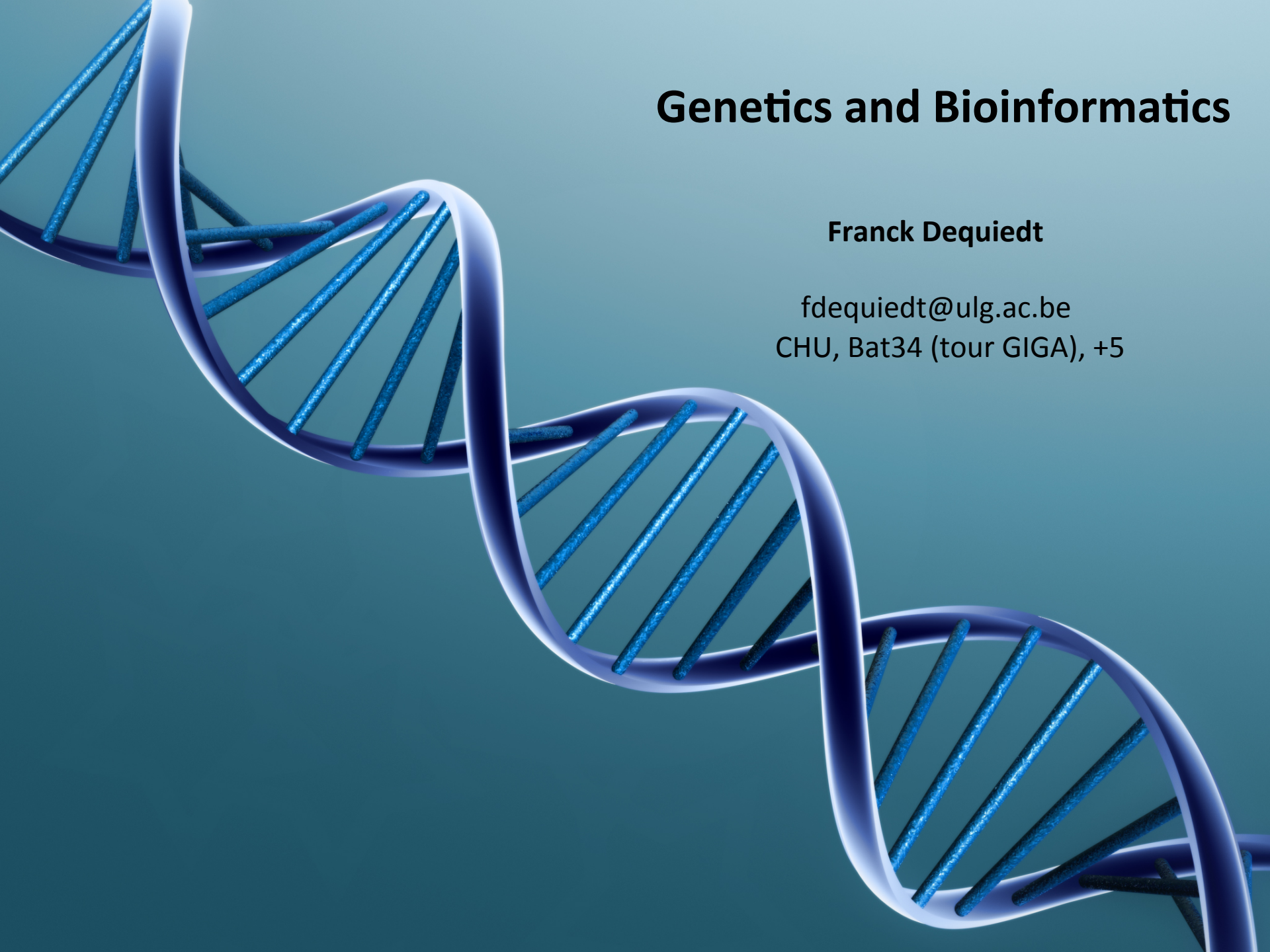


Genetics and Bioinformatics

Franck Dequiedt

fdequiedt@ulg.ac.be
CHU, Bat34 (tour GIGA), +5

INTRODUCTION

Les PRINCIPES FONDAMENTAUX

Caractère : particularité qualitative ou quantitative transmissible à la descendance

Evolution : Modification des caractères de génération en génération sous l'action de la sélection naturelle et de la dérive génétique

Hérédité : transmission des caractères (qualitatifs ou quantitatifs) innés d'un individu à sa descendance

I - La naissance de la Biologie Moléculaire

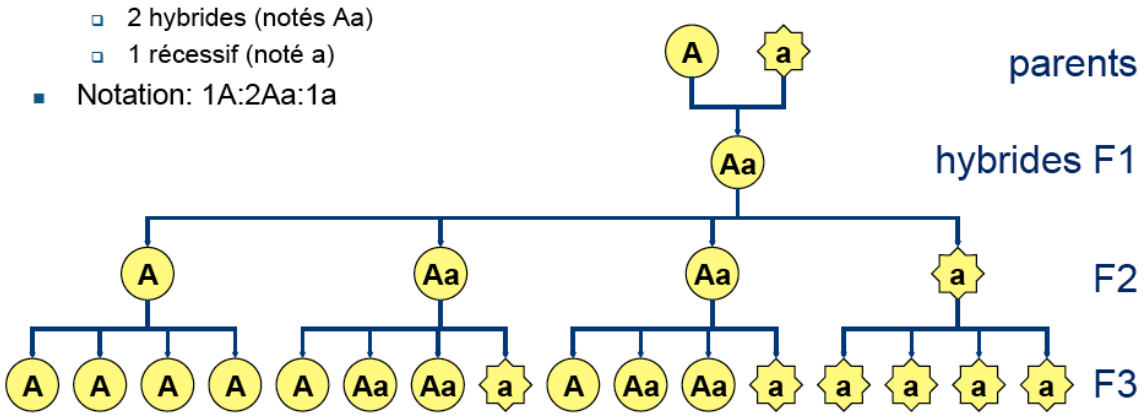
- L'émergence de la génétique formelle

Gregor Mendel (1822 - 1884)

les Lois de l'hérédité

Loi n°1 : uniformité des caractères à la première génération.

Loi n°2 : Ségrégation des caractères



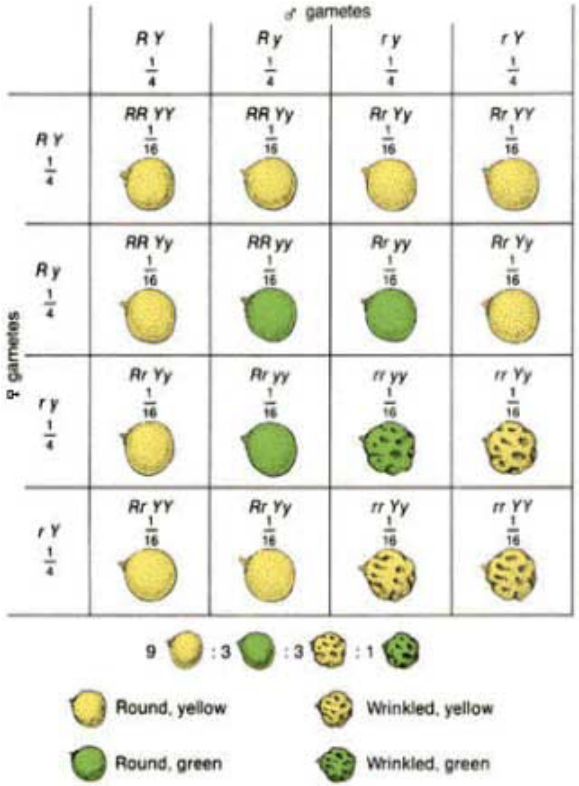
I - La naissance de la Biologie Moléculaire

- L'émergence de la génétique formelle

Gregor Mendel (1822 - 1884)

les Lois de l'hérédité

Loi n°3 : Assortiment indépendant des caractères



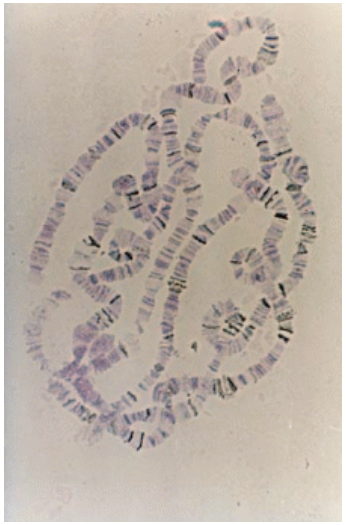
I - La naissance de la Biologie Moléculaire

- L'émergence de la génétique formelle Gregor Mendel (1822 - 1884)
- Le chromosome, support de l'hérédité

Thomas Morgan (1866 - 1945)

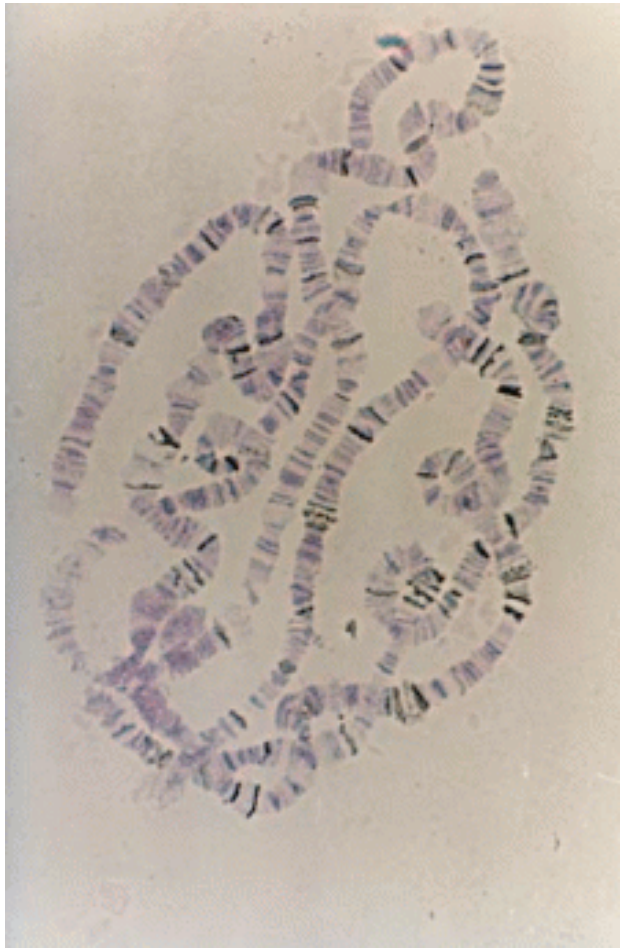
Théorie chromosomique de l'hérédité

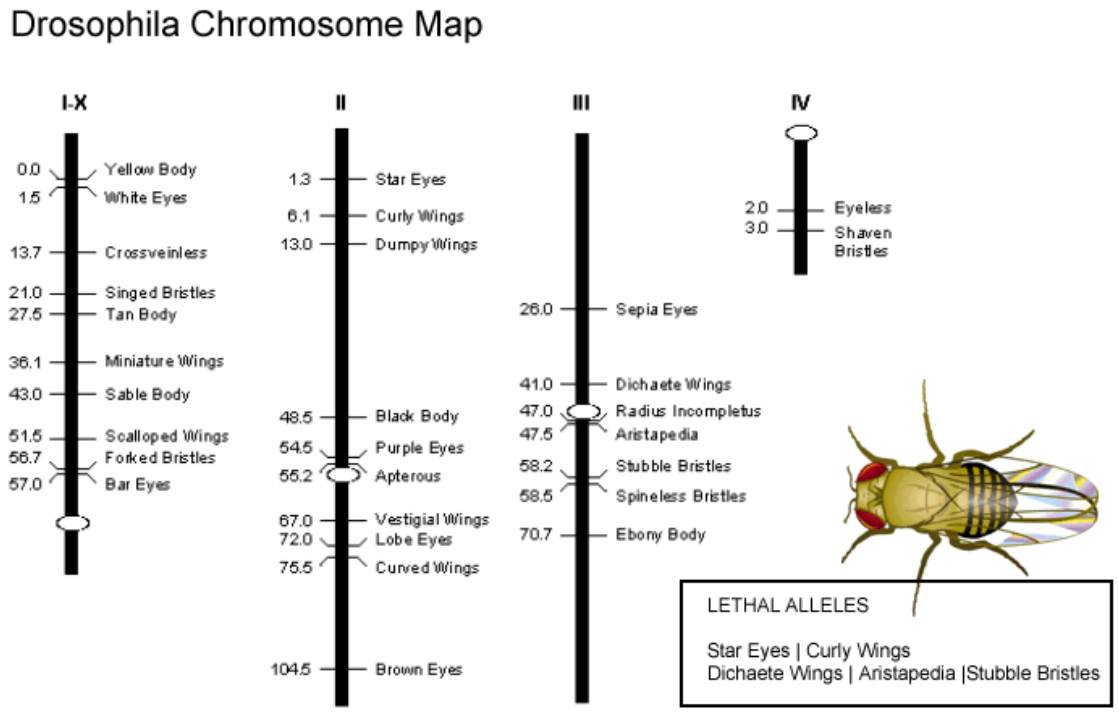
Drosophila melanogaster



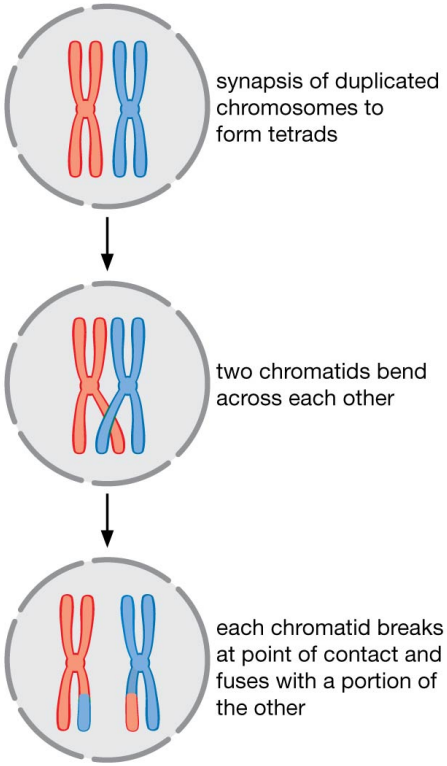
Chromosomes polytènes

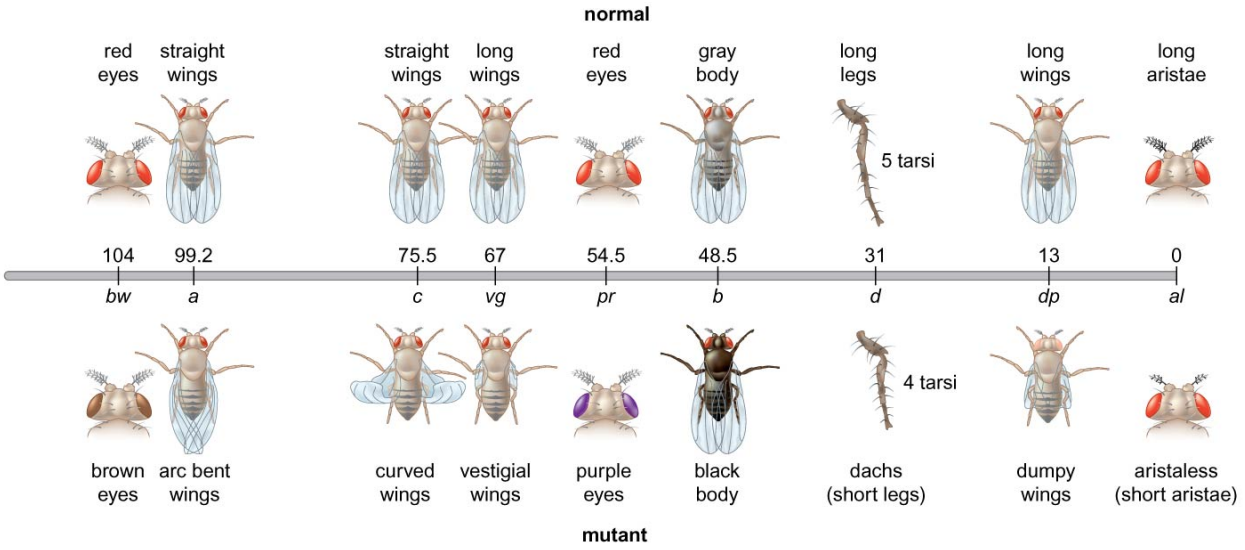
- Le chromosome, support de l'hérédité (Thomas Morgan)





