the DNA sequence of another gene on chromosome X.

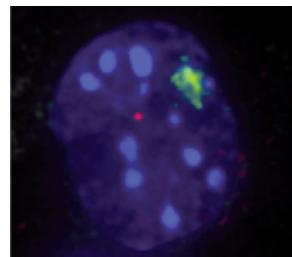
These sequences are injected in stem cells (undifferentiated, still without Barr body) and we can see:



1. The two red dots correspond to the alleles of the random gene of the X chromosome, they are both in an euchromatin state.



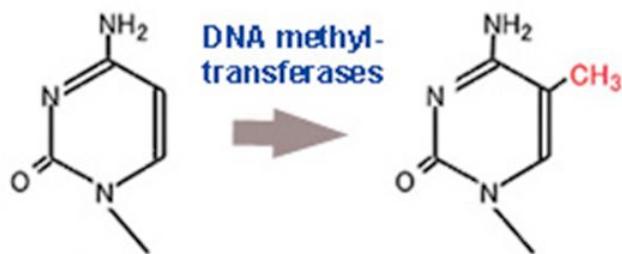
2. the inactivation starts from the centre of the chromosome, we can still see the two red dots because the Xist sequence still hasn't reached the red gene loci.



3. only one red allele has remained, from now on this cell and all its future generations will express only that allele.
The Barr body is completely formed.

- **DNA methylation**

Now that DNA is compacted and that transcription is repressed, it is necessary to create a memory mechanism that transmits the epigenetic information through cell generations. One hypothesis for the epigenetic transmission is based on the use of DNA methylation; DNA methylation consists in the addition of a methyl group to the DNA nucleobase (base azotata) usually in correspondence of C-G sequences because these sequences identify regulatory genes, they are the promoters of those genes, if the promoter is methylated the gene is not expressed.



The DNA methylation is grouped in the epigenetic modifications because it only changes the structure of DNA not its sequence.

Another important function of DNA methylation is the memory mechanism:

During S phase, when the DNA is replicated, each single filament of DNA is used as a mold to synthesise the new filaments, when the DNA polymerase finds a methylated nucleotide it cannot reproduce the methylation but soon after another enzyme (a methyltransferase) that is able to recognise the methylations and to copy them on the newformed strand.

In this way the inhibitions in the genetic information are maintained and the new DNA is a perfect epigenetic copy of the original one.

PEDIGREE ANALYSIS

Human being is not a suitable model organism, we actually know many anatomic and structural informations but:

- Controlled matings are not possible
- Long generation time
- Small families
- Few monogenic traits that follow the mendelian genetics (tongue rolling, freckles red-green colorblindness)

Human genetics has been possible only thanks to molecular genetics, it was not possible to have a mendelian approach (experiments).

Garrod's work

He studied a human trait that seemed to follow the Mendelian model, it was a metabolic disorder (alkaptonuria) that produced a brownish color in the urine.

At first it was thought that it could have been caused by the environment but the correlation between individuals of the same family was too strong.

Garrod discovered that the trait was transmitted to the offspring following the Mendelian recessive model, and that it was due to a genetic defect.

The successive correlation that Garrod was able to find was that the genetic defect brought affected people to be lacking an enzyme, involved in the chemical breakdown of protein, this generated the metabolic disorder.

This was the first Mendelian trait found in humans and for the first tiny genes were associated with enzymes; until 80'-90' there wasn't much to do, the only way to study inheritance was to look at family trees.

Pedigrees

Notation

- Female:  male: 

- If female and male have children there is a line between them



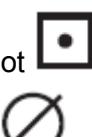
- If parents are consanguineous there is a double line between them



- People affected by the trait of interest are colored



- Heterozygous individuals present a dot



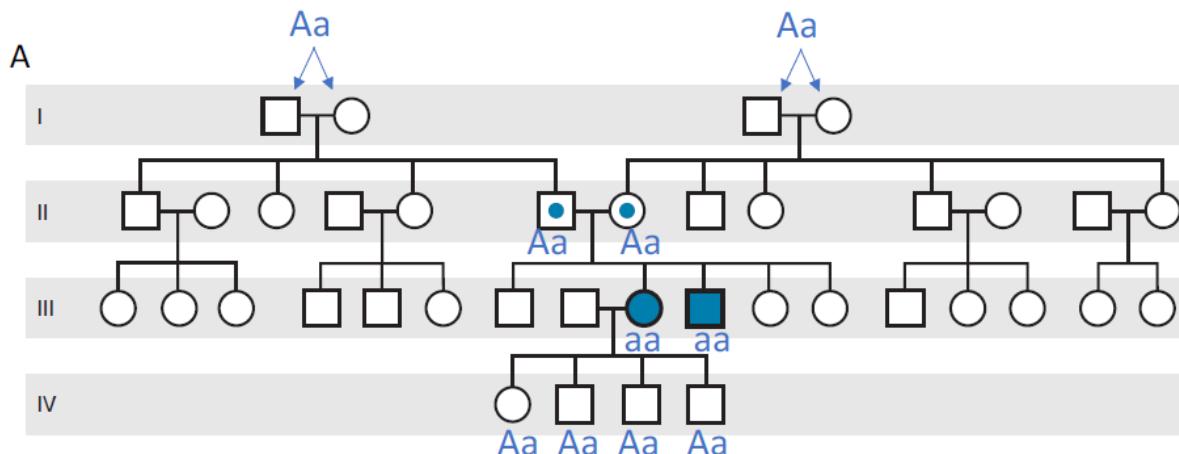
- Deceased individual are crossed out



- The older generations are at the top, while younger ones at the bottom, they are identified by roman numbers from the top to the bottom

- Along the same generation individuals are identified from left to right with Arabic numbers.
- Every trait studied through pedigree is a rare trait and follows a mendelian pattern
- By the pedigree analysis we can infer the inheritance pattern of the trait:
 - Autosomal / sex-linked
 - Dominant / recessive
- We ALWAYS consider the model that has less external carriers of the rare genes from the principal pedigree.

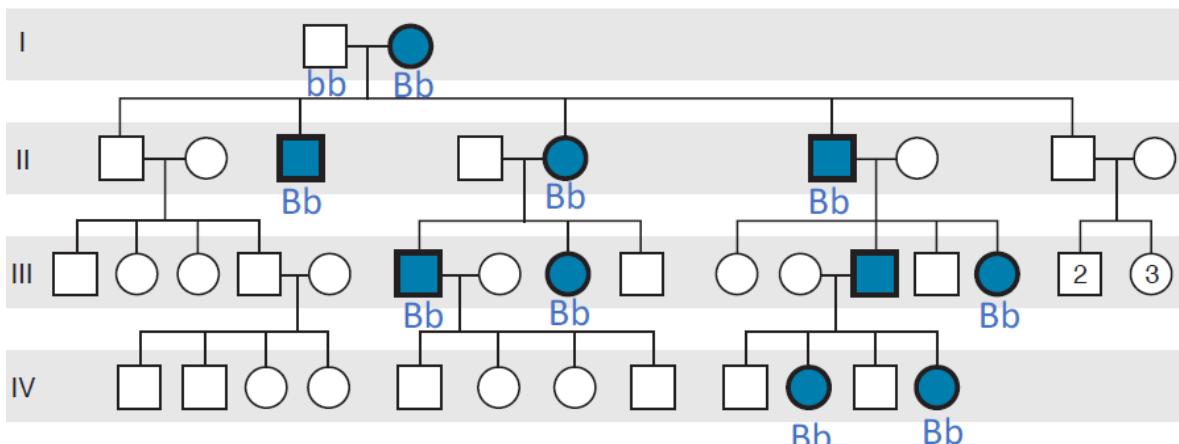
Autosomal recessive



To recognise an autosomal recessive trait we need to determine some assumption and check if the pedigree matches.

