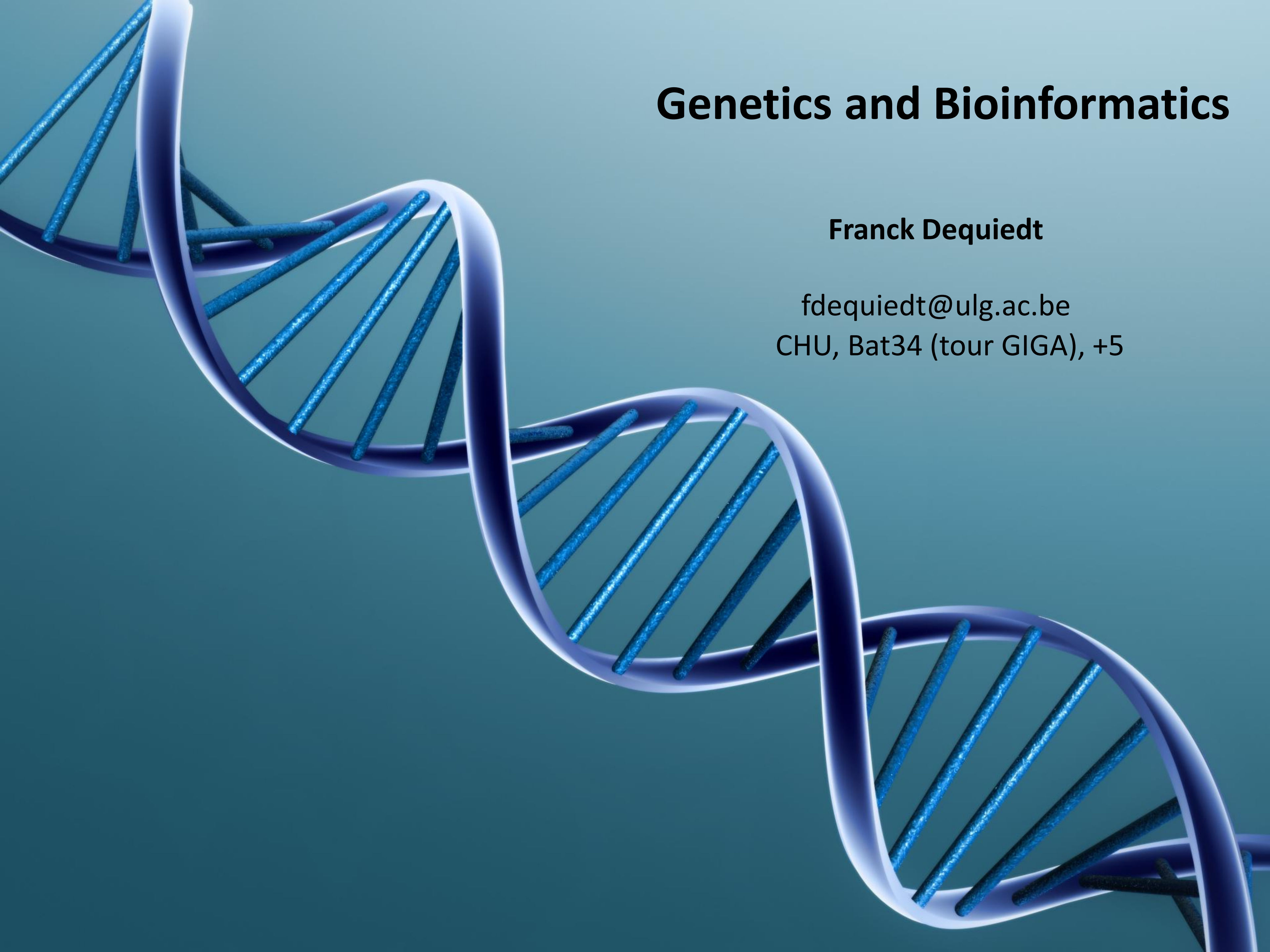


Genetics and Bioinformatics

Franck Dequiedt

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

INTRODUCTION

Fundamental principles

Trait (Character) : qualitative or quantitative feature transmissible to the next generations

Evolution : Modification of traits from generation to generation driven par natural selection and genetic drift

Heredity : transmission of inherited traits from one individual to its progeny

Genetic: science that studies individuals' inherited characters, their transmission throughout generations and their alterations (mutations).

I – The emergence of Molecular Biology

- Classical genetics

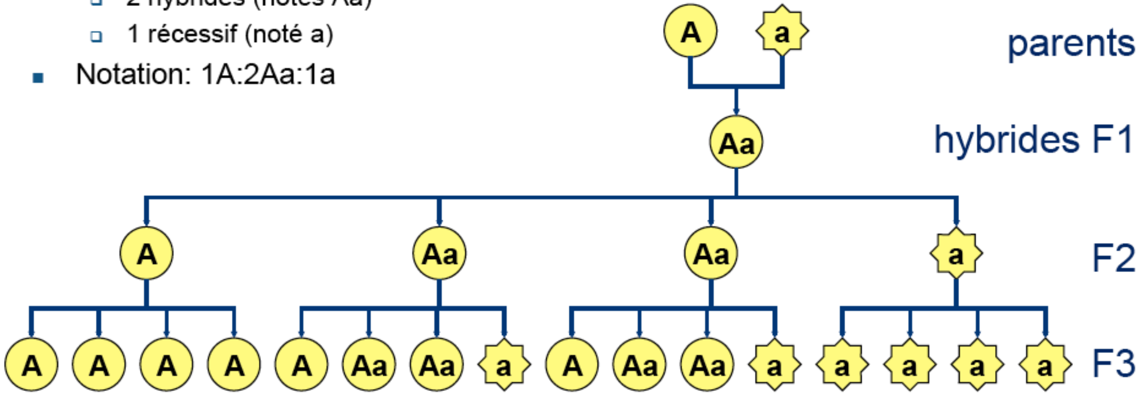
Gregor Mendel (1822 - 1884)

Heredity’s (Mendel’s) laws (1866)

Law n°1 : uniformity of characters at first generation.

Loi n°2 : Segregation of characters

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



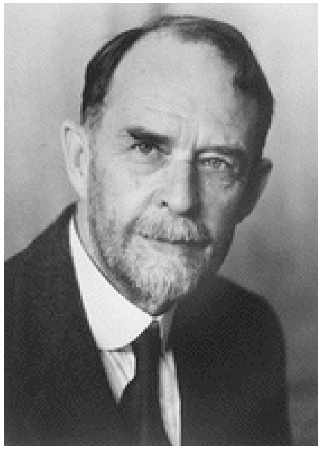
Genes and alleles : (concepts 1 to 6) : http://www.dnaftb.org/1/

I – The emergence of Molecular Biology

- Classical genetics

Gregor Mendel (1822 - 1884)

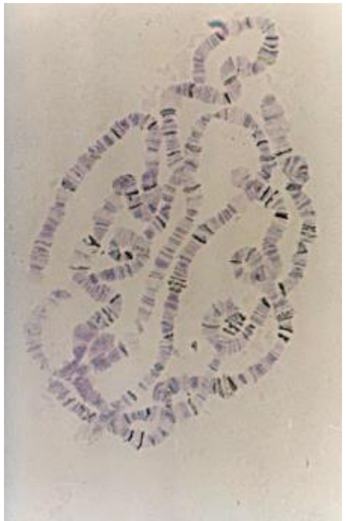
- Chromosome is the support of heredity (1910)



Thomas Morgan (1866 - 1945)

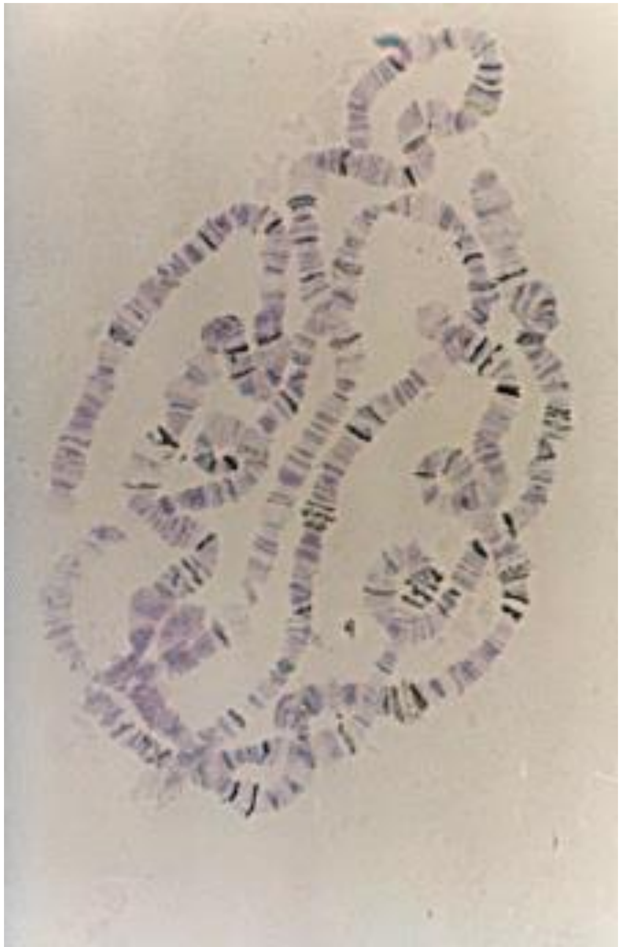
Chromosmal theory of heredity

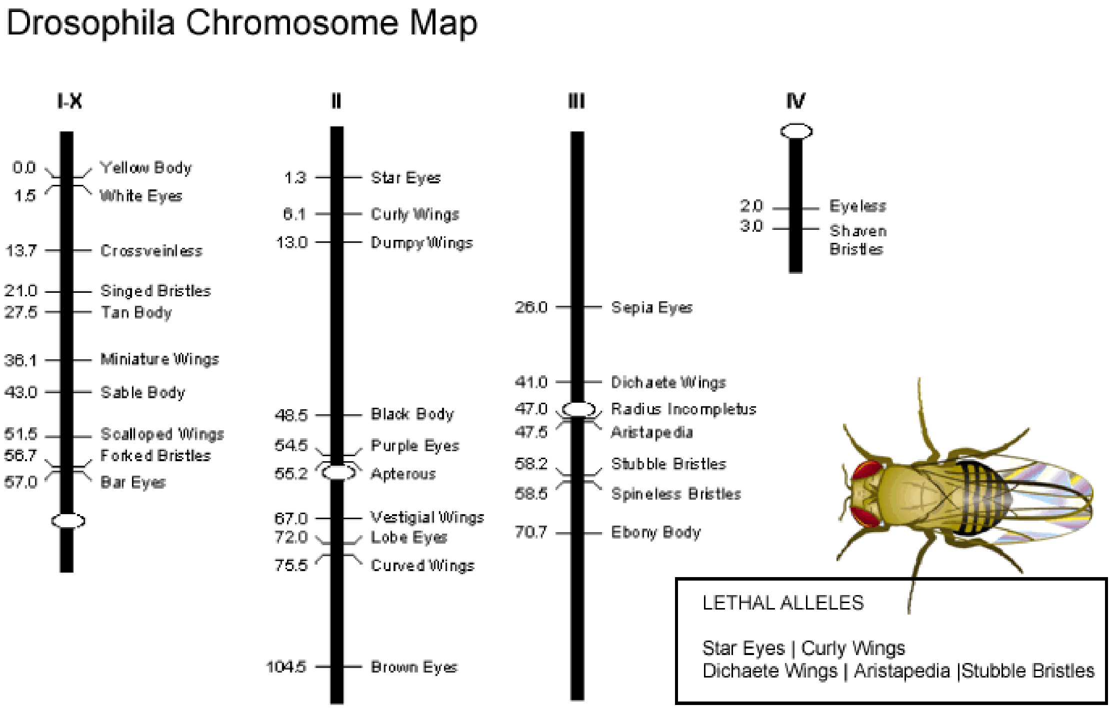
Drosophila melanogaster



Chromosomes polytènes

- Chromosome is the support of heredity (Thomas Morgan)



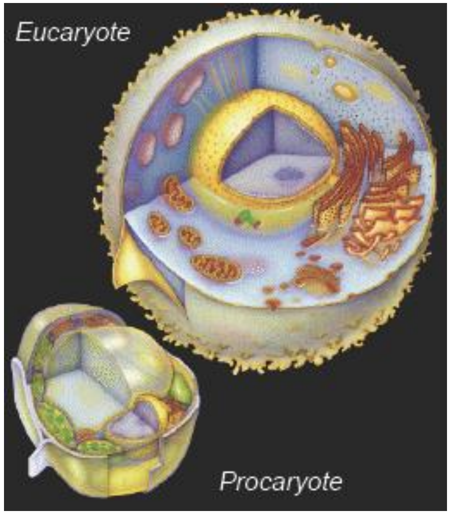


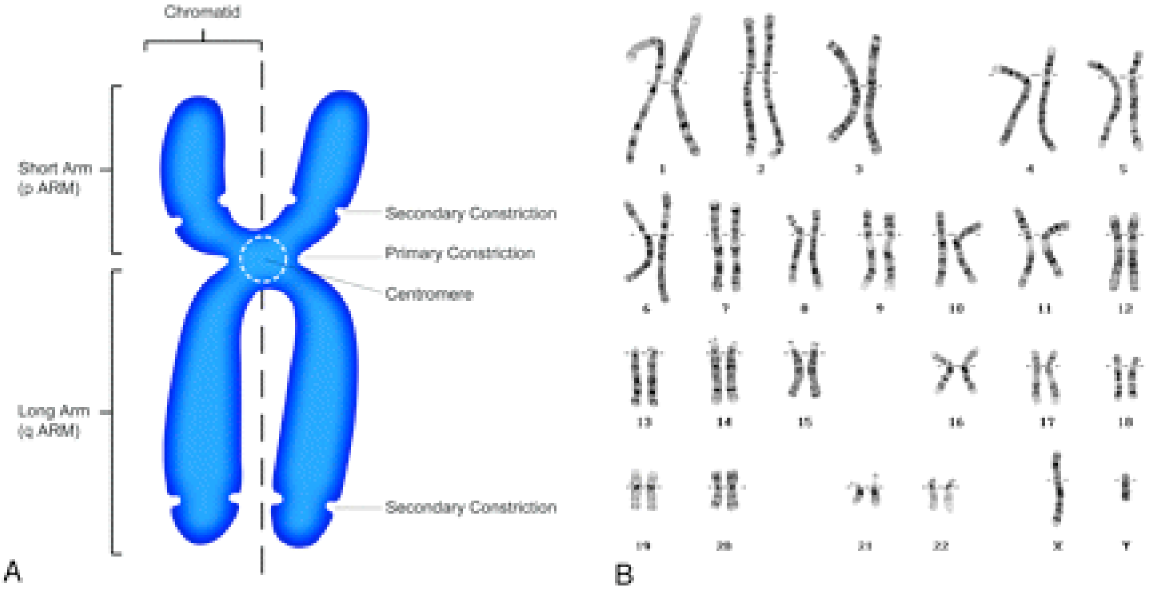
Drosophila Chromosome Map

LETHAL ALLELES

Star Eyes | Curly Wings
Dichaete Wings | Aristapedia |Stubble Bristles

- Chromosome is the support of heredity





Genes and Chromosomes : (concepts 7 to 14) : http://www.dnaftb.org/1/

https://www.youtube.com/watch?v=PLaDJMx88FI

I – The emergence of Molecular Biology

- Classical genetics

Gregor Mendel (1822 - 1884)

- Chromosome is the support of inheritence

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

- DNA is the support of inheritence

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

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

- Composition and structure of DNA

Erwin Chargaff (1905 - 1992)

A+T / C+G = K

A / T = C / G = 1

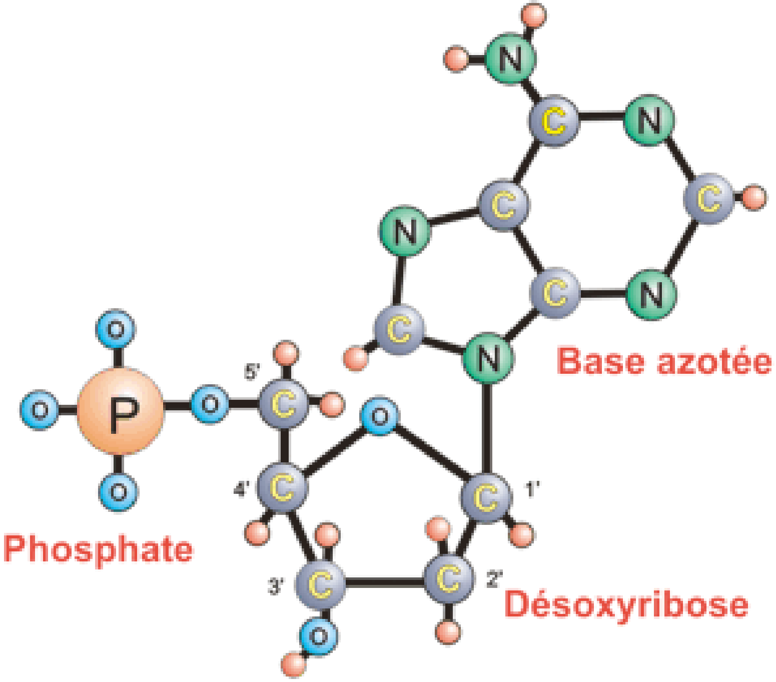
https://www.youtube.com/watch?v=a3_1AOXCMmo

II – Structure of nucleic acids

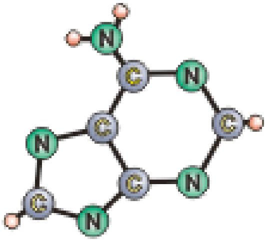
II.I Simple molecules

Desoxyribonucleic acid (DNA)

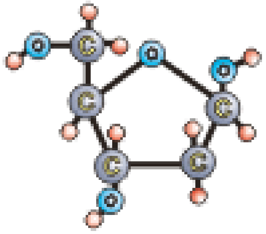
= polymers of small sub-units called **nucleotides**



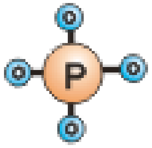
BASE



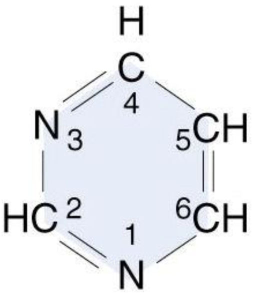
SUGAR



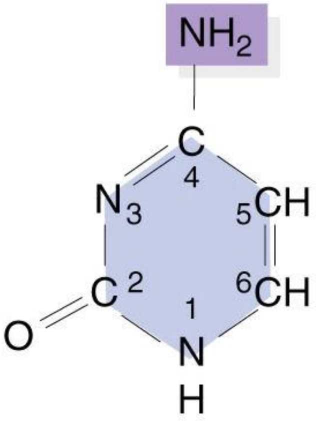
PHOSPHORIC ACID



II.I.1 Nitrogenous bases

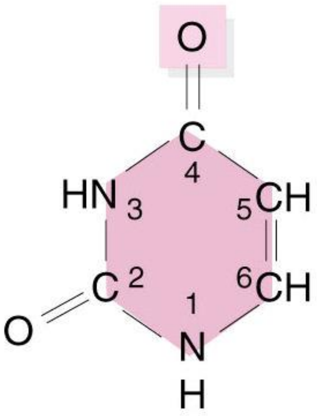


Pyrimidine



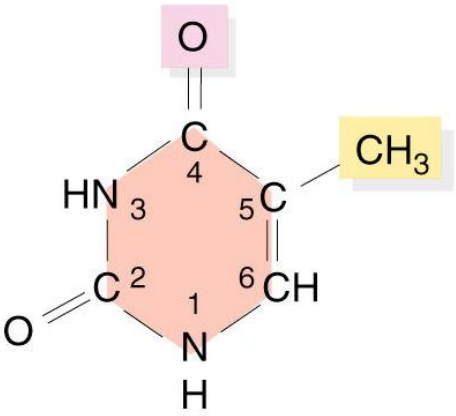
Cytosine (C)

