

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



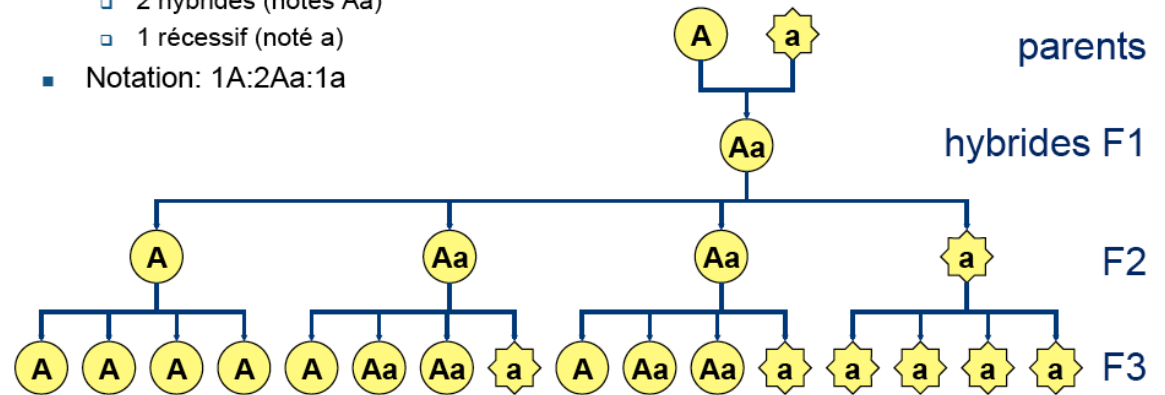
Gregor Mendel (1822 - 1884)

Heredity's (Mendel's) laws (1866)

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

Law 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)

Chromosomal theory of heredity

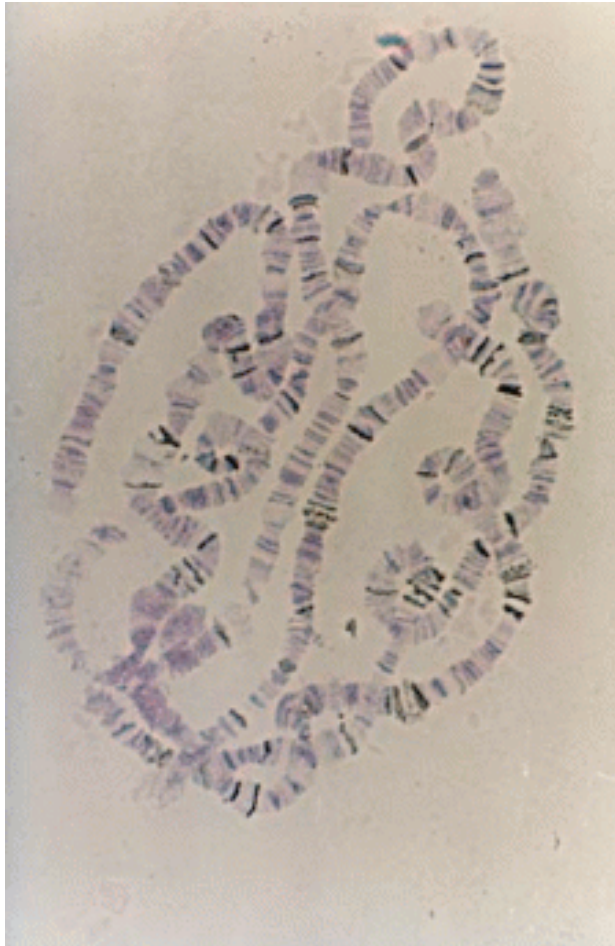


Drosophila melanogaster

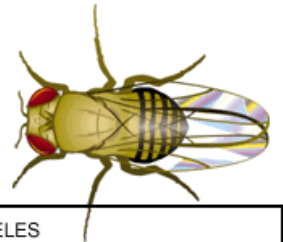
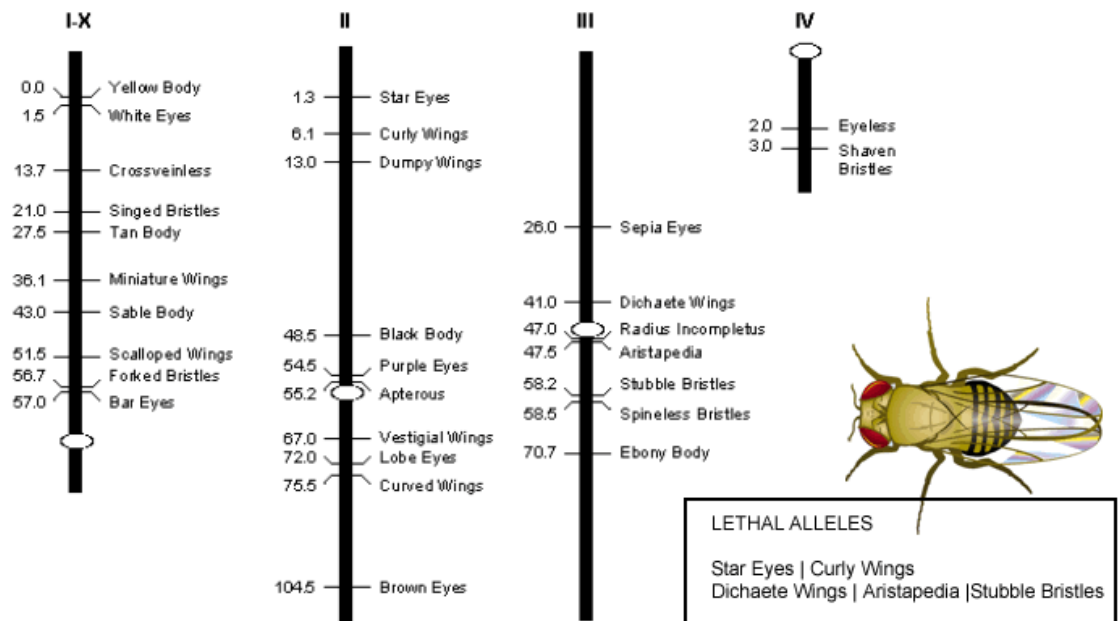


Chromosomes polytènes

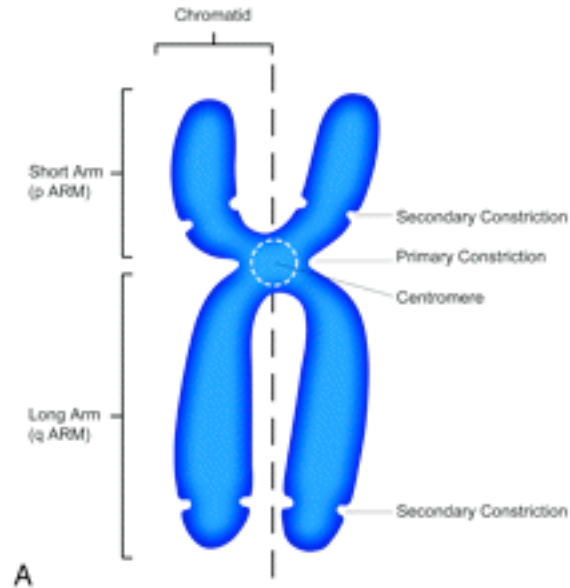
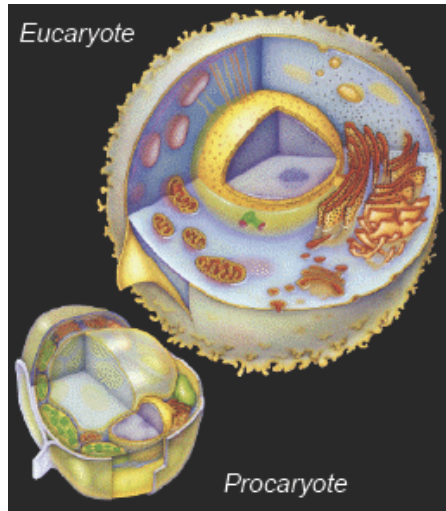
- Chromosome is the support of heredity (Thomas Morgan)



Drosophila Chromosome Map



- Chromosome is the support of heredity



Genes and Chromosomes : (concepts 7 to 14) : <http://www.dnafb.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 inheritance
Thomas Morgan (1866 - 1945)
Alfred Sturtevant (1891 - 1970)
- DNA is the support of inheritance
Fred Griffith (1877 - 1941)
Oswald Avery (1877 – 1955)
- Composition and structure of DNA
Alfred Hershey (1908-1997)
Martha Chase (1927-2003)



Erwin Chargaff (1905 - 1992)

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

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

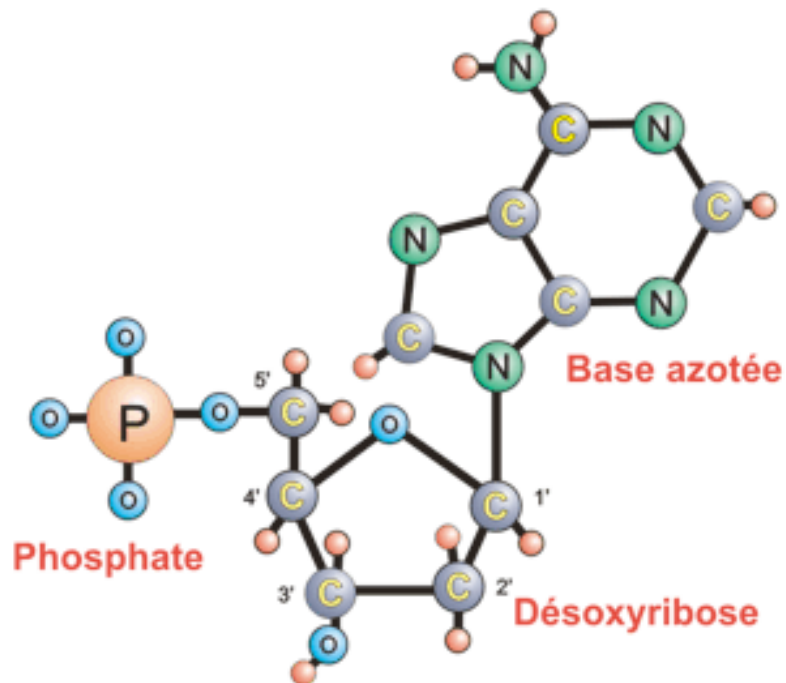
https://www.youtube.com/watch?v=a3_1AOXCMmo

II – Structure of nucleic acids

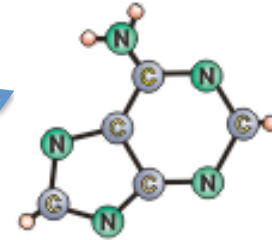
II.I Simple molecules

Desoxyribonucleic acid (DNA)

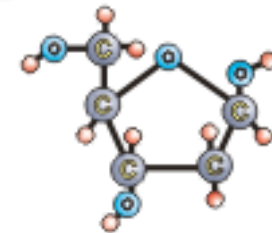
= polymers of small sub-units called **nucleotides**



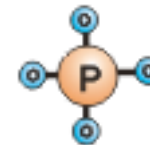
BASE



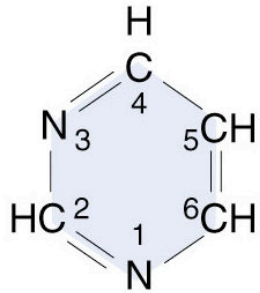
SUGAR



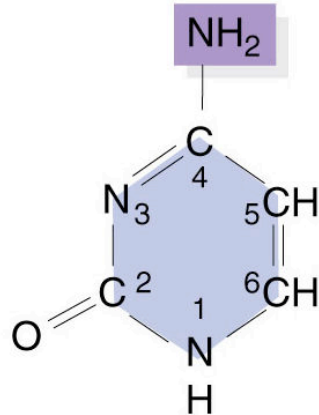
PHOSPHORIC ACID



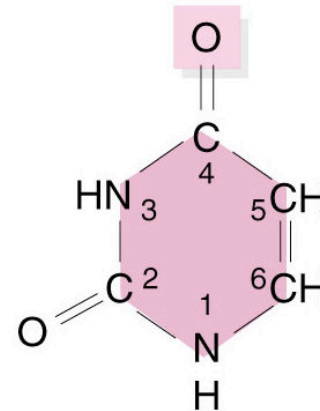
II.I.1 Nitrogenous bases



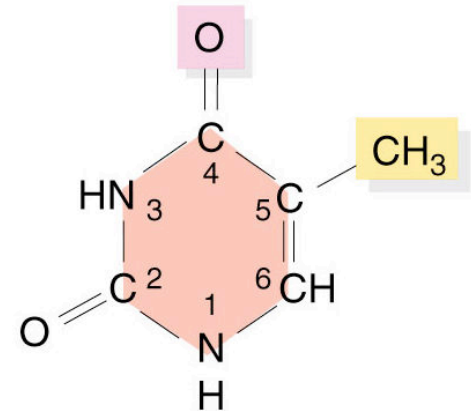
Pyrimidine



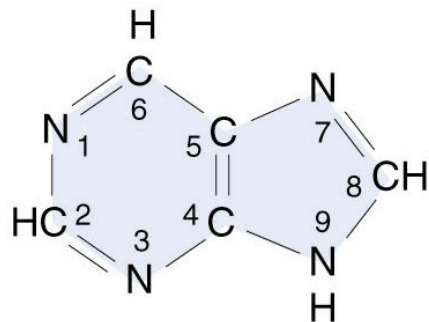
Cytosine (C)
2-oxy-4-amino-pyrimidine



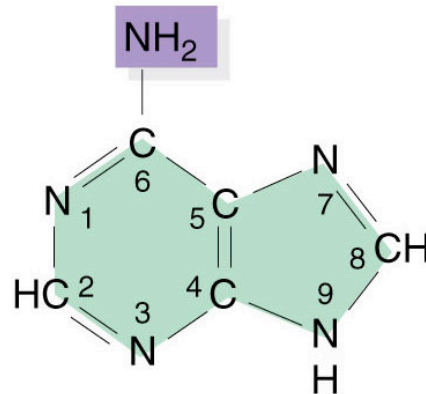
Uracil (U)
(found in RNA)
2,4-dioxy-pyrimidine



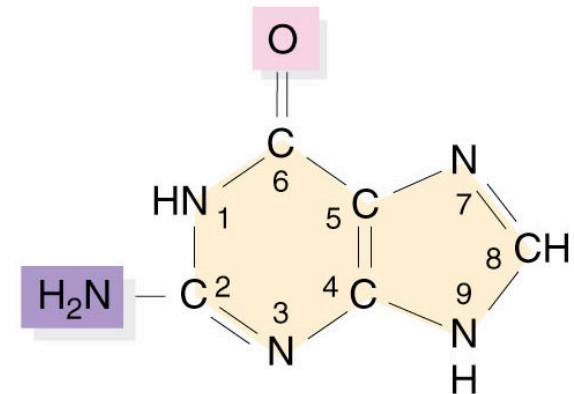
Thymine (T)
(found in DNA)
2,4-dioxy-5-méthyl-pyrimidine



