BIOINFORMATICS

Kristel Van Steen, PhD^2

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

CHAPTER 2: INTRODUCTION TO GENETICS

1 Basics of molecular genetics

1.a Where is the genetic information located?

The structure of cells, chromosomes, DNA and RNA

1.b What does the genetic information mean?

Reading the information, reading frames

1.c How is the genetic information translated?

The central dogma of molecular biology

2 Overview of human genetics

2.a How is the genetic information transmitted from generation to generation?

Review of mitosis and meiosis, recombination and cross-over

2.b How do individuals differ with regard to their genetic variation?

Alleles and mutations

2.c How to detect individual differences?

Sequencing and amplification of DNA segments

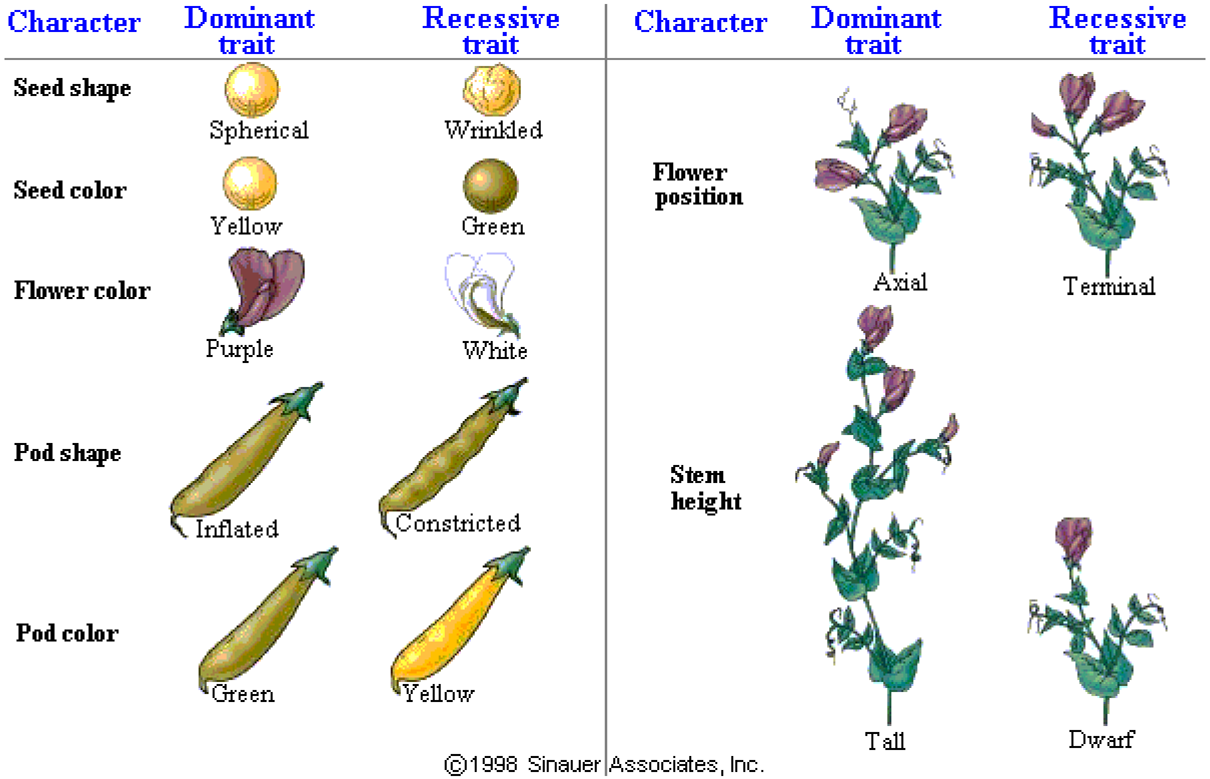
1 Basics of molecular genetics

1.a Where is the genetic information located?

Mendel

- Many traits in plants and animals are heritable; genetics is the study of these heritable factors
- Initially it was believed that the mechanism of inheritance was a masking of parental characteristics
- Mendel developed the theory that the mechanism involves random transmission of discrete “units” of information, called genes. He asserted that,
    - when a parent passes one of two copies of a gene to offspring, these are transmitted with probability 1/2, and different genes are inherited independently of one another (is this true?)

Mendel’s pea traits

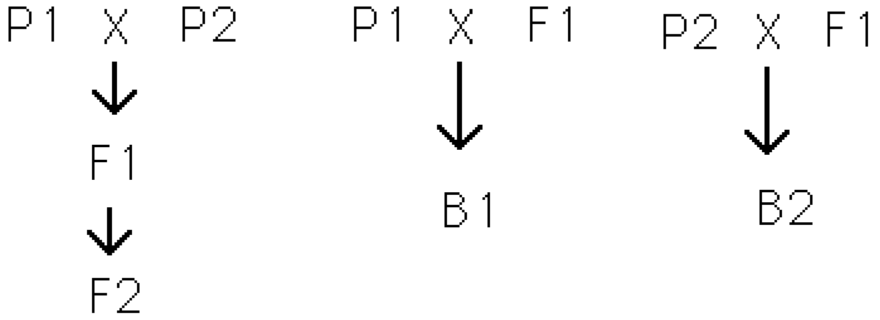


Character	Dominant trait	Recessive trait
Seed shape	Spherical	Wrinkled
Seed color	Yellow	Green
Flower color	Purple	White
Pod shape	Inflated	Constricted
Pod color	Green	Yellow
Flower position	Axial	Terminal
Stem height	Tall	Dwarf

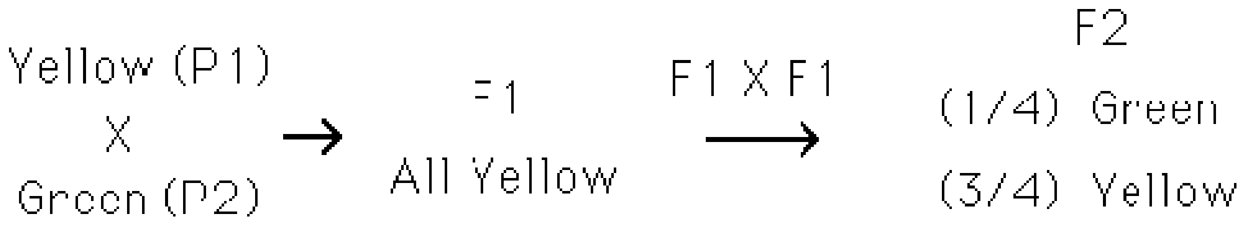
©1998 Sinauer Associates, Inc.

Some notations for line crosses

- Parental Generations (P_1 and P_2)
- First Filial Generation $F_1 = P_1$ X P_2
- Second Filial Generation $F_2 = F_1$ X F_1
- Backcross one, $B_1 = F_1$ X P_1
- Backcross two, $B_2 = F_1$ X P_2



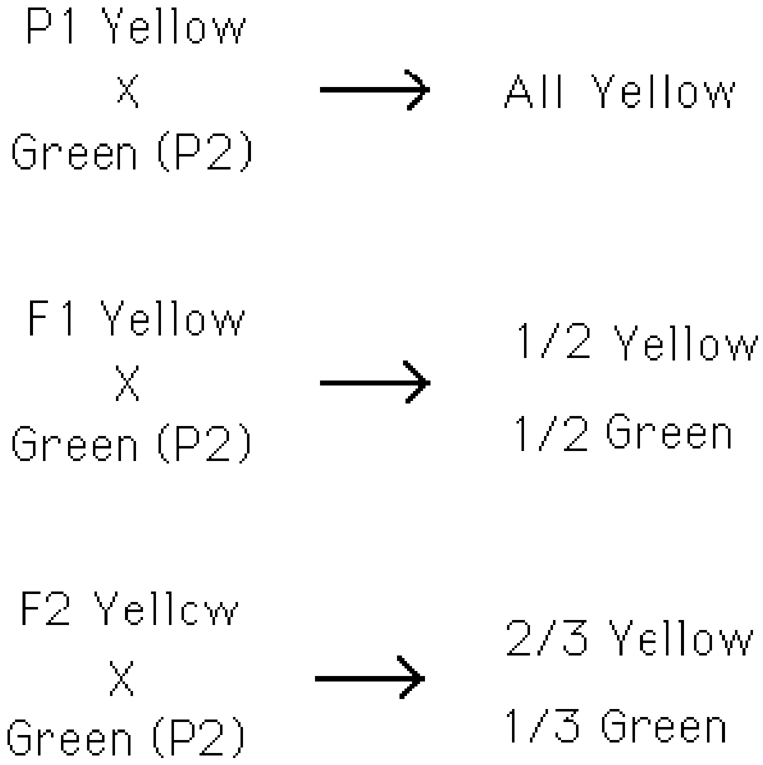
What Mendel observed



- The F_1 were all Yellow
- Strong evidence for discrete units of heredity , as "green" unit obviously present in F_1, appears in F_2
- There is a 3:1 ratio of Yellow : Green in F2

What Mendel observed (continued)

- Parental, F_1 and F_2 yellow peas behave quite differently



P1 Yellow
X
Green (P2)
→ All Yellow

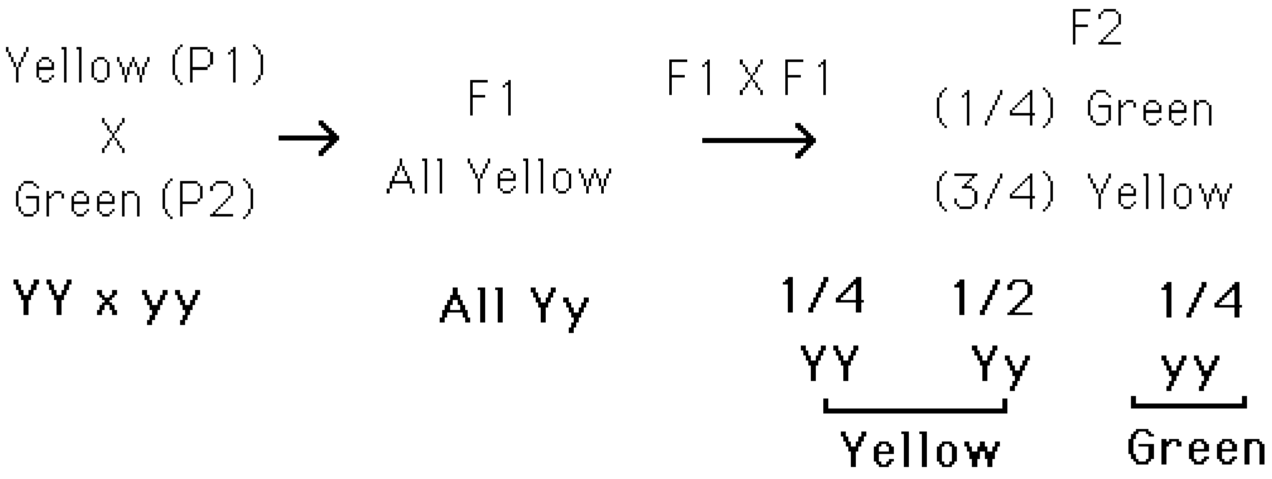
F1 Yellow
X
Green (P2)
→ 1/2 Yellow
1/2 Green

F2 Yellcw
X
Green (P2)
→ 2/3 Yellow
1/3 Green

Mendel’s conclusions

- **Mendel’s first law** (law of segregation of characteristics)

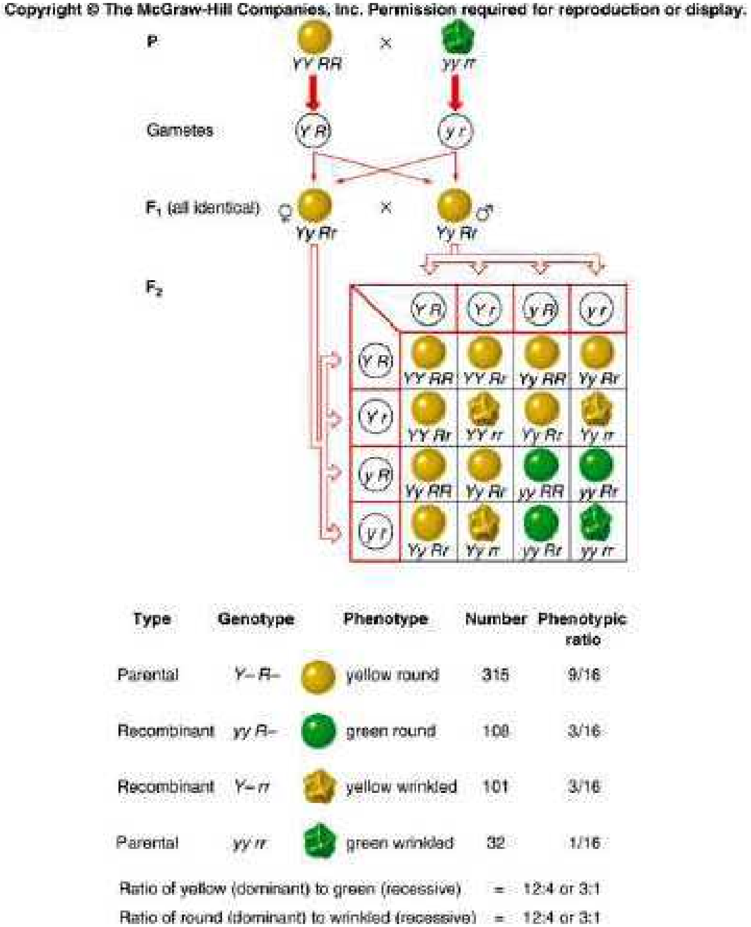
  This says that of a pair of characteristics (e.g. blue and brown eye colour) only one can be represented in a gamete. What he meant was that for any pair of characteristics there is only one gene in a gamete even though there are two genes in ordinary cells.



Mendel's conclusions (continued)

- **Mendel's second law** (law of independent assortment)

  This says that for two characteristics the genes are inherited independently.



	Y R	Y r	y R	y r
Y R	YY RR	YY Rr	Yy RR	Yy Rr
Y r	YY Rr	YY rr	Yy Rr	Yy rr
y R	Yy RR	Yy Rr	yy RR	yy Rr
y r	Yy Rr	Yy rr	yy Rr	yy rr

Type	Genotype	Phenotype	Number	Phenotypic ratio
Parental	Y– R–	yellow round	315	9/16
Recombinant	yy R–	green round	108	3/16
Recombinant	Y– rr	yellow wrinkled	101	3/16
Parental	yy rr	green wrinkled	32	1/16

Ratio of yellow (dominant) to green (recessive) = 12:4 or 3:1

Ratio of round (dominant) to wrinkled (recessive) = 12:4 or 3:1

Mendelian transmission in simple words

- One copy of each gene is inherited from the mother and one from the father. These copies are not necessarily identical
- Mendel postulated that mother and father each pass one of their two copies of each gene independently and at random
- At a given locus, the father carries alleles a and b and the mother carries c and d, the offspring may be a/c, a/d, b/c or b/d, each with probability 1/4
- Transmission of genes at two different positions, or loci, on the same *chromosome* (see later) may not be independent. If not, they are said to be *linked*

The cell as the basic unit of biological functioning

- Let us take it a few levels up …
- Although the tiniest bacterial cells are incredibly small, weighing less than 10-12 grams, each is in effect a veritable micro-miniaturized factory containing thousands of exquisitely designed pieces of intricate molecular machinery, made up altogether of one hundred thousand million atoms, far more complicated than any machinery built by man and absolutely without parallel in the non-living world.
- Each microscopic cell is as functionally complex as a small city. When magnified 50,000 times through electron micrographs, we see that a cell is made up of multiple complex structures, each with a different role in the cell's operation.

(http://www.allaboutthejourney.org/cell-structure.htm)

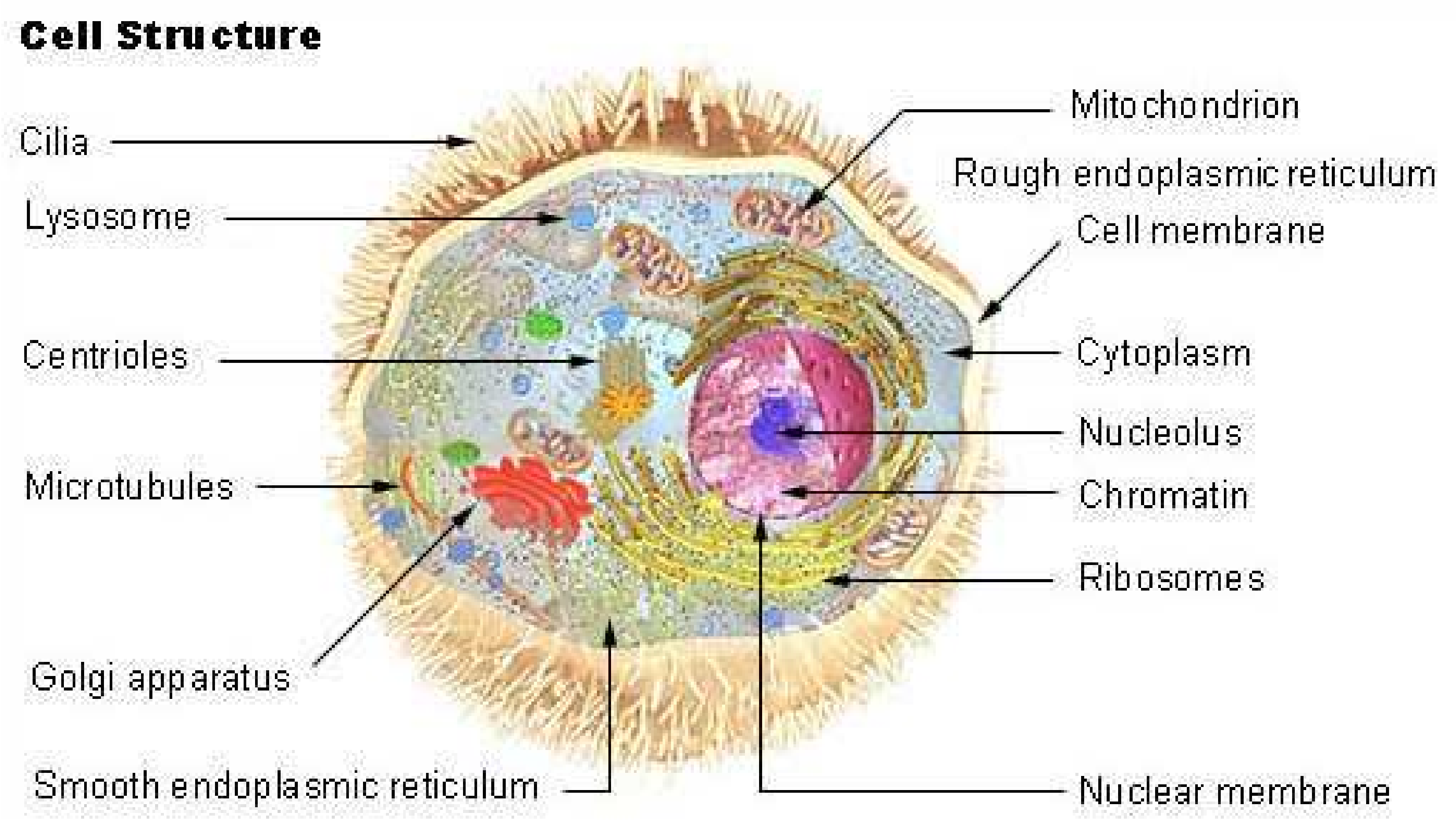
The cell as the basic unit of biological functioning

- Using the city comparison, here's a simple chart that reveals the design of a typical human cell:

City	Cell
Workers	Proteins
Power plant	Mitochondria
Roads	Actin fibers, Microtubules
Trucks	Kinesin, Dinein
Factories	Ribosomes
Library	Genome
Recycling center	Lysosomes
Police	Chaperones
Post office	Golgi Apparatus

(http://www.allaboutthejourney.org/cell-structure.htm)

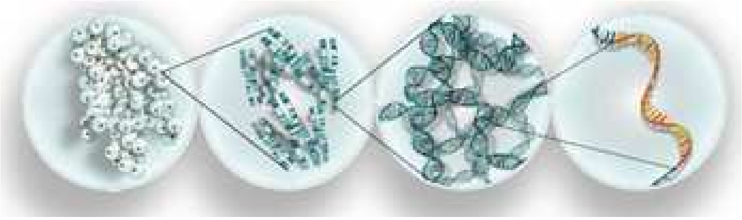
The cell as the basic unit of biological functioning



(http://training.seer.cancer.gov/anatomy/cells_tissues_membranes/cells/structure.html)

The miracle of lifel

- There are three explanatory platforms:



(VIB, Biotechnology)

  - The cells of the living organism. The cells are thus the basic unit of all biological functions
  - The genetic instructions that are responsible for the properties of the cell
  - The biological mechanisms that are used by the cells to carry out the instructions.
- The genetic instructions are stored in code in the DNA. The collection of all possible genetic instructions in a cell is called the *genome.*

History revealed how that genes involved DNA

Geneticists already knew that DNA held the primary role in determining the structure and function of each cell in the body, but they did not understand the mechanism for this or that the structure of DNA was directly involved in the genetic process.

British biophysicist Francis Crick and American geneticist **James Watson** undertook a joint inquiry into the structure of DNA in 1951.

(http://www.pbs.org/wgbh/nova/genome)

Watson and Crick

> *"We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A). This structure has novel features which are of considerable biological interest."*

(Watson JD and Crick FHC. A Structure for DNA, *Nature*, 1953)

No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, 40, 149 (1920).

[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, 5, 285 (1949).

[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., 11 (3) (1950).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), 2 (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

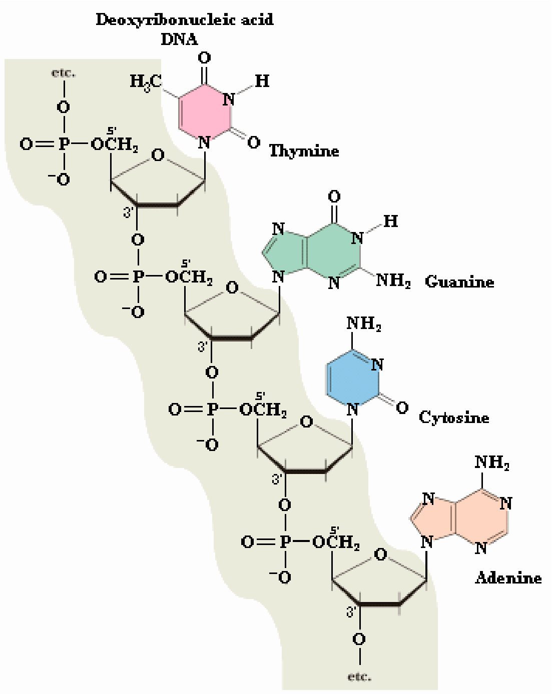
We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

What does “DNA” stand for?

- Deoxyribonucleic acid (DNA) IS the genetic information of most living organisms. In contrast, some viruses (called retroviruses) use ribonucleic acid as genetic information. “Genes” correspond to sequences of DNA
- DNA is a polymere (i.e., necklace of many alike units), made of units called nucleotides.
- Some interesting features of DNA include:
    - DNA can be copied over generations of cells*: DNA replication*
    - DNA can be translated into proteins: *DNA transcription* into RNA, further translated into proteins
    - DNA can be repaired when needed: *DNA repair*.

What does “DNA” stand for?