1. If an individual is infected his parents must be heterozygous, the trait appears only in few generations
2. If an individual is heterozygous one of his parents must be heterozygous
3. If parents are consanguineous, it's more likely that they are both carrier of a rare recessive trait, they have inherited the gene from a common ancestor (generally in a whole population is hard to find an individual with the same recessive allele)

Autosomal dominant



assumptions:

1. If an individual is affected he is heterozygous
2. If an individual is affected one of the parents is recessive and the other one is heterozygous, the phenotype appears in every generation
3. If an individual is affected his children have a chance of 50% to inherit the trait

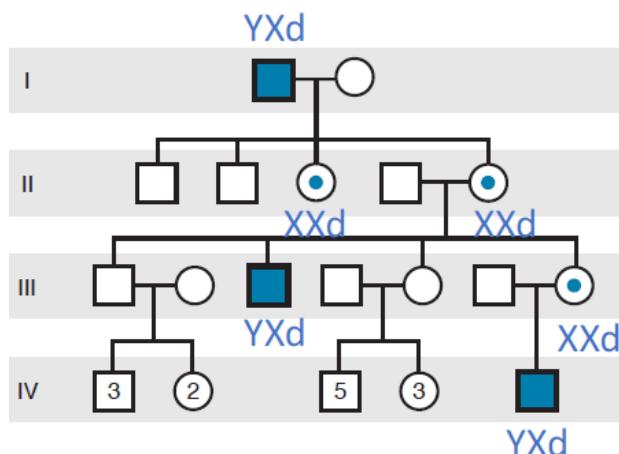
Isolated cases of a dominant genetic disease usually due to a new mutation, de novo mutation, it usually seems a recessive trait but by sequencing the DNA we can see that both parents do not have the trait.

The risk of having another child with the same mutation is very low, it depends from the stage of the germline in which the mutation happened, it is difficult to pervade.

X-linked recessive

It has no sense to talk about dominance or recessiveness in males, the distinction is among females.

The incidence of recessive traits is much higher in males than in females.

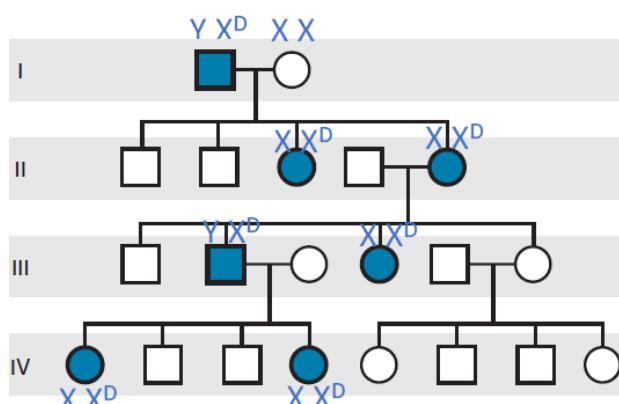


Assumptions:

1. If a male is affected his mother is heterozygous, the father doesn't influence the outcome
2. If a female is affected her mother is heterozygous and her father is affected

X-linked dominant

More probability of female affected



Assumptions:

1. If male is affected his daughters will be affected
2. If a female is affected she has 50% chance to pass the phenotype to the offspring with no distinction

Effect of X inactivation

In reality even in females it's difficult to determine if the trait is dominant or recessive.

Due to X inactivation each cell of the female organism expresses only one allele, so in both recessive and dominant heterozygous organisms the phenotype is mosaic.

Recessive: half of the cell of an heterozygous female will show the recessive phenotype

Dominant: half of the cells of an heterozygous individual will not show the phenotype, in the case that the protein synthesized by the gene remains localized in the cell the organism results patched; instead, if the gene codes for a secreted protein there will be an averaging quantity of the protein in the whole organism.

Ex. hemophilia

In hemophilia the affected organism lacks a coagulant factor in blood, if a woman is heterozygous for that gene her levels of coagulants are still sufficient to result as a non affected individual.

In Europe's royal families there has been a huge spread of hemophilia and a lot of males died, because of consanguineous matings.

Ex. muscular dystrophy

This disease causes degeneration of muscles, it is caused by a miss functionality of dystrophin, proteins that connect the actin of a muscle fiber to the surrounding extracellular matrix through the cell membrane.

The gene that causes this disease is situated on the X chromosome so 1 male over 3000 females is affected.

Ex. red-green colorblindness

Another group of genes situated on the X chromosome are the ones responsible for the sight, usually the blue gene is not affected and one between red and green loses its functionality.

EXTENSION OF MENDELIAN LAWS

Classical Mendelian modes of inheritance

Phenotype is determined by a single gene with just 2 alleles, in the simple dominant/recessive relationship.

Many scientists started investigate traits in different organisms to find extensions and exceptions to mendelian laws, infact, in the reality the **majority of genes are complex genes**, influenced by multiple factors:

- Multiple loci for the same phenotype
- Environment influences

Basic extensions (complications) of Mendelian inheritance

For now we always consider that **every phenotype is mainly influenced from 1 locus**.

Types of dominance:

	A^1/A^1	A^2/A^2	A^1/A^2 hybrid	
complete				A^1 dominant on A^2
complete				A^1 recessive vs A^2
incomplete				The hybrid is a mesh
codominance				The hybrid express simultaneously the two alleles

Incomplete/partial dominance

This phenomenon was observed by mendel in a cross between late-blooming peas and early-blooming peas, the result were plants that **bloomed** in between of the two extremes, for this reason mendel discarded this trait.

Another perfect example is the flower color of snapdragons, they have red and white flower genes; if we cross homozygous white flowers with homozygous red flowers we obtain all heterozygous pink flowers.

If we self breed the filial generation we obtain $\frac{1}{4}$ of red flowers, $\frac{1}{2}$ of pink flowers and $\frac{1}{4}$ of white flowers, we can infer that the **heterozygous phenotype is different from both homozygous ones and it is a mesh of the parental colors**.

This type of dominance is highly common, it is actually observable in much more phenomena if we consider a deeper level of observation.

Ex. hypercholesterolemia

Normally on the membrane of some cells of an individual it is present an LDL receptor that is useful to locate and absorb bad cholesterol molecules from the blood.

The wild genotype is recessive and it is identified by HH, affected individual present h gene, which removes the receptors, this results in high levels of LDL in blood.

The disease is considered as a dominant inheritance because heterozygous individuals result with high levels of LDL too; in the reality the inheritance is partial dominant, the **heterozygous individuals present half of the receptors of a healthy individual but they are not enough to prevent them from accumulating LDL.**

The disease results to have a severe form hh and a mild disease Hh

Ex. round/wrinkled pea seeds

Mendel considered this trait as a complete dominance of the round trait over the wrinkle one.

In the reality the texture of the seed is based on the amount of starch that it contains, RR individuals have a huge amount of starch, the shape results round, **Rr individuals have an intermediate amount of starch that still permits the seed to look round**, while rr individuals have a little amount of starch and the seed is highly wrinkled.

Codominance

Heterozygotes display both characteristics

Ex. ABO blood group system

Our blood type is controlled from **3 types of alleles**, that are able to determine 4 blood types to which correspond 4 types of antibodies.

- **Antigens:** molecules on the surface of the blood cells, they are what characterises the blood type
- **Antibodies:** molecules of the immune system, they are able to attack and eliminate foreign antigens (different from the ones belonging to the organism)

The genes that characterise the blood type are 3: I^A , I^B , i

I^A and I^B are codominant, they encode for a glycosyltransferase, an enzyme that modifies the carbohydrate tails that are present on the surface of the blood cells.

i gene is recessive in respect to the other ones, in homozygous individuals there is no modification to the carbohydrate tail.

Blood types

- A: $I^A I^A$ or $I^A i$, this gene adds a particular sugar to the carbohydrate tail, making it become an antigen A.
anti-B antibodies are produced, it can receive from A or O.
- B: $I^B I^B$ or $I^B i$, this phenotype has a different sugar bonded to the tails on the blood cells, the set of carbohydrate tail and banded sugar identifies the antigen B.
anti-A antibodies are produced, it can receive blood from B or O
- AB: $I^A I^B$ both antigens are produced.
None antibody is produced, it can receive blood from every blood type

- O: i i, the carbohydrate tails don't undergo any addition.
Both anti-A and anti-B antibodies are produced, it can receive blood just from O blood types

Multiple alleles of a gene

Ex. coat color in rabbits

4 different genes:

- C encodes for black coat, it is dominant to all the others alleles
- c^{ch} encodes for chinchilla (grey) coat, it is incompletely dominant to c^h and c alleles
- c^h encodes for himalayan (spots black and white) coat, it is completely dominant to c allele
- c encodes for albino coat and it is completely recessive

Genetic explanation:

C gene encodes for an enzyme needed for the synthesis of the melanin pigment.

c^{ch} allele encodes for a partially functioning enzyme.

c^h encodes for a temperature sensitive allele, it works only at low temperature, so the melanin pigment is produced just at the extremities of the rabbit body.

c allele encodes for a completely non functional enzyme.

Recessive lethal alleles

If specific genes are present in an organism they may cause an arrest of development, that's the case of agouti gene in mice.

Ex. agouti gene

The agouti gene controls the distribution of melanin pigment in the hair of mammals, when agouti gene works yellow pigment is deposited as a band on the black hair of a mouse.

