Synopsis

Chromosome mapping in Eukaryote cells

Contents

Required knowledge:

Biology 4, chapter 14, mendelian genetics

Knowledge about meiosis 1 & 2, mainly that chiasma form between homologous chromosomes, see image 1 & 2:

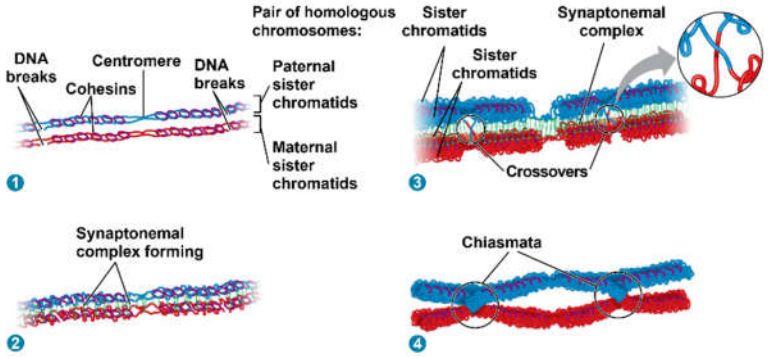

Image 1 Graphical view for Chiasmata

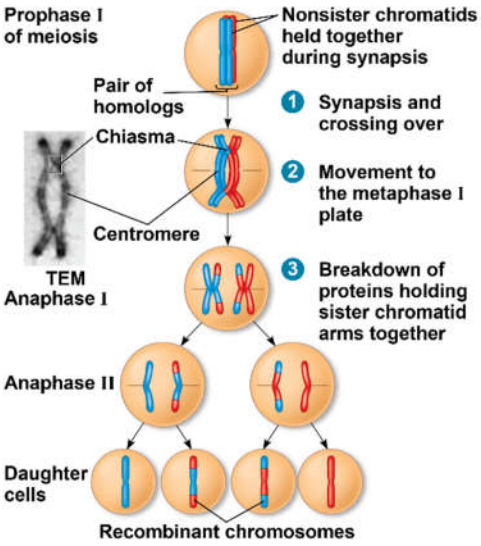

Image 2 Meiosis overview

Knowledge about the Mendel's Laws for genes is also needed to map chromosomes, they are recapped below:

- Law of segregation of genes (the "First Law")
  - During gamete formation, the alleles for each gene segregate from each other so that each gamete carries only one allele for each gene.
- Law of Independent Assortment (the "Second Law")
  - Genes for different traits can segregate independently during the formation of gametes.
- Law of Dominance (the "Third Law")
  - Some alleles are dominant while others are recessive, an organism with at least one dominant allele will display the effect of the dominant allele.

Chapter 5, Chromosome Mapping in Eukaryotes

5.1 Genes linked on the same chromosome segregate together.

Depending on how the genes are linked on chromosomes, there are 3 different concepts on how genes can be transferred and how gametes can be formed, also known as Meiotic consequences:

Independent assortment
No linkage exhibited, which means that the genes are on two different homologue pairs of chromosomes.

Linkage without crossing over
Complete linkage, which means that two genes exist on a single pair of homologs, where no exchange occurs.

Linkage with crossing over
Generates recombinant (crossover) gametes, which means that two genes exist on a single pair of homologs, but exchange between two nonsister chromatids occurs.

See image 3 below for a graphical overview:

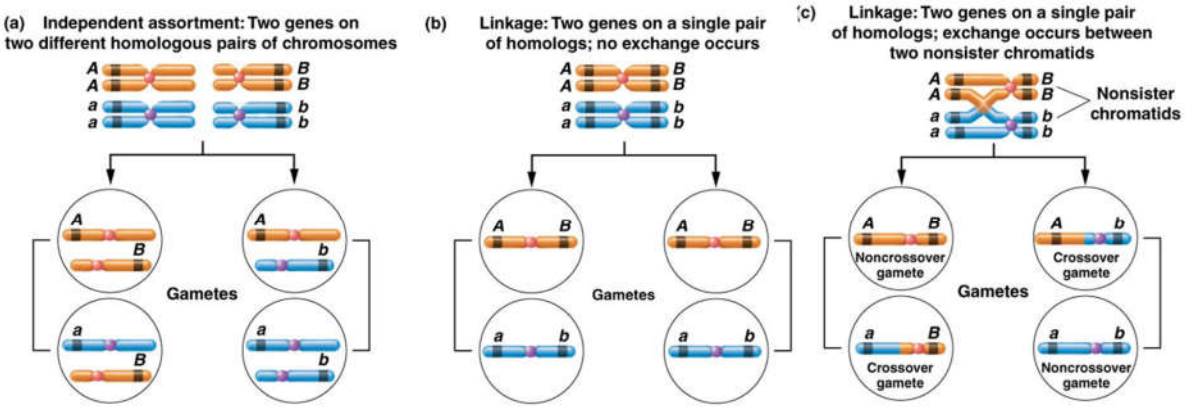

Image 3 Meiotic consequences

If complete linkage exists between two genes because of their close proximity, and organisms heterozygous at both loci are mated, a unique F_2 phenotypic ratio results, which we designate as the **linkage ratio**. See P_1 in image 4. Here we can see that, we can also write P_1 as the following, with the following designations:

Recessive genes; Heavy wing vein (hv), brown eye (bw)
Dominant genes: thin wing veins (hv^+), red eyes (bw^+)

$$P_1 = \frac{hv^+\ bw}{hv^+\ bw} * \frac{hv\ bw^+}{hv\ bw^+}$$

If these parents were to receive kids, the kids would all have the following F_1 gene notation:

$$\mathrm{F}_1 = \frac{hv^+\ bw}{hv\ bw^+}$$

If the F_1 generation would be interbred, you'd get the following ratio:
$1/4^{th}$ have thin wing veins, with brown eyes, $2/4^{th}$ would have thin wing veins with red eyes and $1/4^{th}$ would have heavy wing veins, with red eyes. A **1:2:1** ratio, this all is showcased in image 4 as a graphical overview:

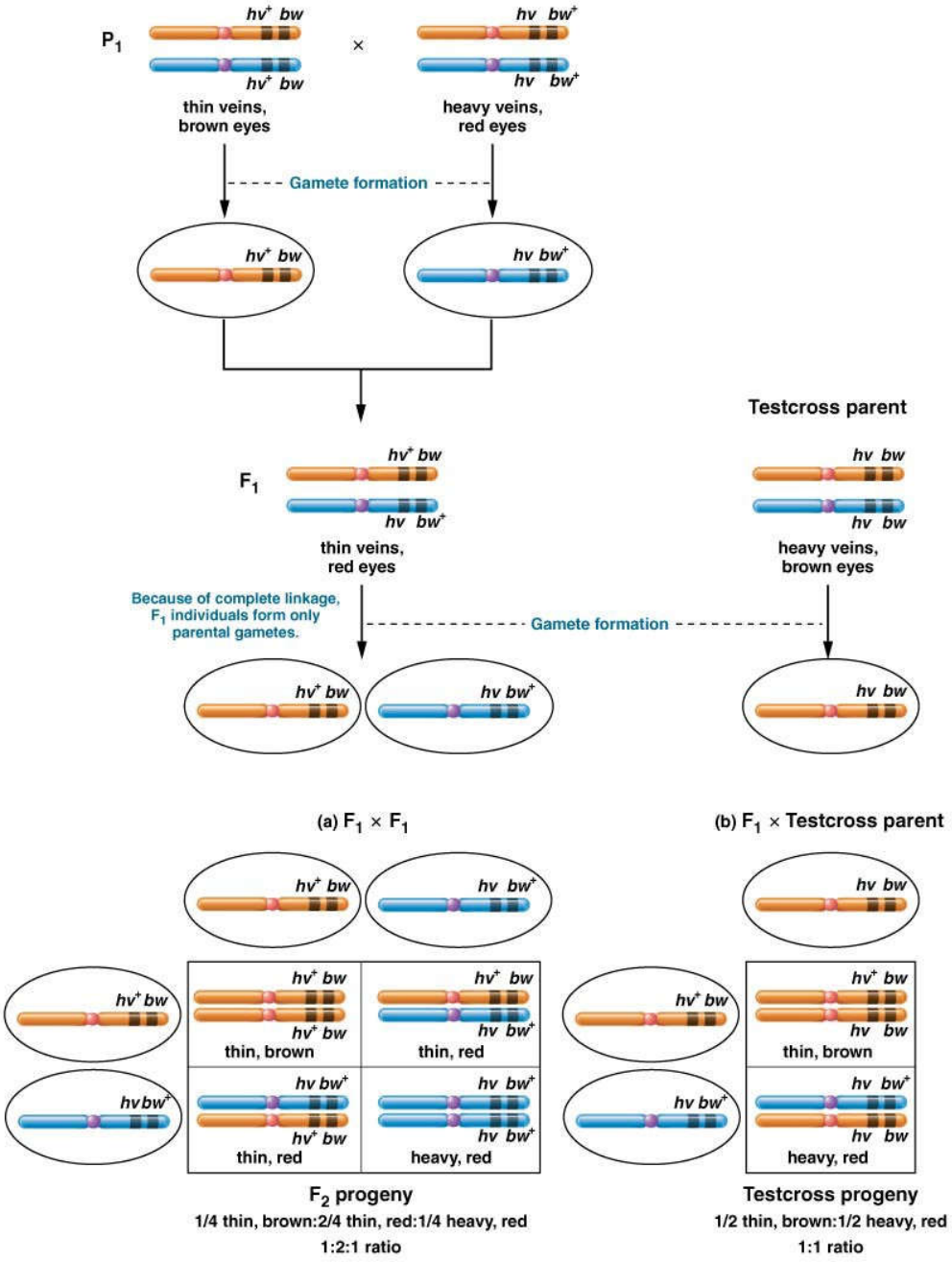

Image 4 Linkage ratio graphical overview.

5.2 Crossing over serves as the basis for determining the distance between genes in chromosome mapping.

It is highly improbable that two randomly selected genes linked on the same chromosome will be so close to one another along the chromosome that they demonstrate complete linkage. Instead, they will almost always produce a percentage of offspring resulting from recombinant gametes. The percentage varying depending on the distance between the two genes.
Crossing over happening through a chiasma, the closer the genes are located, the less likely they have a chance to form a chiasma.

Sturtevant was one of the first people to realise that the variations in the strength of the linkage, could determine the sequence in the linear dimension of a chromosome.