- There are 4 nucleotide *bases*, denoted A (adenine), T (thymine), G (guanine) and C (cytosine)
- A and G are called purines,
  T and C are called pyrimidines
  (smaller molecules than purines)
- The two strands of DNA in the double helix structure are complementary
  (sense and anti-sense strands); A binds with T and G binds with C



(Biochemistry 2nd Ed. by Garrett & Grisham)

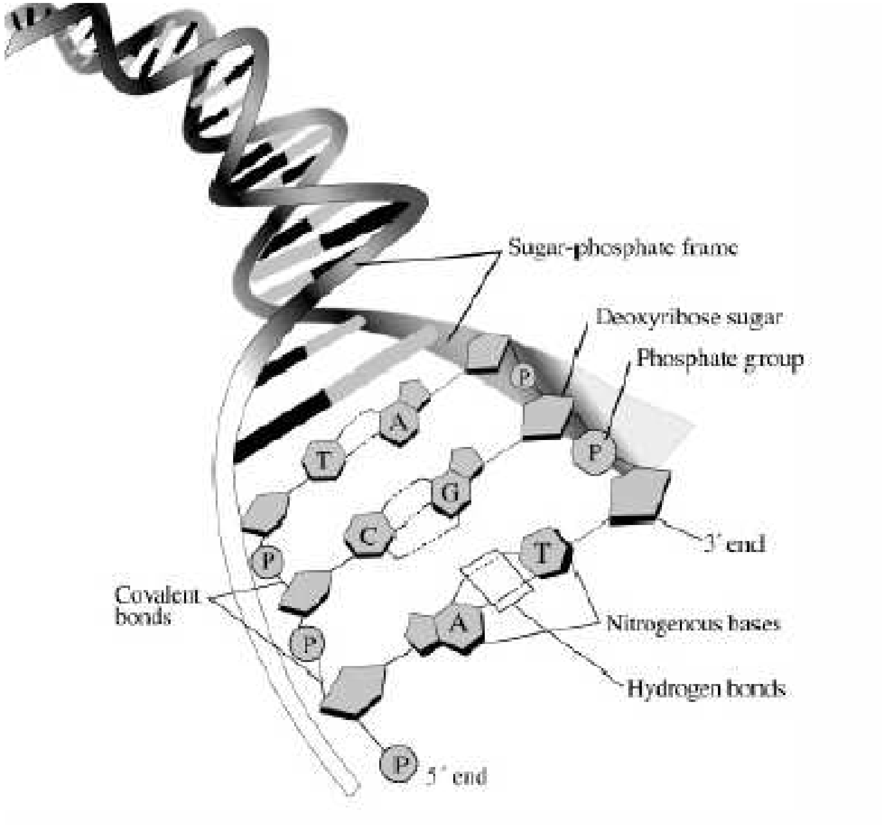
Primary structure of DNA

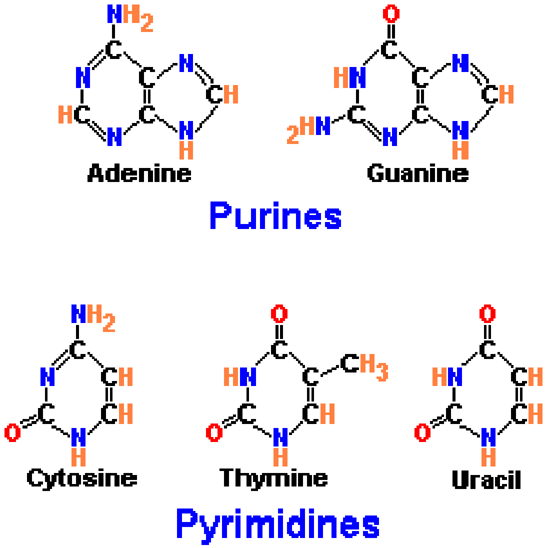
The 3 dimensional structure of DNA can be described in terms of primary, secondary, tertiary, and quaternary structure.

- The primary structure of DNA is the sequence itself - the order of nucleotides in the deoxyribonucleic acid polymer.
- A *nucleotide* consists of
  - a phosphate group,
  - a deoxyribose sugar and
  - a nitrogenous base.
- Nucleotides can also have other functions such as carrying energy: ATP
- Note: Nucleo **s** ides are made of a sugar and a nitrogenous base…

Nucleotides

Nitrogenous bases

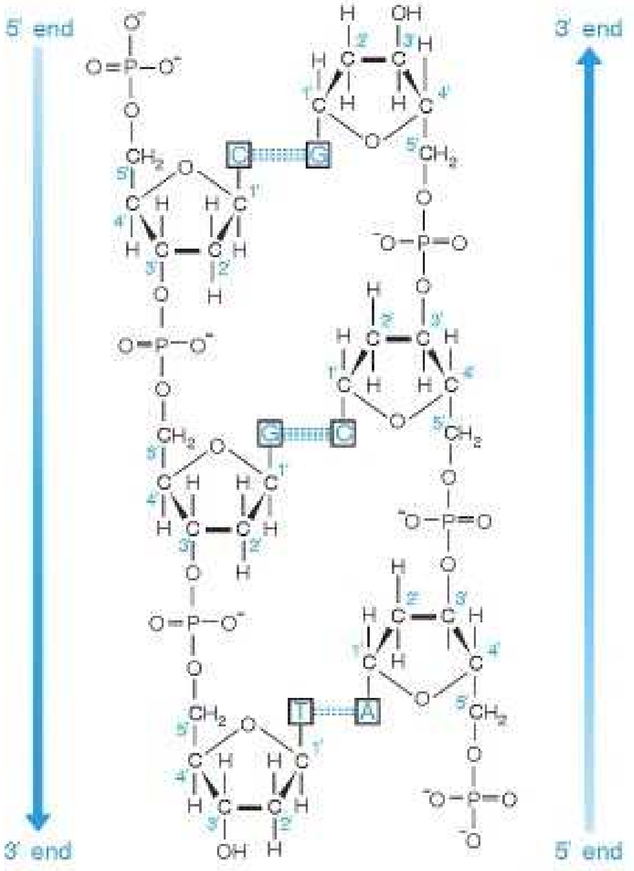




(http://www.sparknotes.com/101/index.php/biology)

Secondary structure of DNA

- The secondary structure of DNA is relatively straightforward - it is a double helix.
- It is related to the hydrogen bonding
- The two strands are anti-parallel.
  - *The 5' end* is composed of a phosphate group that has not bonded with a sugar unit.
  - *The 3' end* is composed of a sugar unit whose hydroxyl group has not bonded with a phosphate group.



Major groove and minor groove

- The double helix presents a major groove and a minor groove (Figure 1).
  - The major groove is deep and wide
  - The minor groove is narrow and shallow.
- The chemical groups on the edges of GC and AT base pairs that are available for interaction with proteins in the major and minor grooves are color-coded for different types of interactions (Figure 2)

Figure 1

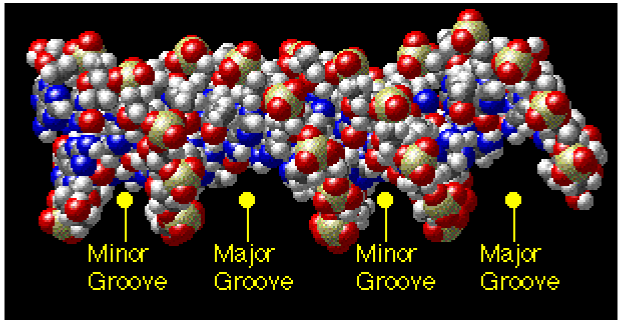
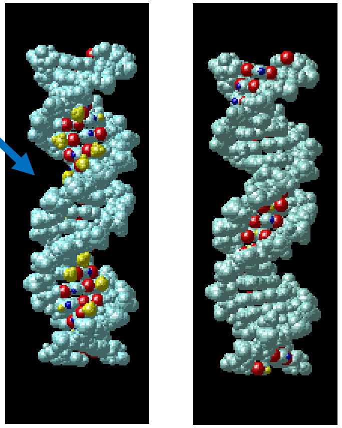


Figure 2



Tertiary structure of DNA

- This structure refers to how DNA is stored in a confined space to form the chromosomes.
- It varies depending on whether the organisms prokaryotes and eukaryotes:
  - In prokaryotes the DNA is folded like a super-helix, usually in circular shape and associated with a small amount of protein. The same happens in cellular organelles such as mitochondria .
  - In eukaryotes, since the amount of DNA from each chromosome is very large, the packing must be more complex and compact, this requires the presence of proteins such as histones and other proteins of non-histone nature
- Hence, in humans, the double helix is itself super-coiled and is wrapped around so-called histones (see later).

- *Eukaryotes*: organisms with a rather complex cellular structure. In their cells we find organelles, clearly discernable compartments with a particular function and structure.
  - The organelles are surrounded by semi-permeable membranes that compartmentalize them further in the cytoplasm.
  - The Golgi apparatus is an example of an organelle that is involved in the transport and secretion of proteins in the cell.
  - Mitochondria are other examples of organelles, and are involved in respiration and energy production

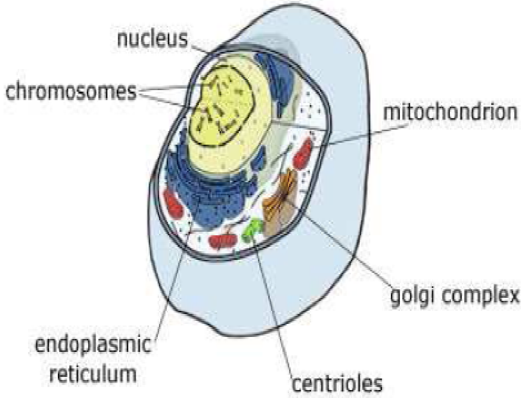
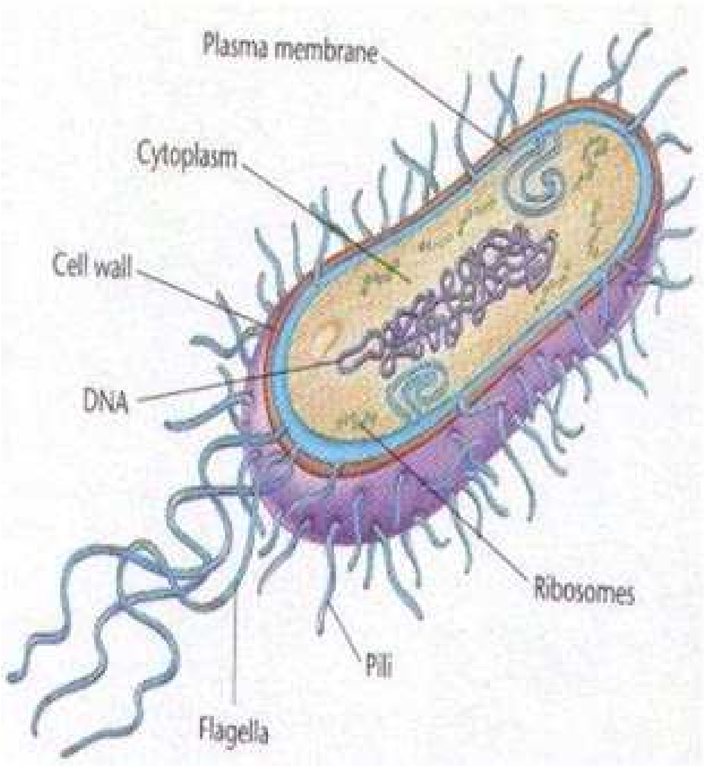


Image adapted from: National Human Genome Research Institute.

A typical eukaryotic cell.

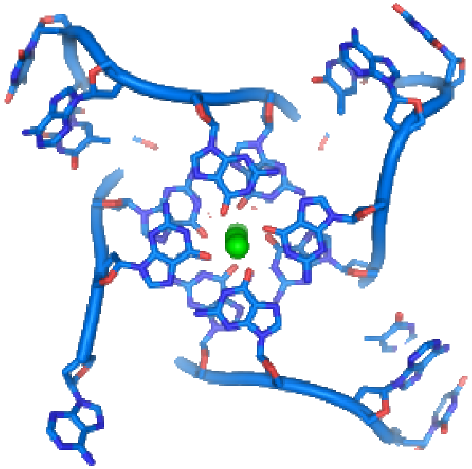
- *Prokaryotes*: cells without organelles where the genetic information floats freely in the cytoplasm



Quaternary structure of DNA

- At the ends of linear chromosomes are specialized regions of DNA called telomeres.
- The main function of these regions is to allow the cell to replicate chromosome ends using the enzyme telomerase, since other enzymes that replicate DNA cannot copy the 3 'ends of chromosomes.

- In human cells, telomeres are long areas of single-stranded DNA containing several thousand repetitions of a single sequence TTAGGG.

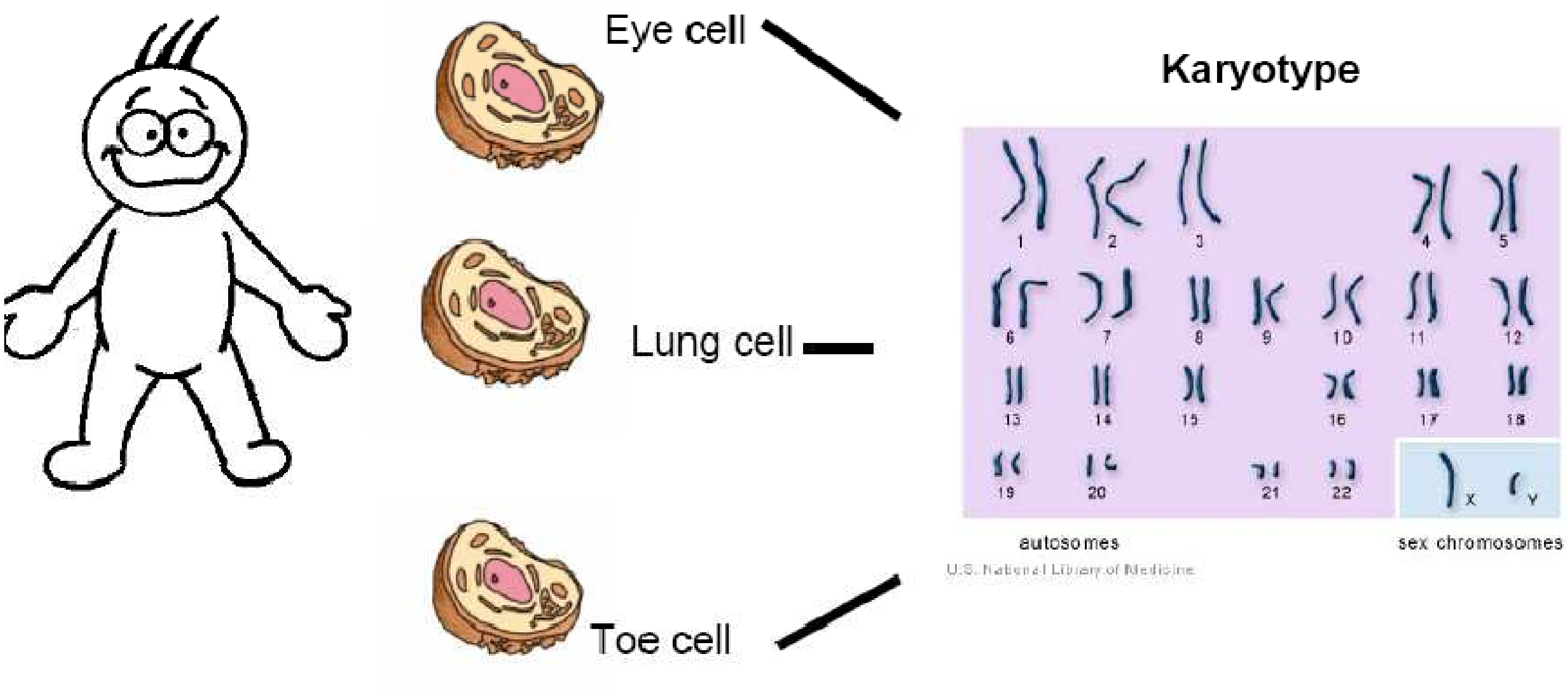


(http://www.boddunan.com/miscellaneous)

The structure of DNA

- A wide variety of proteins form complexes with DNA in order to replicate it, transcribe it into RNA, and regulate the transcriptional process (central dogma of molecular biology).
  - P*roteins* are long chains of amino acids
  - An *amino acids* being an organic compound containing amongst others an amino group (NH_2) and a carboxylic acid group (COOH))
  - Think of aminco acids as 3-letter words of nucleotide building blocks (letters).

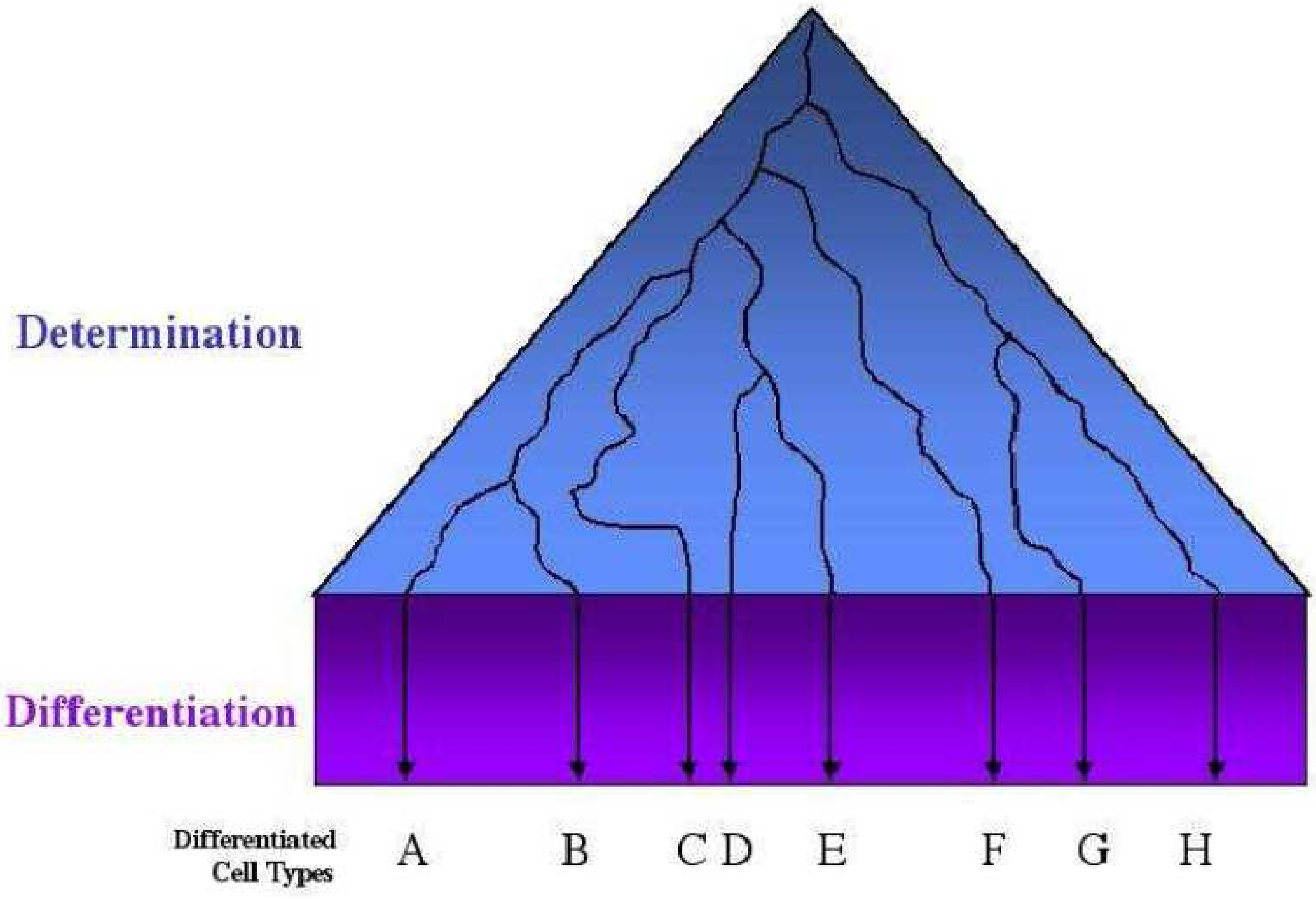
Every cell in the body has the same DNA



- One base pair is 0.00000000034 meters
- DNA sequence in any two people is 99.9% identical – only 0.1% is unique!

Differential expression

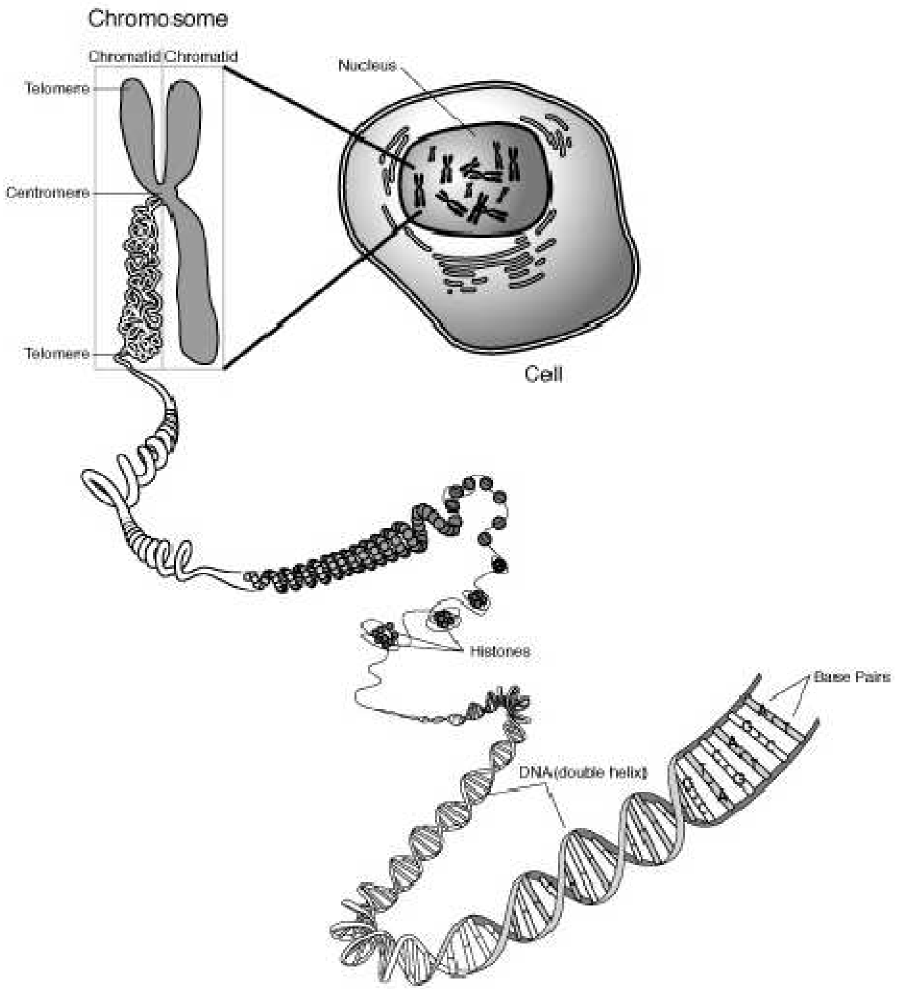
- The determination of different cell types (cell fates) involves progressive restrictions in their developmental potentials. When a cell “chooses” a particular fate, it is said to be determined, although it still "looks" just like its undetermined neighbors. Determination implies a stable change - the fate of determined cells does not change.
- Differentiation follows determination, as the cell elaborates a cell-specific developmental program. Differentiation results in the presence of cell types that have clear-cut identities, such as muscle cells, nerve cells, and skin cells.
- Differentiation results from differential gene expression


Determination
Differentiation
Differentiated Cell Types
A
B
C
D
E
F
G
H

Chromosomes

- In the nucleus of each cell, the DNA molecule is packaged into thread-like structures called chromosomes. Each chromosome is made up of DNA tightly coiled many times around proteins called histones (see later) that support its structure.
- Chromosomes are not visible in the cell’s nucleus—not even under a microscope—when the cell is not dividing.
- However, the DNA that makes up chromosomes becomes more tightly packed during cell division and is then visible under a microscope. Most of what researchers know about chromosomes was learned by observing chromosomes during cell division.

