

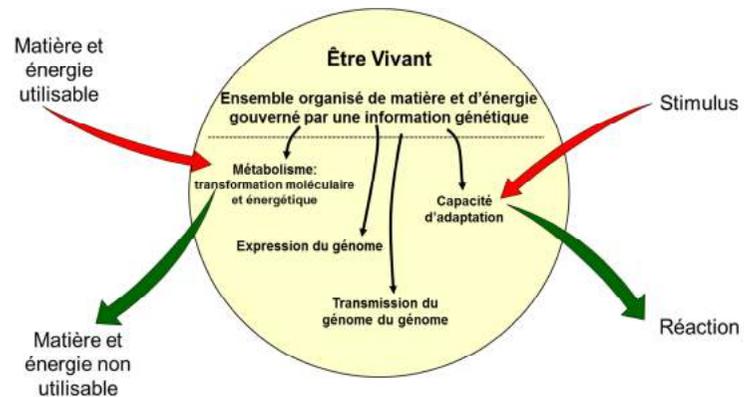
Une introduction au système d'information génétique des êtres vivants

Introduction: Rappel sur les êtres vivants et leur diversité

Être vivant: Système organisé de matière et d'énergie gouverné par une information génétique (ou genome) permettant le développement de caractéristiques propre à la vie:

- Capacité d'auto-organisation en puisant les ressources extérieures. Cette capacité est basée sur un ensemble de réaction chimique (appelé métabolisme) catalysée par des protéines (appelées enzymes)
- Capacité de reproduction assurant la transmission du matériel génétique. Cette transmission peut se faire à l'identique ou impliquer des processus assurant l'évolution du matériel génétique.
- Capacité (variable selon les êtres vivants) d'adaptation à l'environnement dans une gamme restreinte par les capacités du matériel génétique.

Le génome est le matériel génétique transmis de génération en génération et porté par la molécule d'ADN. C'est donc le matériel héréditaire des organismes vivants: l'ADN possède donc des propriétés qui assurent sa transmission au cours des générations cellulaires. Sa structure lui permet aussi de contenir des informations (appelés gènes) permettant le fonctionnement de l'organisme vivant: transmission et expression du patrimoine génétique, mise en place de l'organisation cellulaire, adaptation au milieu.



1. Les bases de l'information chez les êtres vivants : de l'ADN aux protéines

1.1. L'ADN support de l'information génétique

- 1.1.1. Expérience de mise en évidence du rôle de l'ADN
- 1.1.2. Structure de l'ADN
- 1.1.3. La copie de la molécule d'ADN, base de la transmission génétique

1.2. L'ADN, une molécule sémantique : un code pour la synthèse des protéines

- 1.2.1. Les protéines, acteurs majeurs de l'activité cellulaire
- 1.2.2. La relation « un gène-une protéine »
- 1.2.3. La nature du code

1.3. De l'ADN aux protéines : le dogme de la biologie moléculaire

- 1.3.1. Le dogme de la biologie moléculaire: la transcription
- 1.3.2. Le dogme de la biologie moléculaire: la traduction

1.4. Information génétique et contrôle des activités cellulaires

- 1.4.1. Mécanismes de contrôle de l'expression génétique
- 1.4.2. Mécanismes de contrôle post-traductionnel de l'activité des protéines

2. Quelles informations rechercher dans les génomes ?

2.1. Préalable : Les différentes séquences présentes dans l'ADN des êtres vivants

- 2.1.1. Les séquences codantes
- 2.1.2. Les séquences non codantes: ARNt, ARNr
- 2.1.3. Les séquences régulatrices (réplication, transcription, traduction)
- 2.1.4. Les autres types de séquences (ADN non codant, intron, ADN répété, transposon...)

2.2. Analyse de la taille des génomes et de leur contenu en gènes codants

- 2.2.1. Diversité des tailles et des contenus en gènes des génomes procaryotes
- 2.2.2. Le paradoxe de la taille des génomes eucaryotes : l'importance des régions non codantes

2.3. Analyse du pourcentage en G/C des génomes

- 2.3.1. Répartition des pourcentages en G/C dans les génomes
- 2.3.2. Etude du pourcentage en G/C pour comprendre l'évolution des bactéries

2.4. Comparaison de séquence : l'alignement et ses enseignements

- 2.4.1. Qu'est-ce qu'un alignement de séquences ?
- 2.4.2. Alignement de séquence et étude du polymorphisme génétique dans une espèce
- 2.4.3. Alignement des séquences et étude de l'évolution des êtres vivants (phylogénie)
- 2.4.4. Alignement des séquences et annotation des génomes

Qu'est-ce qu'un génome?
Une introduction au système d'information
génétique des êtres vivants

Quelles informations contient un génome ?

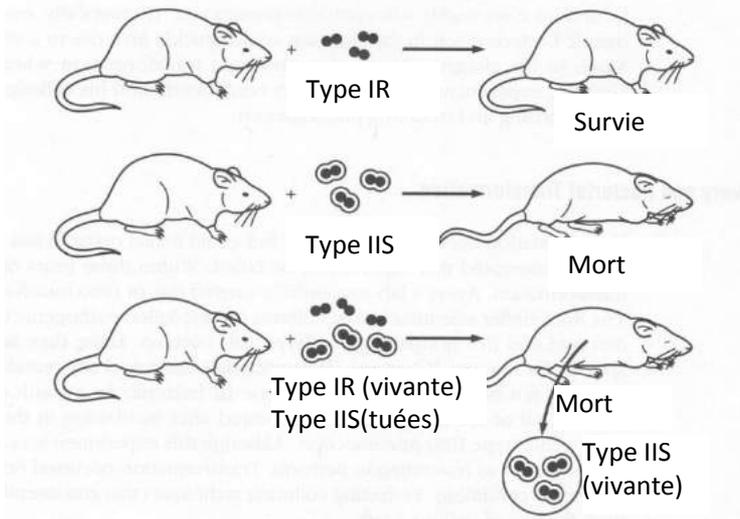
- Structure de l'information génétique
- Analyse de génome: bases de la recherche d'informations dans les génomes

Qu'est-ce qu'un génome?
Une introduction au système d'information
génétique des êtres vivants

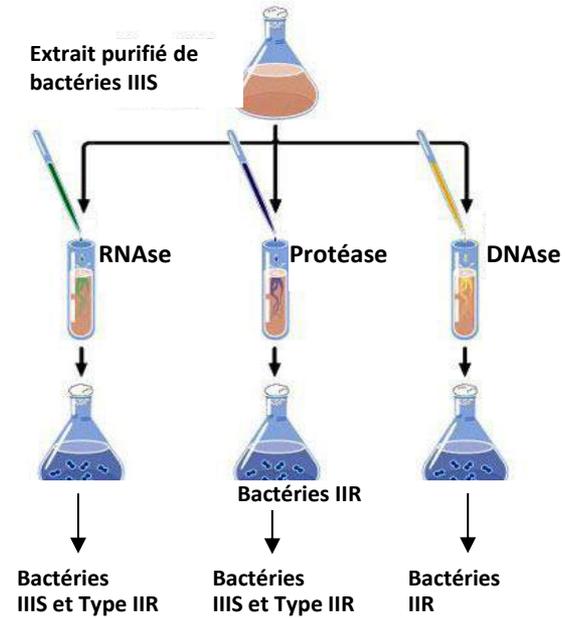
Quelles informations contient un génome ?
Structure de l'information génétique