- Le chromosome, support de l'hérédité (Thomas Morgan)





I La naissance de la Biologie Moléculaire

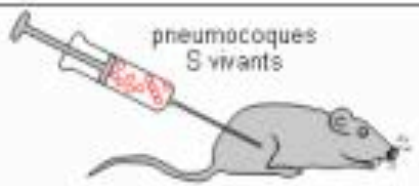
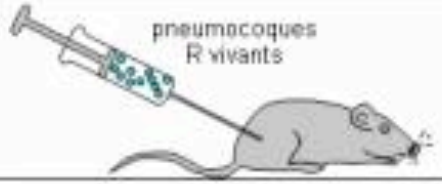
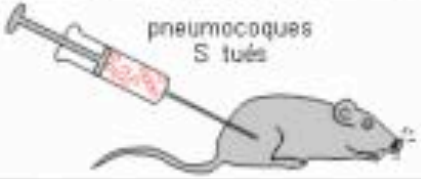
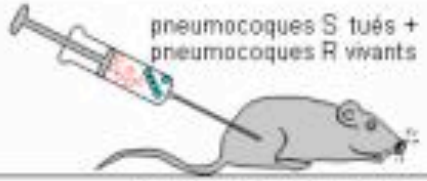
- L'émergence de la génétique formelle — Gregor Mendel (1822 - 1884)
- Le chromosome, support de l'hérédité — Thomas Morgan (1866 - 1945), Alfred Sturtevant (1891 - 1970)
- Le chromosome, support de l'hérédité

Fred Griffith (1877 - 1941)

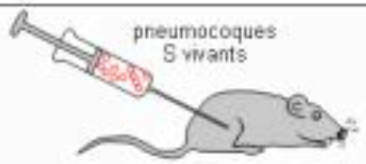
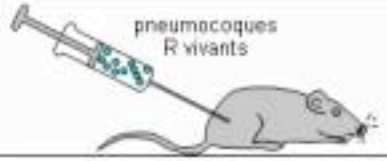
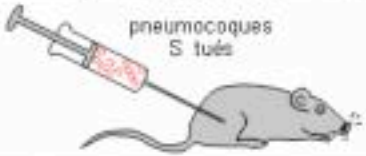
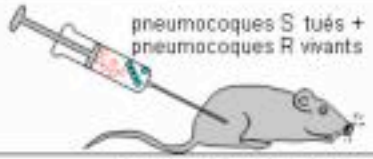
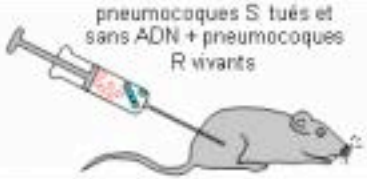
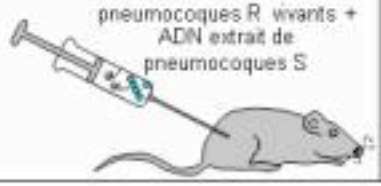
L'expérience de Griffith : le facteur transformant

L'expérience de Griffith (1928) : le facteur transformant

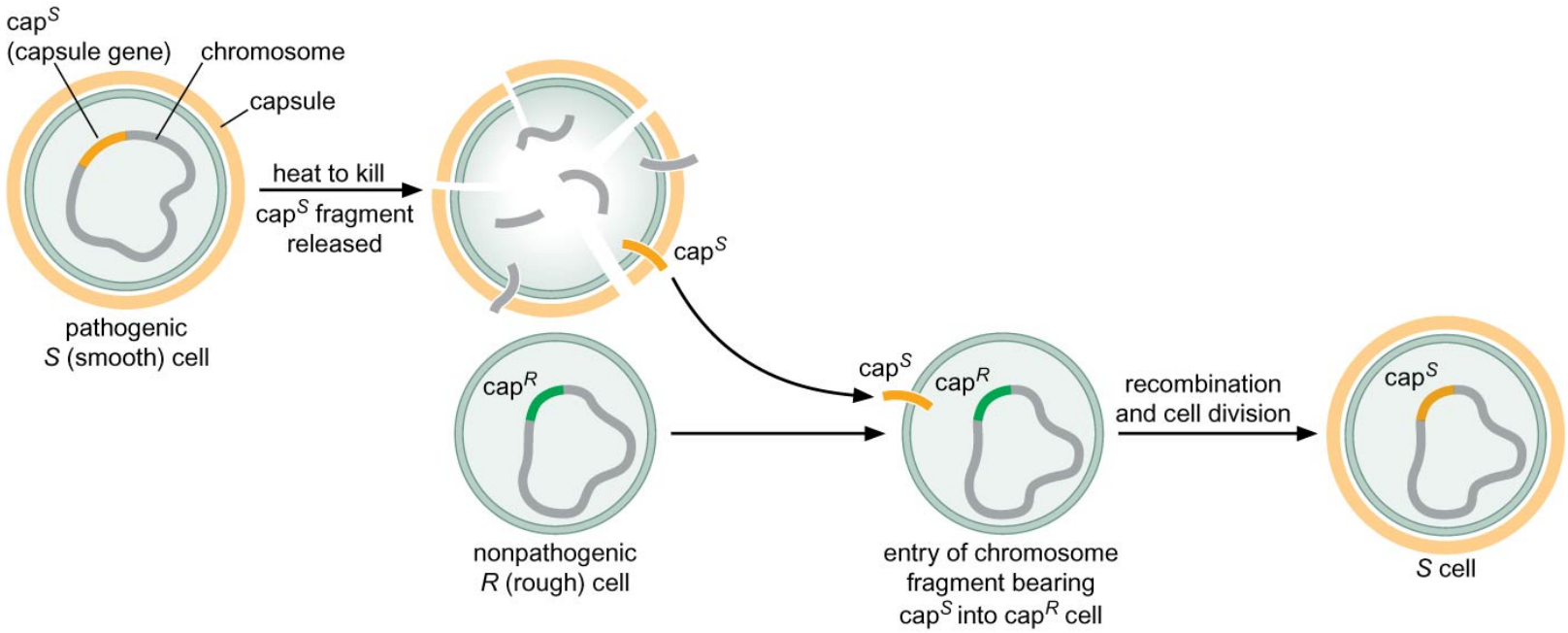
Streptococcus pneumoniae : bactérie responsable de pneumonies

n°	expériences	état de la souris	analyse du sang de la souris
1	pneumocoques S vivants   pneumocoques S vivants	mort 	présence de très nombreux pneumocoques S vivants 
2	 pneumocoques R vivants 	survie 	absence de tout pneumocoque
3	capsule détruite  pneumocoques S tués 	survie 	absence de tout pneumocoque
4	 pneumocoques S tués + pneumocoques R vivants  pneumocoques S tués + pneumocoques R vivants	mort 	Présence de très nombreux pneumocoques S vivants 

L'expérience de Griffith complétée par Avery

n°	expériences	état de la souris	analyse du sang de la souris
1	 pneumocoques S vivants 	mort 	présence de très nombreux pneumocoques S vivants 
2	 pneumocoques R vivants 	survie 	absence de tout pneumocoque
3	capsule détruite  pneumocoques S tués 	survie 	absence de tout pneumocoque
4	 pneumocoques S tués + pneumocoques R vivants 	mort 	Présence de très nombreux pneumocoques S vivants 
5	 pneumocoques S tués et sans ADN + pneumocoques R vivants 	survie 	absence de tout pneumocoque
6	 pneumocoques R vivants +ADN extrait de pneumocoques S 	mort 	Présence de très nombreux pneumocoques S vivants 

L'expérience de Griffith complétée par Avery : Le Principe de la Transformation Bactérienne



"Both Francis (Crick) and I had no doubts that DNA was the gene. But most people did. And again, you might say, "Why didn't Avery get the Nobel Prize?" Because most people didn't take him seriously. Because you could always argue that his observations were limited to bacteria, or that [the transformation of Pneumococcus that he described was caused by] a protein resistant to proteases and that the DNA was just scaffolding".
James Watson, in Nature, 302, 21 (April 1983): 654

I - La naissance de la Biologie Moléculaire

- L'émergence de la génétique formelle — Gregor Mendel (1822 - 1884)
- Le chromosome, support de l'hérédité — Thomas Morgan (1866 - 1945), Alfred Sturtevant (1891 - 1970)
- L'ADN, support de l'hérédité — Fred Griffith (1877 - 1941), Oswald Avery (1877 – 1955)

Alfred Hershey (1908-1997)
Martha Chase (1927-2003)

- La composition et structure de l'ADN

Erwin Chargaff (1905 - 1992)

$$A+T / C+G = K$$

$$A / T = C / G = 1$$

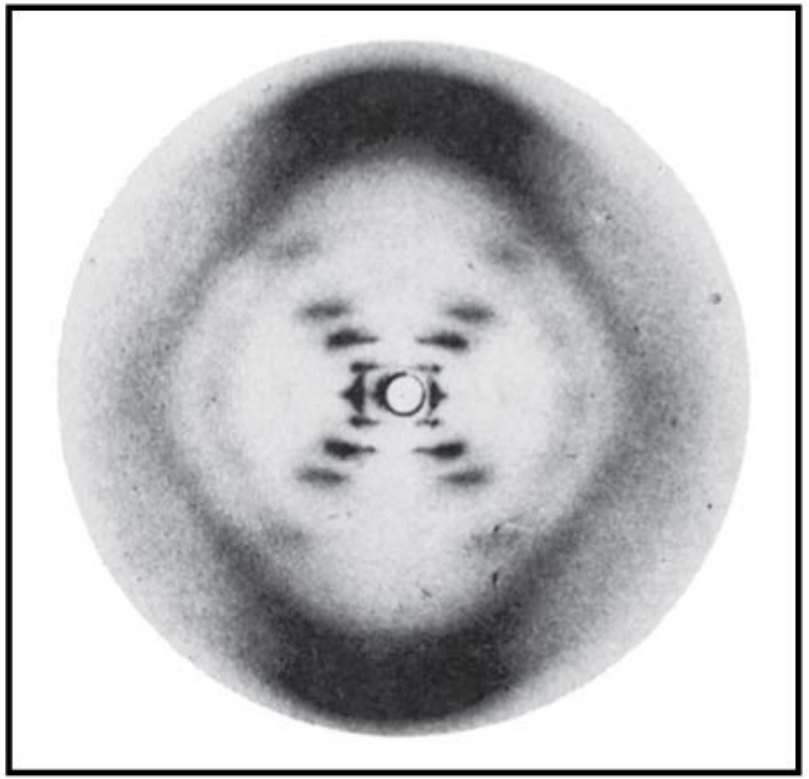
La structure hélicoïdale de l'ADN (1952)

Rosalind Franklin,
1920–1958

Maurice Wilkins,
1916–2004

"Conclusion: Big helix in several chains, phosphates on outside, phosphate-phosphate inter-helical bonds disrupted by water. Phosphate links available to proteins."

Rosalind Franklin



I - La naissance de la Biologie Moléculaire

- L'émergence de la génétique formelle — Gregor Mendel (1822 - 1884)
- Le chromosome, support de l'hérédité — Thomas Morgan (1866 - 1945), Alfred Sturtevant (1891 - 1970)
- Le chromosome, support de l'hérédité — Fred Griffith (1877 - 1941), Oswald Avery (1877 - 1955)
- La composition et structure de l'ADN

Erwin Chargaff (1905 - 1992)

Rosalind Franklin (1920-1958)

James Dewey Watson (1928 -)

Francis Harry Compton Crick (1916 - 2004)

http://www.radio-canada.ca/actualite/decouverte/reportages/2003/04-2003/03-04-27/adn.html

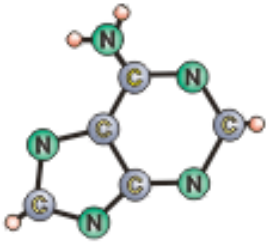
II – Structure des acides nucléiques

II.I Les molécules simples

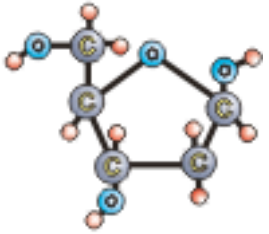
L'acide désoxyribonucléique (ADN)

= polymères de sous-unités appelées **nucléotides**

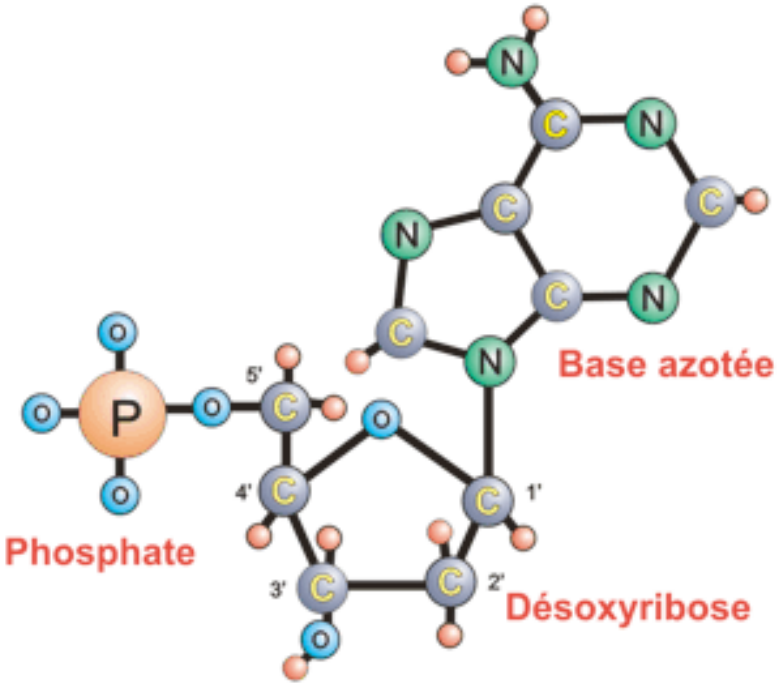
une **BASE**



un **SUCRE**

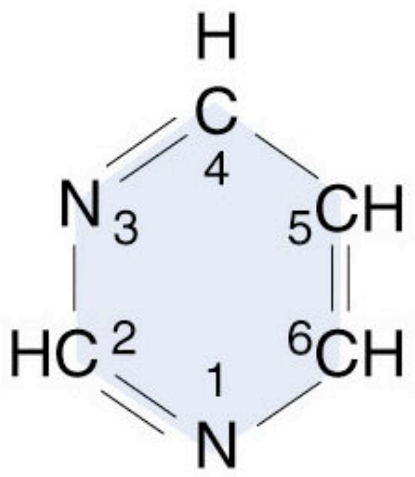


un **ACIDE PHOSPHORIQUE**

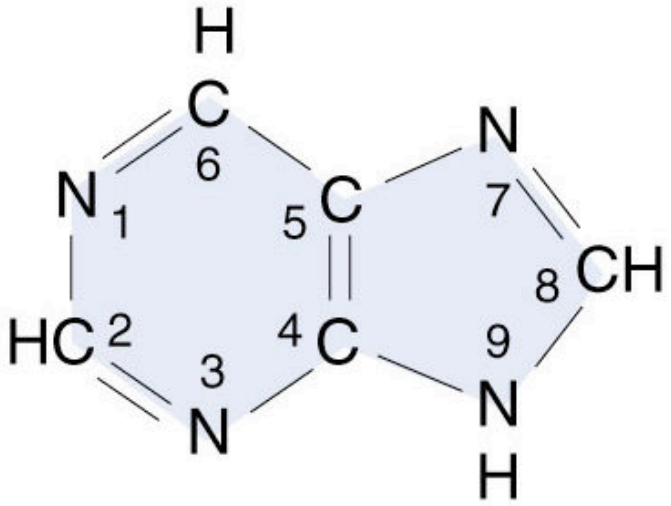


II.I Les molécules simples

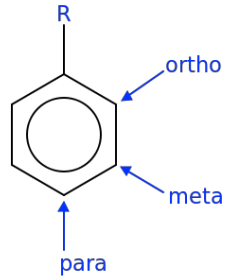
II.I.1 Les bases azotées



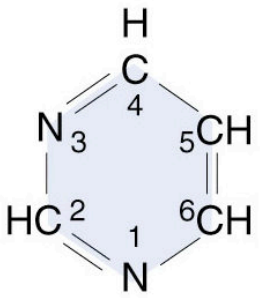
Pyrimidine
(1,3 Diazine)