Melanin in mammals:

In mammals there is a biological pathway that can result in production of:

- Eumelanin: brown-black
- Pheomelanin: yellow-red

Agouti blocks a signal in this pathway and pheomelanin is produced, usually this occurs just in a short period in the life of a mouse and just on its skin.

When the agouti yellow gene is produced agouti gets expressed in all tissues and for the whole hair growth of a mouse.

The alleles that control agouti are:

- a allele: normal functionality of agouti
- A^Y allele: completely yellow hairs and obesity

If we cross two agouti mice the offspring will be normal agouti mice.

If we cross one agouti and one yellow mouse the result will be 50% of times an agouti mouse and the other 50% a yellow mouse.

This result seems to be in consonant with a Mendelian scenario where yellow gene is dominant over agouti.

In reality something strange happens when we breed two agouti mice, we obtain $\frac{1}{3}$ agouti and $\frac{2}{3}$ yellow mice, we can deduce that we are missing the homozygous yellow mouse.

	a	A^Y
a	aa	aA^Y
A^Y	aA^Y	$A^Y A^Y$

The gene is recessive lethal because life is not allowed when an individual is homozygous for that gene.

Why is $A^Y A^Y$ not possible?

The physiological reason of the death of homozygous individual is the lack of raly gene; the A^Y allele is an alternative version of the gene that does not modify the main instructions of the agouti gene but it removes the agouti promoter and the raly gene.

By removing the agouti promoter the agouti gene is no more finely controlled, it is expressed in every tissue of the mouse for his whole life, it is actually controlled by the raly promoter.

The raly gene encodes for an important protein, required for embryonic development, the lack of this protein is the true reason for the death of the mouse.

Multiple phenotypes

If the same genes have effect on multiple genotypes of an organism their dominance relations can vary depending on the phenotype under consideration

Ex. $Hb\beta^A$ and $Hb\beta^S$

Phenotypes at Different Levels of Analysis	Normal $Hb\beta^A Hb\beta^A$	Carrier $Hb\beta^A Hb\beta^S$	Diseased $Hb\beta^S Hb\beta^S$	Dominance Relations at Each Level of Analysis
Red blood cell shape at sea level	Normal	Normal	Sickled cells present	$Hb\beta^A$ is dominant $Hb\beta^S$ is recessive
Red blood cell concentration at sea level	Normal	Normal	Lower	
β -globin polypeptide production	A protein	A and S proteins	S proteins	$Hb\beta^A$ and $Hb\beta^S$ are codominant
Red blood cell shape at high altitudes	Normal	Sickled cells present	Severe sickling	
Red blood cell concentration at high altitudes	Normal	Lower	Very low, anemia	$Hb\beta^A$ and $Hb\beta^S$ show incomplete dominance
Susceptibility to malaria	Normal susceptibility	Resistant	Resistant	$Hb\beta^S$ is dominant $Hb\beta^A$ is recessive

Environment influences

consequences of how gene expression is affected by environment and genetic background

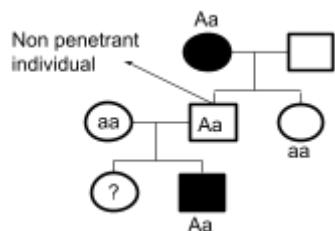
Incomplete penetrance

Penetrance: **percentage of individuals** having a particular genotype **that express** the expected **phenotype**.

Ex. the phenotype Pp of a particular gene has a percentage of 80% to show the P phenotype and a 20% chance to show p phenotype.

Dominant traits that admit non-penetrant individuals (in heterozygous genotypes), individuals that show an alternative phenotype, present incomplete penetrance, it is a population value.

Ex. non penetrant parent



The probability that the unknown individual is unaffected are:
The probability that he is homozygous recessive + the probability that he is heterozygous but non penetrant.
Considering a 80% penetrance, we have that 20% of heterozygous do not present the phenotype so the probability results:
 $\frac{1}{2} + (\frac{1}{2} * 0.2) = 0.6$

Ex. sixth finger

The abundance of fingers in hands or feet is a dominant allele in humans, this trait has a penetrance of 75%, that is to say that the heterozygous version of this gene shows the extra finger only 75% of the times.

Ex. ovarian cancer

The inheritance of cancer follows a particular mechanism.

To develop a cancer in an organism it is necessary to have a double mutation in the genes that function as the “brakes” of the cancer development, these genes are the **tumor suppressors**; when an individual gets both tumor suppressor alleles mutated the cell loses its possibility to control the cell division and it verifies an **excessive proliferation**; a cancer is developed.

For this reason tumors can be considered as dominant autosomal traits with a low penetrance: If an individual is affected by cancer he will pass a mutated allele to the offspring so the probability that the offspring will spontaneously develop a tumor are multiplied (inherited tumor). The penetrance of the trait can be seen as the ratio of individuals that develop a cancer from the initial mutation over those that just bring one mutation.

The penetrance percentage depends on the environment, for example the age of an individual highly influences the possibility that a mutation occurs, so going on with the age the probability of developing a tumor grows even for those ones who have not inherited a mutation (sporadic tumors), obviously if the tumor verifies when the organism is old there is no possibility to transmit the mutation to the offspring.

Variable expressivity

Some genotypes can cause different phenotypes for no known reason; this scenario verifies in many genetic diseases, different individuals, with equal mutations, undergo complications of very different degrees (from light to very severe symptoms).

Sex limited and influenced characters

Sex limited traits are autosomal genes that can be present in both males and females but they are expressed only in the sex that has a specific hormonal activation.

Ex. breast cancer

Sex influenced traits are traits that can appear in both sexes but that are expressed to a different degree in each: the character seems to act as dominant in one sex and recessive in the other

Ex. male pattern baldness

GENE INTERACTIONS

It's now time to consider characters influenced by multiple loci.

The interaction between multiple genes to outcome a single phenotype is usually due to a role in the same metabolic pathway.

Ex. signal transduction pathway

This metabolic pathway consists in the amplification of a signal coming from a receptor of a cell to activate multiple cellular responses; to amplify the signal are involved many different enzymes, encoded from different genes that have to cooperate to grant the right cellular response.

The genetic interaction can be very difficult to infer.

Genetic analysis

The same or indistinguishable phenotype can be due to mutation in different genes so we need a **series of techniques to build a genetic map of a phenotype**.

Ex. hereditary deafness

Often deafness is caused by environmental causes but there is also a congenital deafness; hearing is determined by a huge amount of organic components, we can clearly comprehend how many genes are involved at very different levels.

A single mutation in one of those genes can cause the same final outcome, deafness.

A great experimental process to understand the mechanism behind a character is studying the mutant organisms.

Two main techniques can help us in digging behind the mutation of one phenotypic character: complementation analysis and dihybrid crosses

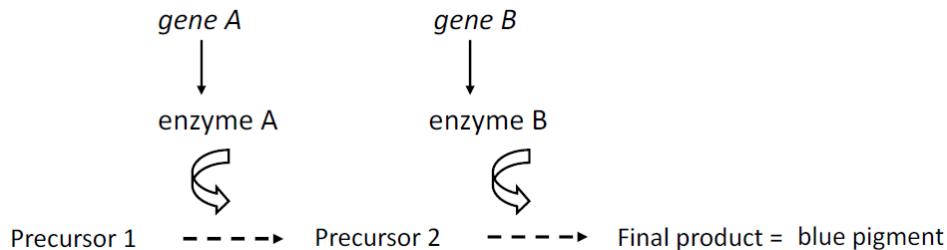
Complementation analysis

This technique is **only useful** in the case of **recessive alleles** that cause the mutation.

Ex. two equal phenotypic mutations

Suppose we have a wild type flower, having a blue color, we find 2 plants that are producing white flowers; we want to know if those two plants present the same mutation on the same gene or if they are affecting different genes of the same metabolic pathway.

Let's assume that we have a scenario like this one:



The lack of one or the other enzyme causes the non production of blue pigment, the flower result white.

We assume that the mutation of the first flower (M1) is affecting the gene A, inhibiting the production of enzyme A (aa genotype) and that the second mutation (M2) is affecting gene B, causing a lack of enzyme B (bb genotype).

