



# 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

**Génétique** : science qui étudie les caractères héréditaires des individus, leur transmission au fil des générations et leur variations (mutations).

# I - La naissance de la Biologie Moléculaire

- L'émergence de la génétique formelle



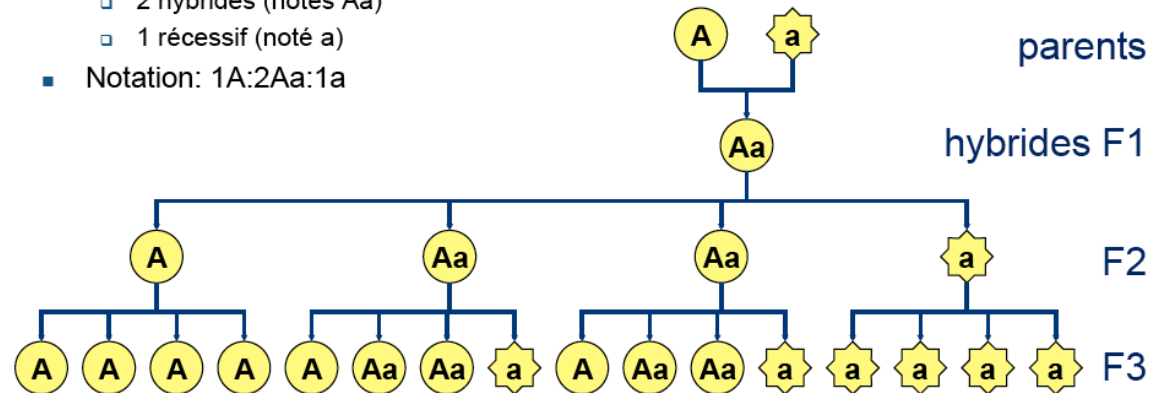
Gregor Mendel (1822 - 1884)

les Lois de l'hérédité (1866)

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

*Loi n°2 : Ségrégation des caractères*

- 2 hybrides (notés Aa)
- 1 récessif (noté a)
- Notation: 1A:2Aa:1a



**La notion de gènes et d'allèles**

# I - La naissance de la Biologie Moléculaire

- L'émergence de la génétique formelle Gregor Mendel (1822 - 1884)
- Le chromosome, support de l'hérédité (1910)



Thomas Morgan (1866 - 1945)

Théorie chromosomique de l'hérédité



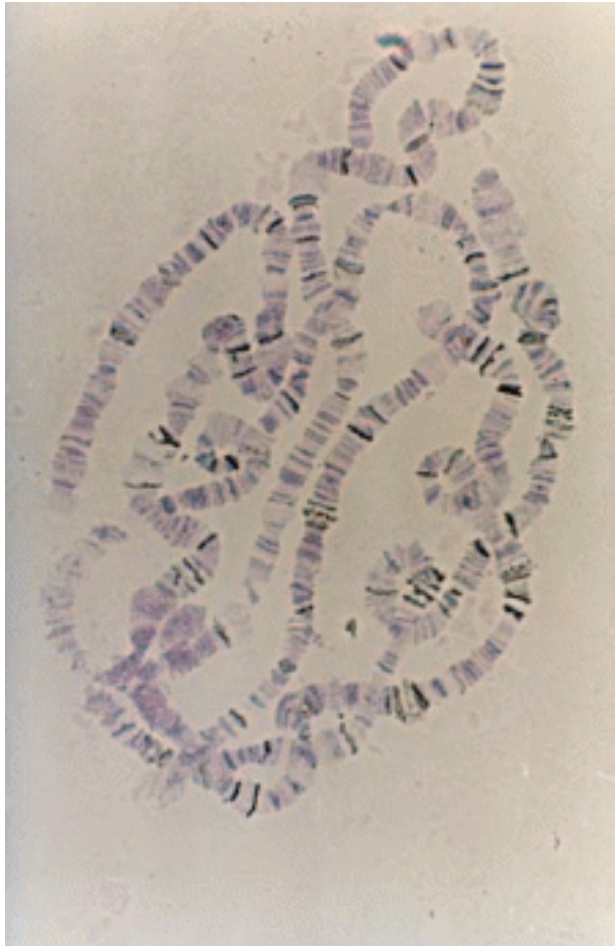
*Drosophila melanogaster*



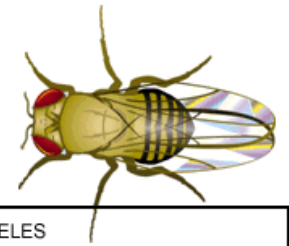
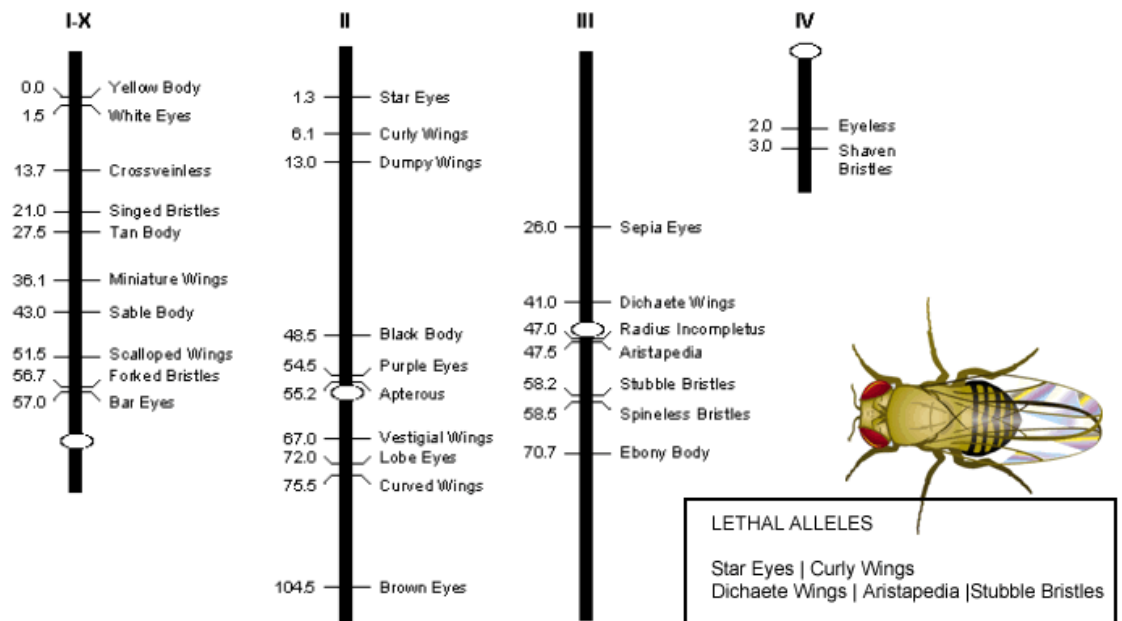
*Chromosomes polytènes*



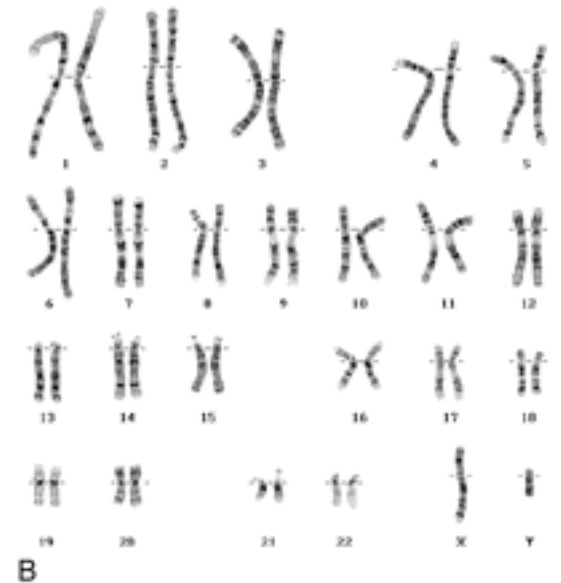
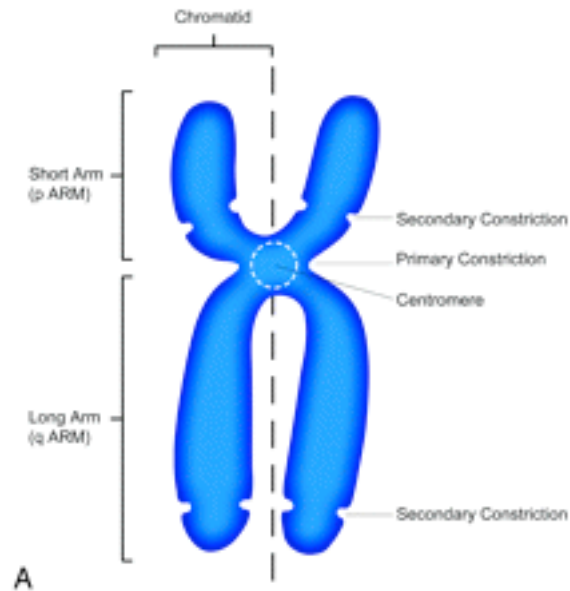
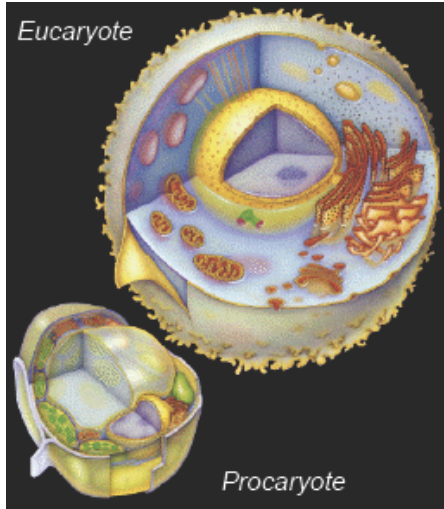
- Le chromosome, support de l'hérédité (Thomas Morgan)



Drosophila Chromosome Map



- Le chromosome, support de l'hérédité



# I - La naissance de la Biologie Moléculaire

- L'émergence de la génétique formelle
- Le chromosome, support de l'hérédité
- L'ADN, support de l'hérédité
- La composition et structure de l'ADN

Gregor Mendel (1822 - 1884)

Thomas Morgan (1866 - 1945)  
Alfred Sturtevant (1891 - 1970)

Fred Griffith (1877 - 1941)  
Oswald Avery (1877 - 1955)

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



Erwin Chargaff (1905 - 1992)

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

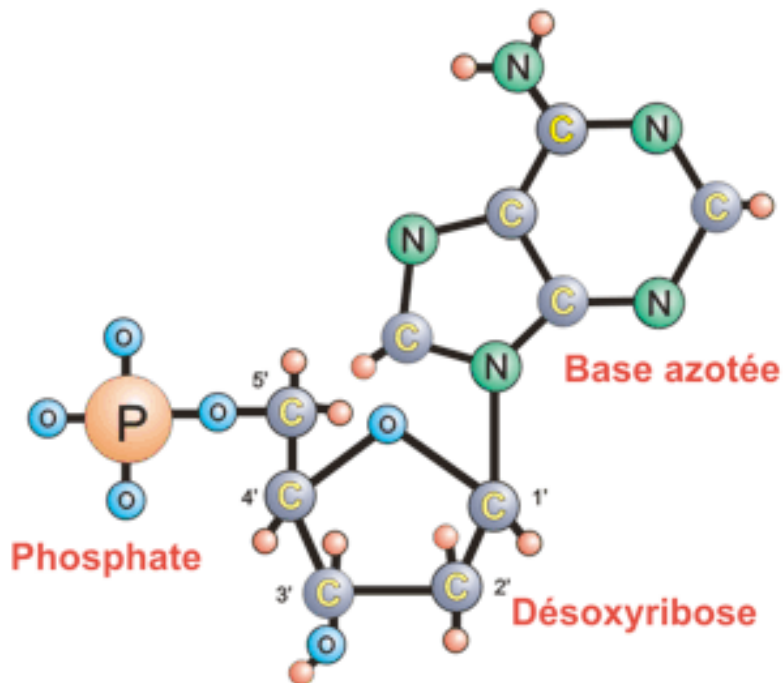
$$A / T = C / G = 1$$

## II – Structure des acides nucléiques

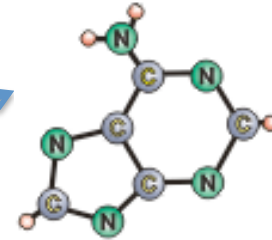
### II.I Les molécules simples

L'acide désoxyribonucléique (ADN)

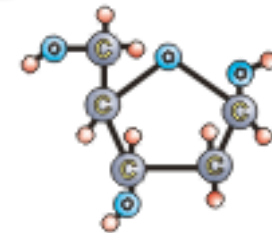
= polymères de sous-unités appelées **nucléotides**



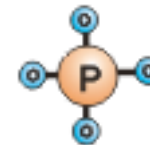
une **BASE**



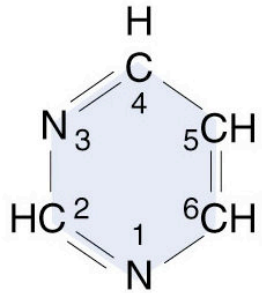
un **SUCRE**



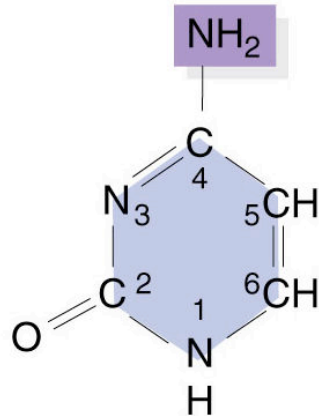
un **ACIDE PHOSPHORIQUE**



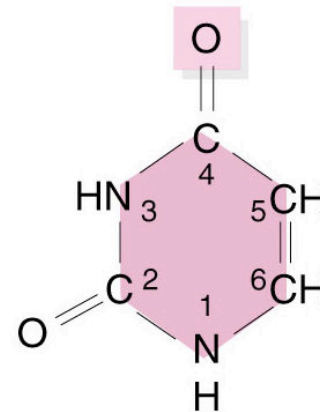
## II.I.1 Les bases azotées



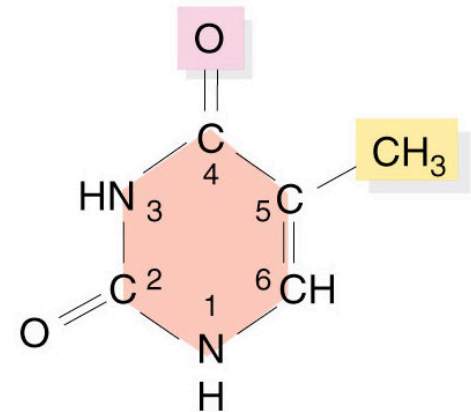
**Pyrimidine**



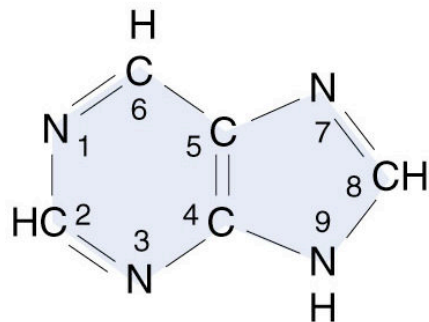
**Cytosine (C)**  
*2-oxy-4-amino-pyrimidine*