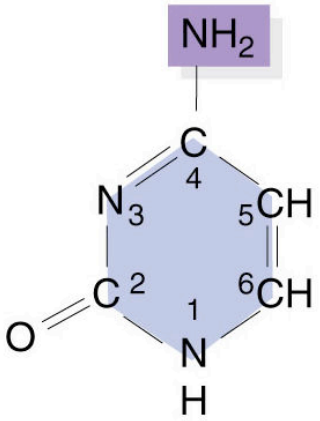
Purine
(Imidazole(4,5)Pyrimidine)



II.I.1 Les bases azotées

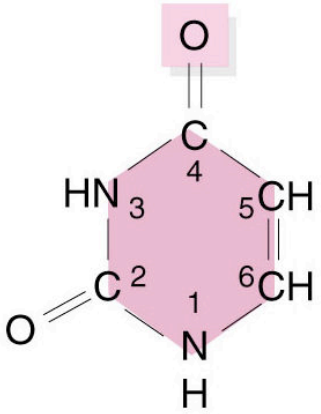


Pyrimidine



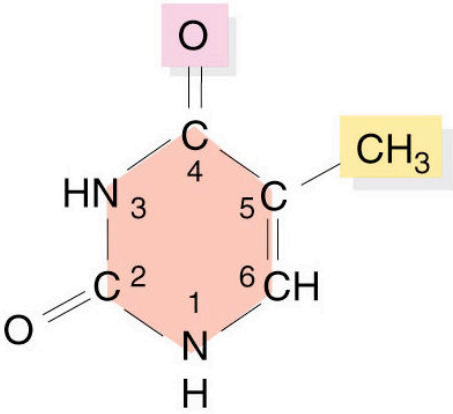
Cytosine (C)

2-oxy-4-amino -pyrimidine



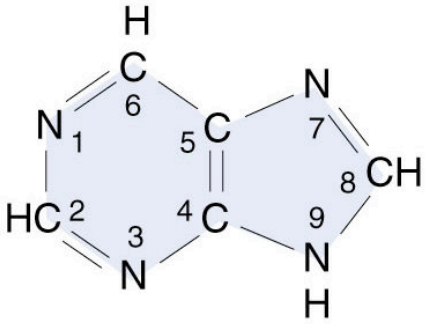
Uracil (U) (found in RNA)

2,4-dioxy-pyrimidine

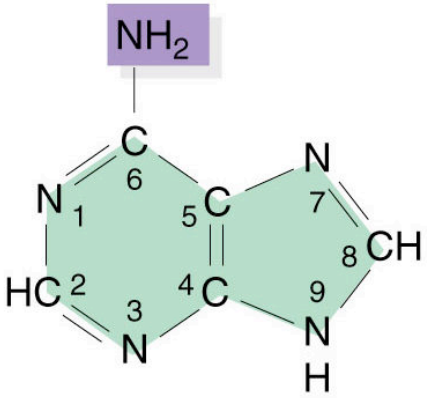


Thymine (T) (found in DNA)

2,4-dioxy-5-méthyl-pyrimidine

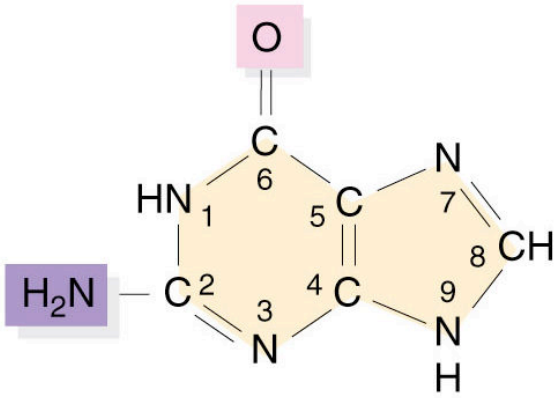


Purine



Adenine (A)

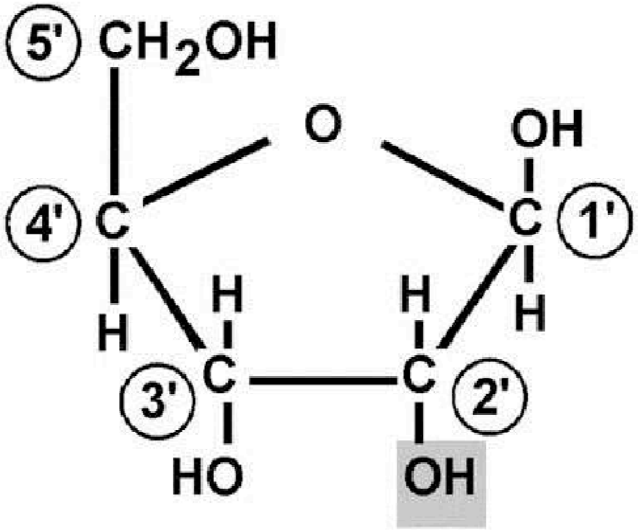
6-amino-purine



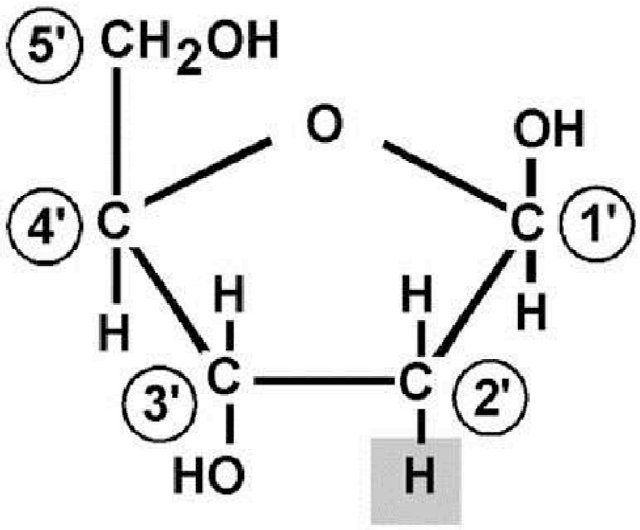
Guanine (G)

2-amino-6-oxy -purine

II.I.2 Les sucres



Pentose présent dans l'ARN :
le ribose

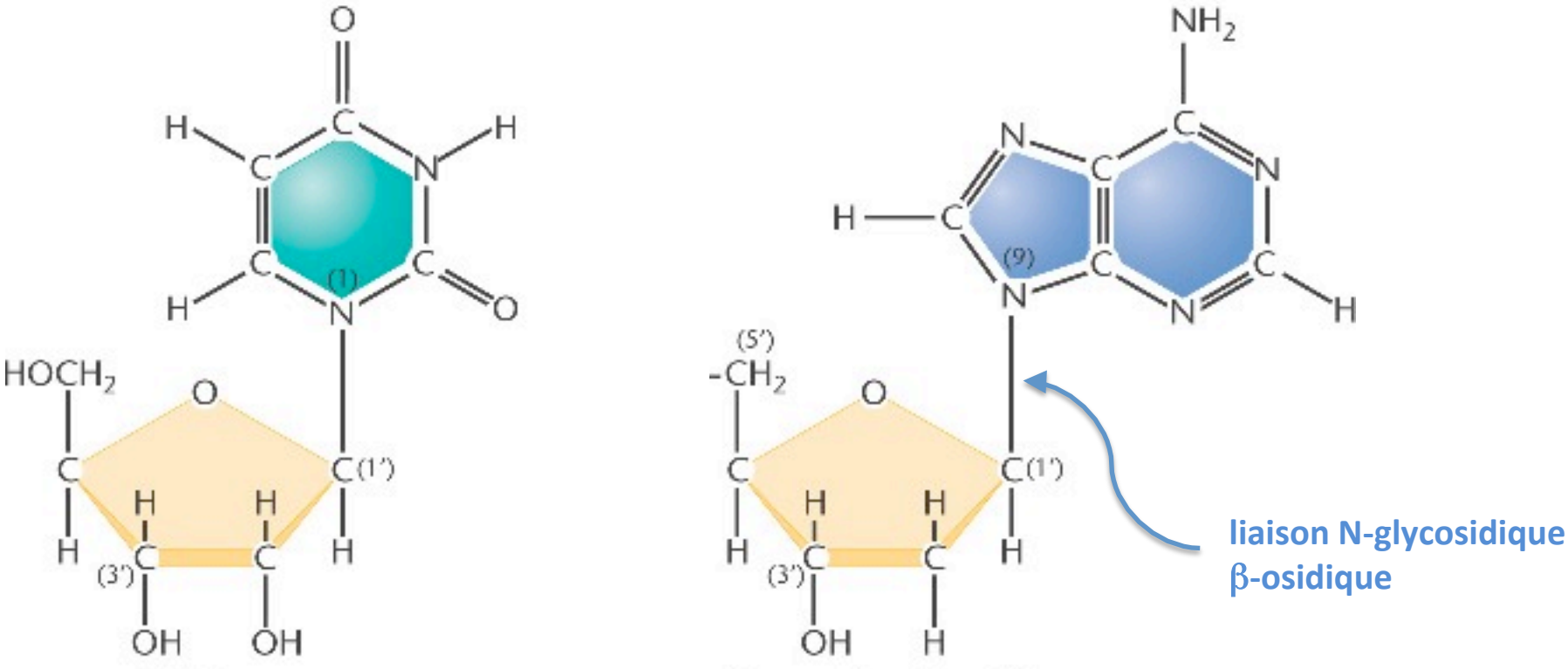


Pentose présent dans l'ADN :
le désoxyribose

II.2 Les nucléosides

Les nucléoSides

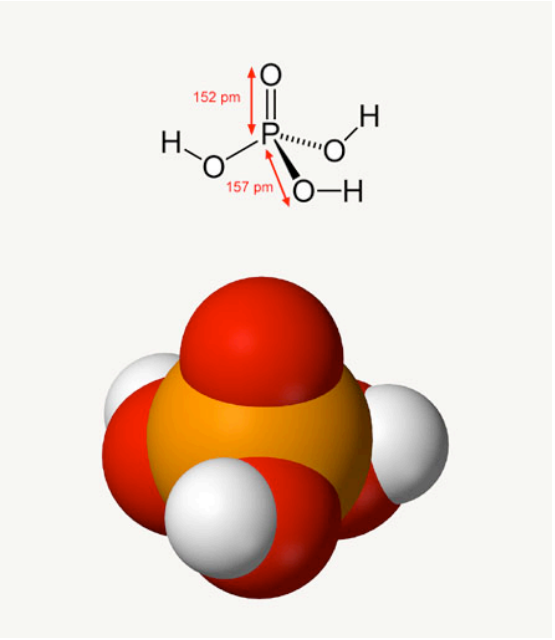
SUCRE + BASE



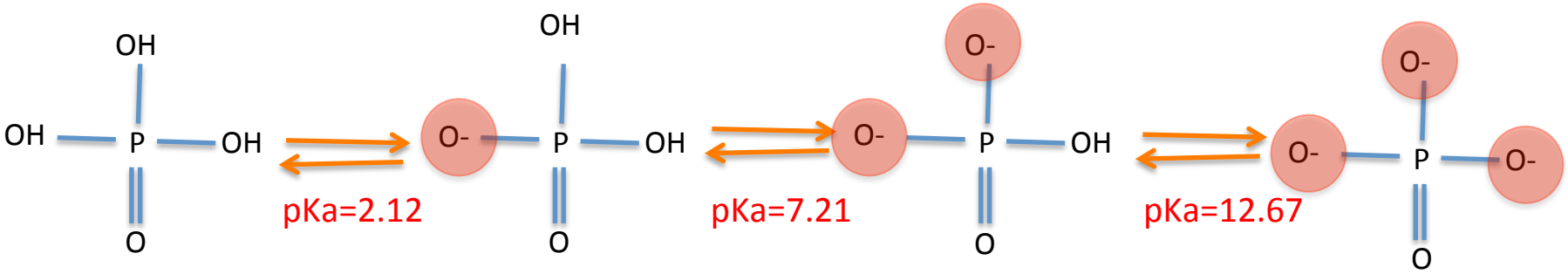
II.2 Les nucléosides

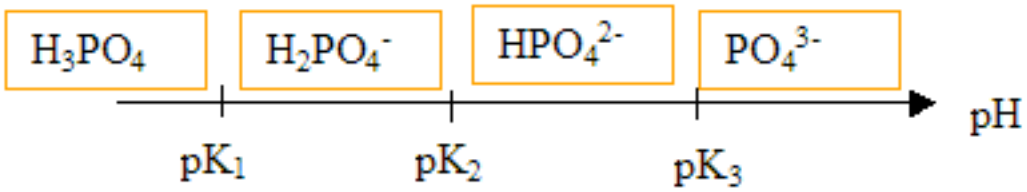
Base	Ribonucleoside	Desoxyribonucleoside
Adénine	Adén**osine**	Désoxyadénosine
Guanine	Guan**osine**	Désoxyguanosine
Uracile	Ur**idine**	Désoxyuridine
Cytosine	Cyt**idine**	Désoxycytidine
Thymine	Thymine ribonucléoside	Désoxythymidine ou thymidine