2-oxy-4-amino
-pyrimidine



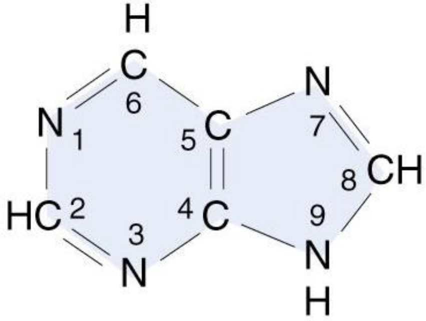
Uracil (U)
(found in RNA)

2,4-dioxy-pyrimidine

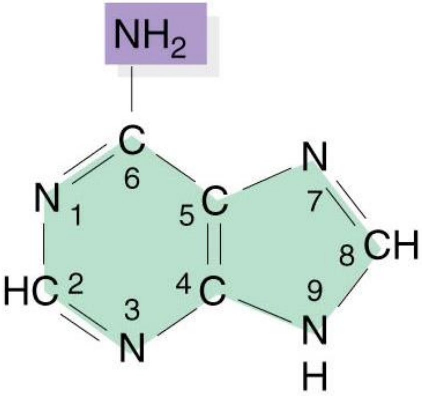


Thymine (T)
(found in DNA)

2,4-dioxy-5-méthyl-
pyrimidine

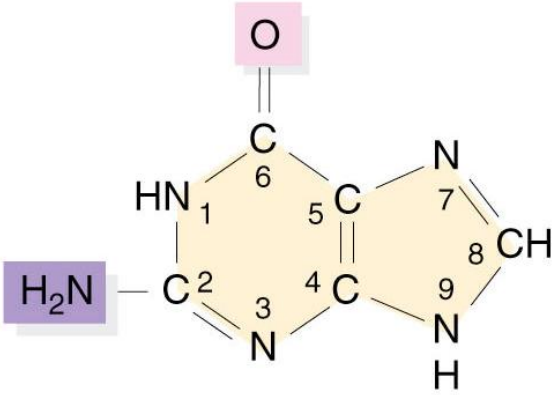


Purine



Adenine (A)

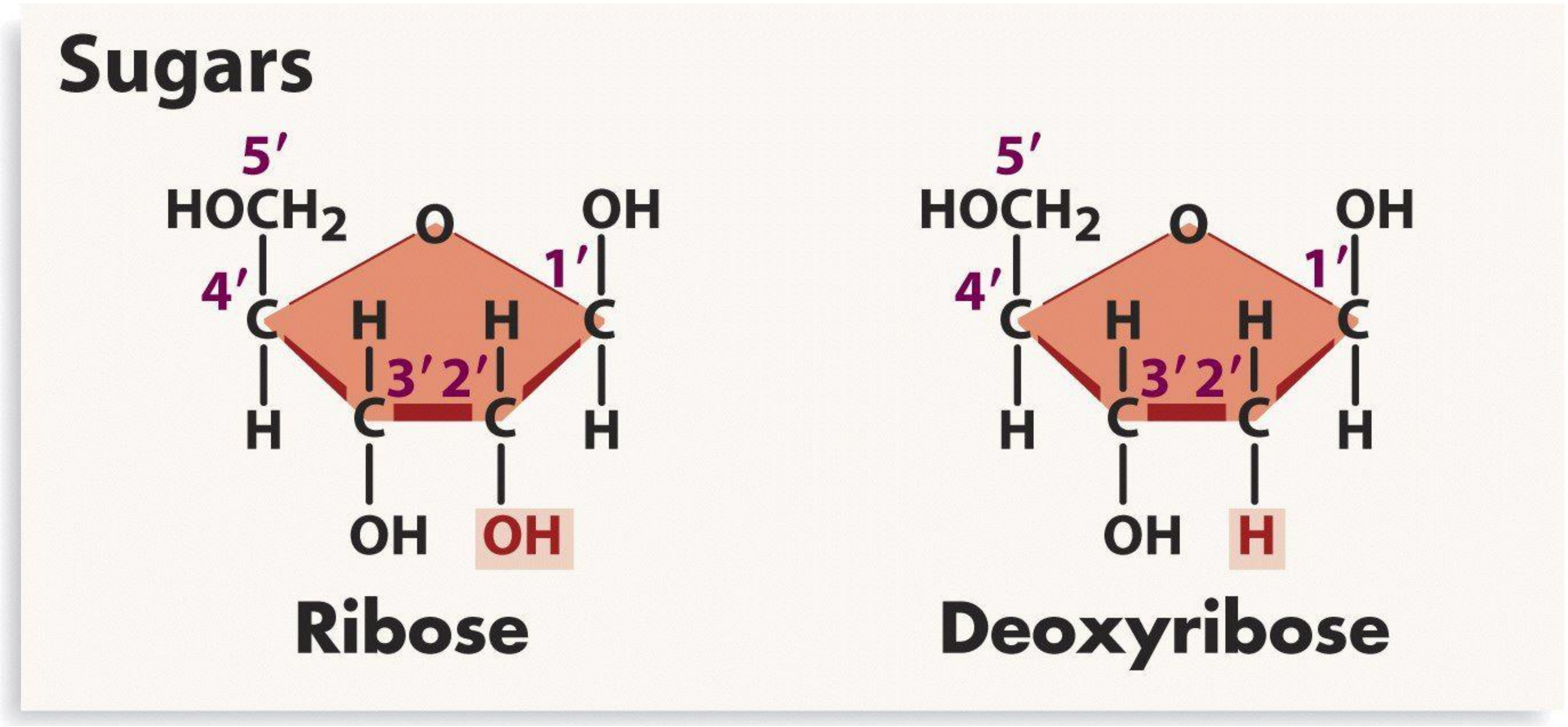
6-amino-purine



Guanine (G)

2-amino-6-oxy
-purine

II.I.2 Sugars



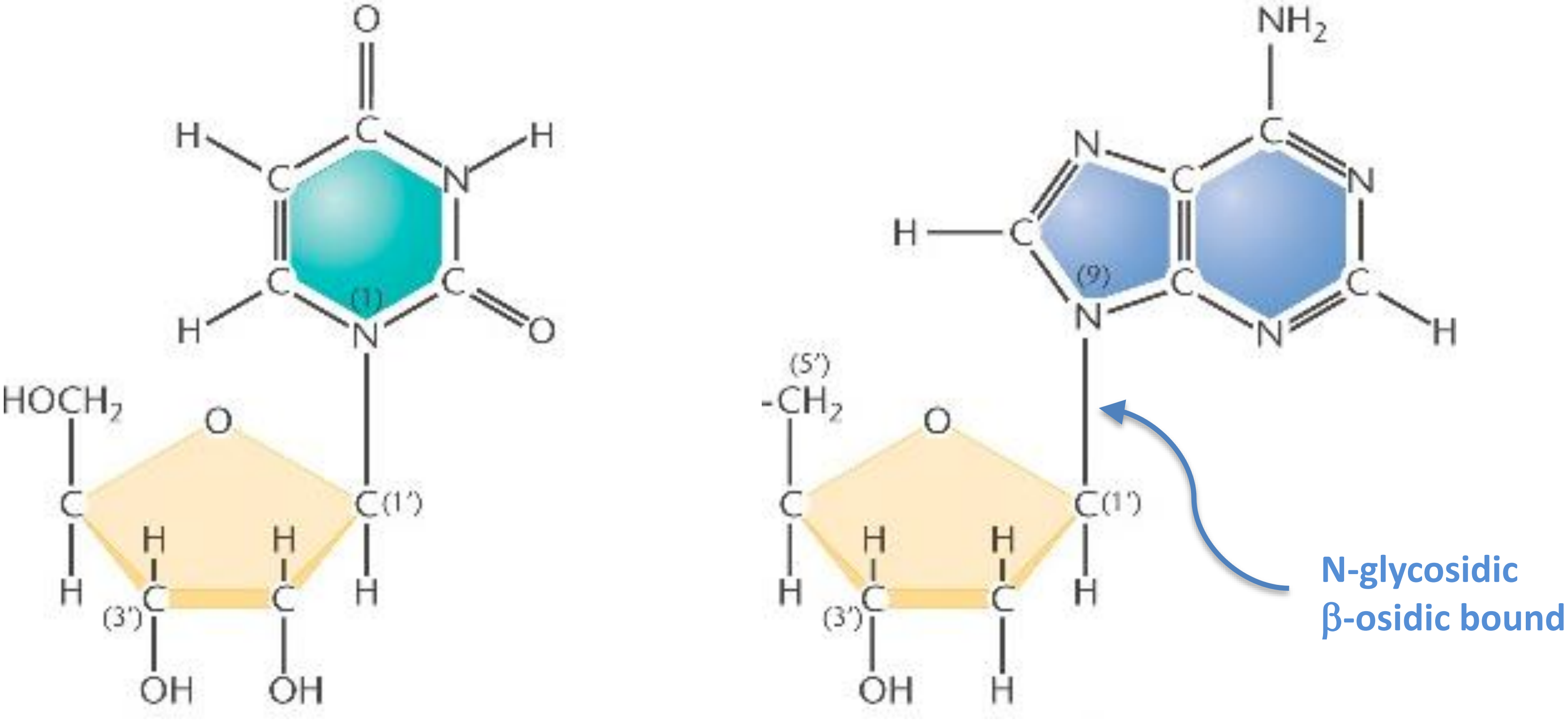
RNA

DNA

II.2 Nucleosides

NucleoSides

SUGAR + BASE

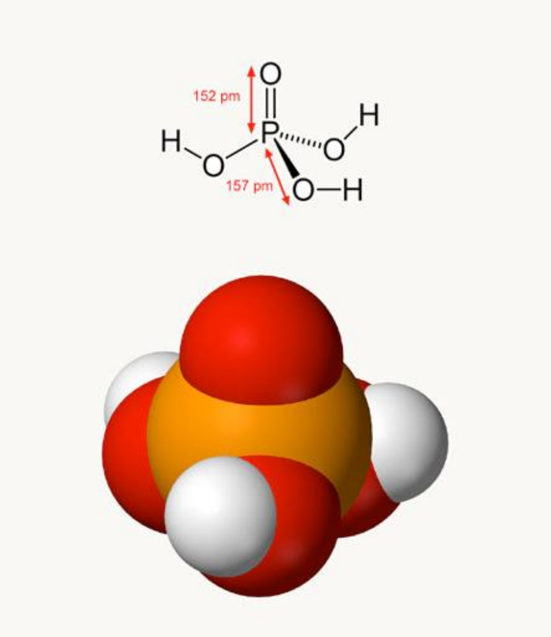


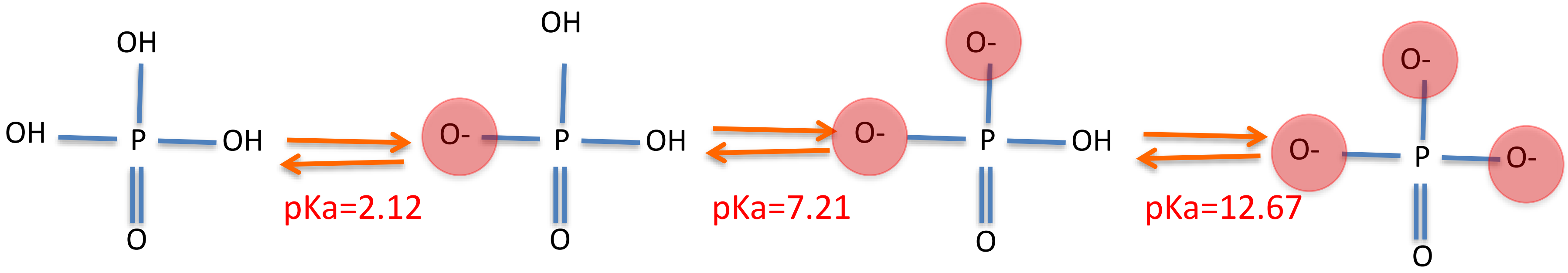
N-glycosidic
β-osidic bound

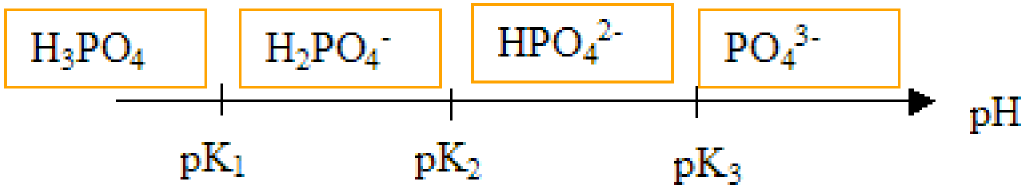
II.2 Nucleosides

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

II.3 Phosphoric Acid





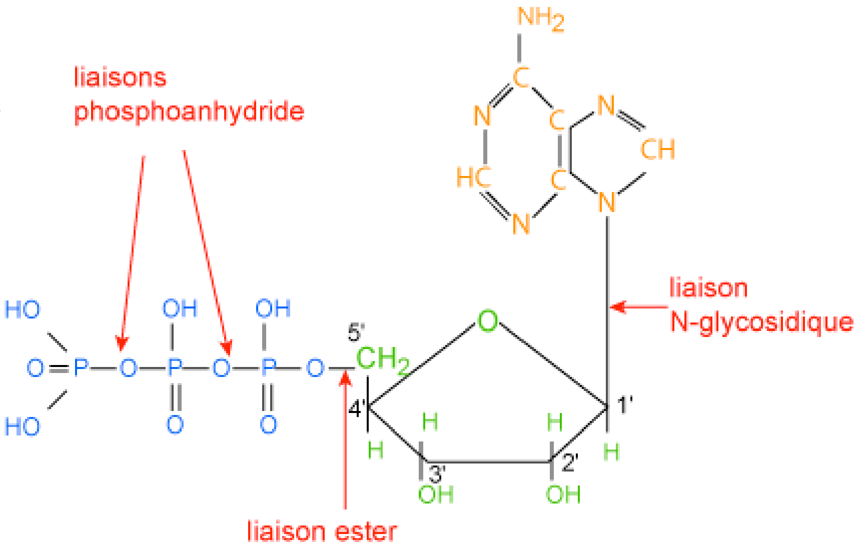


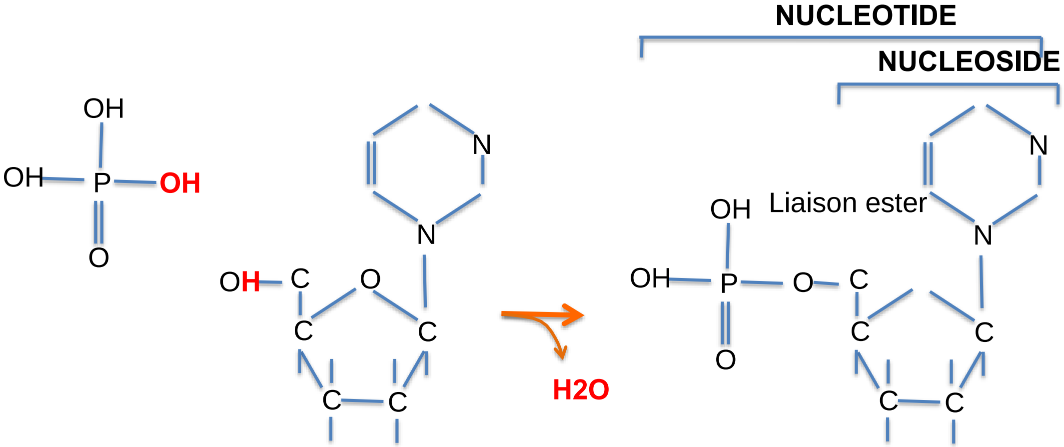
H_3PO_4 | $H_2PO_4^-$ | HPO_4^{2-} | PO_4^{3-}

pK_1 | pK_2 | pK_3 | pH

II.4 Nucleotides

SUGAR + BASE+ PHOSPHATE = NUCLEOTIDE





Nucléoside-5’-Phosphate

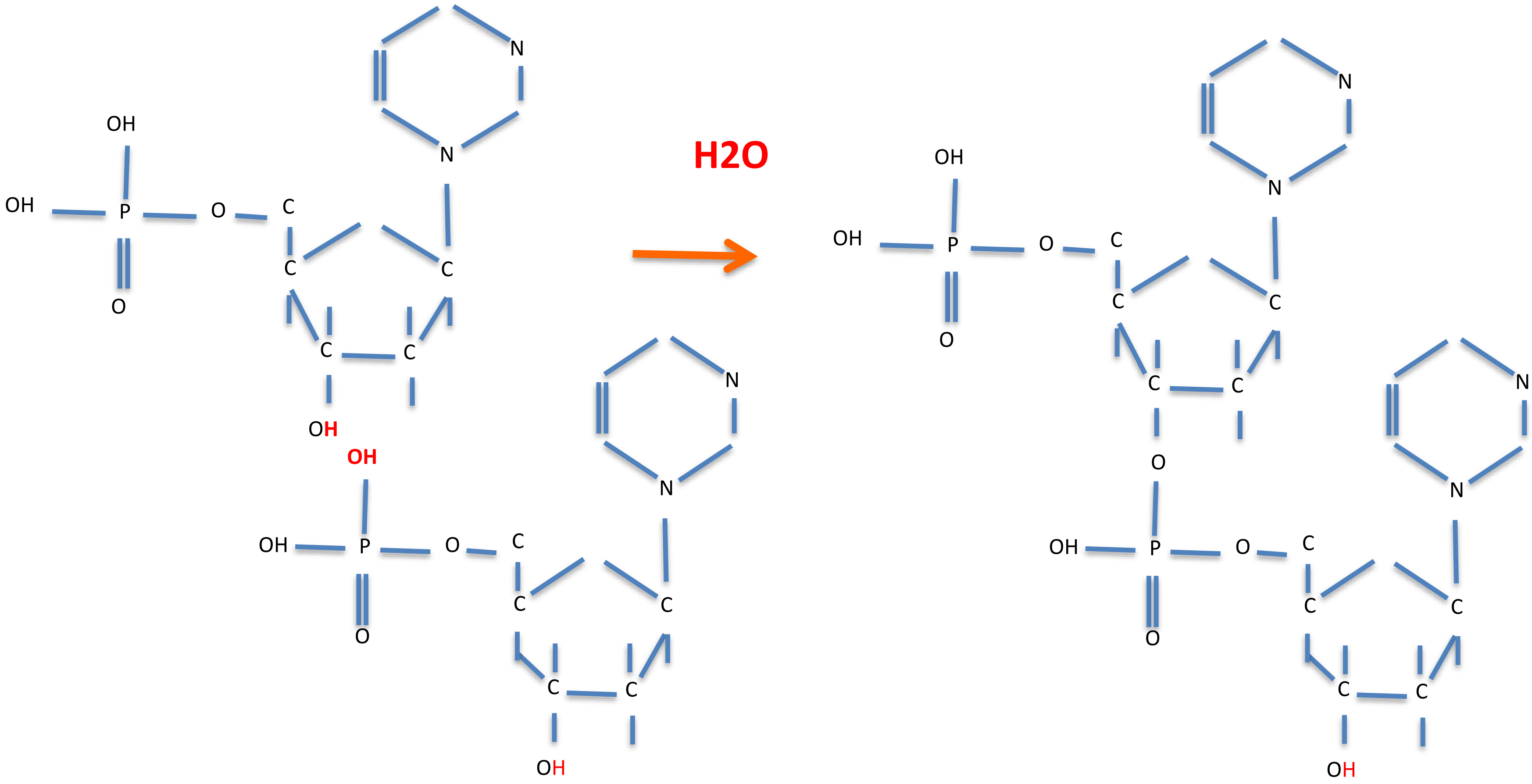
II.4 Nucleotides

Nucleo**T**ides : nucleosides-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)