Purine



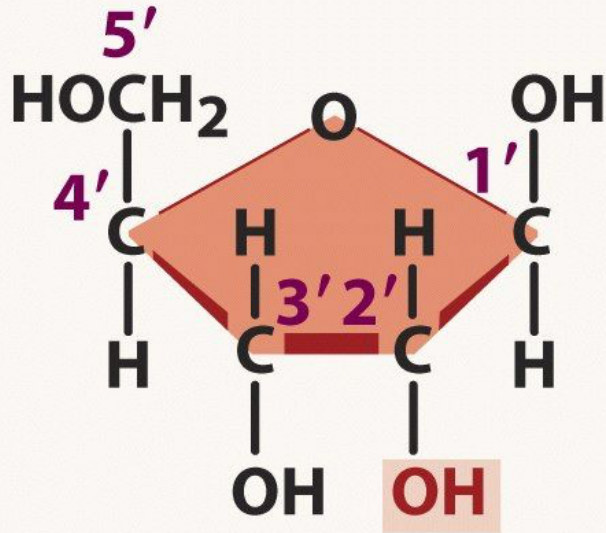
Adenine (A)
6-amino-purine



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

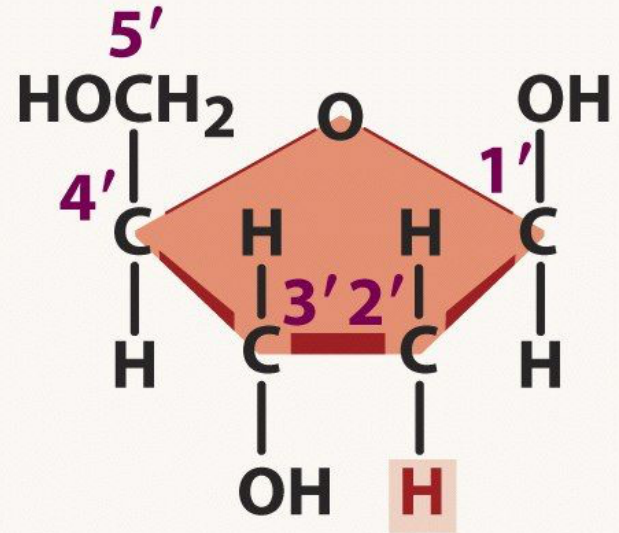
II.1.2 Sugars

Sugars



Ribose

RNA



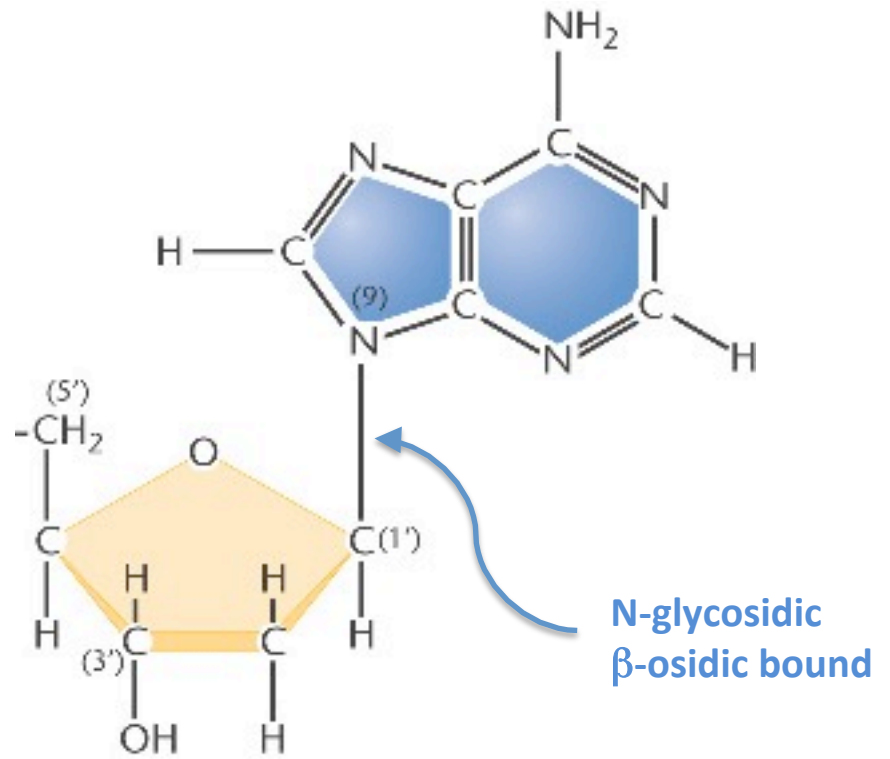
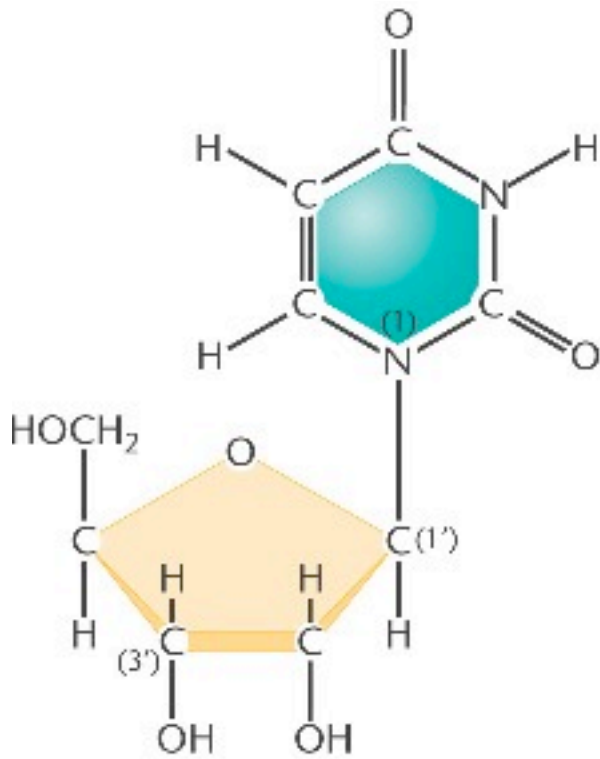
Deoxyribose

DNA

II.2 Nucleosides

NucleoSides

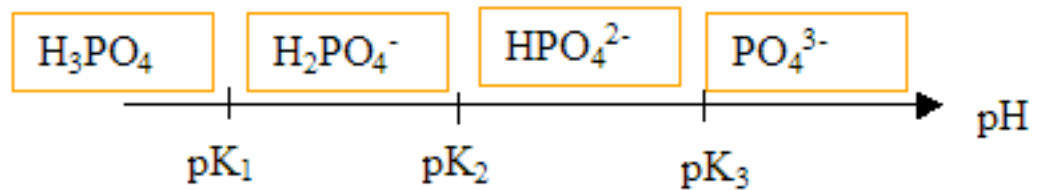
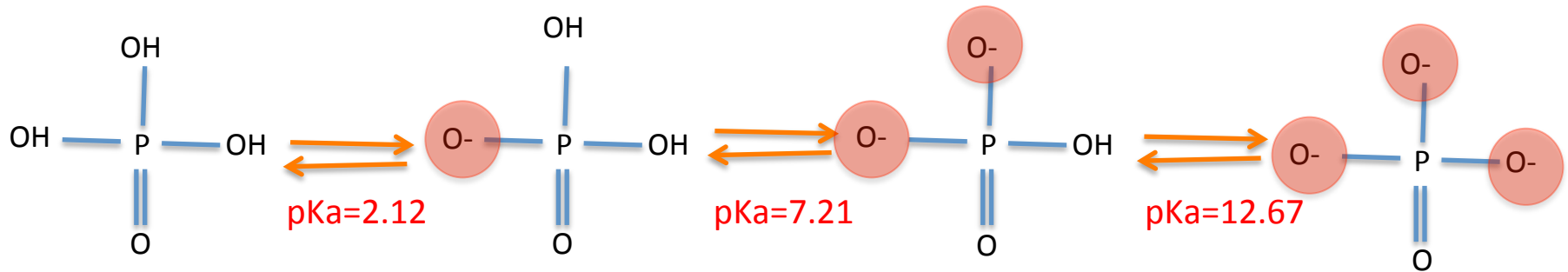
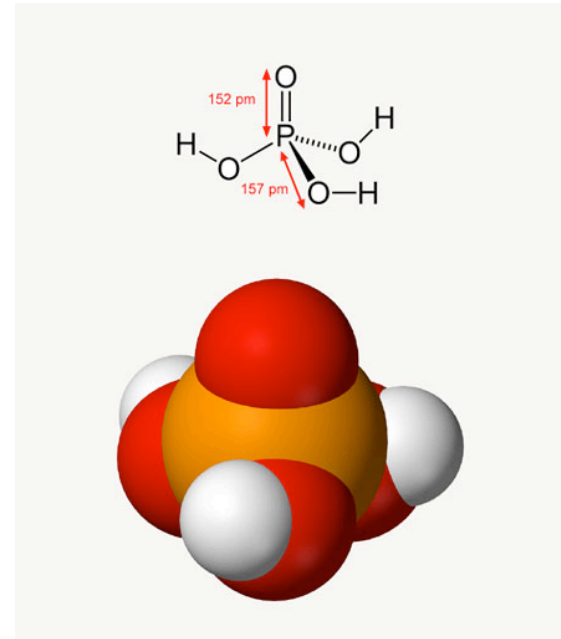
SUGAR + BASE



II.2 Nucleosides

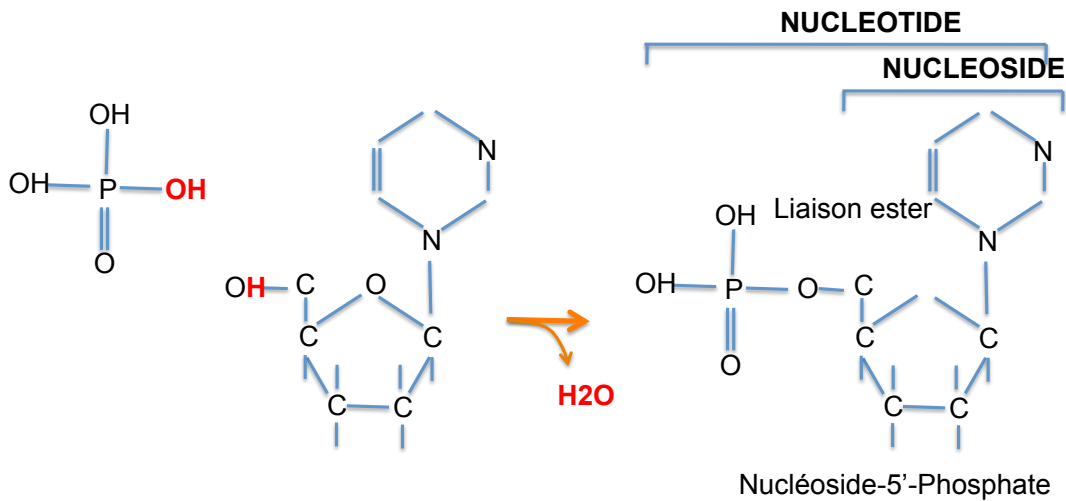
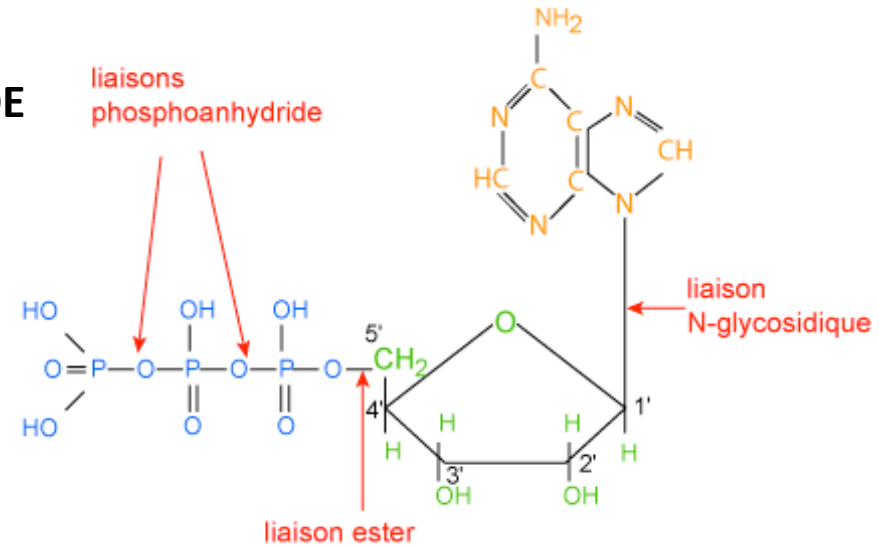
	Base	Ribonucleoside	Desoxyribonucleoside
Purines	Adénine Guanine	Adénosine Guanosine	Désoxyadénosine Désoxyguanosine
Pyrimidines	Uracile Cytosine Thymine	Uridine Cytidine Thymine ribonucléoside	Désoxyuridine Désoxycytidine Désoxythymidine ou thymidine