According to this hypothesis the two mutants should have the genotypes:

M1: aa BB M2: AA bb

In this case the cross of these two mutant gives

F1: aA Bb

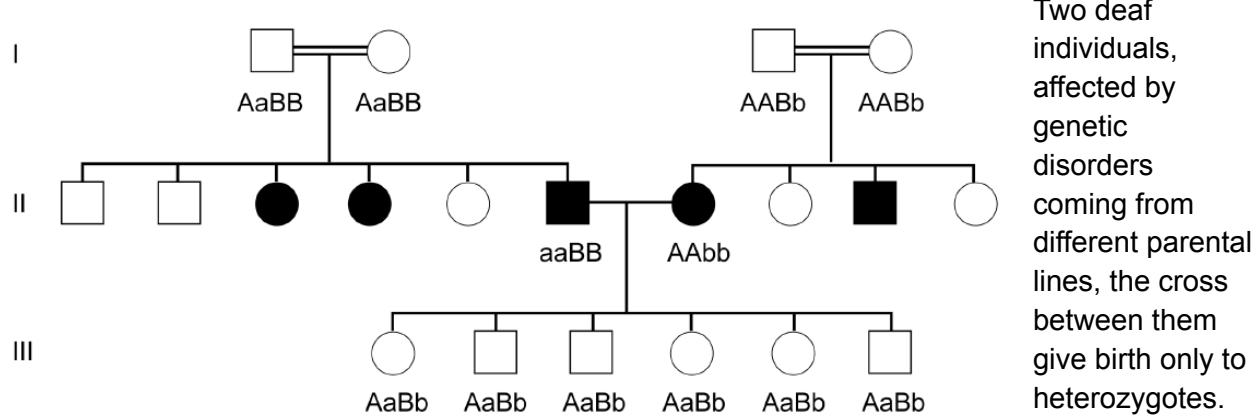
The **resulting individual has at least one functioning dominant allele** so both enzymes are produced and the resulting flower color is blue.

In this case we had **complementation**.

Supposing the two mutants were affected by a mutation on the same gene we should have had:

M1: aa BB M2: aa BB → F1: aa BB still a white flower

Ex. recessive hereditary deafness



Ex. drosophila eyes

Other than the white eyes mutants discovered by morgan drosophila can have a lot more of wye colors, in particular, we want to investigate the apricot eyes mutation.

We try the complementation test:

White female apricot male

$$X^W X^W \times X^A Y$$

$$F1: X^W X^A \quad (\text{and } X^W Y)$$

The result are white eyed female, there is **no complementation** between the characters and **white allele is dominant** over apricot: W dominant over W^a

The two mutations are just different alleles of the same gene.

Now let's try the complementation test on all eye colors available:

Mutation	Apricot	Brown	Buff	Carnation	Cherry	Claret	Coral	Vermilion	White
Apricot	-	+	-	+	-	+	-	+	-
Brown		-	+	+	+	+	+	+	+
Buff			-	+	-	+	-	+	-
Carnation				-	+	+	+	+	+
Cherry					-	+	-	+	-
Claret						-	+	+	+
Coral							-	+	-
Vermilion								-	+
White									-

Complementation

Group	Mutant (allele)
I	Apricot (w^a), buff (w^b), cherry (w^{ch}), coral (w^{co}), white (w)
II	Carnation (c)
III	Claret (cl)
IV	Brown (b)
V	Vermilion (v)

The **positive** outcomes are the successful **complementations**, when a complementation has success we can infer that the two traits are due to mutation on different genes, they belong to different complementation groups.

Complementation group: group of alleles of the same gene that do not complement each other.

Traits from different complementation groups give complementation when crossed.

In drosophila eyes we can find 5 different groups, in particular the fourth and the fifth group will show up again.

To display the complementation groups we just have to put together the traits that show a negative result in the table; complementation analysis gives us an idea of how many genes are controlling the phenotype at least.

Dihybrid crosses

This technique is the same used by Mendel to formulate his second law, dihybrid crosses can give us way more information than what we think.

Types of interaction:

- None

Each gene controls different characters, they are in different loci that are able to assort independently, dominant/recessive relationship (Mendel)

9 genotypes

4 phenotypes, 9:3:3:1

- ### • Addition

The two genes produce different products, both used to build up a phenotypic character. The sum of the effects of the single genes result in the product.

Ex. additive interaction in eye color of drosophila

From what we have studied before:

We consider 2 complementation groups:

IV: Brown (bw)

V: scarlet (st)

These two mutant genes are both recessive to the wild type bright red eyes.

They are also complementary to each other, they are mutations of different genes, if an individual has both dominant forms it results in the wild type.

We can observe:

bw / bw st+ / st+ x bw+ / bw+ st / st

F1: bw+ / bw st+ / st

Wild type organism, bright red eyes

Now we perform the dihybrid cross:

b) $F_1 \times F_1$ bw+/bw st+/st x bw+/bw st+/st

F_2 generation

F_2 phenotypic ratio for $bw^+/bw \times bw^+/bw$	F_2 phenotypic ratio for $st^+/st \times st^+/st$	Combined F_2 ratios	Expected F_2 phenotypic proportions
$\frac{3}{4} bw^+/-$	$\frac{3}{4} st^+/-$	$\frac{9}{16} bw^+/-$	$\frac{9}{16}$ red (wild type)
	$\frac{1}{4} st/st$	$\frac{3}{16} bw^+/-$	$\frac{3}{16}$ scarlet
$\frac{1}{4} bw/bw$	$\frac{3}{4} st^+/-$	$\frac{3}{16} bw/bw$	$\frac{3}{16}$ brown
	$\frac{1}{4} st/st$	$\frac{1}{16} bw/bw$	$\frac{1}{16}$ white

We can see how the bright red phenotype is composed of brown and scarlet pigment, while the eyes of a fly without pigment result white, that's exactly additive interaction. The phenotypic proportions are equal to the independent genes but the product of two

genes is used for the same character.

N.B.

The white mutation of the first group works with a different mechanism, it affects the gene responsible for the transport of scarlet and brown pigments to the eye, so a direct in that gene doesn't permit to the pigments to migrate from where they are produced to the eyes.

White mutation has an epistatic (recessive) effect of these genes, it is necessary for their expression

- **Epistasis**

The two genes are part of the same pathway, they need one another for their expression, the expression of the gene that comes first in the pathway can influence the second one.

- Recessive epistasis: one recessive gene prevents the expression of another gene.

Ex. coloration of labrador dogs

1. TYRP synthesis of eumelanin:

B allele = black

b allele = white

2. MC1R produces the signal for eumelanin production:

E allele = ON, eumelanin pigment is expressed

e allele = OFF, yellow pigment is expressed

To perform a dihybrid cross we need to obtain a hybrid individual:

BB EE x bb ee

Black yellow

Bb Ee

The result is a black labrador

Now we can perform the dihybrid cross:

BbEe	X	BbEe		
BE	Be	bE	be	
BE	BB EE	BB Ee	Bb EE	Bb Ee
Be	BB Ee	BB ee	Bb Ee	Bb ee
bE	Bb EE	Bb Ee	bb EE	bb Ee
be	Bb Ee	Bb ee	bb Ee	bb ee

The obtained phenotypes are:

9/16 black

3/16 brown

4/16 yellow

Every dog that has ee genes has a yellow coat, it doesn't matter what types of B alleles it has.

So to have a brown coat we need dominant E and two recessive b.

Ex. ABO blood group

What we said a few paragraphs ago about blood groups is true, but according to that model, when we record the blood types of a population we should have a prevalence of A and B groups.

In reality the most common blood type is O, that's because there exists a precursor gene that permits the carbohydrate tails of blood cells to attach to the membrane of the cell itself.