La découverte de l'ADN comme support de l'information génétique

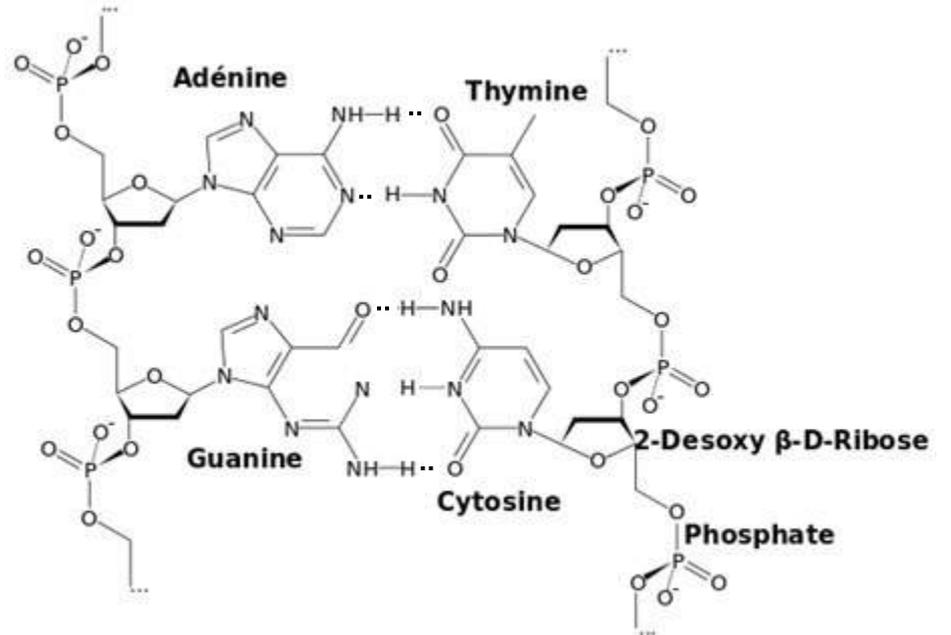
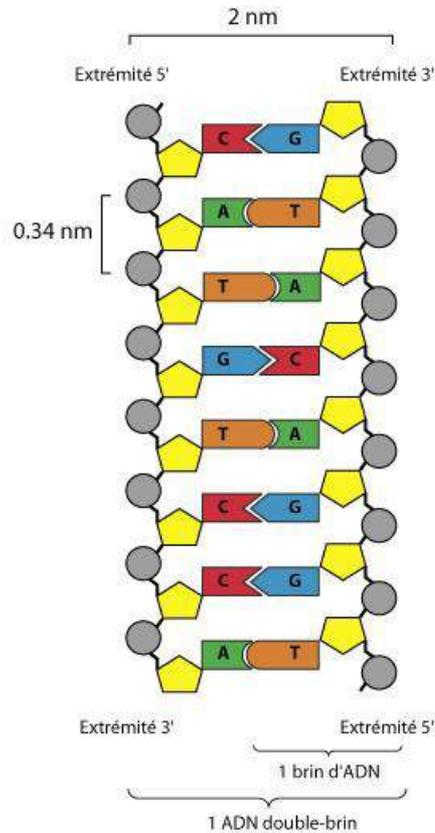
Expérience de transformation de Griffith (1928)



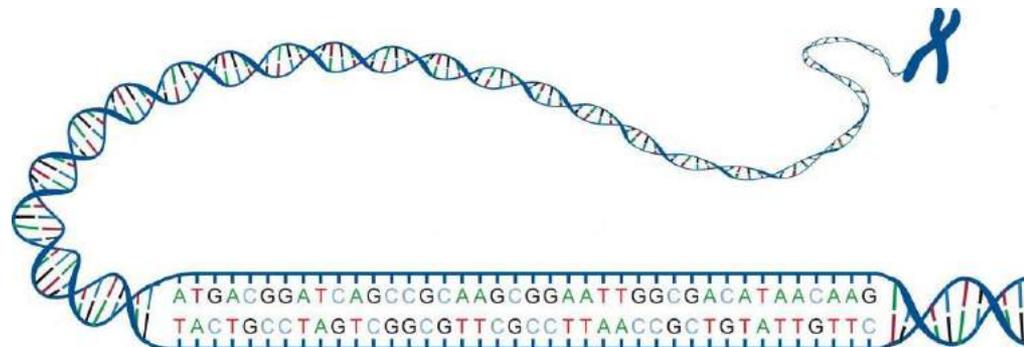
Purification du facteur transformant par Avery (1943)



La molécule d'ADN (structure élucidée en 1953 par Watson et Crick)



Chromosome

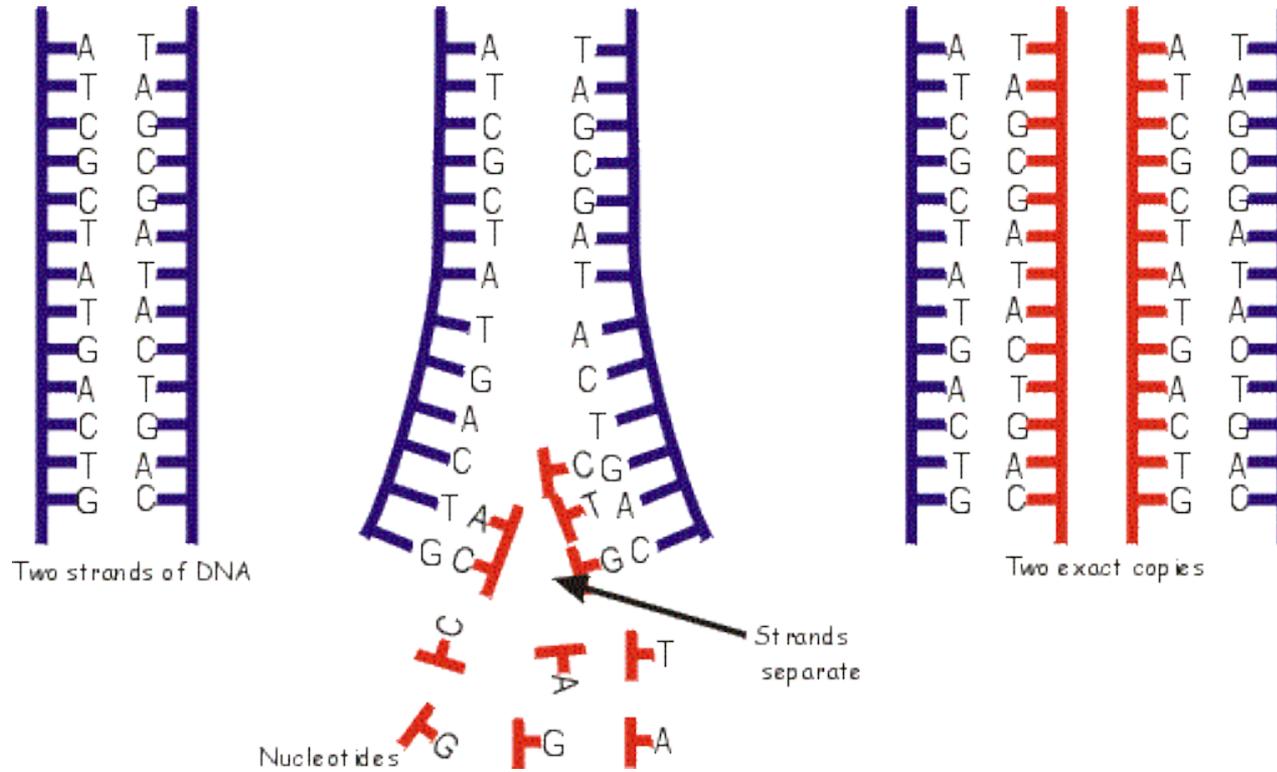


bases azotées:

	A adénine		phosphate
	C cytosine		sucré
	G guanine		nucléotide
	T thymine		

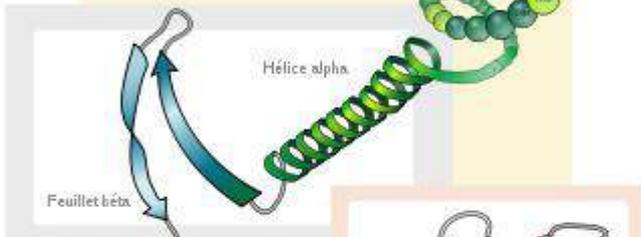
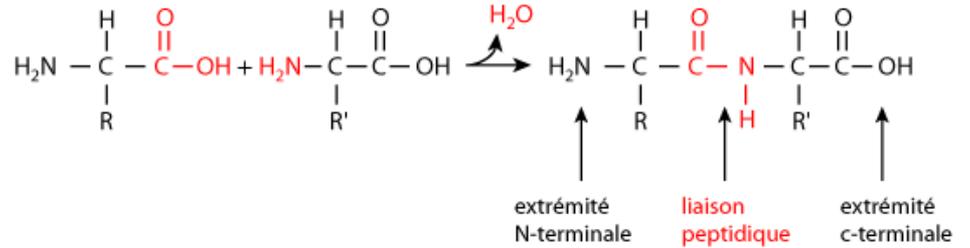
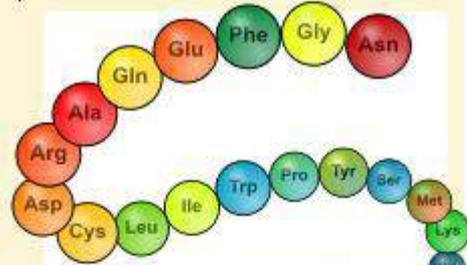
L'adénine ne s'apparie qu'avec la thymine
la cytosine ne s'apparie qu'avec la guanine

La réplication semi-conservative de la molécule d'ADN

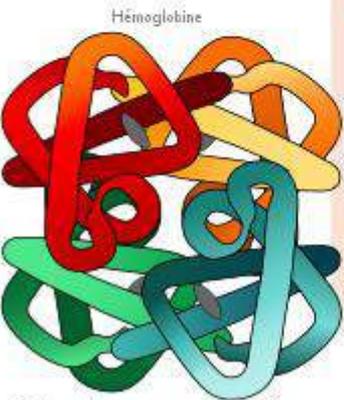


Les protéines

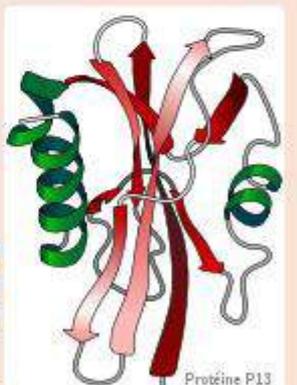
Structure primaire
Séquence d'acides aminés



Structure secondaire
Repliement local de la chaîne principale



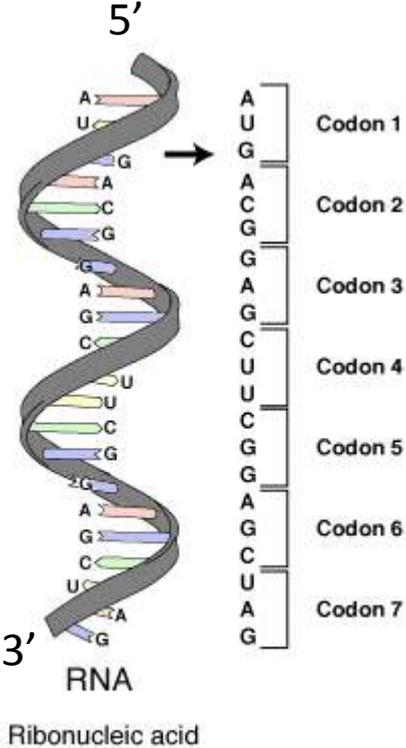
Structure quaternaire
Association de plusieurs chaînes polypeptidiques



Structure tertiaire
Structure tridimensionnelle

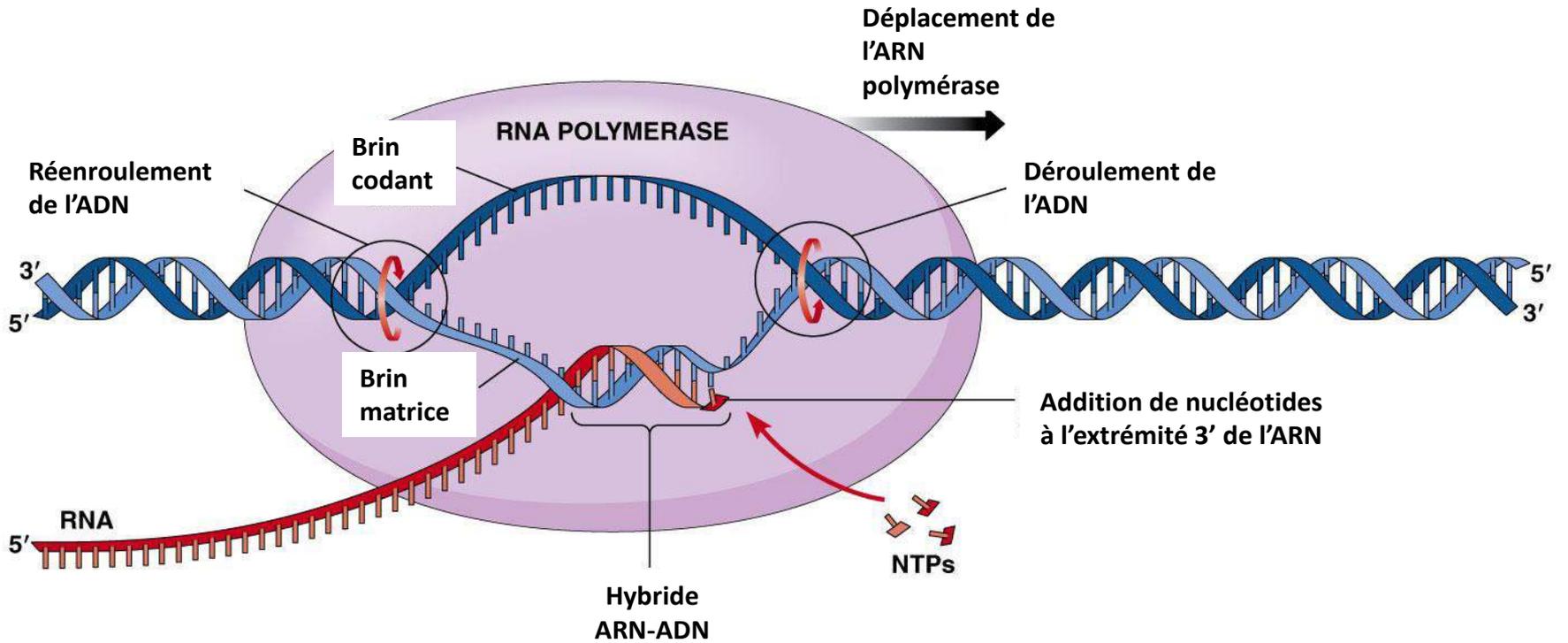
non polaires				aromatiques		
$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{H} \end{array}$ glycine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_3 \end{array}$ alanine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH} \\ \quad \\ \text{CH}_3 \quad \text{CH}_3 \end{array}$ valine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_5 \end{array}$ phénylalanine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_4 - \text{OH} \end{array}$ tyrosine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_8\text{H}_6\text{N}_2 \end{array}$ tryptophane	
$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{C}_5\text{H}_7\text{N} \end{array}$ proline	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH} \\ \quad \\ \text{CH}_3 \quad \text{CH}_3 \end{array}$ leucine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{S} \\ \\ \text{CH}_3 \end{array}$ méthionine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{H} - \text{C} - \text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH}_3 \end{array}$ isoleucine			
polaires mais non-chargé				chargés positif		
$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2\text{OH} \end{array}$ sérine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{H} - \text{C} - \text{OH} \\ \\ \text{CH}_3 \end{array}$ thréonine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{SH} \end{array}$ cystéine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH}_2^+ \end{array}$ lysine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH} \\ \\ \text{C} = \text{NH}_2^+ \end{array}$ arginine	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{C} - \text{NH} \\ \quad \\ \text{C} \quad \text{CH} \\ \quad \\ \text{C} \quad \text{N} \\ \quad \\ \text{H} \end{array}$ histidine	
				chargés négatif		
				$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{COO}^- \end{array}$ aspartate	$\begin{array}{c} \text{COO}^- \\ \\ \text{H}_3\text{N}^+ - \text{C} - \text{H} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{COO}^- \end{array}$ glutamate	