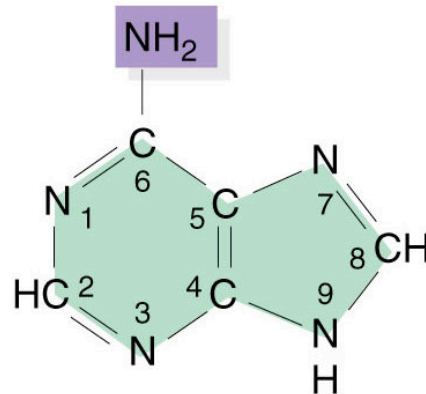
**Uracil (U)**  
**(found in RNA)**  
*2,4-dioxy-pyrimidine*



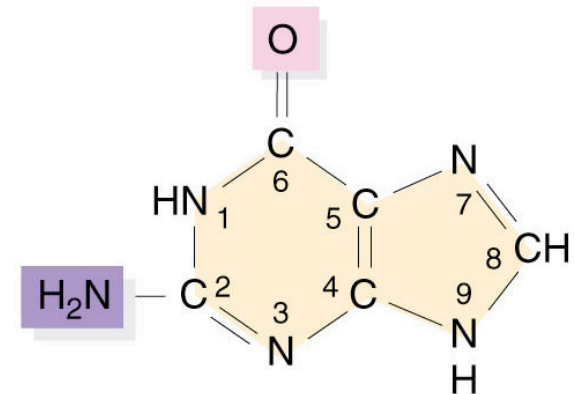
**Thymine (T)**  
**(found in DNA)**  
*2,4-dioxy-5-méthyl-pyrimidine*



**Purine**



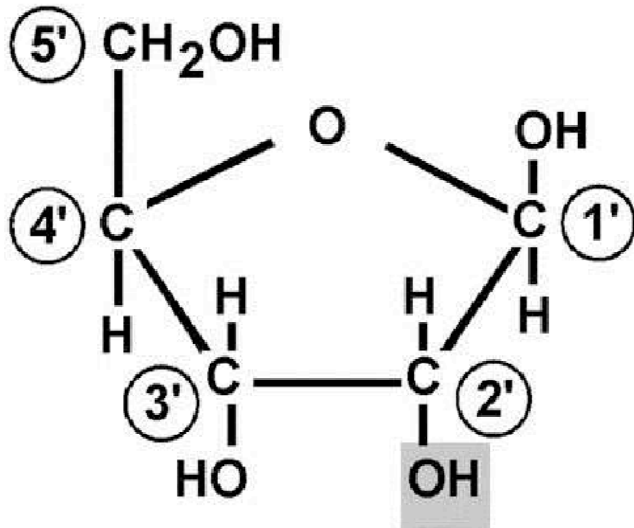
**Adenine (A)**  
*6-amino-purine*



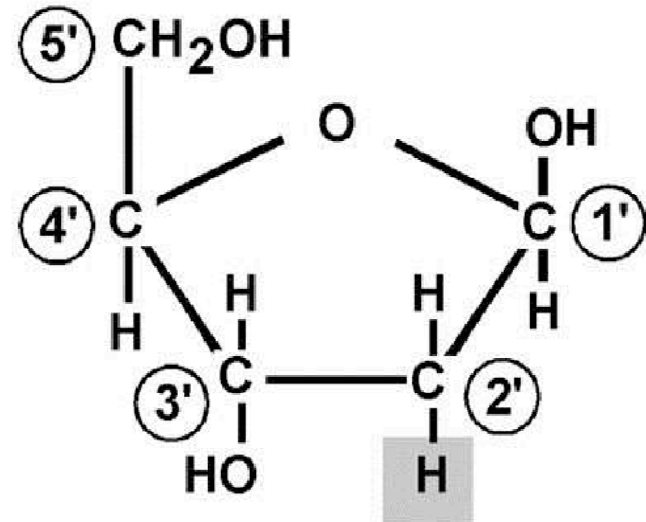
**Guanine (G)**  
*2-amino-6-oxy-purine*



## II.1.2 Les sucres



Pentose présent dans l'ARN :  
le ribose

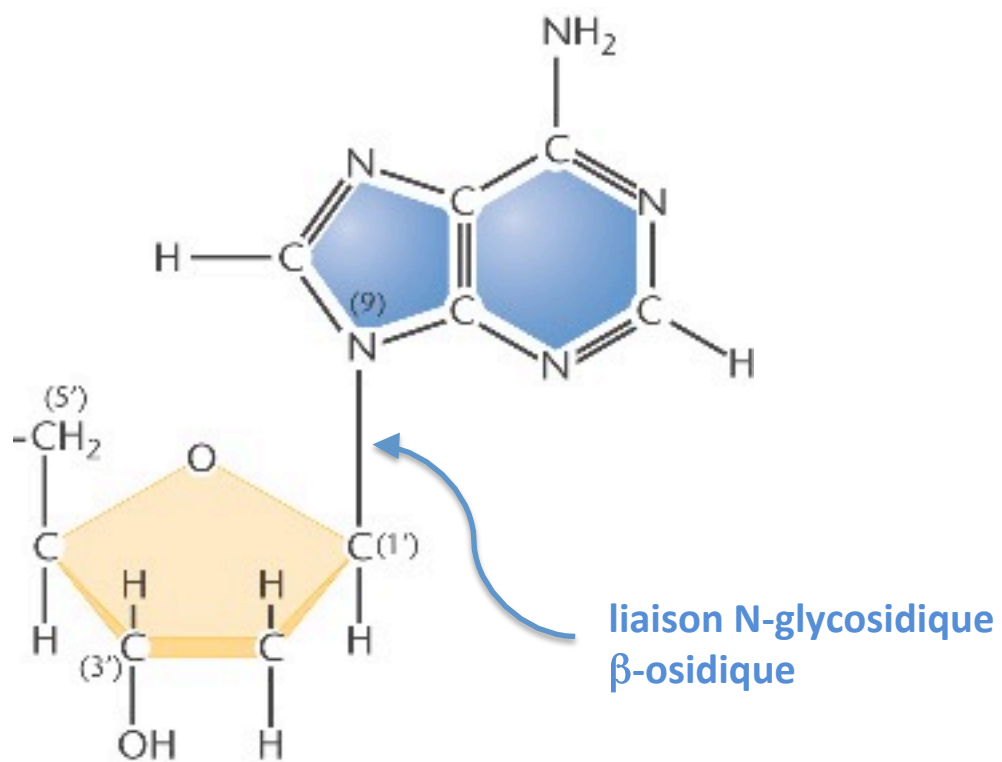
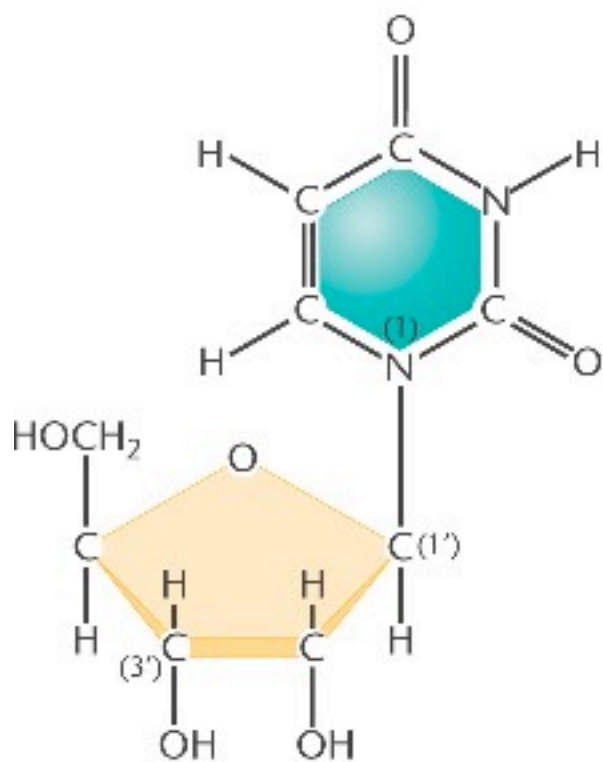


Pentose présent dans l'ADN :  
le désoxyribose

## II.2 Les nucléosides

### Les nucléosides

SUCRE + BASE

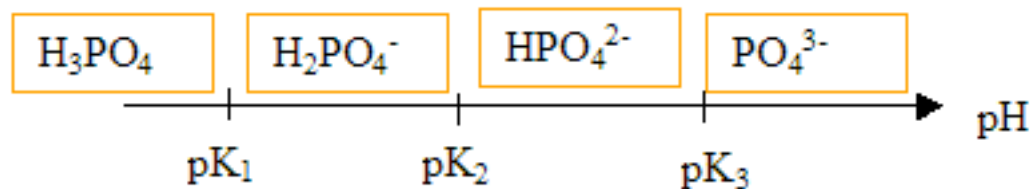
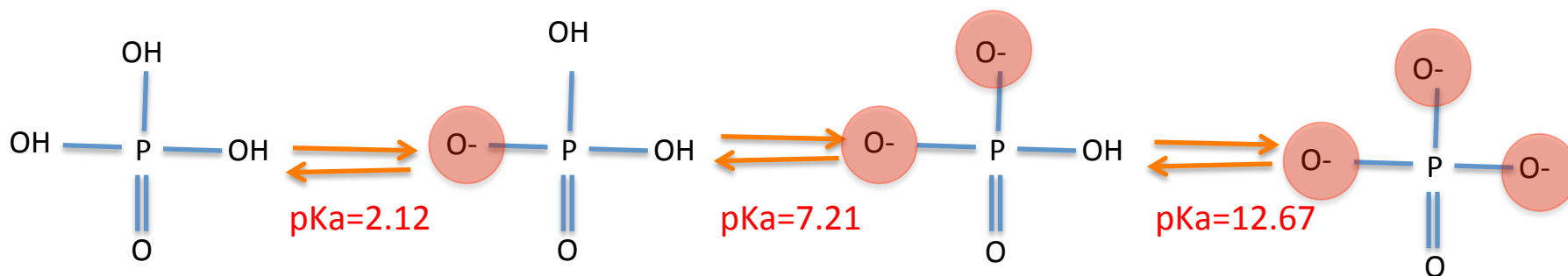
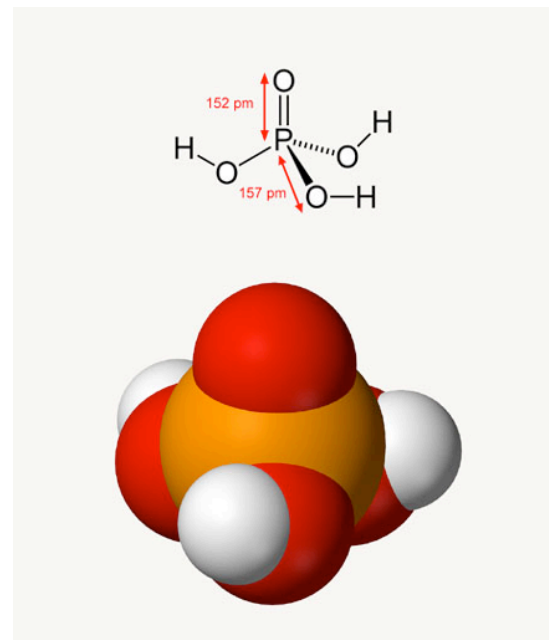


## II.2 Les nucléosides

	<b>Base</b>	<b>Ribonucleoside</b>	<b>Desoxyribonucleoside</b>
Purines	Adénine Guanine	<b>Adénosine</b> <b>Guanosine</b>	Désoxyadénosine Désoxyguanosine
Pyrimidiques	Uracile Cytosine Thymine	<b>Uridine</b> <b>Cytidine</b> Thymine ribonucléoside	Désoxyuridine Désoxycytidine Désoxythymidine ou thymidine

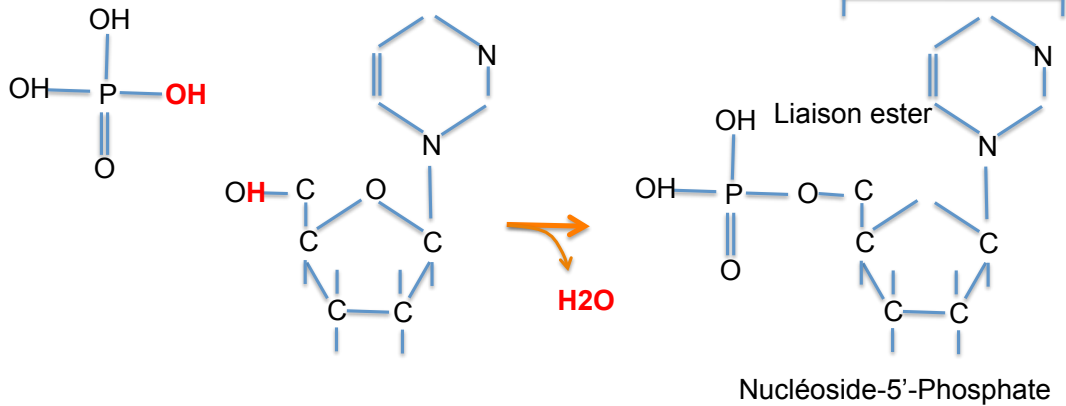
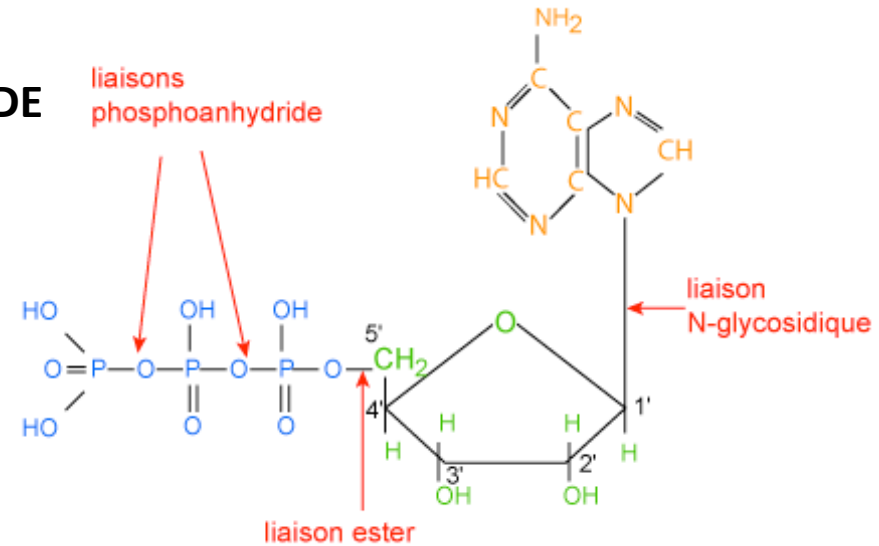
## II.3 L'acide phosphorique