To test this, he compiled a lot of data from crossing Drosophila regarding the genes coding for yellow, white, and miniature mutants. He discovered that there were 3 different frequencies of recombination between each pair of these three genes as follow:
1: yellow, white = 0.5%
2: white, miniature = 34.5%
3: yellow, miniature = 35.4%

A subset (so not everything, just for frequencies 1 & 2) is shown below as an example in image 5:

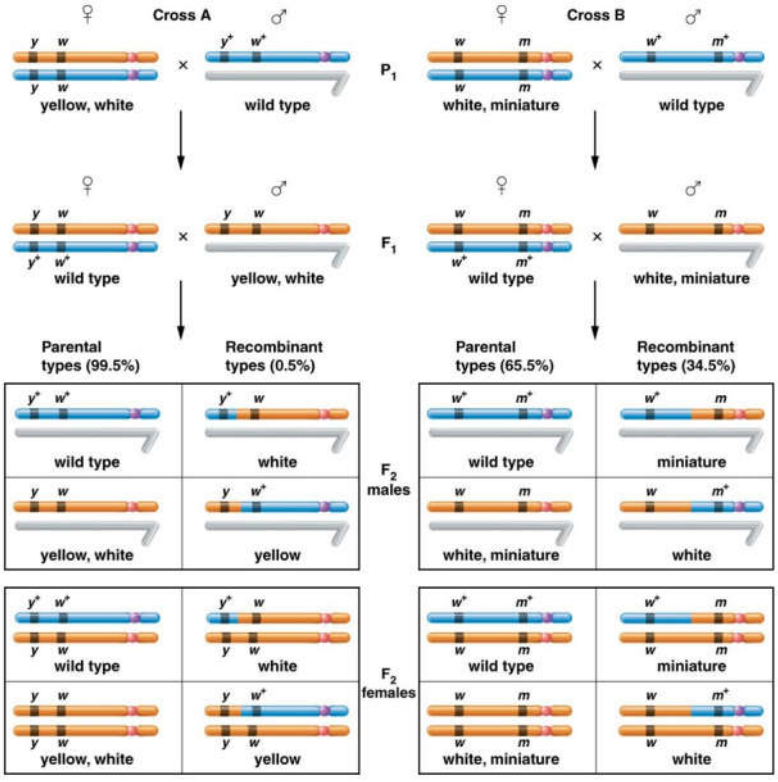

Image 5 Overview Sturtevant frequencies 1 & 2.

Because the sum of frequencies 1 (0.5%) & 2 (34.5%), are almost similar to 3 (35.4%), Sturtevant suggested that the recombinant frequencies between linked genes are additive. He thus concluded that the order of the genes on the X chromosome is as follows: yellow – white – miniature.

He came to the said conclusion that the yellow and white genes are apparently close to each other due to the low recombination frequency between them. With this information, he mapped the chromosome as follows in image 6, where 1% recombination is 1 map unit (mu), which is also called **centimorgans (cM)**:

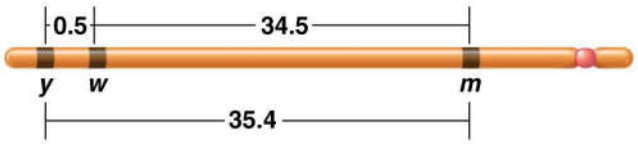

Image 6 Mapped chromosome with the recombinant frequencies as distance

Then we can also conclude that the closer two loci reside along the axis of the chromosome, the less likely that any **single crossover (SCO)** will occur between them, see the following image 7 for an example of an SCO in 2 settings:

In example 7.a, the SCO happens between two nonsister chromatids, but the linkage between A & B alleles are unchanged due to their distance being small.

In example 7.b, the SCO happens between two nonsister chromatids and the alleles of B & b are exchanged, causing recombination in 2 of the 4 gametes.

With this, we ca also conclude, that if SCO happens 100% of the time, we can only observe recombination in 50% of the gametes.

Using that information, we can conclude that if two linked genes are more than 50 map units apart, a crossover can theoretically be expected to occur between them in 100% of the tetrads, as we can see in image 8 below, as if the genes would be on different chromosomes:

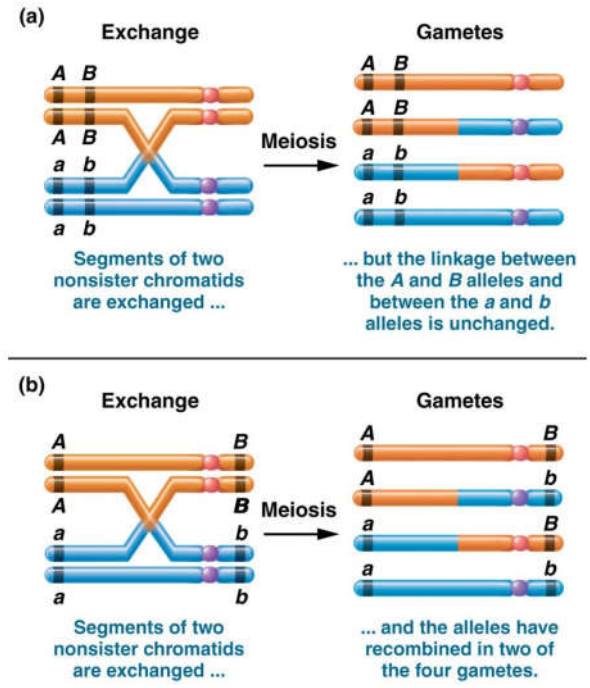

Image 7 SCO example.

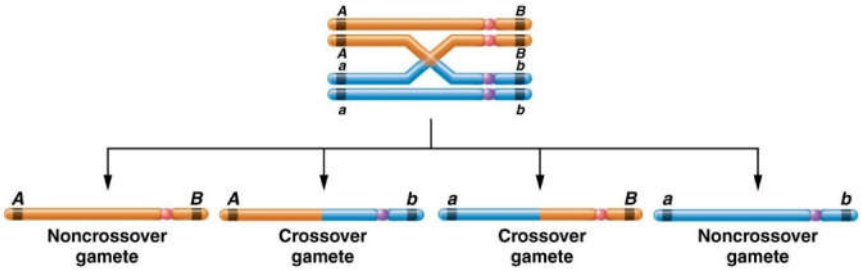

Image 8 50 map unites example.