Le code génétique

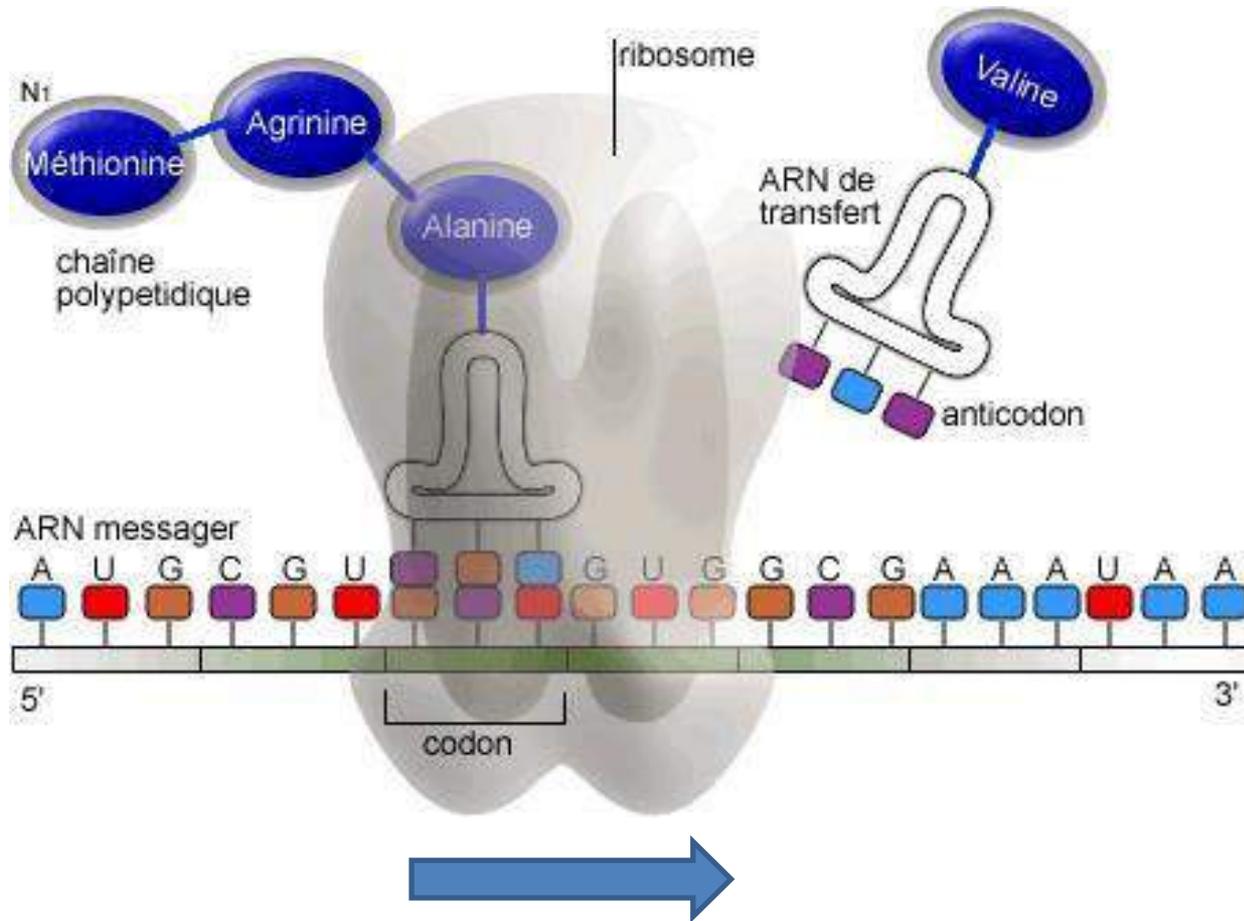


		Second Letter					
		U	C	A	G		
1st letter	U	UUU Phe UUC UUA Leu UUG	UCU UCC Ser UCA UCG	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G	
	C	CUU CUC Leu CUA CUG	CCU CCC Pro CCA CCG	CAU His CAC CAA Gln CAG	CGU CGC Arg CGA CGG	U C A G	
	A	AUU AUC Ile AUA AUG Met	ACU ACC Thr ACA ACG	AAU 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	
						3rd letter	