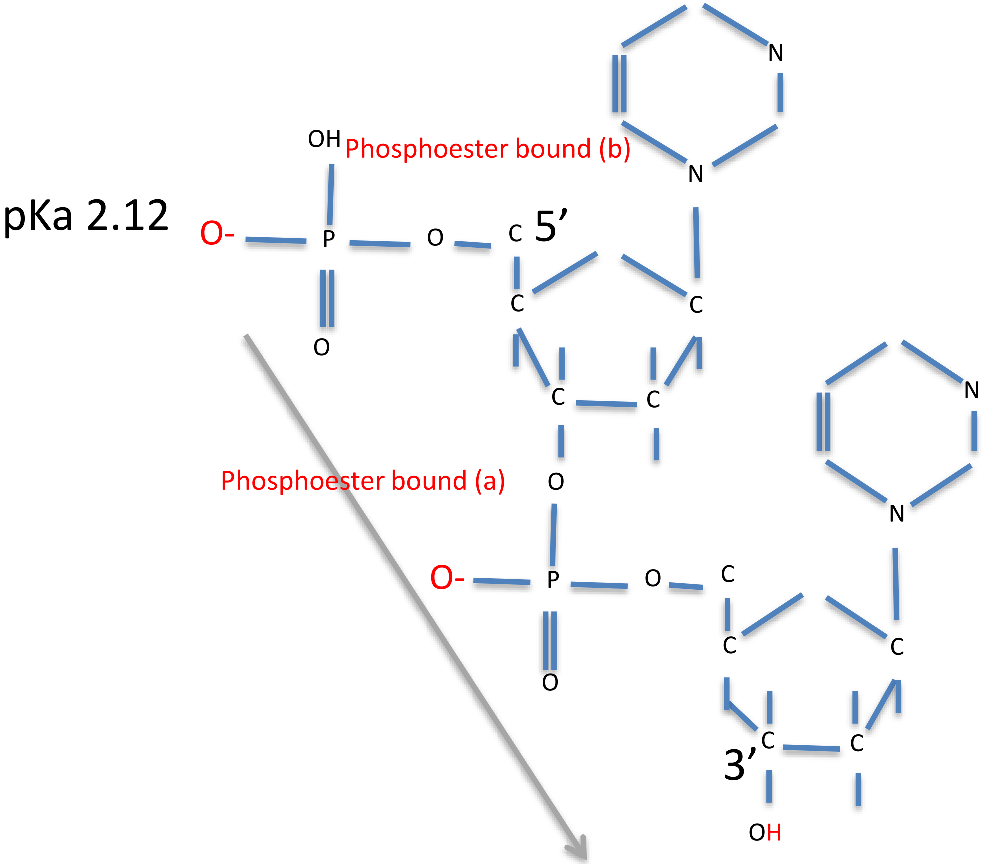
II.5 Phosphodiester bound

Polynucleotidic chain



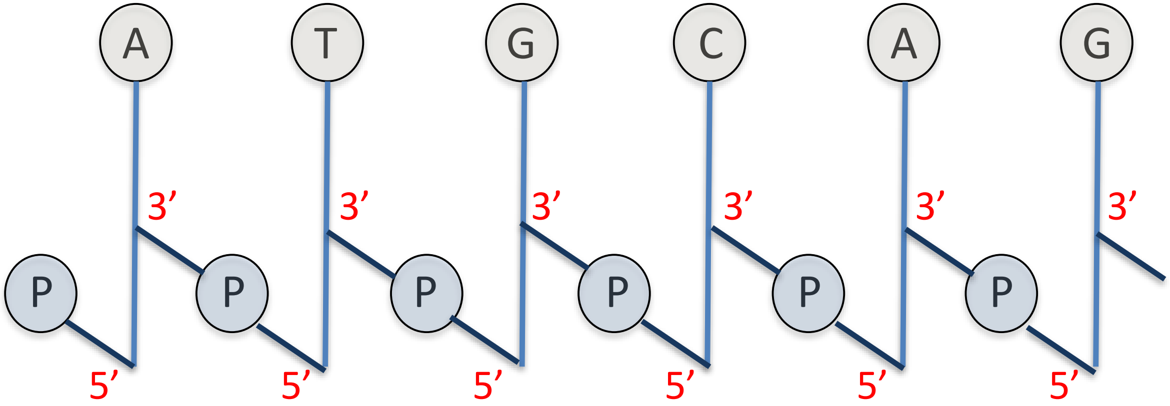
II.5 Phosphodiester bound

Polynucleotidic chain



II.5 Phosphodiester bound

Polynucleotidic chain

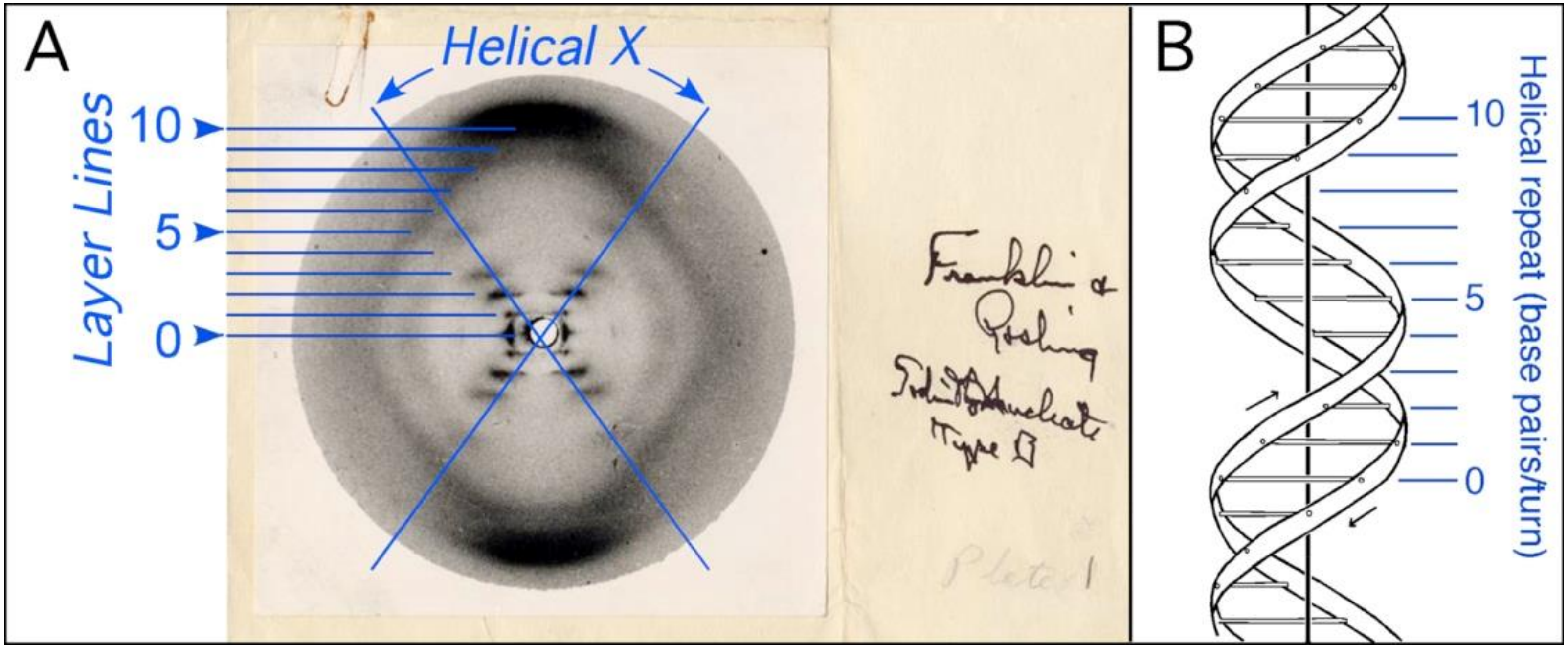


p5’dAp5’dTp5’dGp5’dCp5’dAp5’dG ou 5’-dATGCAG-3’

5’-ATGCAG-3’

II.6 Secondary Structure of DNA

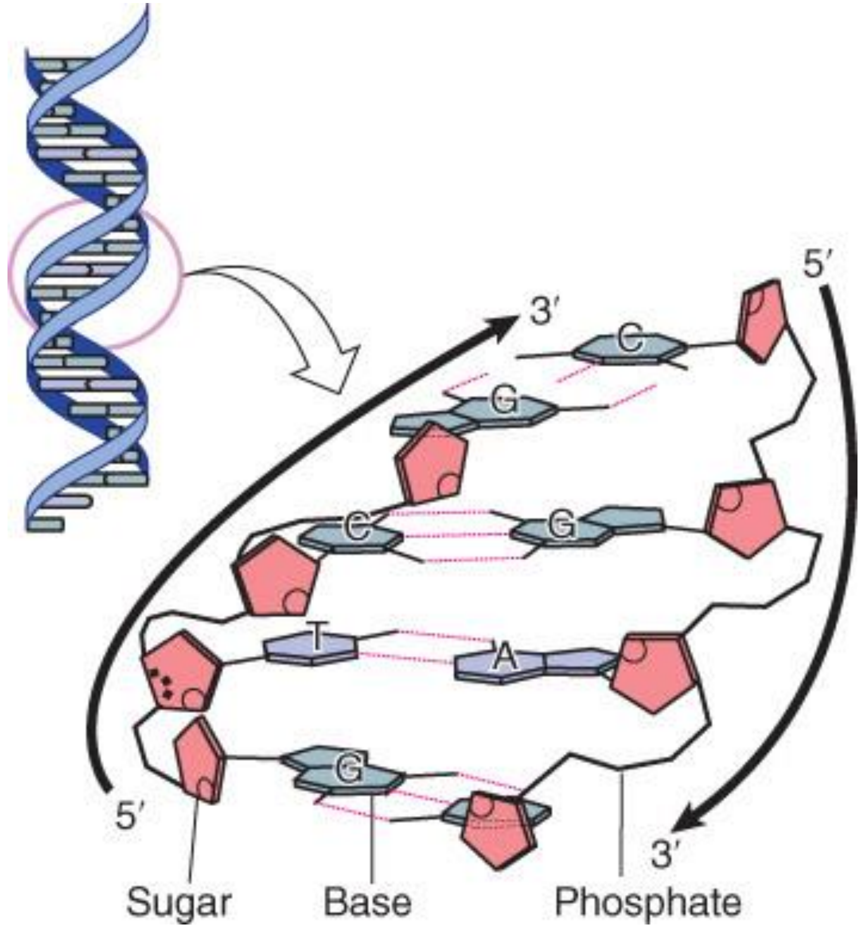
Secondary Structure : The Watson et Crick model



Rosalind Franklin

II.6 Structure Secondaire de la molécule d'ADN

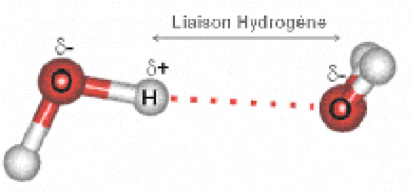
La molécule d'ADN : structure secondaire



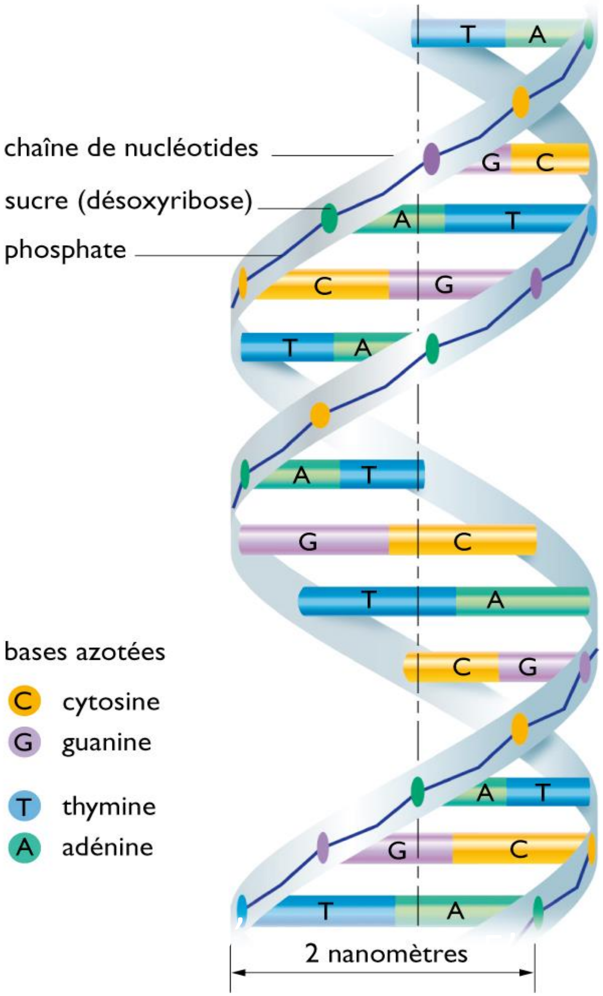
- **Bicaténaire** (2 chaînes polynucléotidiques)
- **Antiparallèle**
- **Hélicoïdale**
- Bases (intérieur) Sucres/phosphates (extérieur)
- **Complémentaire**

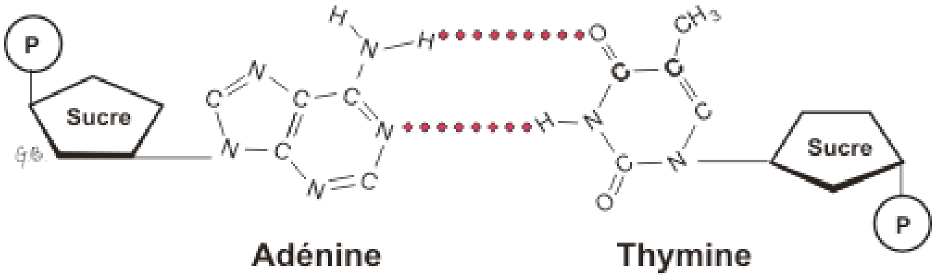
http://biomodel.uah.es/en/model4/dna_fr/dnapairs.htm

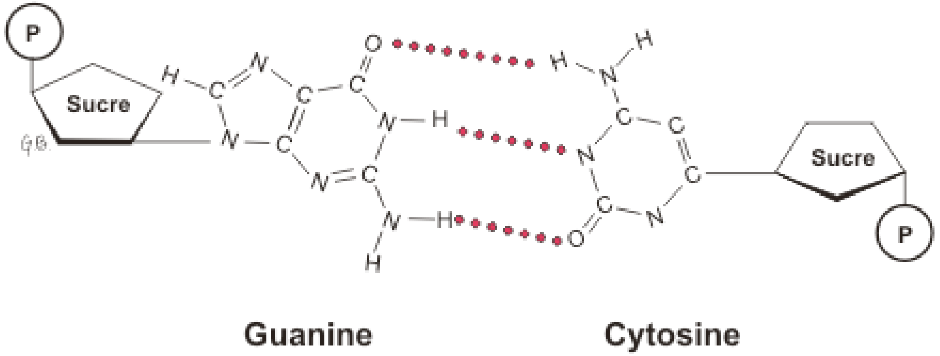
II.6 Secondary Structure of DNA



Rules of complementarity

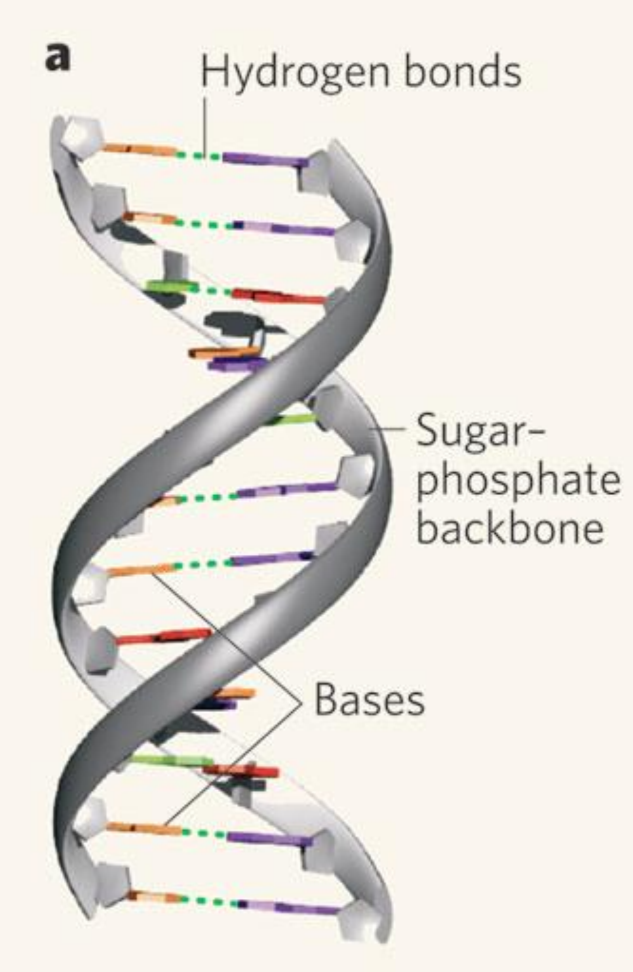






II.6 Secondary Structure of DNA

The DNA molecule : secondary structure and stability



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

https://www.youtube.com/watch?v=o_-6JXLYS-k

Exercise

The proportion of nitrogenous bases in a single stranded DNA was calculated. The G-C content of that molecule is 39,8%.