5.3 Determining the gene sequence during mapping requires the analysis of multiple crossovers.

Crossovers.

Single crossovers are used to determine the distance between two linked genes.

However, it is also possible for more than one exchange to occur between nonsister chromatids because of several crossing over events.

Double exchanges of genetic material result from **double crossovers (DCOs)**, as seen in image 9 below, as we can see that, due to a double crossing over, usually the middle gene gets exchanged, the requirement being that the genes must be heterozygous for two alleles:

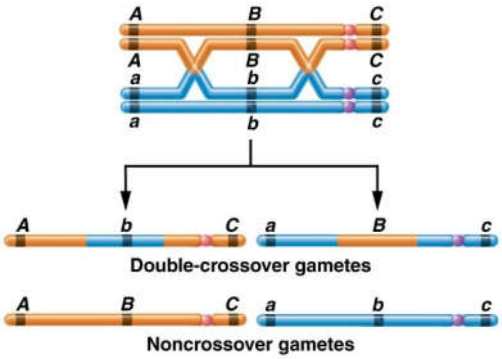

Image 9 DCO example

DCOs are therefore used to determine the distance between three linked genes.

This also means that the expected frequency of a double-crossover gamete is lower than that of either single-crossover gamete class and if three genes are close together along one chromosome.
For example, the SCO for A & B is 20% (0.2), the SCO for B & C is 30%(0.3), to get a double crossover for between A/B and B/C, the calculation would be the following: 0.2*0.3 = 0.06 = 6%.
If the genes were already closely related, the calculation could be the following: 0.02*0.03 = 0.0006 = 0.06%

Three-point mapping

Now for successful three-point (three gene) mapping three criteria must be met:

1. The parent must be heterozygous for all three genes under consideration.
2. Phenotypic class must reflect genotype of gametes of parents.
3. Sufficient number of offspring must be produced for representative sample.

Below in image 10, would be an example of three-point mapping involving the yellow (y), white (w) and echinus (ec) genes in fruit flies. The hypothetical sequence is y-w-ec and if incorrect, the three point mapping would otherwise reveal the correct sequence.

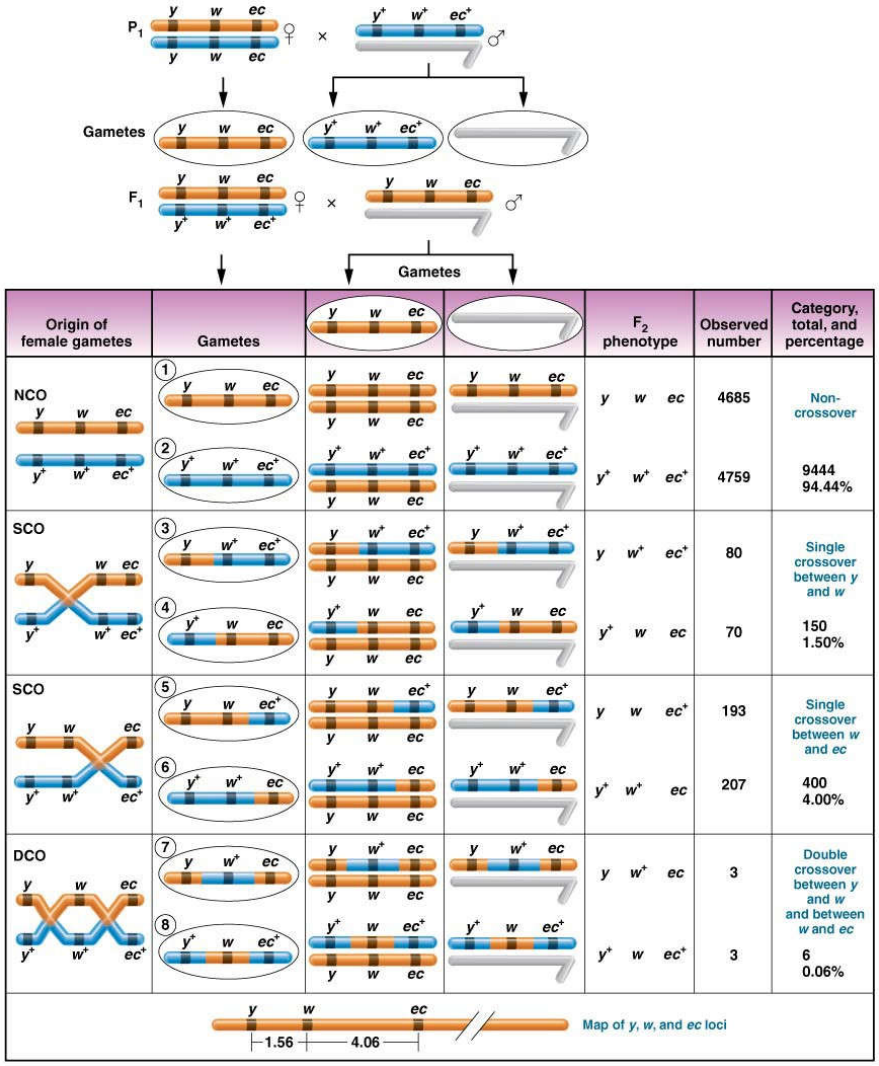

Image 10 Three point mapping example

Now, to discover the **noncrossover F_2 phenotypes** in the table, we look for the phenotypes with the highest observed number.
To discover the **double-crossover (DCO) phenotypes**, we look at the phenotypes which are the least common.