Histones: packaging of DNA in the nucleus

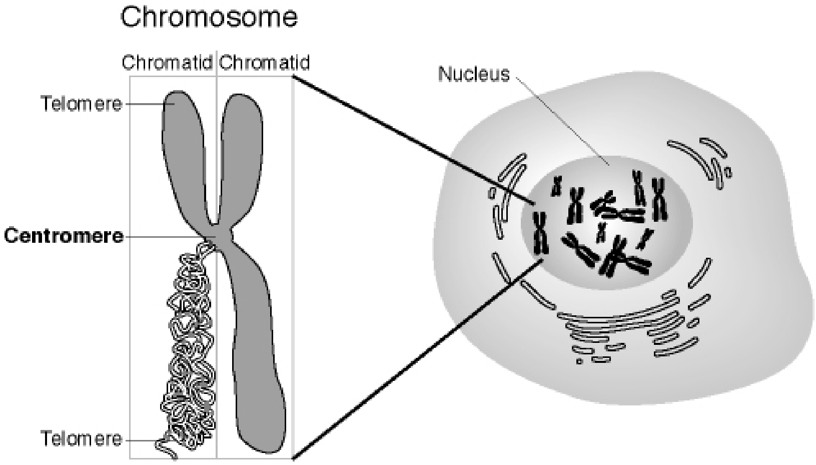


http://www.accessexcellence.org/AB/GG/chromosome.html

- *Histones* are proteins rich in lysine and arginine residues and thus positively-charged.
- For this reason they bind tightly to the negatively-charged phosphates in DNA.

Chromosomes

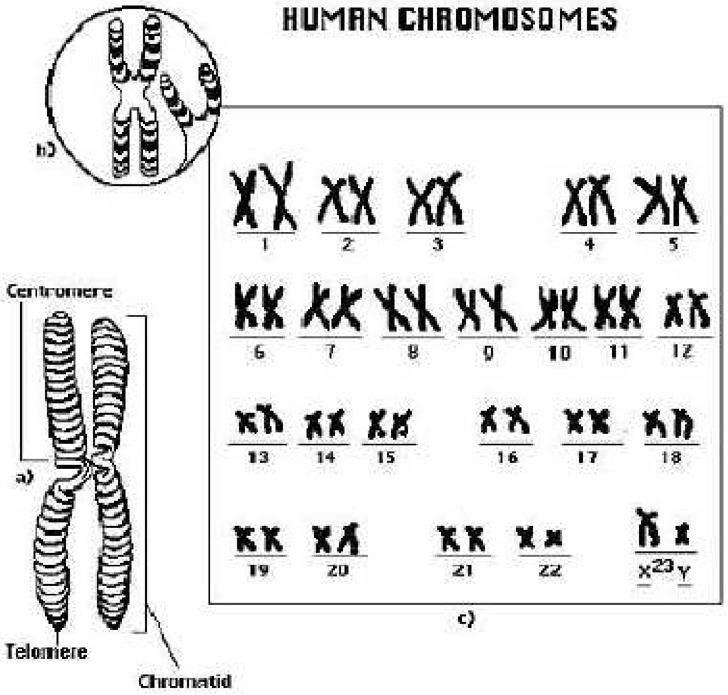
- All chromosomes have a stretch of repetitive DNA called the centromere. This plays an important role in chromosomal duplication before cell division.
- If the centromere is located at the extreme end of the chromosome, that chromosome is called acrocentric.
- If the centromere is in the middle of the chromosome, it is termed metacentric
- The ends of the chromosomes (that are not centromeric) are called telomeres. They play an important role in aging.



(www.genome.gov)

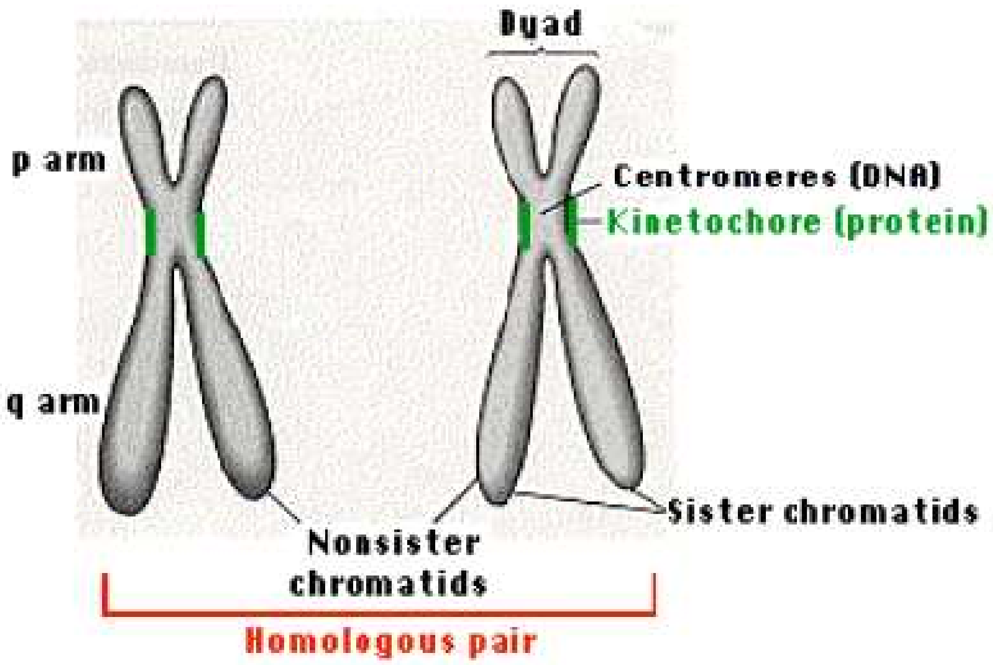
Chromosomes

- The short arm of the chromosome is usually termed *p* for petit (small), the long arm, *q*, for queue (tall).
- The telomeres are correspondingly referred to as *pter* and *qter*.



Chromatids

- A chromatid is one among the two identical copies of DNA making up a replicated chromosome, which are joined at their centromeres, for the process of cell division (mitosis or meiosis – see later).



Sex chromosomes

- Homogametic sex :

  that sex containing two like sex chromosomes

  - In most animals species these are females (XX)
  - Butterflies and Birds, ZZ males

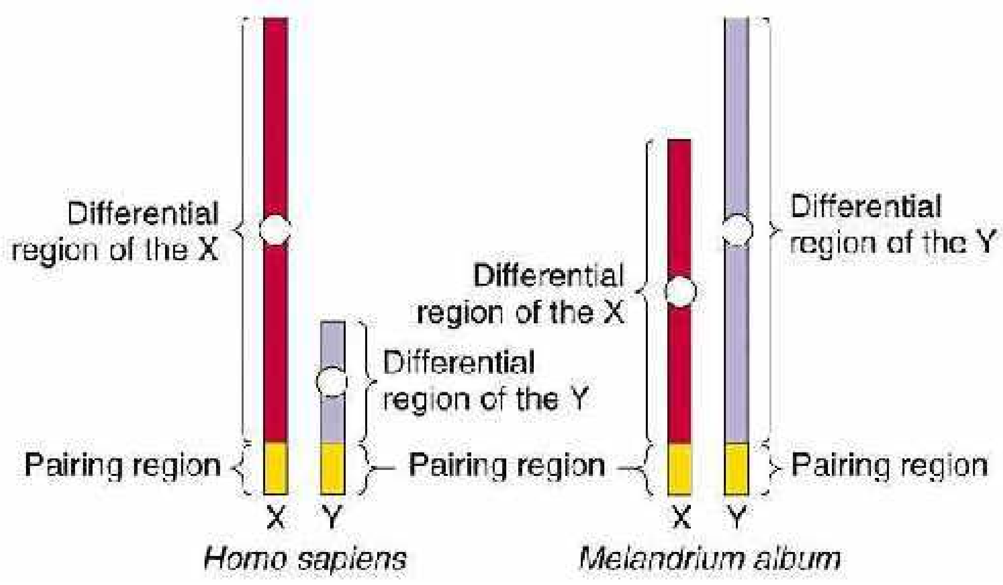
- Heterogametic sex:

  that sex containing two different sex chromosomes

  - In most animal species these are XY males
  - Butterflies and birds, ZW females
  - Grasshopers have XO males

Pairing of sex chromosomes

- In the homogametic sex: pairing happens like normal autosomal chromosomes
- In the heterogametic sex: The two sex chromosomes are very different, and have special pairing regions to insure proper pairing at meiosis


Differential region of the X
Differential region of the Y
Pairing region
Pairing region
X
Y
Homo sapiens
Differential region of the X
Differential region of the Y
Pairing region
X
Y
Melandrium album

X-inactivation

- X-inactivation (also called lyonization) is a process by which one of the two copies of the X chromosome present in female mammals is inactivated
- X-inactivation occurs so that the female, with two X chromosomes, does not have twice as many X chromosome gene products as the male, which only possess a single copy of the X chromosome

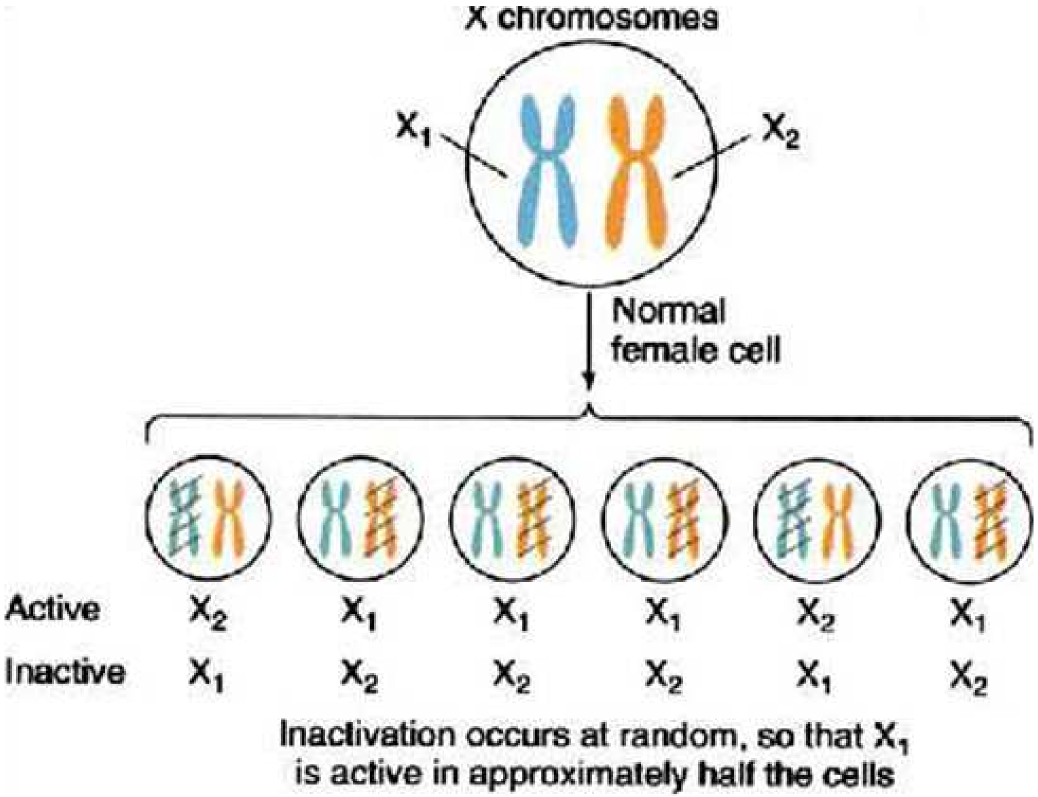
The ginger colour of cats (known as "yellow", "orange" or "red" to cat breeders) is caused by the "O" gene. The O gene changes black pigment into a reddish pigment. The O gene is carried on the X chromosome. A normal male cat has XY genetic makeup; he only needs to inherit one O gene for him to be a ginger cat. A normal female is XX genetic makeup. She must inherit two O genes to be a ginger cat. If she inherits only one O gene, she will be tortoiseshell. The O gene is called a sex-linked gene because it is carried on a sex chromosome. Tortoiseshell cats are therefore heterozygous (not true-breeding) for red colour.

The formation of red and black patches in a female with only one O gene is through a process known as X-chromosome inactivation. Some cells randomly activate the O gene while others activate the gene in the equivalent place on the other X chromosome.

(wikipedia)

X-inactivation

- The choice of which X chromosome will be inactivated is random in placental mammals such as mice and humans, but once an X chromosome is inactivated it will remain inactive throughout the lifetime of the cell.

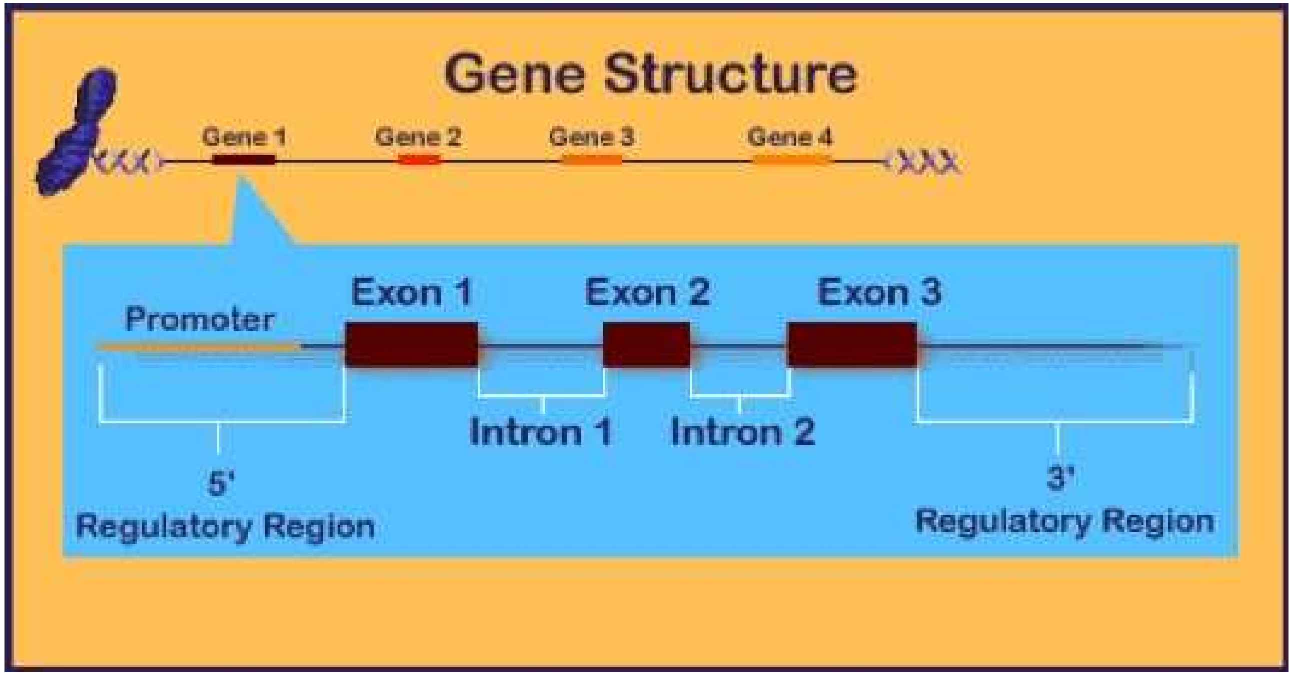

X chromosomes
X_1
X_2
Normal female cell
Active X_2 X_1 X_1 X_1 X_2 X_1
Inactive X_1 X_2 X_2 X_2 X_1 X_2
Inactivation occurs at random, so that X_1 is active in approximately half the cells

The human genome

- The human genome consists of about 3 ×10^9 base pairs and contains about 22,000 genes
- Cells containing 2 copies of each chromosome are called diploid (most human cells).

  Cells that contain a single copy are called haploid.
- Humans have 23 pairs of chromosomes: 22 autosomal pairs and one pair of sex chromosomes
- Females have two copies of the X chromosome, and males have one X and one Y chromosome
- Much of the DNA is either in introns or in intragenic regions … which brings us to study the transmission or exploitation of genetic information in more detail.

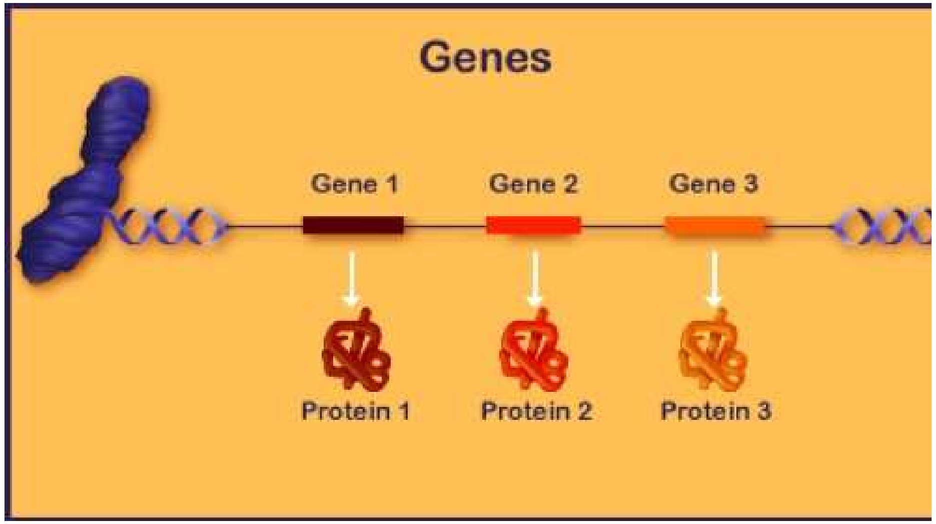
1.b What does the genetic information mean?

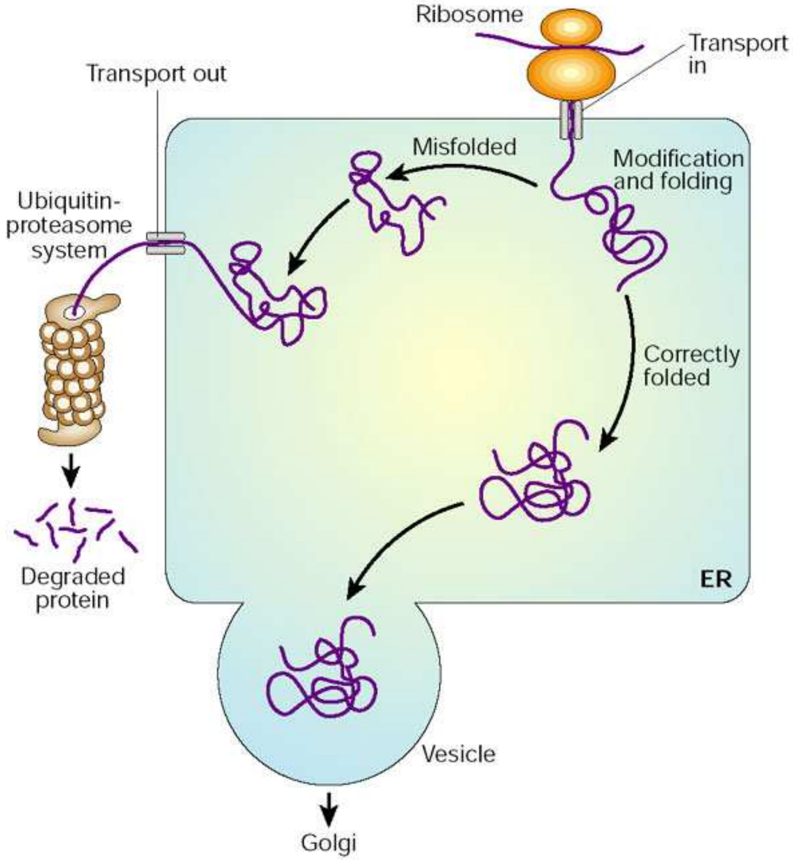


(Roche Genetics)

- *Promoter*: Initial binding site for RNA polymerase in the process of gene expression. First transcription factors bind to the promoter which is located 5' to the transcription initiation site in a gene.

Genes and Proteins

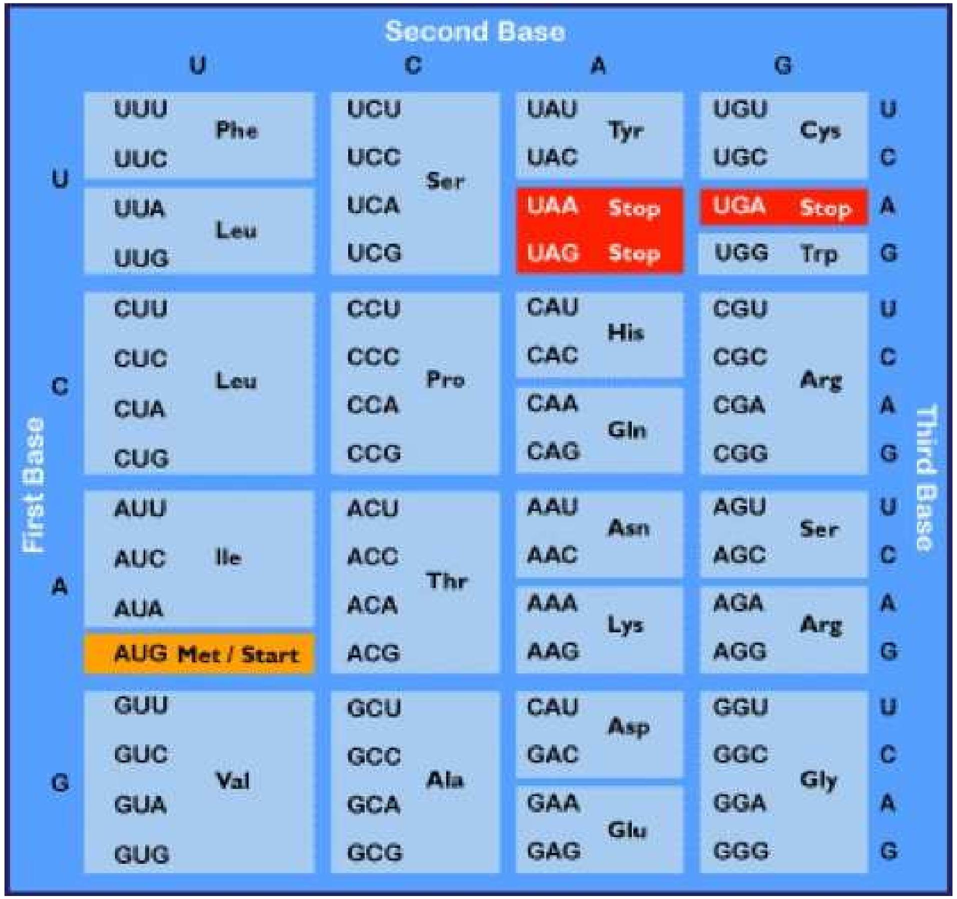




(Roche Genetics)

(http://www.nature.com/nature/journal/v426/n6968/images/nature02261-f2.2.jpg)

Translation table from DNA building stones to protein building stones



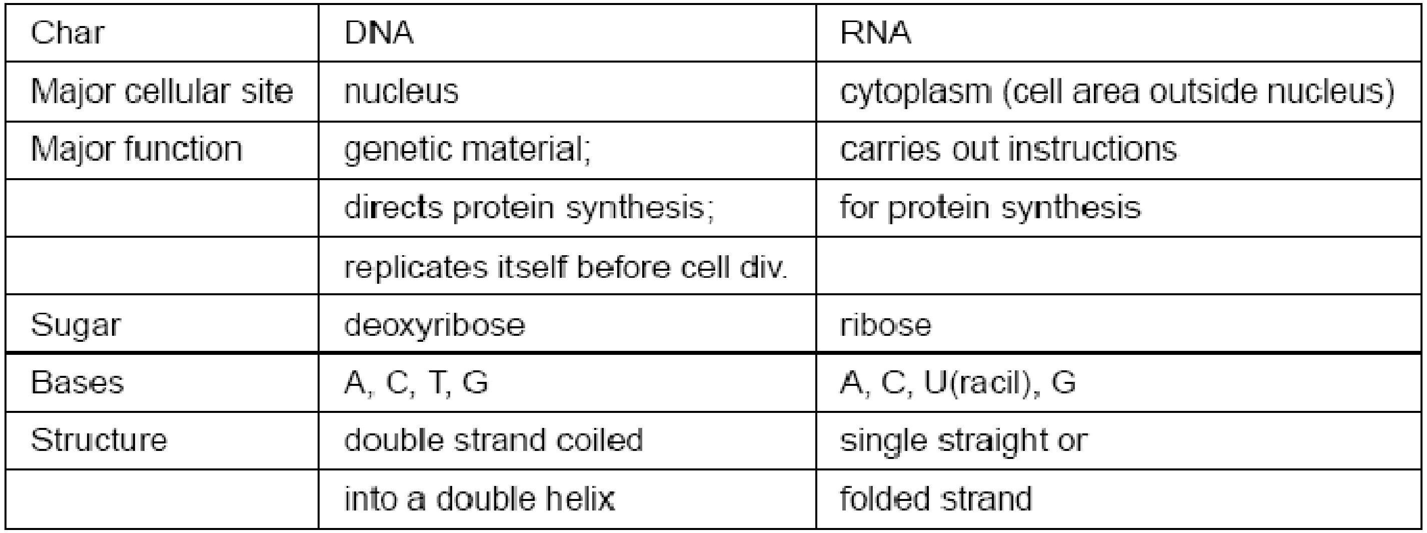
	Second Base				
First Base	U	C	A	G	Third Base
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA Stop	UGA Stop	A
	UUG	UCG	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Gln	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys	AGA Arg	A
	AUG Met / Start	ACG	AAG	AGG	G
G	GUU Val	GCU Ala	CAU Asp	GGU Gly	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu	GGA	A
	GUG	GCG	GAG	GGG	G

(Roche Genetics)

- Where does the U come from?