### L'acide phosphorique



## II.4 Les nucléotides

**SUCRE + BASE+ PHOSPHATE = NUCLEOTIDE**





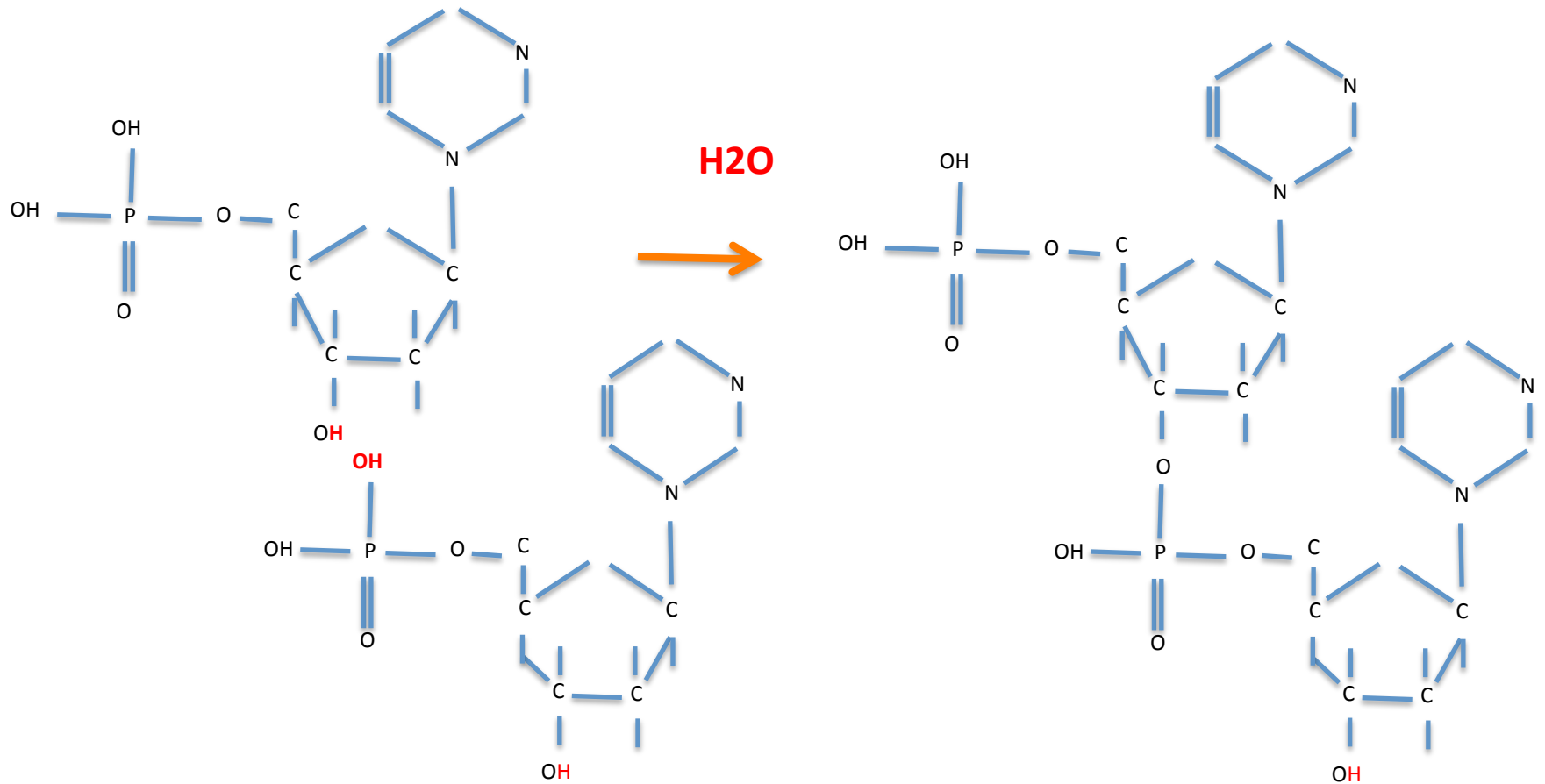
## II.4 Les nucléotides

### Les nucléotides : nucléosides-5'-Monophosphate

Base	Ribonucleotide	Desoxyribonucleotide
Adénine	Adénosine-5'-Monophosphate (AMP)	Déoxyadénosine-5'-Monophosphate (dAMP)
Guanine	Guanosine-5'-Monophosphate (GMP)	Déoxyguanosine-5'-Monophosphate (dGMP)
Uracile	Uridine-5'-Monophosphate (UMP)	Déoxyuridine-5'-Monophosphate (dUMP)
Cytosine	Cytidine-5'-Monophosphate (CMP)	Déoxycytidine-5'-Monophosphate (dCMP)
Thymine	Thymine riboside -5'-Monophosphate (TMP)	Déoxythymidine -5'-Monophosphate (dTMP)

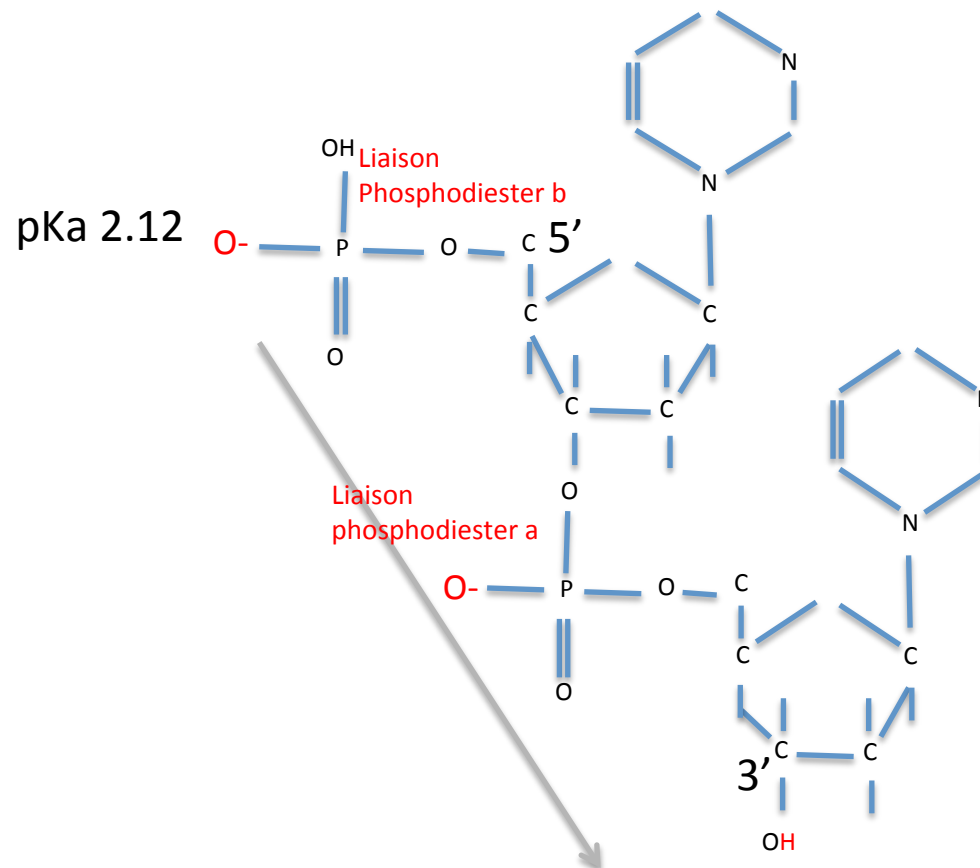
## II.5 La liaison phosphodiester

La chaîne polynucléotidique



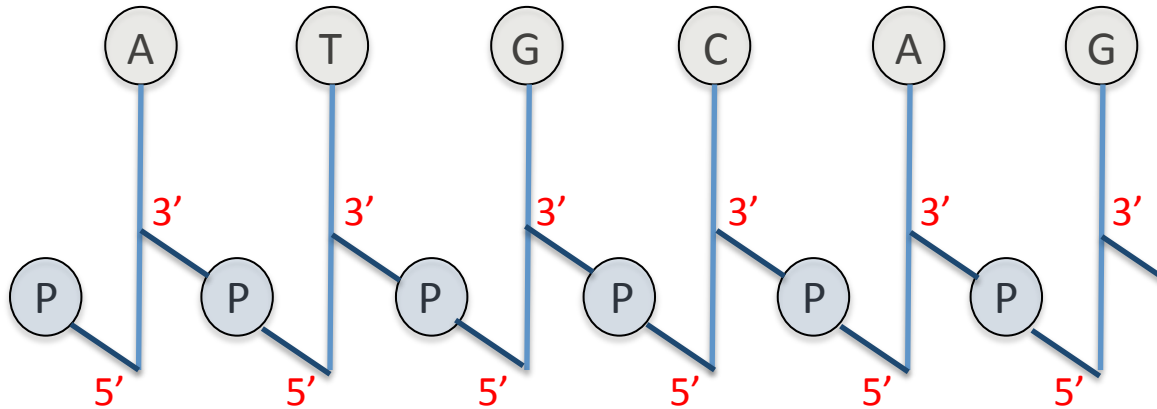
## II.5 La liaison phosphodiester

La chaîne polynucléotidique



## II.5 La liaison phosphodiester

La chaîne polynucléotidique

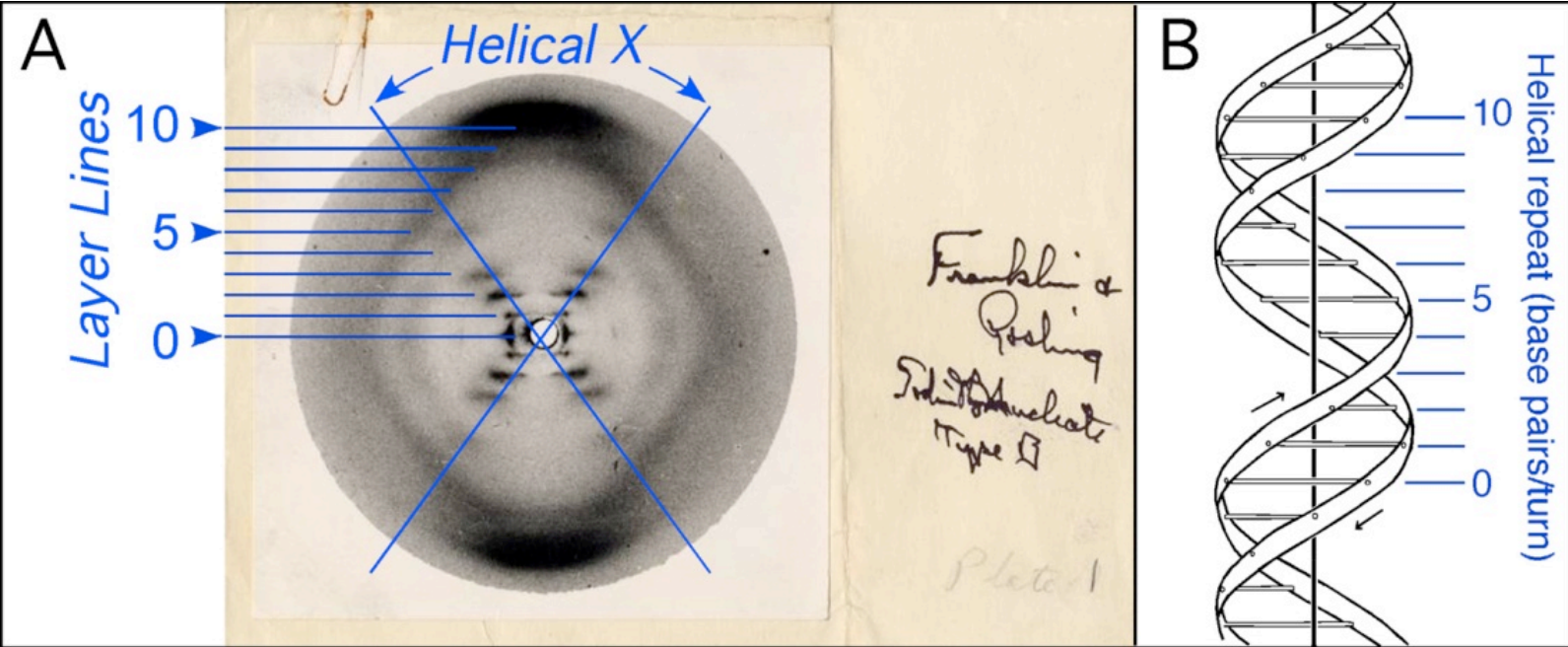


p5'dAp5'dTp5'dGp5'dCp5'dAp5'dG

ou 5'-dATGCAG-3'  
5'-ATGCAG-3'