Reciprocal classes of phenotypes are F_2 phenotypes that complement each other and are derived from the heterozygote, they're usually an NCO F_2 phenotype and have a wild type and mutant type for all three genes. The gamete would for example contain y^+-w^+-ec^+ and y-w-ec, as can be seen in the blue-orange gamete at the NCO row in image 10.

Methods for determining the gene sequence

There are 2 methods to determine the Gene sequence, as an example, using the gene information from the image 10 example, the assumed sequence of the three genes along the chromosome was y-w-ec. Now if the gene order was unknown, that's where the methods come into play:

Method 1: There are only 3 possible arrangements, each one containing a different one of the three genes between the two:
(1) w-y-ec (y is in the middle)
(2) y-ec-w (ec is in the middle)
(3) y-w-ec (w is in the middle)

Then there are 3 steps for this method:
1. Assuming any of the three orders, first determine the arrangement of alleles along each homolog of the heterozygous parent giving rise to noncrossover and crossover gametes (the F_1 female in the example)

2. Determine whether a double-crossover event occurring within that arrangement will produce the observed double-crossover phenotypes. These are the least occurring phenotypes.

3. If this order does not produce the correct phenotype, try each of the other two orders. One should work!

So using arrangement (1):
1. Assuming that y is between w and ec, the districution of alleles between the homologs of the F_1 heterozygote is w-y-ec / w^+-y^+-ec^+.

2. The double crossover between said arrangement gives the following gametes:
w-y^+-ec / w^+-y-ec^+, which gives the following white, echinus, phenotype & yellow phenotype, these don't correspond with the phenotype findings needed: yellow, echinus phenotype and white phenotype. (The gene sequence for these would be (y-w^+-ec/ y^+-w-ec^+).

3. Try again with the other 2 arrangements (2) & (3).

See the following image 11 for a graphical overview of method 1.

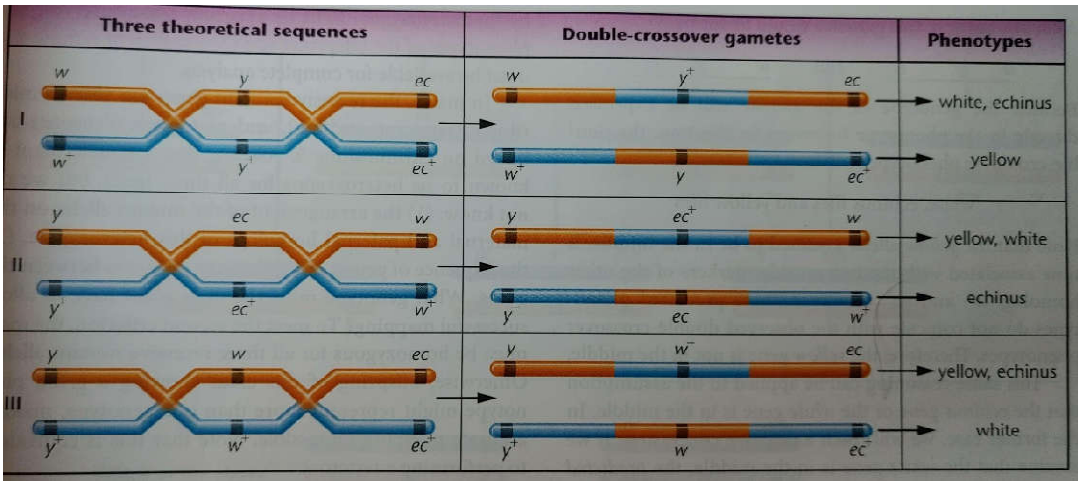

Image 11 Example of method 1.

Method 2 also uses the possible three arrangements, albeit they also consider the double crossover event from the start, you look at the phenotypes that are observed, in this case yellow, echinus and white.
From that you can conclude that the gene encoding for white (w) should be in the middle.

The following is an example in going from a genemap to determining the linkage distances.
The table on the right is the genemap.

Genotype	Observed
v cv$^+$ *ct*$^+$	580
v$^+$ *cv ct*	592
v cv ct$^+$	45
v$^+$ *cv*$^+$ *ct*	40
v cv ct	89
v$^+$ *cv*$^+$ *ct*$^+$	94
v cv$^+$ *ct*	3
v$^+$ *cv ct*$^+$	5
Total	**1448**

Step 1: Determine the parental genotype.

The most abundant genotypes are always the parental genotypes, in this case being: v-cv$^+$-ct$^+$ and v$^+$-cv-ct.

Step 2: Determine the gene order.

To determine the gene order, we need the parental genotypes as well as the double crossover genotypes. As we mentioned above, the least frequent genotypes are the double-crossover genotypes. These genotypes are *v cv*$^+$ *ct* and *v*$^+$ *cv ct*$^+$.

From the first double crossover, *v cv*$^+$ *ct*, the *ct* allele is associated with the *v* and *cv*$^+$ alleles, two alleles it was not associated with in the original cross. Therefore, *ct* is in the middle, and the gene order is *v ct cv.*

Step 3: Determining the linkage distances.
For the calculation, you need to count all the SCO & DCO and divide them by the total number.

v - *ct* distance calculation. This distance is derived as follows: 100*((89+94+3+5)/1448) = 13.2 cM

ct - *cv* distance calculation. This distance is derived as follows: 100*((45+40+3+5)/1448) = 6.4 cM

5.4 As the distance between two genes increases, mapping estimates become more inaccurate.

So far, we have assumed that crossover frequencies are directly proportional to the distance between any two loci along the chromosome. However, it is not always possible to detect all crossover events. For example, if we look at image 12 below.