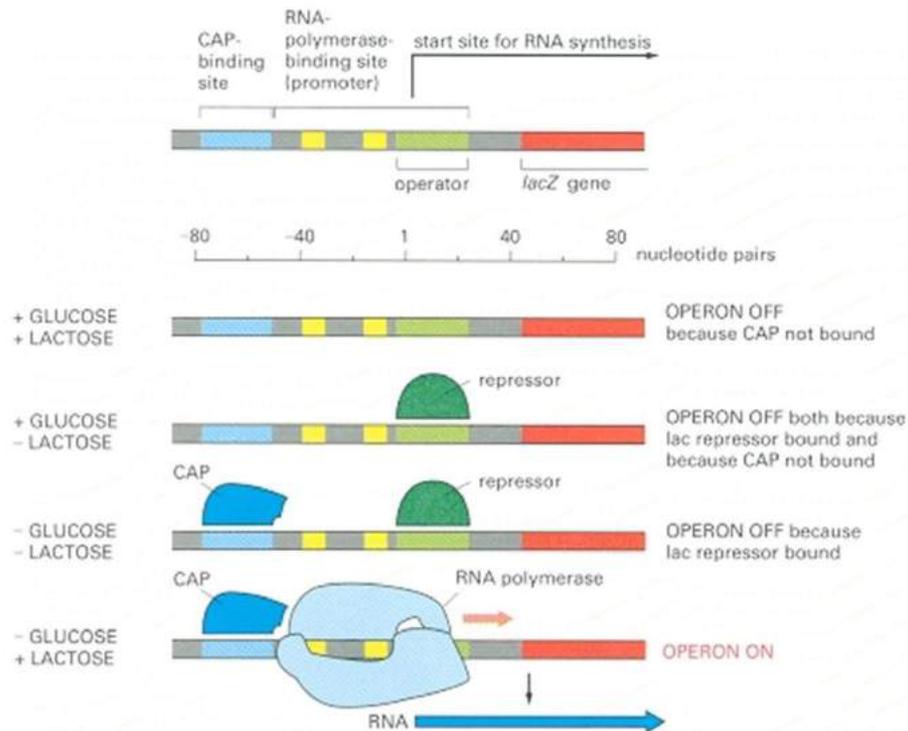
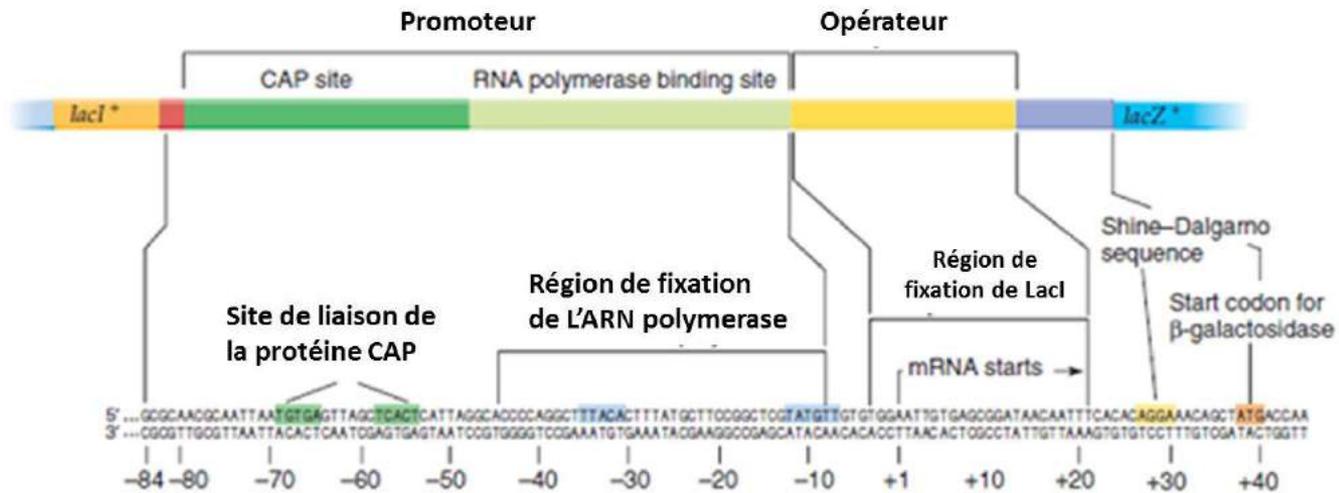
La transcription



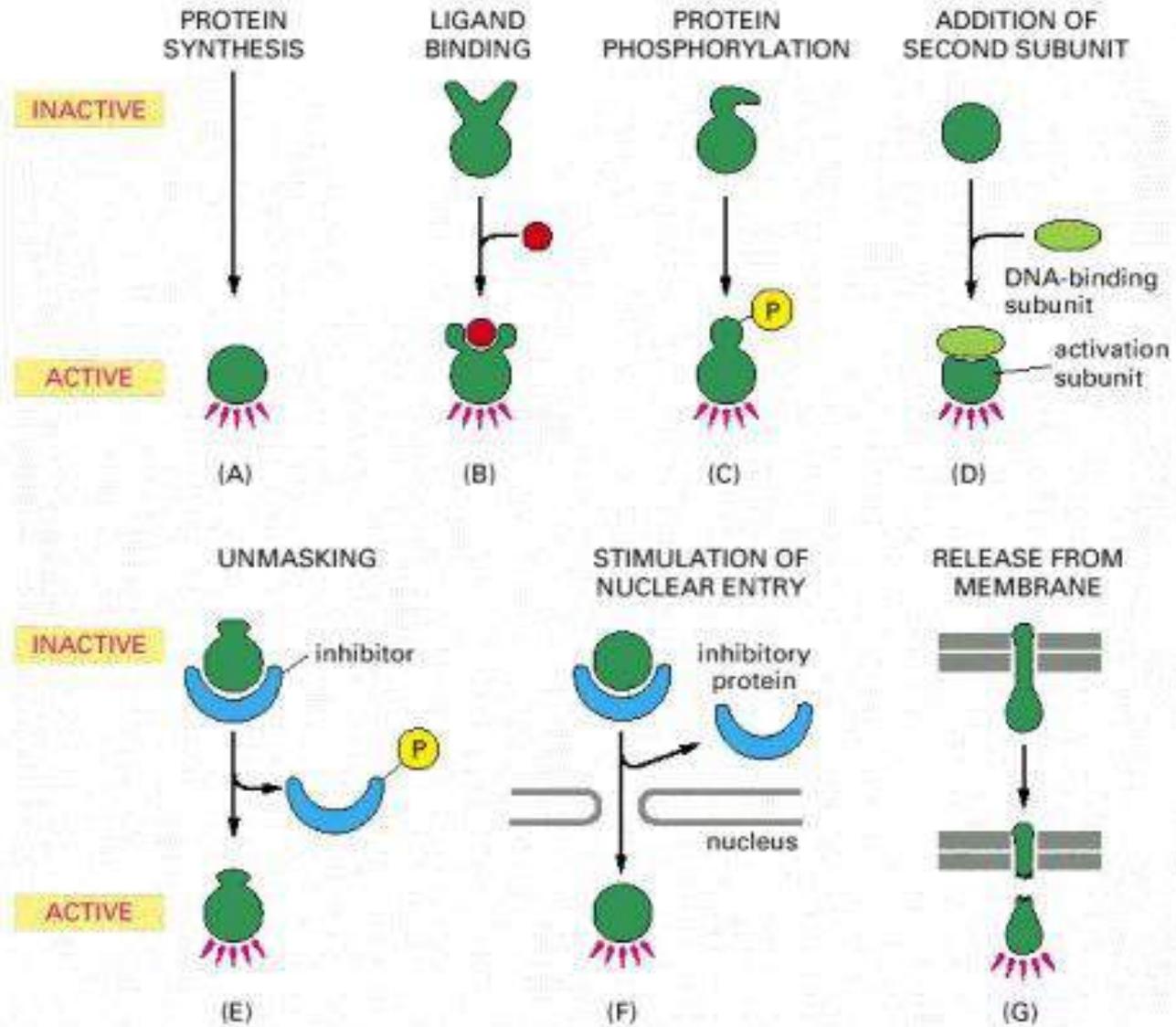
La traduction



Un exemple de séquences régulatrices de la transcription chez les bactéries



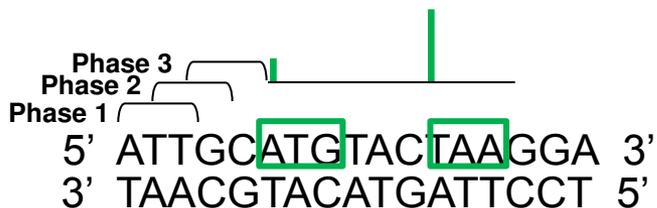
Divers mécanismes de contrôle de l'activité des protéines



Qu'est-ce qu'un génome?
Une introduction au système d'information
génétique des êtres vivants