Comparison between DNA and RNA

- Pieces of coding material that the cells needs at a particular moment, is transcribed from the DNA in RNA for use outside the cell nucleus.



Char	DNA	RNA
Major cellular site	nucleus	cytoplasm (cell area outside nucleus)
Major function	genetic material;	carries out instructions
	directs protein synthesis;	for protein synthesis
	replicates itself before cell div.	
Sugar	deoxyribose	ribose
Bases	A, C, T, G	A, C, U(racil), G
Structure	double strand coiled	single straight or
	into a double helix	folded strand

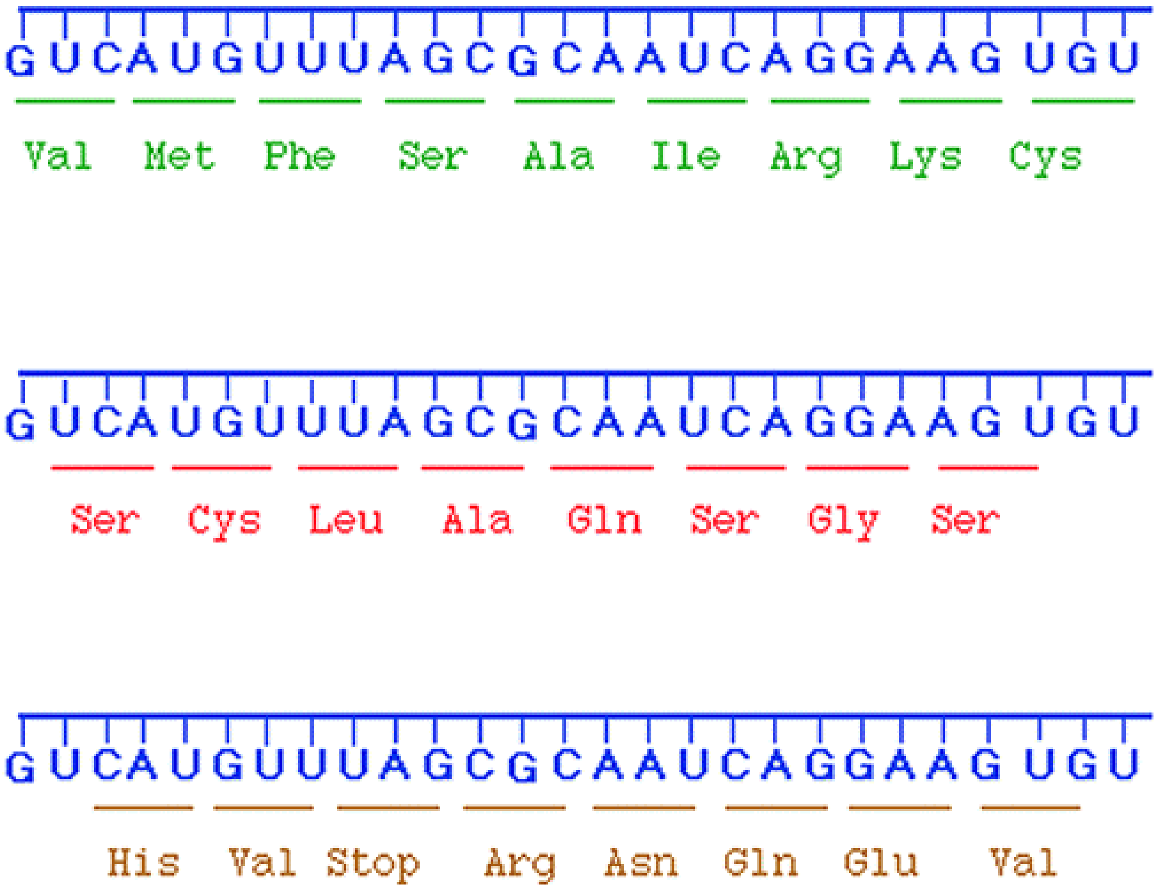
(Human Anatomy & Physiology - Addison-Wesley 4th ed)

- Note that in RNA U(racil), another pyrimidine, replaces T in DNA

Reading the code

- Because there are only 20 amino acids that need to be coded (using A, C, U or G), the genetic code can be said to be degenerate, with the third position often being redundant
- The code is read in triplets of bases.
- Depending on the starting point of reading, there are three possible variants to translate a given base sequence into an amino acid sequence. These variants are called *reading frames*

Reading the code

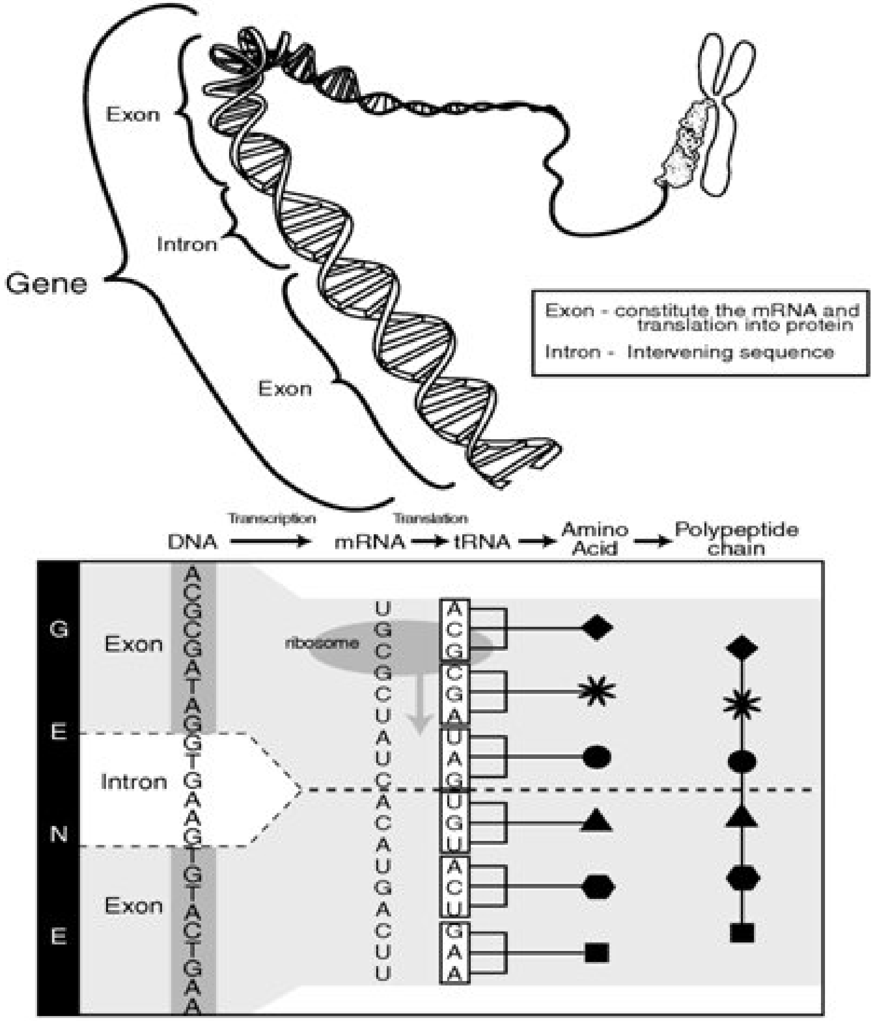

G U C A U G U U U A G C G C A A U C A G G A A G U G U
Val Met Phe Ser Ala Ile Arg Lys Cys
G U C A U G U U U A G C G C A A U C A G G A A G U G U
Ser Cys Leu Ala Gln Ser Gly Ser
G U C A U G U U U A G C G C A A U C A G G A A G U G U
His Val Stop Arg Asn Gln Glu Val

1.c How is the genetic information translated?

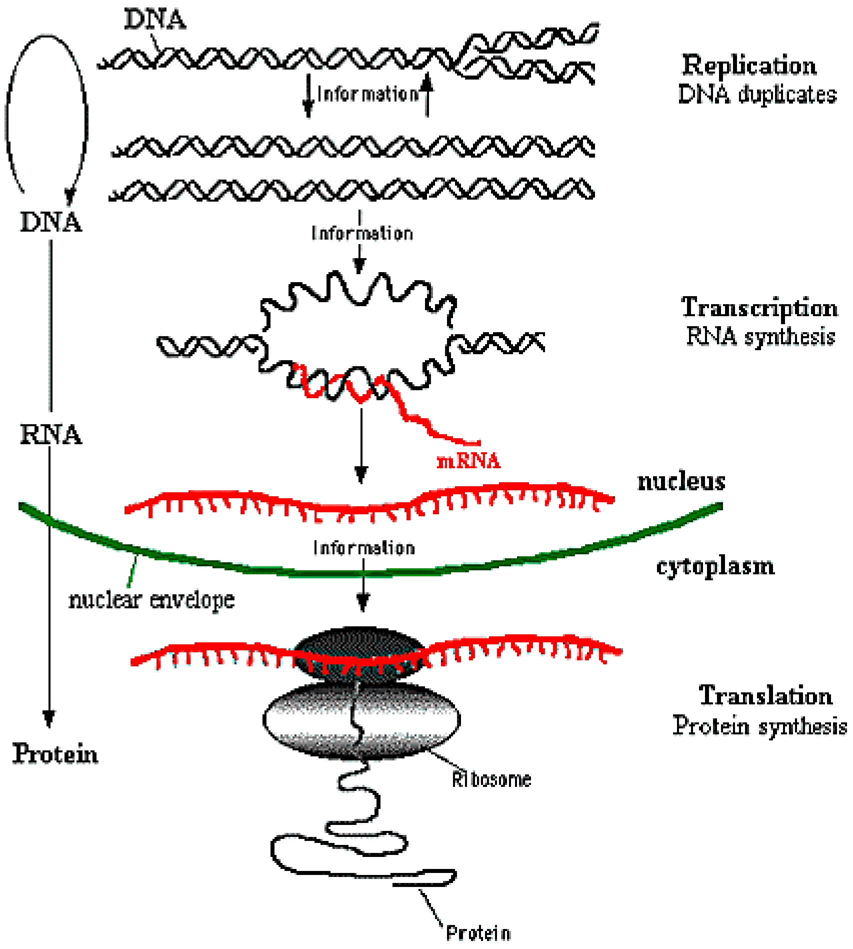
The link between genes and proteins: nucleotide bases

- A gene codes for a protein, but also has sections concerned with gene expression and regulation (E.g., promoter region)
- The translation of bases into amino acids uses RNA and not DNA; it is initiated by a START codon and terminated by a STOP codon.
- Hence, it are the three-base sequences (codons) that code for amino acids and sequences of amino acids in turn form proteins

DNA makes RNA, RNA makes proteins, proteins make us

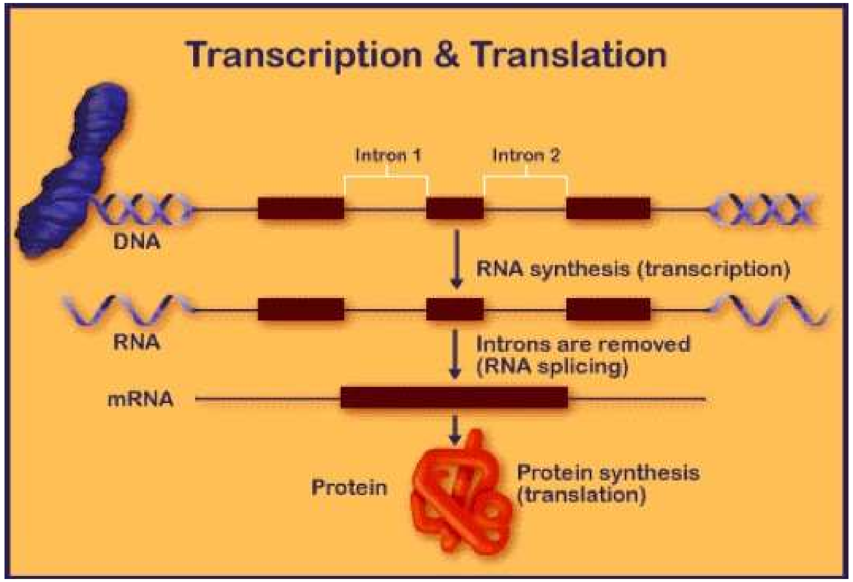


Central dogma of molecular biology



Central dogma of molecular biology

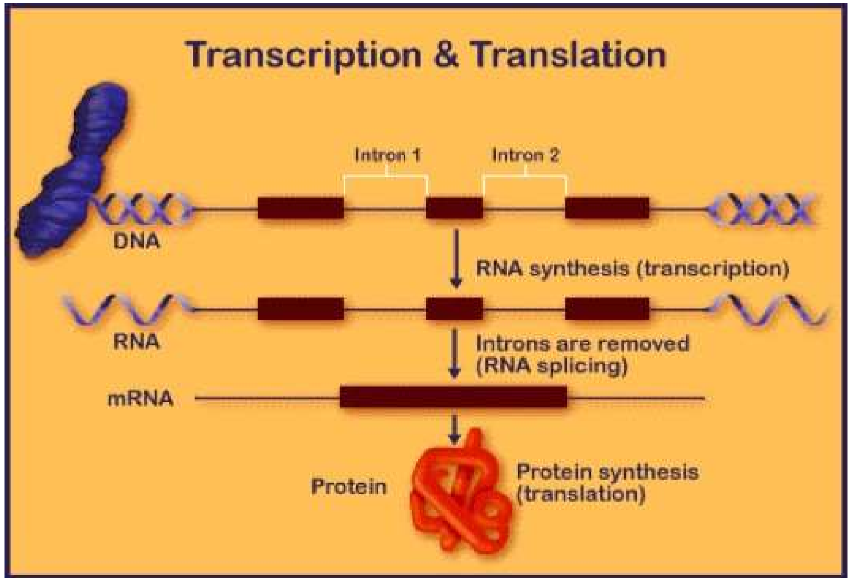
- Stage 1: DNA replicates its information in a process that involves many enzymes. This stage is called the ***replication*** stage.
- Stage 2: The DNA codes for the production of messenger RNA (mRNA) during ***transcription*** of the sense strand (coding or non-template strand)



(Roche Genetics)

So the *coding strand* is the DNA strand which has the same base sequence as the RNA transcript produced (with thymine replaced by uracil). It is this strand *which contains codons*, while the non-coding strand (or anti-sense strand) contains anti-codons.

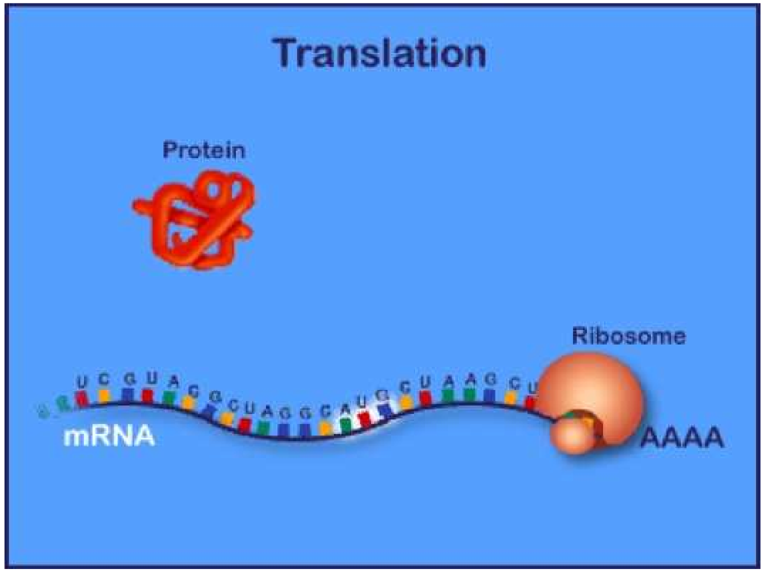
- Stage 3: In eukaryotic cells, the mRNA is ***processed*** (essentially by splicing) and migrates from the nucleus to the cytoplasm



(Roche Genetics)

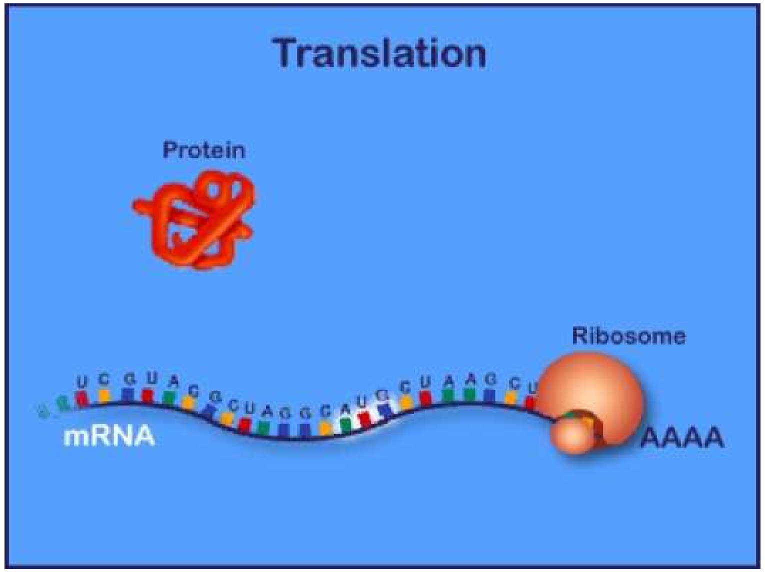
- Stage 4: mRNA carries coded information to ribosomes. The ribosomes "read" this information and use it for protein synthesis. This stage is called the ***translation*** stage.

Translation is facilitated by two key molecules



- *Transfer RNA* (tRNA) molecules transport amino acids to the growing protein chain. Each tRNA carries an amino acid at one end and a three-base pair region, called the anti-codon, at the other end. The anti-codon binds with the codon on the protein chain via base pair matching.

Translation is facilitated by two key molecules (continued)

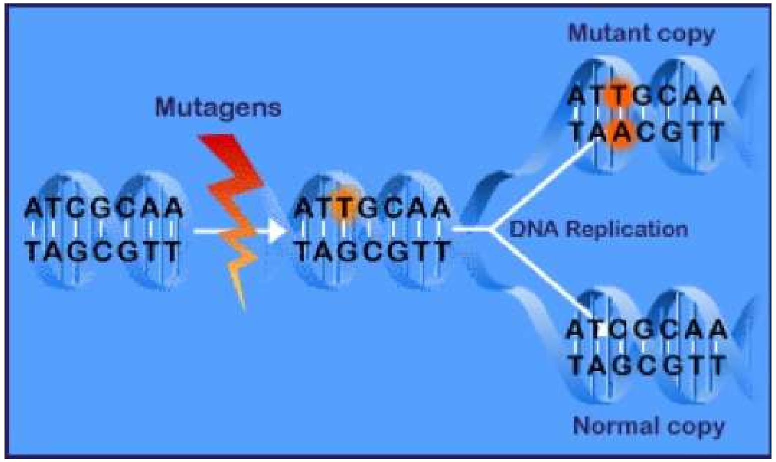


(Roche Genetics)

- *Ribosomes* bind to the mRNA and facilitate protein synthesis by acting as docking sites for tRNA. Each ribosome is composed of a large and small subunit, both made of ribosomal RNA (rRNA) and proteins. The ribosome has three docking sites for tRNA

DNA repair mechanisms

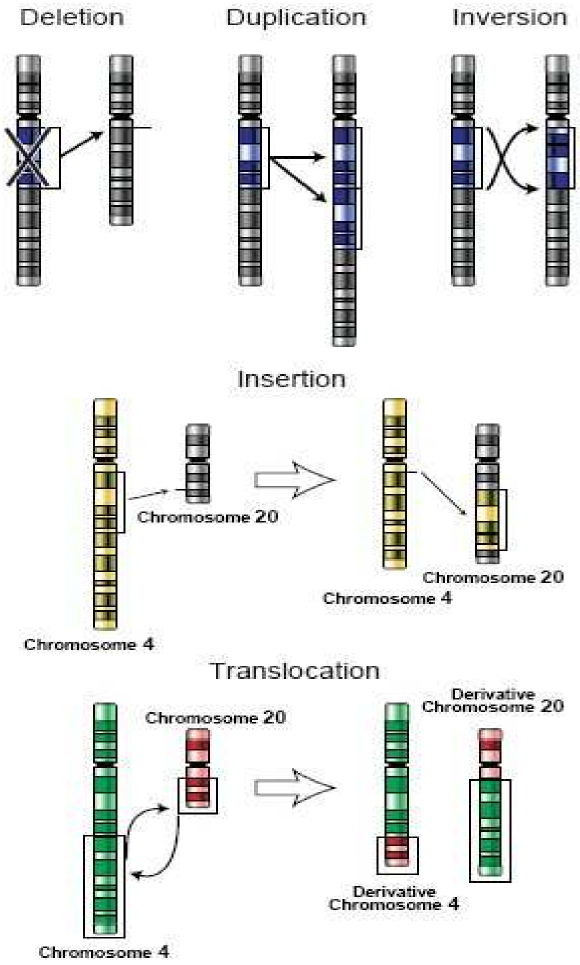
- In biology, a mutagen (Latin, literally origin of change) is a physical or chemical agent that changes the genetic material (usually DNA) of an organism and thus increases the frequency of mutations above the natural background level.
- As many mutations cause cancer, mutagens are typically also carcinogens.
- Not all mutations are caused by mutagens: so-called "spontaneous mutations" occur due to errors in DNA replication, repair and recombination.