If that gene (FUT1-) is expressed in a recessive form the other genes are not expressed, in the individuals that presents the dominant form (FUT+) we observe a normal distribution of the blood types

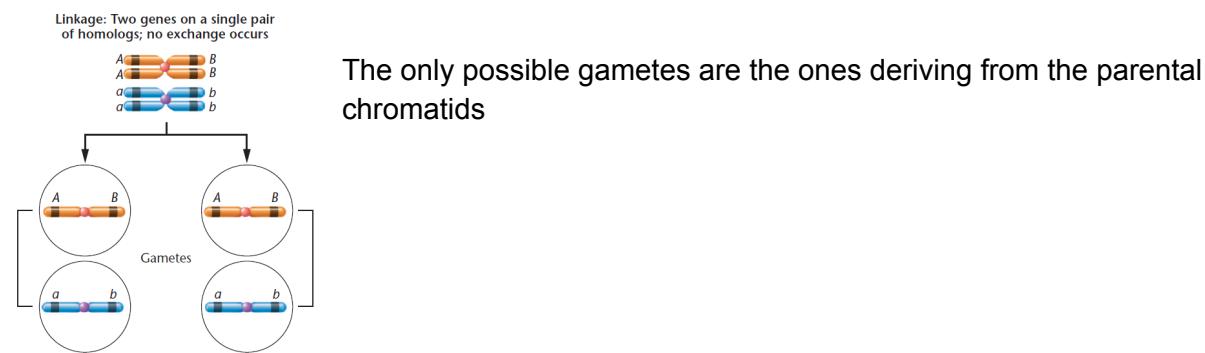
- **Dominant epistasis**: a dominant allele causes the inhibition of other genes
- **Duplicate recessive epistasis**: complementary genes, this case is the one used to explain complementary genes.
The same phenotypic mutation can be caused by two different gene modifications, in the case that these mutations are recessive, it is enough that one of the two genes is recessive to express the mutant phenotype.
The resulting ratios are 9/16 for the wild type phenotype and 7/16 for the mutant phenotype.
- **Duplicate dominant epistasis**: two dominant alleles cause a reciprocal inhibition, the only way to express a different phenotype is that neither of the genes has a dominant allele, the ratio is 15:1

TABLE 3.2 Summary of Gene Interactions

Gene Interaction	Example	F ₂ Genotypic Ratios from an F ₁ Dihybrid Cross				F ₂ Phenotypic Ratio
		A-B-	A-bb	aa B-	aa bb	
None: Four distinct F ₂ phenotypes	Lentil: seed coat color (see Fig. 3.10a)	9	3	3	1	9:3:3:1
Complementary: One dominant allele of each of two genes is necessary to produce phenotype	Sweet pea: flower color (see Fig. 3.12b)	9	3	3	1	9:7
Recessive epistasis: Homozygous recessive of one gene masks both alleles of another gene	Labrador retriever: coat color (see Fig. 3.14b)	9	3	3	1	9:3:4
Dominant epistasis I: Dominant allele of one gene hides effects of both alleles of another gene	Summer squash: color (see Fig. 3.17a)	9	3	3	1	12:3:1
Dominant epistasis II: Dominant allele of one gene hides effects of dominant allele of other gene	Chicken feathers: color (see Fig. 3.18a)	9	3	3	1	13:3
Redundancy: Only one dominant allele of either of two genes is necessary to produce phenotype	Maize: leaf development (see Fig. 3.19b)	9	3	3	1	15:1

LINKAGE ANALYSIS

The linkage analysis is a technique to create a primitive map of a chromosome. According to the Mendelian theory of inheritance during gametes formation each gene is independent from the others. In our previous models we considered genes as belonging to different chromosomes and the random alignment of these chromosomes created an equal chance for each combination. If we consider genes belonging to the same chromosome we have the following situation:



Morgan

Morgan was able to recognise which genes were present on the X chromosome of drosophila, so he started to study the behaviour of dihybrid crosses between genes on the X chromosome. In the first experiment he considered the mutant recessive genes white eyes (w) and miniature wings (m).

Now we will introduce a new representation for alleles, the two homologous chromosomes are separated by an horizontal line, in each line we put all the genes that we are considering.

P: $\frac{w\ m}{w\ m}$ X $\frac{w+m+}{Y}$
 ♀ white eyes, miniature wings ♂ wild type

F1: $\frac{w\ m}{w+m+}$ X $\frac{w\ m}{Y}$
 ♀ wild type ♂ white eyes, miniature wings

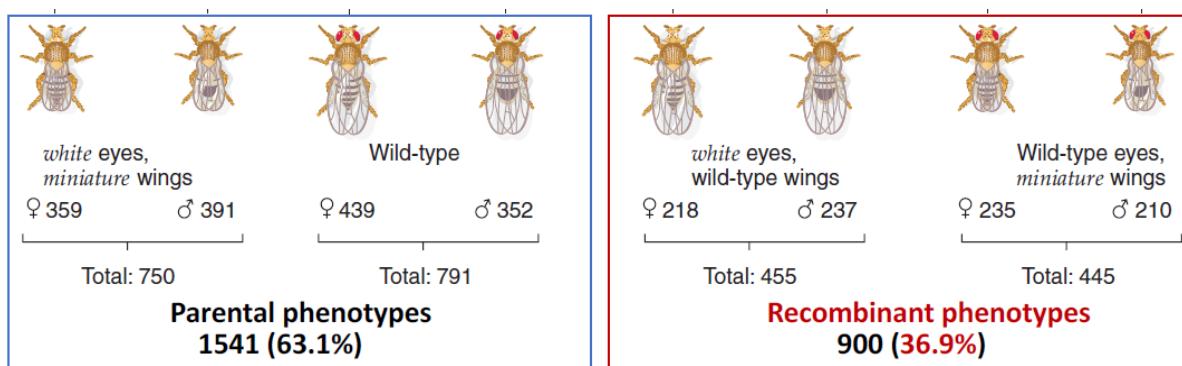
Using our knowledge about dihybrid crosses, considering a complete linkage, we expect:

F2:	$\frac{w\ m}{w\ m}$	$\frac{w+m+}{w\ m}$	$\frac{w\ m}{Y}$	$\frac{w+m+}{Y}$
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
	♀ white, miniature	♀ wild type	♂ white, miniature	♂ wild type

These results are obviously different from a standard mendelian cross because the possible gametes are reduced by the impossibility to create all possible combinations of alleles (we can have just wm or $w+m+$ couples); if the two genes were independent we would have had 8 possible genotypes with equal probability in F2.

In reality Morgan reported intermediate results between a linked and an independent inheritance.

F2:



Approximately 60% of individuals showed the parental phenotypes while nearly 40% the recombinant ones, so the main inheritance pattern is the linked one.

Actually we are not experimenting different inheritance patterns, the inheritance is always linked but in 40% of cases the combinations of alleles swap.

Now let's check if these results are consistent for each gene pair:

We'll now consider the linked behaviour between white eyes (*w*) and yellow body (*y*).

P: $\frac{w\ y}{w\ y}$ X $\frac{w+y+}{Y}$
♀ white eyes, yellow body ♂ wild type

F1: $\frac{w\ m}{w+y+}$ X $\frac{w\ m}{Y}$
♀ wild type ♂ white eyes, yellow body

white eyes and yellow body	$\frac{w\ y}{w\ y}$	$\frac{w\ y}{Y}$	{	Parental phenotypes 99%
wt eyes and wt body	$\frac{w+y+}{w+y+}$	$\frac{w+y+}{Y}$		
white eyes and wt body	$\frac{w\ y}{w\ y}$	$\frac{w\ y}{Y}$	{	Recombinant phenotypes 1%
wt eyes and yellow body	$\frac{w+y}{w+y}$	$\frac{w+y}{Y}$		

In this case the recombinant phenotypes are just 1 %, this means that the two genes were much more linked to one another than the previous couple.

The next step for Morgan was to see if the outcome of his experiments was influenced by the initial genetic combinations of parents.

He discovered that, no matter which phenotype the parental generation had, the F2 presented the parental phenotypes always in the same ratio.

The conclusions that Morgan took from this set of experiments were:

1. The parental phenotypes are always the most frequent and the recombinant ones occur less frequently.
2. Each of the two parental classes appear approximately in an equal number of individuals, as well as each of the two recombinant classes have approximately equal numbers.
3. The percentage of recombinants varies according to the gene pairs analysed
4. The frequency of recombination is independent of the arrangement of alleles.

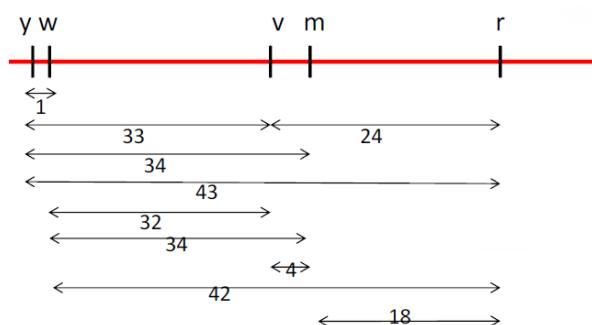
Morgan's hypothesis

- Genes are linearly arranged on chromosomes.
- Recombinants are due crossing-over between homologous chromosomes during meiosis; it consists in a reciprocal exchange of genetic material, the regions that get exchanged are delimited by chiasmata.
- The closer two genes are, the less likely is that a crossing-over will occur between them: the frequency of recombination is related to the distance between the genes on a chromosome.

First linkage map

By doing reciprocal dihybrid crosses between 5 different genes on the X chromosome Alfred Henry Sturtevant was able to put on an ordered line all the genes, with a distance proportional to the recombination frequency.