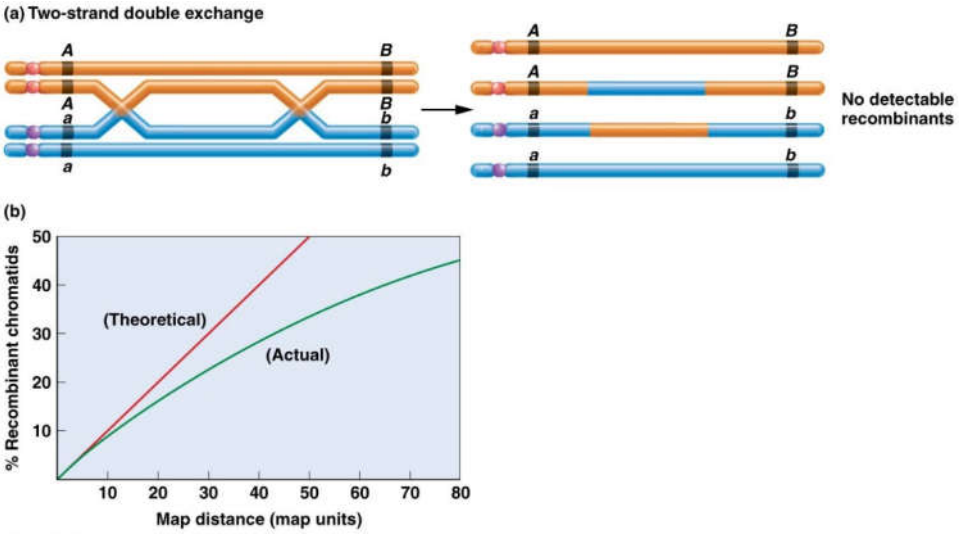

Image 12 Interference in mapping

We can see that due to a double crossing over, the recombinants aren't detectable at all, because the original arrangement of the alleles on each nonsister chromatid is recovered, as if the DCO never happened.

So due to this we can conclude that the further the genes are apart from each other, the more undetected crossovers will happen.

As such, using an example, after having determined the distances between the genes:
The double crossover (DCOs) amount was 7.8% of the total amount.
genes are v-pr-bm, distance between v-pr = 22.3 mu and distance between pr-bm = 43.4 mu.
Knowing the distances we can calculate the expected amount of DCOs:

DCO_{exp} = (0.223) * (0.434) = 0.097 = 9.7%
As we can see there is a clear difference between the expected amount of DCOs (9.7%) and the observed DCOs (7.8%).

Also, therefore we can conclude that there is **interference (I)** possible on detecting crossing overs detectable for the study. Interference can happen in the following ways:
Inhibition of further crossover events and inhibition by another crossing over event nearby.
This reduces the expected number of multiple crossovers.

To quantify the disparities that result from interference, we calculate the **coefficient of coincidence (C)**:
$C = \frac{Observed\ DCO}{Expected\ DCO}$

With the information from the example:
C = 0.078/0.097 = 0.804

Now we can quantify the interference (I) by using the following equation:
I = 1 – C

So we have:
I = 1.0 - 0.804 = 0.196

Now there are 3 different interferences:
Complete interference, which means that there are no double crossovers, needing I to be 1.
Positive interferences, which mean that there are more DCOs expected than occur, needing I to be a negative number.
Negative interferences, which mean that there are less DCOs expected than they occur, needing I to be a positive number.

In our example I is a positive number, and the observed amount is indeed less than the expected amount, we're dealing with a negative interference.

So, we can conclude that if two genes are together, positive influence can occur and that the accuracy of mapping is high. And that if the distance between two genes increases, that there will be more negative interference and the accuracy of mapping to decrease, as we can see in the graph in image 12.

5.5 Drosophila genes have been extensively mapped

Basically a large number of mutants in organisms such as listed below have been found.
- Drosophila
- Maize
- Mice

This allows for construction of extensive chromosome mapping, as we can see an example of that in image 13.