# II.6 Structure Secondaire de la molécule d'ADN

Structure secondaire : le modèle de Watson et Crick

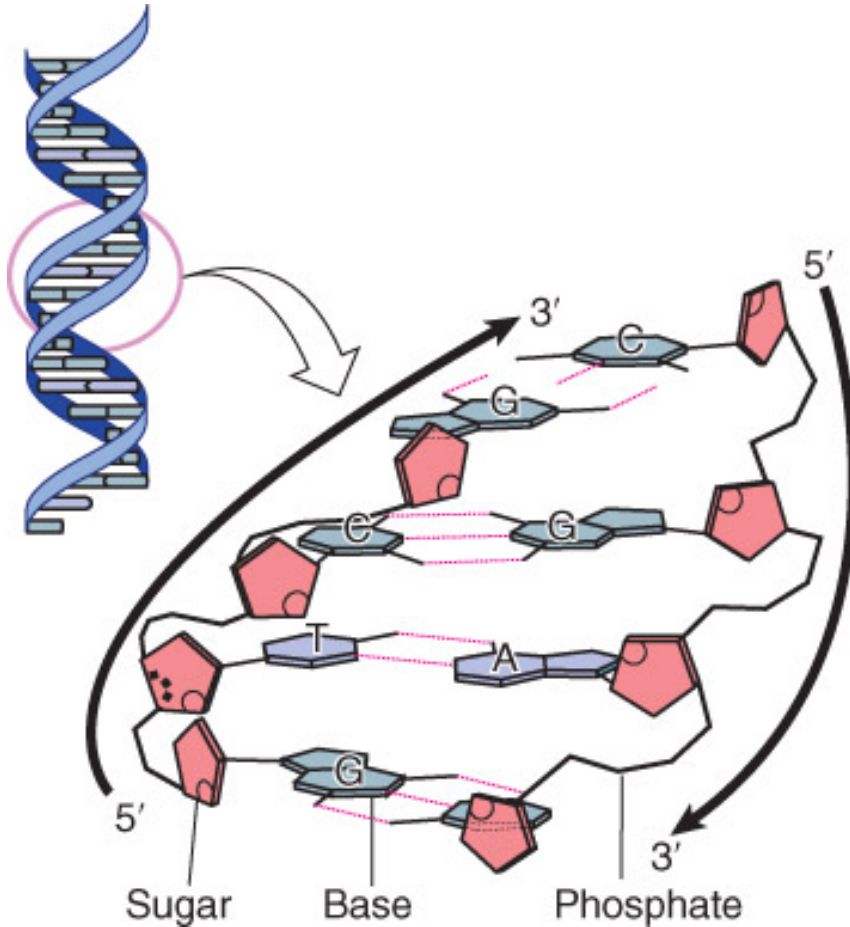


Rosalind Franklin



## II.6 Structure Secondaire de la molécule d'ADN

La molécule d'ADN : structure secondaire

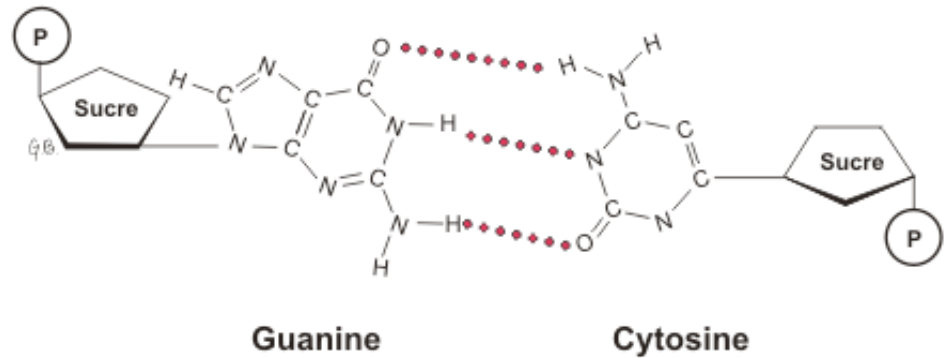
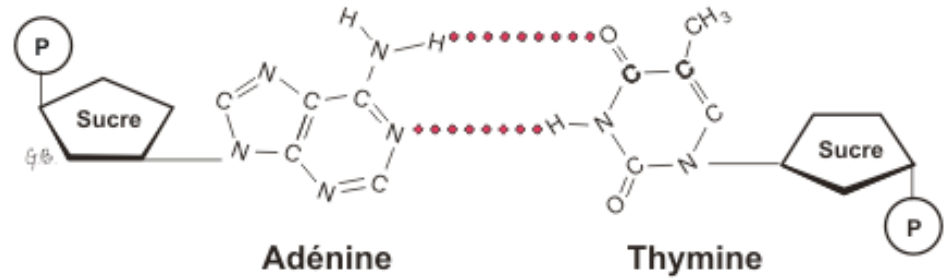
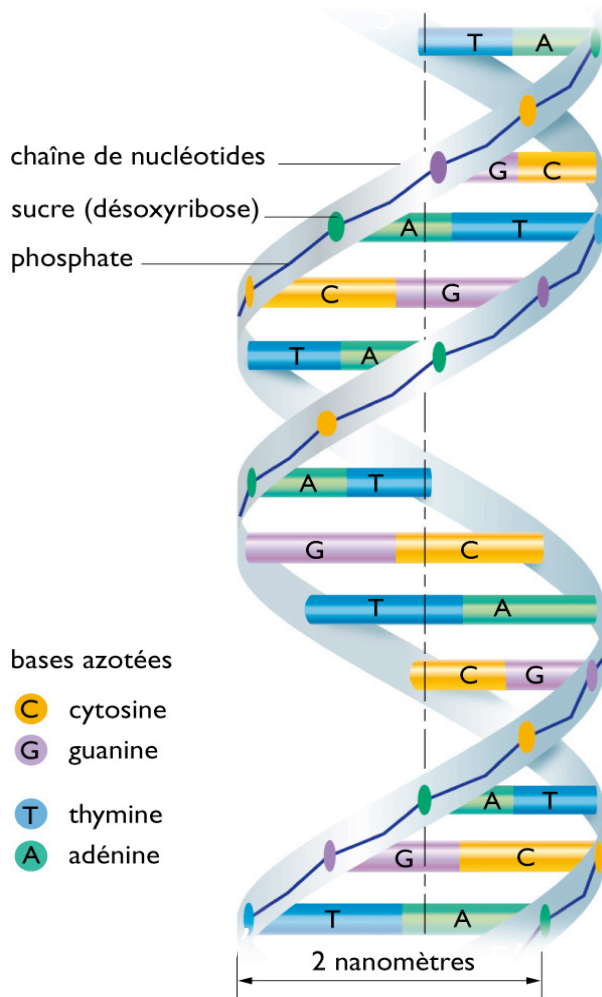
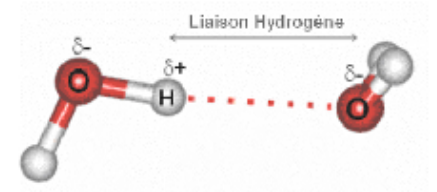


- **Bicaténaire** (2 chaînes polynucléotidiques)
- **Antiparallèle**
- **Hélicoïdale**
- Bases (intérieur) Sucres/phosphates (extérieur)
- **Complémentaire**

## II.6 Structure Secondaire de la molécule d'ADN

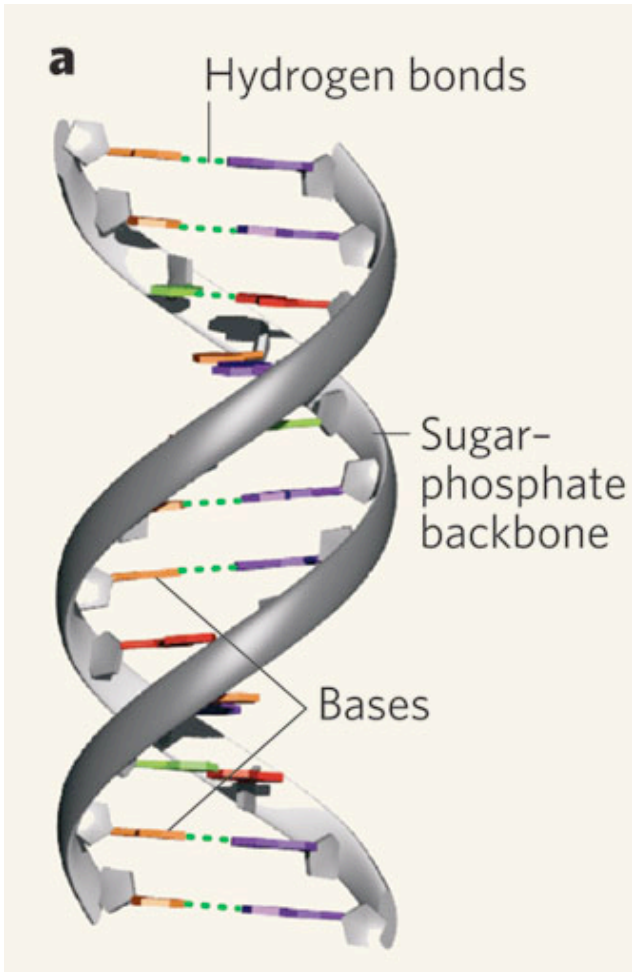
### La molécule d'ADN : structure secondaire

### 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>

## Exercice

La quantité de bases azotées d'une molécule d'ADN monocaténaire a été calculée. Le contenu en bases G--C de cette molécule est de 39,8%.

1. Déduisez les pourcentages respectifs de chaque base azotée (A, T, G, C).
2. Proposez une séquence de 20 nucléotides issue d'une molécule d'ADN bicaténaire respectant ces pourcentages.

### 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..) : **Chromosome**

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 $\mu\text{m}$	One single-stranded RNA	2 $\mu\text{m}$ = 6.4 kb
Phage fd	filament	0.006 x 0.85 $\mu\text{m}$	One single-stranded DNA	2 $\mu\text{m}$ = 6.0 kb
Adenovirus	icosahedron	0.07 $\mu\text{m}$ diameter	One double-stranded DNA	11 $\mu\text{m}$ = 35.0 kb
Phage T4	icosahedron	0.065 x 0.10 $\mu\text{m}$	One double-stranded DNA	55 $\mu\text{m}$ = 170.0 kb
<i>E. coli</i>	cylinder	1.7 x 0.65 $\mu\text{m}$	One double-stranded DNA	1.3 mm = 4.2 x 10 <sup>3</sup> kb
Mitochondrion (human)	oblate spheroid	3.0 x 0.5 $\mu\text{m}$	~10 identical double-stranded DNAs	50 $\mu\text{m}$ = 16.0 kb
Nucleus (human)	spheroid	6 $\mu\text{m}$ diameter	46 chromosomes of double-stranded DNA	1.8 m = 6 x 10 <sup>6</sup> kb

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



### III. Structure du génome

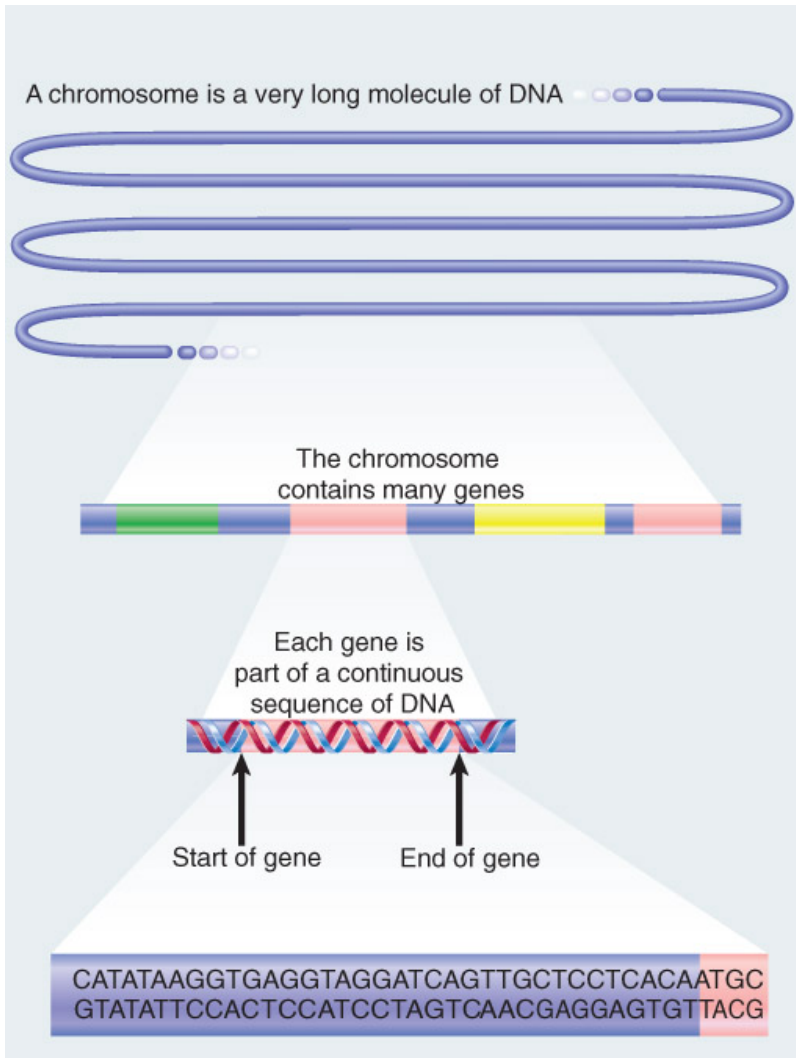
Le génome de chaque organisme est organisé en un nombre spécifique de chromosomes, de forme caractéristique

**TABLE 8-1** Variation in Chromosome Makeup in Different Organisms

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

### III. Structure du génome

#### Notion de gène



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

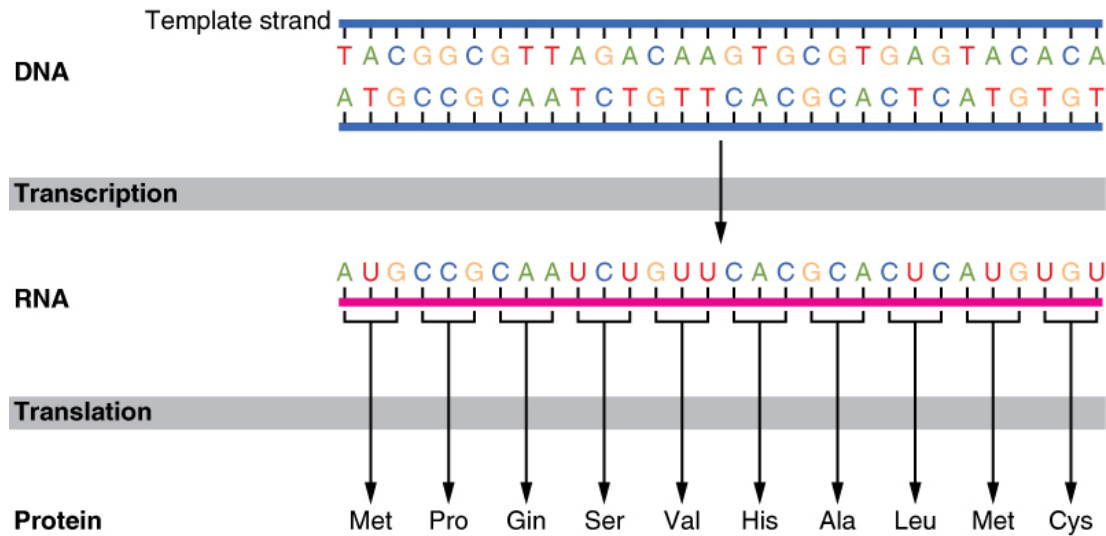
Each allele has a different phenotype	
Allele	Phenotype of homozygote
$w^+$	red eye (wild type)
$w^{bl}$	blood
$w^{ch}$	cherry
$w^{bf}$	buff
$w^h$	honey
$w^a$	apricot
$w^e$	eosin
$w^l$	ivory
$w^z$	zeste (lemon-yellow)
$w^{sp}$	mottled, color varies
$w^1$	white (no color)

# Chapitre II – Le gène

## II.4 Le code génétique

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

Information génétique : gènes + séquences régulatrices



## II.4 Le code génétique

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

**Troisi me lettre : wobble**

### III. Structure du génome

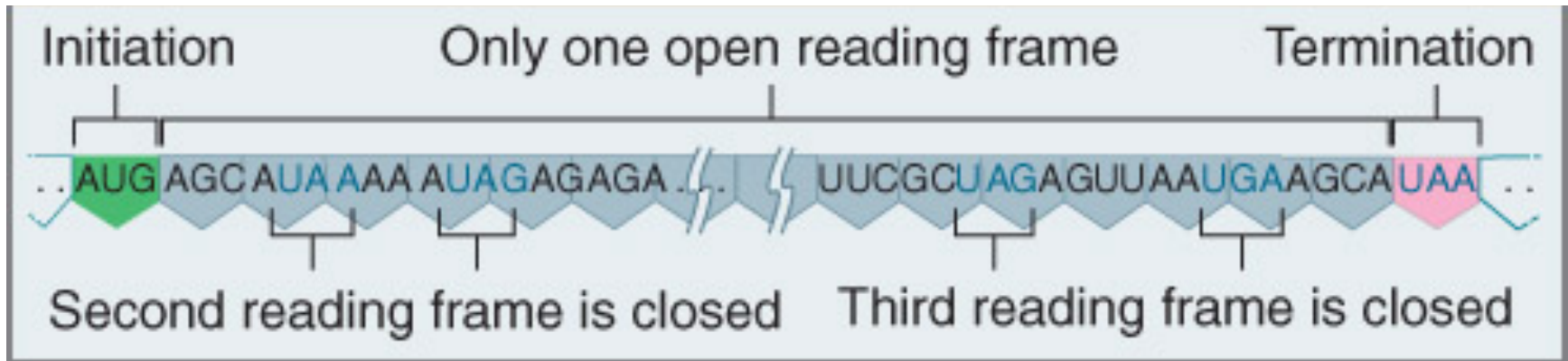
#### Le code génétique : La phase de lecture