II.3 Phosphoric Acid



II.4 Nucleotides

SUGAR + BASE + PHOSPHATE = NUCLEOTIDE



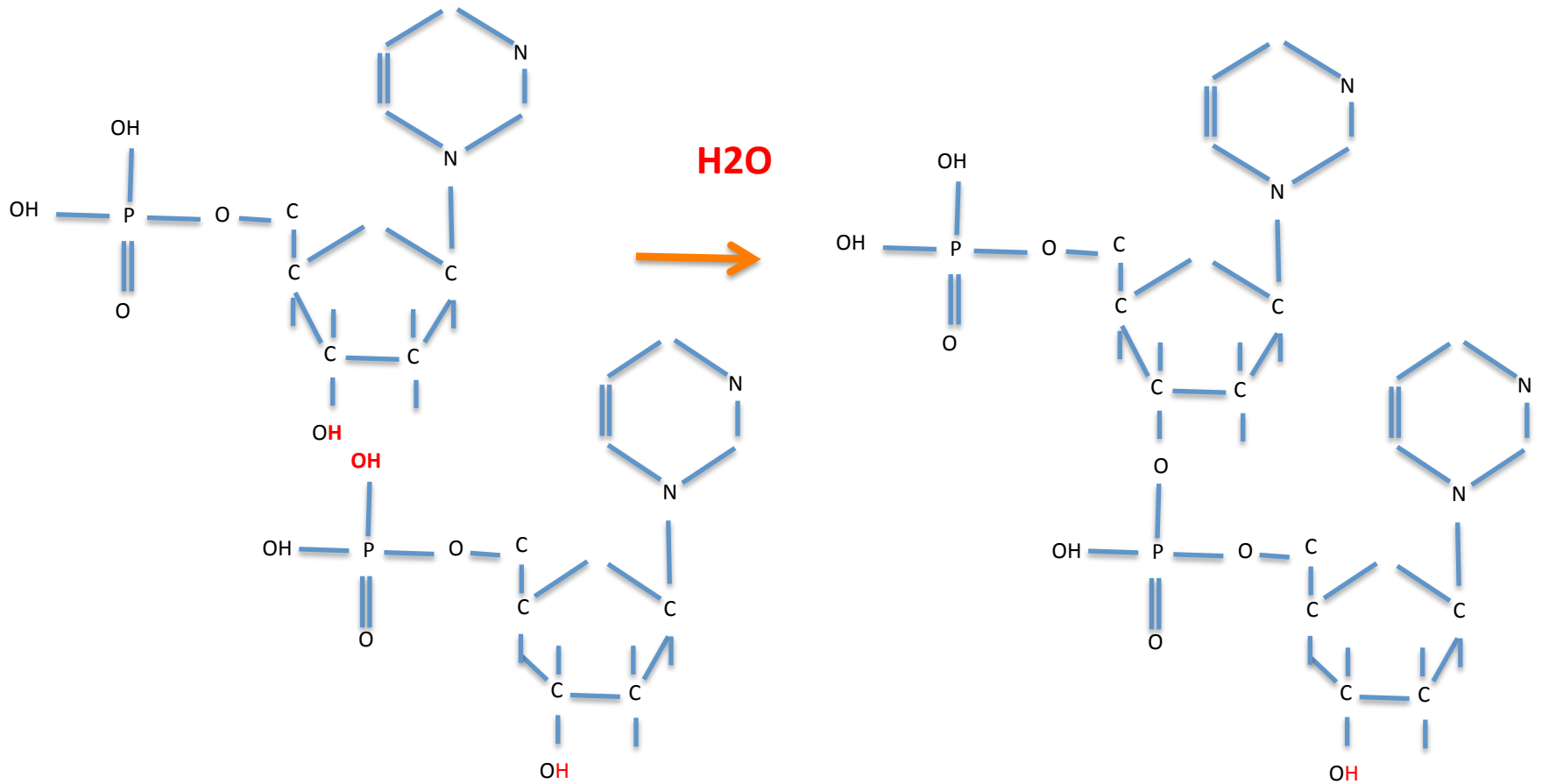
II.4 Nucleotides

NucleoTides : nucleosides-5'-Monophosphate

Base	RibonucleoTide	DesoxyribonucleoTide
Adénine	Adé no sine-5'-Monophosphate (AMP)	Désoxyadénosine-5'-Monophosphate (dAMP)
Guanine	Gua no sine-5'-Monophosphate (GMP)	Désoxyguanosine-5'-Monophosphate (dGMP)
Uracile	U ri dine-5'-Monophosphate (UMP)	Désoxyuridine-5'-Monophosphate (dUMP)
Cytosine	Cy ti dine-5'-Monophosphate (CMP)	Désoxycytidine-5'-Monophosphate (dCMP)
Thymine	Thymine riboside -5'-Monophosphate (TMP)	Désoxythymidine -5'-Monophosphate (dTMP)

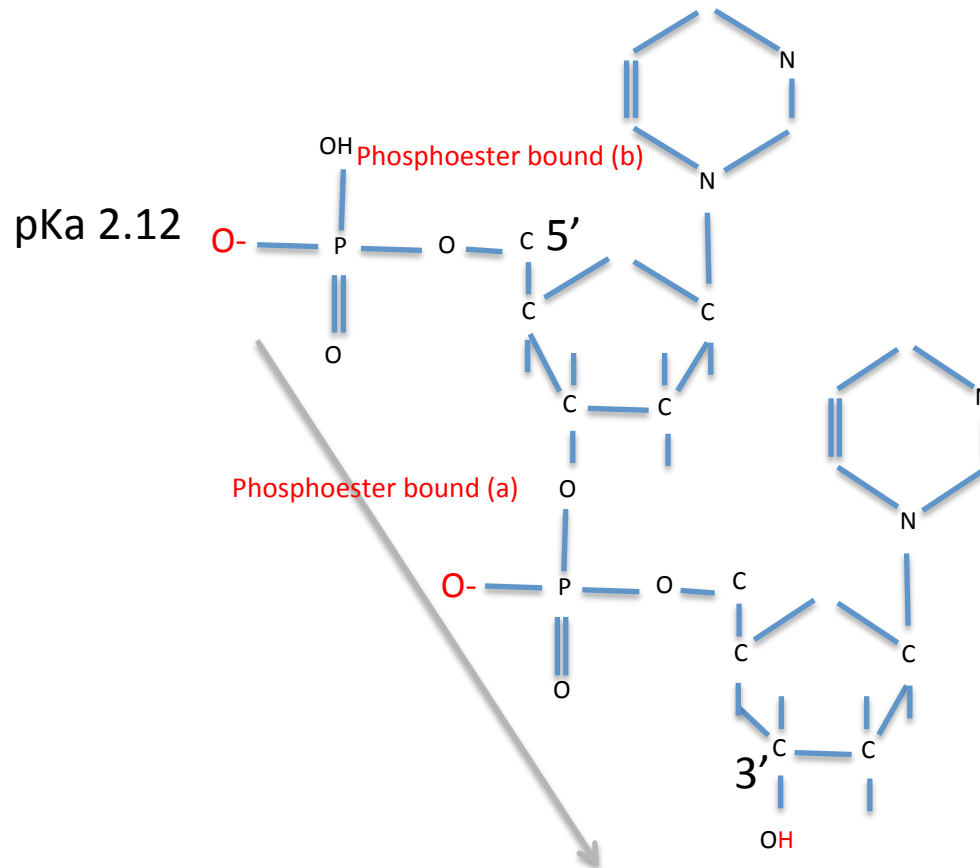
II.5 Phosphodiester bound

Polynucleotidic chain



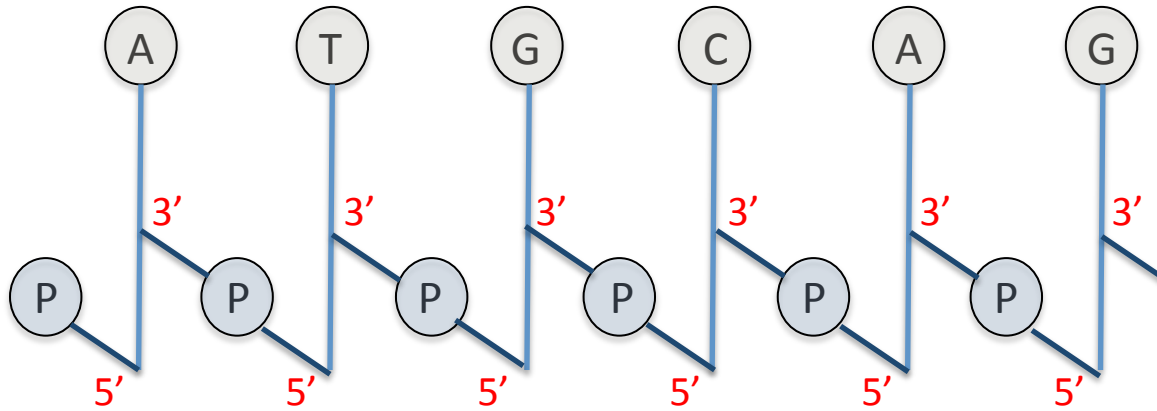
II.5 Phosphodiester bound

Polynucleotidic chain



II.5 Phosphodiester bound

Polynucleotidic chain

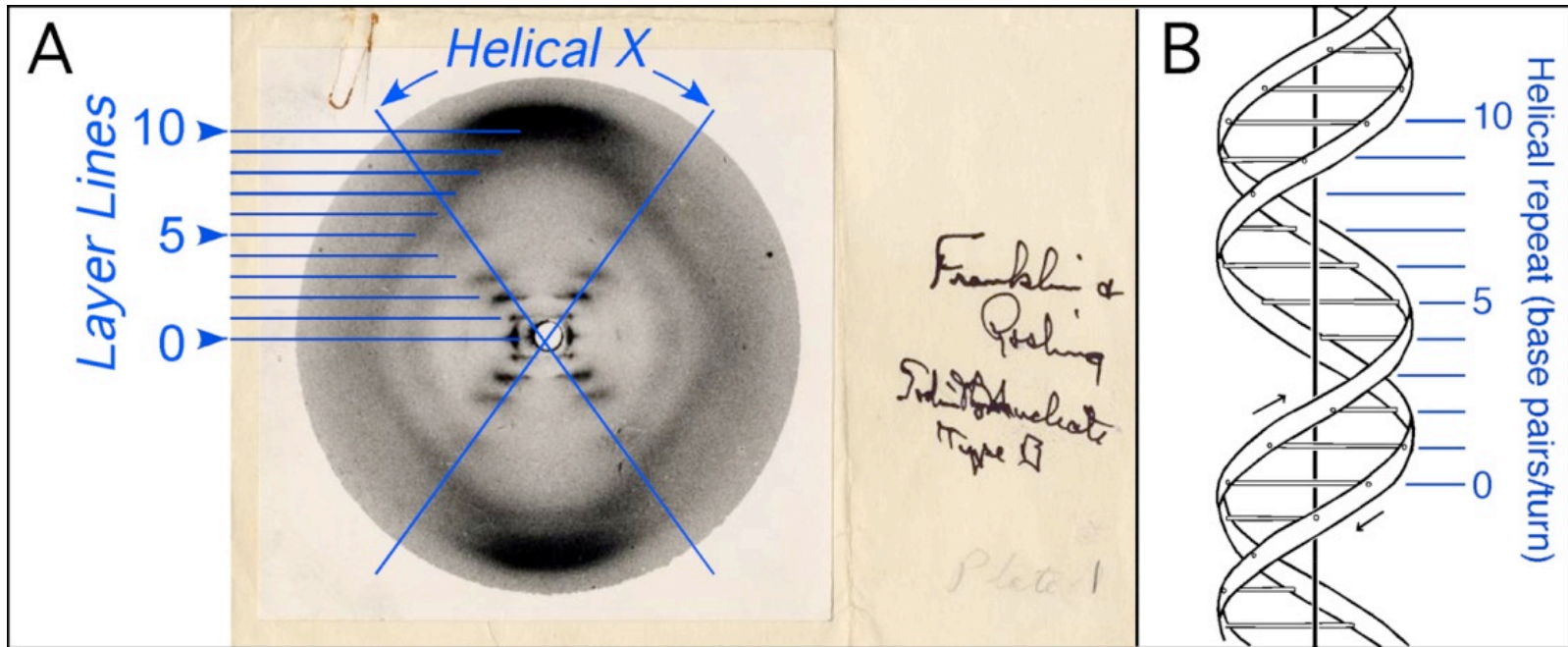


p5'dAp5'dTp5'dGp5'dCp5'dAp5'dG

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

II.6 Secondary Structure of DNA

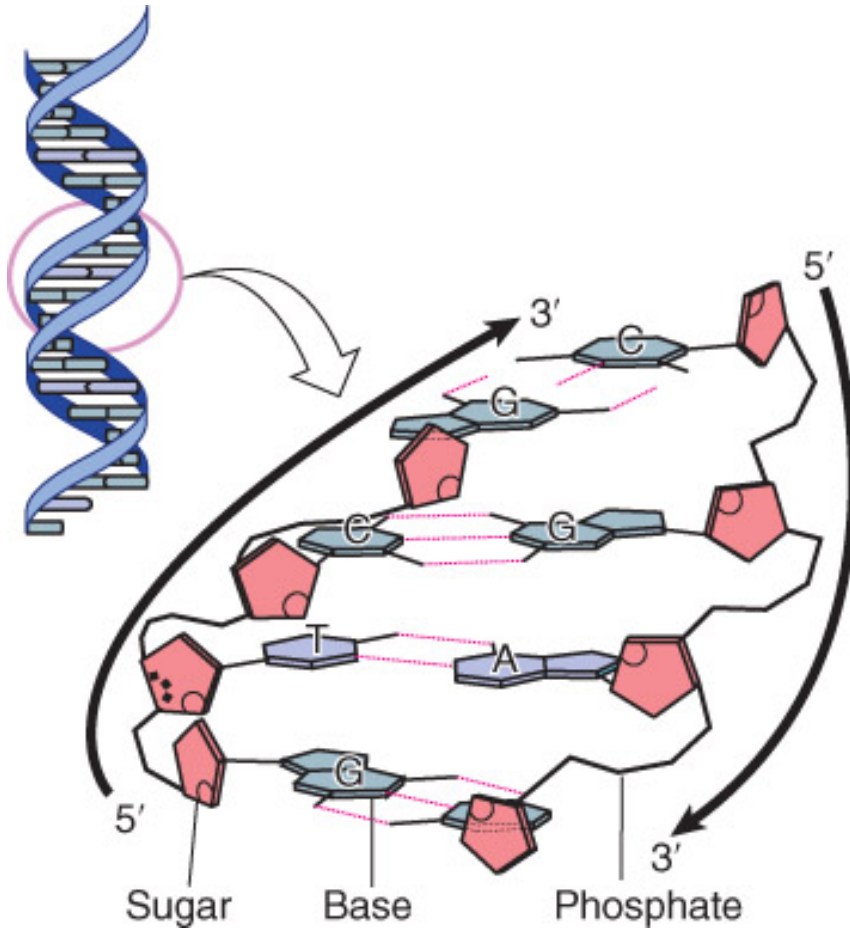
Secondary Structure : The Watson et Crick model