II.3 L'acide phosphorique



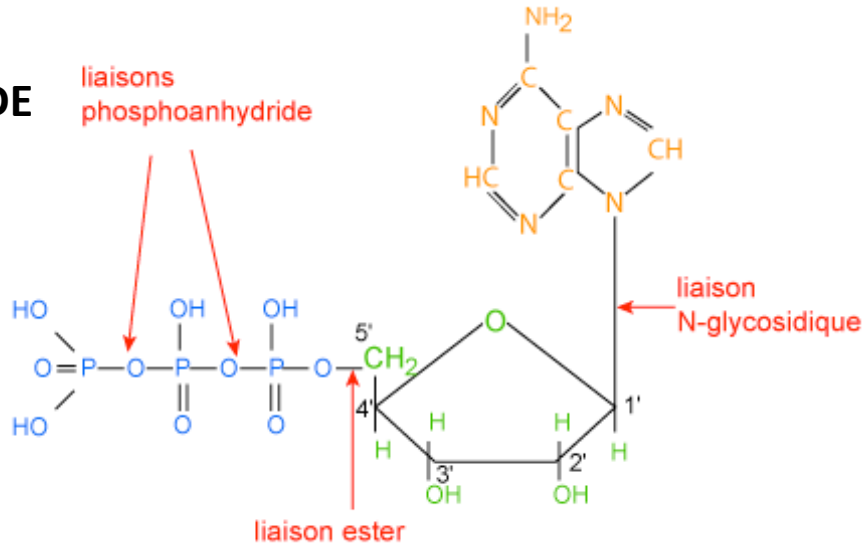
L'acide phosphorique

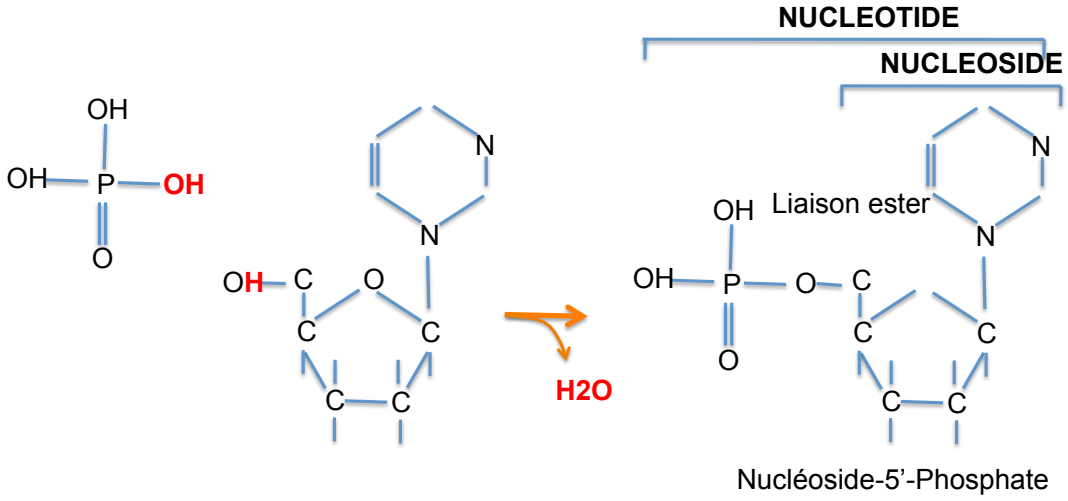




II.4 Les nucléotides

SUCRE + BASE+ PHOSPHATE = NUCLEOTIDE





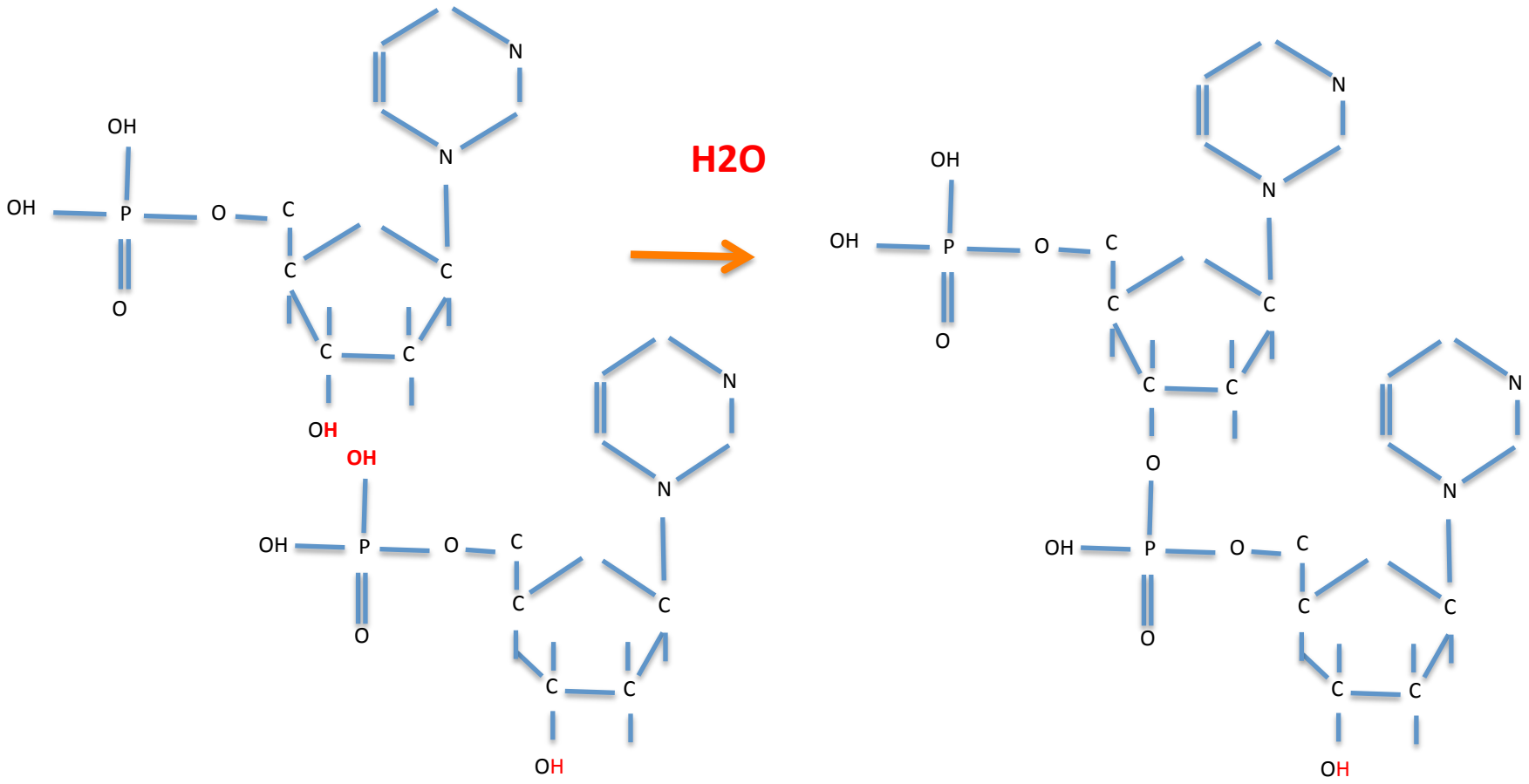
II.4 Les nucléotides

Les nucléo**T**ides : nucléosides-5'-Monophosphate

Base	RibonucleoTide	DesoxyribonucleoTide
Adénine	Adén**osine**-5'-Monophosphate (AMP)	Désoxyadénosine-5'-Monophosphate (dAMP)
Guanine	Guan**osine**-5'-Monophosphate (GMP)	Désoxyguanosine-5'-Monophosphate (dGMP)
Uracile	Ur**idine**-5'-Monophosphate (UMP)	Désoxyuridine-5'-Monophosphate (dUMP)
Cytosine	Cyt**idine**-5'-Monophosphate (CMP)	Désoxycytidine-5'-Monophosphate (dCMP)
Thymine	Thymine riboside -5'-Monophosphate (TMP)	Désoxythymidine -5'-Monophosphate (dTMP)

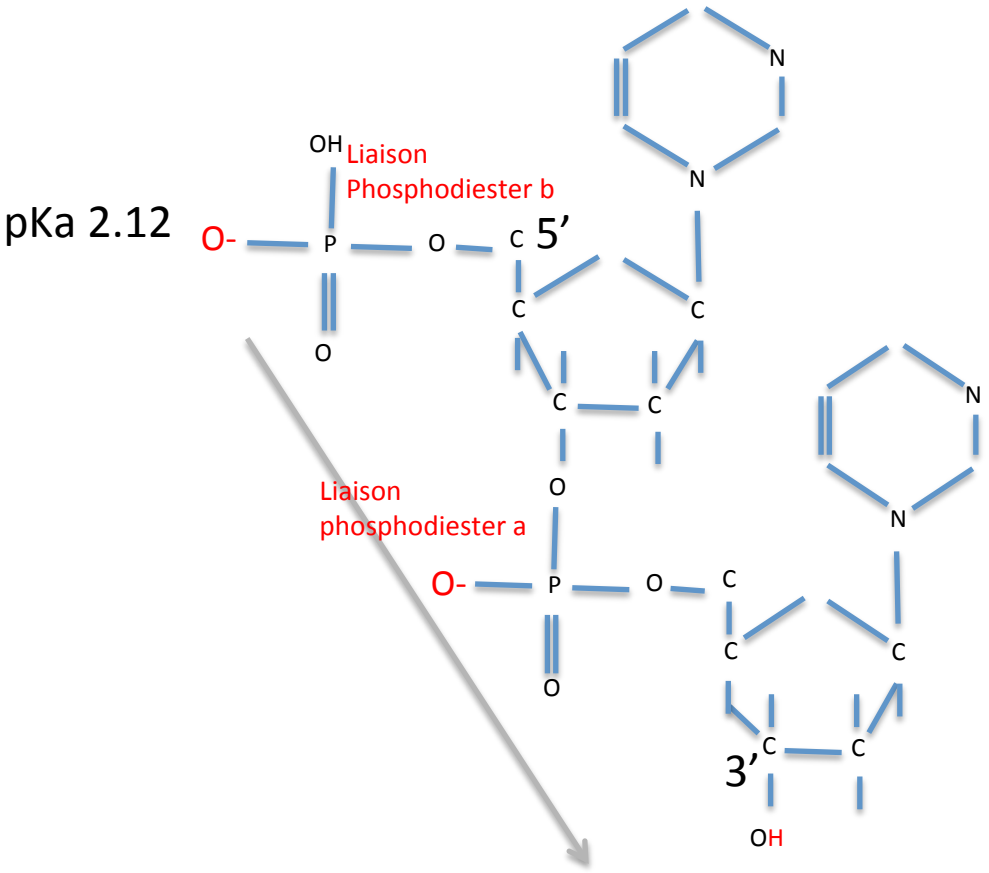
II.5 La liaison phosphodiester

La chaine polynucléotidique



II.5 La liaison phosphodiester

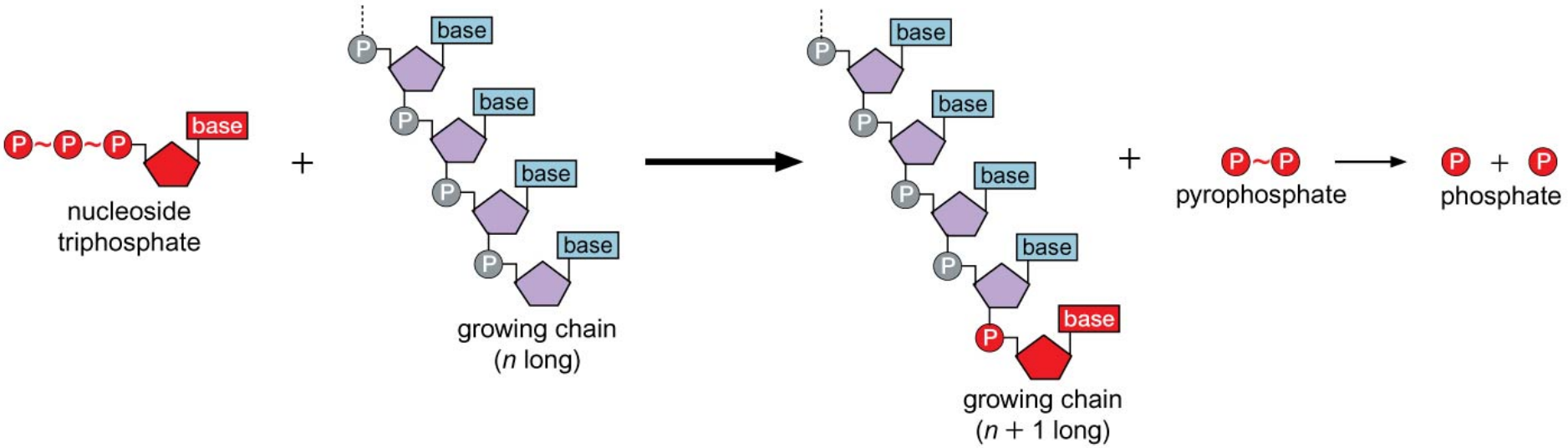
La chaine polynucléotidique



II.5 La liaison phosphodiester

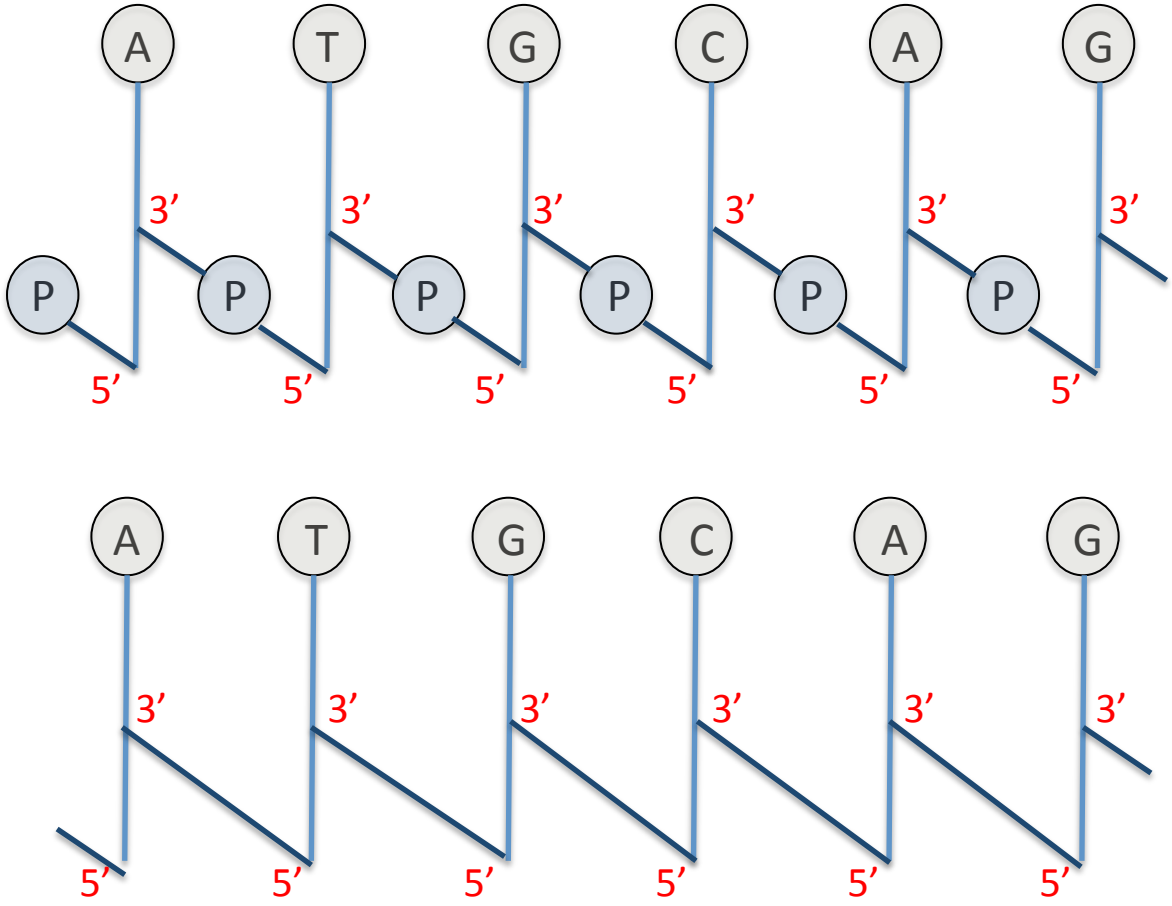
(Désoxy)nucléotide-P-P-P + Chaine polynucléotidique à n nucléotides