(Roche genetics)

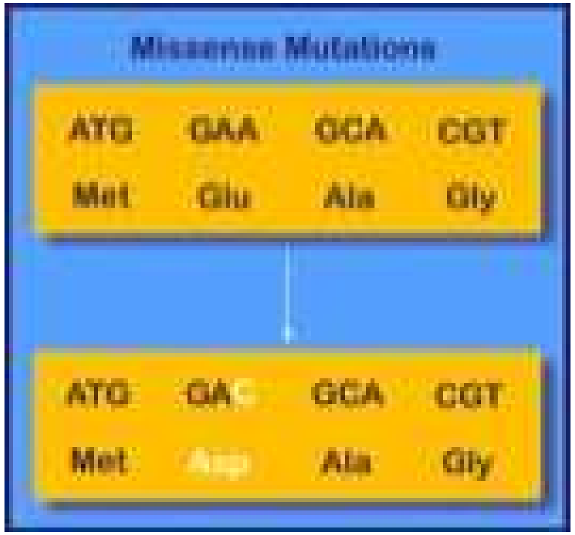
Types of mutations

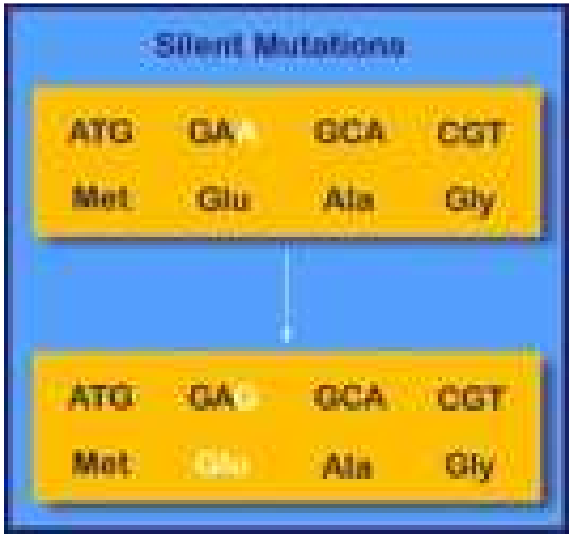
- Deletion
- Duplication
- Inversion
- Insertion
- Translocation

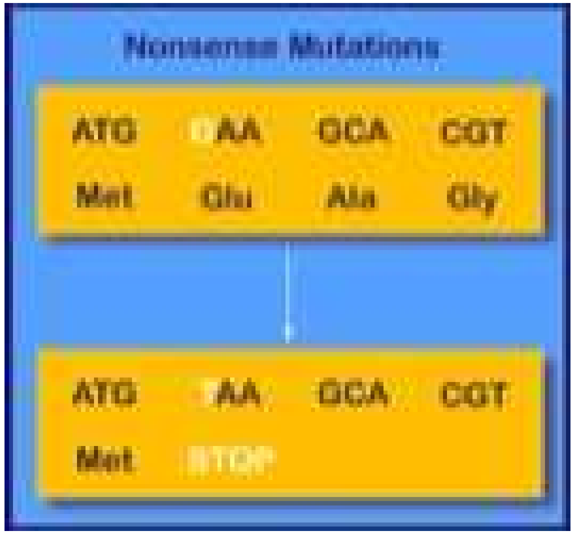


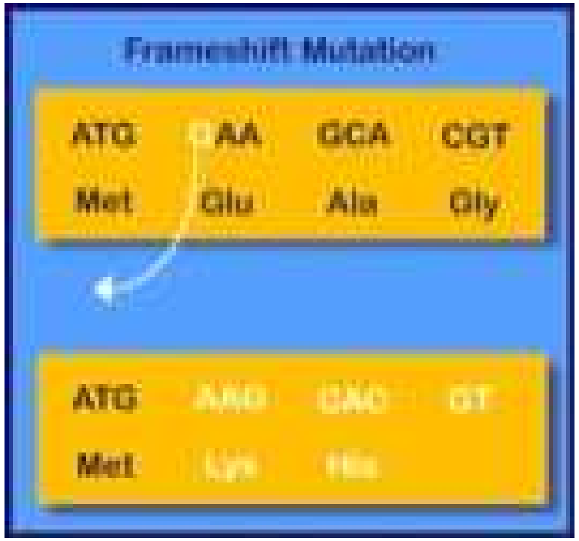
(National Human Genome Research Institute)

Types of mutations (continued)



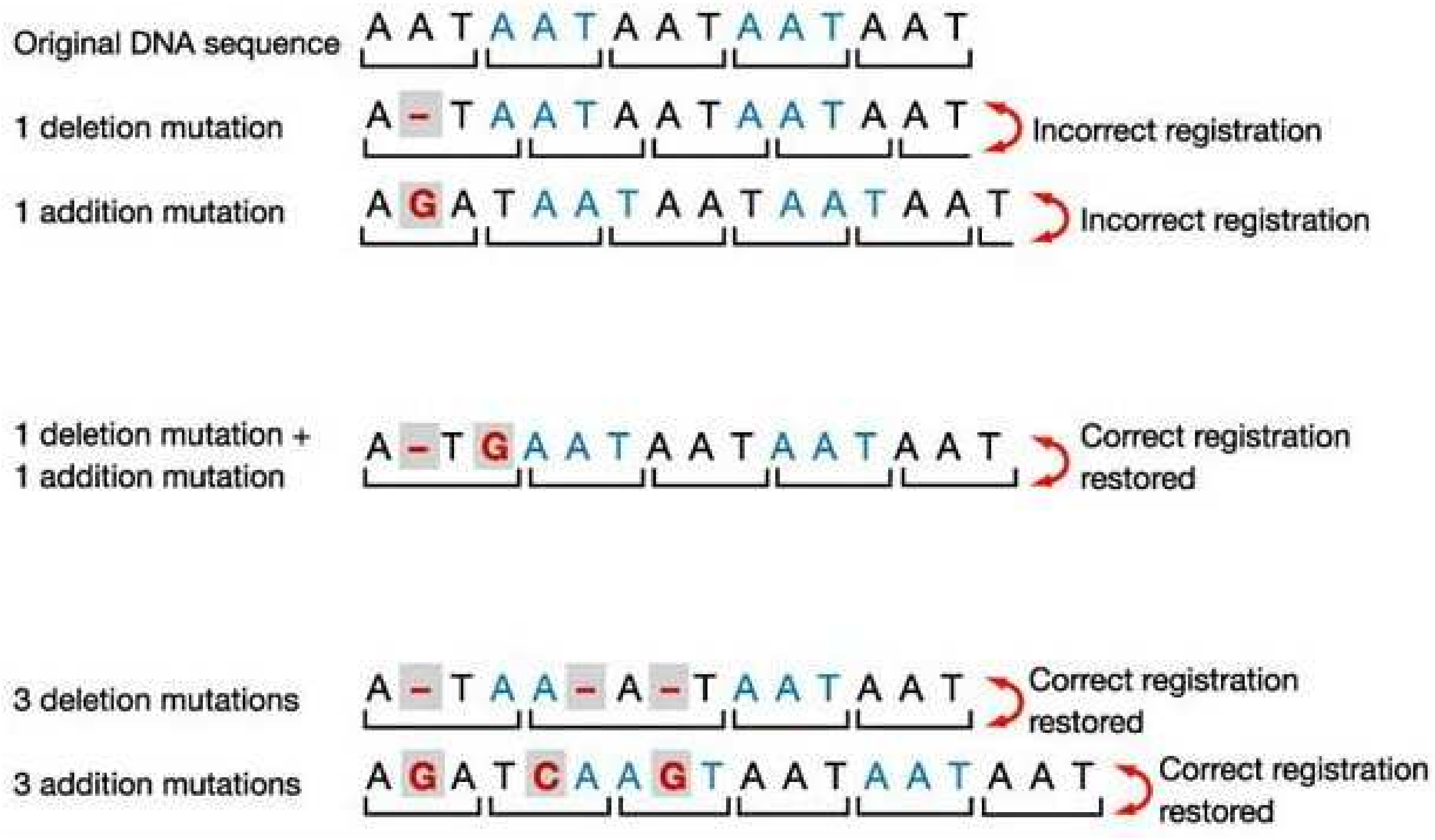






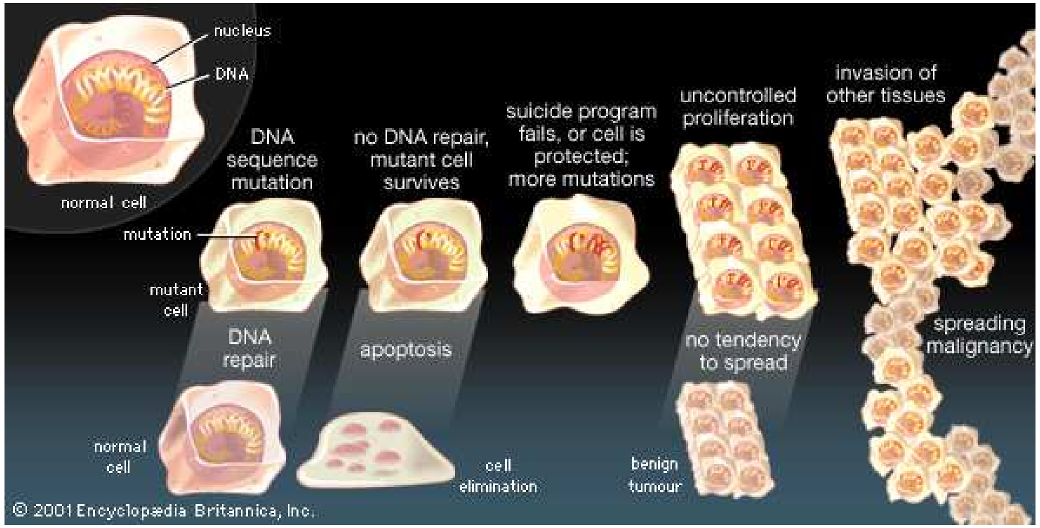
DNA repair mechanisms

- Where it can go wrong when reading the code …


Original DNA sequence
A A T A A T A A T A A T A A T
1 deletion mutation
A – T A A T A A T A A T A A T
Incorrect registration
1 addition mutation
A G A T A A T A A T A A T A A T
Incorrect registration
1 deletion mutation +
1 addition mutation
A – T G A A T A A T A A T A A T
Correct registration restored
3 deletion mutations
A – T A A – A – T A A T A A T
Correct registration restored
3 addition mutations
A G A T C A A G T A A T A A T A A T
Correct registration restored

DNA repair mechanisms

- damage reversal: simplest; enzymatic action restores normal structure without breaking backbone
- damage removal: involves cutting out and replacing a damaged or inappropriate base or section of nucleotides
- damage tolerance: not truly repair but a way of coping with damage so that life can go on



© 2001 Encyclopædia Britannica, Inc.

2 Overview of human genetics

2.a How is the genetic information transmitted from generation to generation

Understanding heredity

- Pythagoras
- Empedocles
- Aristotle
- Harvey
- Leeuwenhoek
- de Maupertuis
- Darwin

- Mendel
- Morgan
- Crick & Watson
- McClintock

(http://www.pbs.org/wgbh/nova/genome)

Pythagoras (580-500 BC)

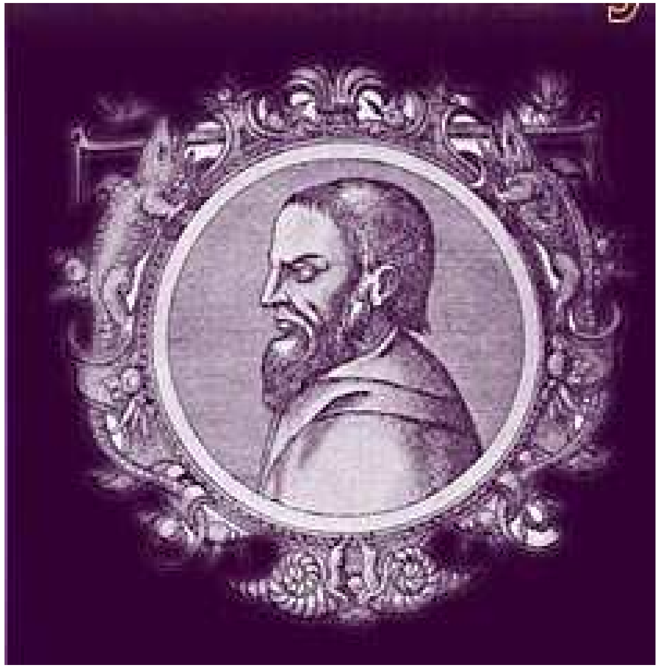
Pythagoras surmised that all hereditary material came from a child’s father. The mother provided only the location and nourishment for the fetus.

Semen was a cocktail of hereditary information, coursing through a man’s body and collecting fluids from every organ in its travels. This male fluid became the formative material of a child once a man deposited it inside a woman.

Aristotle (384-322 BC)

Aristotle’s understanding of heredity, clearly following from Pythagorean thought, held wide currency for almost 2,000 years.

The Greek philosopher correctly believed that both mother and father contribute biological material toward the creation of offspring, but he was mistakenly convinced that a child is the product of his or her parents’ commingled blood.

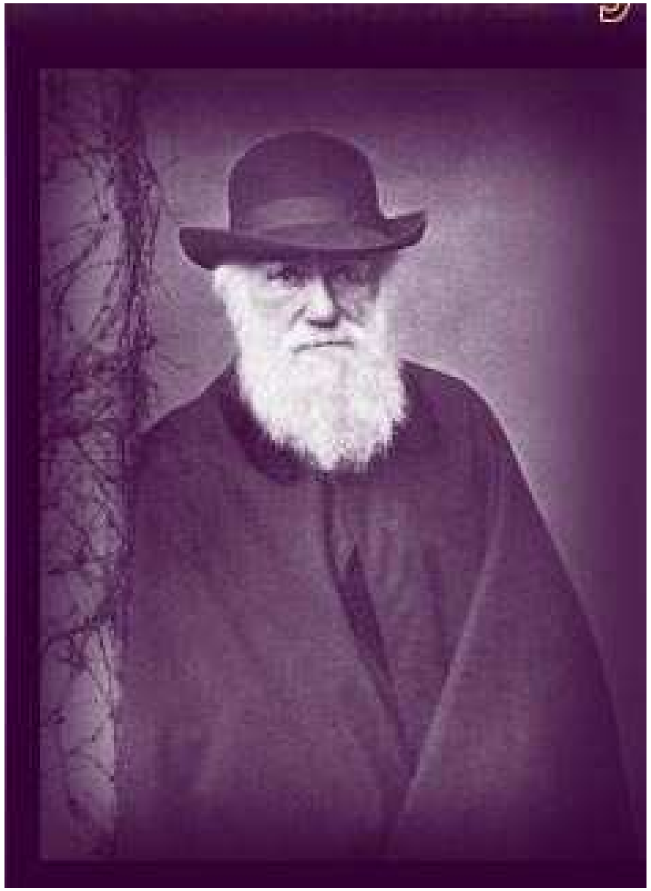


De Maupertuis (1698-1759)

In his 1751 book, Système de la nature (System of Nature), French mathematician, biologist, and astronomer Pierre-Louis Moreau de Maupertuis initiated the first speculations into the modern idea of dominant and recessive genes. De Maupertuis studied the occurrences of polydactyly (extra fingers) among several generations of one family and showed how this trait could be passed through both its male and female members.

Darwin (1809-1882)

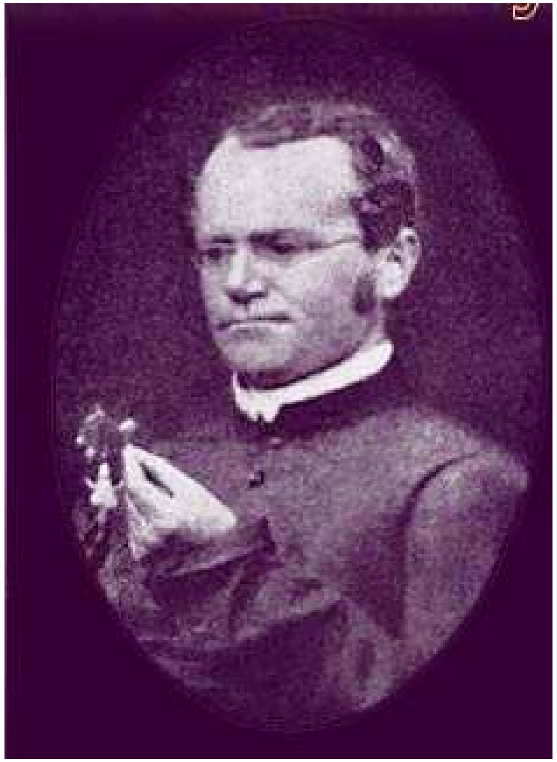
Darwin’s ideas of heredity revolved around his concept of "pangenesis." In pangenesis, small particles called pangenes, or gemmules, are produced in every organ and tissue of the body and flow through the bloodstream. The reproductive material of each individual formed from these pangenes was therefore passed on to one’s offspring.



Here we meet again … our friend Mendel (1822-1884)

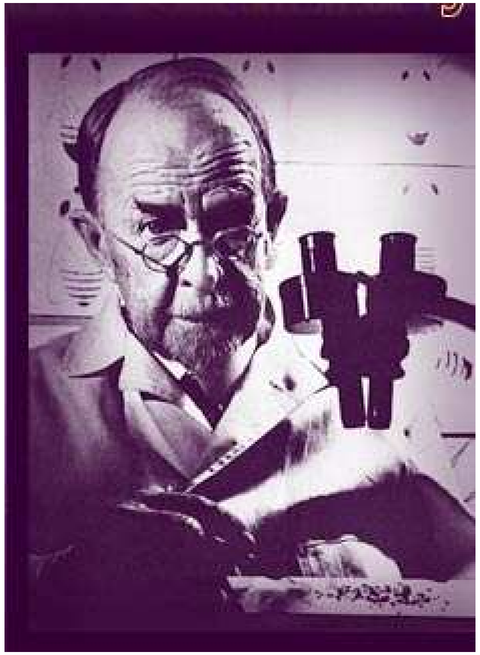
Gregor Mendel, an Austrian scientist who lived and conducted much of his most important research in a Czechoslovakian monastery, stablished the basis of modern genetic science. He experimented on pea plants in an effort to understand how a parent passed physical traits to its offspring. In one experiment, Mendel crossbred a pea plant with wrinkled seeds and a pea plant with smooth seeds.

All of the hybrid plants produced by this union had smooth seeds...



Morgan (1866-1945)

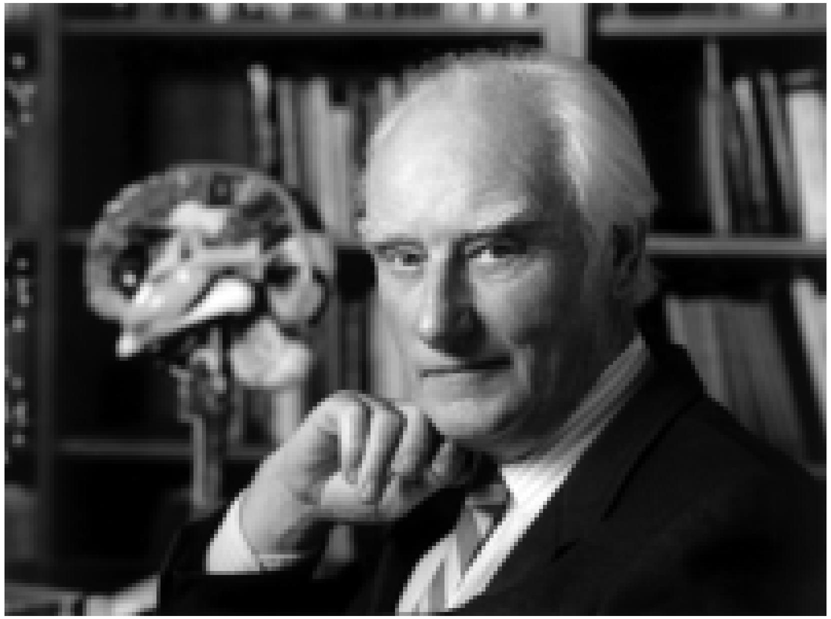
Thomas Hunt Morgan began experimenting with Drosophilia, the fruit fly, in 1908. He bred a single white-eyed male fly with a red-eyed female. All the offspring produced by this union, both male and female, had red eyes. From these and other results, Morgan established a theory of heredity that was based on the idea that genes, arranged on the chromosomes, carry hereditary factors that are expressed in different combinations when coupled with the genes of a mate.



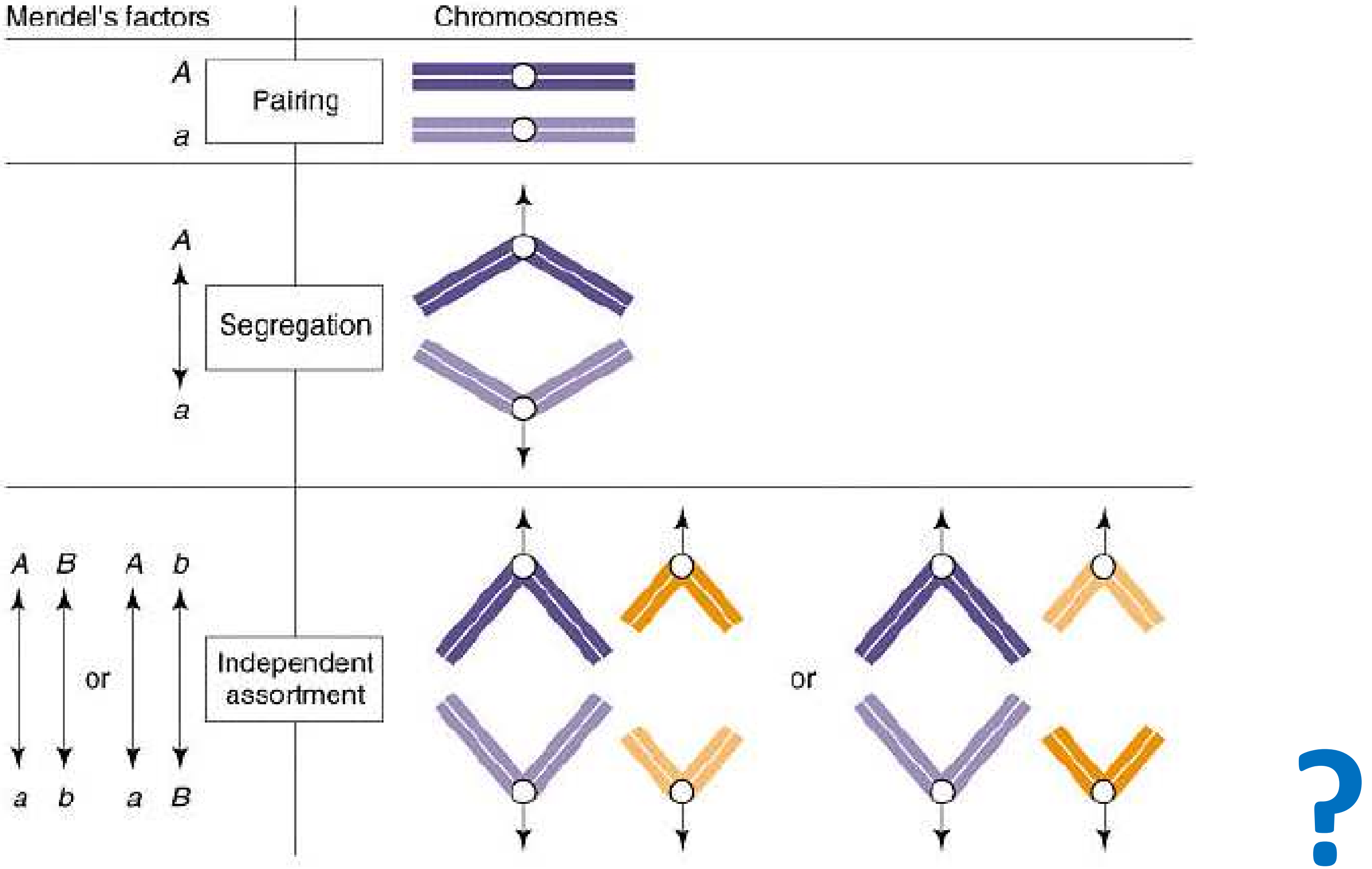
Crick (1916-2004) and

Watson (1928-)

Employing X-rays and molecular models, Watson and **Crick** discovered the double helix structure of DNA. Suddenly they could explain how the DNA molecule duplicates itself by forming a sister strand to complement each single, ladder-like DNA template.



Mendel hits the modern world: Chromosomes contain the units of heredity



Formal work definition of heredity

- Heredity is always linked to the trait under investigation:
  - The *phenotype* is the characteristic (e.g. hair color) that results from having a specific genotype ;
  - The *trait* is a coded (e.g. for actual statistical analysis) of the phenotype.
- The concept of "heritability" was introduced in order to measure the importance of genetics in relation to other factors in causing the variability of a trait in a population
  - What could these other factors be?