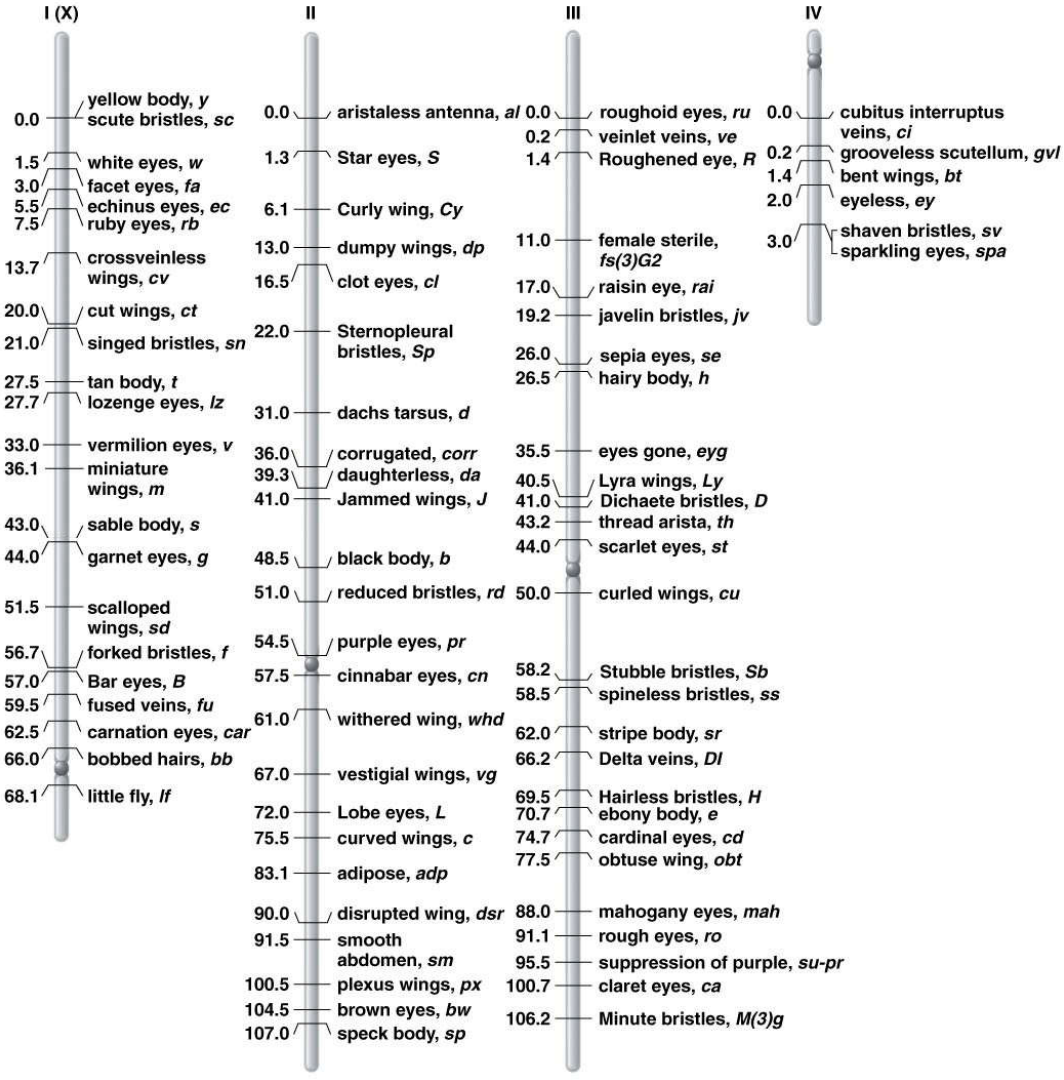

Image 13 Example of the large chromosome mapping.

5.6 Lod score analysis and somatic cell hybridization were historically important in creating human chromosome maps

Due the fact that some genes of interest in human chromosomes are separated on a chromosome to such a degree that recombinant gametes are formed, an approach relying on probability calculations, called the **lod score method**, helps to demonstrate linkage.
Lod score (Log of the odds favouring linkage) method:
- Relies on probability calculations
- Demonstrates linkage between two genes when linkage analysis relies primarily on pedigrees
- Assesses probability that pedigree with two traits reflects genetic linkage between them.

In the 1960s a new technique, **somatic cell hybridization**, proved to be an immense aid in assigning human genes to their respective chromosomes. This technique relies on the fact that two cells in culture can be induced to fuse into a single hybrid cell. When fusion occurs, an initial cell type called a **heterokaryon** is produced, when they're cultivated in vivo, two interesting changes occur. Eventually the nuclei fuse together, creating a **synkaryon**. Then, as the culturing is continued for many generations, chromosomes from one of the two parental species is gradually lost.

Somatic cell hybridization:
- Made assigning of human genes to their respective chromosomes possible
- Involves fusing two cells into a single hybrid cell: **heterokaryon**

Synkaryon:
-Heterokaryons cultured in vivo where the nuclei are fused together
-Chromosomes from 1 of the two parental species gradually fades over generations.

For example, a human/mouse hybrid cell loses human chromosomes over time.
Due to the human chromosomes fading over generations, with multiple hybrid cell lines we are able to assign genes to chromosomes they reside on, by looking at the gene products the cell lines make, this is also called **synteny testing**, see image 14 below.

Syntenty testing
- Looking at the presences or absence of each chromosome, with presence or absence of each gene product.

Hybrid cell lines	Human chromosomes present								Gene products expressed			
	1	2	3	4	5	6	7	8	A	B	C	D
23									−	+	−	+
34									+	−	−	+
41									+	+	−	+

Image 14 Syntenty testing

As we can see in image 14, looking at the gene products and the human chromosomes present, we'll take a look in detail at Gene product A.

1. Product A is not produced by cell line 23, but chromosomes 1-4 are present, so we can rule out the presence of gene A on those four chromosomes.
2. Product A is produced by cell line 34, which contains chromosomes 5 and 6, but not 7 & 8. So Gene product A is produced on either chromosome 5 or 6.
3. Product A is also produced by cell line 41, which contain chromosome 5, but not chromosome 6. Therefore, gene A is on chromosome 5.

5.7 Chromosome mapping is now possible using DNA markers and annotated computer databases.

DNA markers:
- Short segments of DNA with known sequence and location
- Useful landmarks for mapping
- Earliest examples of DNA markers: **Restriction fragment length polymorphisms** and **microsatellites.**

RFLPs: Restriction fragment length polymorphisms:
- Polymorphic sites
- Generated when specific DNA sequences are recognized and cut by a restriction enzyme.