1. Calculate the proportion of each of the 4 bases (A, T, G, C) in the molecule.

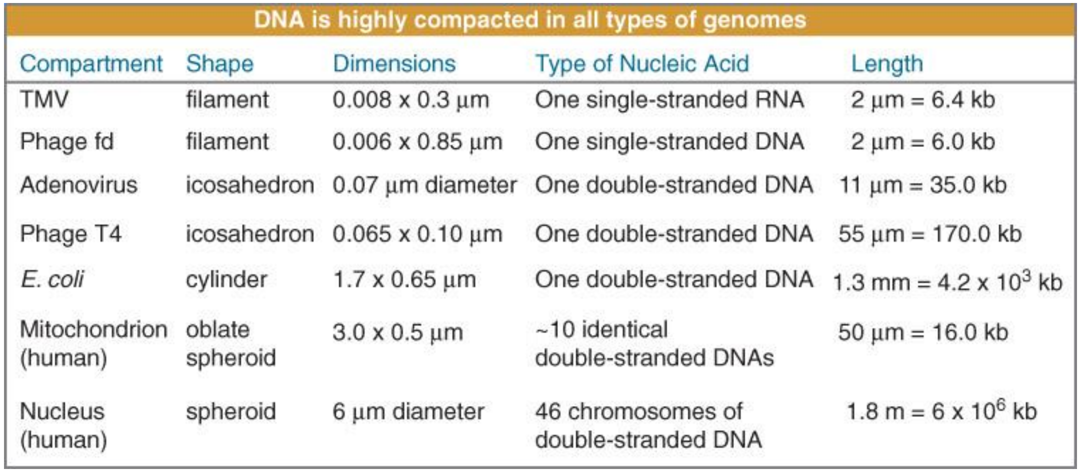
2. Propose a 20 nucleotide long double-stranded DNA molecule that fits this criteria.

III. Structure of genomes

In cells, the molecule of DNA is associated to specialized proteins which will determine the structure by imposing a series of constraints : **Chromosome**

The structuration of DNA will have an impact on :

- its size



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

- its accessibility (protection, replication, transcription, etc..)

III. Structure of genomes

The genome of each organism is organized into a specific number of chromosomes, whose shape is caracteristic

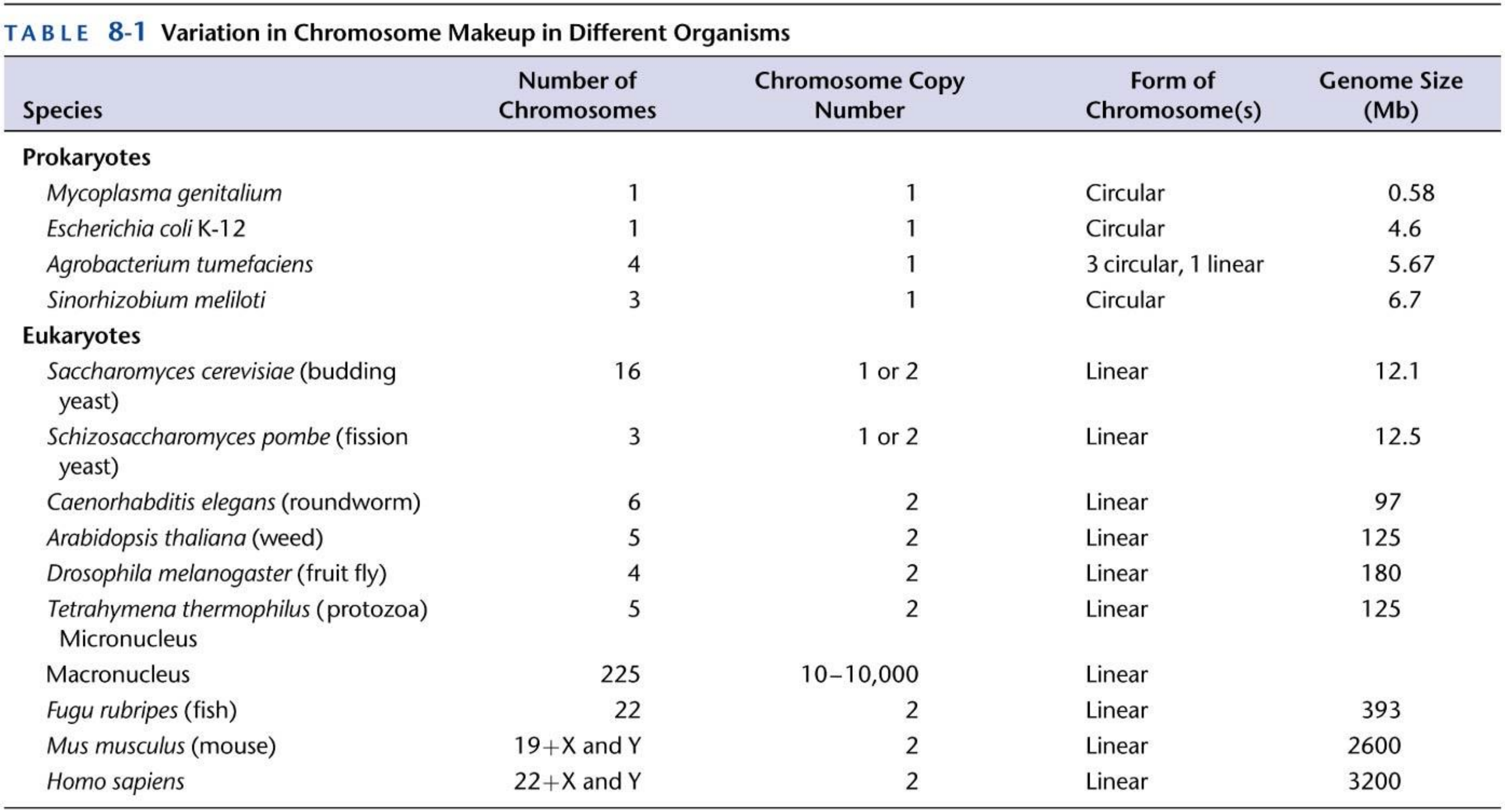
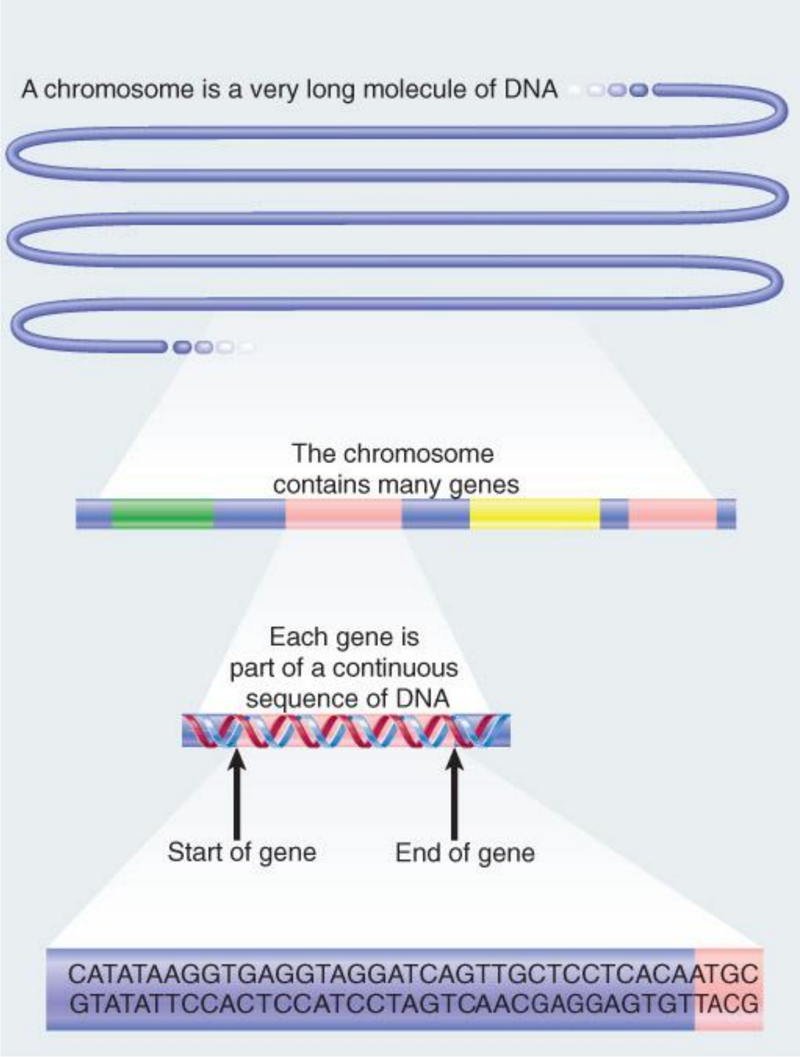


TABLE 8-1 Variation in Chromosome Makeup in Different Organisms

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 of genomes

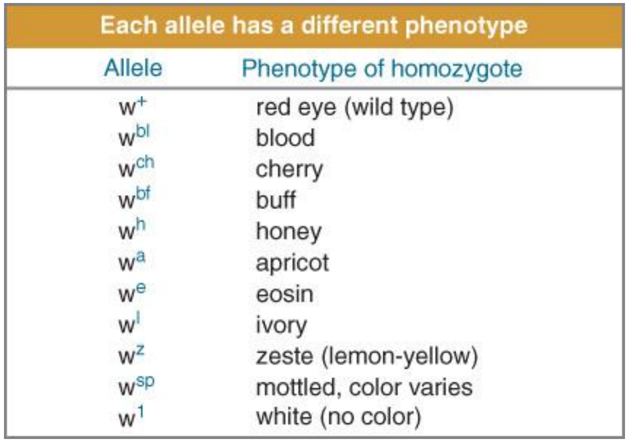
What's a gene



<u>Allele</u> : one of the versions of a gene

<u>Locus</u> : position of a gene on the chromosome

In an individual, or a population, more than one allele can be found at a single locus



Each allele has a different phenotype

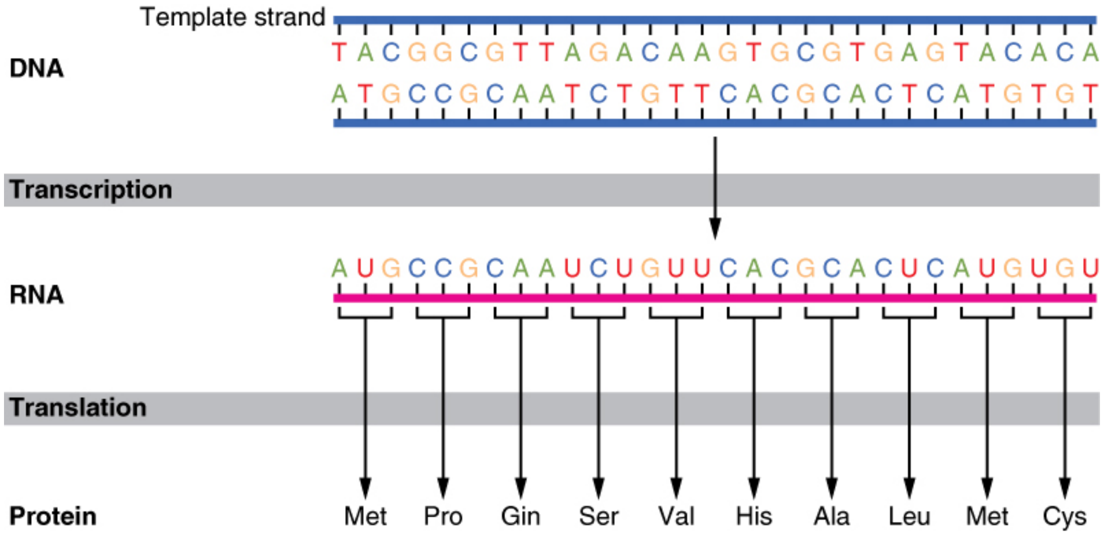
Allele	Phenotype of homozygote
w^+	red eye (wild type)
w^{bl}	blood
w^{ch}	cherry
w^{bf}	buff
w^h	honey
w^a	apricot
w^e	eosin
w^i	ivory
w^z	zeste (lemon-yellow)
w^{sp}	mottled, color varies
w^1	white (no color)

The gene

The genetic code

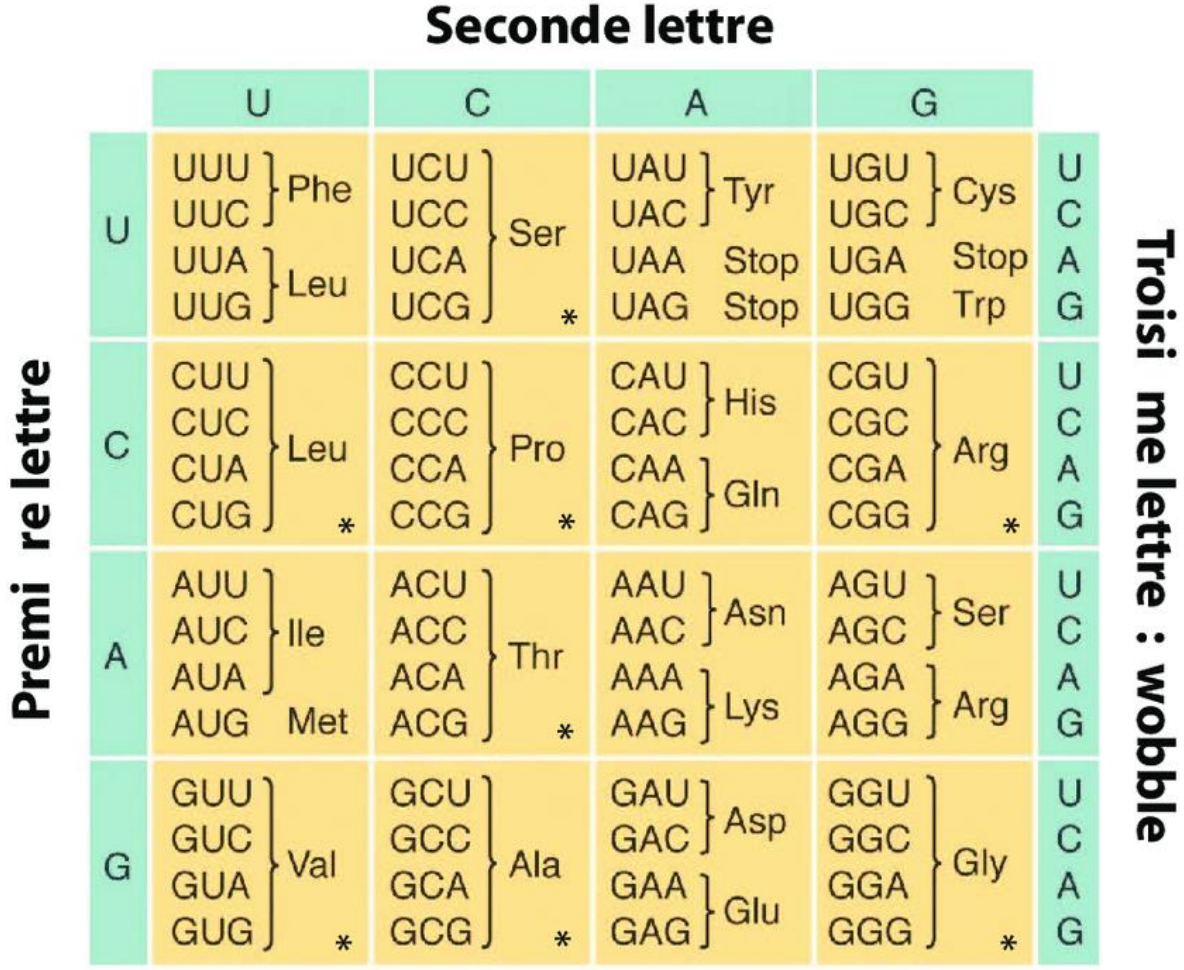
<u>Genetic Code</u> : the relation between the sequence of nucleotides in the DNA and the sequence of amino acids in the polypeptides

<u>Genetic Information</u> : genes + regulatory sequences



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

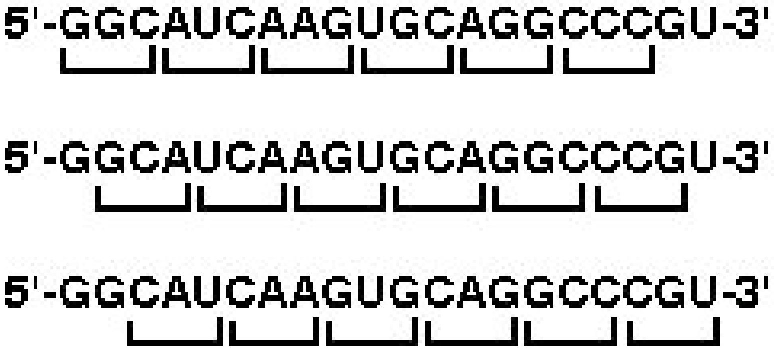
The genetic code



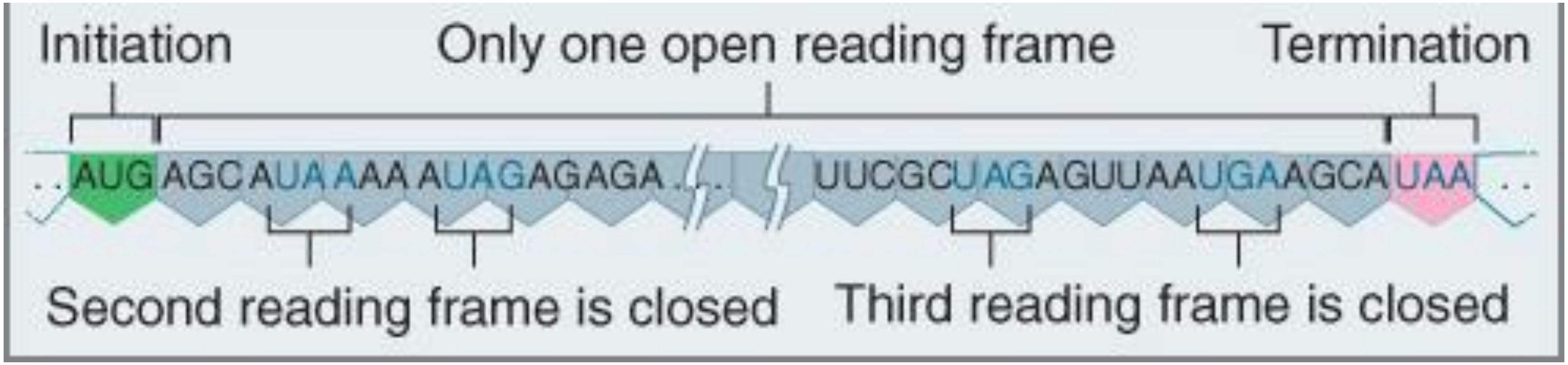
	Seconde lettre				
Première lettre	U	C	A	G	Troisième lettre : wobble
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA Stop	UGA Stop	A
	UUG	UCG *	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Gln	CGA	A
	CUG *	CCG *	CAG	CGG *	G
A	AUU Ile	ACU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

III. Structure of genomes

genetic code : the open reading frame (ORF)