Rosalind Franklin

II.6 Secondary Structure of DNA

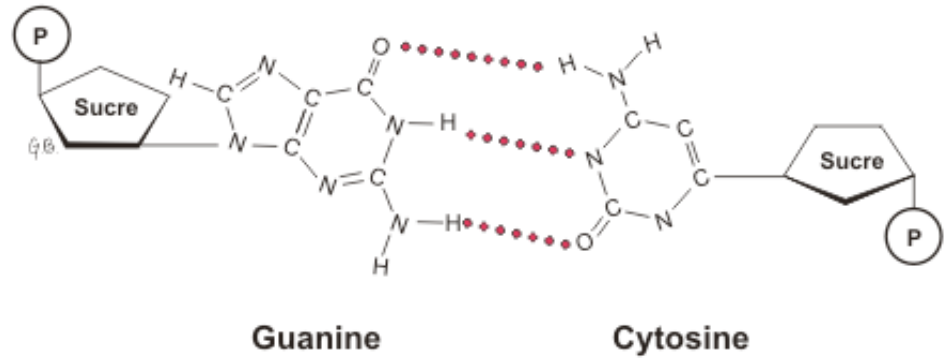
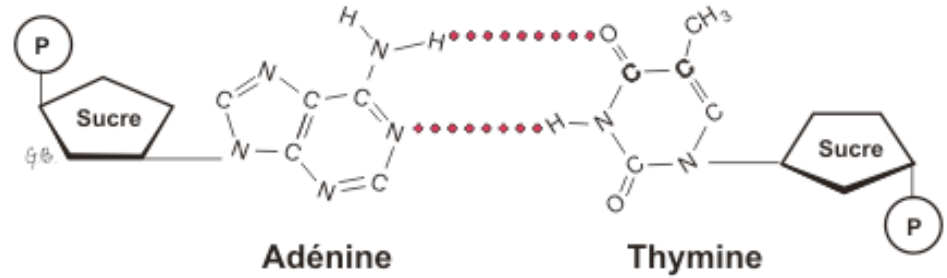
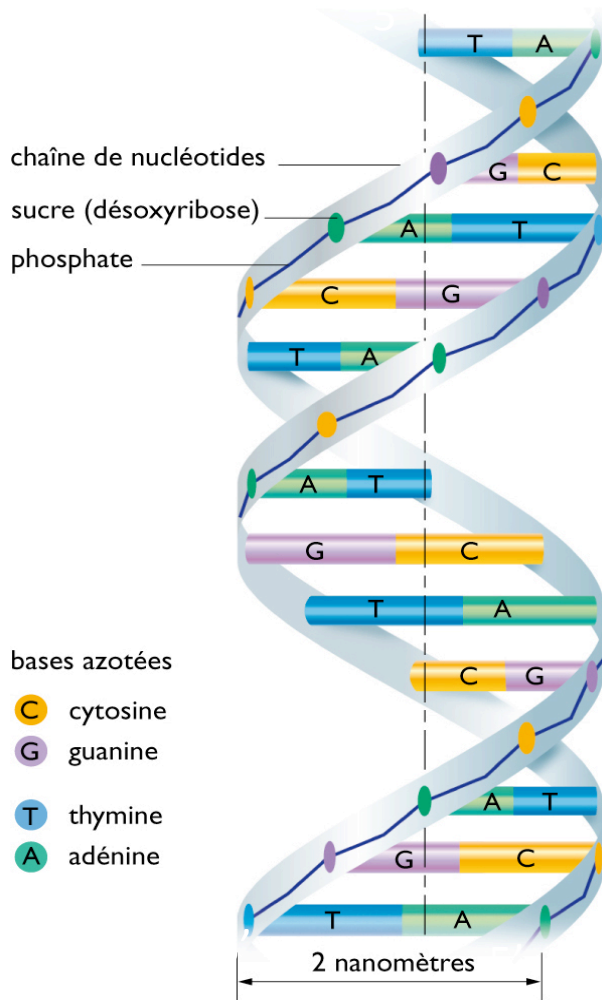
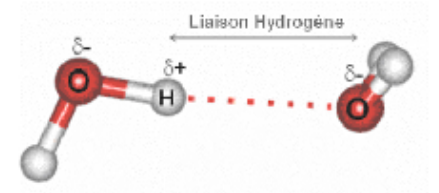
The Molecule of DNA : secondary structure



- **Double stranded** (2 strands of polynucleotide chains)
- **Antiparallel**
- **Helicoidal**
- Bases (inside) Sugar/phosphates (outside)
- **Complementary**

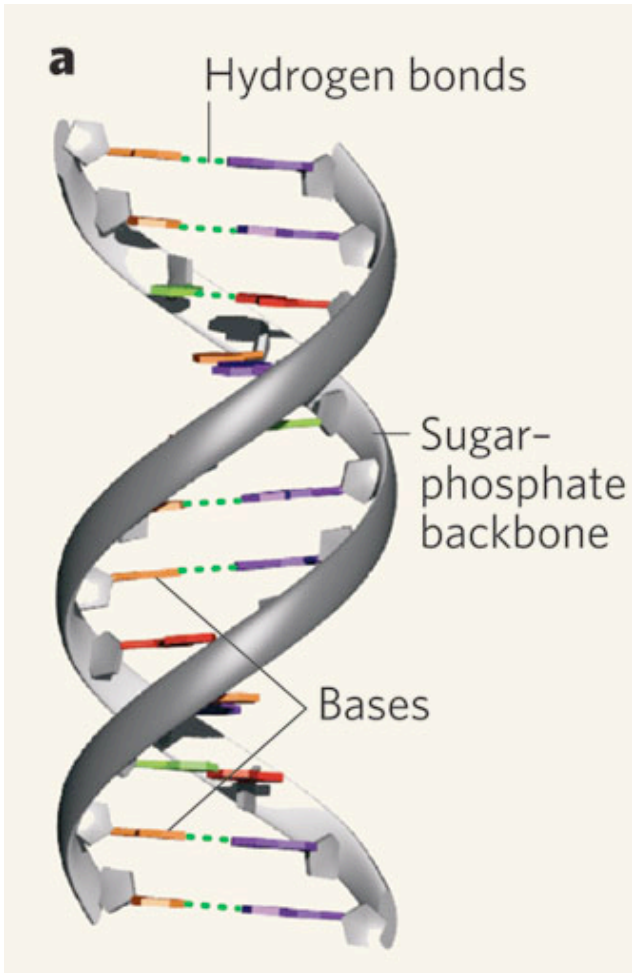
II.6 Secondary Structure of DNA

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 that 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
<i>E. coli</i>	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

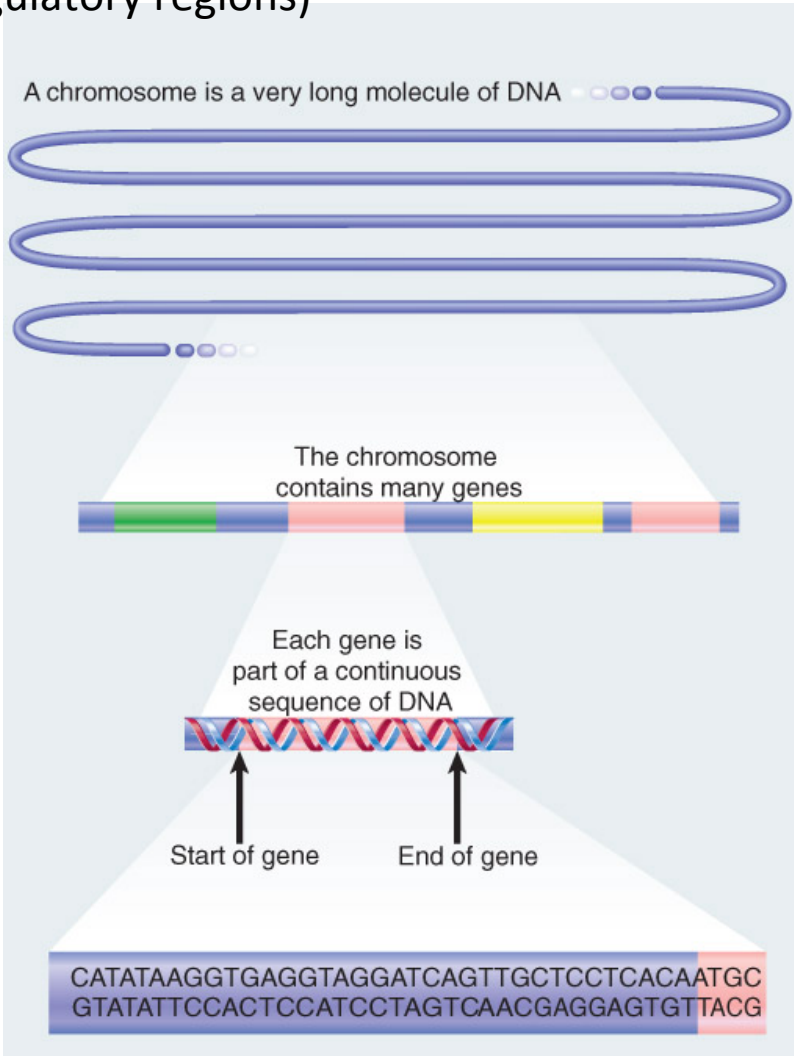
The genome of each organism is organized into a specific number of chromosomes, whose shape is characteristic

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				
<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
Eukaryotes				
<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 of genomes

What's a gene? : a gene is a region of DNA encoding for a polypeptide (+ proximal upstream regulatory regions)



Allele : one of the versions of a gene

Locus : position of a gene on the chromosome

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

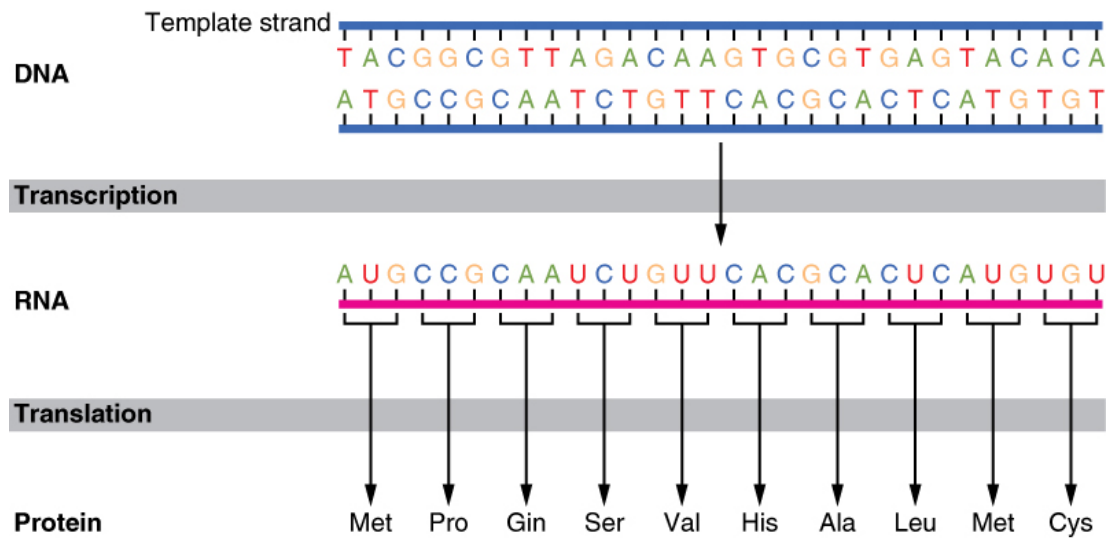
Each allele has a different phenotype	
Allele	Phenotype of homozygote
w^+	red eye (wild type)
w^{bl}	blood
w^{ch}	cherry
w^{bf}	buff
w^h	honey
w^a	apricot
w^e	eosin
w^l	ivory
w^z	zeste (lemon-yellow)
w^{sp}	mottled, color varies
w^1	white (no color)

The gene

The genetic code

Genetic Code : the relation between the sequence of nucleotides in the DNA and the sequence of amino acids in the polypeptides

Genetic Information : genes + regulatory sequences



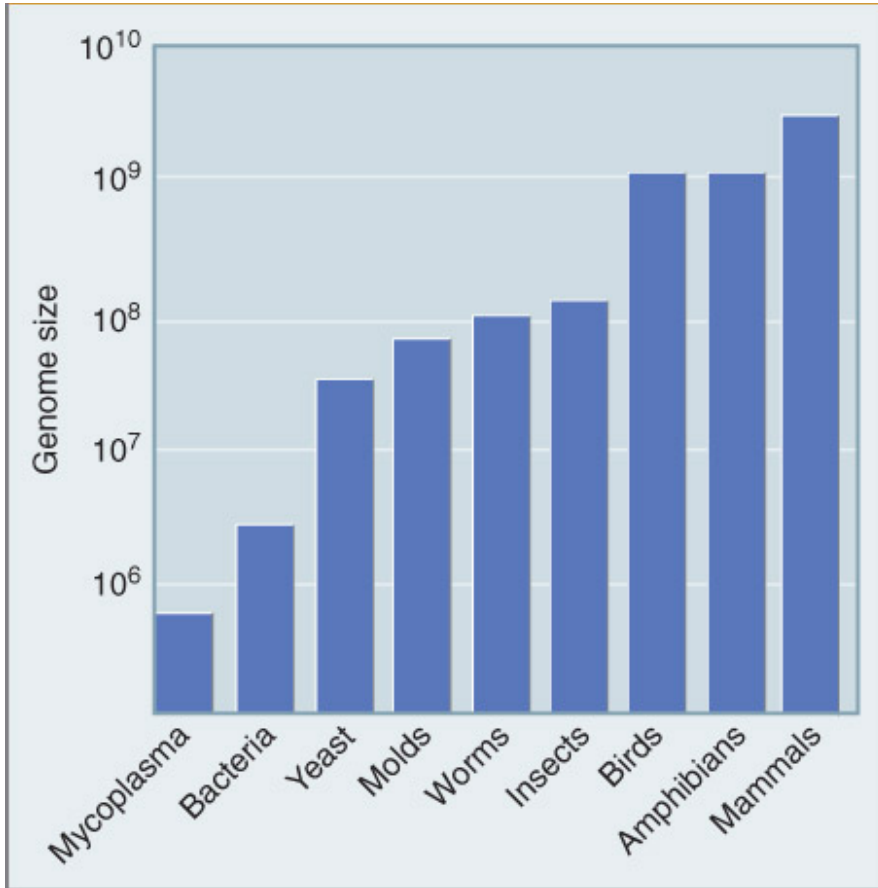
The genetic code

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	Troisi me lettre : wobble
	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 } AUC } Ile 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 } GUC } Val GUA } GUG } *	GCU } GCC } Ala GCA } GCG } *	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG } *	U C A G	

III. Structure of genomes

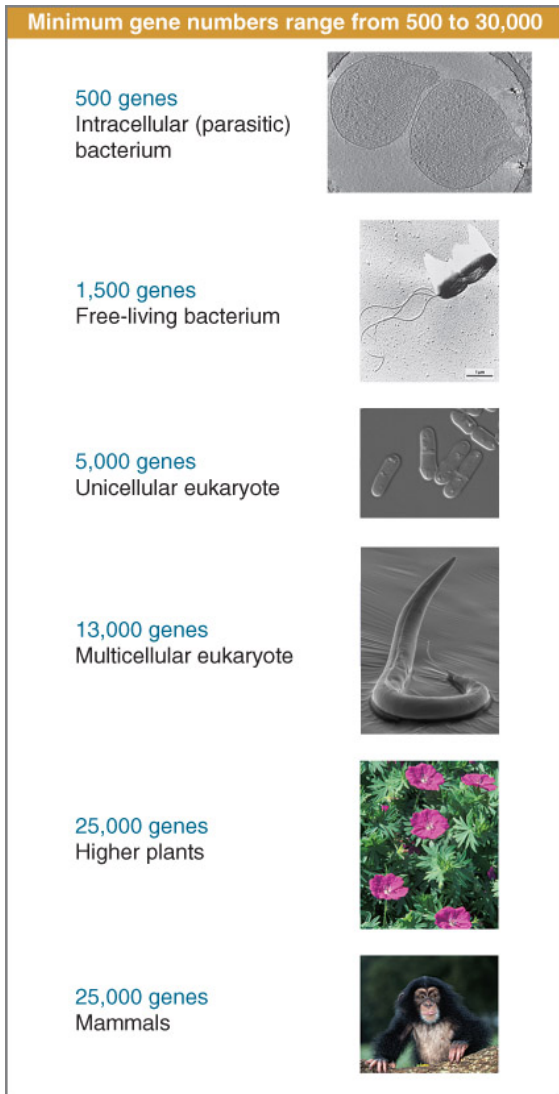
The size of the genome is roughly correlated to the complexity of the organism