5'-GGCAUCAAGUGCAGGCCCGU-3'  
└──┬──┬──┬──┬──┬──┘

5'-GGCAUCAAGUGCAGGCCCGU-3'  
└──┬──┬──┬──┬──┬──┘

5'-GGCAUCAAGUGCAGGCCCGU-3'  
└──┬──┬──┬──┬──┬──┘

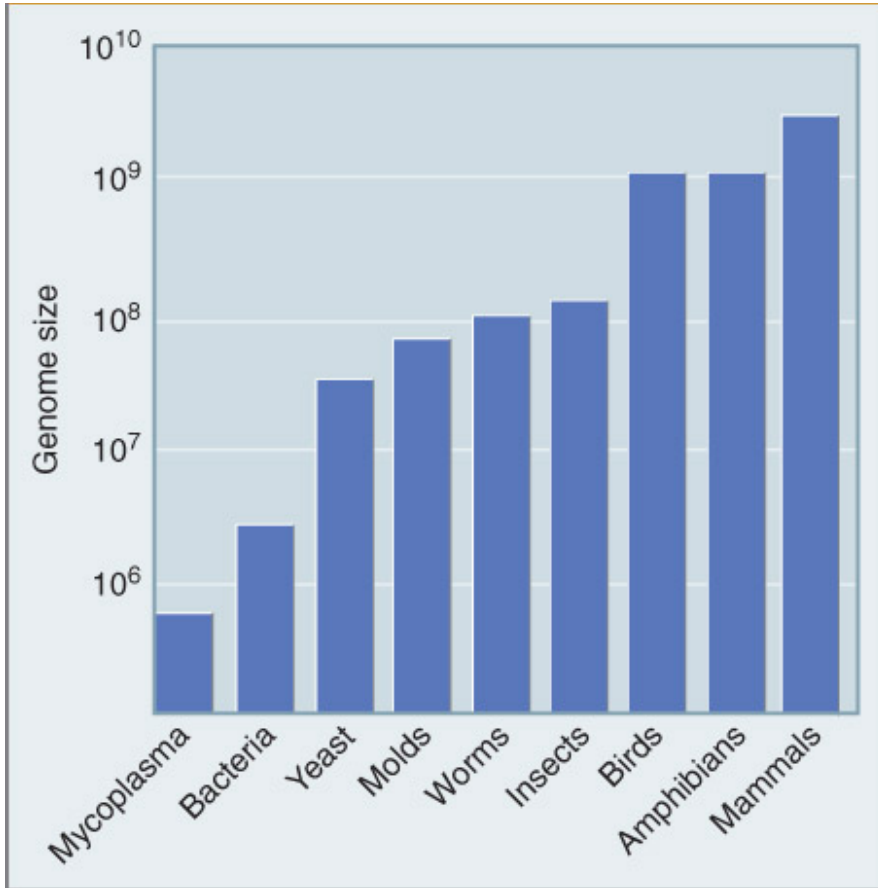
Une séquence d'ADN correspond à 3 phases de lecture indépendantes



En général, une seule de ces phases est fonctionnelle (ouverte)

### III. Structure du génome

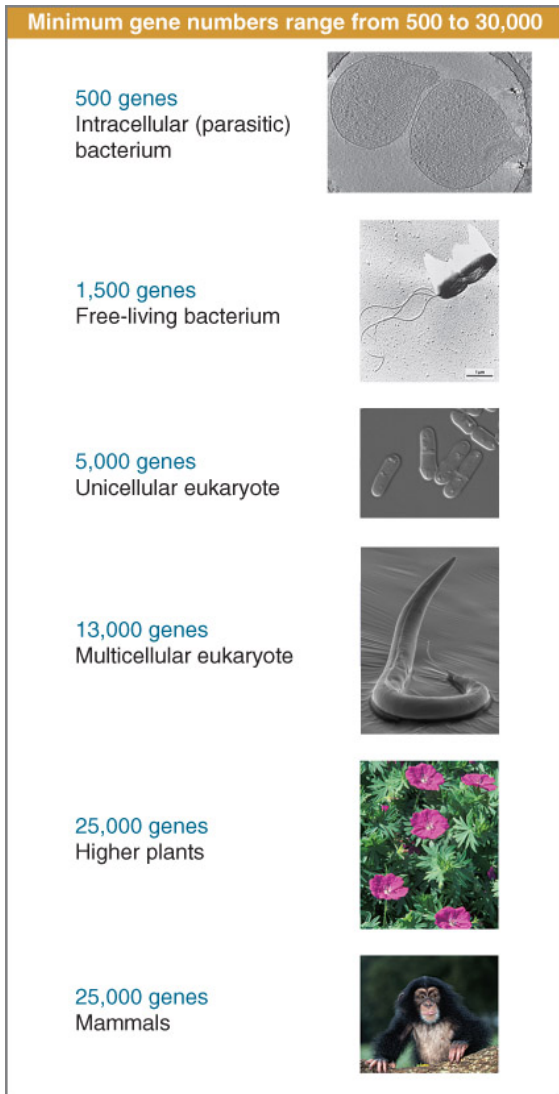
La taille du génome est corrélée à la complexité de l'organisme



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

### III. Structure du génome

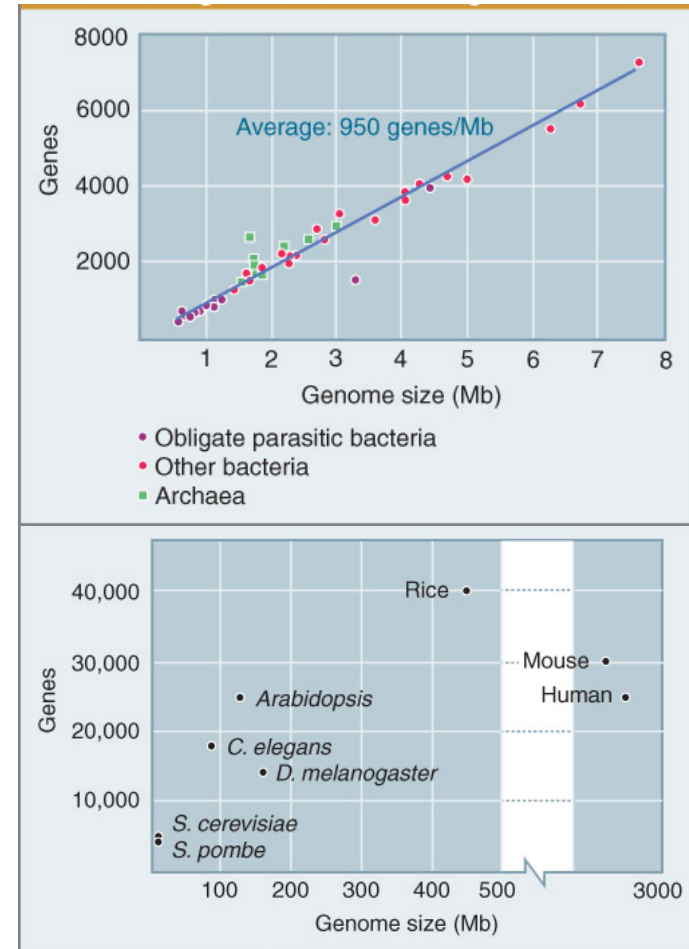
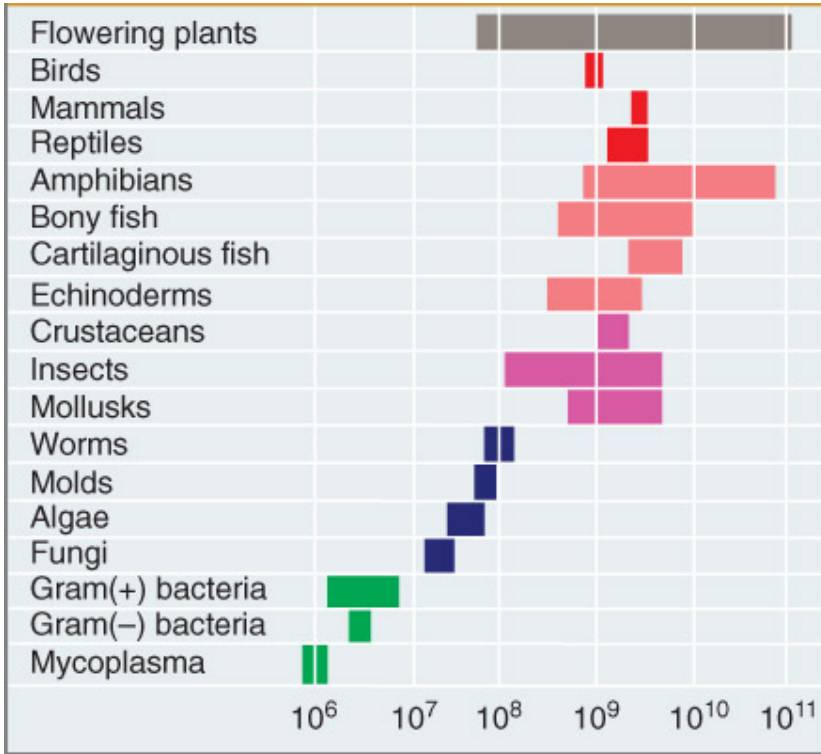
Le nombre de gènes augmente avec la complexité de l'organisme





### III. Structure du génome

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





### III. Structure du génome

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

Minimum gene numbers range from 500 to 30,000

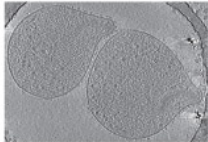


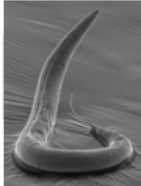


500 genes Intracellular (parasitic) bacterium	
1,500 genes Free-living bacterium	
5,000 genes Unicellular eukaryote	
13,000 genes Multicellular eukaryote	
25,000 genes Higher plants	
25,000 genes Mammals	

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

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

nd, Not determined.

### III. Structure du génome

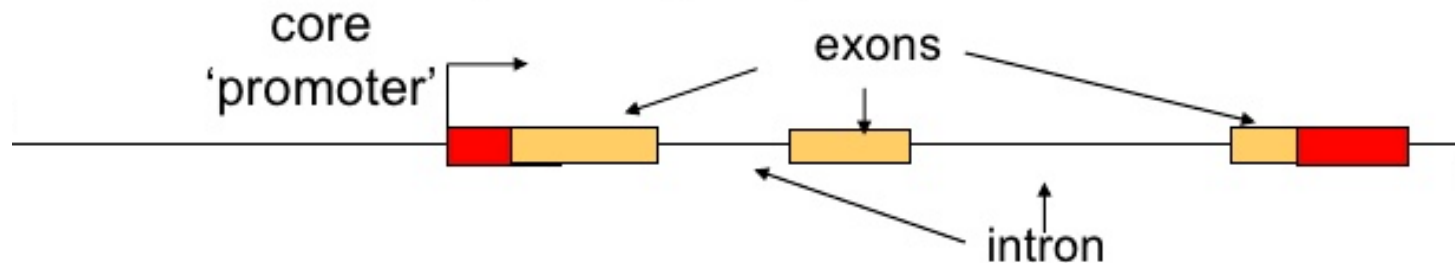
#### Structure du gène

Chez les **procaryotes**, les gènes sont **ininterrompus** (monocistroniques)



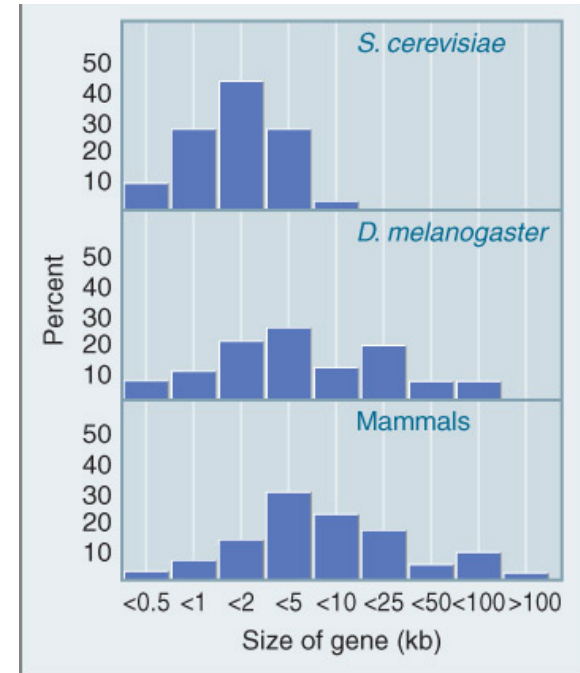
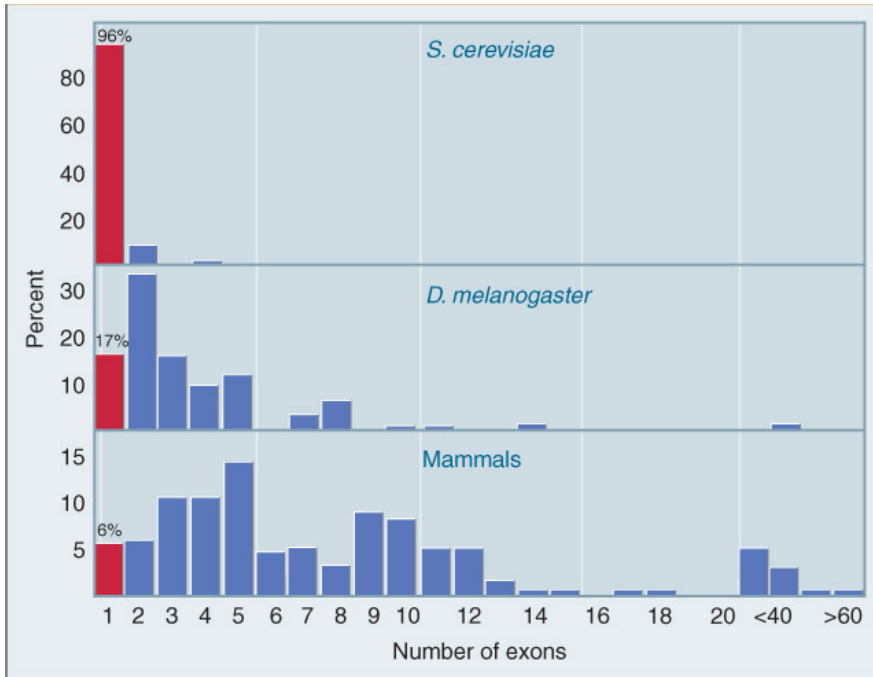
→ ADN et ARNm sont colinéaires

Chez les eucaryotes, les gènes sont devenus **plus longs** et majoritairement **interrompus** (introns/exons)



### III. Structure du génome

#### Structure du gène

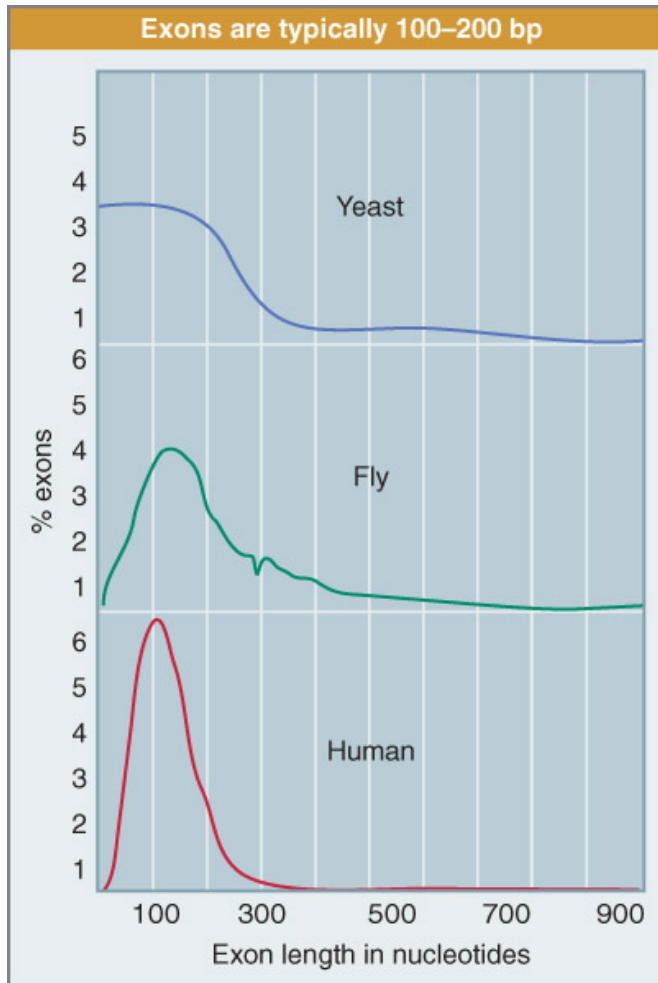


Au cours de l'évolution des eucaryotes, on est passé de **gènes courts, majoritairement ininterrompus**, à des gènes **longs majoritairement interrompus**