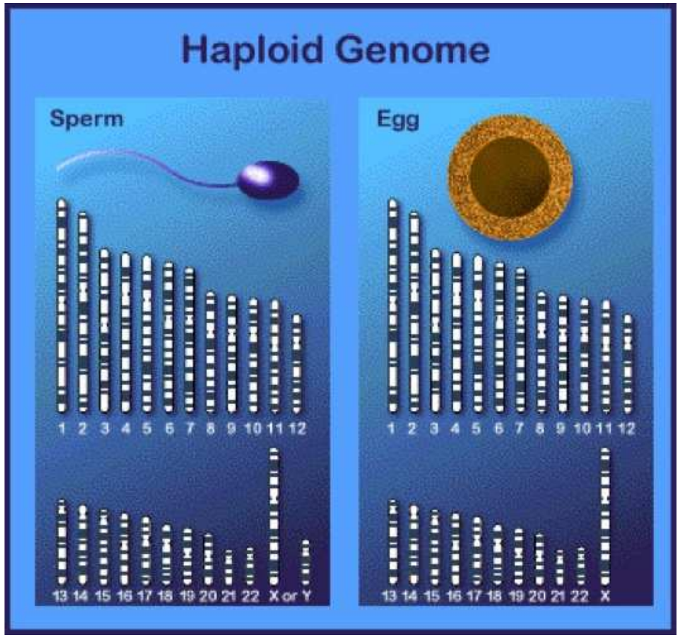
Formal work definition of heredity (continued)

- There are two main different measures for heredity:
    - *Broad heritability*:
      proportion of total phenotypic variance accounted for by all genetic components (coefficient of genetic determination)
    - *Narrow heritability*:
      proportion of phenotypic variance accounted for by the additive genetic component

- Popular study design to estimate heritability is the twins design.
    - Can you come up with reasons?

Genetic information is inherited via meiosis

- Paternal genes (via sperm) and maternal genes (via egg) are donated to offspring
- Yet, parents won’t lose genetic information, nor offspring will have too much genetic information



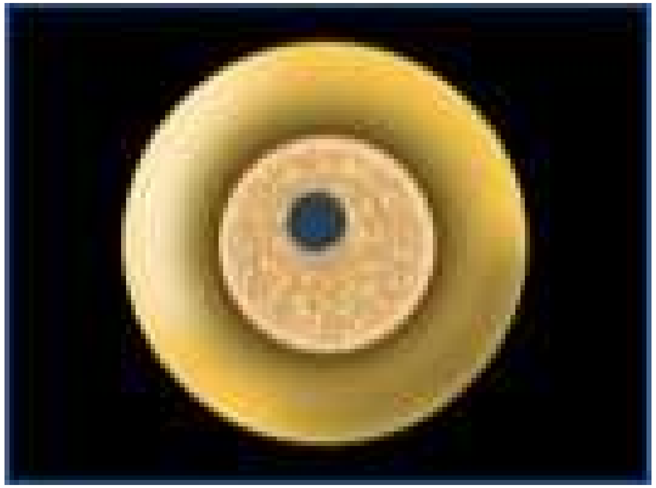
(Roche Genetics)

Meiosis in detail

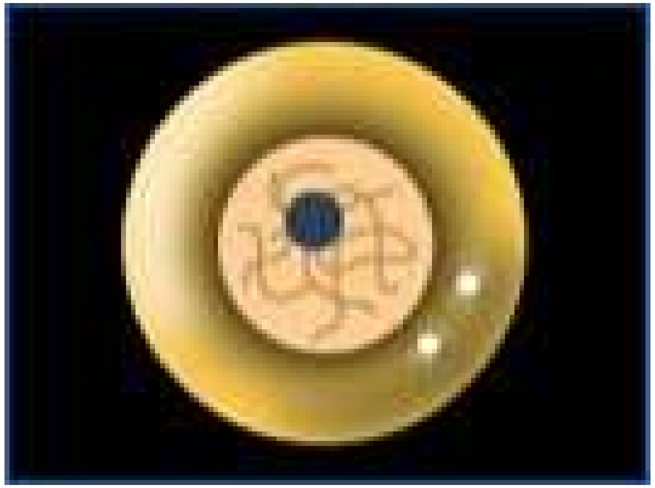
- Meiosis is a process to convert a diploid cell to a haploid gamete, and causes a change in the genetic information to increase diversity in the offspring.
- In particular, meiosis refers to the processes of cell division with two phases resulting in four haploid cells (gametes) from a diploid cell. In meiosis I, the already doubled chromosome number reduces to half to create two diploid cells each containing one set of replicated chromosomes. Genetic recombination between homologous chromosome pairs occurs during meiosis I. In meiosis II, each diploid cell creates two haploid cells resulting in four gametes from one diploid cell (mitosis).

- Check out a nice demo to differentiate meiosis from mitosis: http://www.pbs.org/wgbh/nova/miracle/divide.html

Meiosis in detail

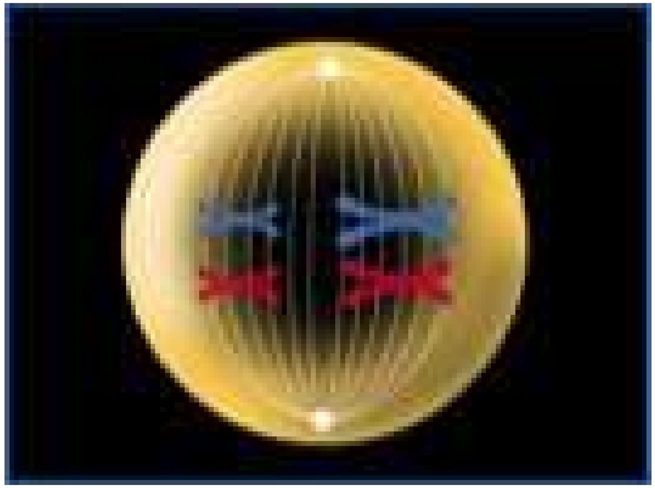


1



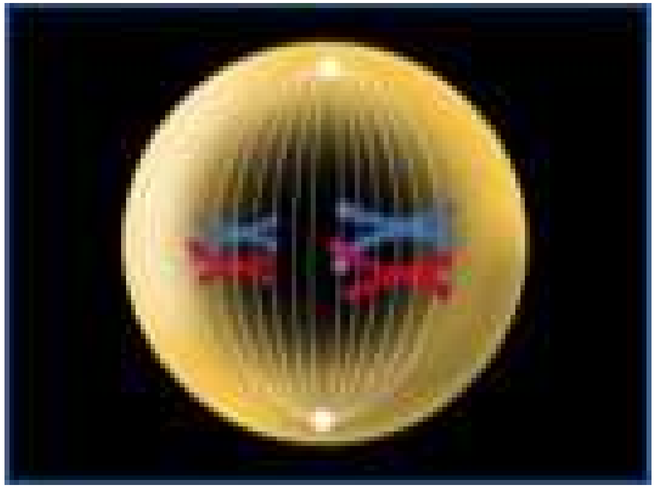
2

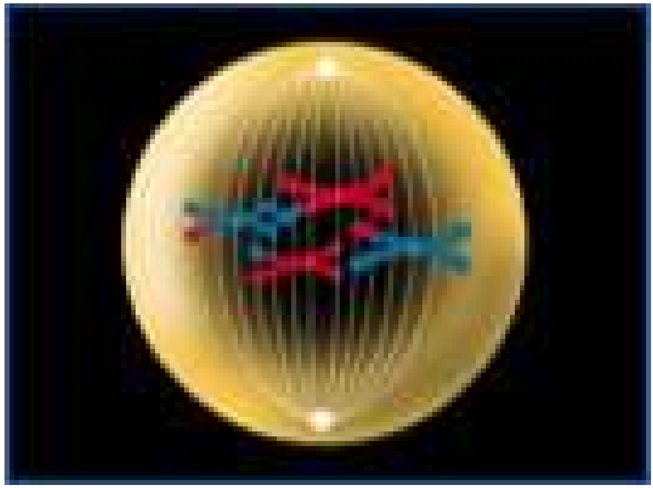
3

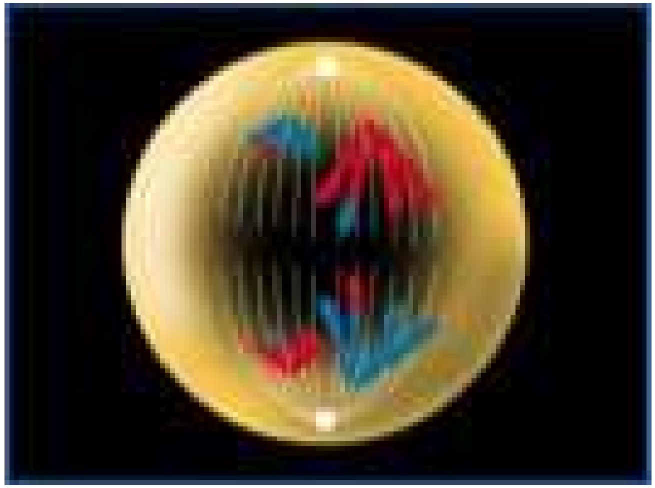


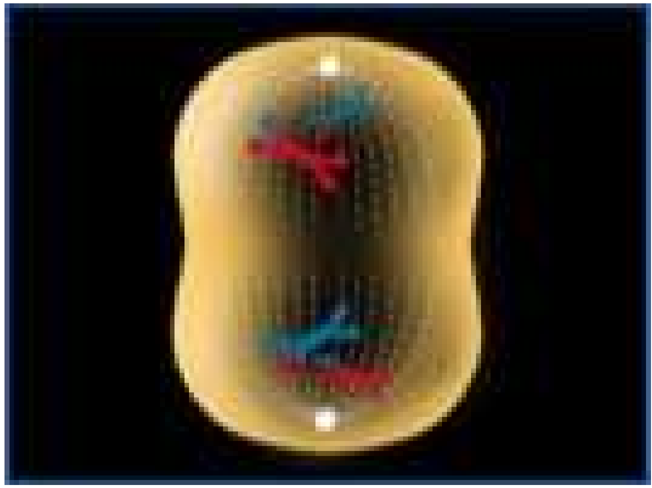
4

Meiosis in detail

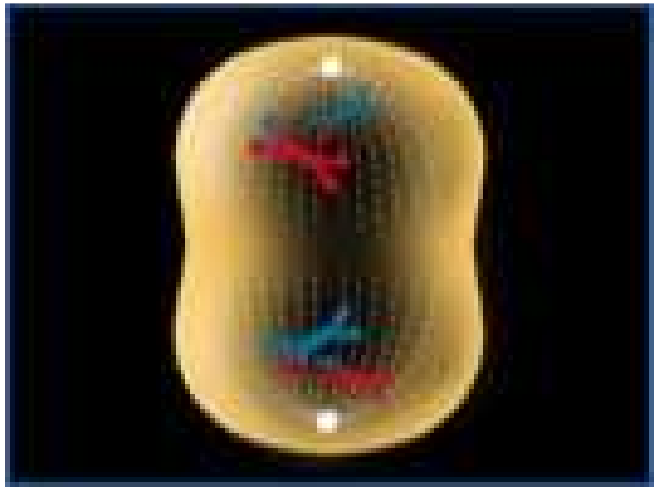

5


6


7

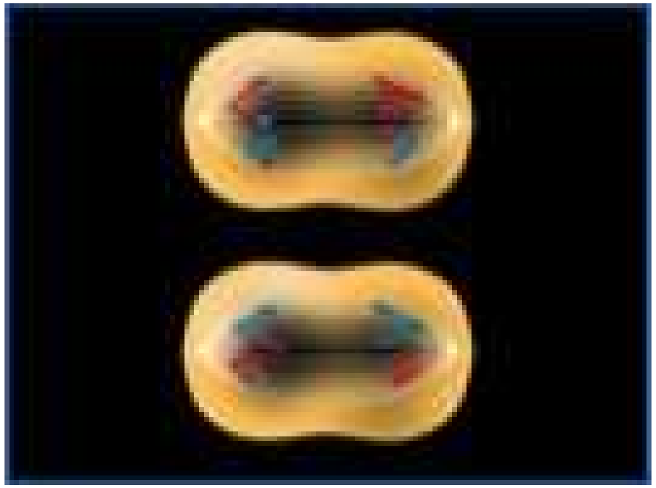

8

Meiosis in detail

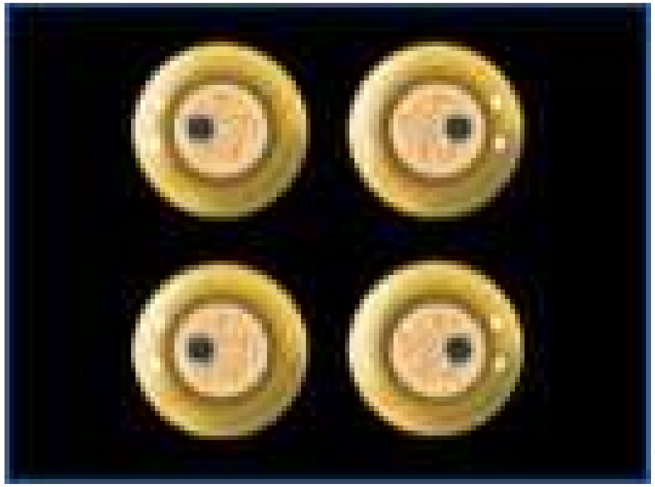


8

9



10



11

Recombination introduces extra variation

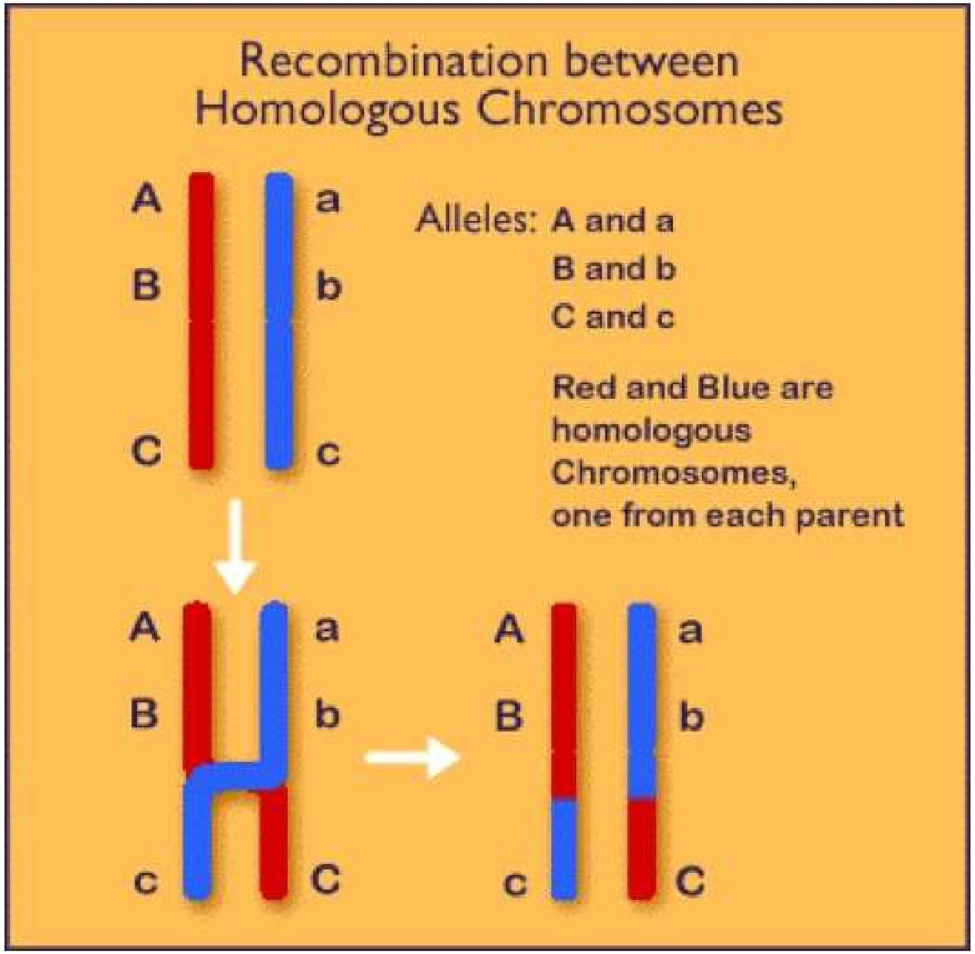
- A collection of linked loci (loci that tend to be inherited together) is called a *haplotype*
- Immediately before the cell division that leads to gametes, parts of the homologous chromosomes may be exchanged

  An individual with haplotypes A-B and a-b may produce gametes

  A-B and a-b or A-b and a-B. This process is called *recombination*.

- The probability of recombination during meiosis is termed the *recombination fraction*, and is usually denoted by ϑ**.**
  - What are the extreme values of the recombination fraction?

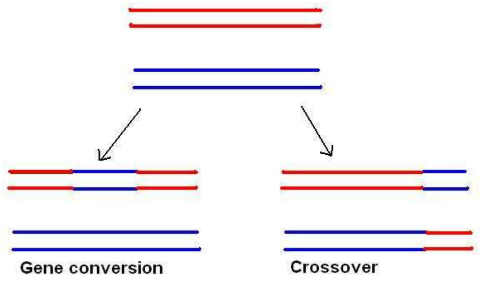
Recombination and haplotypes


Recombination between Homologous Chromosomes
A
B
C
a
b
c
Alleles: A and a
B and b
C and c
Red and Blue are homologous Chromosomes, one from each parent
A
B
c
a
b
C
A
B
c
a
b
C

(Roche genetics)

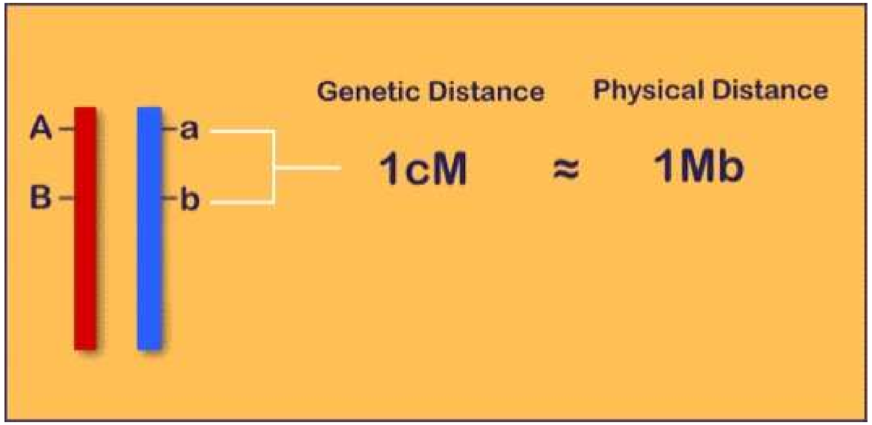
Recombination is different from gene conversion

- What has been described historically, and above, as recombination should, more properly, be called cross-over (i.e. the process by which two chromosomes pair up and exchange sections of their DNA; recombination refers to the result of such a process, namely genetic recombination).
- Although cross-over is indeed caused by breaking and rejoining of chromosomes, they more often rejoin nearly the same way around.
- Often a short segment of DNA (< 50 base pairs) is exchanged, where one double helix remains unaltered but the other has changed. This is called *gene conversion*:



Recombination is related to genetic distance

- The greater the physical distance between two loci, the more likely it is that there will be recombination.
- This forms the basis of mapping strategies such as linkage and association.
- So recombination is related to "distance" D. In a way, it forms a bridge between "physical distance" and "genetic distance"



(Roche Genetics)

Genetic distance (continued)

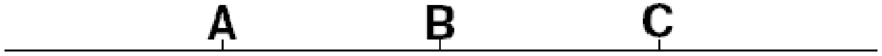
- In general, a genetic map function M(D) = ϑ provides a mapping from the additive genetic distance D to the non-additive recombination fraction ϑ between a given pair of loci, where the recombination fraction ϑ is, as before, the proportion of gametes that are recombinant between the two loci.
- Genetic map functions are needed because in most experiments all we can directly observe are the recombination events.
- However, since a recombination event is only observed if there are an odd number of crossovers between the two loci, recombination fractions are not additive.
- One of the most widely used map functions is *Haldane's map function*, and has been in widespread use since 1919.

Genetic distance (continued)

- Several models exist for recombination rates, but the "constant recombination rate" model is the simplest:
    - A simplified model is that loci can be arranged along a line in such a way that, with each meiosis, recombinations occur at a constant rate.

    - In the simplest setting, the relationship between the recombination frequency and the genetic distance is then given by Haldane's map function as follows:

$$D_{\mathrm{AB}} = -\frac{1}{2}\log_e(1-2\theta_{\mathrm{AB}})$$



$$D_{\mathrm{AC}} = D_{\mathrm{AB}} + D_{\mathrm{BC}}$$

Genetic distance (continued)

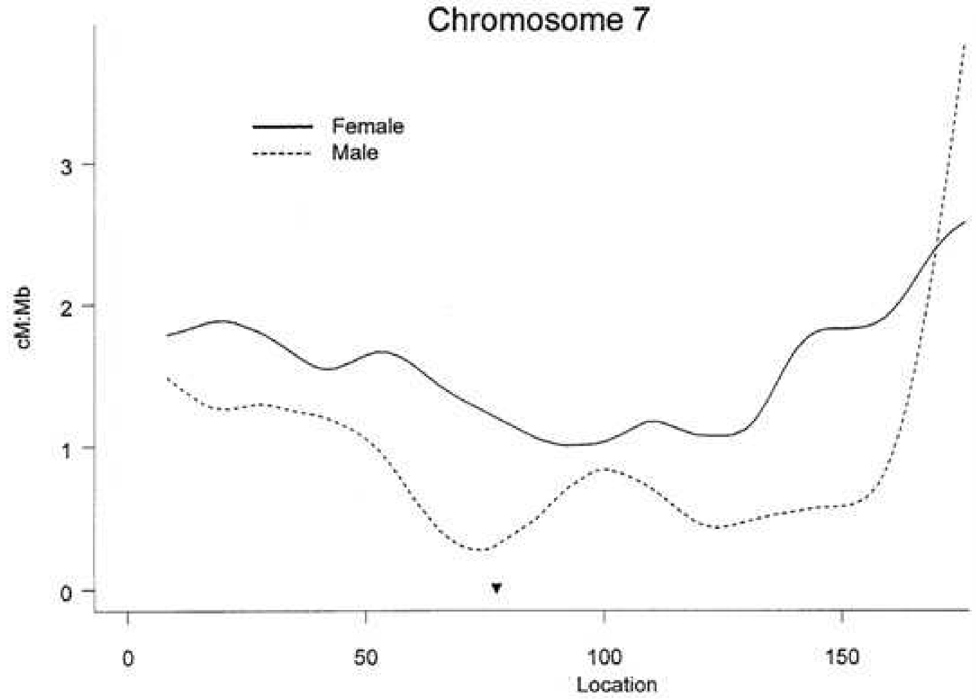
- In practice, real-life is more complicated, due to settings for which the model of independence of recombinations does not fit
    - Under the *Kosambi map function*, complete *interference* is assumed for small map distances and a decreasing amount of interference accompanies increasing distances.
    - *Hot spots* cause uneven relationship between physical and genetic distances

Genetic distance (continued)

- The unit of genetic distance D_{AB} is called a *Morgan*.
    - At each meiosis the expected number of recombinations is one per Morgan (definition)
- An extra real-life complication is that recombination appears to be more frequent in females than in males:
    - Total female map length: 44 Morgans
    - Total male map length: 27 Morgans
    - Total sex-averaged map length: 33 Morgans
- On average, 1 cM corresponds to about 10^6 bases.
    - The total length of the human genome is “on average” 33 Morgans ($\approx 3 \times 10^9$ bases)

Sex differences in cross-over events

- Plot of sex-specific genetic distance to physical distance ratio (in cM/Mb) against genetic location.