⟶ Chaine polynucléotidique à n+1 + P-P



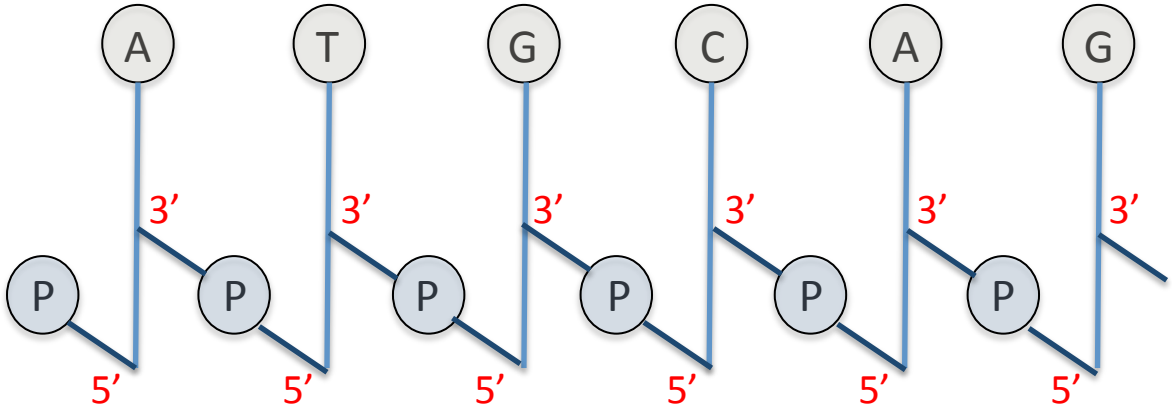
II.5 La liaison phosphodiester

La chaine polynucléotidique



II.5 La liaison phosphodiester

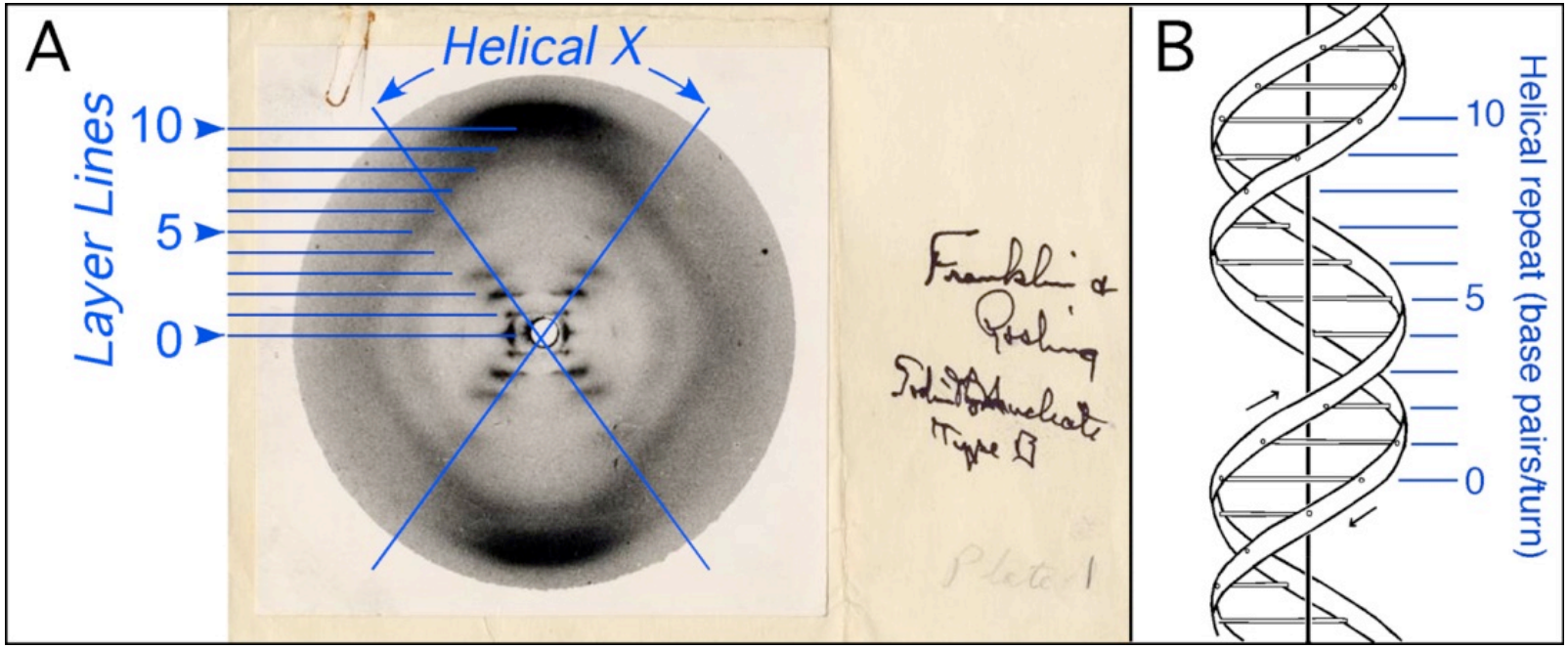
La chaine polynucléotidique



p5’dAp5’dTp5’dGp5’dCp5’dAp5’dG ou 5’-dATGCAG-3’
5’-ATGCAG-3’

II.6 Structure Secondaire de la molécule d'ADN

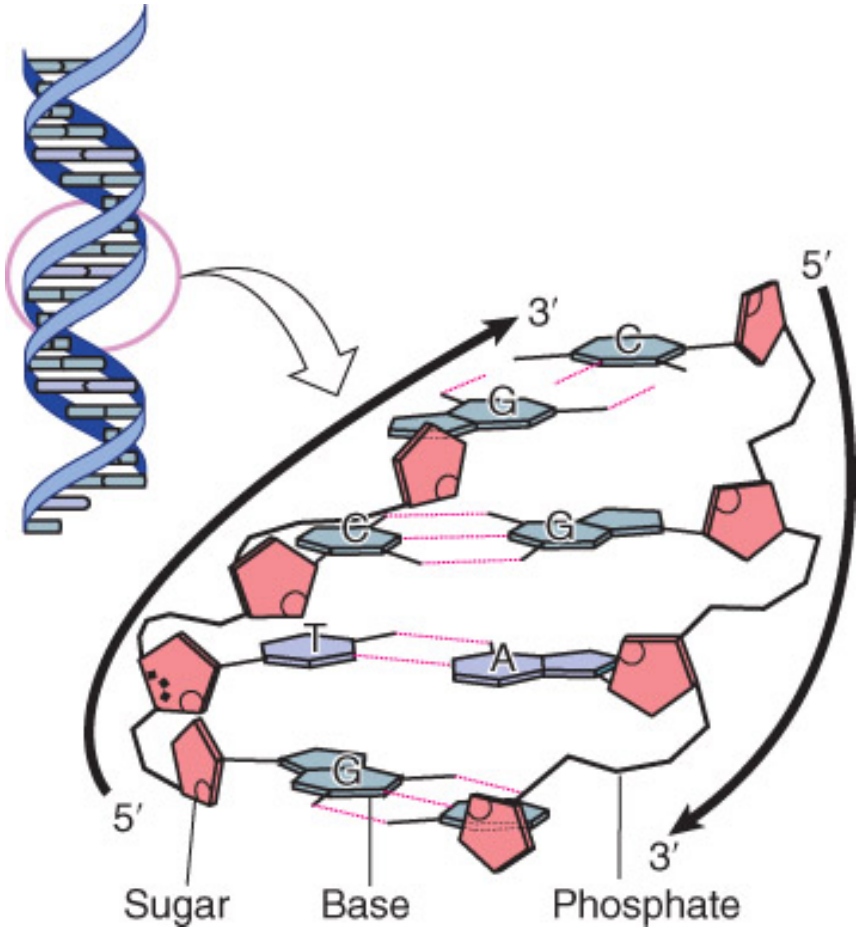
Structure secondaire : le modèle de Watson et Crick



Rosalind Franklin

II.6 Structure Secondaire de la molécule d'ADN

La molécule d'ADN : structure secondaire



- **Bicaténaire** (2 chaînes polynucléotidiques)

- **Antiparallèle**

- **Hélicoïdale**

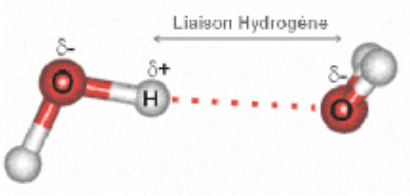
- Bases (intérieur) Sucres/phosphates (extérieur)

- **Complémentaire**

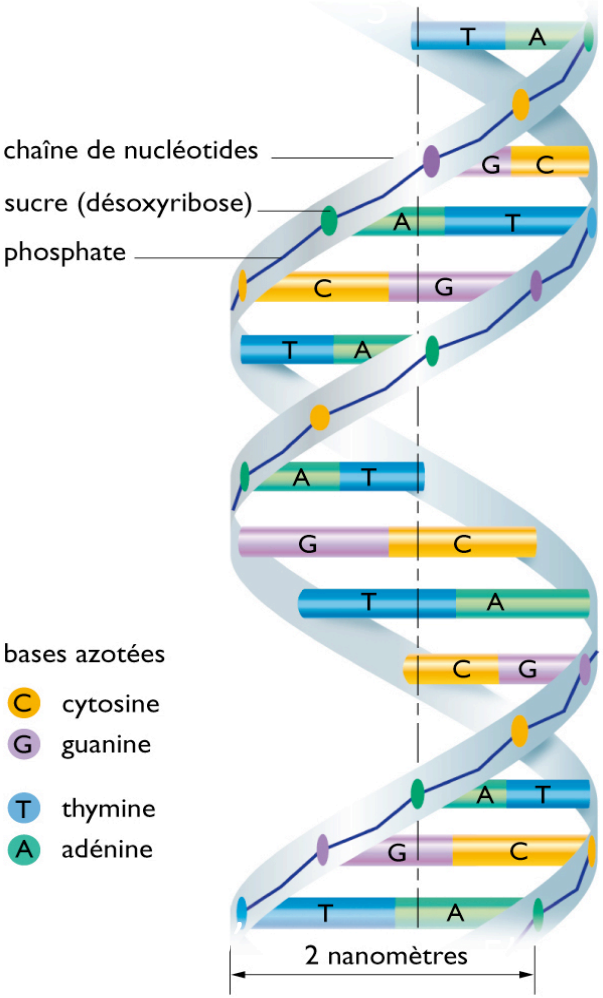
http://biomodel.uah.es/en/model4/dna_fr/dnapairs.htm

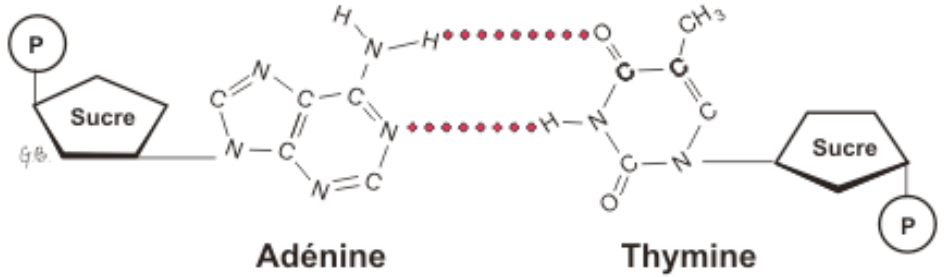
II.6 Structure Secondaire de la molécule d'ADN

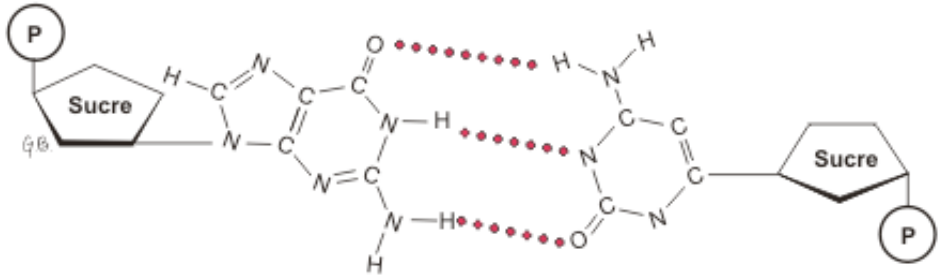
La molécule d'ADN : structure secondaire



La règle de complémentarité

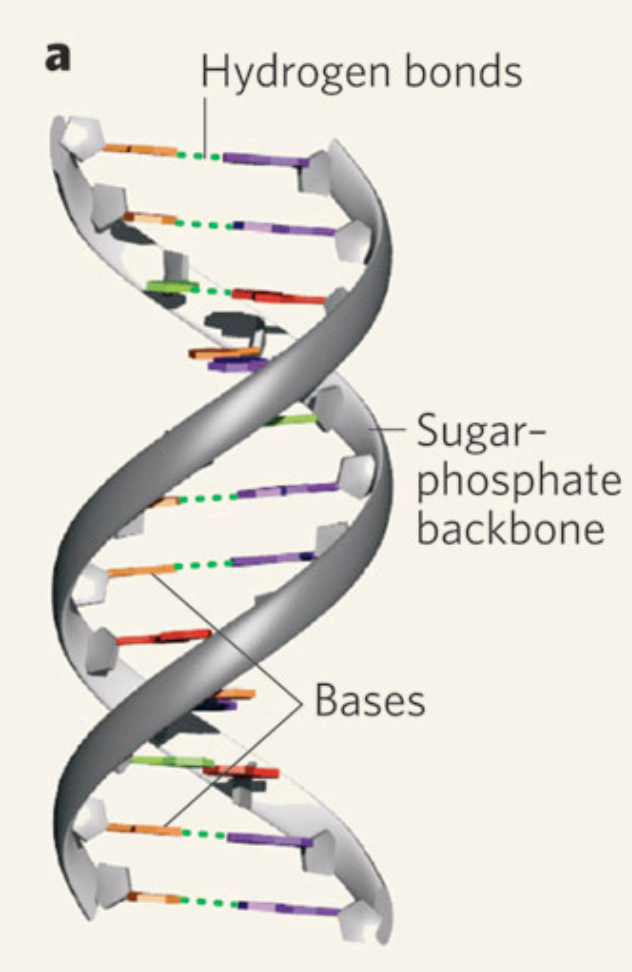






II.6 Structure Secondaire de la molécule d'ADN

La molécule d'ADN : structure secondaire et stabilité



https://www.youtube.com/watch?v=q6PP-C4udkA

III. Structure du génome

Dans les cellules, la molécule d'ADN est associées à des protéines qui vont déterminer sa structure en imposant à la molécule d'ADN toute une série de contraintes (torsions, enchevêtrements, etc..).

Cette structuration de l'ADN va avoir des conséquences sur :

- son encombrement

DNA is highly compacted in all types of genomes

Compartment	Shape	Dimensions	Type of Nucleic Acid	Length
TMV	filament	0.008 x 0.3 µm	One single-stranded RNA	2 µm = 6.4 kb
Phage fd	filament	0.006 x 0.85 µm	One single-stranded DNA	2 µm = 6.0 kb
Adenovirus	icosahedron	0.07 µm diameter	One double-stranded DNA	11 µm = 35.0 kb
Phage T4	icosahedron	0.065 x 0.10 µm	One double-stranded DNA	55 µm = 170.0 kb
E. coli	cylinder	1.7 x 0.65 µm	One double-stranded DNA	1.3 mm = 4.2 x 10^3 kb
Mitochondrion (human)	oblate spheroid	3.0 x 0.5 µm	~10 identical double-stranded DNAs	50 µm = 16.0 kb
Nucleus (human)	spheroid	6 µm diameter	46 chromosomes of double-stranded DNA	1.8 m = 6 x 10^6 kb

- son accessibilité (protection, réplication, transcription, etc..)