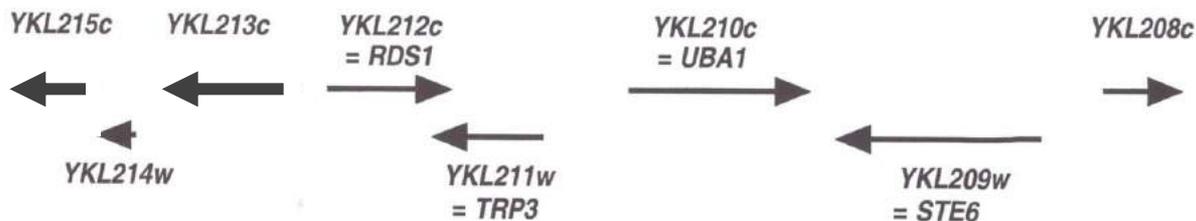
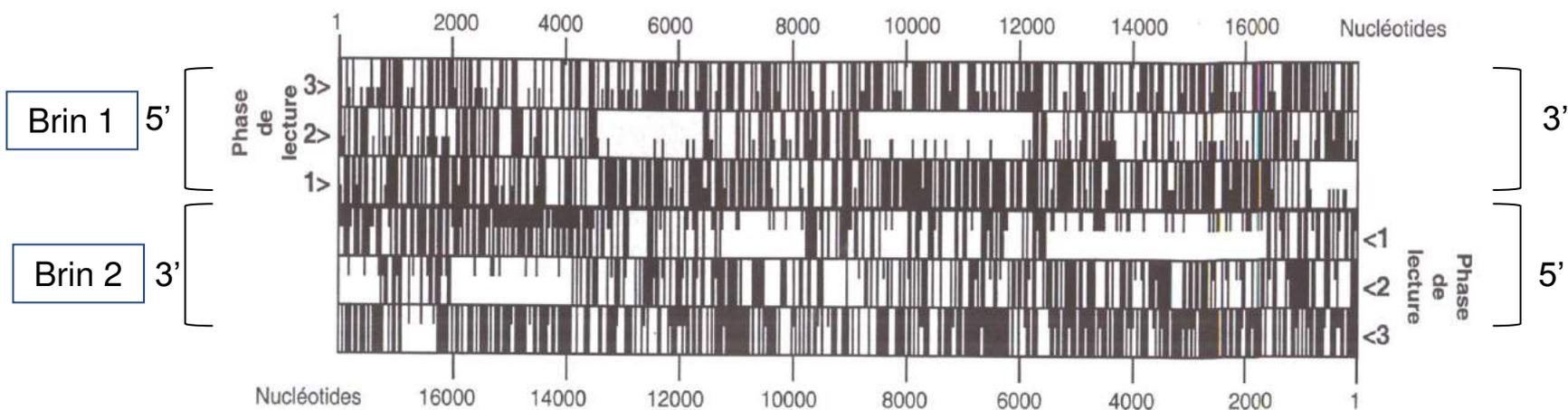
Quelles informations contient un génome ?
Analyse de génome: bases de la recherche d'informations dans les génomes

Identification des séquences codante dans un génome



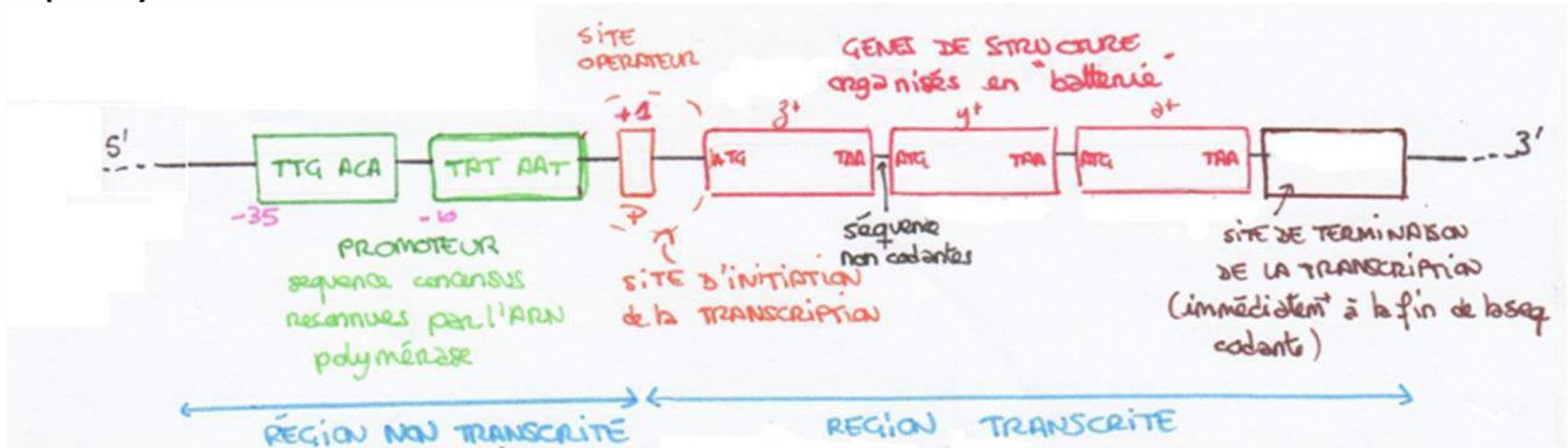
ORF: Open Reading Frame: entre deux codon stop