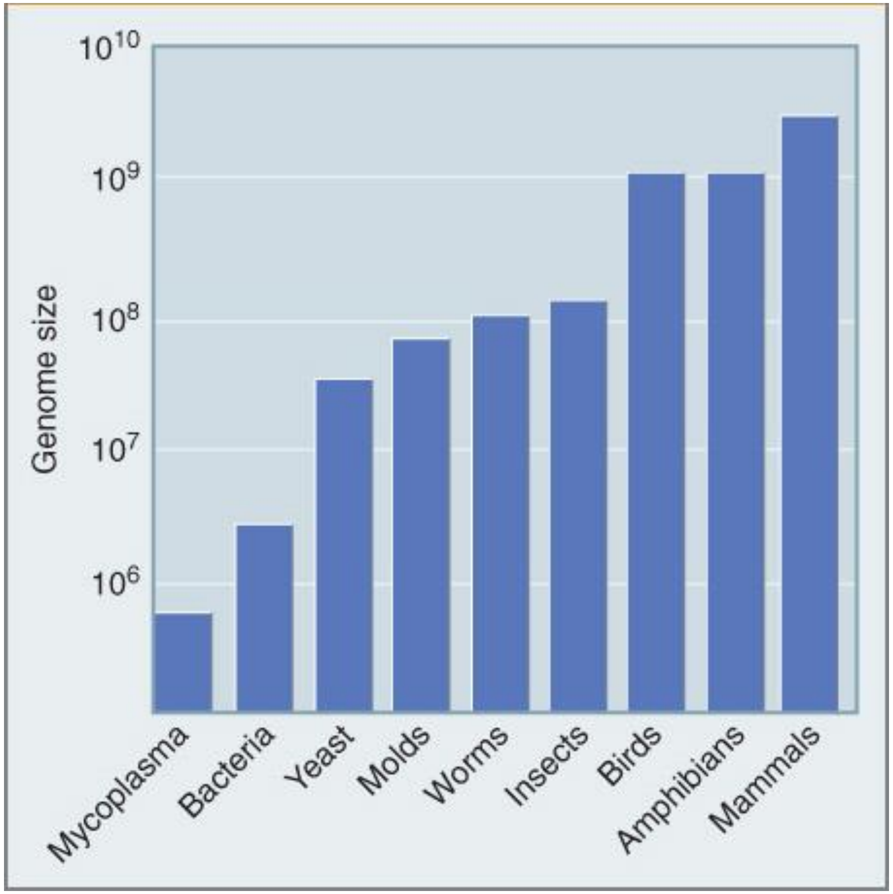
<u>A particular DNA sequence encompases 3 independent ORFs</u>



<u>In general, only one of them is used</u>

III. Structure of genomes

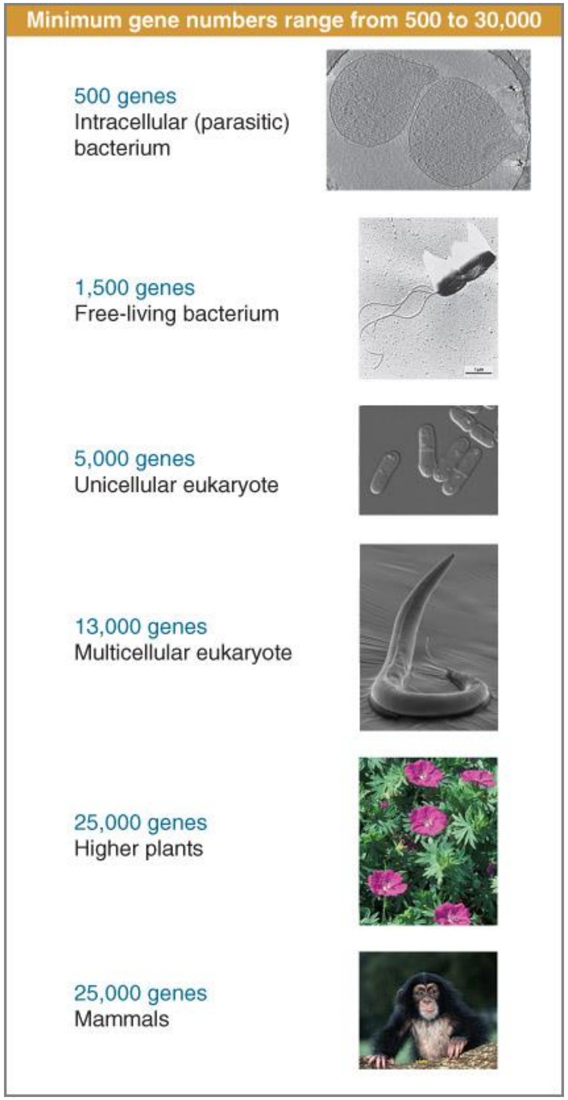
The size of the genome is roughly correlated to the complexity of the organism



C-value : size of a genome in bas pairs (bp) or pico-gram

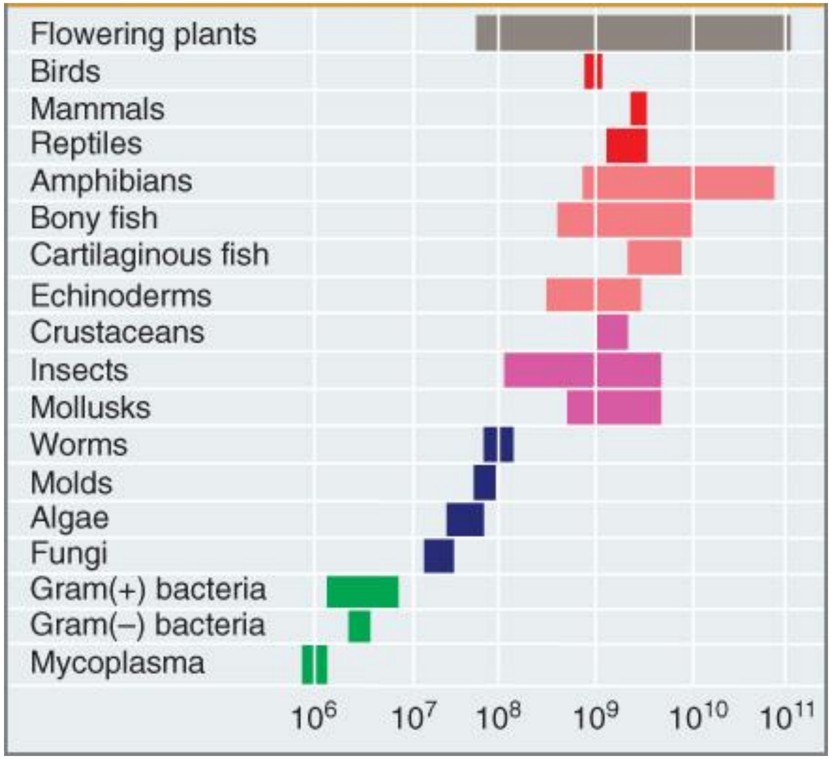
III. Structure of genomes

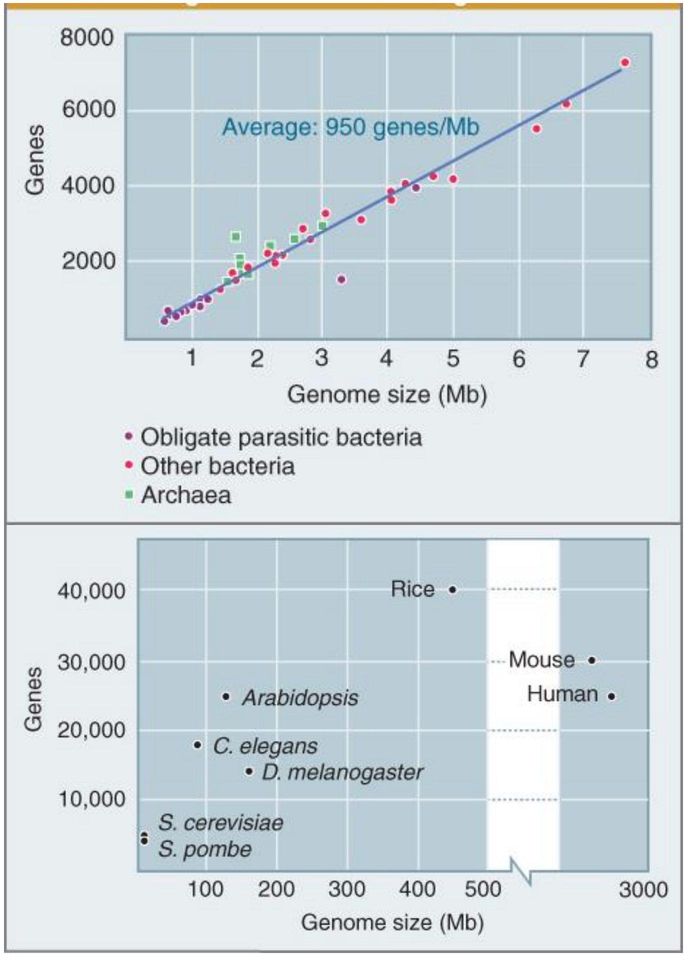
The number of genes increases as the complexity of the organism increases



III. Structure of genomes

The C-value paradox : There is not a strict correlation between the size of the genome and the complexity of the organism or the number of genes

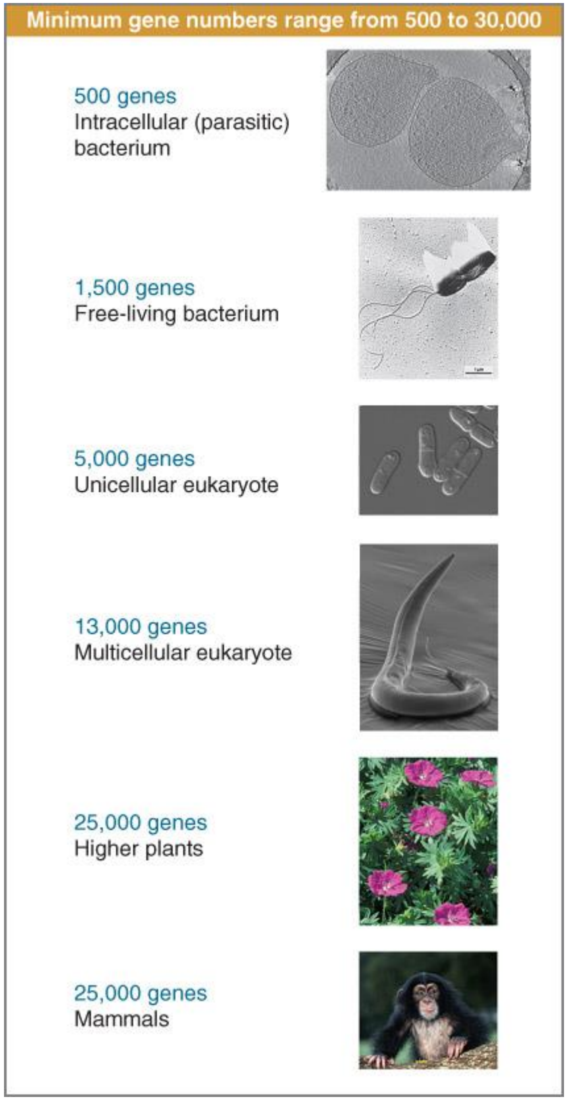




https://bio.libretexts.org/TextMaps/Genetics/Book%3A_Working_with_Molecular_Genetics_(Hardison)/Unit_I%3A_Genes%2C_Nucleic_Acids%2C_Genomes_and_Chromosomes/4%3A_Genomes_and_Chromosomes/4.5%3A_Sizes_of_genomes_-_The_C%E2%80%91value_paradox

III. Structure of genomes

The number of genes roughly increases with the complexity of the organism



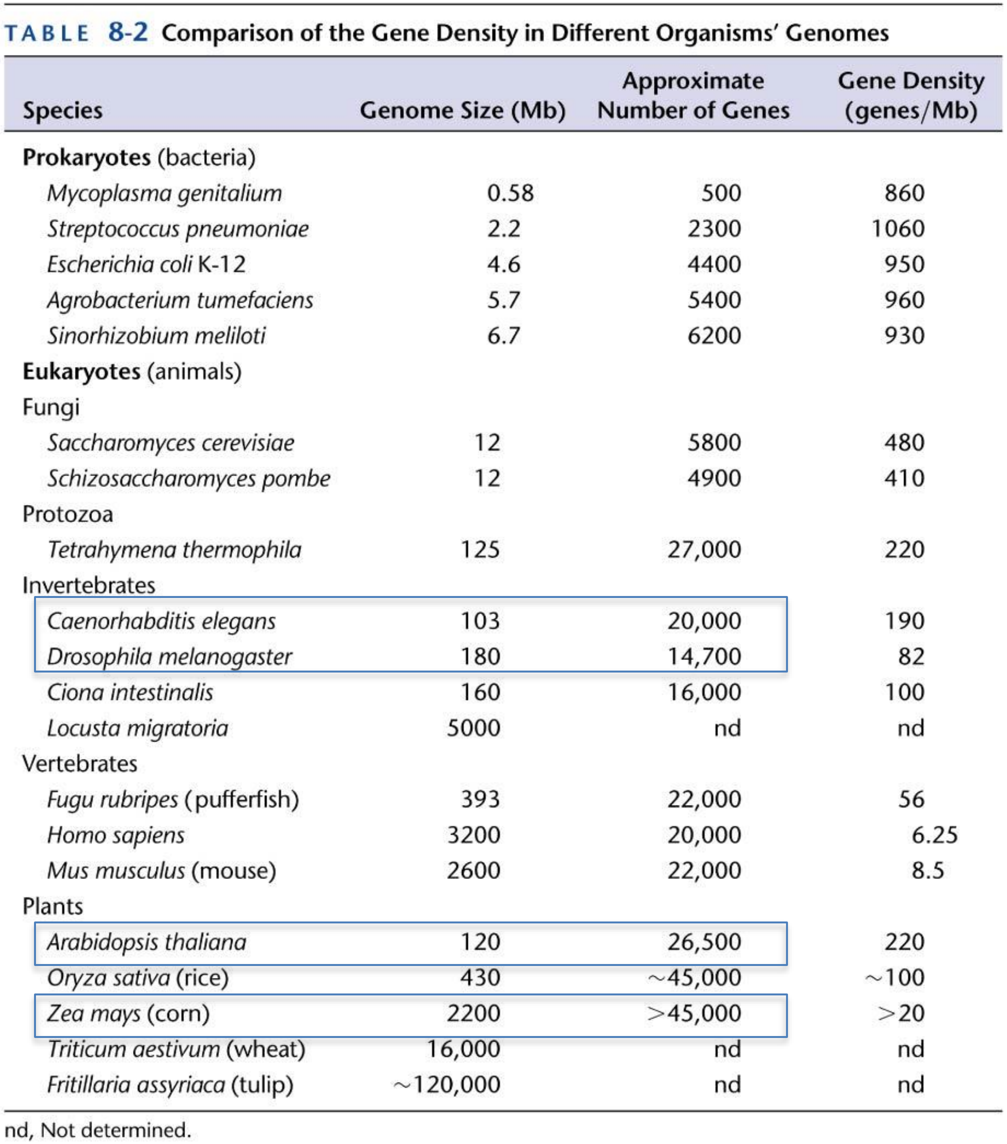


TABLE 8-2 Comparison of the Gene Density in Different Organisms' Genomes

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.

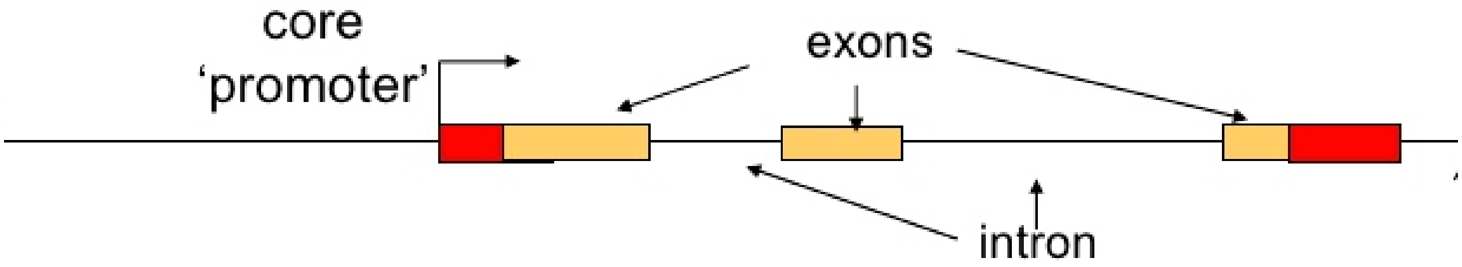
III. Structure of genomes

Structure of genes

In **procaryotes**, genes are not interrupted (monocistronic)

→ ADN et ARNm sont colinéaires

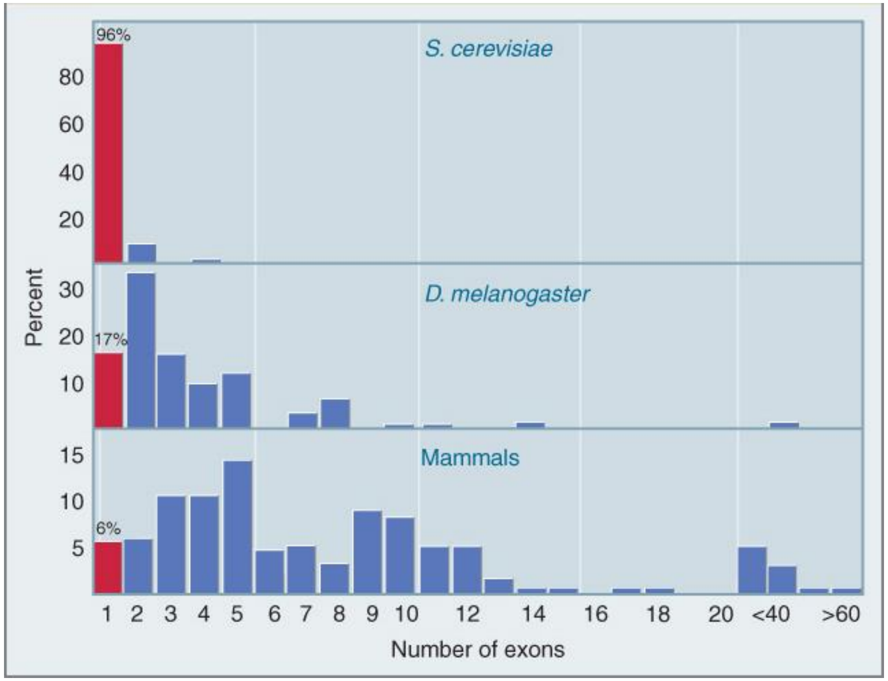
In eukaryotes, genes are longer and mostly interrupted (introns/exons)

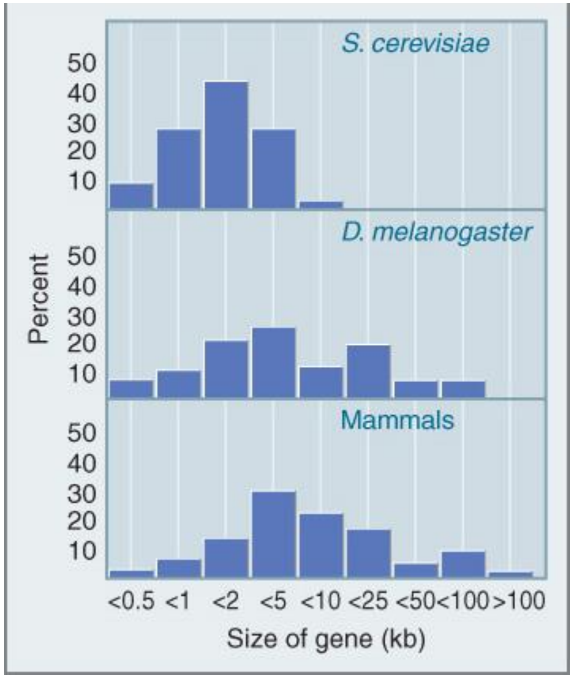


http://www.nature.com/scitable/topicpage/what-is-a-gene-colinearity-and-transcription-430