He considered as distance units of the map the centiMorgan (cM), 1 centimorgan corresponded to 1 percentage point of recombination frequency.

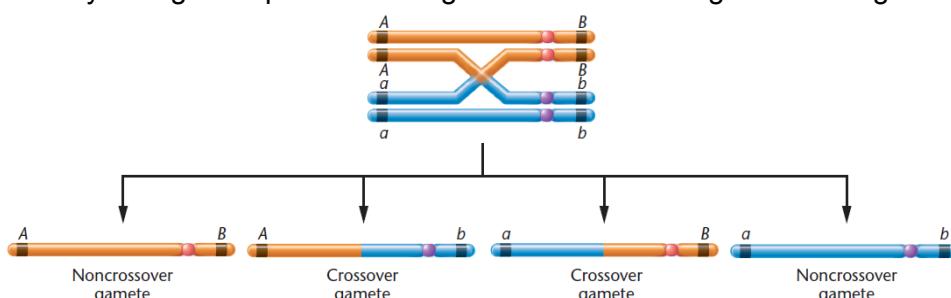


Unfortunately this is a brutal simplification of the true nature of chromosomes

Single crossing over

Until now we supposed that all the gametes that are produced after a single crossing over present the recombinant genotypes; so that the recombination frequency corresponds to the frequency of occurrence of crossing overs.

In reality crossing overs occur in prophase of meiosis I, when the chromosomes are organised in tetrads, and only a single couple of homologous at a time undergoes crossing over.



So after a single crossing over the result is that just 2 sister chromatids of 2 different homologs are recombinant.

We can clearly infer that if a single crossing over happens 100% of the time between two genes (they have to be very far from each other) the recombination frequency (recombinant gametes) that we'll see is 50%.

↓

We cannot see more than 50% recombinant individuals in a dihybrid cross.

If recombination frequency (θ):

- $\theta=0 \rightarrow$ complete linkage
- $\theta<50\% \rightarrow$ partial linkage
- $\theta=50\% \rightarrow$ genes are segregating independently, they could be very far from each other on the same chromosome or on different chromosomes, we cannot tell the difference.

Autosomal chromosomes

In order to reach a genetic map of the whole genome of the individual we need to map all the other genes of drosophila.

Let's consider mutant recessive characters black body (b) and vestigial wings (vg).

P: $\frac{b+vg}{b+vg} \times \frac{b vg+}{b vg+}$
wt body, vestigial wings black body, wt wings

TEST CROSS

F1: $\frac{b+vg}{b vg+} \times \frac{b vg}{b vg}$
wt body, wt wings black body, vestigial wings

We use a test cross to study the linkage because it is the most handy way, the result will depend only on the hybrid individual.

F2:

283	$b^+ vg^+$	ricombinant	8.7%
wt body, wt wings	$\underline{\underline{b \quad vg}}$		
1294	$b^+ vg$	parental	40%
wt body, vestigial wings	$\underline{\underline{b \quad vg}}$		
1418	$b \quad vg^+$	parental	43.8%
black body, wt wings	$\underline{\underline{b \quad vg}}$		
241	$b \quad vg$	ricombinant	7.4%
black body, vestigial wings	$\underline{\underline{b \quad vg}}$		

Total Progeny : 283 + 1294 + 1418 + 241 = 3236

Recombination frequency: $283 + 241 / 3236 \times 100 = 16.2\%$

To state whether or not the two genes can be considered independent we can use the chi-square test:

Phenotypes	Observed (o)	Expected (e)	d (o - e)	d^2	d^2/e
wt body, wt wings.	283	809	- 526	276676	342,00
wt body, vestigial wings.	1294	809	485	235225	290,76
black body, wt wings	1418	809	609	370881	458,44
black body, vestigial wings	241	809	- 568	322624	398,79
Totale	3236	3236			1490,00

In a normal independent segregation we should have obtained $\frac{1}{4}$ of every phenotypic combination.

In this case we have a $\chi^2 = 1490$ with 3 degrees of freedom, it corresponds to a p value smaller than 0.001: we have to reject the hypothesis that the genes are independent.

Through this method we can easily study all the autosomal genes, applying all the rules that we have seen before.

Physical proof of crossing over

Creighton and McClintock carried out a specific experiment on maize chromosomes to see if they actually exchanged DNA portions between each other.

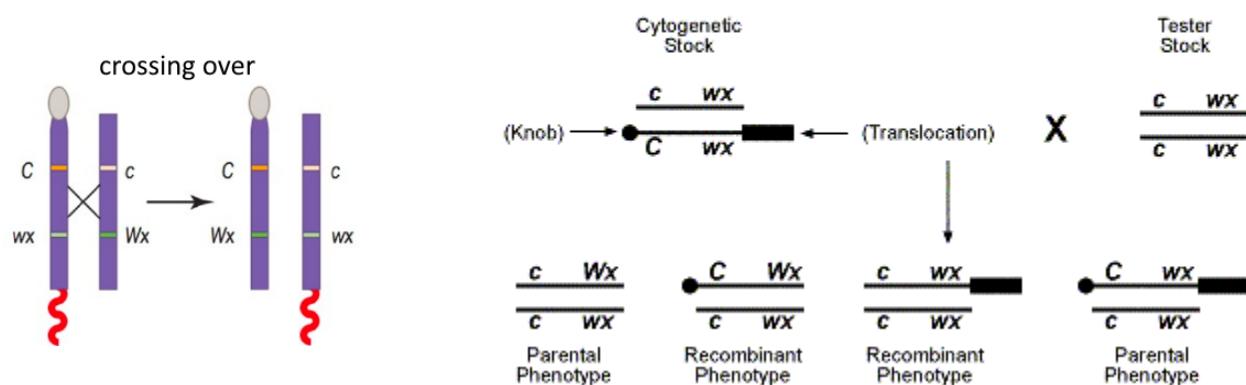
They created an organism that had a couple of chromosomes in which one chromosome had 2 cytological markers at the extremes (variations in chromosome morphology).

Near each marker was present a gene that has a phenotypic expression, during the experiment they kept track of the endosperm coloration (colored: C; colourless: c) and carbohydrate composition (starchy: Wx; waxy: wx).

The core of the experiment consisted in a test cross between the individual with the marked chromosome and a neutral recessive individual.

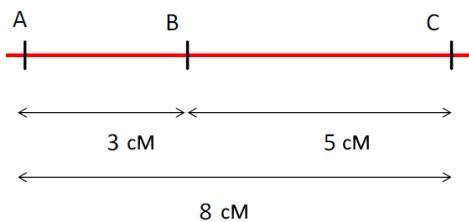
If crossing over involves the physical exchange of chromosomal material, then when crossing over occurs between the marked homolog and the normal one we should obtain two chromosomes with just one marked ending.

These recombinant chromosomes will then form the recombinant phenotypes.

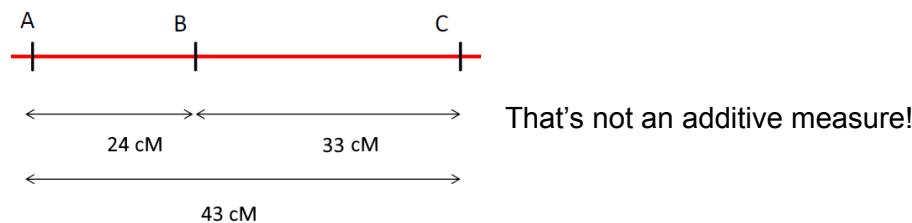


Additiveness of recombination frequencies

Recombination frequencies are additive only within short distances (~10cM).



Otherwise the measured distance between two genes is always less than the sum of the distances between all the intermediate genes.