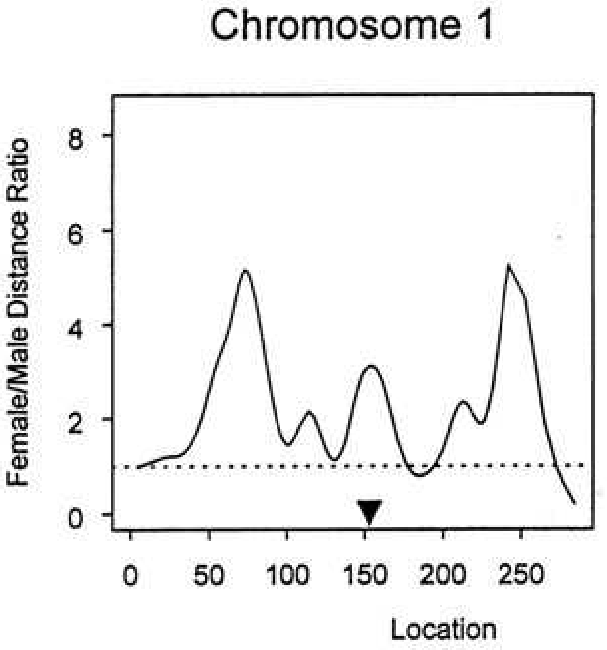


The full line was obtained by use of female genetic distance; the dashed line was obtained by use of male genetic distance. Triangle: approximate location of the centromere.

(Broman et al, *AJHG*, 1998)

Sex differences in cross-over events

- At the telomeres of nearly all chromosomes, the female:male genetic-distance ratio approaches and often dips below 1, so that males exhibit equal or greater recombination rates in the telomeric regions.



(Broman et al, *AJHG*, 1998)

2.b How do individuals/animals/plants differ with regard to their genetic variation?

Variation in chromosome numbers between species

Diploid numbers

- All animals have a characteristic number of chromosomes in their body cells called the diploid (or 2n) number.
- These occur as homologous pairs, one member of each pair having been acquired from the gamete of one of the two parents of the individual whose cells are being examined.
- The gametes contain the haploid number (n) of chromosomes.

Diploid numbers of commonly studied organisms

Homo sapiens (human)	46
Mus musculus (house mouse)	40
Drosophila melanogaster (fruit fly)	8
Caenorhabditis elegans (microscopic roundworm)	12
Saccharomyces cerevisiae (budding yeast)	32
Arabidopsis thaliana (plant in the mustard family)	10
Xenopus laevis (South African clawed frog)	36
Canis familiaris (domestic dog)	78

Gallus gallus (chicken)	78
Zea mays (corn or maize)	20
Muntiacus reevesi (the Chinese muntjac, a deer)	23
Muntiacus muntjac (its Indian cousin)	6
Myrmecia pilosula (an ant)	2
Parascaris equorum var. univalens (parasitic roundworm)	2
Cambarus clarkii (a crayfish)	200
Equisetum arvense (field horsetail ; a plant)	216

Haploid, haplotypes and phase

- Phase refers to the haplotypic configuration of linked loci.
- The diplotype U1U3–V1V2 is consistent with two possible phases: (1) U1–V1 on one chromosome and U3–V2 on the other; or (2) U1–V2 on one chromosome and U3–V1 on the other.
- If a child receives U1–V1 on a paternally derived chromosome from a father with diplotype U1U3–V1V2, it either implies that the father was in phase (1) and no recombination has occurred, or he was in phase (2) and there has been recombination.
- This concept is extremely important in genetic linkage and association studies (see later)
- Variation in phase is related to variation at composite loci

Variation at genetic loci

What is a locus?

- A locus is a unique chromosomal location defining the position of an individual gene or DNA sequence.
    - Hence, it does not necessarily refer to one particular base-pair position!
- In genetic linkage studies, the term can also refer to a larger region involving several genes, perhaps even including non-coding parts of the DNA.

What is an allele?

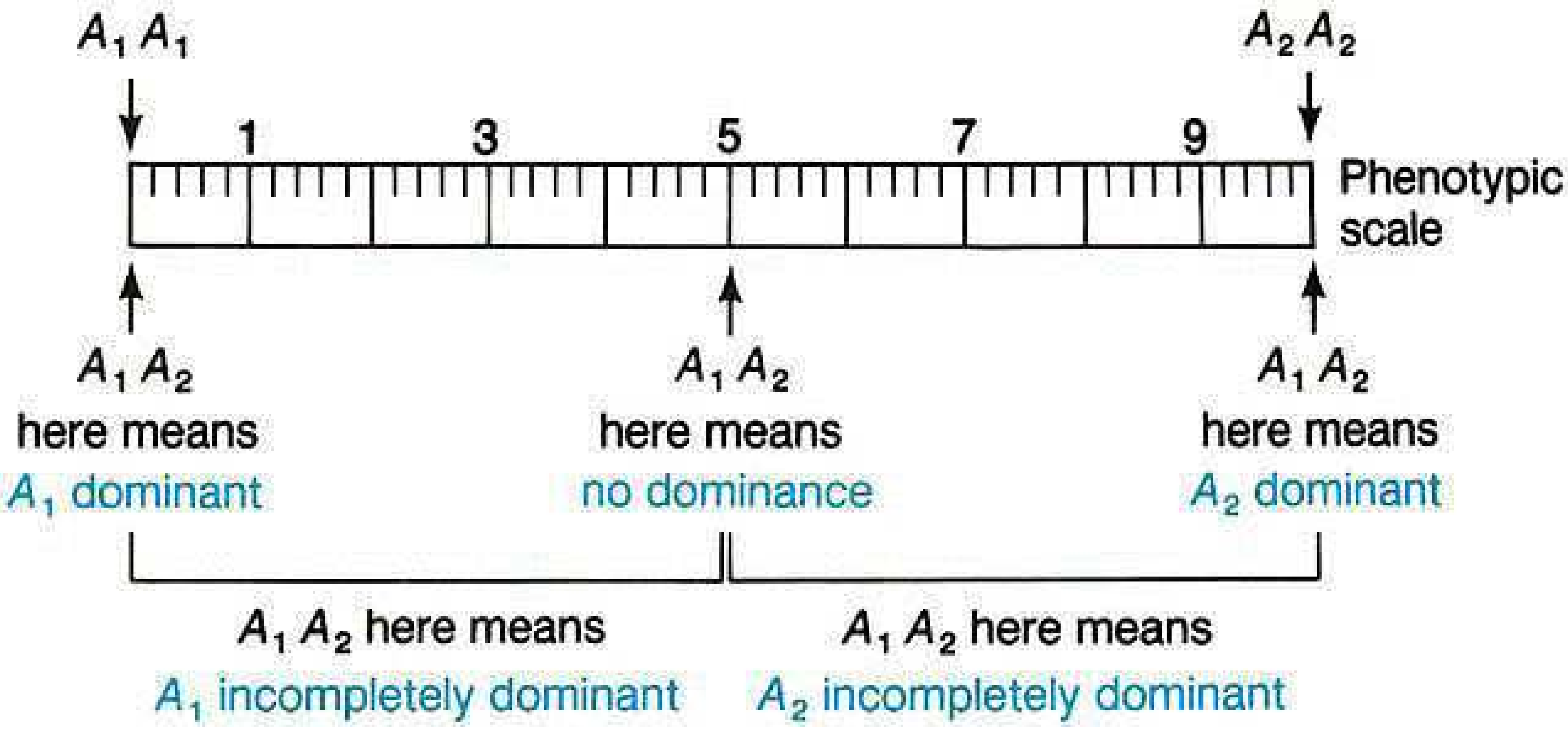
- Because human cells are diploid, there are two alleles at each genetic locus
- This pair of alleles is called the individual's *genotype* at that locus
- If the 2 alleles are the same, the individual is said to be *homozygous* at the locus. If they are different, he/she is said to be *heterozygous* at the locus
- The heterozygosity of a marker is defined as the probability that two alleles chosen at random are different. If π is the (relative) frequency of the *i*-th allele, then heterozygosity can be expressed as:

$$\text{Heterozygosity} = 1 - \sum_i \pi_i^2$$

Associating alleles to traits?

- If a single copy of an allele results in the same phenotype as two copies irrespective of the second allele, the allele is said to be dominant over the second allele
- Likewise, an allele which must occur in both copies of the gene to yield the phenotype is termed recessive
- Alleles which correspond to mutations which destroy the coding of a protein tend to be recessive
- If the phenotype for genotype *i/j* is intermediate between the phenotypes for *i/i* and *j/j*, the alleles *i* and *j* are co-dominant

Associating alleles to traits?


$A_1 A_1$
$A_2 A_2$
1
3
5
7
9
Phenotypic scale
$A_1 A_2$ here means A_1 dominant
$A_1 A_2$ here means no dominance
$A_1 A_2$ here means A_2 dominant
$A_1 A_2$ here means A_1 incompletely dominant
$A_1 A_2$ here means A_2 incompletely dominant

Associating alleles to traits?

- Recall: The *phenotype* is the characteristic (e.g. hair color) that results from having a specific genotype
- Often we require probability models to describe phenotypic expression of genotypes. Probabilities of phenotype conditional upon genotype are called *penetrances*
- In many cases, the same phenotype can result from a variety of different genotypes (sometimes termed *phenocopies*)
- Equally, the same gene may have several different phenotypic manifestations. This phenomenon is called *pleiotropy*.

Using proxies to capture genetic variation at loci

- Framework maps of the chromosomes are actually built using polymorphic *markers*. These may or may not have any function at all.
- A marker is *polymorphic* if it can exist in different forms (meaning, with slightly different sequences). The different forms are called *alleles*. Some polymorphic markers may have 20 or more distinct alleles
- Random mutations within the marker sequence may lead to a new allele or to the conversion of one allele into another (see before)

Distinguish between polymorphisms and mutations

- The verb mutation describes the process by which new variants of a gene arise. As a noun it is used to describe a rare variant of a gene.
- Polymorphisms are more common variants (more than 1%).
- Most mutations will disappear but some will achieve higher frequencies due either to random genetic drift or to selective pressure
- The most common forms of variants are:
    - repeated sequences of 2, 3 or 4 nucleotides (microsatellites)
    - single nucleotide polymorphisms (SNPs) in which one letter of the code is altered

Non-synonymous SNP

- A SNP that alters the DNA sequence in a coding region such that the amino acid coding is changed.
- The new code specifies an alternative amino acid or changes the code for an amino acid to that for a stop translation signal or vice versa.
- Non-synonymous SNPs are sometimes referred to as coding SNPs.

Synonymous SNP

- Synonymous SNPs alter the DNA sequence but do not change the protein coding sequence as interpreted at translation, because of redundancy in the genetic code.
- Exonic SNPs may or may not cause an amino acid change

2.c How to detect individual differences?

- Based on the previous, one obvious way to detect individual differences is by studying differences in an individual’s DNA sequence.
- Hence,
    - How to sequence?
    - How to study sequences? (Chapter 4)
    - How to compare multiple sequences? (Chapter 5)

The sequencing reaction

- The purpose of sequencing is to determine the order of the nucleotides of a gene.
- For sequencing, we mostly start from smaller fragments (PCR fragments; see later) or cloned genes.
- There are three major steps in a sequencing reaction (like in PCR), which are repeated for 30 or 40 cycles.
  - Step 1: Denaturation at 94°C :
    During the denaturation, the double strand melts open to single stranded DNA, all enzymatic reactions stop (for example : the extension from a previous cycle).

(http://users.ugent.be/~avierstr/principles/seq.html)

The sequencing reaction

- Step 2: Annealing at 50°C :

  In sequencing reactions, only one primer is used, so there is only one strand copied (in PCR : two primers are used, so two strands are copied). Ionic bonds are constantly formed and broken between the single stranded primer and the single stranded template. The more stable bonds last a little bit longer (primers that fit exactly) and on that little piece of double stranded DNA (template and primer), the polymerase can attach and starts copying the template. Once there are a few bases built in, the ionic bond is so strong between the template and the primer, that it does not break anymore.

(http://users.ugent.be/~avierstr/principles/seq.html)

The sequencing reaction

- Step 3: extension at 60°C:
  This is the ideal working temperature for the polymerase (normally it is 72 °C, but because it has to incorporate ddNTP's which are chemically modified dNTPs (deoxynucleotide triphosphates: the free nucleotide bases used for DNA strand growing) with a fluorescent label, the temperature is lowered so it has time to incorporate the 'strange' molecules).
  The primers, where there are a few bases built in, already have a stronger ionic attraction to the template than the forces breaking these attractions.

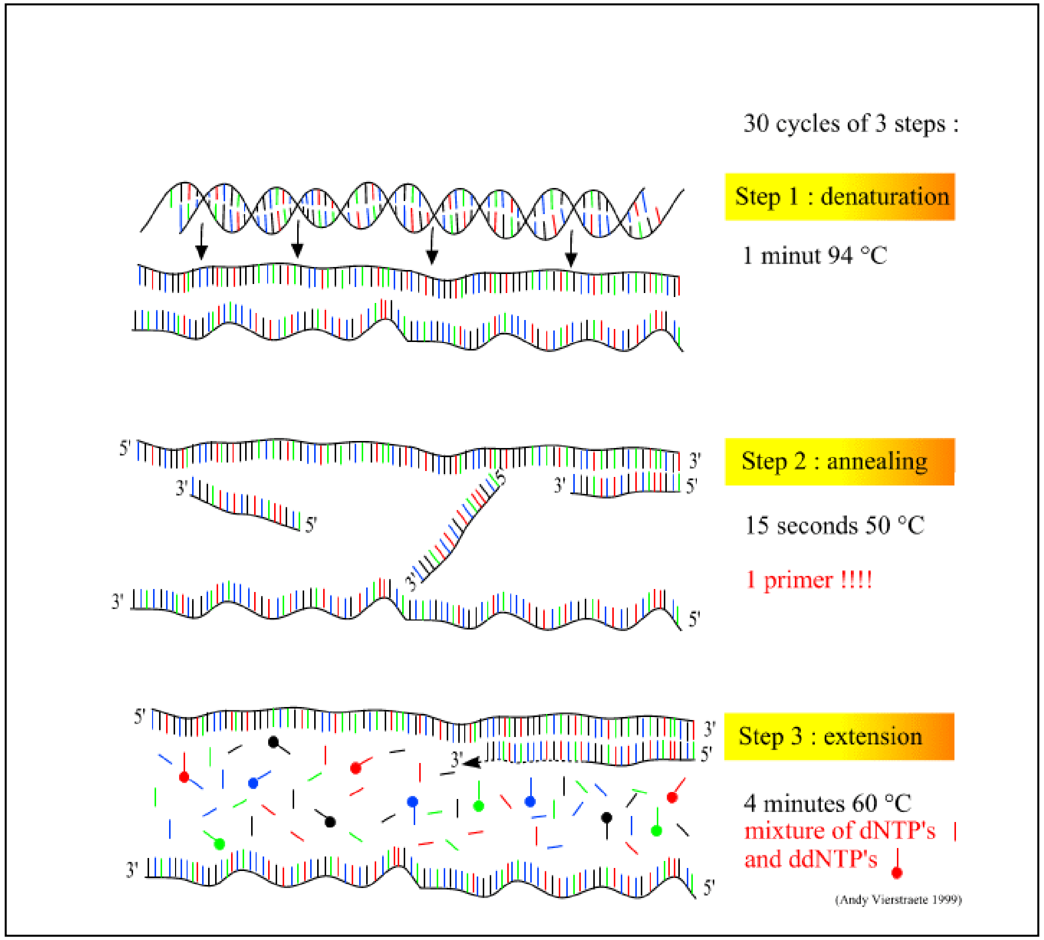
(http://users.ugent.be/~avierstr/principles/seq.html)

The sequencing reaction

- Step 3: extension at 60°C:
  Primers that are on positions with no exact match, come loose again and don't give an extension of the fragment.
  The bases (complementary to the template) are coupled to the primer on the 3'side (adding dNTP's or ddNTP's from 5' to 3)'.
  When a ddNTP is incorporated, the extension reaction stops because a ddNTP contains a H-atom on the 3rd carbon atom. Since the ddNTP's are fluorescently labeled, it is possible to detect the color of the last base of this fragment on an automated sequencer.

(http://users.ugent.be/~avierstr/principles/seq.html)

The sequencing reaction

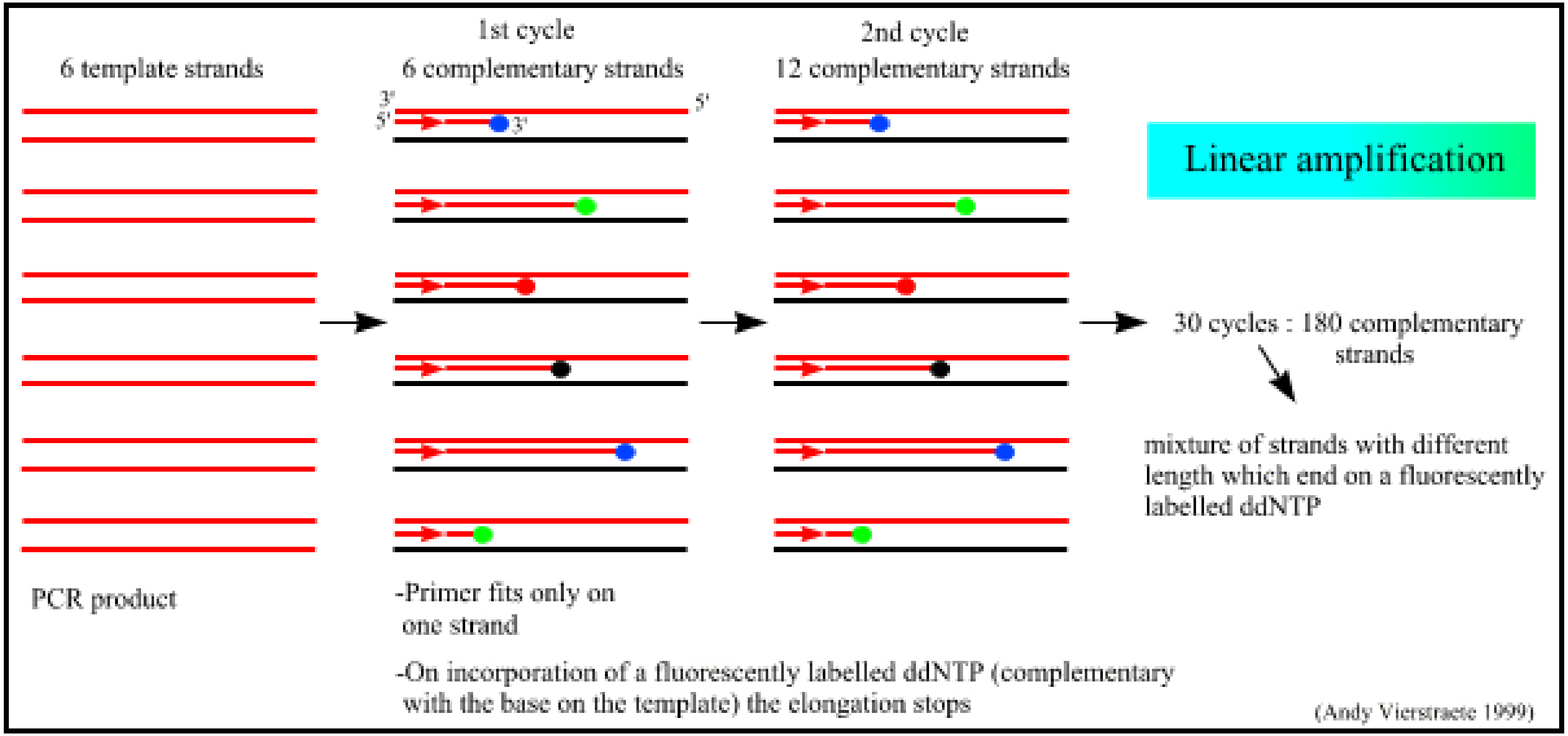


(http://users.ugent.be/~avierstr/principles/seq.html)

The sequencing reaction

- Because only one primer is used, only one strand is copied during sequencing, there is a *linear* increase of the number of copies of one strand of the gene.
- Therefore, there has to be a large amount of copies of the gene in the starting mixture for sequencing.
- Suppose there are 1000 copies of the wanted gene before the cycling starts,
  - after one cycle, there will be 2000 copies: the 1000 original templates and 1000 complementary strands with each one fluorescent label on the last base,
  - after two cycles, there will be 2000 complementary strands,
  - three cycles will result in 3000 complementary strands and so on.

The sequencing reaction


1st cycle
2nd cycle
6 template strands
6 complementary strands
12 complementary strands
3'
5'
5'
3'
Linear amplification
30 cycles : 180 complementary strands
mixture of strands with different length which end on a fluorescently labelled ddNTP
PCR product
-Primer fits only on one strand
-On incorporation of a fluorescently labelled ddNTP (complementary with the base on the template) the elongation stops
(Andy Vierstraete 1999)

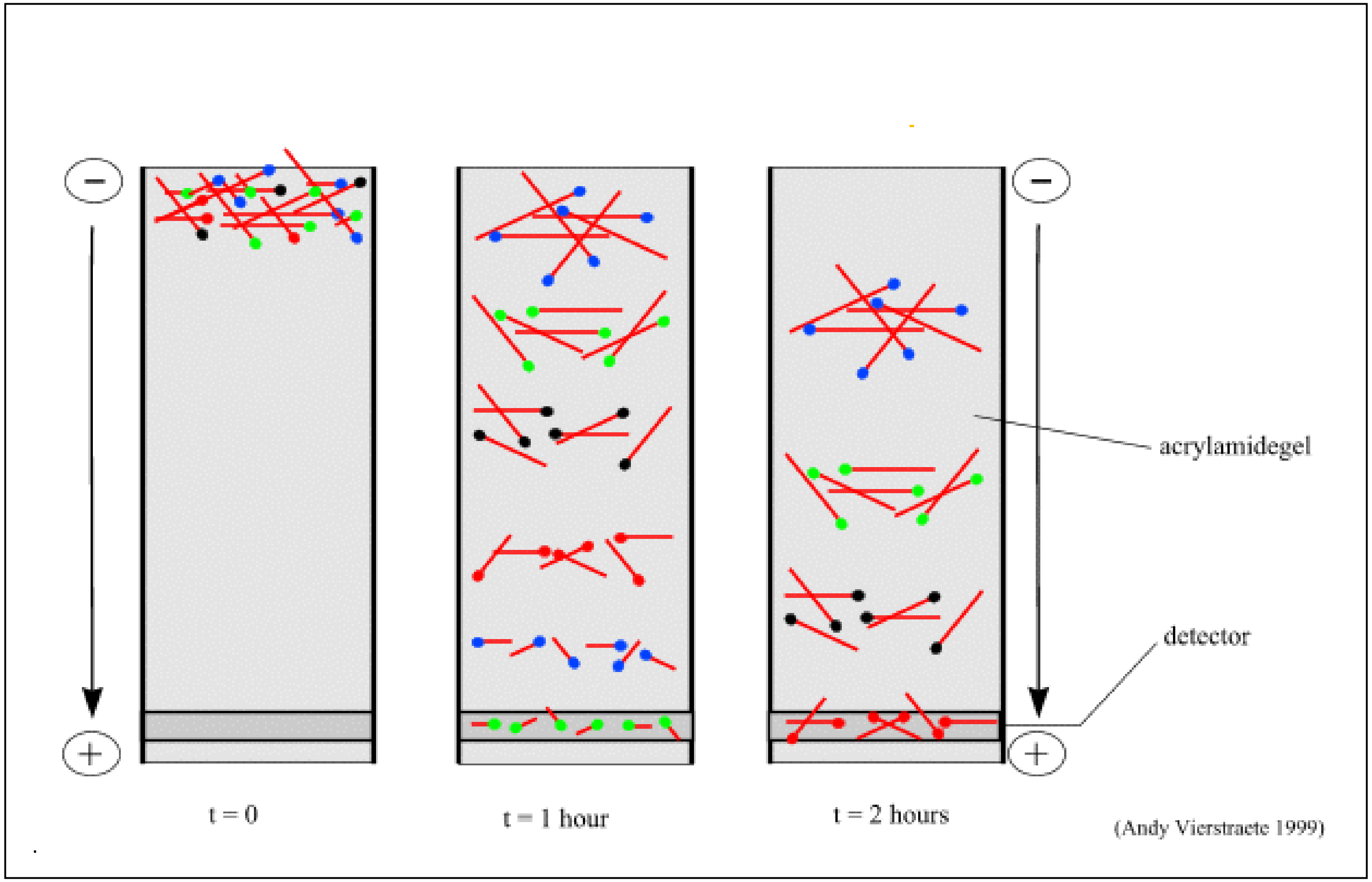
(http://users.ugent.be/~avierstr/principles/seq.html)

Separation of the molecules

- After the sequencing reactions, the mixture of strands, all of different length and all ending on a fluorescently labelled ddNTP have to be separated;
- This is done on an acrylamide gel, which is capable of separating a molecule of 30 bases from one of 31 bases, but also a molecule of 750 bases from one of 751 bases. The separation is done with gel electrophoresis.
- DNA has a negative charge and migrates to the positive side. Smaller fragments migrate faster, so the DNA molecules are separated on their size.