III. Structure of genomes

Structure of genes



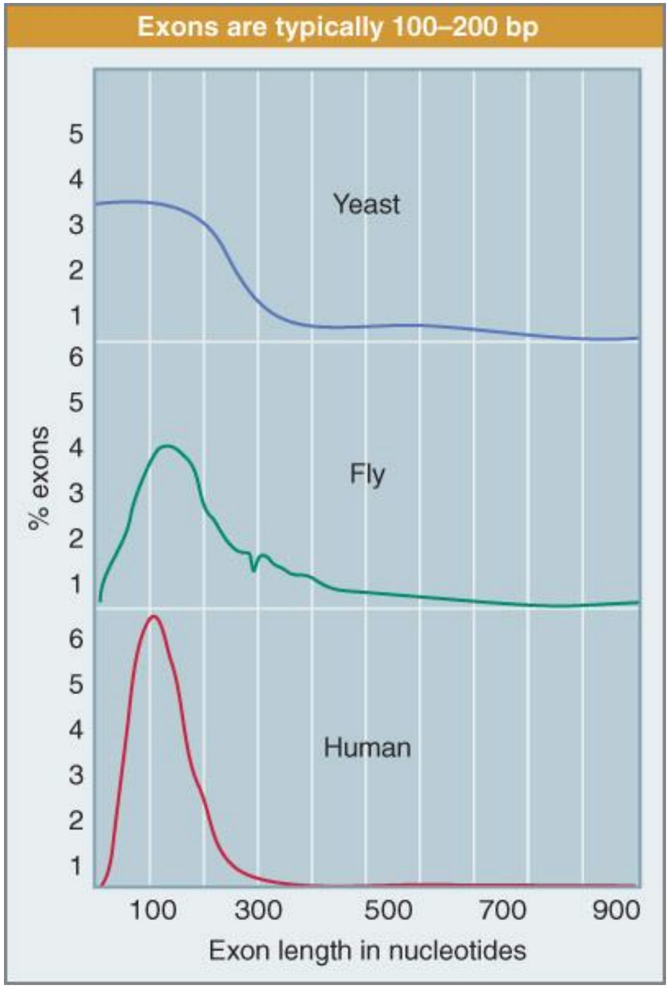


During evolution, genes went from mostly short and not interrupted to mostly longer and interrupted

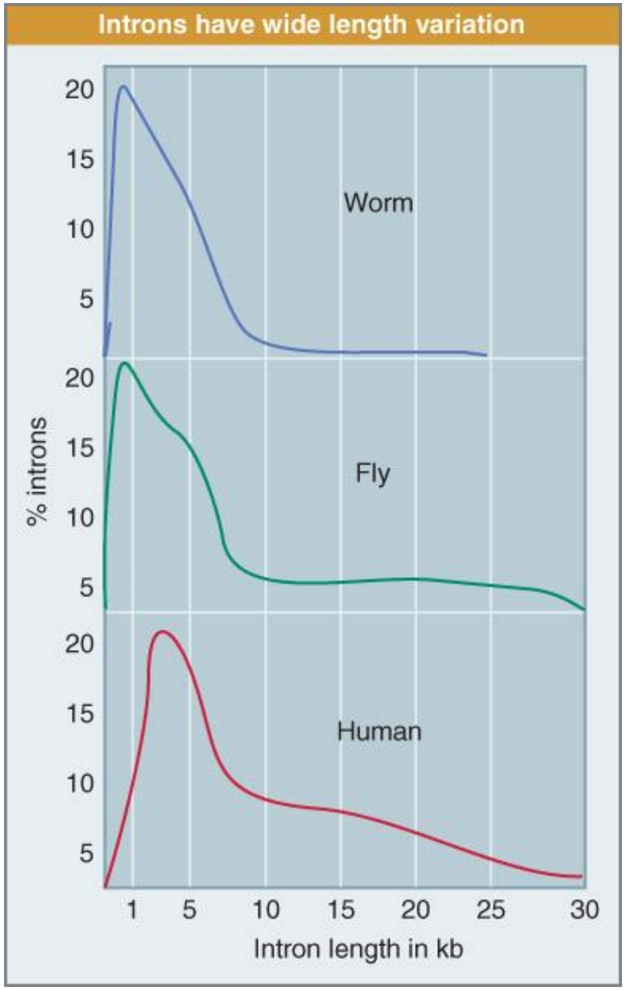
III. Structure of genomes

Structure of genes

exons are short and with constant sizes throughout evolution

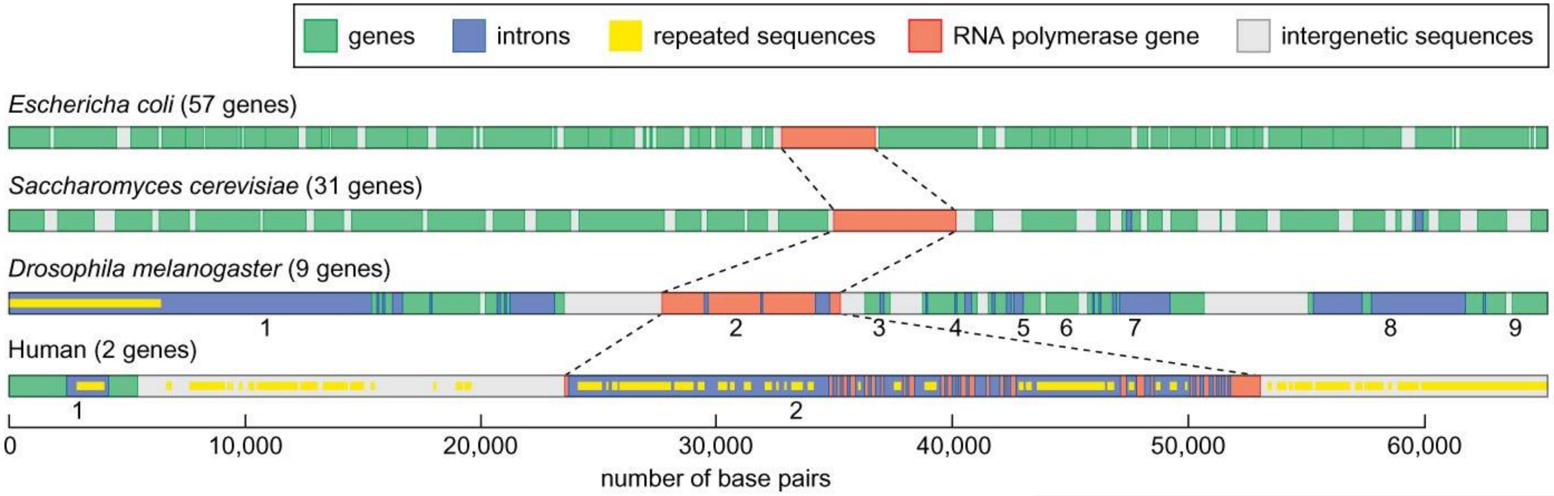


introns are more variable in size and become longer throughout evolution



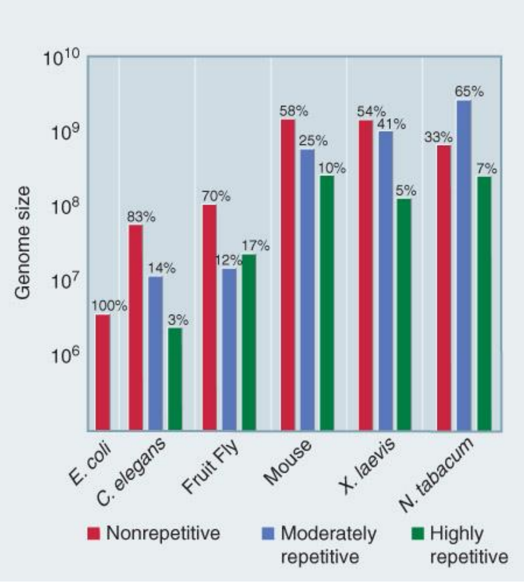
III. Structure du génome

More complex organisms have lower gene densities



Moderatly or highly repeated DNA (junk DNA) : in general, non coding (transposons, satellite DNA, low complexity DNA)

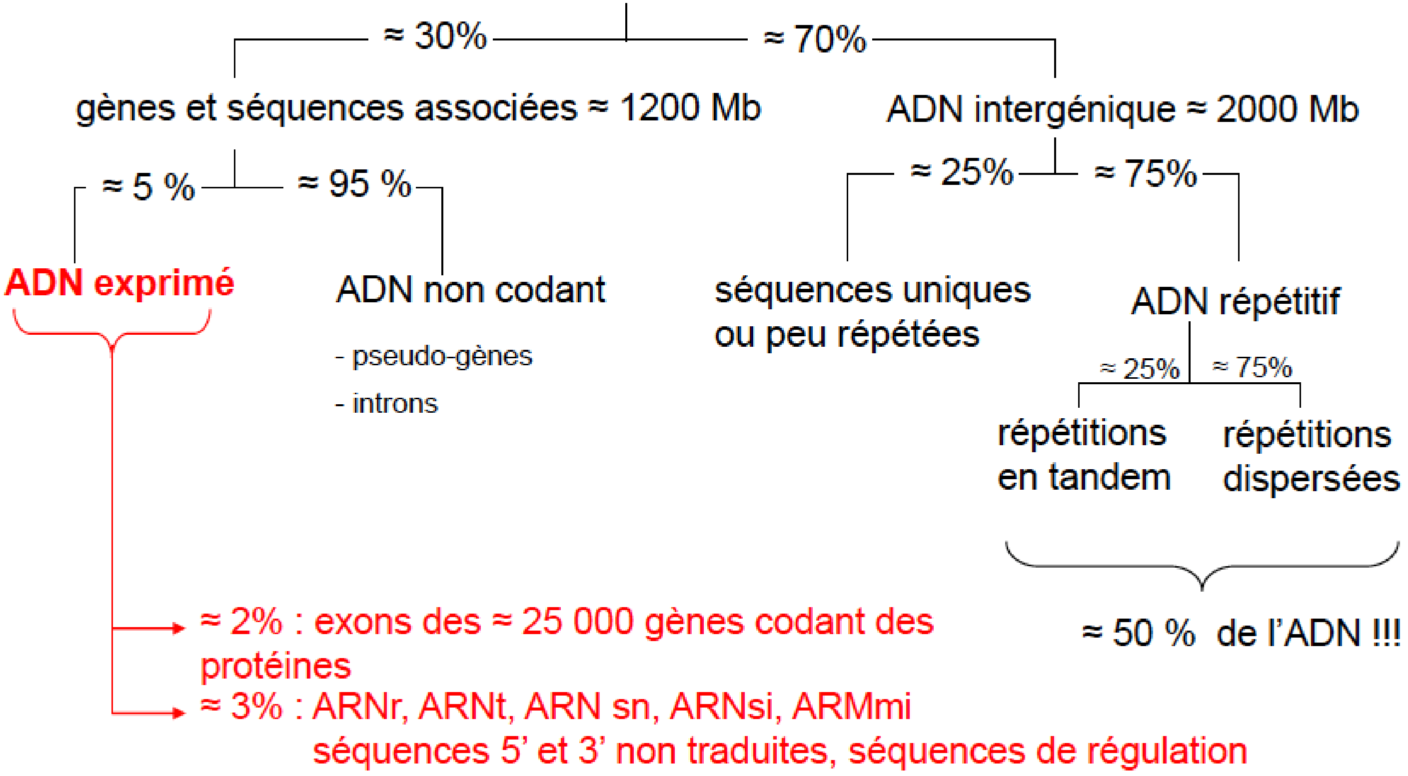
Non repeated DNA : coding regions (genes)



III. Structure du génome

Classical composition of a superior eukaryote genome

The majority of DNA is non coding!



≈ 30%

≈ 70%

gènes et séquences associées ≈ 1200 Mb

ADN intergénique ≈ 2000 Mb

≈ 5 %

≈ 95 %

≈ 25%

≈ 75%

ADN exprimé

ADN non codant

- pseudo-gènes
- introns

séquences uniques
ou peu répétées

ADN répétitif

≈ 25%

≈ 75%

répétitions
en tandem

répétitions
dispersées

≈ 50 % de l'ADN !!!

≈ 2% : exons des ≈ 25 000 gènes codant des protéines

≈ 3% : ARNr, ARNt, ARN sn, ARNsi, ARMmi
séquences 5' et 3' non traduites, séquences de régulation

https://www.youtube.com/watch?v=uaq2XdXgfPY

IV. Mutations

Mutations are modifications of DNA sequences

Spontaneous mutations (mistakes during replication)

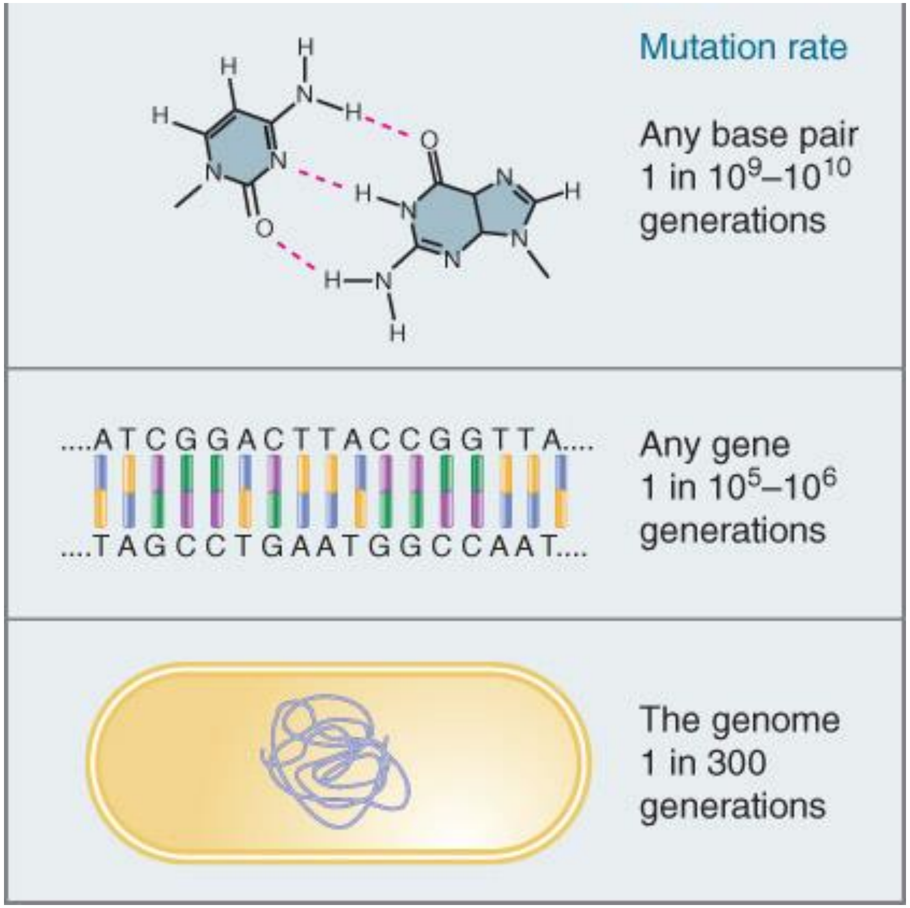
Inducible mutations (chemical modifications)

I**nsertion of exogenous DNA** (viruses, transposons)

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

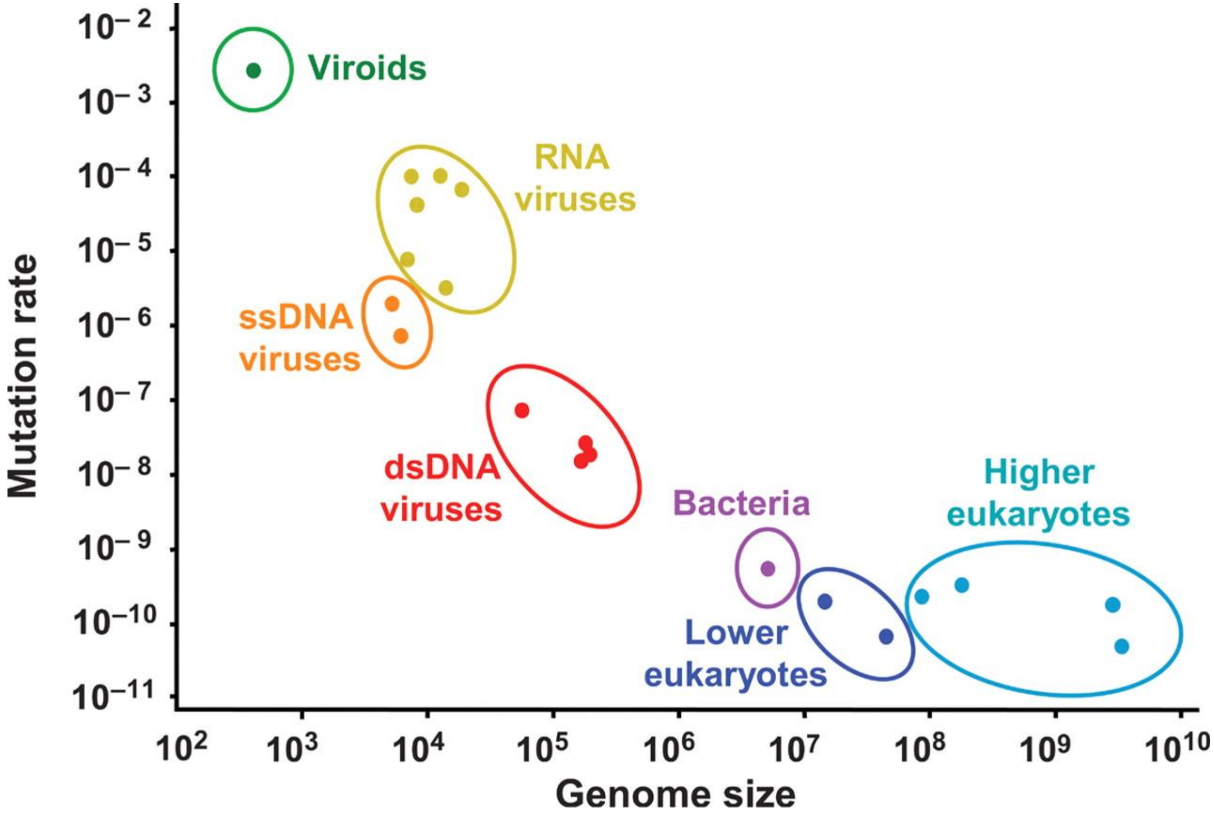
IV. Mutations

Spontaneous Mutations



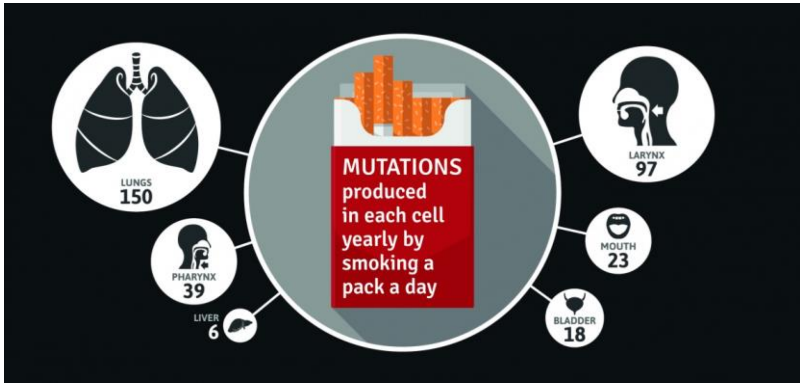
IV. Mutations

Spontaneous Mutations



IV. Mutations

Inducible Mutations and mutagens



(Science, Nov 2016)