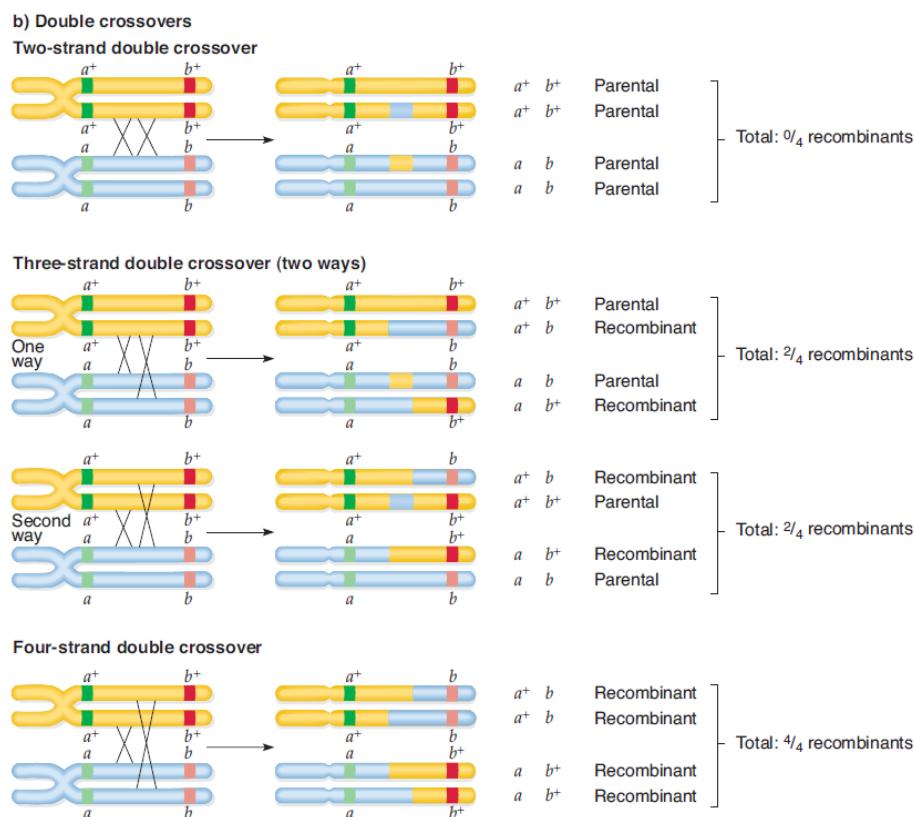
As the distance between genes increases, the crossing over frequencies are not accurate, they underestimate the real distance.

Double crossing over

On long distances between genes the possibility of a double crossing over to happen influences heavily the measured distance.

Only in the absence of multiple crossing over between two loci would there be a linear relationship between map distance and recombination frequency.

Now let's consider all the possible double crossing over:



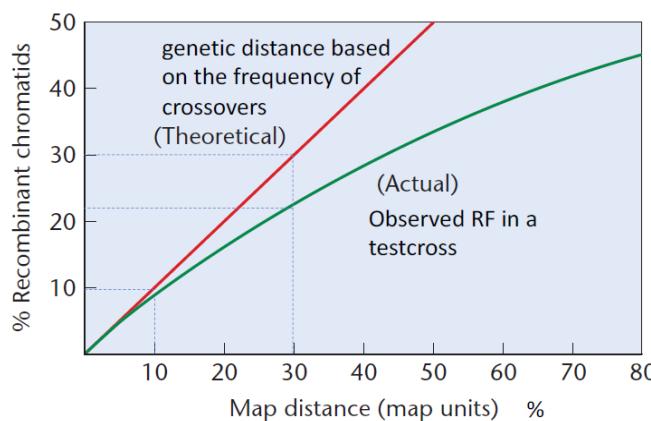
As you can see the average result of a double crossing over is 50% of recombinant gametes; that's the same result as single crossing overs.

The problem in the measurement of distance is that if we consider the parental outcomes of a double crossing over as non-crossing overs we are measuring a much lower distance.

As the distance between two genes increases, the double crossing overs get more probable, all we have to do to take a more precise measure is to **sum the parental double crossing overs to the single crossing overs**.

$$\frac{\text{recombinants} + \text{fake parentals}}{\text{total number of gametes}} * 100 = \text{actual gene distance}$$

The only problem now is that through a **dihybrid cross we cannot distinguish the double crosses phenotypically** because we cannot know if between the two genes there has been an exchange; that's why we can use mathematical equations to do so:



Using this complicated function we can keep into account the presence of false parental gametes, even if we are not physically able to see them.
(page 56)

Three point mapping

By studying 3 genes at a time we can reach a decent grade of precision in the determination of their **order and reciprocal distance**.

All this precise information can only be obtained through a three hybrid cross, keeping in account of the double crossing overs.

We will consider the mutant recessive genes vestigial wings (*vg*), black body (*b*) and (*pr*).

$$P: \quad \frac{vg\ b\ pr}{vg\ b\ pr} \quad \times \quad \frac{vg^+\ b^+\ pr^+}{vg^+\ b^+\ pr^+}$$

Test cross

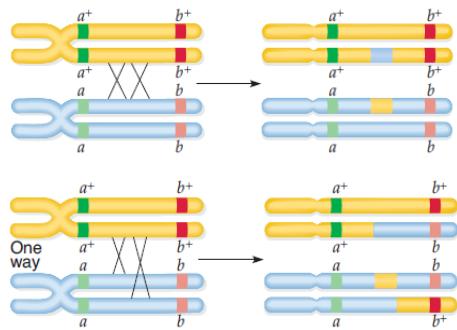
$$F1: \quad \frac{vg\ b\ pr}{vg^+\ b^+\ pr^+} \quad \times \quad \frac{vg\ b\ pr}{vg\ b\ pr}$$

Class	Number	Phenotype	
1	1779	<i>vg</i> <i>b</i> <i>pr</i>	parental phenotypes
2	1654	<i>vg</i> ⁺ <i>b</i> ⁺ <i>pr</i> ⁺	
3	252	<i>vg</i> ⁺ <i>b</i> <i>pr</i>	
4	241	<i>vg</i> <i>b</i> ⁺ <i>pr</i> ⁺	
5	131	<i>vg</i> ⁺ <i>b</i> <i>pr</i> ⁺	
6	118	<i>vg</i> <i>b</i> ⁺ <i>pr</i>	
7	13	<i>vg</i> <i>b</i> <i>pr</i> ⁺	double crossing-over
8	9	<i>vg</i> ⁺ <i>b</i> ⁺ <i>pr</i>	

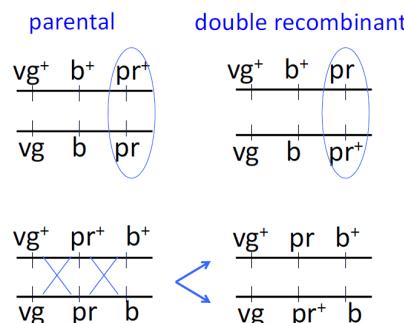
In each one of these crosses we can easily find the parental phenotypes, those are always the most frequent.

Equally easy is to find the double crosses, they are the least frequent and the ones in which it is changed just the central gene (we have to consider always the change in respect to the parental combinations, we do not have always total recessive/dominant parents, in those cases we always have to remember to compare the double crosses with the parental to really get which is the gene that has changed).

Without the monitoring of the central gene we would not be able to distinguish these double crosses with the normal sequences.



In this case we can state that pr is the central gene because it is the one that changes in the double crossing over.



The remaining combinations are due to single crossing over (a few even from three or four strand double crossing overs).

To state where the crossing over has taken place we need to see which terminal gene has been changed, the crossing over happens between a terminal gene and the central one.

correct order	
<i>vg</i>	<i>pr</i> <i>b</i>
<i>vg⁺</i>	<i>pr⁺</i> <i>b⁺</i>
<hr/>	
<i>vg</i> <i>pr</i> <i>b</i>	
<i>vg⁺</i> <i>pr⁺</i> <i>b⁺</i>	
<hr/>	
<i>vg</i> <i>pr</i> <i>b</i>	
<i>vg⁺</i> <i>pr⁺</i> <i>b⁺</i>	
<hr/>	
<i>vg</i> <i>pr</i> <i>b</i>	
<i>vg⁺</i> <i>pr⁺</i> <i>b⁺</i>	
<hr/>	
<i>vg</i> <i>pr</i> <i>b</i>	
<i>vg⁺</i> <i>pr⁺</i> <i>b⁺</i>	

After the analysis we result with this qualitative model. We know the order of the genes and the mechanism of all the possible crossing over.

Now let's talk about distances:

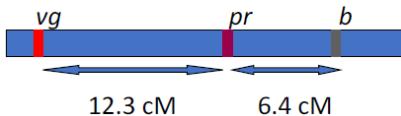
Class		N	Origin
1	<i>vg pr b</i>	1779	
2	<i>vg+ pr+ b+</i>	1654	no crossing-over
3	<i>vg+ pr b</i>	252	single crossing-over
4	<i>vg pr+ b+</i>	241	<i>vg - pr</i>
5	<i>vg+ pr+ b</i>	131	single crossing-over
6	<i>vg pr b+</i>	118	<i>pr - b</i>
7	<i>vg+ pr b+</i>	9	double crossing-over
8	<i>vg pr+ b</i>	13	
Total = 4197			

The distance between first (a) and second (b) gene can be obtained through the following formula:

$$\frac{\text{single crossing over between } a, b + \text{ all double crossing over}}{\text{total}} * 100 = \frac{(252+241)+(9+13)}{4197} * 100 = 12.3\%$$

The distance between second (b) and third (c) gene is:

$$\frac{\text{single crossing over between } b, c + \text{ all double crossing over}}{\text{total}} * 100 = \frac{(131+118)+(9+13)}{4197} * 100 = 6.4\%$$



The distance between the first and the third crossing over is given by the sum of the two intermediate distances.

$$12.3 + 6.4 = 18.7 \text{ cM}$$

Interference

Experimentally we see that double crossing overs occur less than their expected frequency, we can measure this difference and we will call it interference.