III. Structure du génome

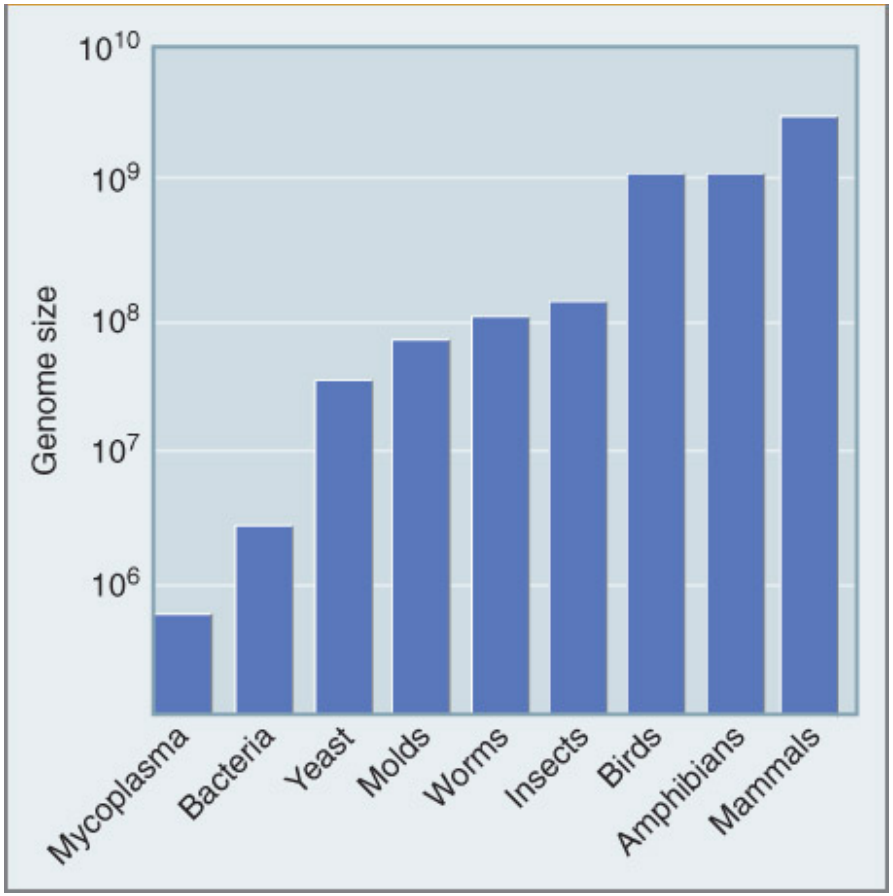
Le génome de chaque organisme est organisé en un nombre spécifique de chromosomes, de forme caractéristique

TABLE 8-1 Variation in Chromosome Makeup in Different Organisms

Species	Number of Chromosomes	Chromosome Copy Number	Form of Chromosome(s)	Genome Size (Mb)
Prokaryotes				
Mycoplasma genitalium	1	1	Circular	0.58
Escherichia coli K-12	1	1	Circular	4.6
Agrobacterium tumefaciens	4	1	3 circular, 1 linear	5.67
Sinorhizobium meliloti	3	1	Circular	6.7
Eukaryotes				
Saccharomyces cerevisiae (budding yeast)	16	1 or 2	Linear	12.1
Schizosaccharomyces pombe (fission yeast)	3	1 or 2	Linear	12.5
Caenorhabditis elegans (roundworm)	6	2	Linear	97
Arabidopsis thaliana (weed)	5	2	Linear	125
Drosophila melanogaster (fruit fly)	4	2	Linear	180
Tetrahymena thermophilus (protozoa) Micronucleus	5	2	Linear	125
Macronucleus	225	10–10,000	Linear	
Fugu rubripes (fish)	22	2	Linear	393
Mus musculus (mouse)	19+X and Y	2	Linear	2600
Homo sapiens	22+X and Y	2	Linear	3200

III. Structure du génome

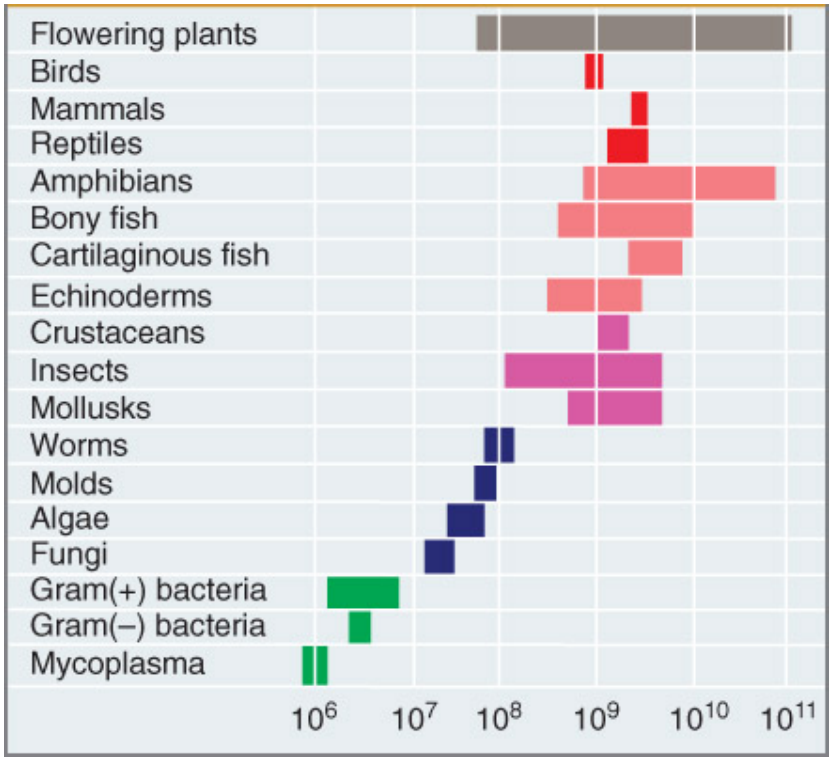
La taille du génome est corrélée à la complexité de l'organisme



C-value (valeur C) : taille d'un génome en paire de bases ou pico-gramme

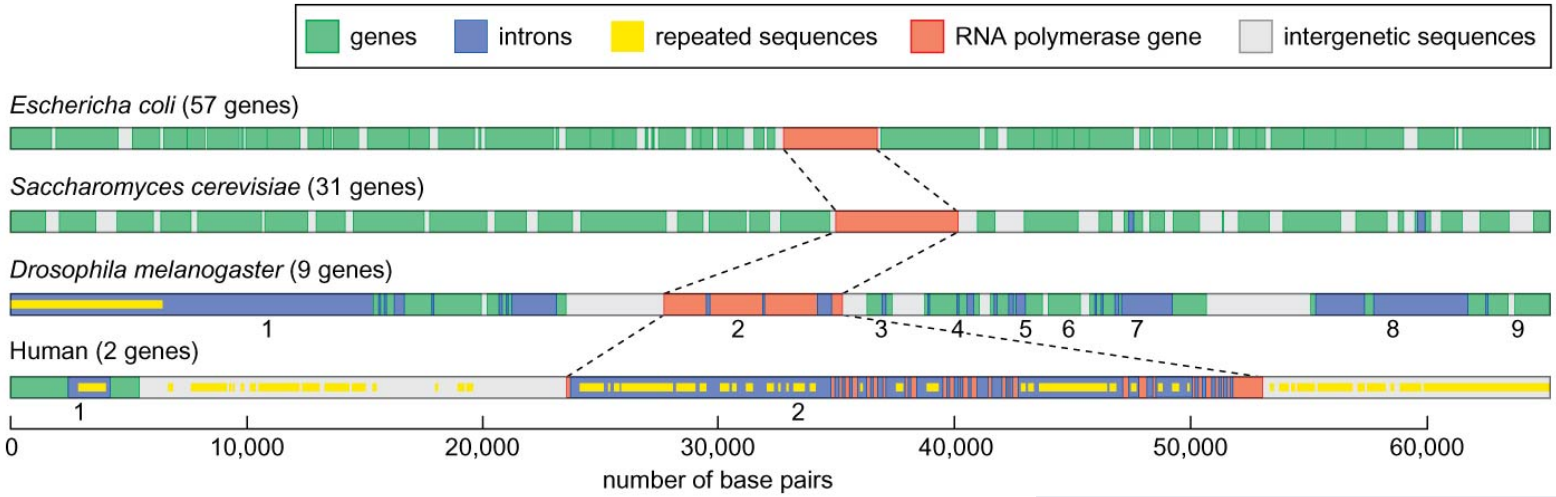
III. Structure du génome

Le paradoxe de la valeur-C : il n'y **a pas** de corrélation exacte entre la **taille** et la **complexité** d'un génome!!!



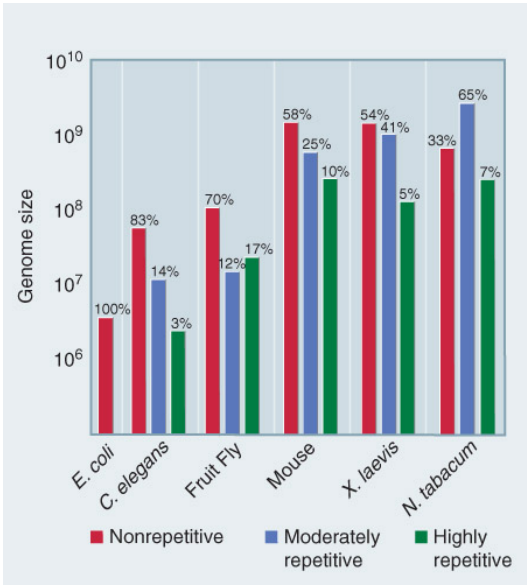
III. Structure du génome

Les organismes les plus complexes ont une densité génique plus faible



<u>ADN modérément ou fortement répété</u> (junk DNA) : en général ce sont des régions non codantes. transposons

<u>ADN non répété</u> : en général ce sont les régions codantes



III. Structure du génome

Le nombre de gènes augmente avec la complexité de l'organisme

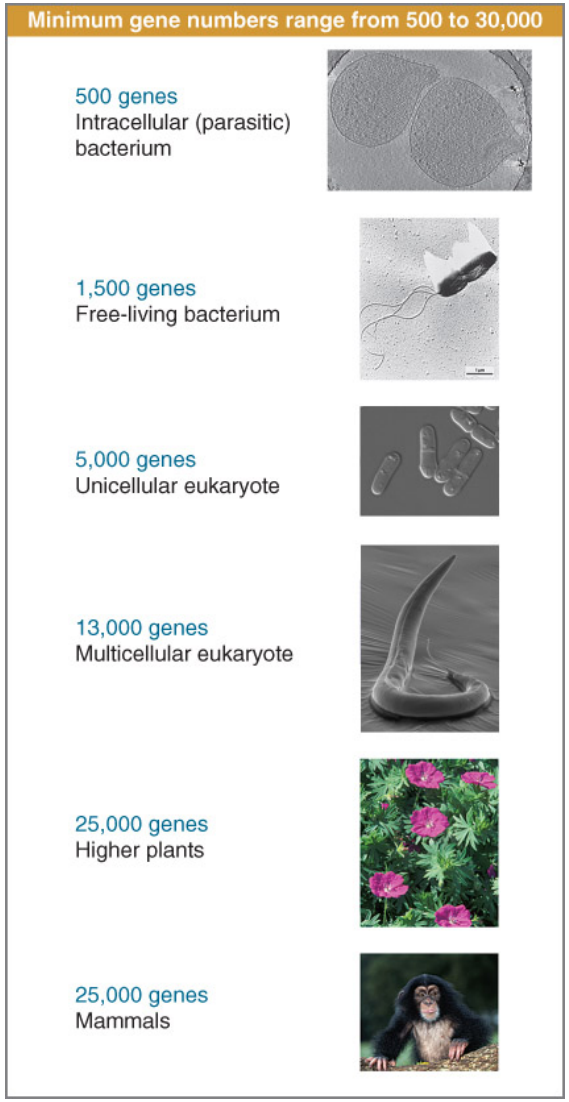
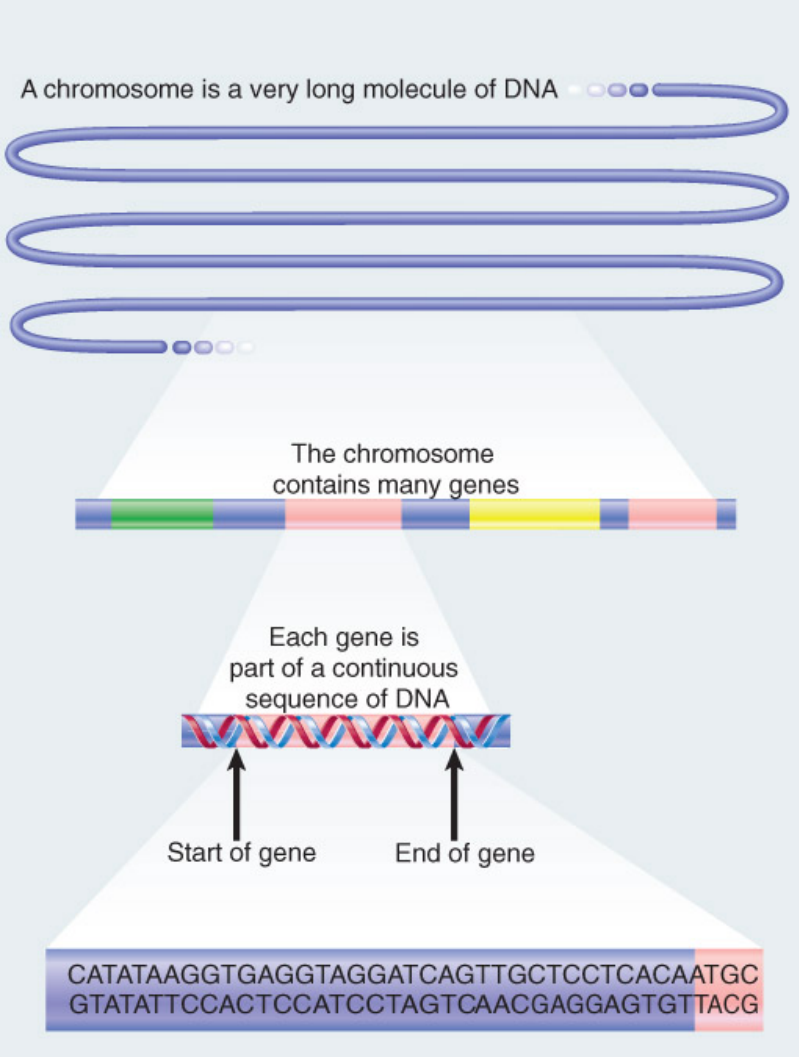


TABLE 8-2 Comparison of the Gene Density in Different Organisms' Genomes

Species	Genome Size (Mb)	Approximate Number of Genes	Gene Density (genes/Mb)
Prokaryotes (bacteria)			
Mycoplasma genitalium	0.58	500	860
Streptococcus pneumoniae	2.2	2300	1060
Escherichia coli K-12	4.6	4400	950
Agrobacterium tumefaciens	5.7	5400	960
Sinorhizobium meliloti	6.7	6200	930
Eukaryotes (animals)			
Fungi			
Saccharomyces cerevisiae	12	5800	480
Schizosaccharomyces pombe	12	4900	410
Protozoa			
Tetrahymena thermophila	125	27,000	220
Invertebrates			
Caenorhabditis elegans	103	20,000	190
Drosophila melanogaster	180	14,700	82
Ciona intestinalis	160	16,000	100
Locusta migratoria	5000	nd	nd
Vertebrates			
Fugu rubripes (pufferfish)	393	22,000	56
Homo sapiens	3200	20,000	6.25
Mus musculus (mouse)	2600	22,000	8.5
Plants			
Arabidopsis thaliana	120	26,500	220
Oryza sativa (rice)	430	~45,000	~100
Zea mays (corn)	2200	>45,000	>20
Triticum aestivum (wheat)	16,000	nd	nd
Fritillaria assyriaca (tulip)	~120,000	nd	nd

nd, Not determined.