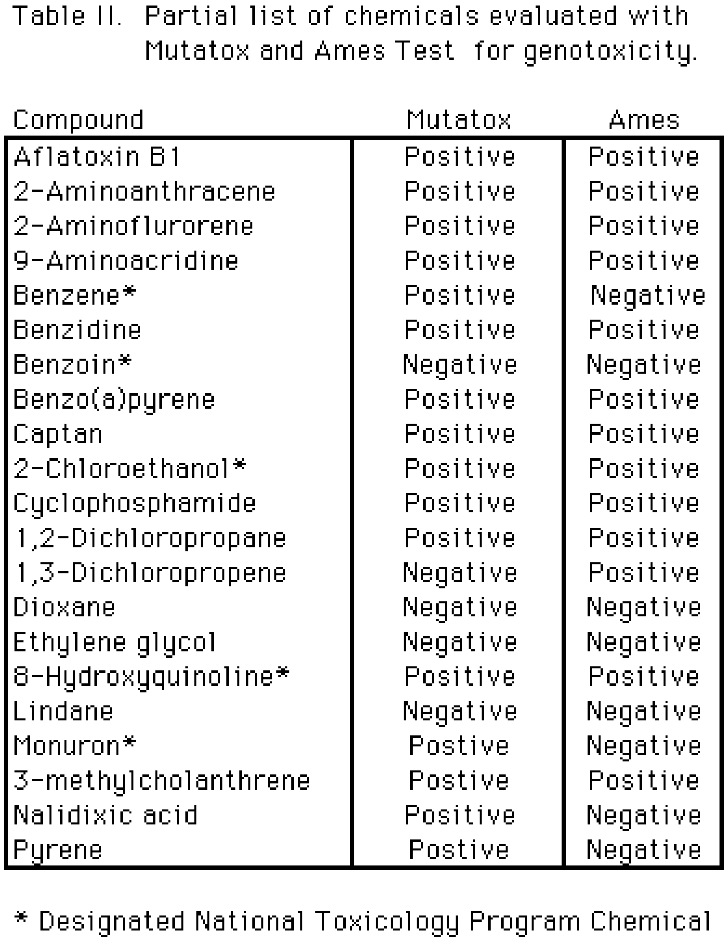
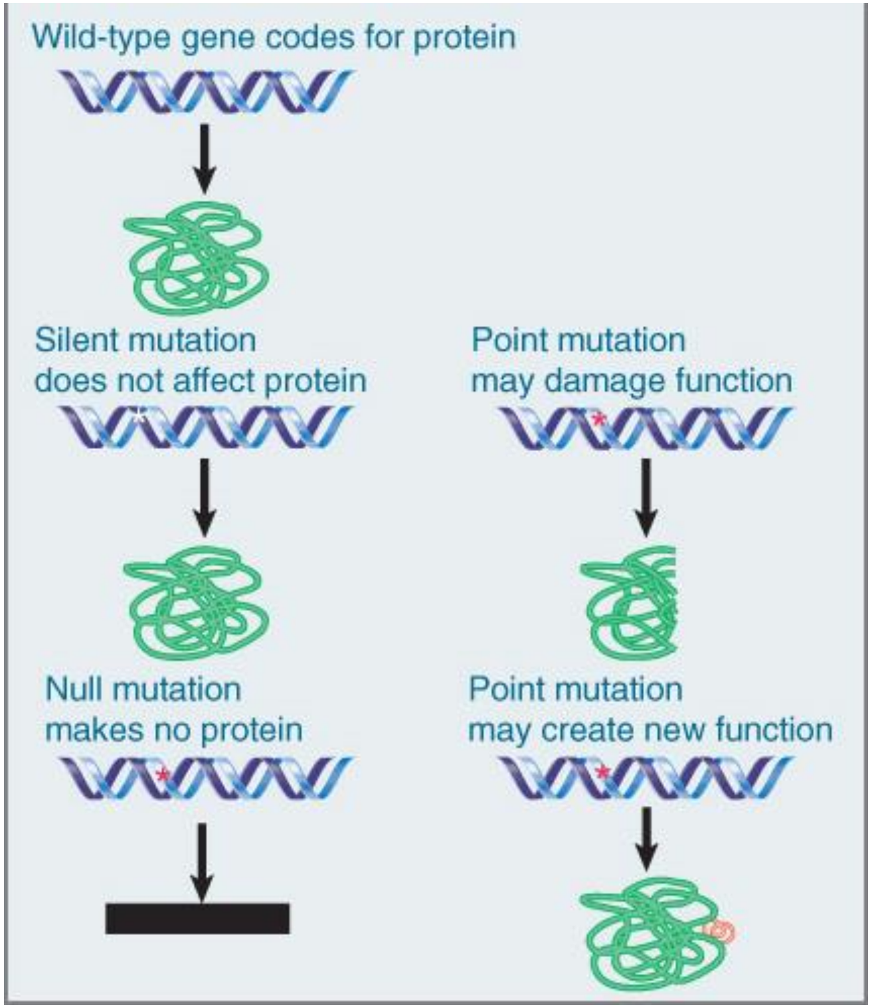


Table II. Partial list of chemicals evaluated with Mutatox and Ames Test for genotoxicity.

Compound	Mutatox	Ames
Aflatoxin B1	Positive	Positive
2-Aminoanthracene	Positive	Positive
2-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

IV. Mutations



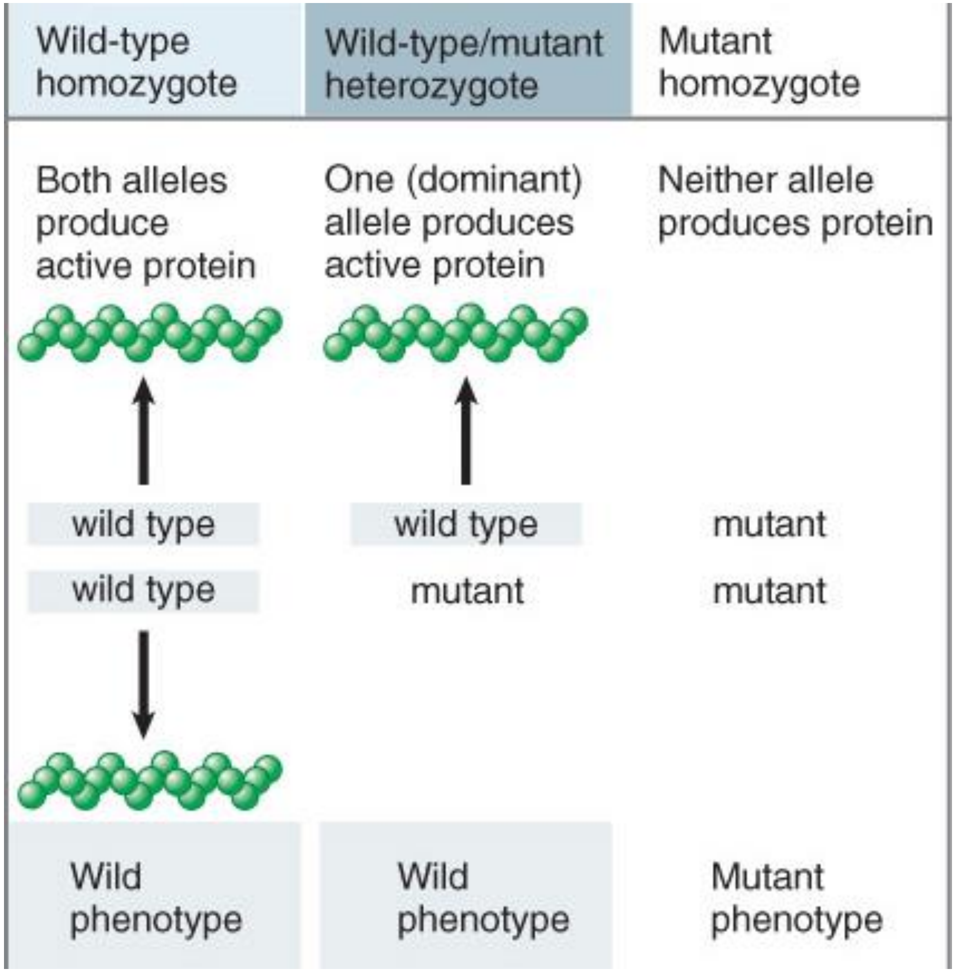
Mutations can be :

Silent (no effect)

Null (loss of function)

New function (gain of function)

IV. Mutations



Wild-type homozygote	Wild-type/mutant heterozygote	Mutant homozygote
Both alleles produce active protein	One (dominant) allele produces active protein	Neither allele produces protein
wild type	wild type	mutant
wild type	mutant	mutant
Wild phenotype	Wild phenotype	Mutant phenotype

- Mutations are in general recessive
- Recessive alleles produce inactive proteins

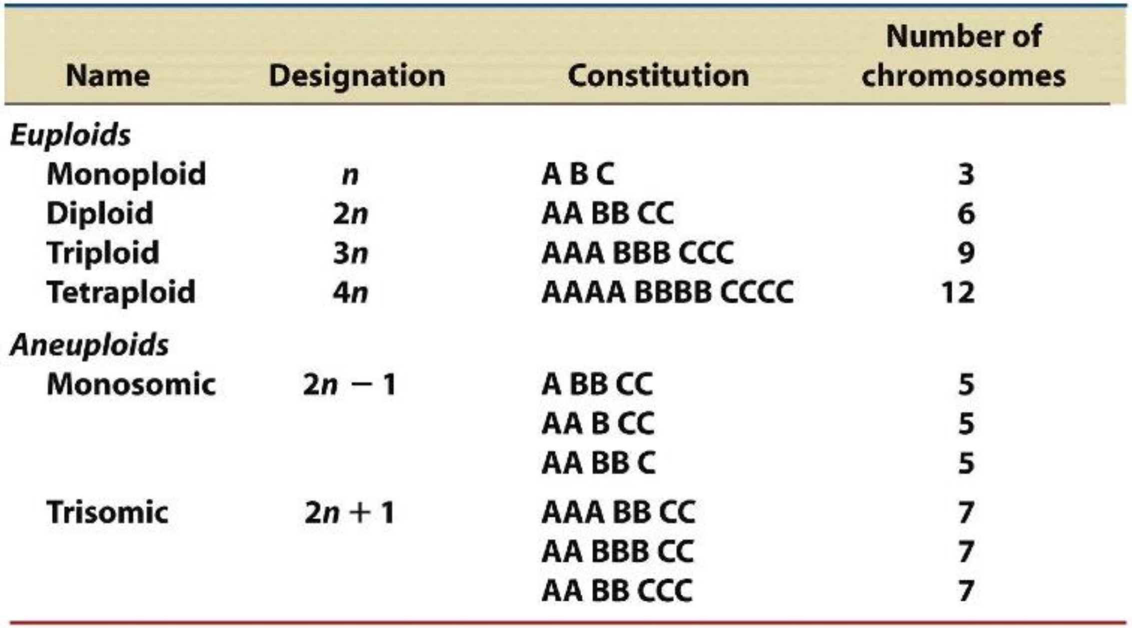
IV. Mutations

1 : **Chromosomal mutations** : modifications of the structure or copy number of chromosomes

n (germinal cell, monoploïd), **2n** (somatic cell, diploïd) : **normal euploïdy**

3n, **4n,…8n** : **abberent euploïdy**

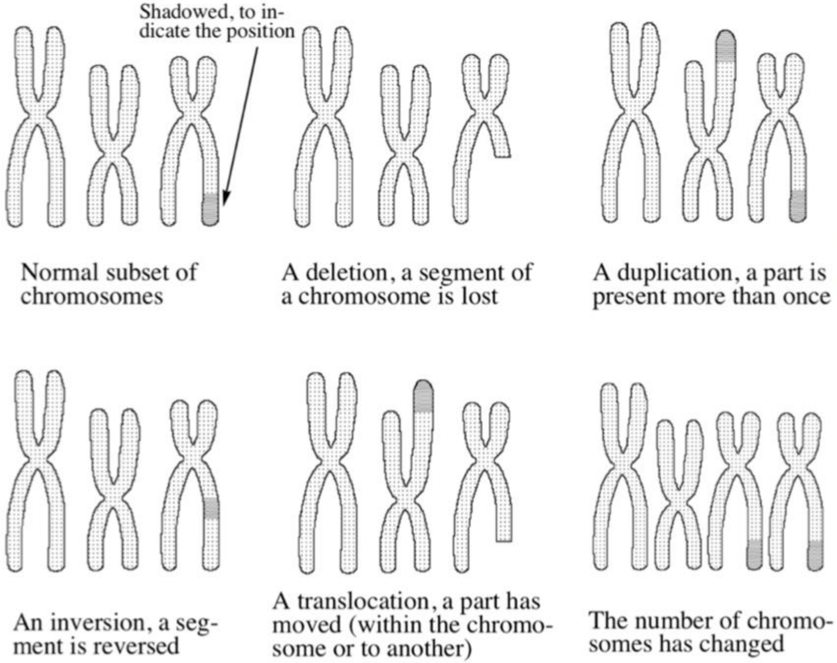
2n-1, **2n+1** : **aneuploïdy**



Name	Designation	Constitution	Number of chromosomes
Euploids			
Monoploid	n	A B C	3
Diploid	$2n$	AA BB CC	6
Triploid	$3n$	AAA BBB CCC	9
Tetraploid	$4n$	AAAA BBBB CCCC	12
Aneuploids			
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. Mutations

1 : **Chromosomal mutations** : modifications of the structure or copy number of chromosomes



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

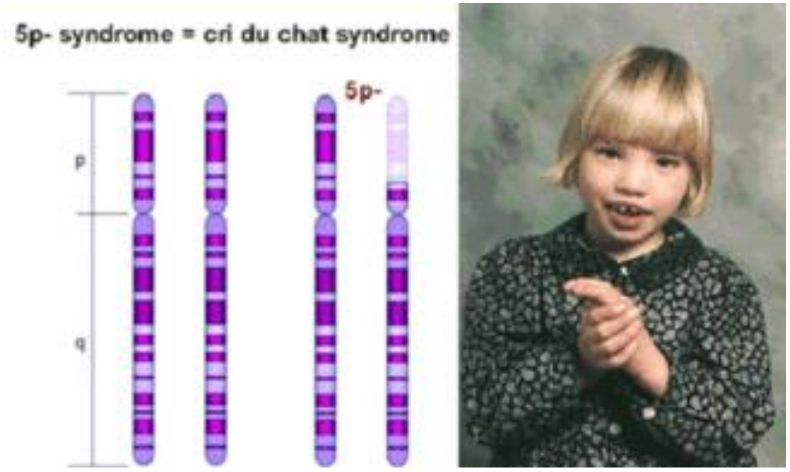
A translocation, a part has moved (within the chromosome or to another)