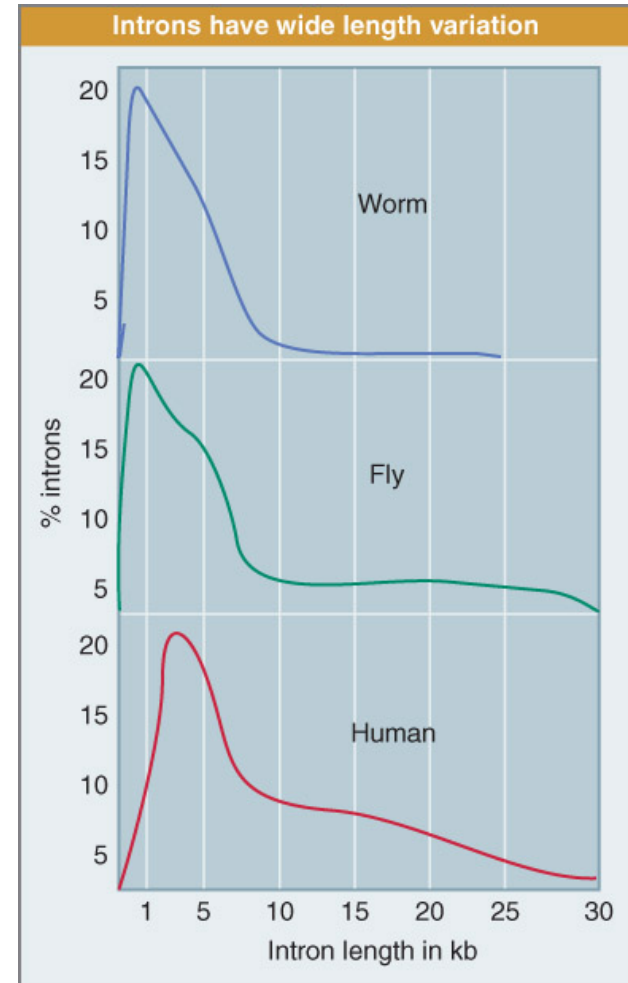
### III. Structure du génome

#### Structure du gène

Les **exons** sont **courts** et de taille relativement **constante** au cours de l'évolution

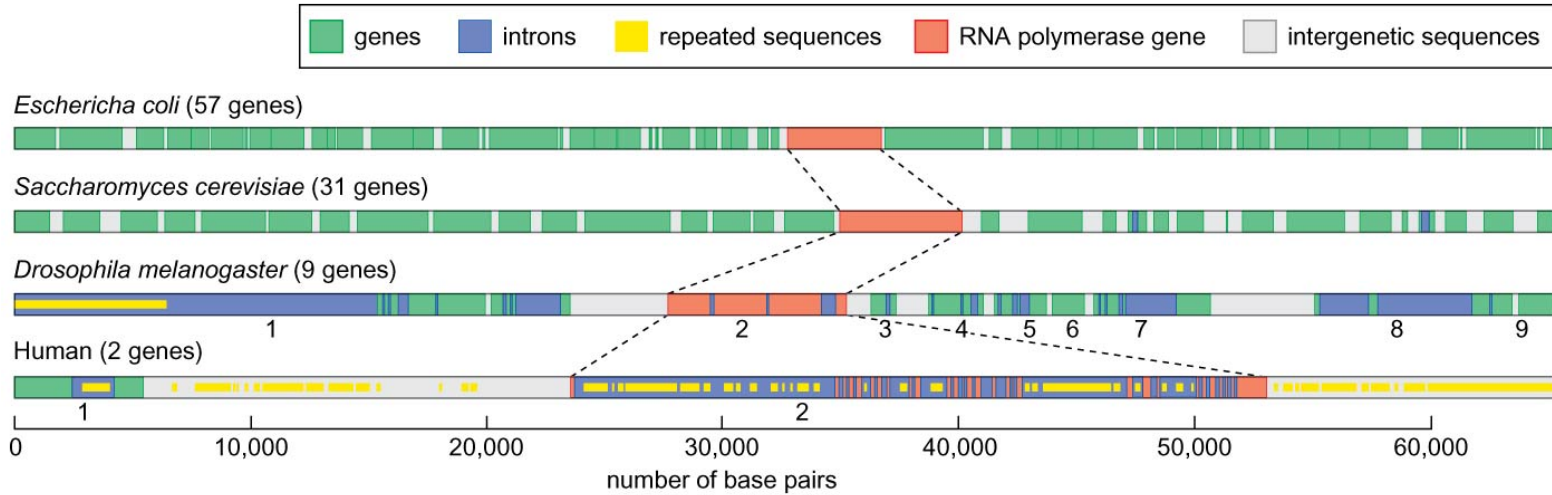


Les **introns** varient en taille et **s'allongent** au cours de l'évolution



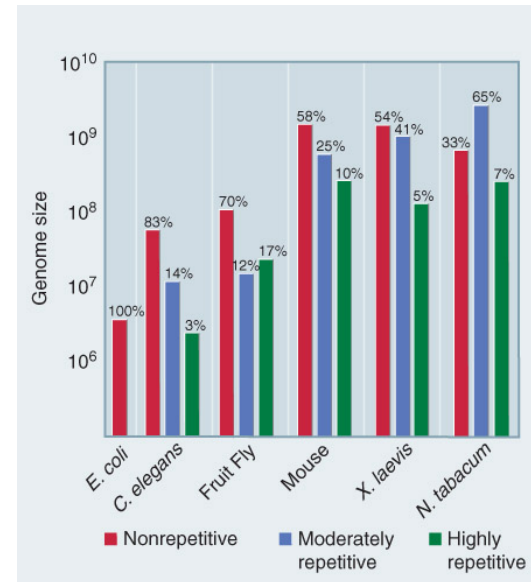
### III. Structure du génome

Les organismes les plus complexes ont une densité génique plus faible



ADN modérément ou fortement répété (junk DNA) : en général ce sont des régions non codantes (transposons, ADN satellitaire, low complexity DNA)

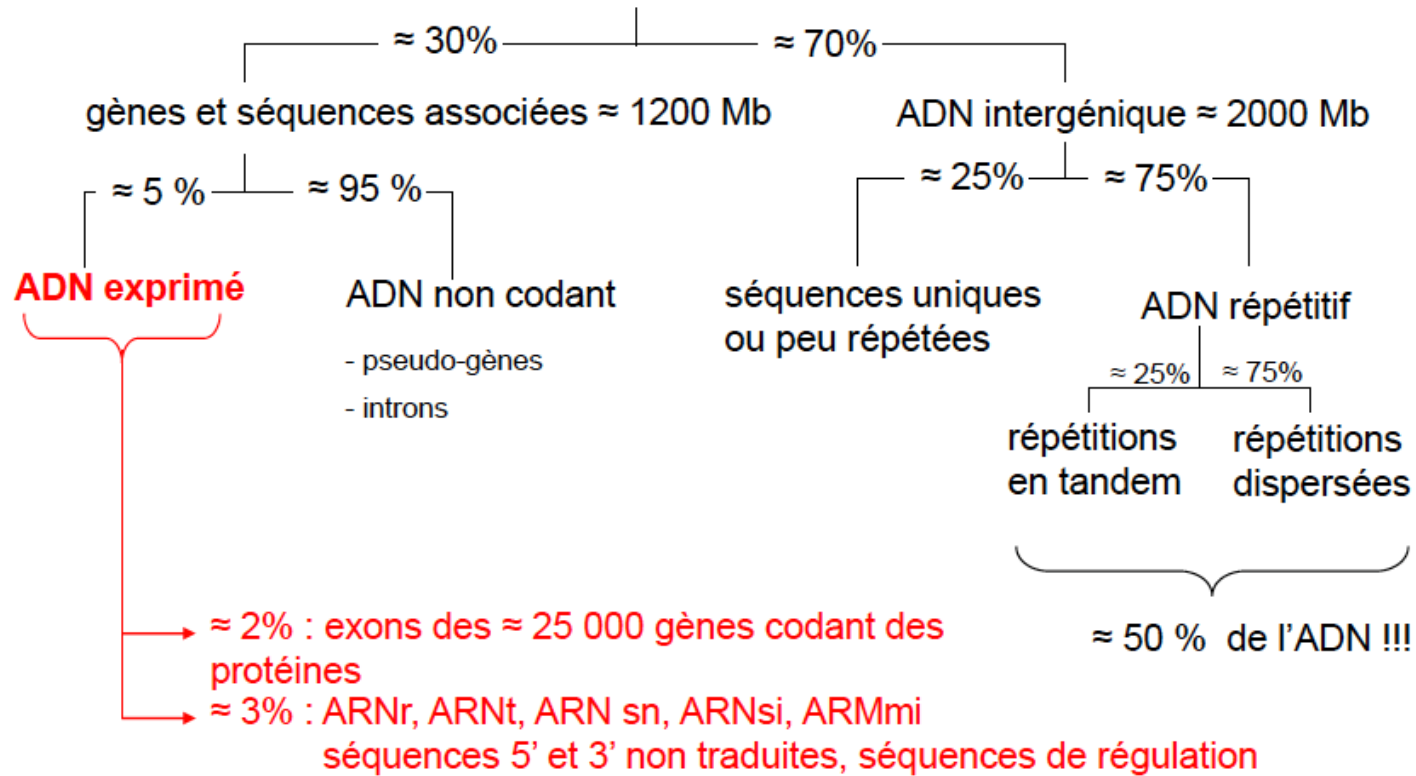
ADN non répété : en général ce sont les régions codantes (gènes)



### III. Structure du génome

#### Composition type d'un génome d'eucaryote supérieur

La majorité de l'ADN humain est non codant!



## IV. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN

**Mutations spontanées** (erreurs de réplication)

**Mutations induites** (altérations chimiques)

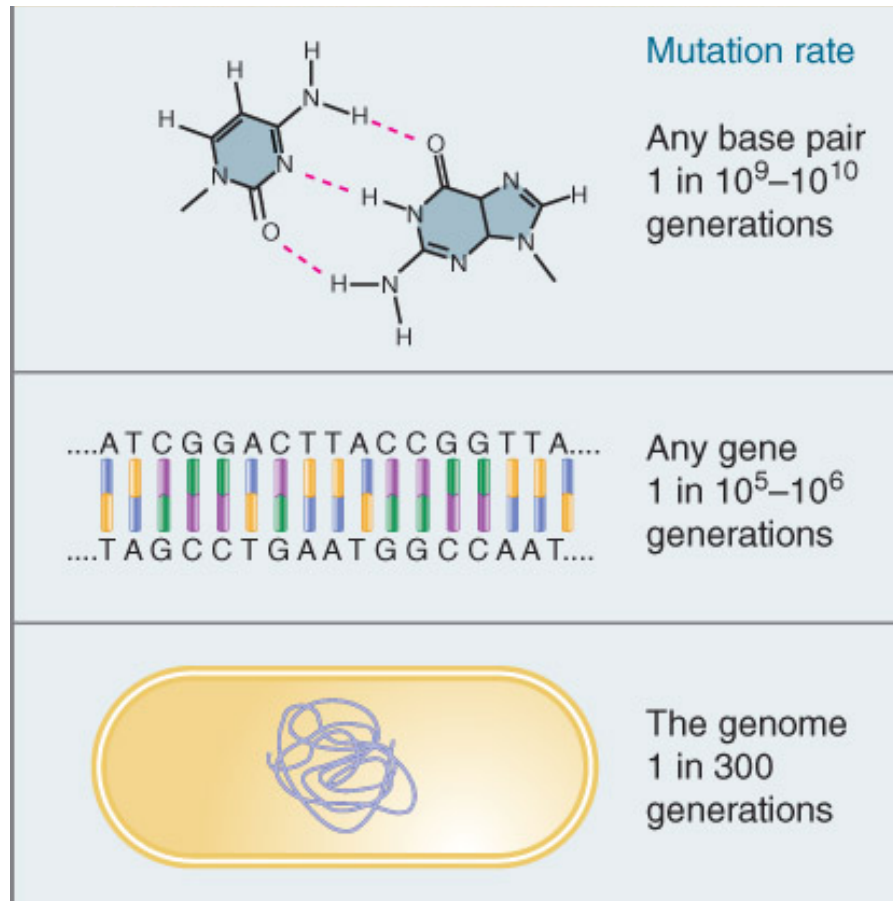
**Insertion de matériel génétique exogène** (virus, transposons)

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

## IV. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN

Mutations spontanées

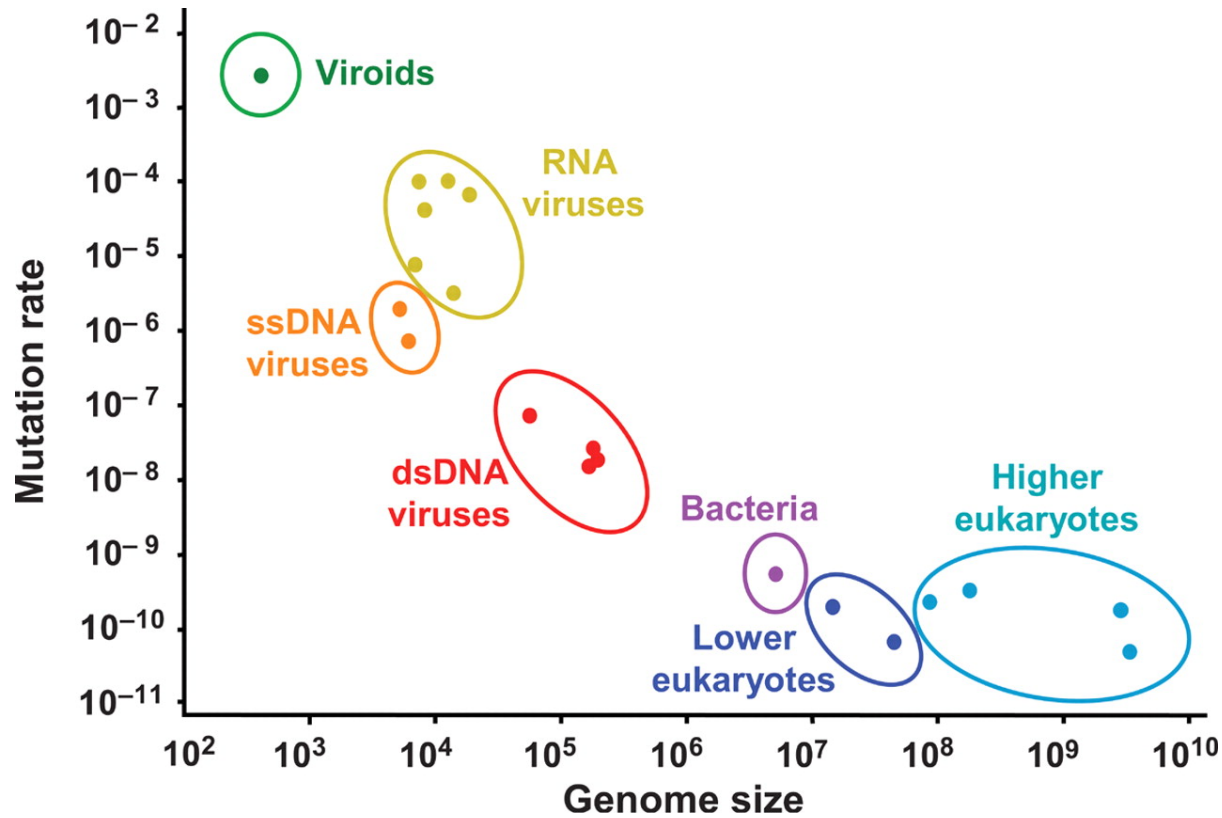




## IV. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN

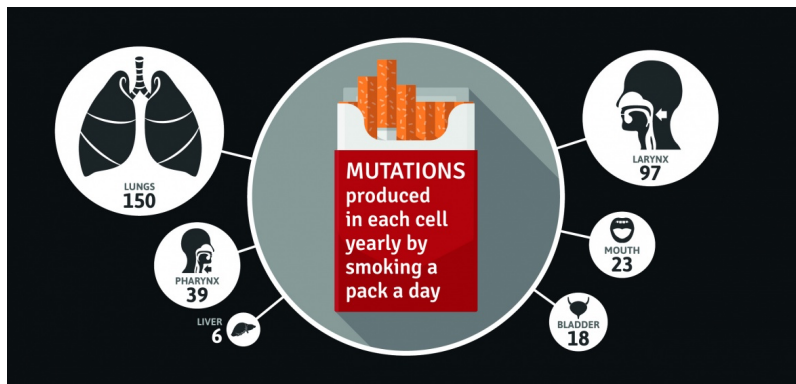
Mutations spontanées



## IV. Les altérations du message

### Les mutations sont des modifications de la séquence d'ADN

#### Mutations induites et mutagènes



(Science, Nov 2016)