IV. Du gène à la protéine



Allèle : une des formes que peut prendre un même gène

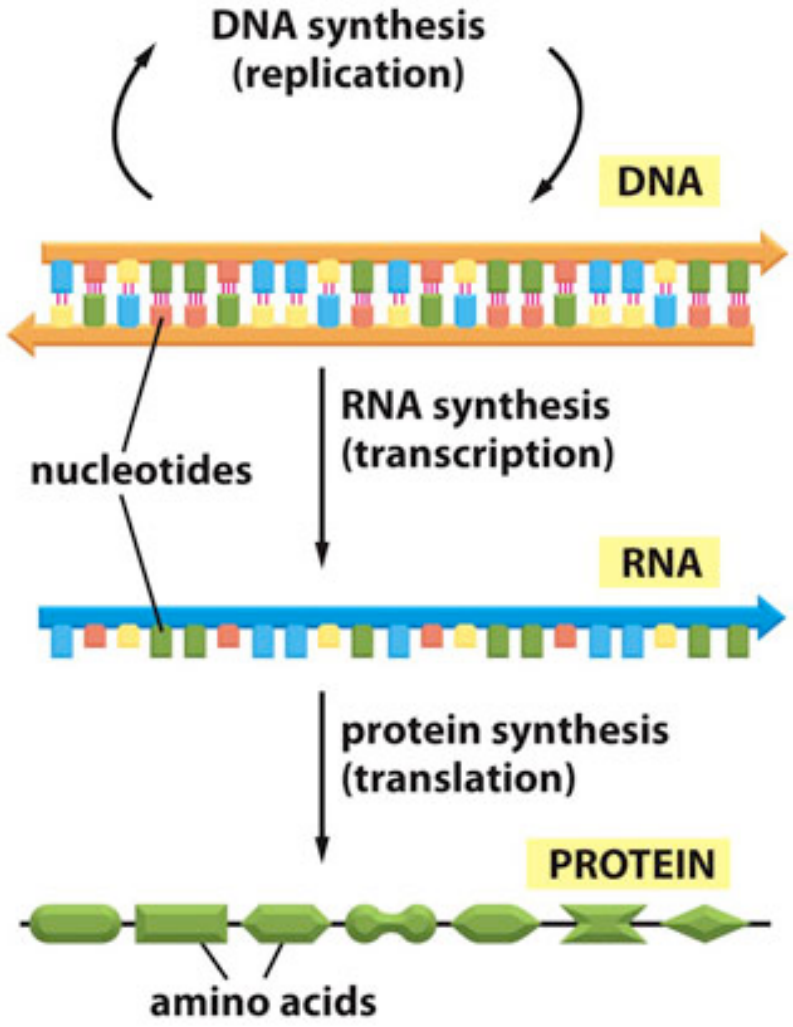
Locus : position occupée par un gène (ou ses allèles) sur le chromosome

Un locus peut avoir plusieurs allèles

Each allele has a different phenotype	
Allele	Phenotype of homozygote
w^+	red eye (wild type)
w^{bl}	blood
w^{ch}	cherry
w^{bf}	buff
w^h	honey
w^a	apricot
w^e	eosin
w^I	ivory
w^z	zeste (lemon-yellow)
w^{sp}	mottled, color varies
w^1	white (no color)

IV. Du gène à la protéine

Le dogme de la biologie moléculaire



Réplication : DNA polymérases ou RNA polymérase RNA-dépendantes

Transcription : RNA polymérarses, DNA-dépendantes

Traduction : Ribosome

V. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN

Mutations spontanées (erreurs de réplication)

Mutations induites (altérations chimiques)

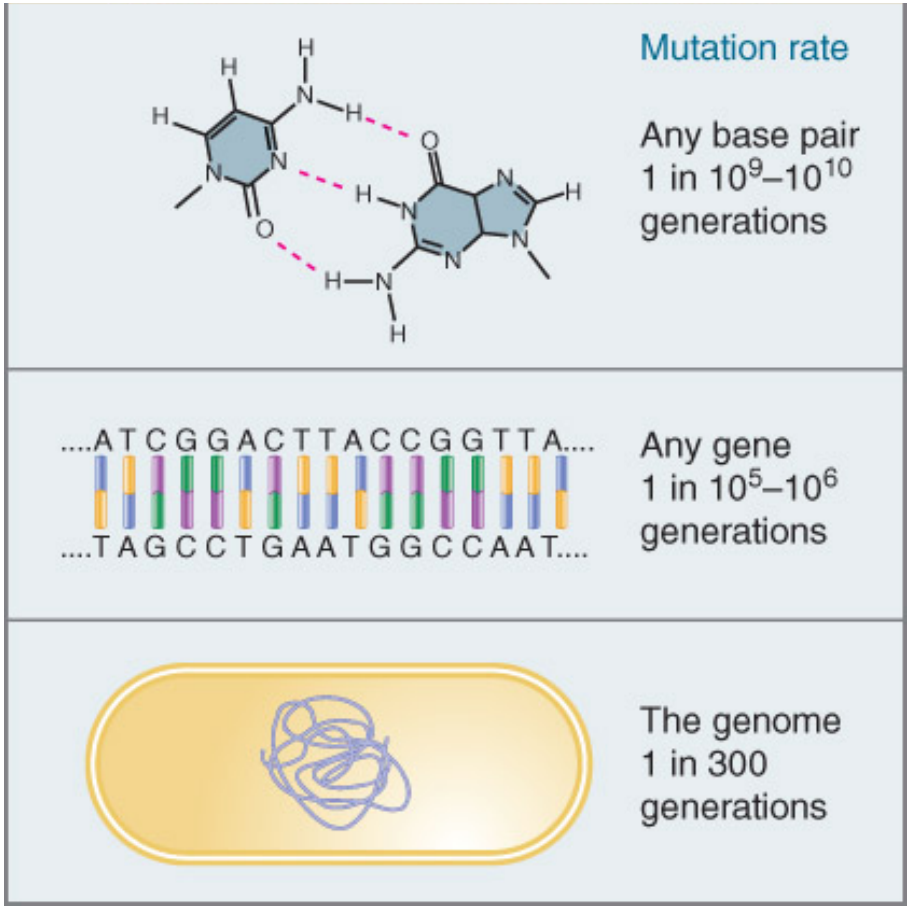
I**nsertion de matériel génétique exogène** (virus, transposons)

http://education-portal.com/academy/lesson/mutagens-how-the-environment-affects-mutation-rates.html

V. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN

Mutations spontanées



V. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN

Mutations spontanées

Organism	G	G_e	μ_b	μ_g
C. elegans	8.0×10^7	1.8×10^7	2.3×10^{-10}	0.018
Drosophila	1.7×10^8	1.6×10^7	3.4×10^{-10}	0.058
Mouse	2.7×10^9	8.0×10^7	1.8×10^{-10}	0.49
Human	3.2×10^9	8.0×10^7	5.0×10^{-11}	0.16

V. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN

Mutations induites et mutagènes

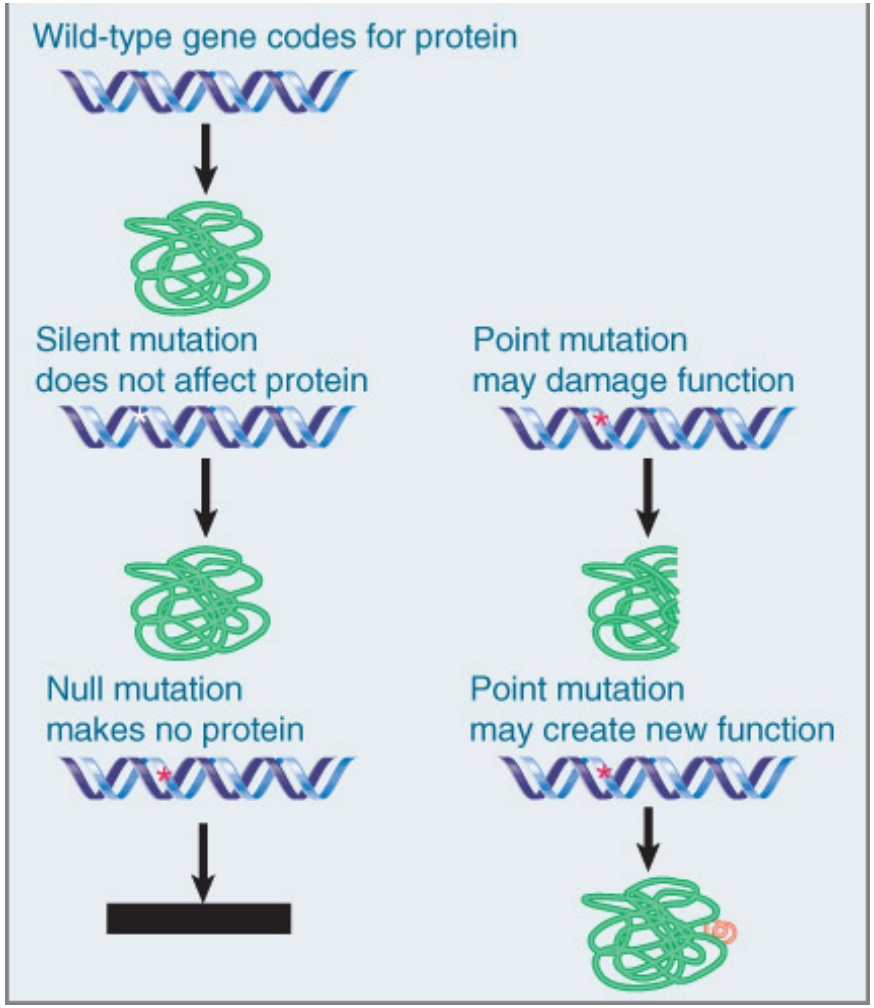
Table II. Partial list of chemicals evaluated with Mutatox and Ames Test for genotoxicity.

Compound	Mutatox	Ames
Aflatoxin B1	Positive	Positive
2-Aminoanthracene	Positive	Positive
2-Aminoflurorene	Positive	Positive
9-Aminoacridine	Positive	Positive
Benzene*	Positive	Negative
Benzidine	Positive	Positive
Benzoin*	Negative	Negative
Benzo(a)pyrene	Positive	Positive
Captan	Positive	Positive
2-Chloroethanol*	Positive	Positive
Cyclophosphamide	Positive	Positive
1,2-Dichloropropane	Positive	Positive
1,3-Dichloropropene	Negative	Positive
Dioxane	Negative	Negative
Ethylene glycol	Negative	Negative
8-Hydroxyquinoline*	Positive	Positive
Lindane	Negative	Negative
Monuron*	Postive	Negative
3-methylcholanthrene	Postive	Positive
Nalidixic acid	Positive	Negative
Pyrene	Postive	Negative

* Designated National Toxicology Program Chemical

V. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN



Les mutations peuvent être :

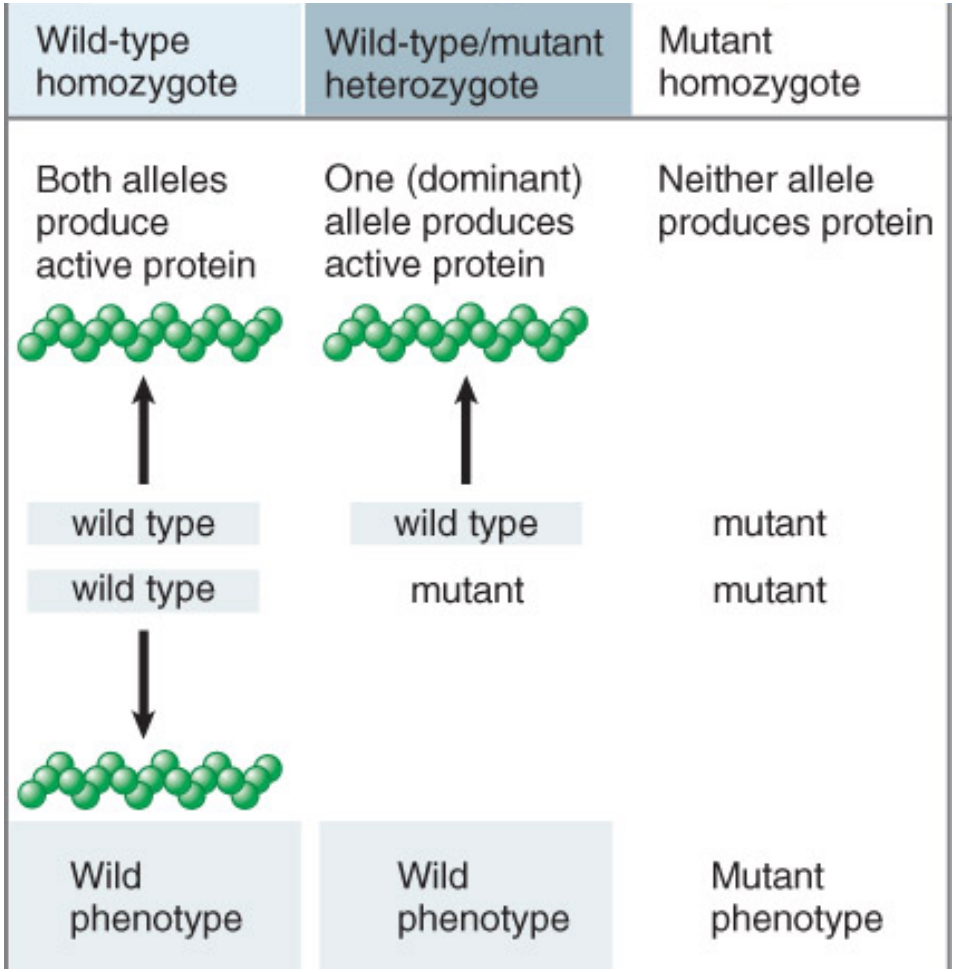
Silencieuses (pas d'effet)

Nulle (perte de fonction)

Nouvelle fonction (gain de fonction)

V. Les altérations du message

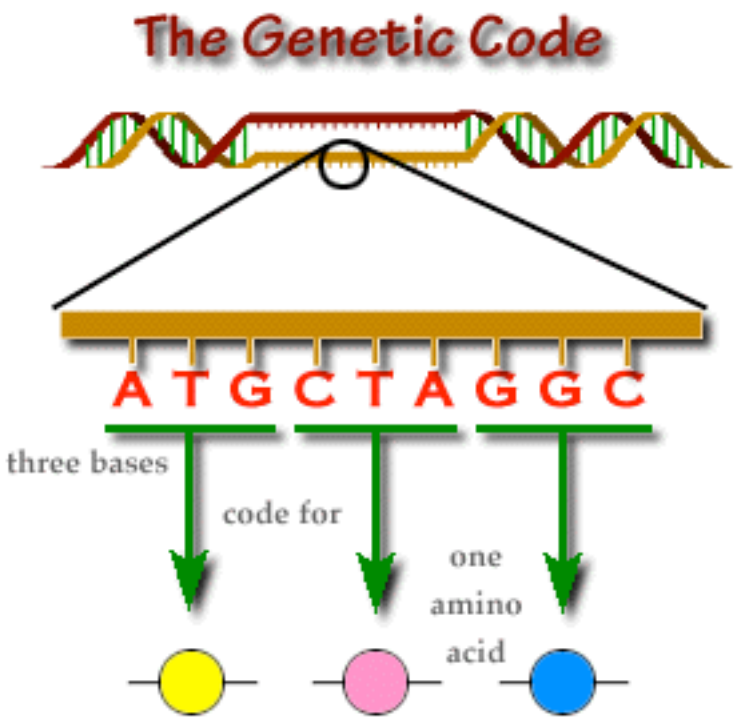
Les mutations sont des modifications de la séquence d'ADN



- Les mutations sont en général récessives
- Les allèles récessifs produisent des protéines inactives