The number of chromosomes has changed

Chromosome mutations

- Loss of DNA
- Move of DNA

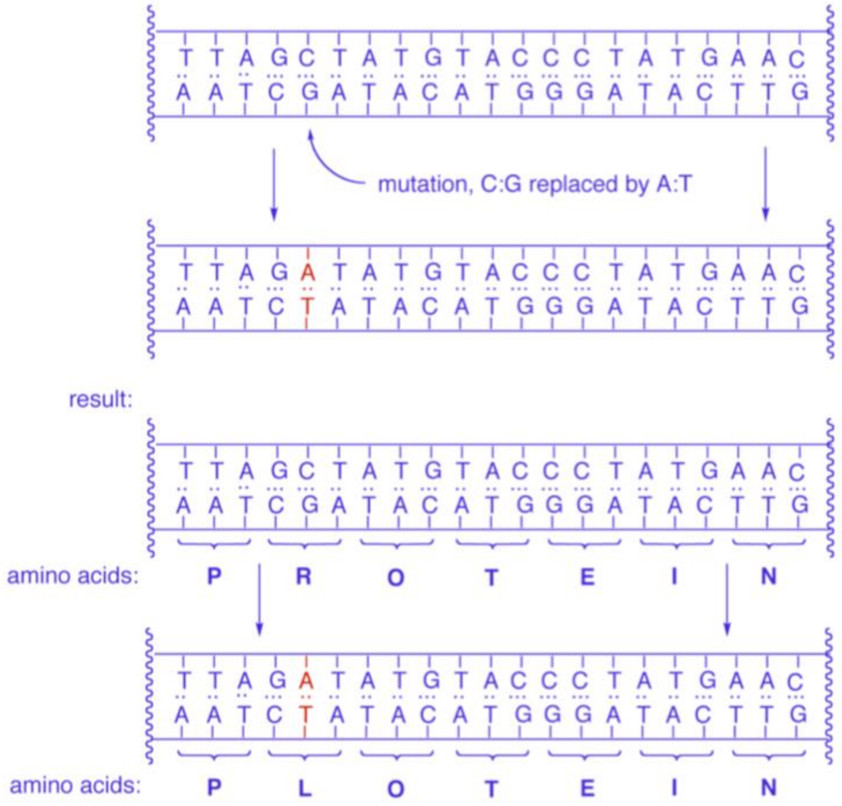
➡Serious for the cell



IV. Mutations

Substitutions

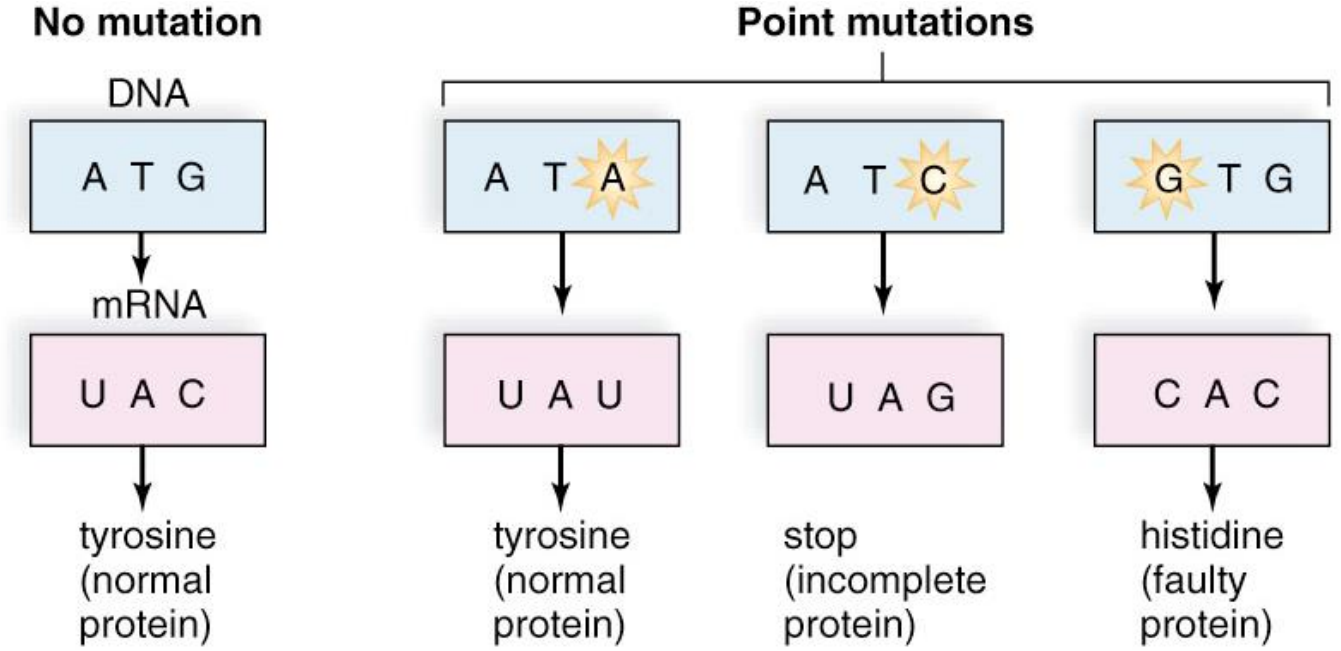
Base pair substitution



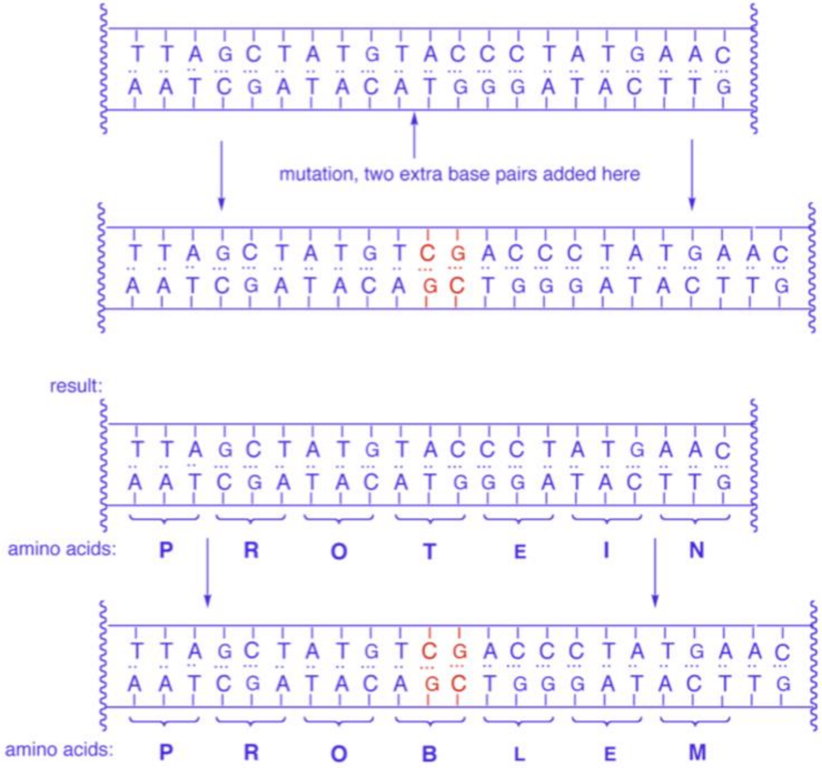
a limited damage

IV. Mutations

Genetic code



Frame shift mutation



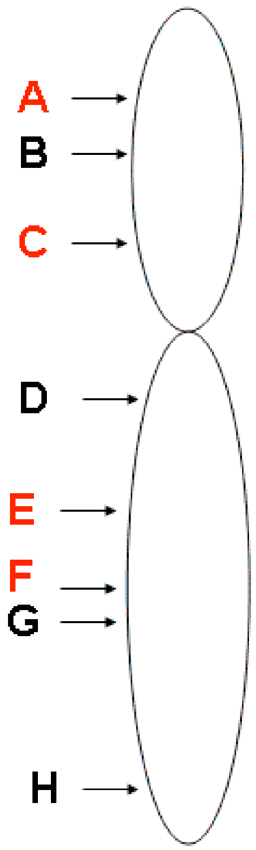
more
serious

Substitutions may change the sequence of the final polypeptide

V. Decoding the genomes

Genetic maps

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).

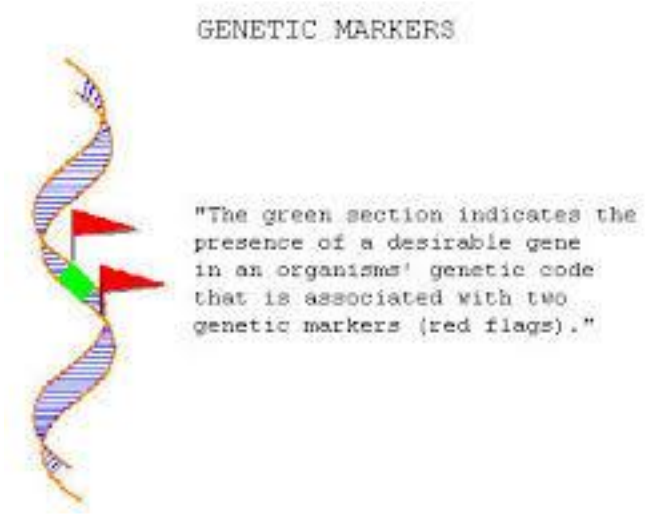
- If pedigree studies show a high incidence of co-occurrence, those genes will be located close together on a genetic map.

V. Decoding the genomes

The genetic/molecular markers :

Définition : fragment (sequence) of polymorphic DNA that is specifically identifiable in individuals

features: multiallelic (polymorphism), universal et inherritable



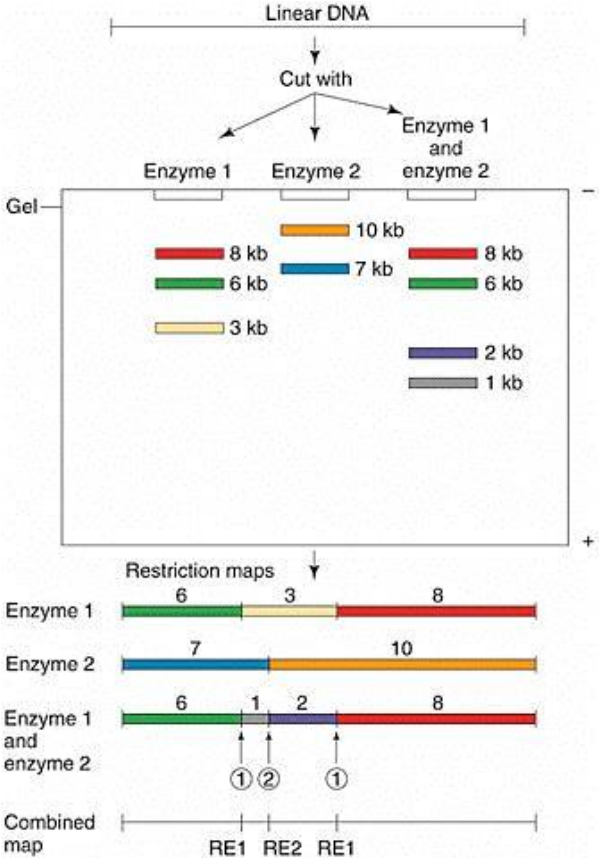
GENETIC MARKERS

"The green section indicates the presence of a desirable gene in an organisms' genetic code that is associated with two genetic markers (red flags)."

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

V. Decoding the genomes

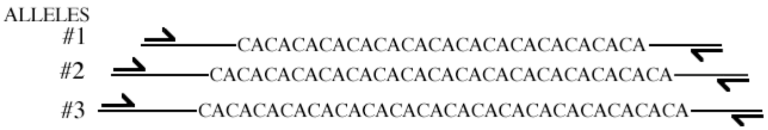
RFLP : Restriction length fragment polymoprhism



Gene coupling

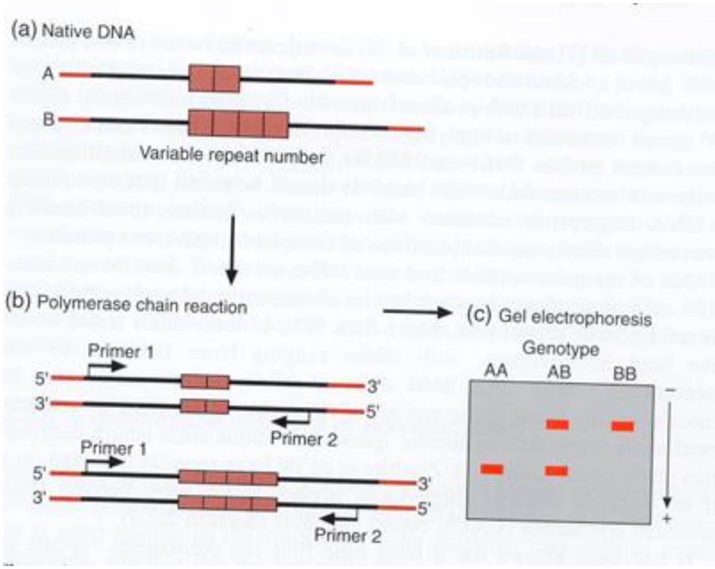
V. Decoding the genomes

STRs : Short tandem repeats (VNTR)



Detected by RFLP :

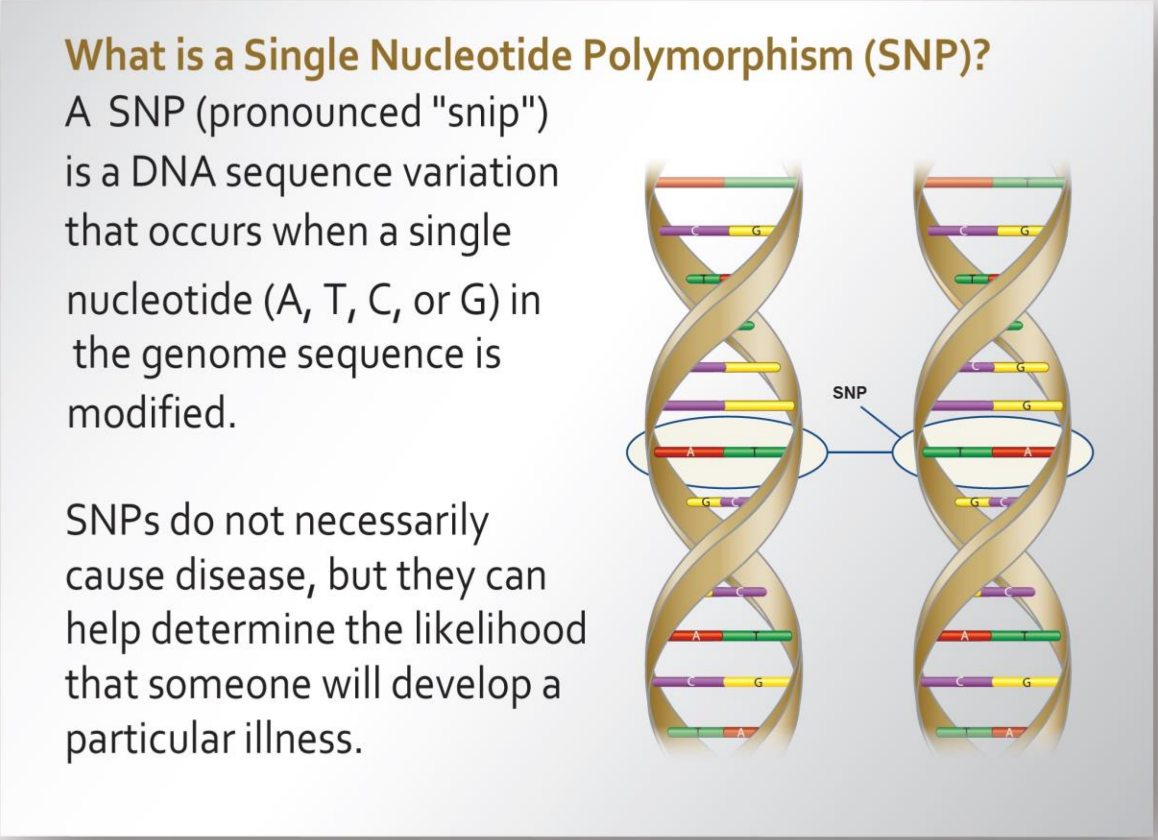
Detected by AFLP :



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

V. Decoding the genomes

SNP : Single nucleotide polymorphism



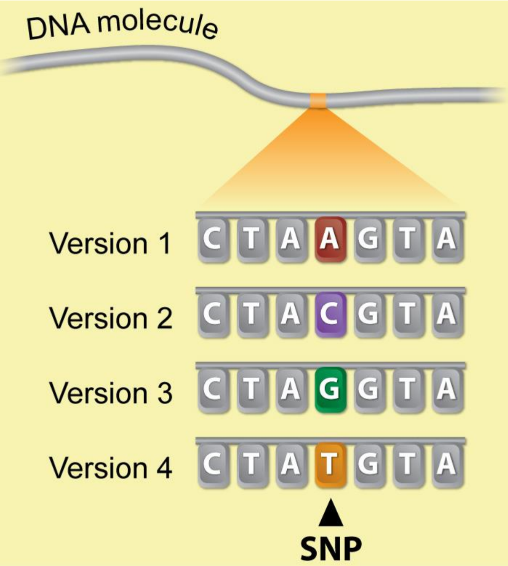
What is a Single Nucleotide Polymorphism (SNP)?

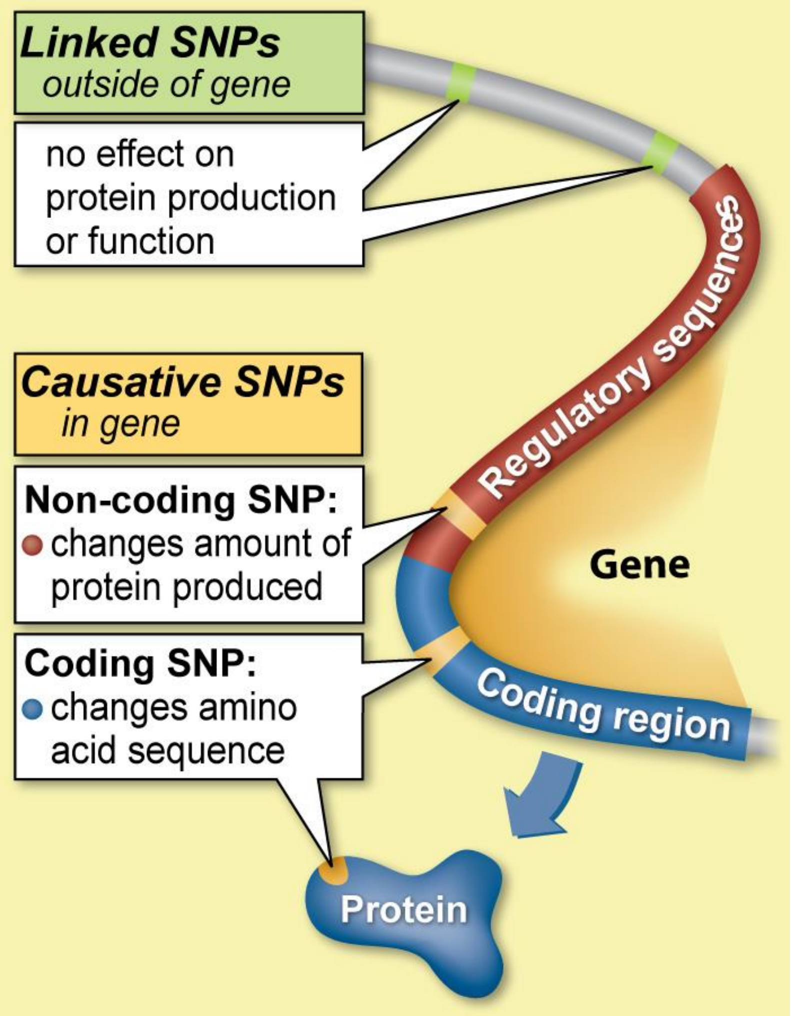
A SNP (pronounced "snip") is a DNA sequence variation that occurs when a single nucleotide (A, T, C, or G) in the genome sequence is modified.

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

V. Decoding the genomes

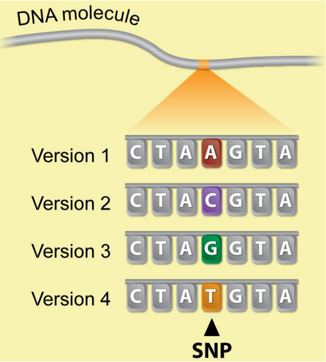
SNP : Single nucleotide polymorphism

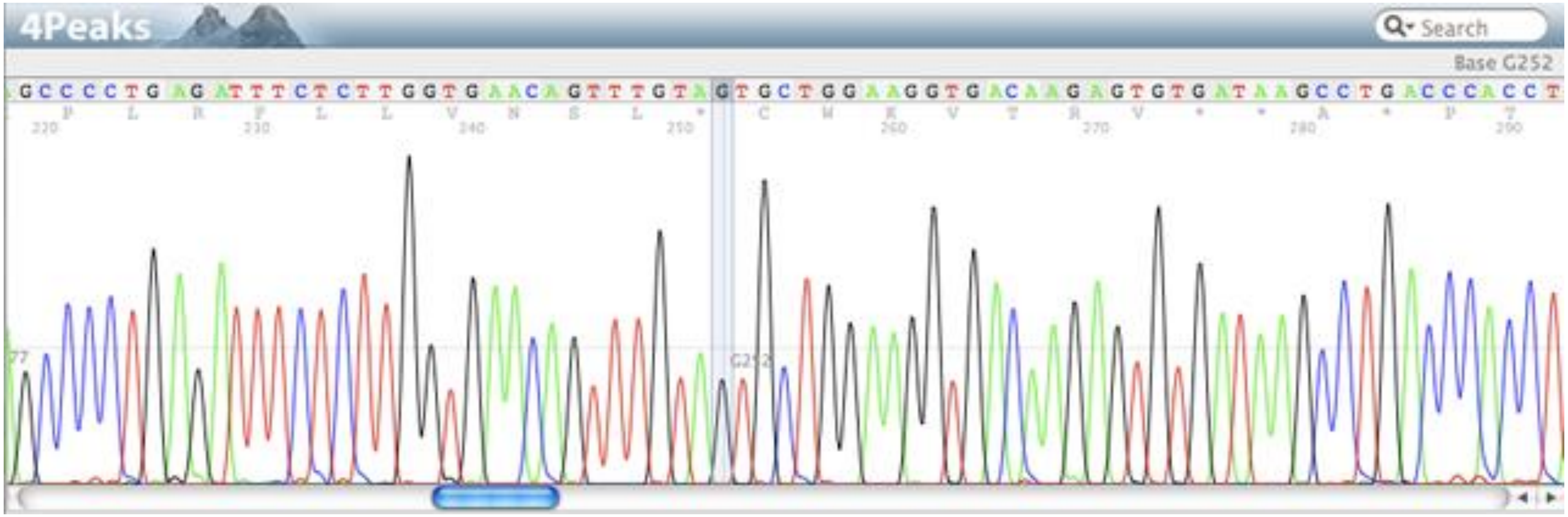




V. Decoding the genomes

SNP : Single nucleotide polymorphism





https://www.youtube.com/watch?v=DE9b1dxy_pE