Microsatellites:
- Short repetitive sequences
- Found throughout the genome

SNPs: Single nucleotide polymorphisms:
- Found throughout the genome
- Used by geneticists to identify and locate related genes
- Used to screen for diseases, for example: **Cystic fibrosis**

Cystic fibrosis
- Gene located by using DNA markers
- Life-shortening autosomal recessive exocrine disorder
- Gene causing disorder found on chromosome 7.

5.8 Crossing over involves a physical exchange between chromatids.

Genetic mapping techniques used to study relationship between chiasmata and crossing over.

Back in the 1930s, scientists weren't sure if chiasmata were a visible manifestation of crossing over events.
So they used a chromosome that had a knob & a translocated element in their experiment to perform a crossing over event, see image 15.
In this experiment they made use of **cytological markers** (the knob & the translocated element) and established that crossing over involves a physical exchange of chromosome regions.

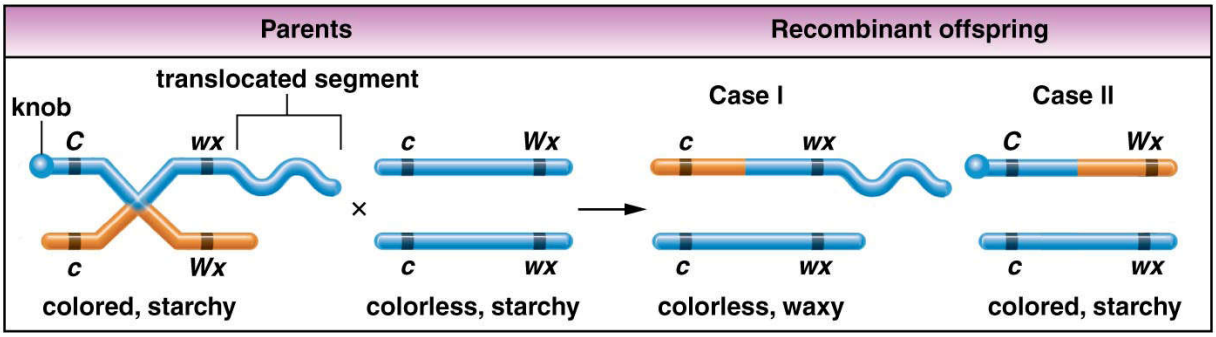

Image 15 Experiment with cytological markers

5.9 Exchanges also occur between sister chromatids during mitosis.

Sister chromatid exchanges (SCEs) occur during mitosis, but do not produce new allelic combinations.

SCEs: Sister chromatid exchanges:
- Reciprocal exchanges similar to crossing over
- Between sister chromatids (crossing over is between NONsisters)

Harlequin chromosomes
- Sister chromatids involved in mitotic exchanges.
-Path-like appearance when stained and viewed under a microscope, see image 16.

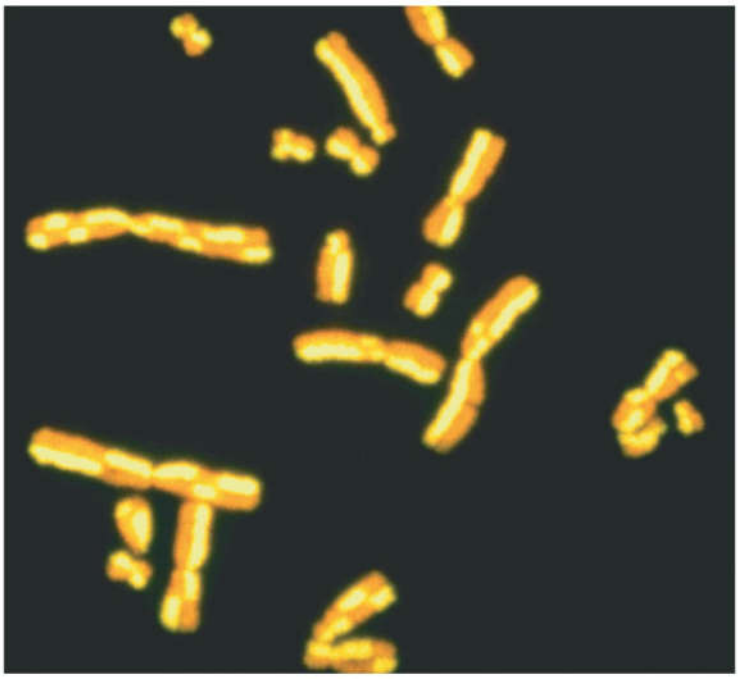

Image 16 Harlequin chromosomes

So far, the significance of SCEs are still uncertain, but agents that increases chromosomal damage **(e.g, viruses, X-rays, ultraviolet light and certain chemical mutagens)** also increase the frequency of SCEs, while also being increased in **Bloom syndrome.**

Bloom syndrome:
- Human disorder
- Mutation in **BLM gene**, located on chromosome 15
- Prenatal and postnatal retardation of growth.

BLM gene
- Encodes enzyme DNA helicase
- DNA helicase's role is DNA replication.

5.10 Did Mendel encounter Linkage?

Mendel did not encounter linkage relationships.

If he had encountered linkage relationships:
- He would've not been able to recognize the basic patterns of inheritance
- He might not have interpreted the basic patterns of inheritance correctly

With later study, it was shown that Mendel had inaccuracy in his hypothesis. He had used seven genes and the pea had only 7 chromosomes, but he most likely was not using 7 genes from 7 chromosomes, but using three genes from chromosome 4, two genes from chromosome 1 and one gene in each of chromosome 5 and 7.