(http://users.ugent.be/~avierstr/principles/seq.html)

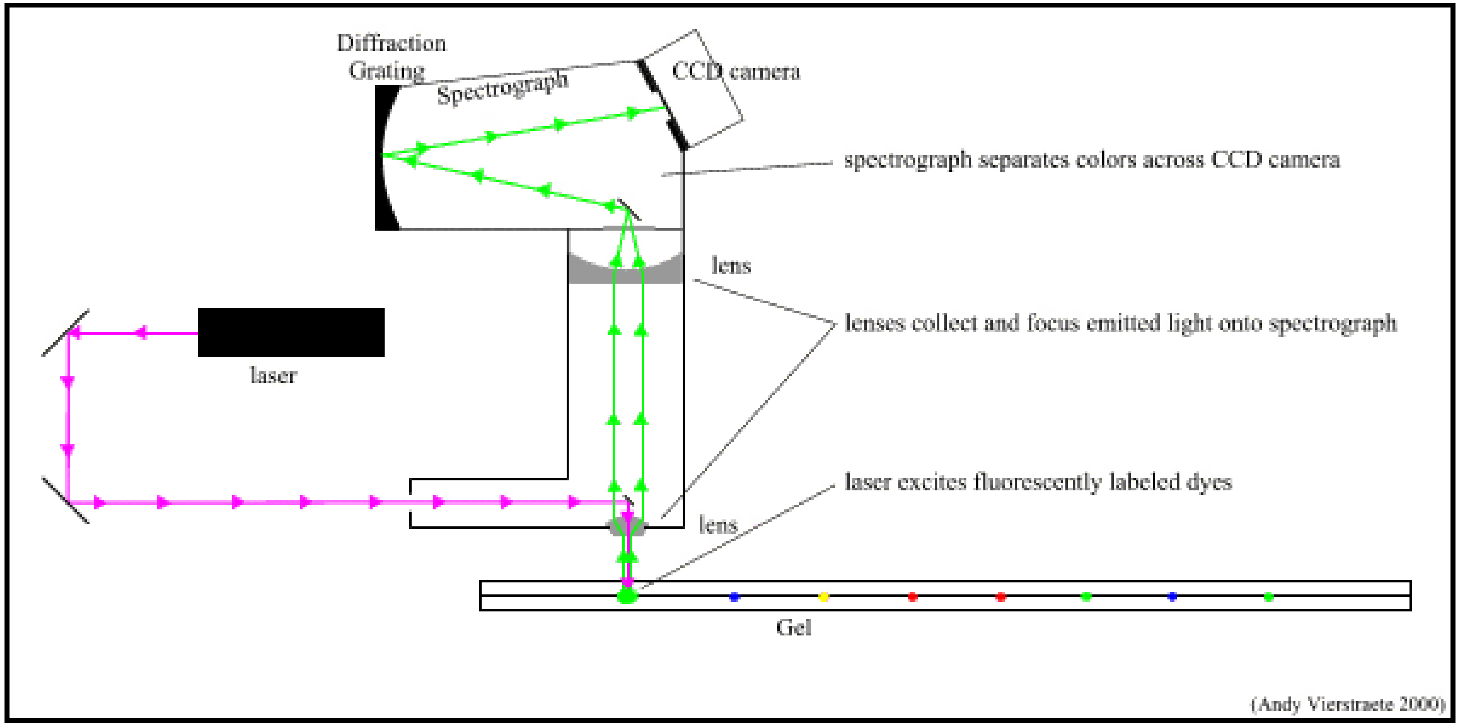
Separation of the molecules via gel electrophoresis



(http://users.ugent.be/~avierstr/principles/seq.html)

Detection on an automated sequencer

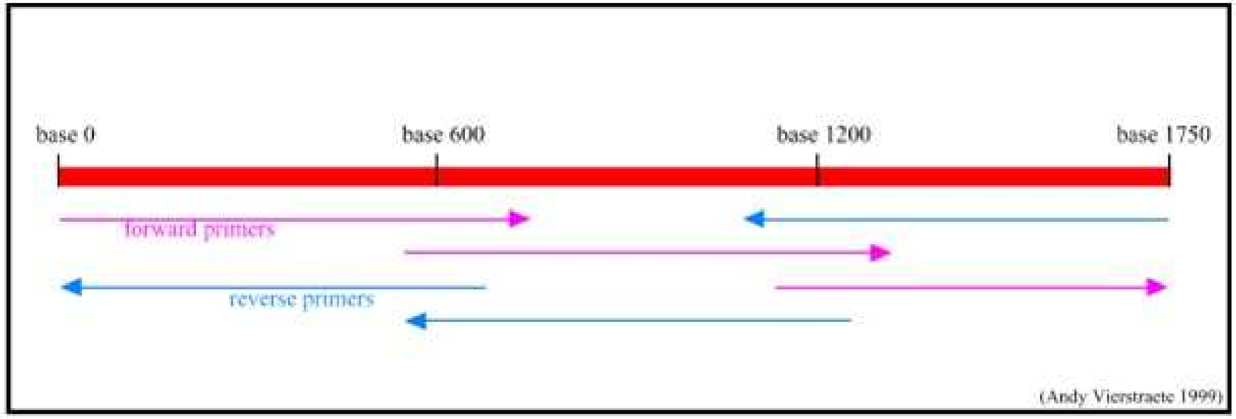
- For a ABI Prism 377 sequencer, the fluorescently labelled fragments that migrate through the gel, are passing a laser beam at the bottom of the gel. The laser exites the fluorescent molecule, which sends out light of a distinct color.


Diffraction Grating
Spectrograph
CCD camera
spectrograph separates colors across CCD camera
lens
lenses collect and focus emitted light onto spectrograph
laser
laser excites fluorescently labeled dyes
lens
Gel

(Andy Vierstraete 2000)

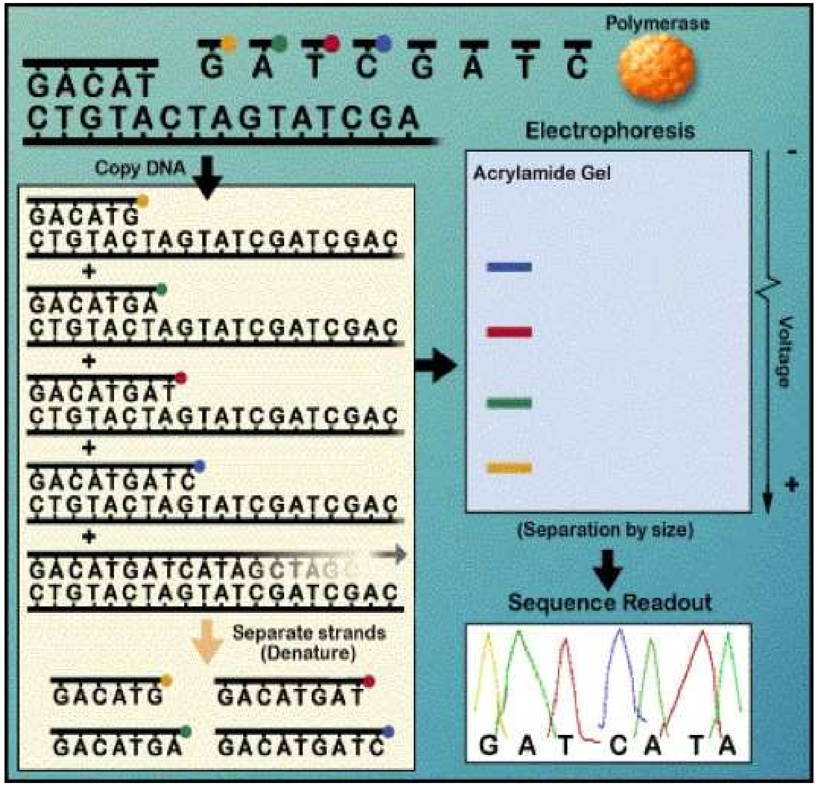
Assembling the sequences parts of a gene

- For publication purposes, each sequence of a gene has to be confirmed in both directions.
- To accomplish this, the gene has to be sequenced with forward and reverse primers.
- Since it is only possible to sequence a part of 750 till 800 bases in one run, a gene of, for example 1800 bases, has to be sequenced with internal primers. When all these fragments are sequenced, a computer program tries to fit the different parts together and assembles the total gene sequence.



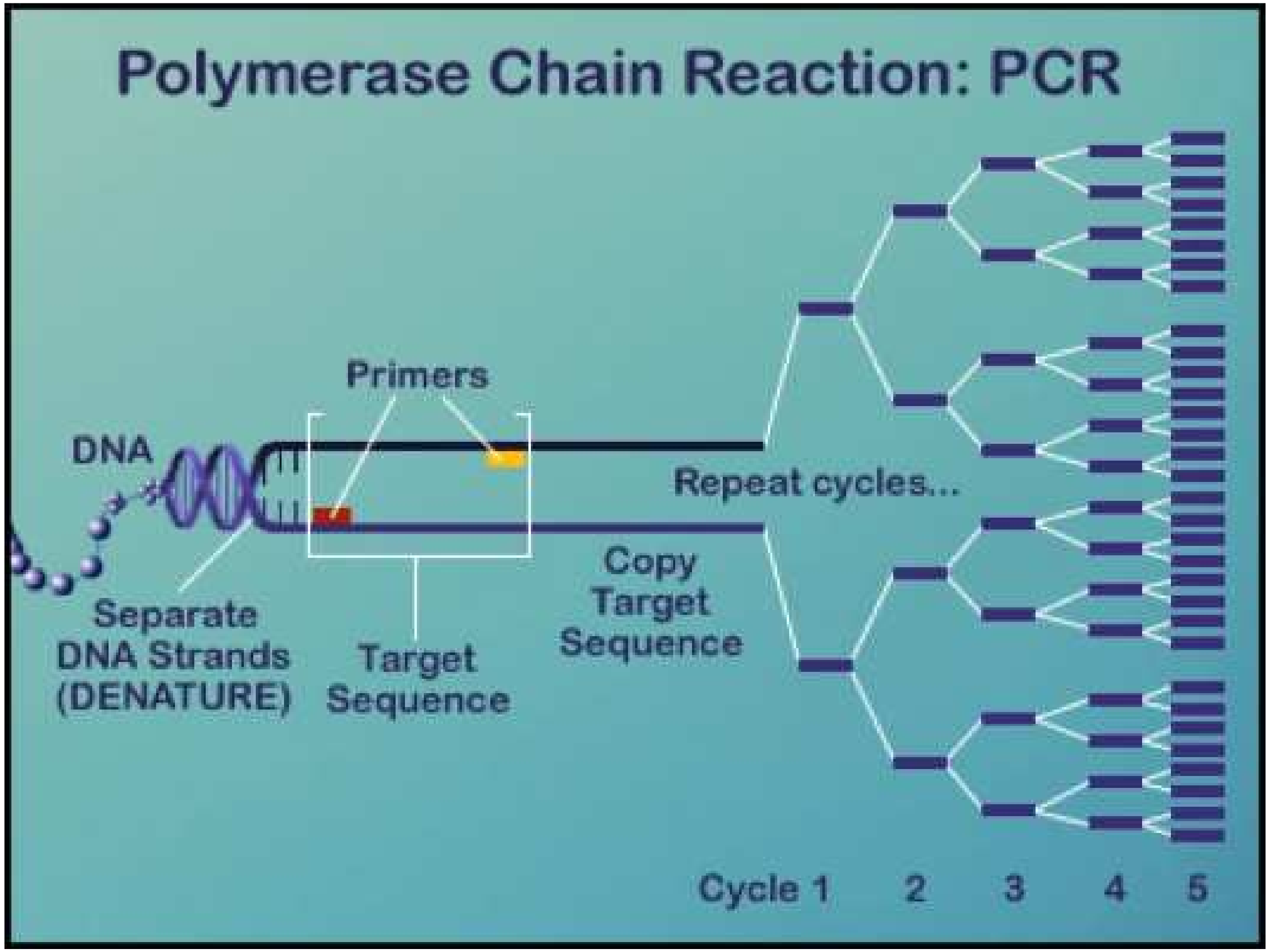
(Andy Vierstraete 1999)

Sequencing of DNA



- The result is an electropherogram showing the fluorescence units over time and fragment positions.

Polymerase chain reaction (PCR): In vitro DNA replication



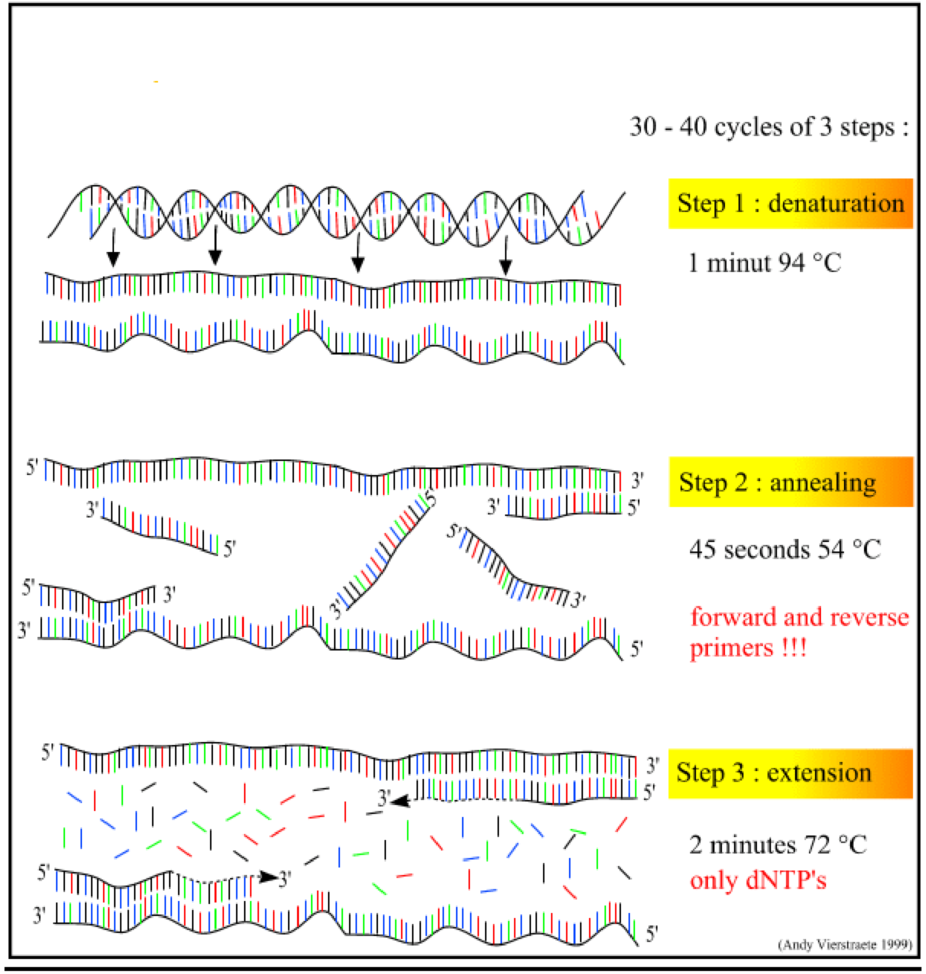
(Roche Genetics)

The cycling reactions of a PCR

- The purpose of a PCR (Polymerase Chain Reaction) is to make a huge number of copies of a gene. This is necessary to have enough starting template for sequencing
- There are three major steps in a PCR, which are repeated for 30 or 40 cycles. This is done on an automated cycler, which can heat and cool the tubes with the reaction mixture in a very short time.
  - Step 1: Denaturation at 94°C
  - Step 2: Annealing at 54°C
  - Step 3: extension at 72°C

(http://users.ugent.be/~avierstr/principles/pcr.html)

The cycling reactions of a PCR



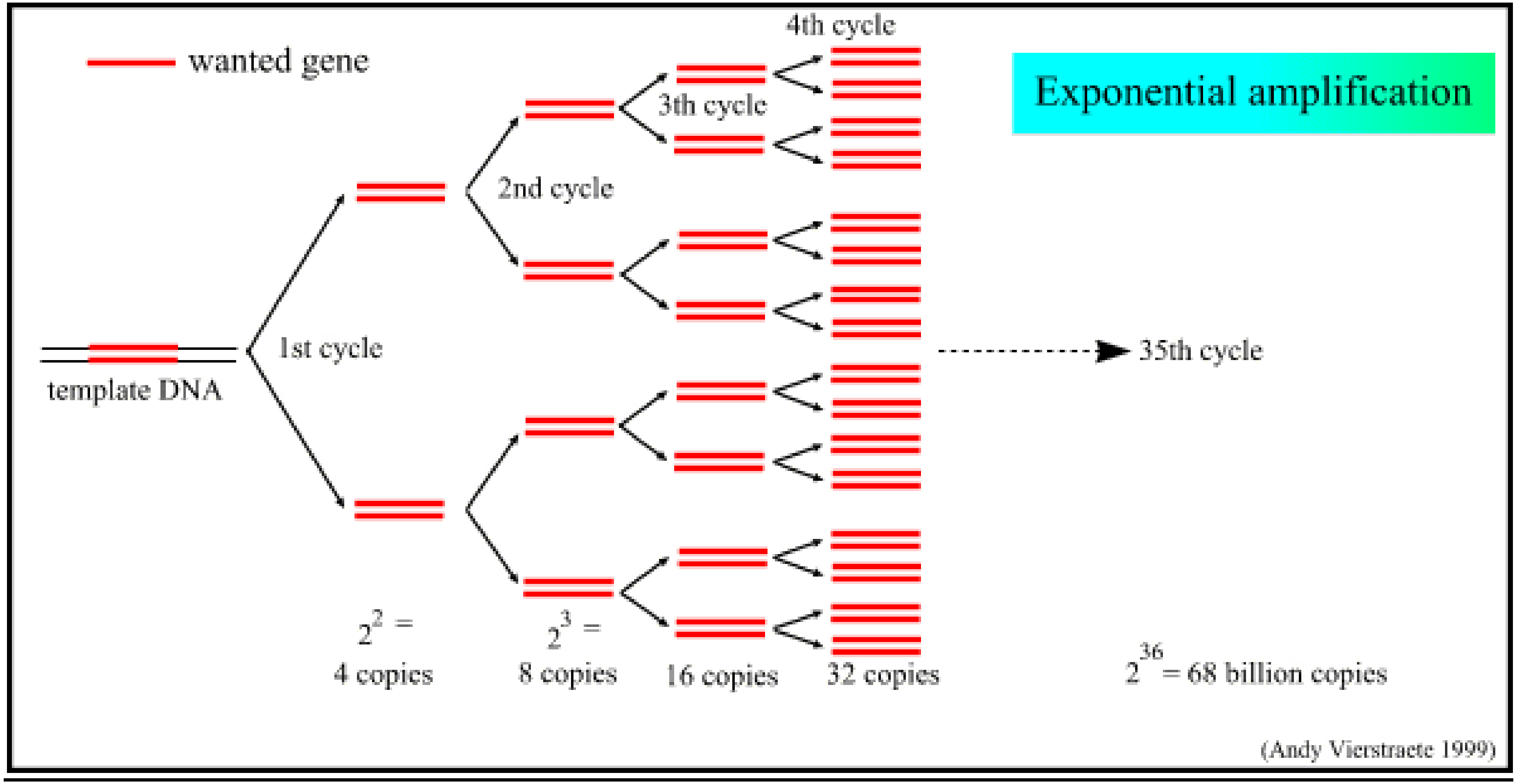
(http://users.ugent.be/~avierstr/principles/pcr.html)

The cycling reactions of a PCR

- Because both strands are copied during PCR, there is an *exponential* increase of the number of copies of the gene.
- Suppose there is only one copy of the wanted gene before the cycling starts,
  - after one cycle, there will be 2 copies,
  - after two cycles, there will be 4 copies,
  - three cycles will result in 8 copies and so on.

(http://users.ugent.be/~avierstr/principles/pcr.html)

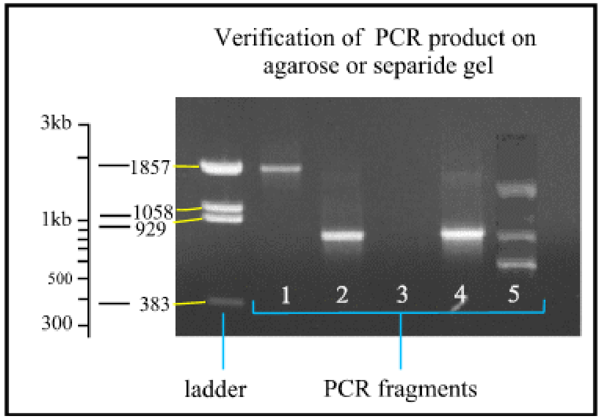
The cycling reactions of a PCR



(http://users.ugent.be/~avierstr/principles/pcr.html)

Quality control of PCR reaction

- Before the PCR product is used in further applications, it has to be checked if :
  - There is a product formed.
  - The product is of the right size
  - Only one band is formed.
    - It is possible that the primers fit on the desired locations, and also on other locations. In that case, you can have different bands in one lane on a gel.



References:

- Ziegler A and König I. *A Statistical approach to genetic epidemiology*, 2006, Wiley.
  (Chapter 1, Sections 2.3.1; 3.1, 3.2.2; 5.1, 5.2.1-5.2.3)
- Burton P, Tobin M and Hopper J. Key concepts in genetic epidemiology. *The Lancet*, 2005
- Clayton D. Introduction to genetics (course slides Bristol 2003)
- URLs:
  - http://www.rothamsted.ac.uk/notebook/courses/guide/
  - http://www.cellbio.com/courses.html
  - http://www.genome.gov/Education/
  - http://www.roche.com/research_and_development/r_d_overview/education.htm
  - http://nitro.biosci.arizona.edu/courses/EEB320-2005/
  - http://atlasgeneticsoncology.org/GeneticFr.html
  - http://www.worthpublishers.com/lehninger3d/index2.html
  - http://www.dorak.info/evolution/glossary.html

For a primer on the Human Genome Project

- http://www.sciencemag.org/content/vol291/issue5507/

Background reading:

- Spurbeck et al 2004. Primer on medical genomics. Part XI: Visualizing human chromosomes. *Mayo Clin Proc*, 79:58-75.
- Wieben 2003. Primer on medical genomics. Part VII: The evolving concept of the gene. *Mayo Clin* Proc, 78:580-587
- URLs:

  A history of the Human Genome Project

  - http://www.sciencemag.org/cgi/content/full/291/5507/1195

  Introduction to genetics

  - http://www.roche.com/research_and_development/r_d_overview/education.htm

In-class discussion document

- Tefferi et al 2002. Primer on medical genomics. Part II: Background principles and methods in molecular genetics. *Mayo Clin Proc*, 77:785-808.
- Schlötterer 2004. The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics*, 5:63-70
- Pompanon et al 2005. Genotyping errors: causes, consequences and solutions. Nature Reviews Genetics, 6:847-859

Questions: In class reading_2.pdf

Preparatory reading:

Bioinformatics explained – biological data bases.