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

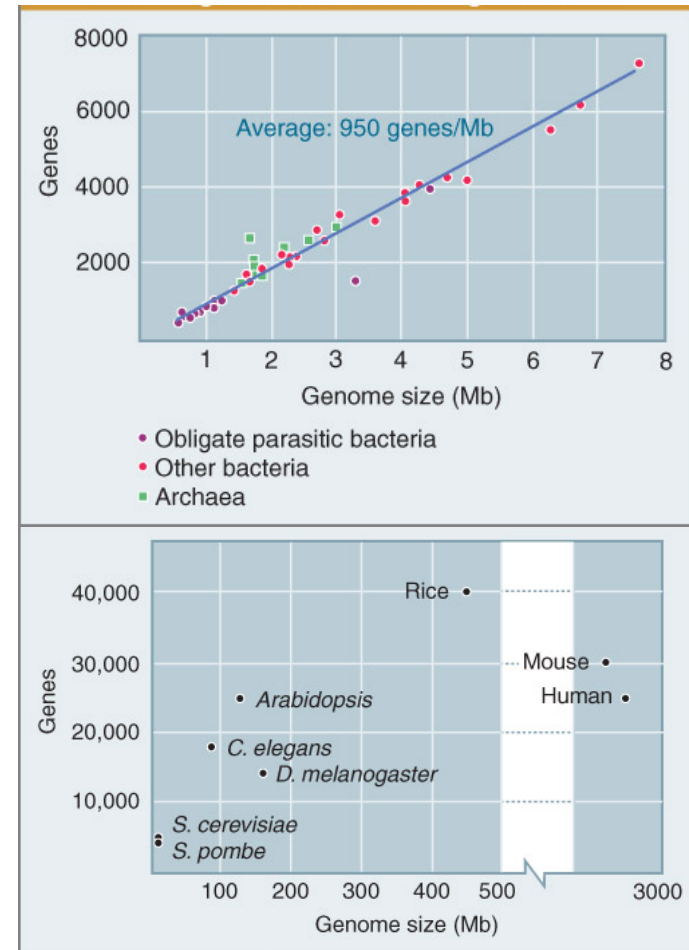
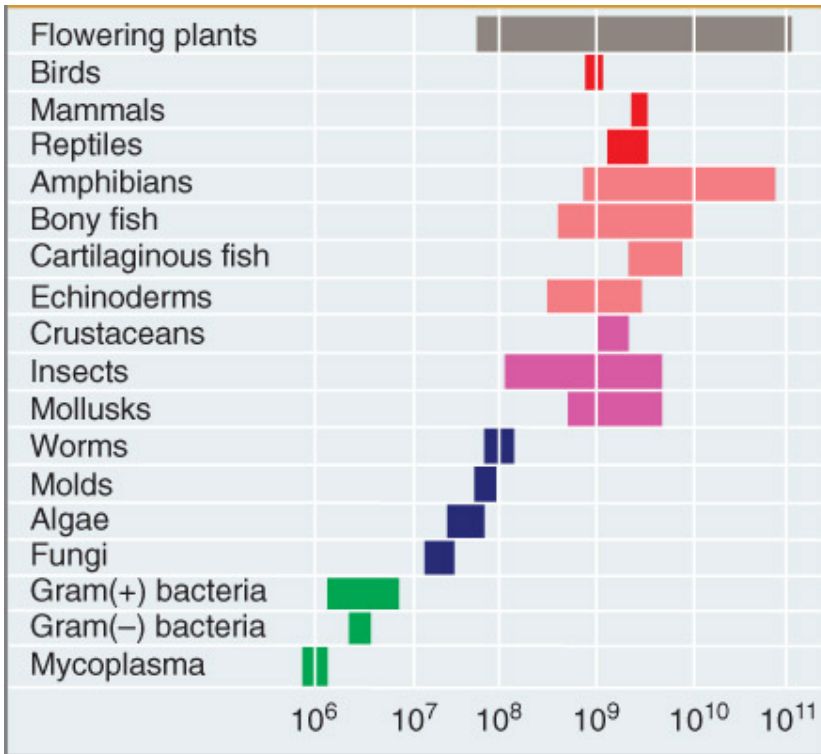
III. Structure of genomes

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_1%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](https://bio.libretexts.org/TextMaps/Genetics/Book%3A_Working_with_Molecular_Genetics_(Hardison)/Unit_1%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

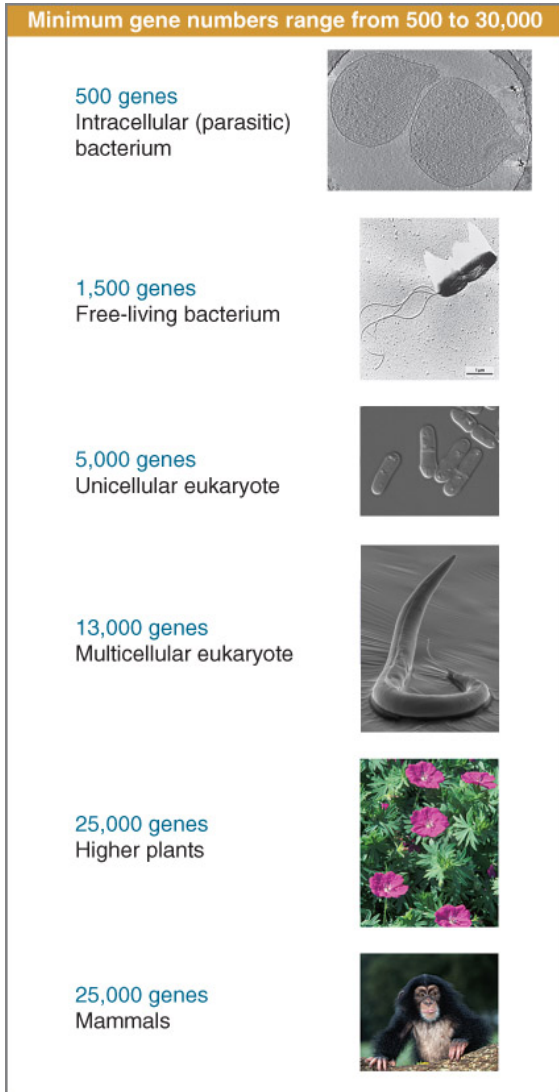


TABLE 8-2 Comparison of the Gene Density in Different Organisms' Genomes

Species	Genome Size (Mb)	Approximate Number of Genes	Gene Density (genes/Mb)
Prokaryotes (bacteria)			
<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
Eukaryotes (animals)			
Fungi			
<i>Saccharomyces cerevisiae</i>	12	5800	480
<i>Schizosaccharomyces pombe</i>	12	4900	410
Protozoa			
<i>Tetrahymena thermophila</i>	125	27,000	220
Invertebrates			
<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
Vertebrates			
<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
Plants			
<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 of genomes

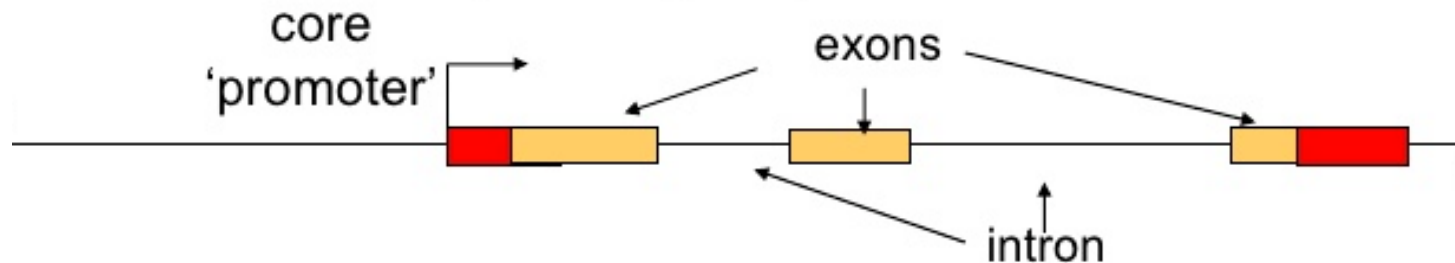
Structure of genes

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



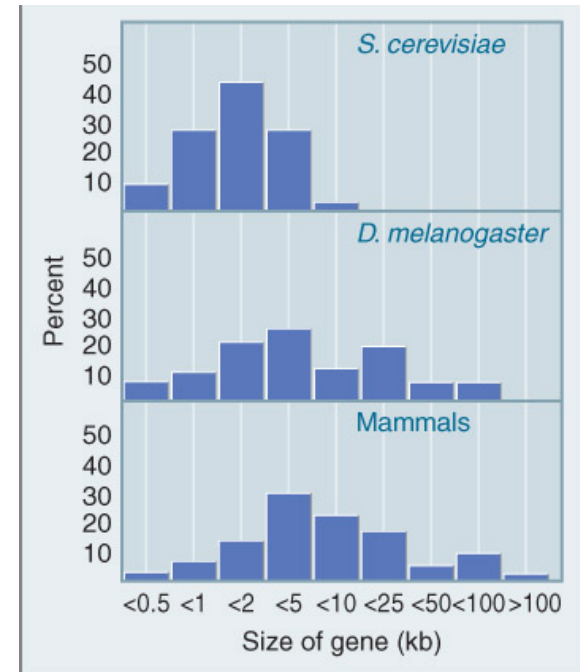
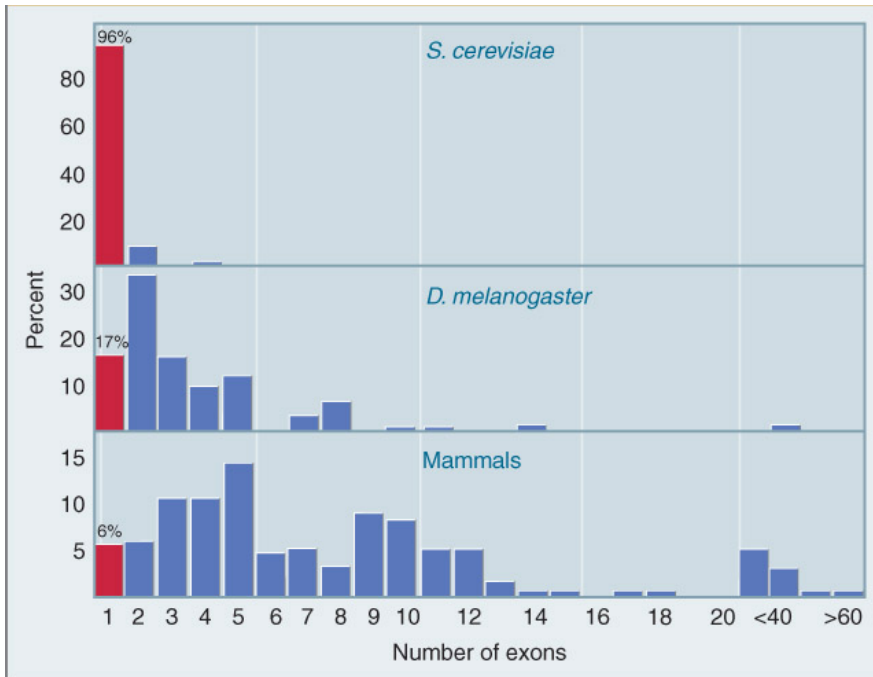
→ ADN et ARNm sont colinéaires

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



III. Structure of genomes

Structure of genes

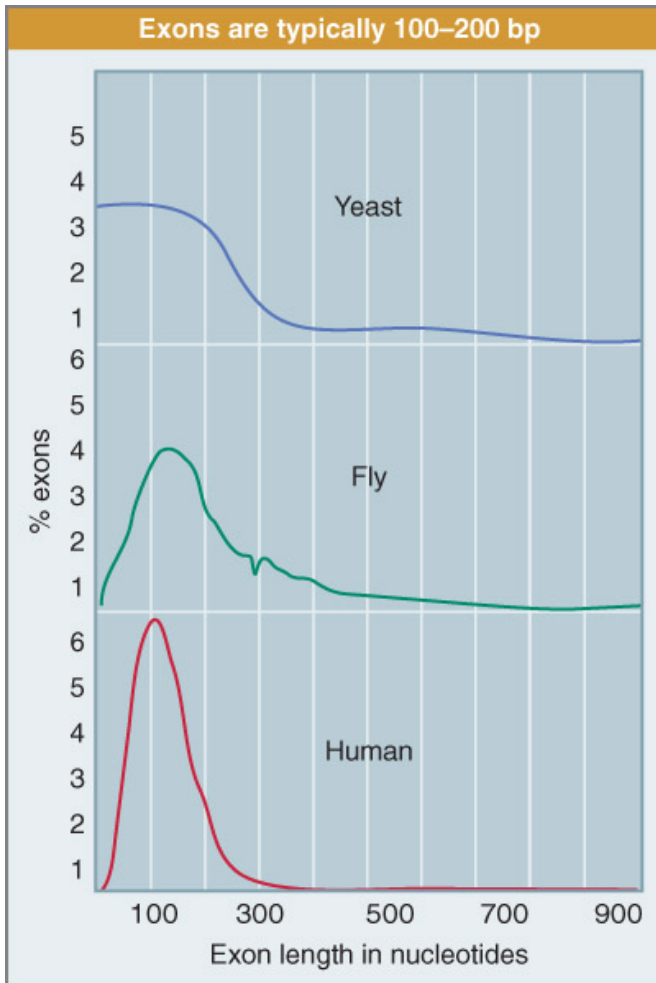


During evolution, genes went from mostly short and not interrupted to mostly longer and interrupted