CDS: Coding Sequence: entre une ATG et un stop

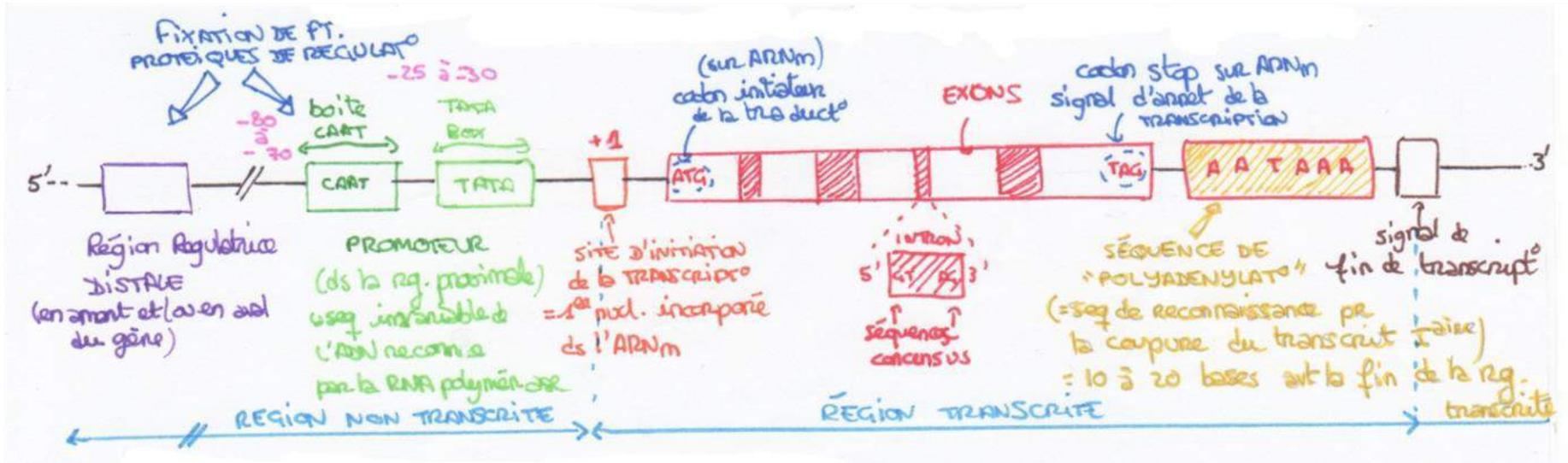


Structure des gènes codant les protéines chez les procaryotes et les eucaryotes

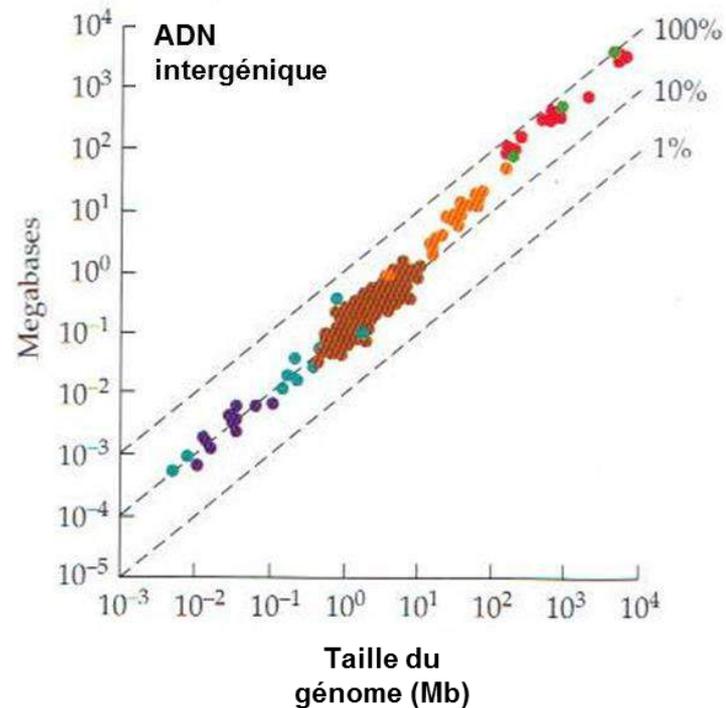
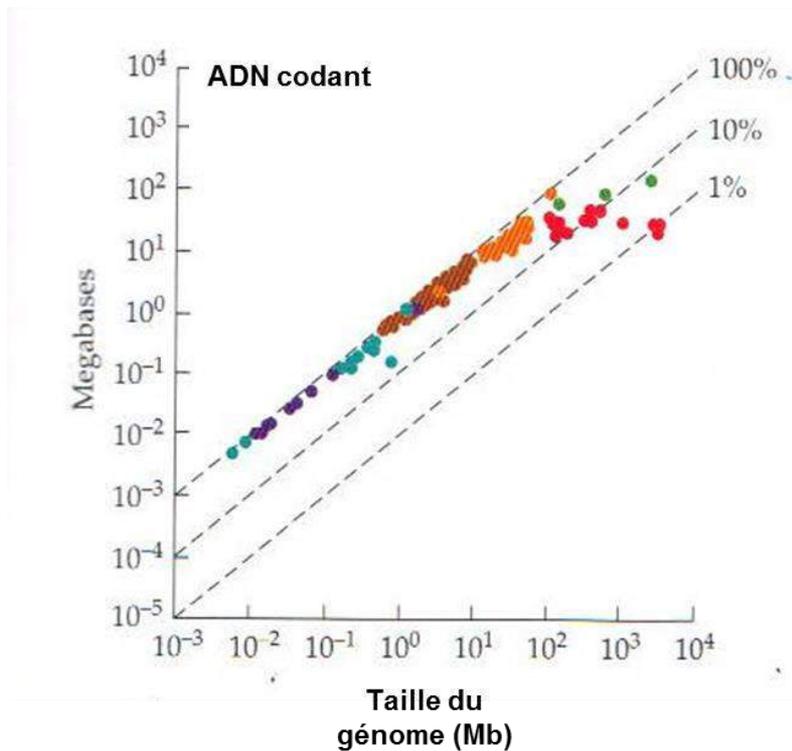
Gène procaryote



Gène eucaryote

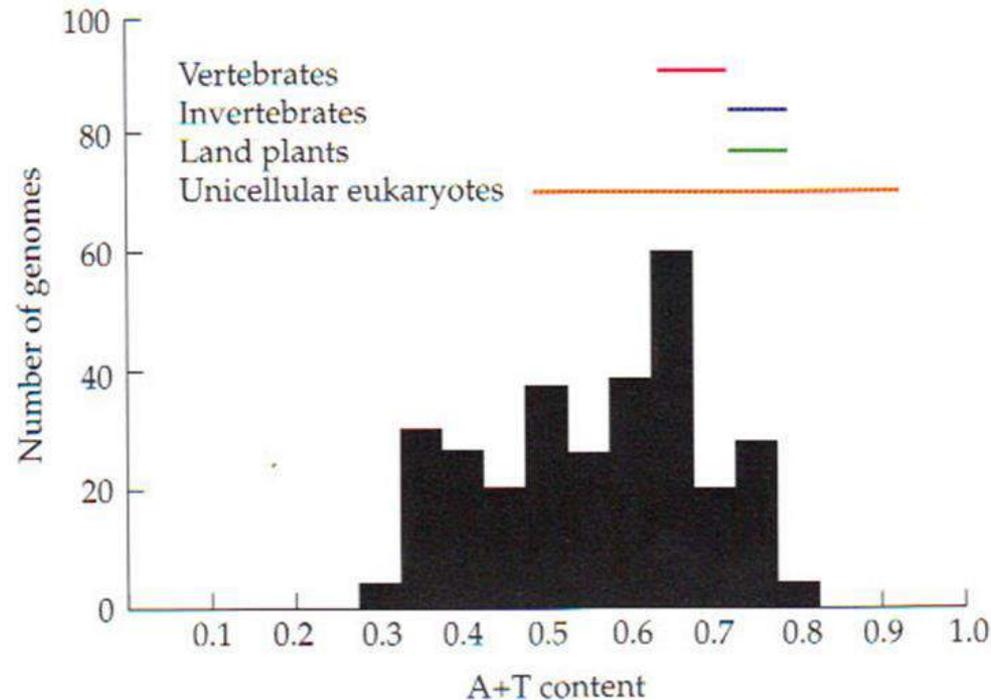


Comparaison de la taille des génomes de diverses espèces séquencés



- Land plant nuclear genome
- Animal nuclear genome
- Unicellular eukaryote nuclear genome
- Prokaryote
- Eukaryotic DNA virus
- Bacteriophage

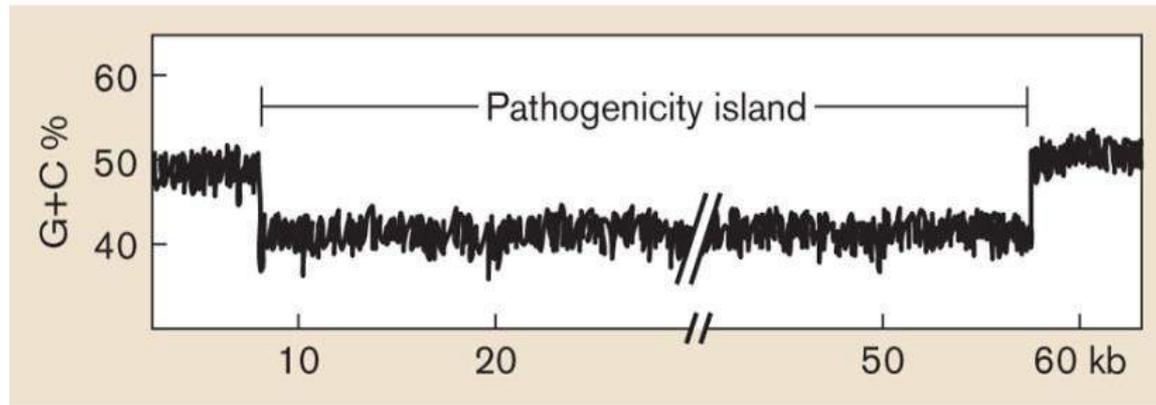
Distribution du pourcentage en (A+T) dans les génomes séquencés