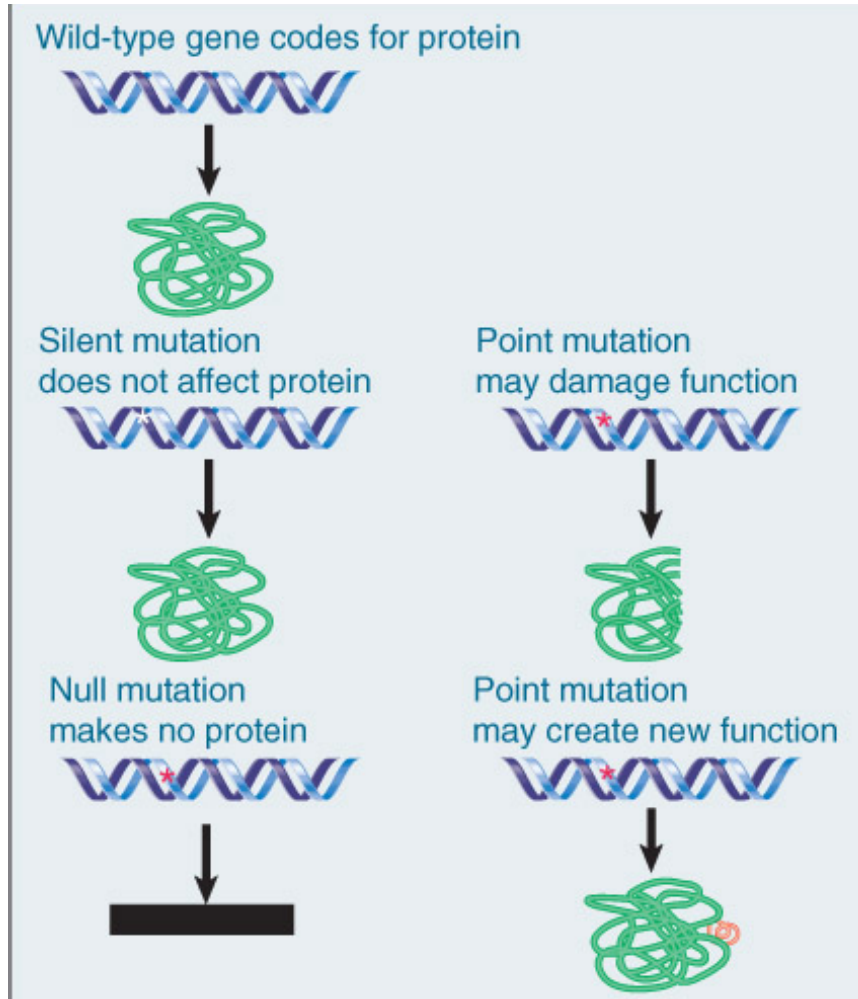
Table II. Partial list of chemicals evaluated with Mutatox and Ames Test for genotoxicity.

Compound	Mutatox	Ames
Aflatoxin B1	Positive	Positive
2-Aminoanthracene	Positive	Positive
2-Aminofluorene	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*	Positive	Negative
3-methylcholanthrene	Positive	Positive
Nalidixic acid	Positive	Negative
Pyrene	Positive	Negative

\* Designated National Toxicology Program Chemical

## IV. Les altérations du message

Les mutations sont des modifications de la séquence d'ADN



Les mutations peuvent être :

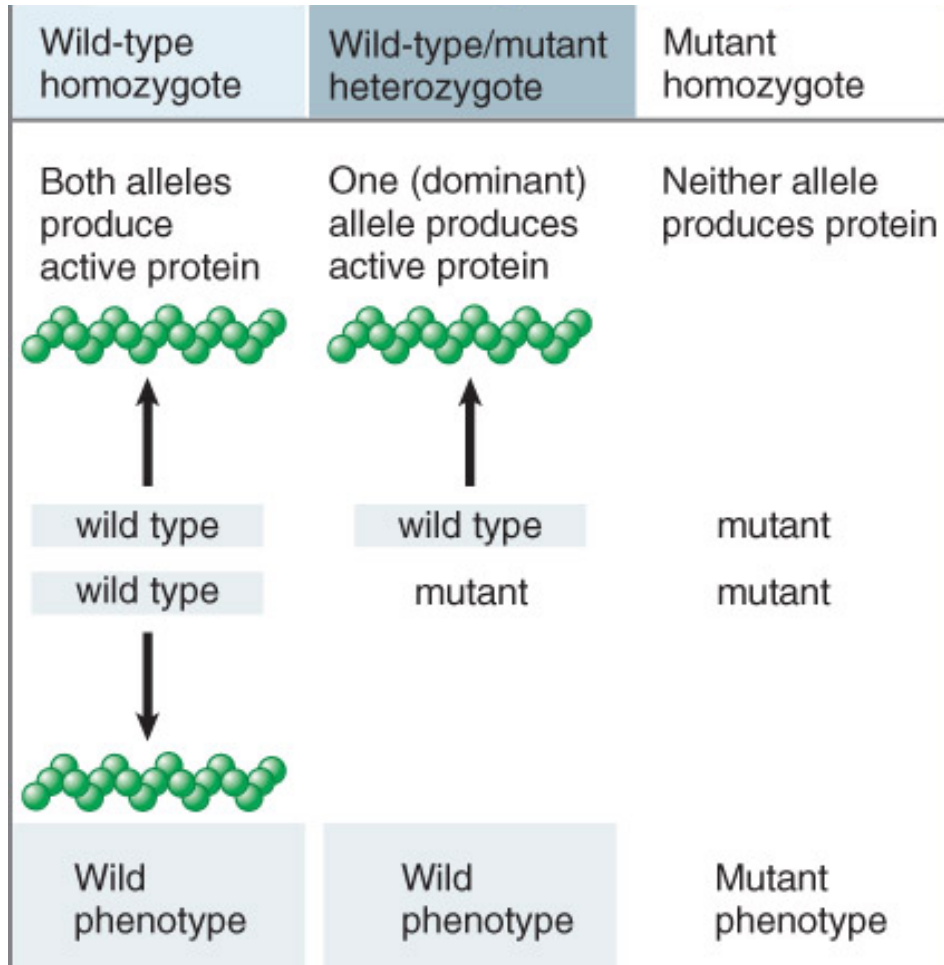
Silencieuses (pas d'effet)

Nulle (perte de fonction)

Nouvelle fonction (gain de fonction)

## IV. 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

## IV. Les altérations du message

1 : **Mutations chromosomiques** : modifications de la structure ou du nombre de copies des chromosomes

**n** (cellule germinale, monoploïde), **2n** (cellule somatique, diploïde) : euploïdie normale

**3n, 4n,...8n** : euploïdie aberrante

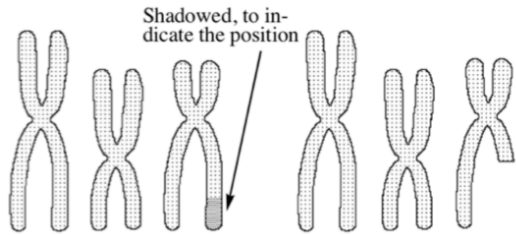
**2n-1, 2n+1** : aneuploïdie

Name	Designation	Constitution	Number of chromosomes
<i>Euploids</i>			
Monoploid	$n$	A B C	3
Diploid	$2n$	AA BB CC	6
Triploid	$3n$	AAA BBB CCC	9
Tetraploid	$4n$	AAAA BBBB CCCC	12
<i>Aneuploids</i>			
Monosomic	$2n - 1$	A BB CC	5
		AA B CC	5
		AA BB C	5
Trisomic	$2n + 1$	AAA BB CC	7
		AA BBB CC	7
		AA BB CCC	7

# IV. Les altérations du message

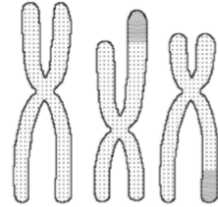
2 : **Mutations chromosomiques** : modifications de la structure ou du nombre de copies des chromosomes

## Chromosome mutations

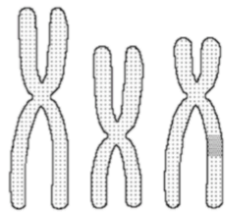


Normal subset of chromosomes

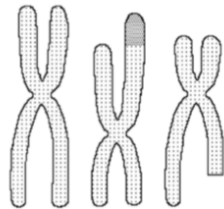
A deletion, a segment of a chromosome is lost



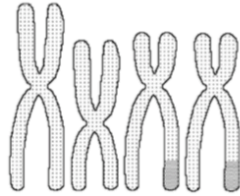
A duplication, a part is present more than once



An inversion, a segment is reversed



A translocation, a part has moved (within the chromosome or to another)