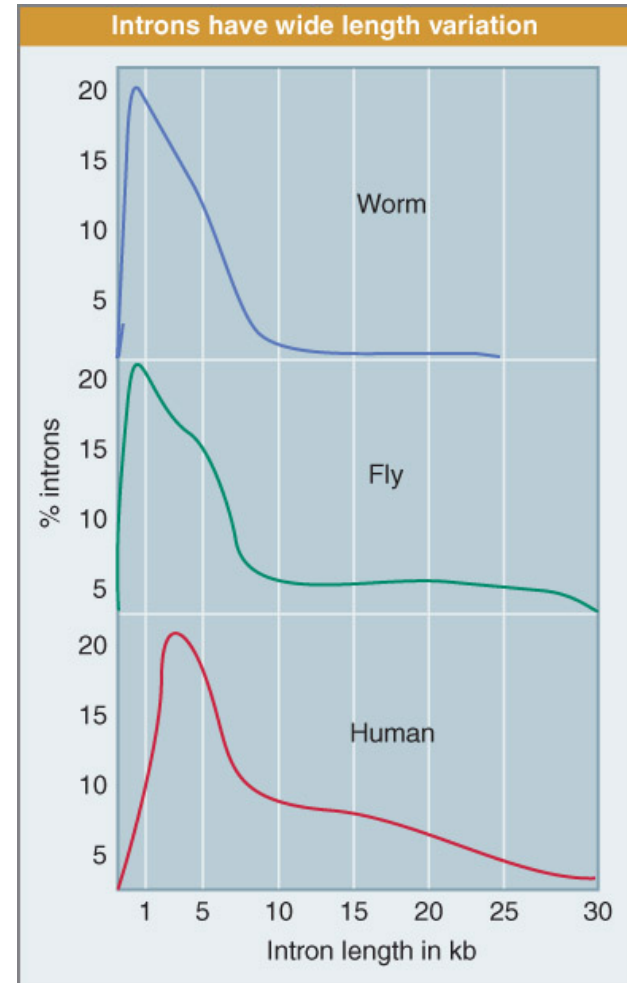
III. Structure of genomes

Structure of genes

exons are short and with constant sizes throughout evolution

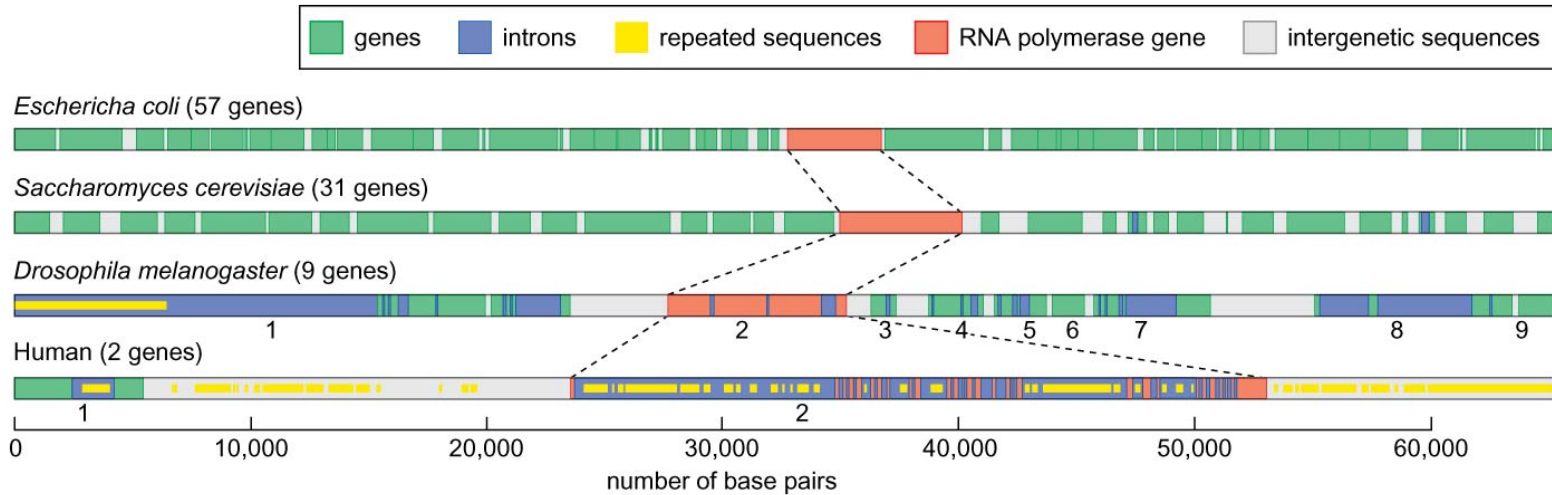


introns are more variable in size and become longer throughout evolution



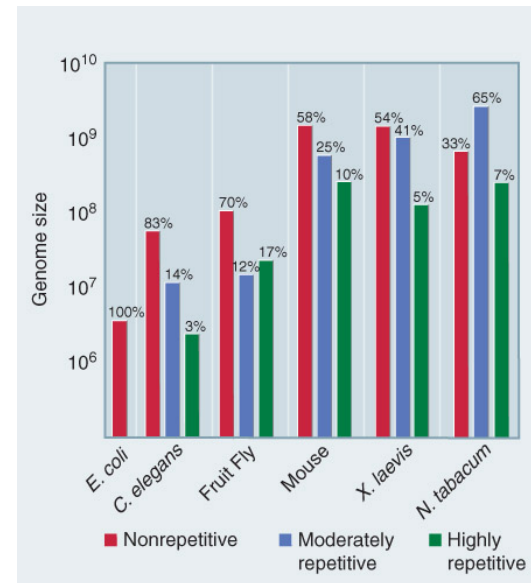
III. Structure du génome

More complex organisms have lower gene densities



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

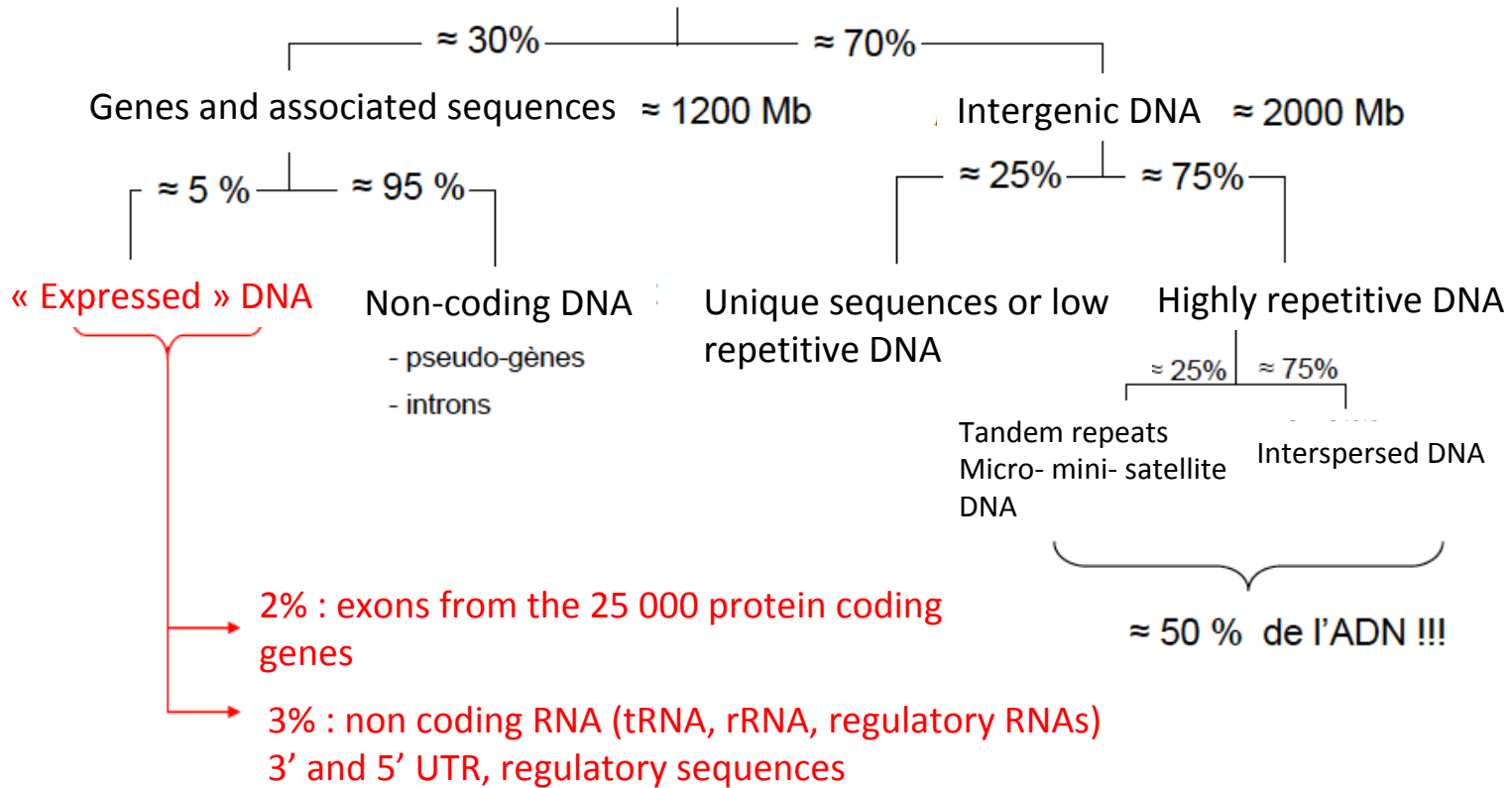
Non repeated DNA : coding regions (genes)



III. Structure du génome

Classical composition of a superior eukaryote genome

The majority of DNA is non coding!



IV. Mutations

Mutations are modifications of the DNA sequence

Spontaneous mutations (mistakes during replication)

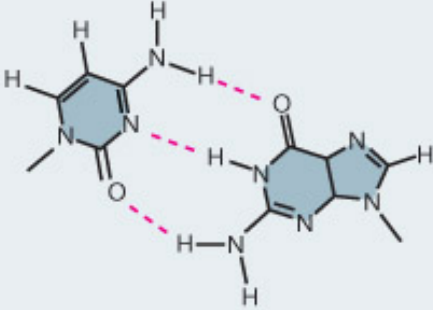


Inducible mutations (chemical modifications)

Insertion of exogenous DNA (viruses, transposons)

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

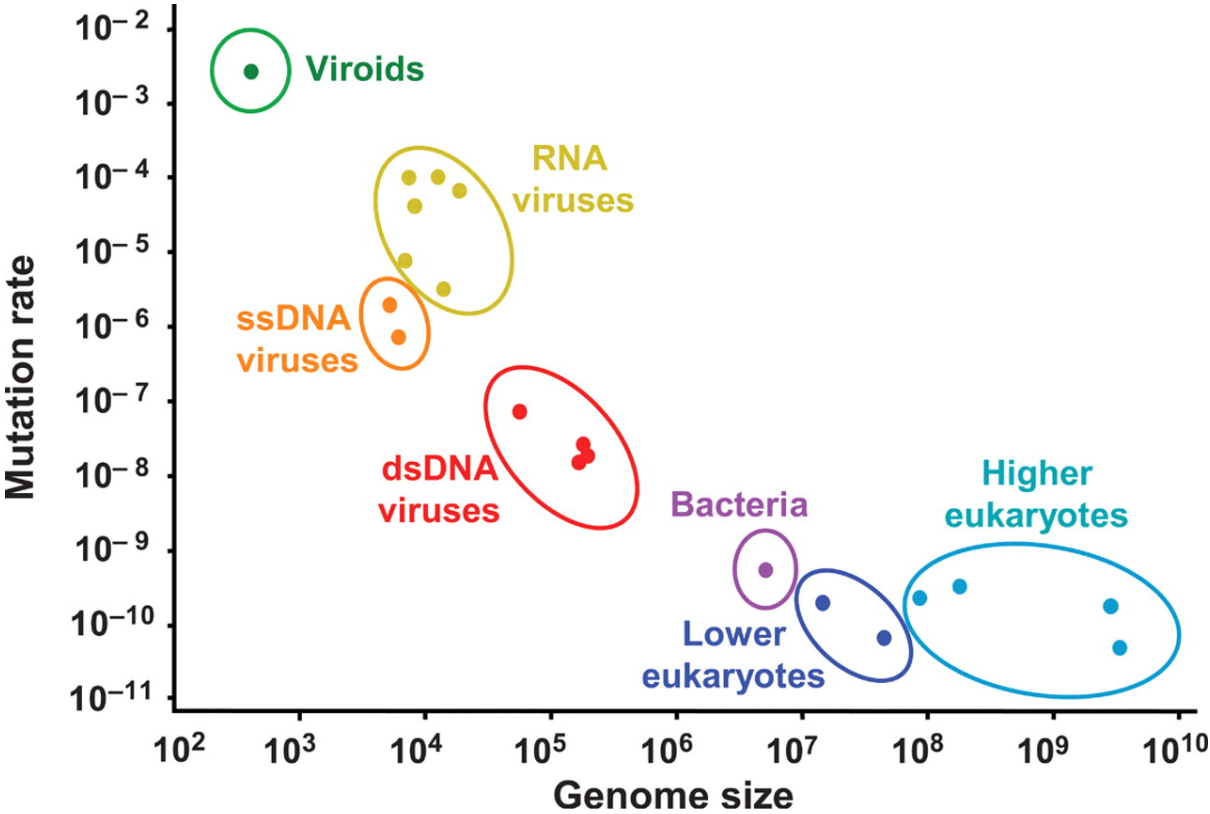
IV. Mutations

Spontaneous Mutations

	<p>Mutation rate</p> <p>Any base pair 1 in 10^9–10^{10} generations</p>
<p>...ATCGGACTTACCGGTTA... ...TAGCCTGAATGGCCAAT...</p> 	<p>Any gene 1 in 10^5–10^6 generations</p>
	<p>The genome 1 in 300 generations</p>

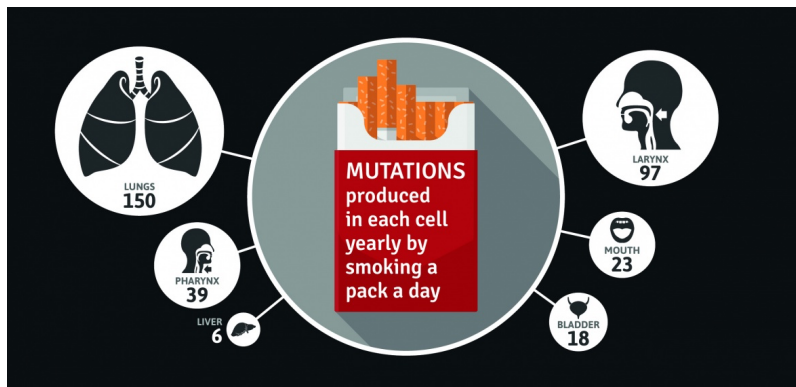
IV. Mutations

Spontaneous Mutations



IV. Mutations

Inducible Mutations and mutagens



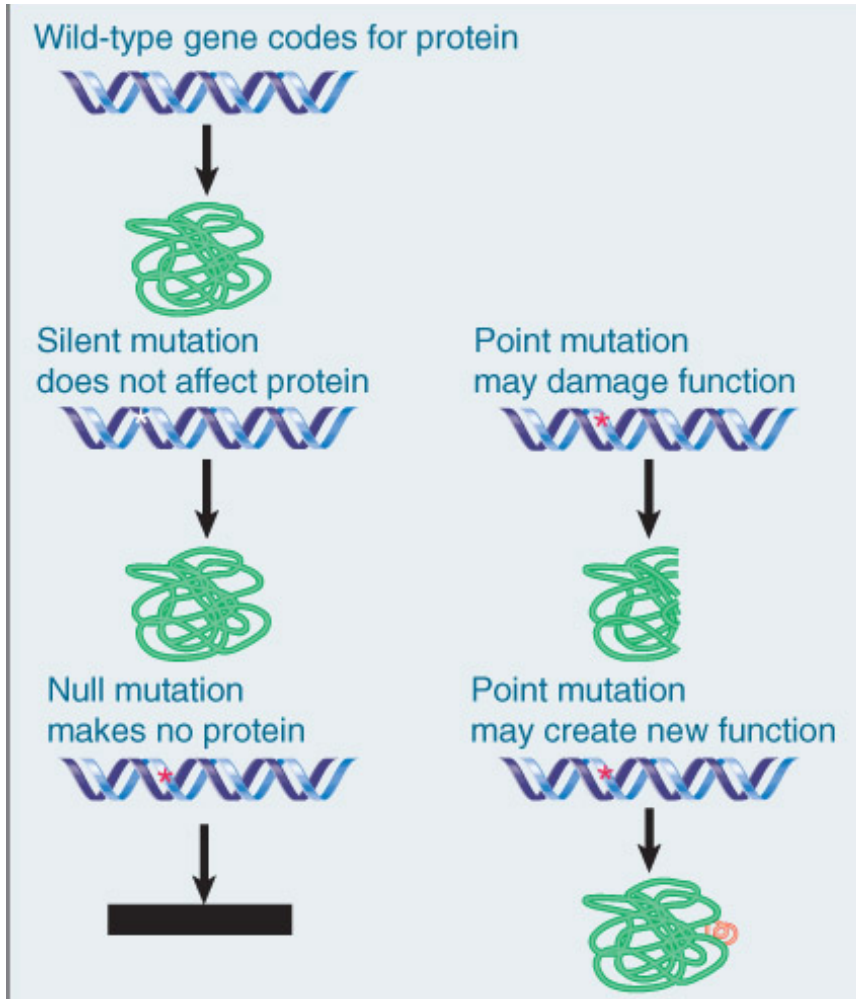
(Science, Nov 2016)