After having determined the distances between the genes we can infer the **expected double crossing over frequency**:

$$\frac{\text{distance } a, b}{100} * \frac{\text{distance } b, c}{100} = \text{theoretical double crossing over frequency}$$

This formula consists in the product between the frequencies of recombinations between the marginal genes and the central one; the reasoning behind this is that we want to find the joint occurrence of both the recombinations (i.e. double crossing overs), so using a normal joint probability we have: $P(\text{co-ab} \cap \text{co-bc}) = P(\text{co-ab} | \text{co-bc}) * P(\text{co-bc})$.

To obtain the theoretical frequency of recombination we suppose that the single crossing overs are independent events, so we reach: $P(\text{co-ab} \cap \text{co-bc}) = P(\text{co-ab}) * P(\text{co-bc})$.

Now that we have obtained the theoretical frequency we can multiply it for the total number of gametes to see how many double crossing over gametes we should have obtained if the single crossing overs were independent events.

$$\text{theoretical d-co frequency} * \text{total individuals} = \text{expected number of d-co gametes}$$

In our previous example we obtain a crossing over frequency of 0.79%, that corresponds approximately to 33 cases of double crossing over; The observed cases are 22, nearly the 0.52%.

The observed double crossing over frequency is always less than the expected number of double crossing over.

That's because of interference (I), it is the inhibition of further crossovers events by a crossover event in a nearby region of the chromosome.

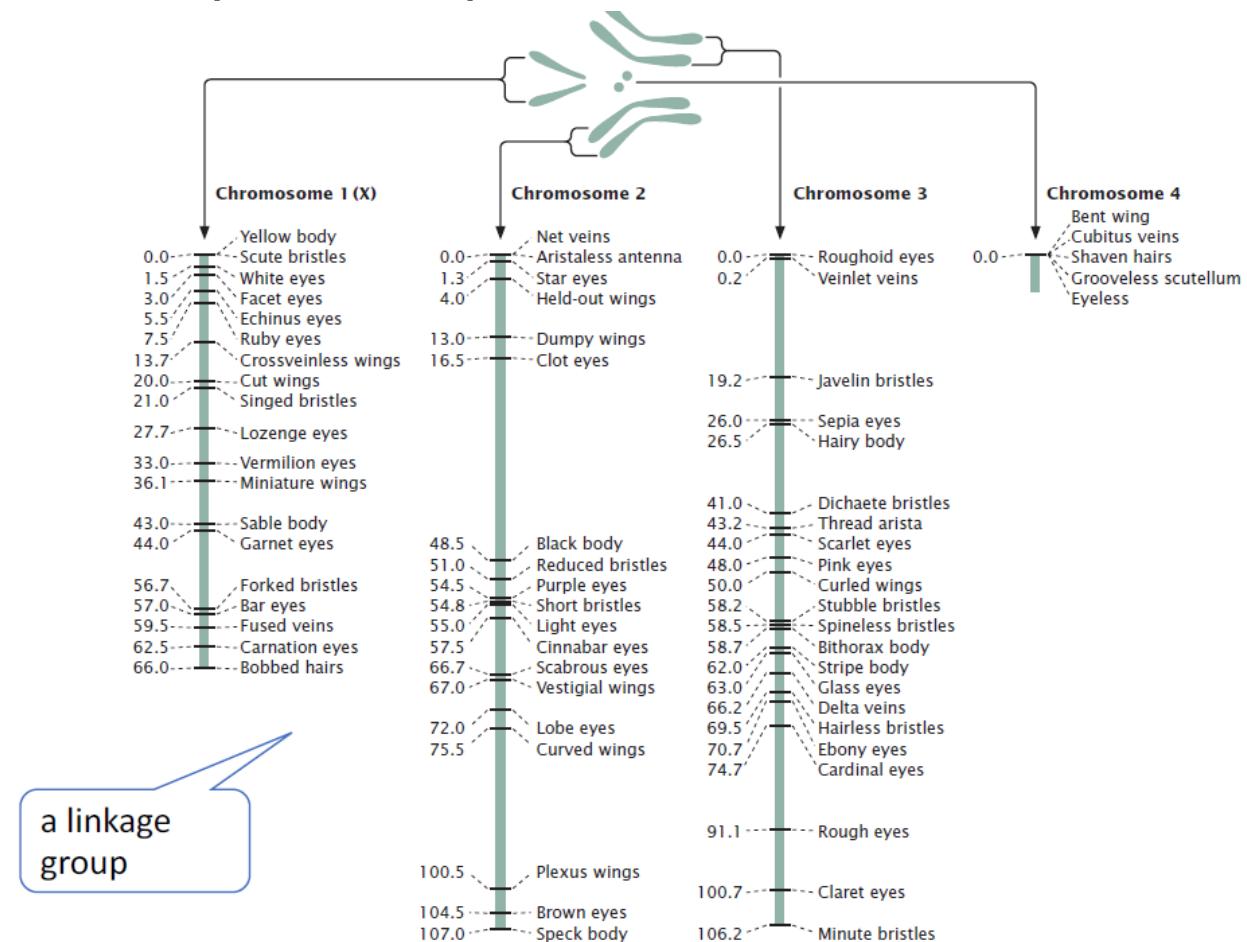
$$\text{Coefficient of coincidence } (C) = \frac{\text{observed number of double crossovers}}{\text{expected number of double crossovers}} = \frac{22}{33} = 0.66$$

$$\text{interference } (I) = 1 - C = 1 - 0.66 = 0.34$$

34% less double crossovers are observed due to interference; higher is the interference less probable is for a double crossing over to verify.

If interference is complete, no double crossovers occur.

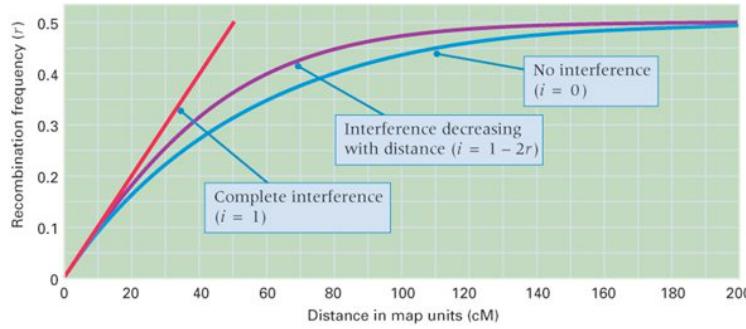
Genetic map of the Drosophila chromosomes



Each gene of the genome that has a phenotypic expression can be localised through these techniques, the set of genes belonging to the same chromosome is called linkage group.

Genetic mapping functions

A mapping function is a mathematical relation between the genetic distance across an interval in map units and the observed frequency of recombination across the interval.



- 1) Assumption of complete interference (no double crossover). Linear relation. For map distances larger than 50 cM, the recombination frequency remains constant at 50 percent.
- 2) Haldane mapping function: No interference
- 3) Kosambi mapping function: The interference is assumed to decrease as a linear function of distance according to $i = 1 - 2r$.

Thanks to these functions we can have a good approximation of the true genetic distance even by just having the frequency of single crossing overs.

Ex. Normally if we do a dihybrid cross and we obtain a frequency of 30% of recombinants we have to assume a distance of 30 cM; through these formulas we know that if we experience 30% of single crossing overs the genes are at a distance of 40cM.

Physical distance

Overall there is correlation between genetic and physical distance, but it is not perfect and not uniform.

The frequency of recombination differs between the sexes (for example, in Drosophila there is no recombination in males and in Human, the recombination rate is higher in females) and the recombination rate is not uniform along the chromosomes, crossing-over is less frequent in regions of heterochromatin, and more frequent in subtelomeric regions

In humans on average 1 cM = 1 Mb of DNA

Recombination hotspots

In human DNA, most crossovers take place in recombination hotspots, small regions of DNA where the frequency of recombination is much higher than average.