V. Les altérations du message

Le code génétique

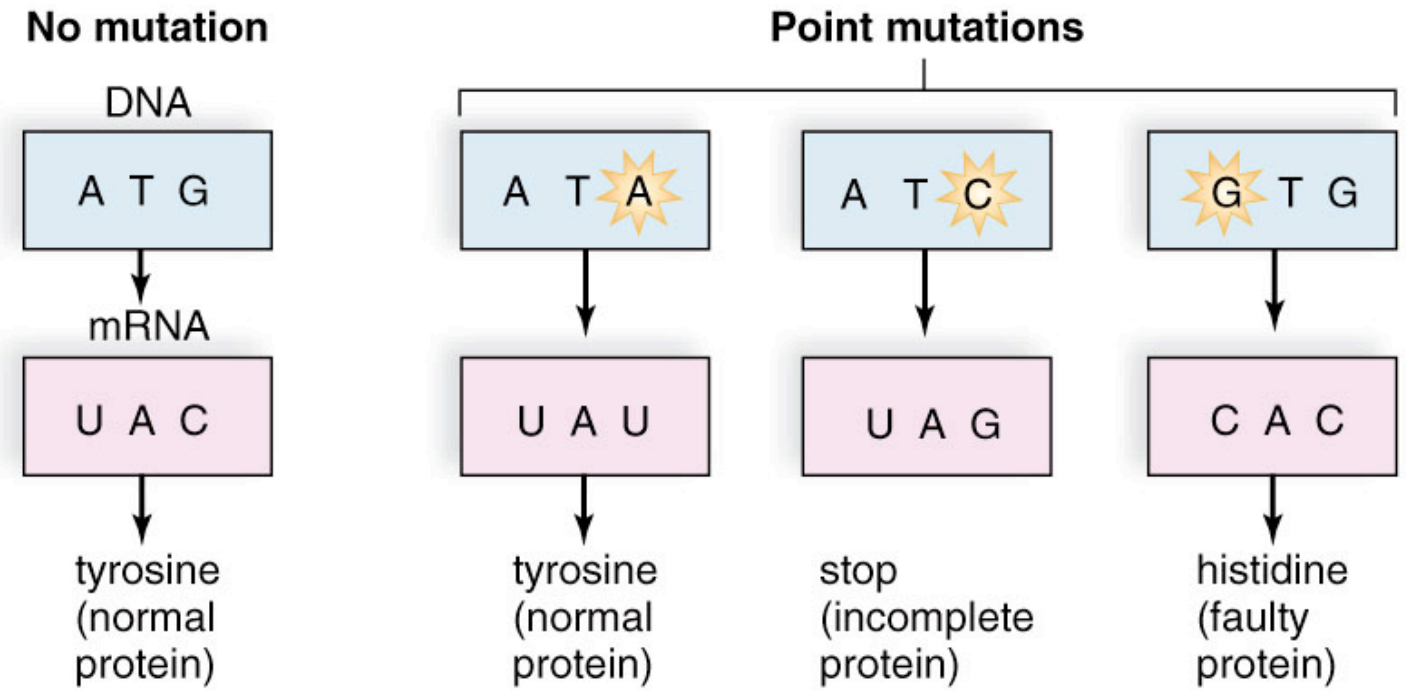


Code génétique : relation entre séquence d'ADN et séquence d'acides aminés

Information génétique : gènes + séquences régulatrices

V. Les altérations du message

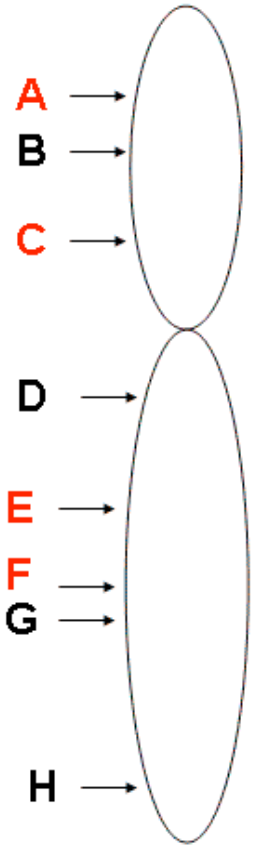
Le code génétique



Les substitutions peuvent ou non induire une changement dans le polypeptide

VI. Le décodage du génome

La cartographie génique



Genetic Maps

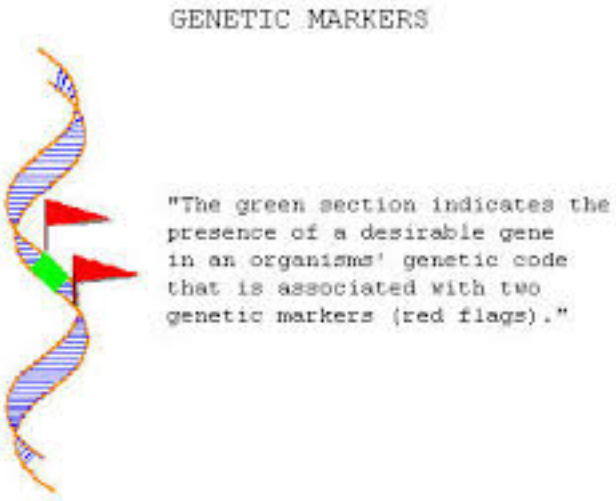
- Genes are shown in relative order and distance from each other based on pedigree studies.
- The chance of the chromosome breaking between A & C is higher than the chance of the chromosome breaking between A & B during meiosis.
- Similarly, the chance of the chromosome breaking between E & F is higher than the chance of the chromosome breaking between F & G.
- The closer two genes are, the more likely they are to be inherited together (co-occurrence).
- If pedigree studies show a high incidence of co-occurrence, those genes will be located close together on a genetic map.

VI. Le décodage du génome

Les marqueurs génétiques/moléculaires :

Définition : fragment (séquence) d'ADN repérable spécifiquement et présentant un polymorphisme chez les individus

Caractéristiques : Neutre phénotypiquement, multiallèlique, universel et transmissible

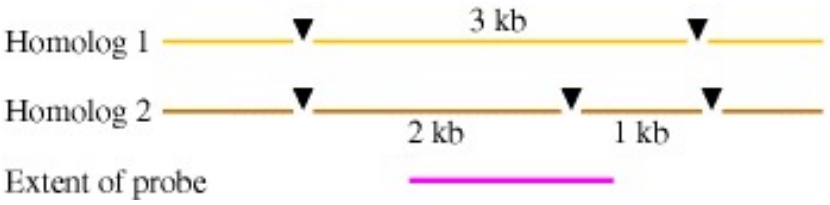


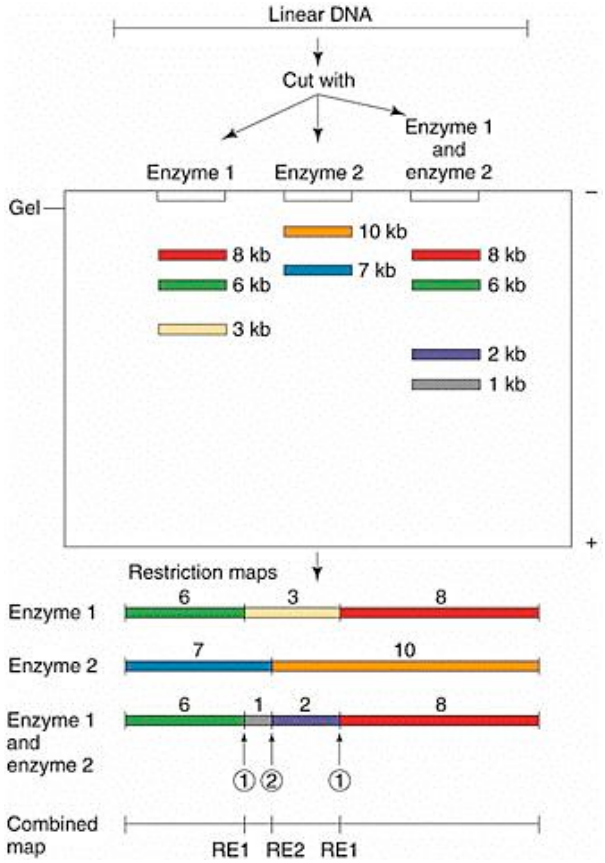
Utilité des marqueurs génétiques :

http://www.youtube.com/watch?v=nrnJPC6e19c

V. Le décodage du génome

RFLP : Restriction length fragment polymoprhism



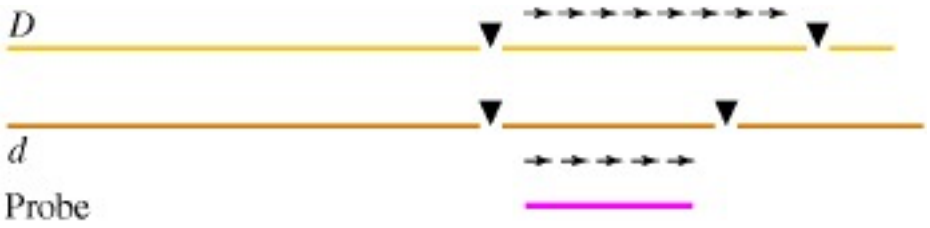


Gene coupling

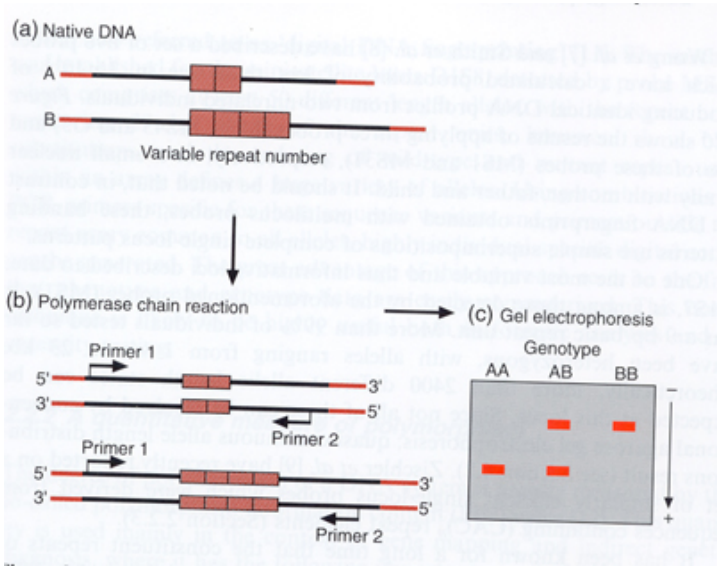
V. Le décodage du génome

STRs : Short tandem repeats (VNTR)

Détecté par RFLP :



Détecté par AFLP :



https://www.youtube.com/watch?v=DbR9xMXuK7c

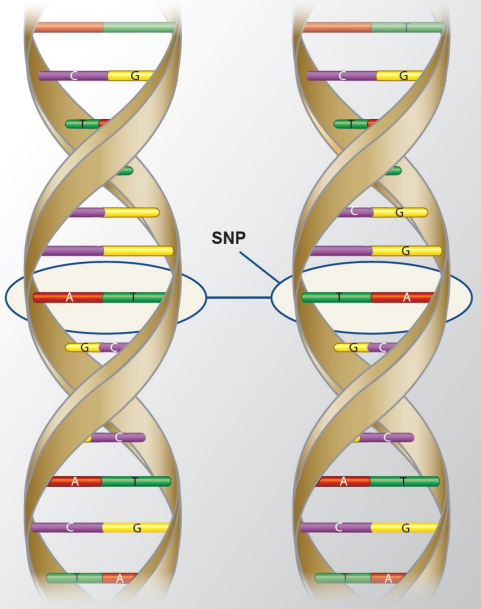
V. Le décodage du génome

SNP : Single nucleotide polymorphism

What is a Single Nucleotide Polymorphism (SNP)?

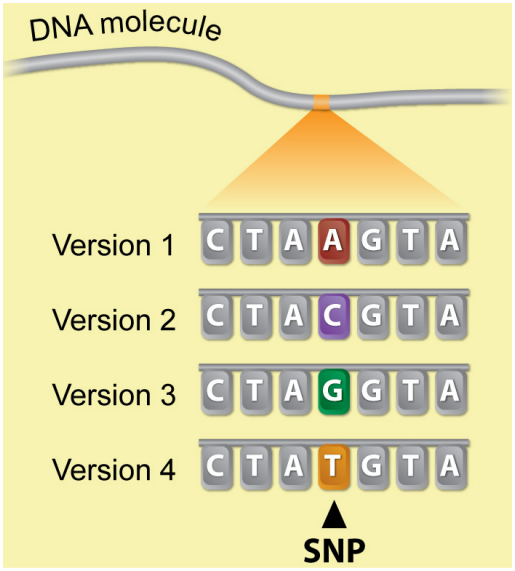
A SNP (pronounced "snip") is a DNA sequence variation that occurs when a single nucleotide (A, T, C, or G) in the genome sequence is modified.

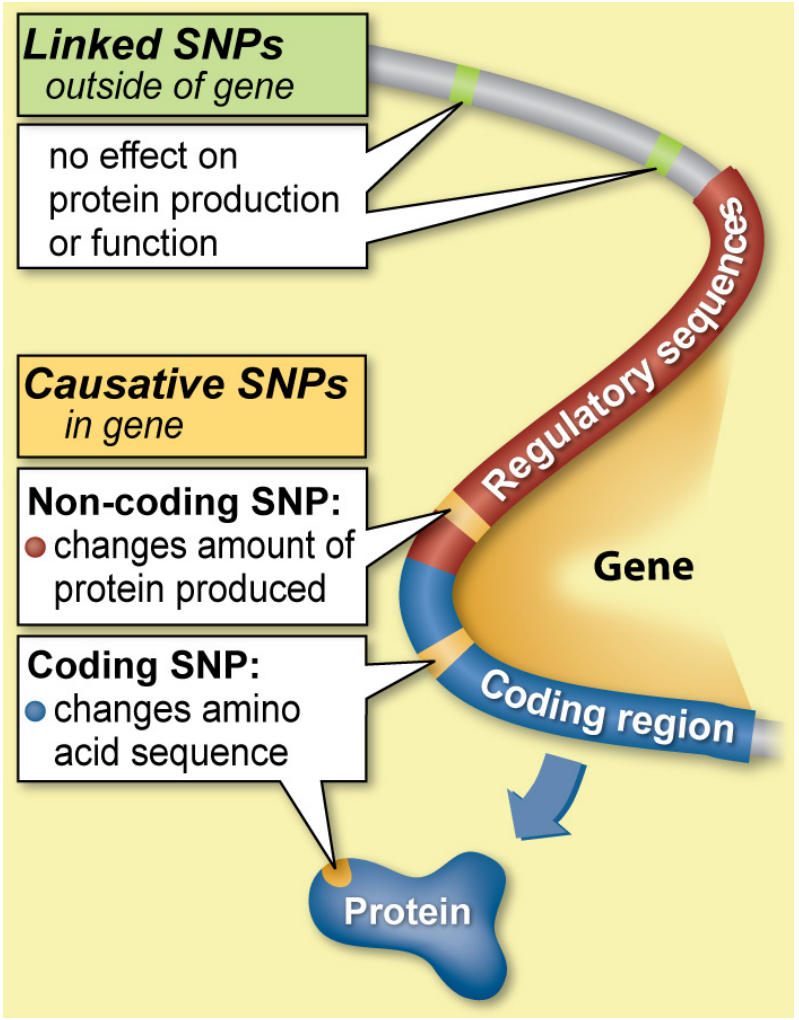
SNPs do not necessarily cause disease, but they can help determine the likelihood that someone will develop a particular illness.



V. Le décodage du génome

SNP : Single nucleotide polymorphism





V. Le décodage du génome

SNP : Single nucleotide polymorphism