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-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. Mutations



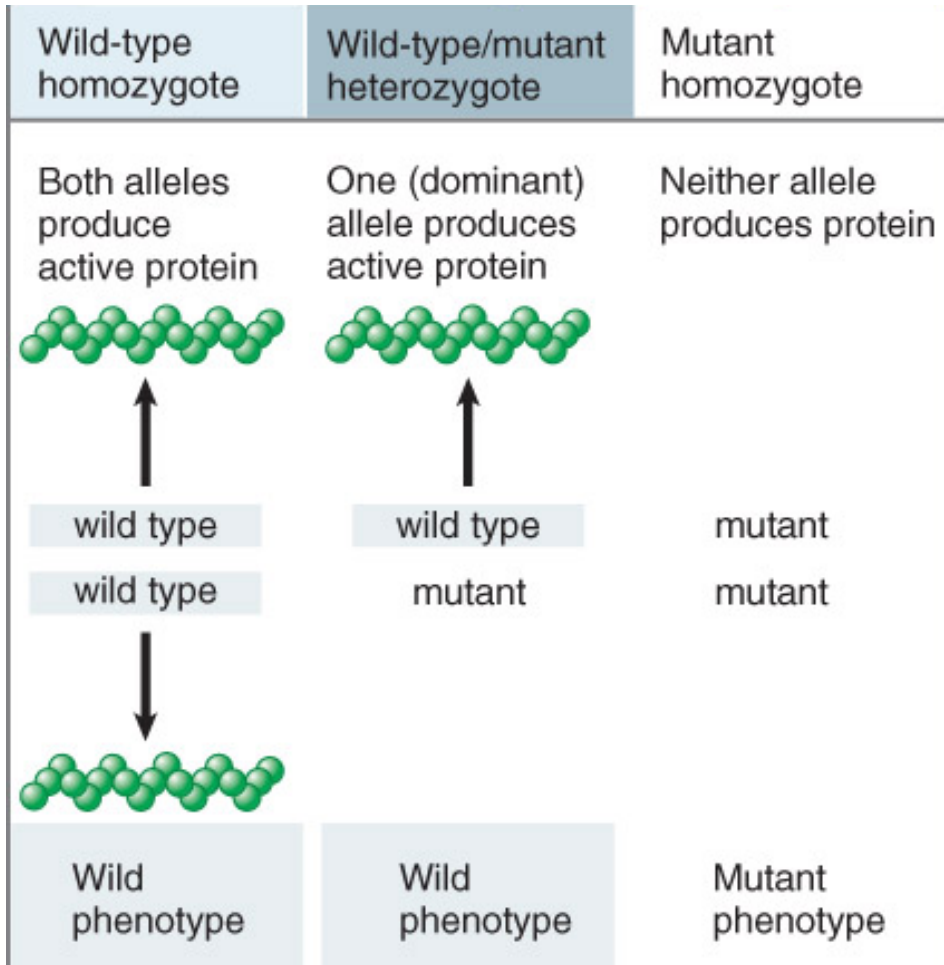
Mutations can be :

Silent (no effect)

Null (loss of function)

Gain of function (new function)

IV. Mutations



- 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, monoploid), **2n** (somatic cell, diploid) : **normal euploidy**

3n, 4n,...8n : **abberent euploidy**

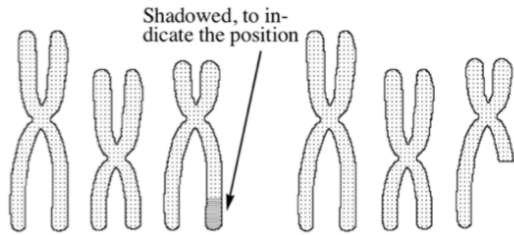
2n-1, 2n+1 : **aneuploidy**

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. Mutations

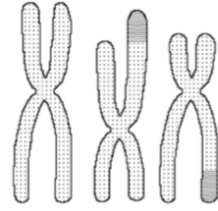
1 : **Chromosomal mutations** : modifications of the structure or copy number of chromosomes

Chromosome mutations

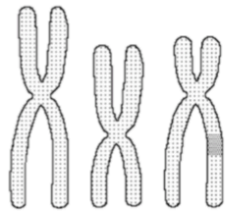


Normal subset of chromosomes

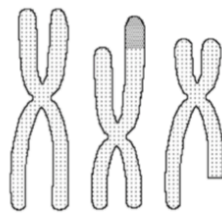
A deletion, a segment of a chromosome is lost



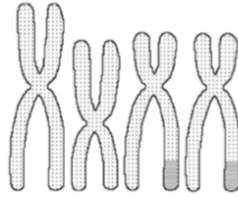
A duplication, a part is present more than once



An inversion, a segment is reversed



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

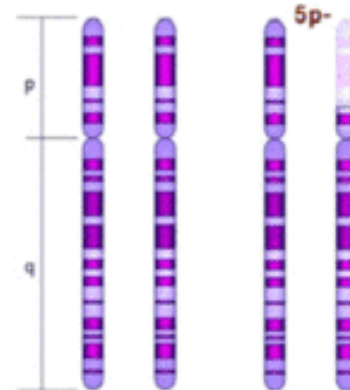


The number of chromosomes has changed

- Loss of DNA
- Move of DNA

➔ Serious for the cell

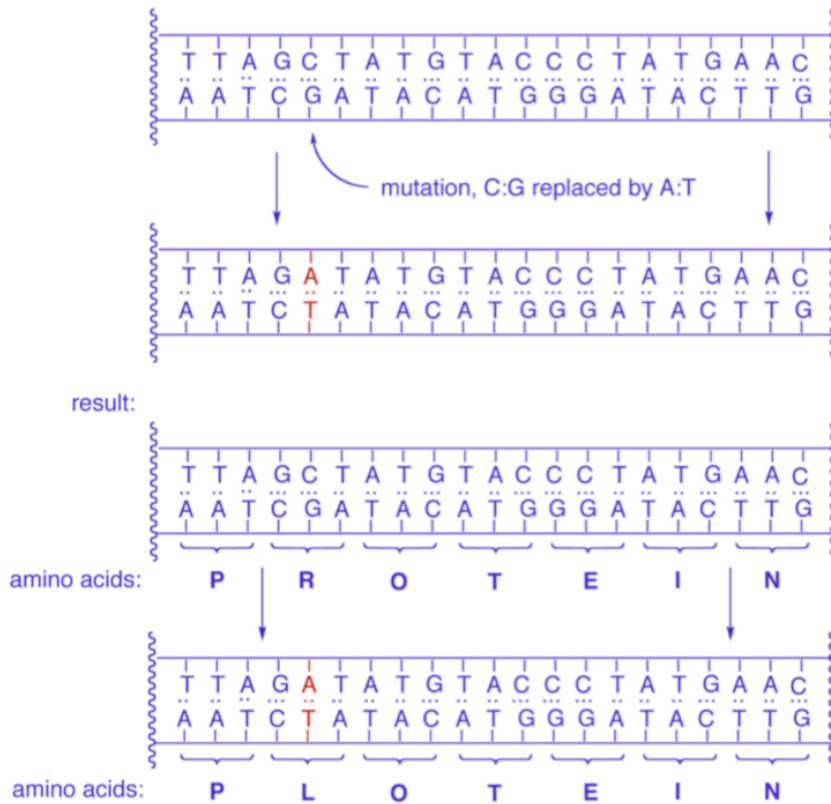
5p- syndrome = cri du chat syndrome



IV. Mutations

Substitutions

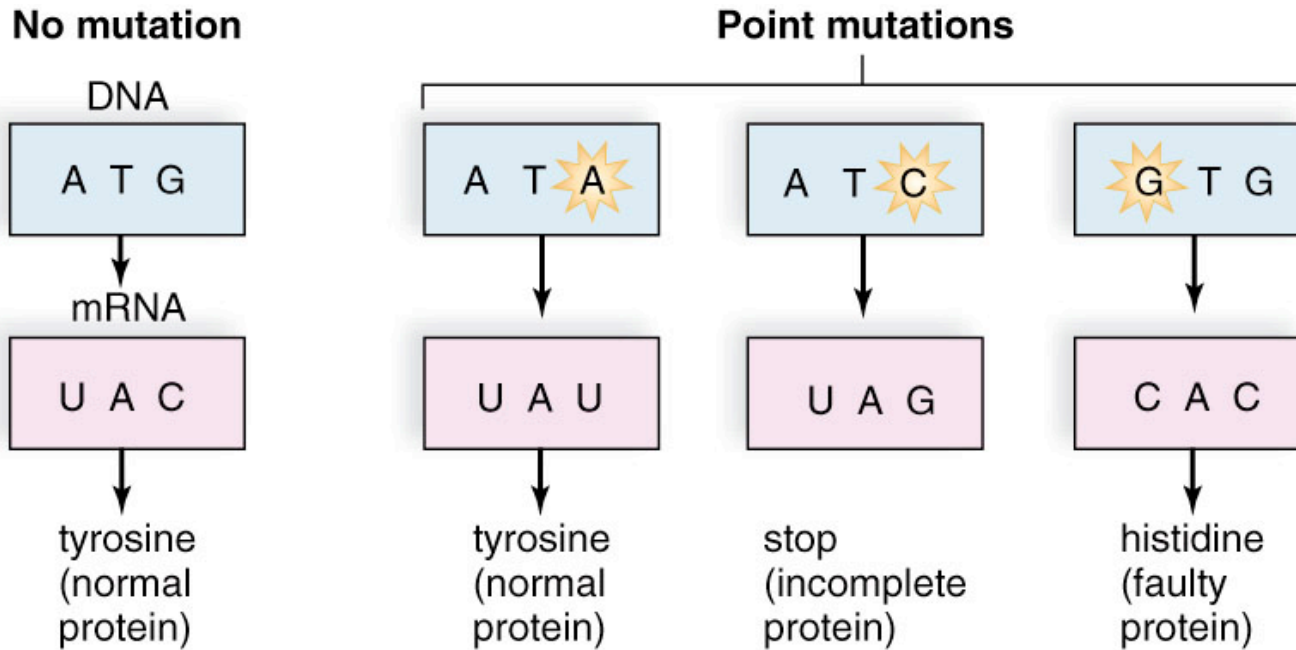
Base pair substitution



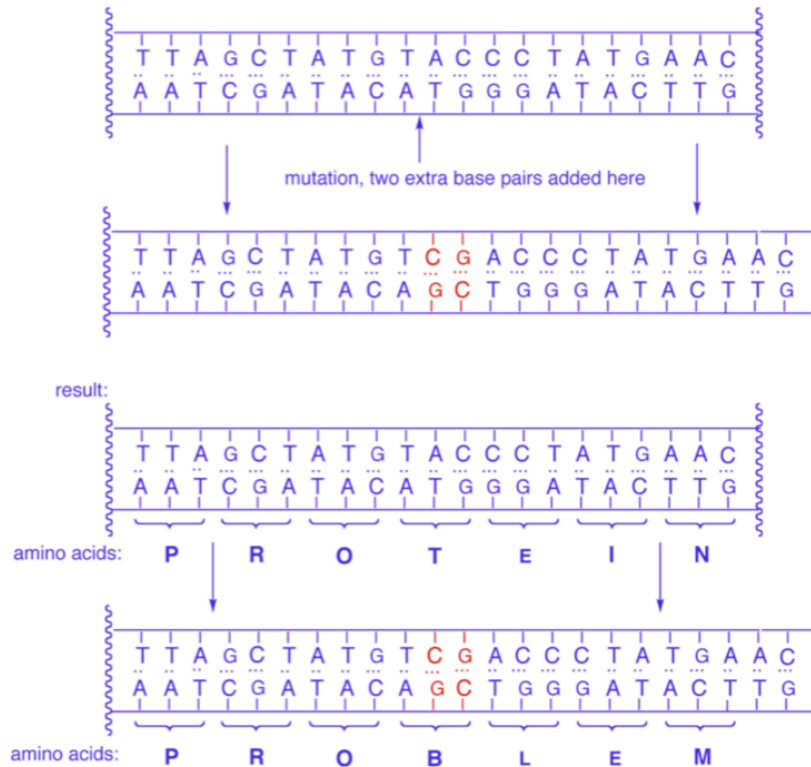
a limited damage

IV. Mutations

Genetic code



Frame shift mutation

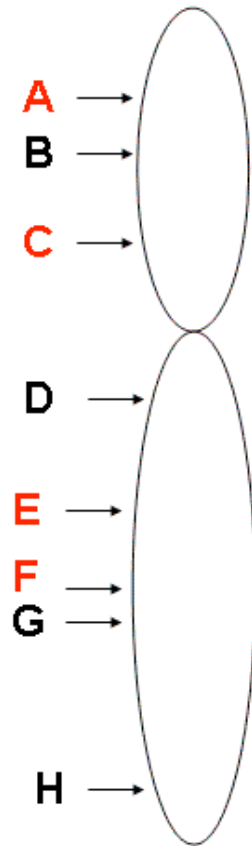


more
serious

Substitutions may change the sequence of the final polypeptide

V. Decoding the genomes

Genetic maps



Genetic Maps

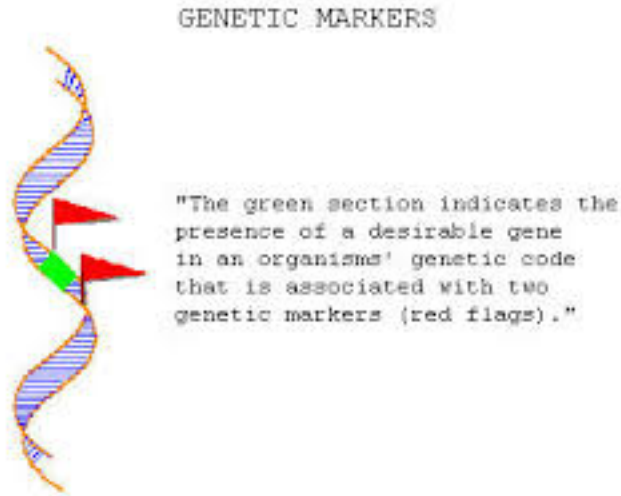
- Genes are shown in relative order and distance from each other based on pedigree studies.
- The chance of the chromosome breaking between A & C is higher than the chance of the chromosome breaking between A & B during meiosis.
- Similarly, the chance of the chromosome breaking between E & F is higher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