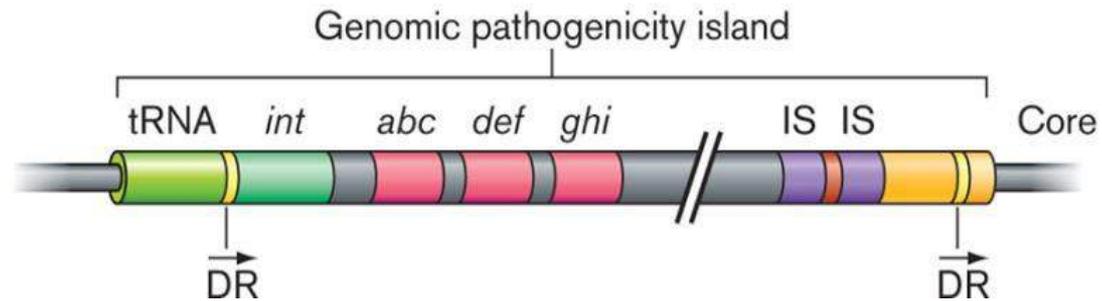
Distribution of genome-wide A+T content for prokaryotes with completely sequenced genomes (histogram) and ranges for eukaryotes (colored lines). The prokaryotic mean and SD are 0.54 and 0.13 (sample size = 293). Note that because of the Watson–Crick base complementarity of As and Ts, the A+T content is identical for both strands.

Détection d'un îlot de pathogénicité par étude du contenu en G+C d'un génome bactérien

A.



B.



Alignement de séquences de différents individus d'une même espèce

```
individu4 ATGGTTAAAACGACTGAAGTAGTAAGCGAAGTTTCAAAGGTGGCAGGTGTAAGCCATGGGCAGGTATAT 70
individu5 ATGGTTAAAACGACTGAAGTAGTAAGCGAAGTTTCAAAGGTGGCAGGTGTAAGACCATGGGCAGGTATAT 70
individu2 ATGGTTAAAACGACTGAAGTAGTAAGCGAAGTTTCAAAGGTGGCAGGTGTAAGACCATGGGCAGGTATAT 70
individu1 ATGGTTAAAACGACTGAAGTAGTAAGCGAAGTTTCAAAGGTGGCAGGTGTAAGACCATGGGCAGGTATAT 70
individu3 ATGGTTAAAACGACTGAAGTAGTAAGCGAAGTTTCAAAGGTGGCAGGTGTAAGACCATGGGCAGGTATAT 70

individu4 TCTTG--TTGAAATTCAAGAGGATATACTCGCGGATGAGTTTACGTTTCGAGGCATTAATGAGAAGCTTTGC 138
individu5 TCGTTGTTGAAATTCAAGAGGATATACTCGCGGATGAGTTTACGTTTCGAGGCATTAATGAGAAGCTTTGC 140
individu2 TCGTTGTTGCA-ATTCAGAGGATATACTCGCGGATGAGTTTACGTTTCGAGGCATTAATGAGAAGCTTTGC 139
individu1 TCGTTGTTGAA-ATTCAGAGGATATACTCGCGGATGAGTTTACGTTTCGAGGCATTAATGAGAAGCTTTGC 139
individu3 TCGTTGTTGAA-ATTCAGAGGATATACTCGCGGATGAGTTTACGTTTCGAGGCATTAATGAGAAGCTTTGC 139
```


Arbre phylogénétique du vivant basé sur l'alignement de l'ARNr 16S

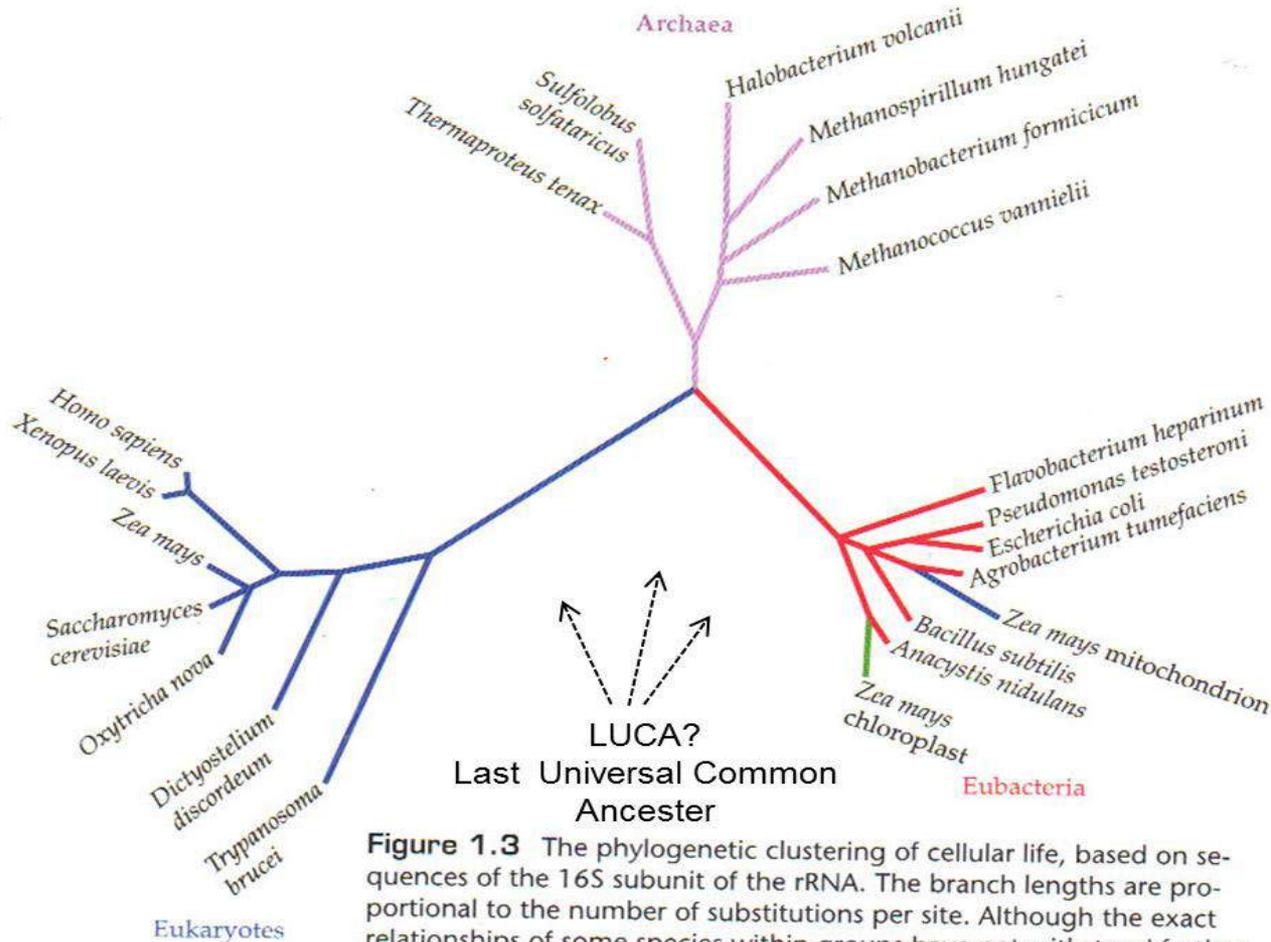


Figure 1.3 The phylogenetic clustering of cellular life, based on sequences of the 16S subunit of the rRNA. The branch lengths are proportional to the number of substitutions per site. Although the exact relationships of some species within groups have not withstood further scrutiny, the distinct nature of the three major domains is well accepted. The presence of mitochondrial and chloroplast sequences in the eubacterial lineage provides compelling evidence for the eubacterial ancestry of these organelles. The tree is unrooted, as the position of the most recent common ancestor of the three major groups is not identified. (Modified from Pace et al. 1986.)