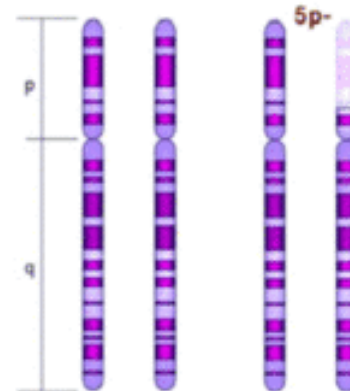


The number of chromosomes has changed

- Loss of DNA
- Move of DNA

➔ Serious for the cell

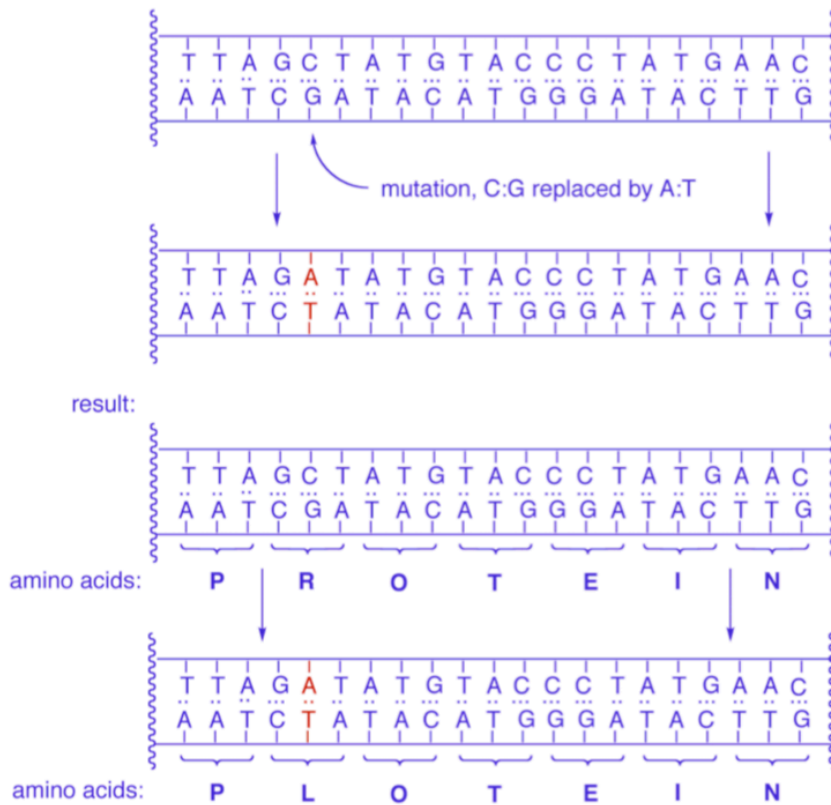
5p- syndrome = cri du chat syndrome



## IV. Les altérations du message

### Les substitutions

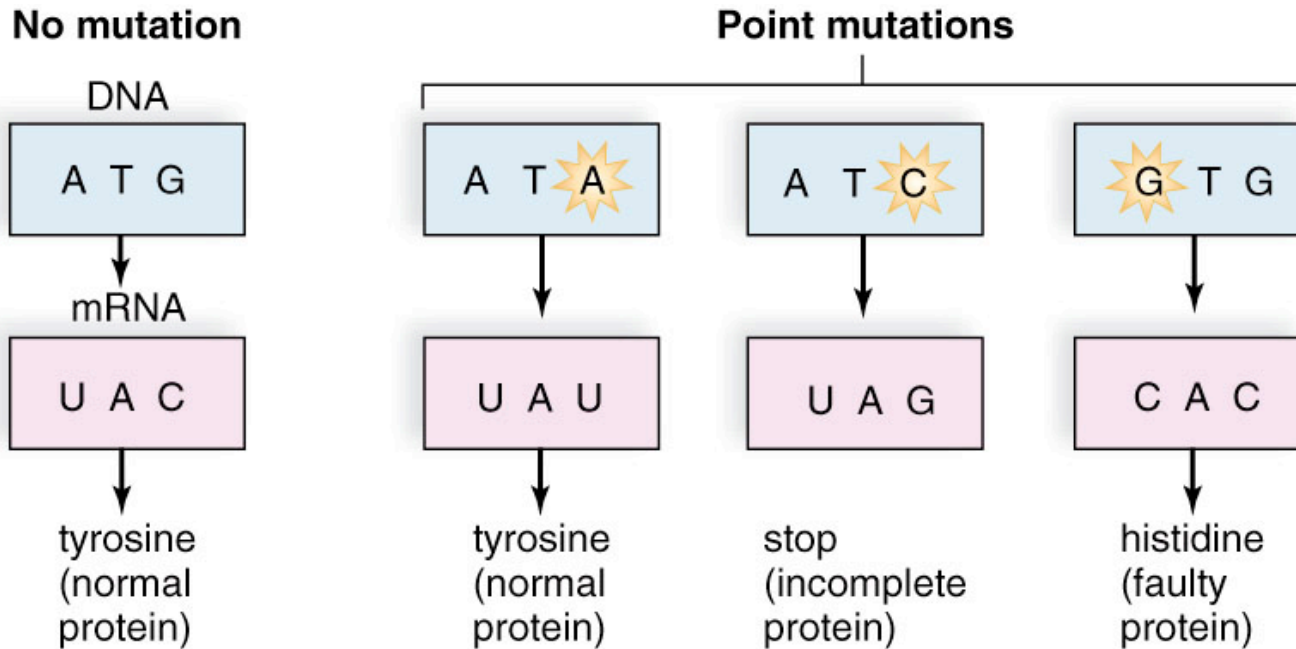
# Base pair substitution



a limited damage

# IV. Les altérations du message

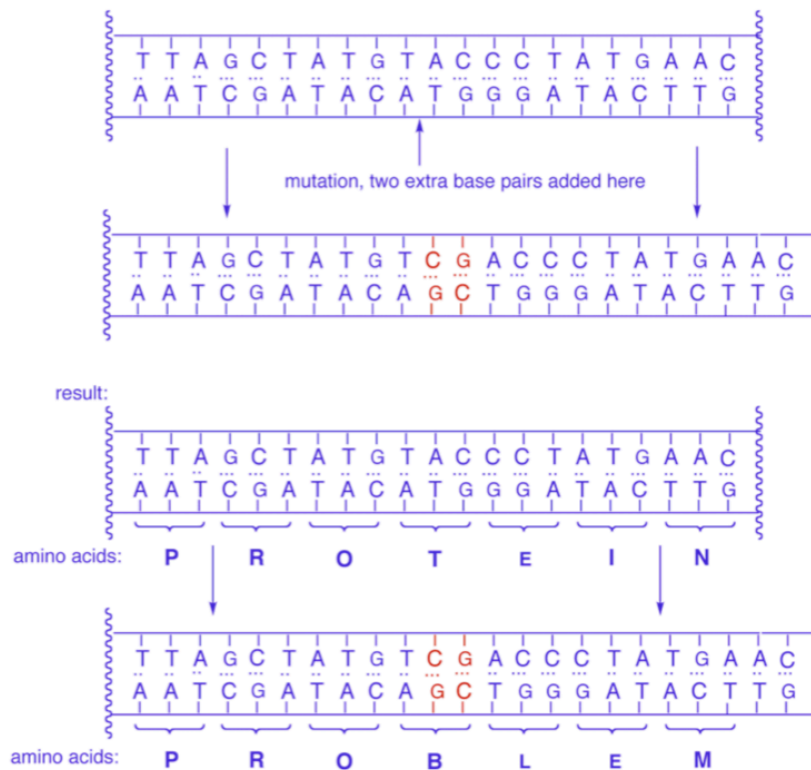
## Le code génétique





## Les mutations saut de phase

# Frame shift mutation

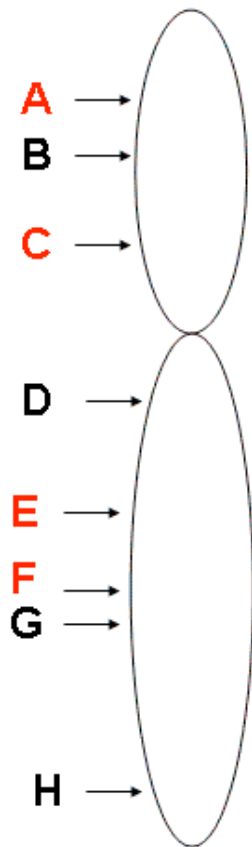


more  
serious

Les substitutions peuvent ou non induire un changement dans le polypeptide

## V. Le décodage du génome

### La cartographie génique



### Genetic Maps

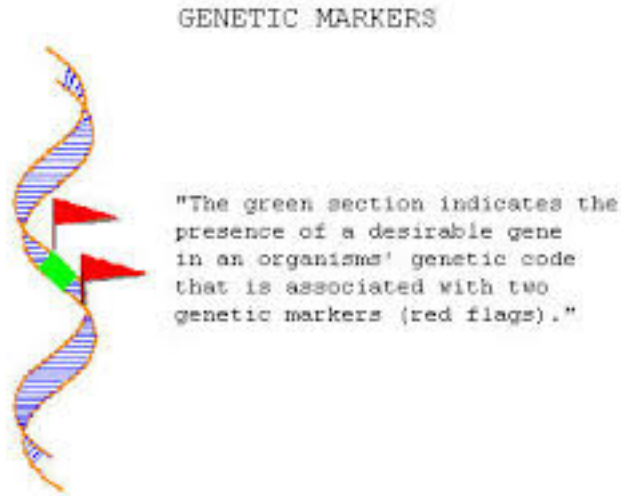
- Genes are shown in relative order and distance from each other based on pedigree studies.
- The chance of the chromosome breaking between A & C is higher than the chance of the chromosome breaking between A & B during meiosis.
- Similarly, the chance of the chromosome breaking between E & F is higher 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.

## V. Le décodage du génome

### Les marqueurs génétiques/moléculaires :

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

Caractéristiques : multiallèlique (polymorphisme), universel et transmissible

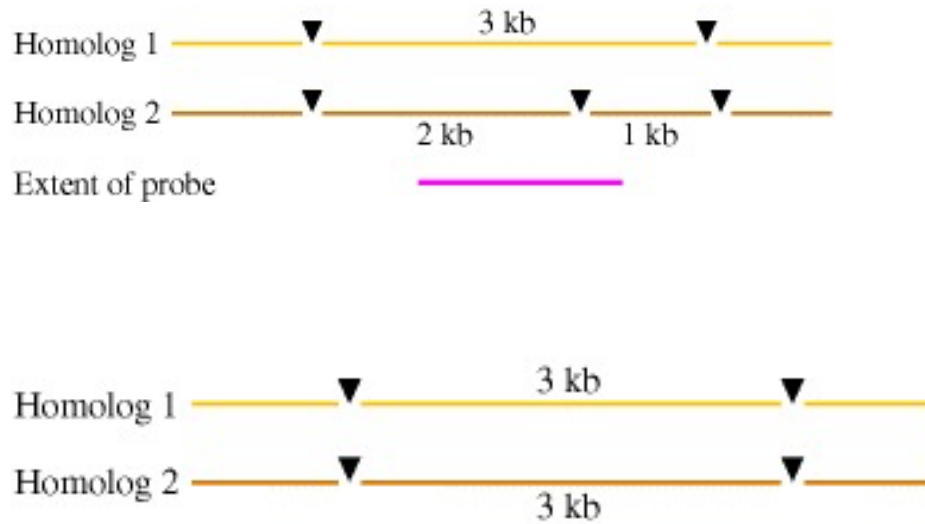


**Utilité des marqueurs génétiques :**

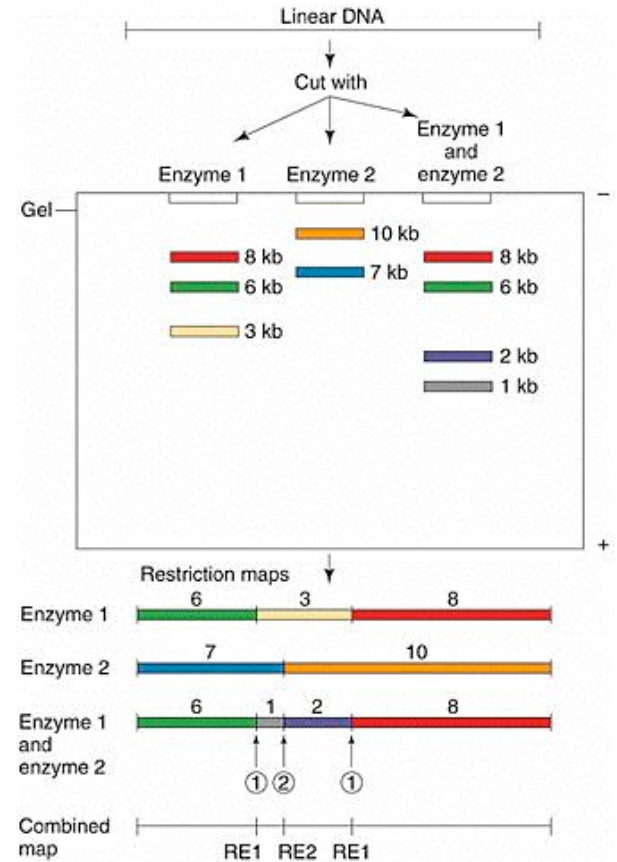
<http://www.youtube.com/watch?v=nrnJPC6e19c>

# V. Le décodage du génome

RFLP : Restriction length fragment polymorphism

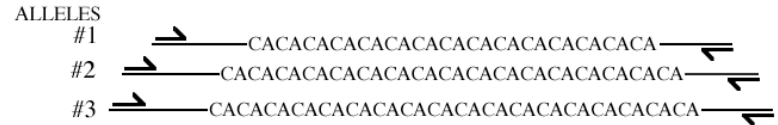


Gene coupling



# V. Le décodage du génome

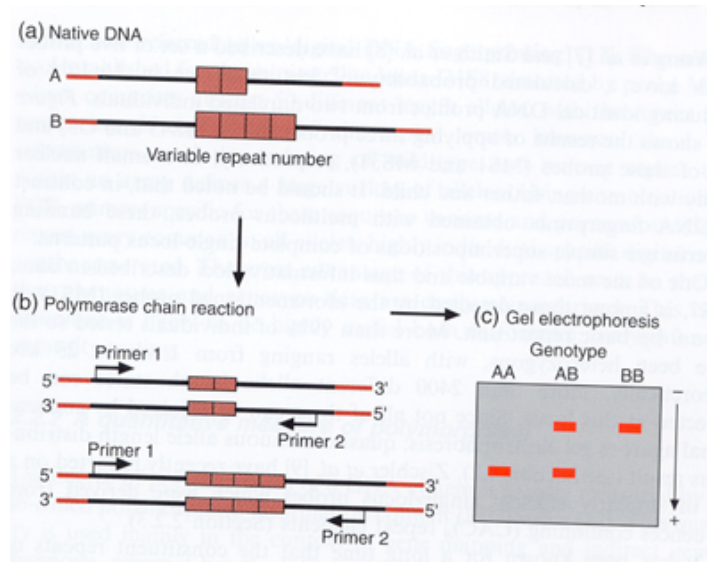
STRs : Short tandem repeats (VNTR)



DéTECTÉ par RFLP :



DéTECTÉ par AFLP :



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

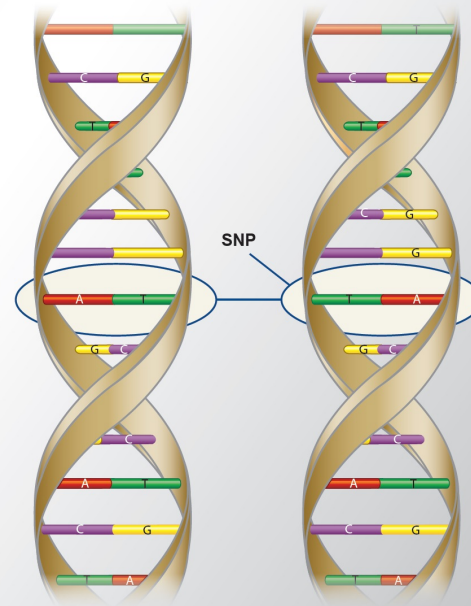
## V. Le décodage du génome

### SNP : Single nucleotide polymorphism

#### What is a Single Nucleotide Polymorphism (SNP)?

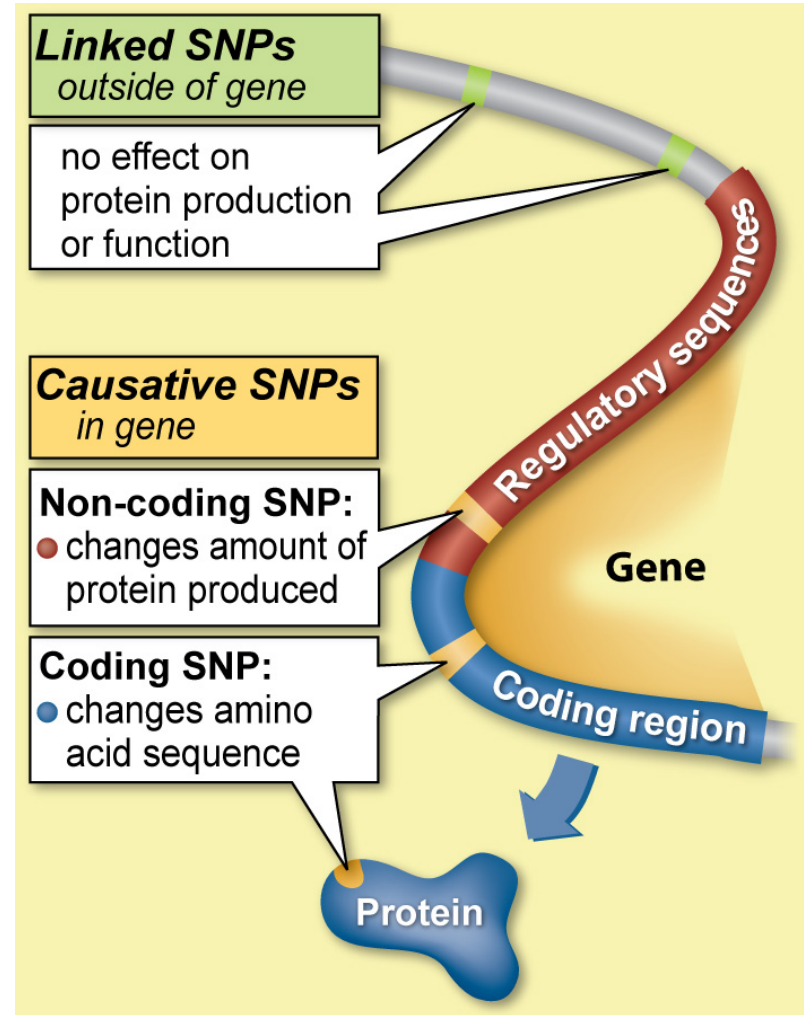
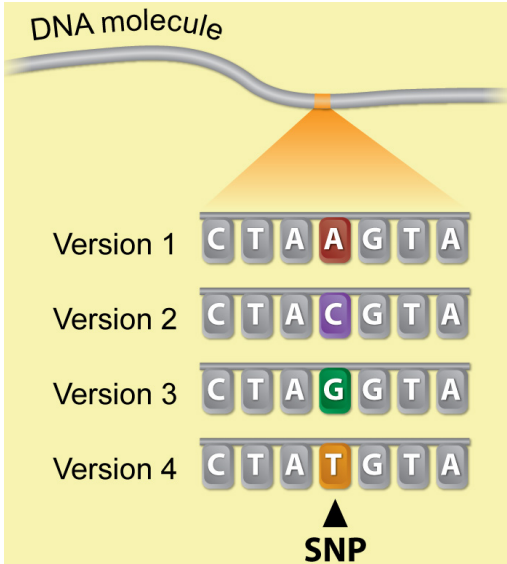
A SNP (pronounced "snip") is a DNA sequence variation that occurs when a single nucleotide (A, T, C, or G) in the genome sequence is modified.

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