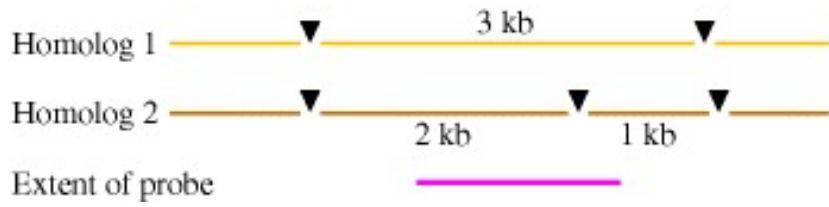
features: multiallelic (polymorphism), universal et inheritable



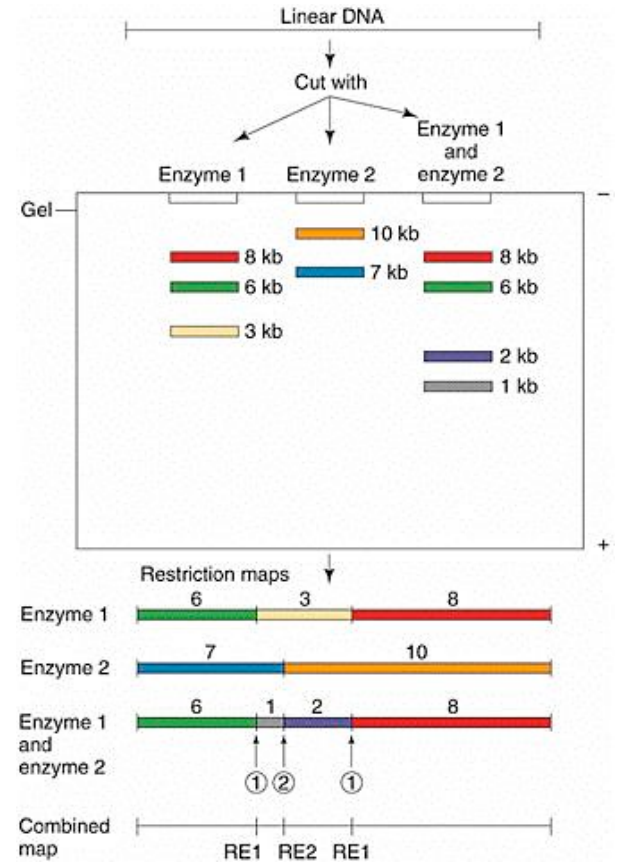
<http://www.youtube.com/watch?v=nrnJPC6e19c>

V. Decoding the genomes

RFLP : Restriction length fragment polymorphism

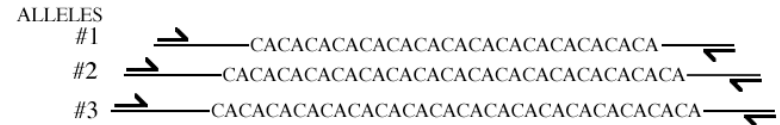


Gene coupling



V. Decoding the genomes

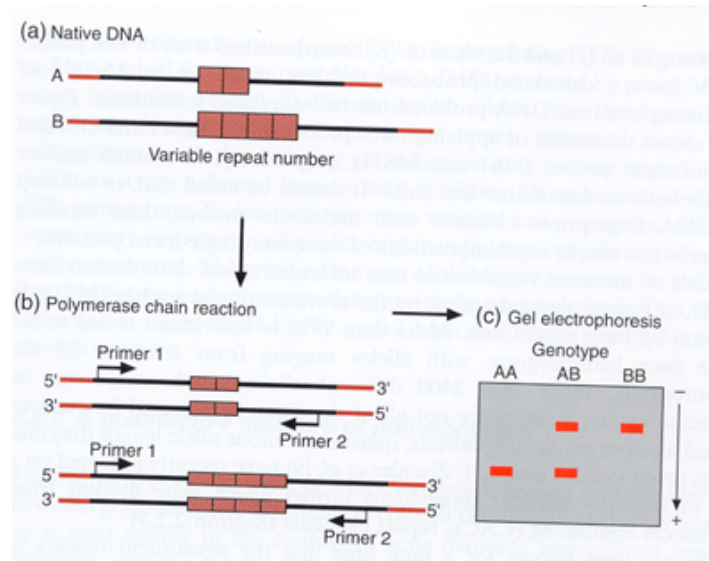
STRs : Short tandem repeats (VNTR)



Detected by RFLP :



Detected by AFLP :



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

V. Decoding the genomes : variant calling

Variant: Variation in a DNA sequence when comparing with a reference sequence

	Substitution	Insertion	Deletion
Wild-Type:	AACGGCC T GTAAC	AACGGCC T GTAAC	AACGGCC T GTAAC
Mutant:	AACGGCC A GTAAC	AACGGCC AGCT AAC	AACGGCC - GTAAC

Substitution:

SNV : Single nucleotide variant : any change of a nucleotide with no criteria of frequency

SNP : Single nucleotide polymorphism : a SNV found in >1% of the population

Individual 1: AACGGCC**T**GTAAC
Individual 2: AACGGCC**T**GTAAC
Individual 3: AACGGCC**T**GTAAC
Individual 4: AACGGCC**A**GTAAC
Individual 5: AACGGCC**T**GTAAC
Individual 6: AACGGCC**A**GTAAC

Individual 7: AACGGCC**T**GTAAC
Individual 8: AACGGCC**T**GTAAC
Individual 9: AACGGCC**T**GTAAC
Individual 10: AACGGCC**A**GTAAC
Individual 11: AACGGCC**T**GTAAC
Individual 12: AACGGCC**A**GTAAC

INDEL:

Insertion or DELetion of one or several nucleotides

MNV: Multi-Nucleotide Variant : multiple SNVs or INDELS in the same region

V. Decoding the genomes : variant calling

Variant: Variation in a DNA sequence when comparing with a reference sequence

SV: Structural Variants : a region of DNA > 1 kbp in which gross rearrangements have occurred



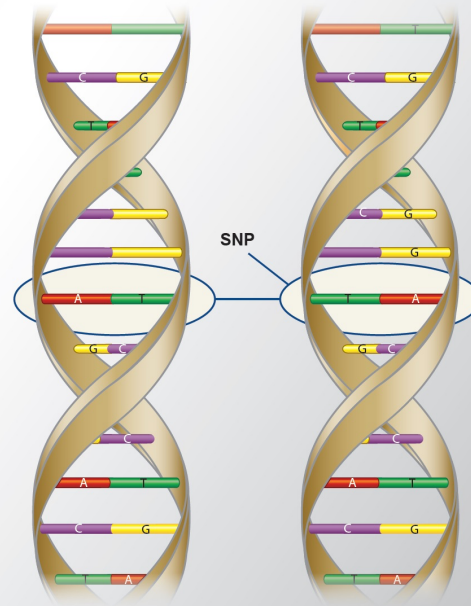
V. Decoding the genomes : variant calling

SNP : Single nucleotide polymorphism (>1% of the population)

What is a Single Nucleotide Polymorphism (SNP)?

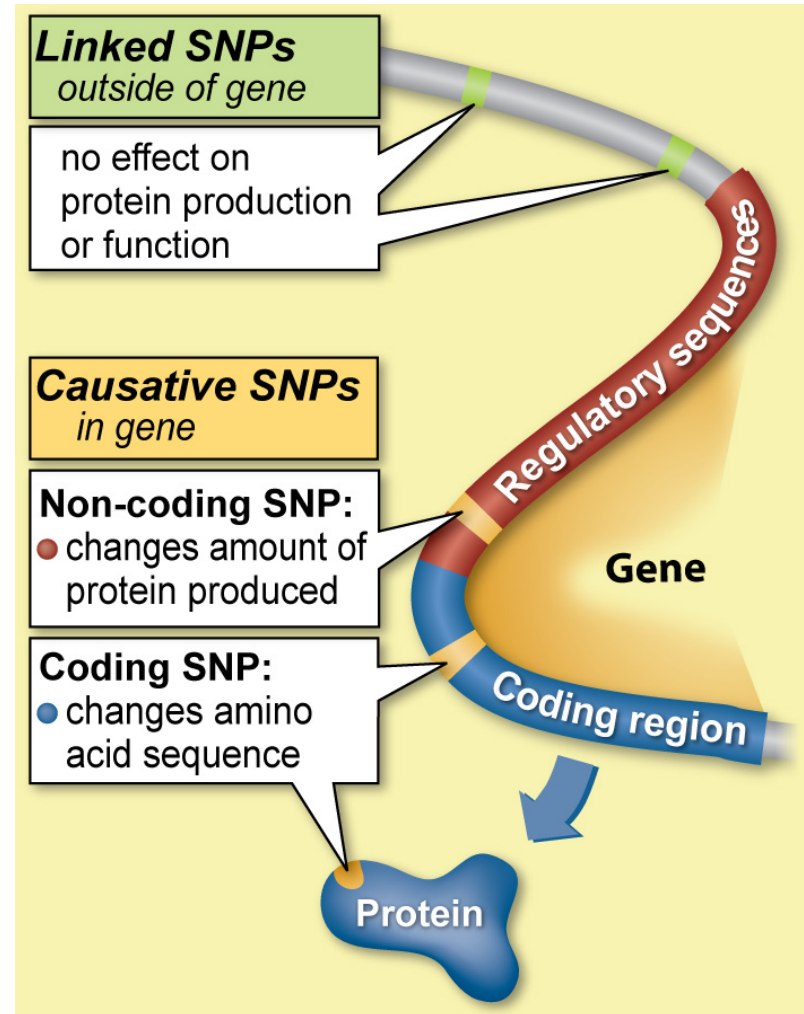
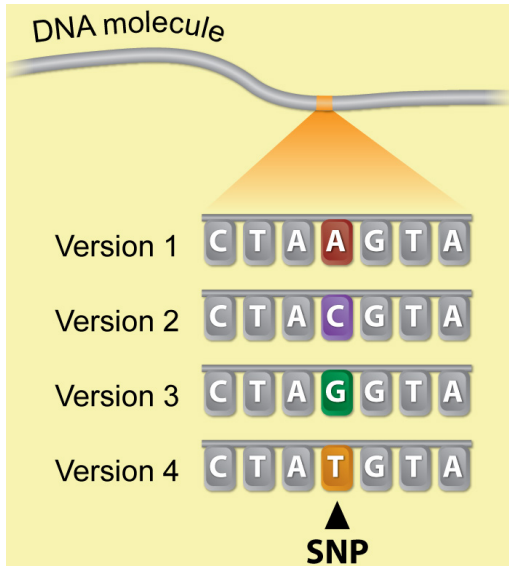
A SNP (pronounced "snip") is a DNA sequence variation that occurs when a single nucleotide (A, T, C, or G) in the genome sequence is modified.

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



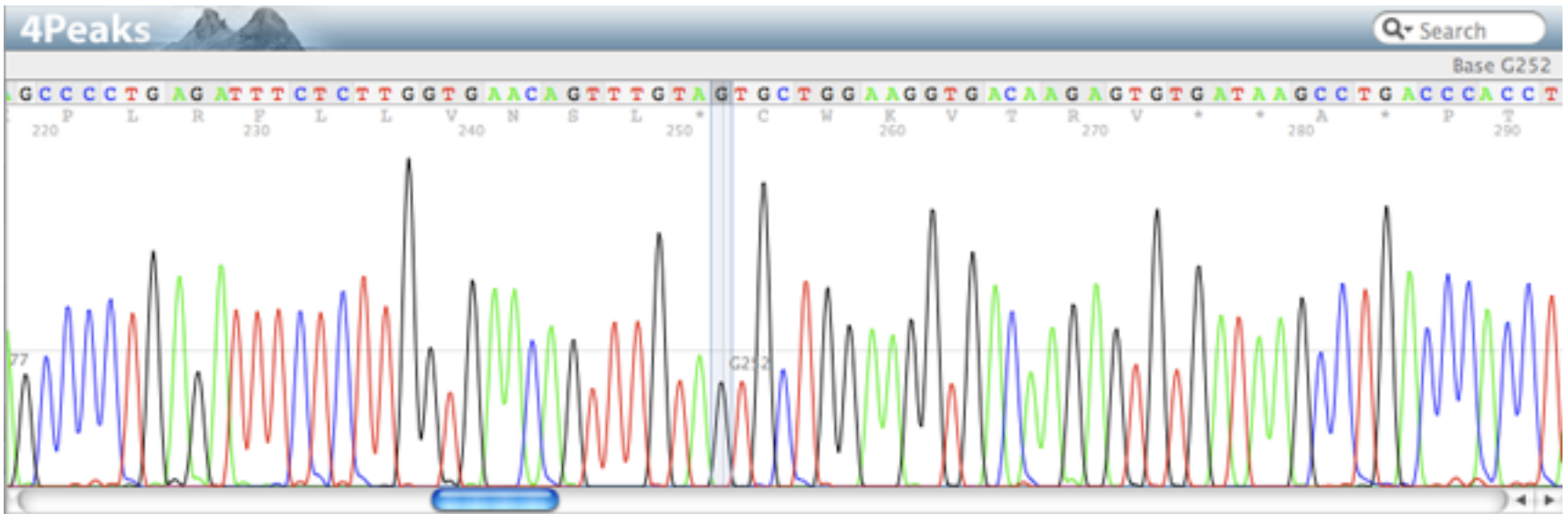
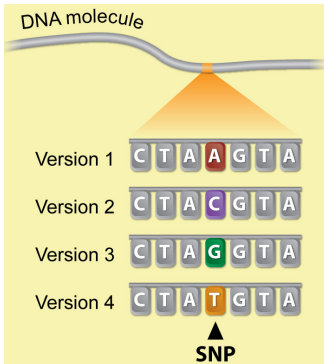
V. Decoding the genomes

SNP : Single nucleotide polymorphism (>1% of the population)



V. Decoding the genomes : variants

SNP : Single nucleotide polymorphism (>1% of the population)



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