

Séance 6: Génomes et séquençage (Partie 1)

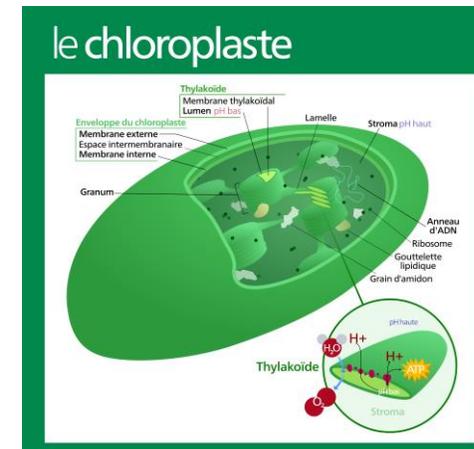
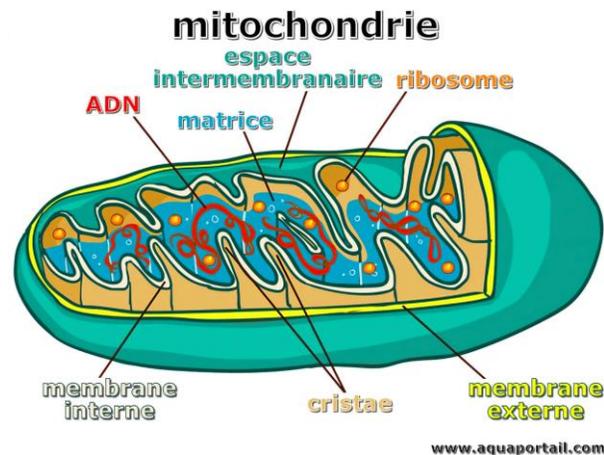
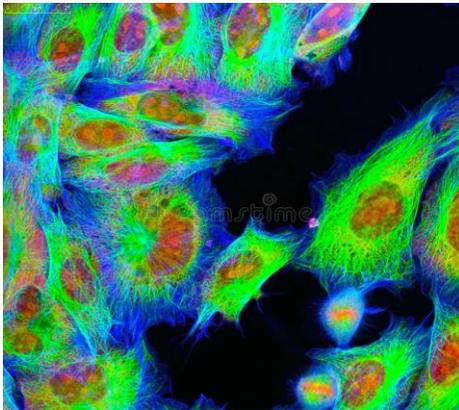
Cours de Biologie UE Codage

SARAH BEN SADOUN

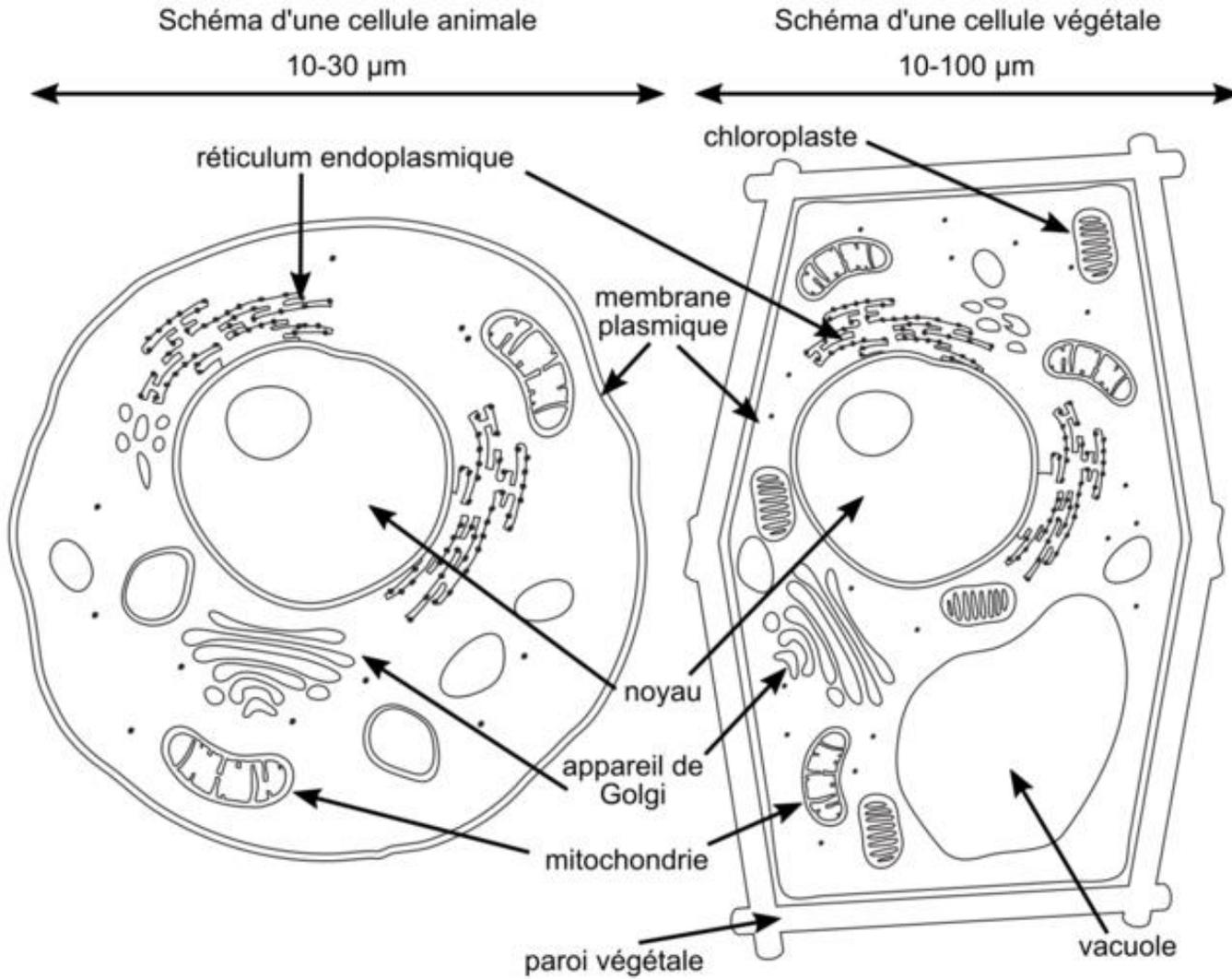
sarah.ben-sadoun@agroparistech.fr

Notion de génome

- Un génome est l'ensemble du matériel génétique, patrimoine héréditaire, d'un individu ou d'une espèce
- Génome présent dans une cellule = **génome nucléaire** (noyau) + **génome des organites** (mitochondries et chloroplastes)



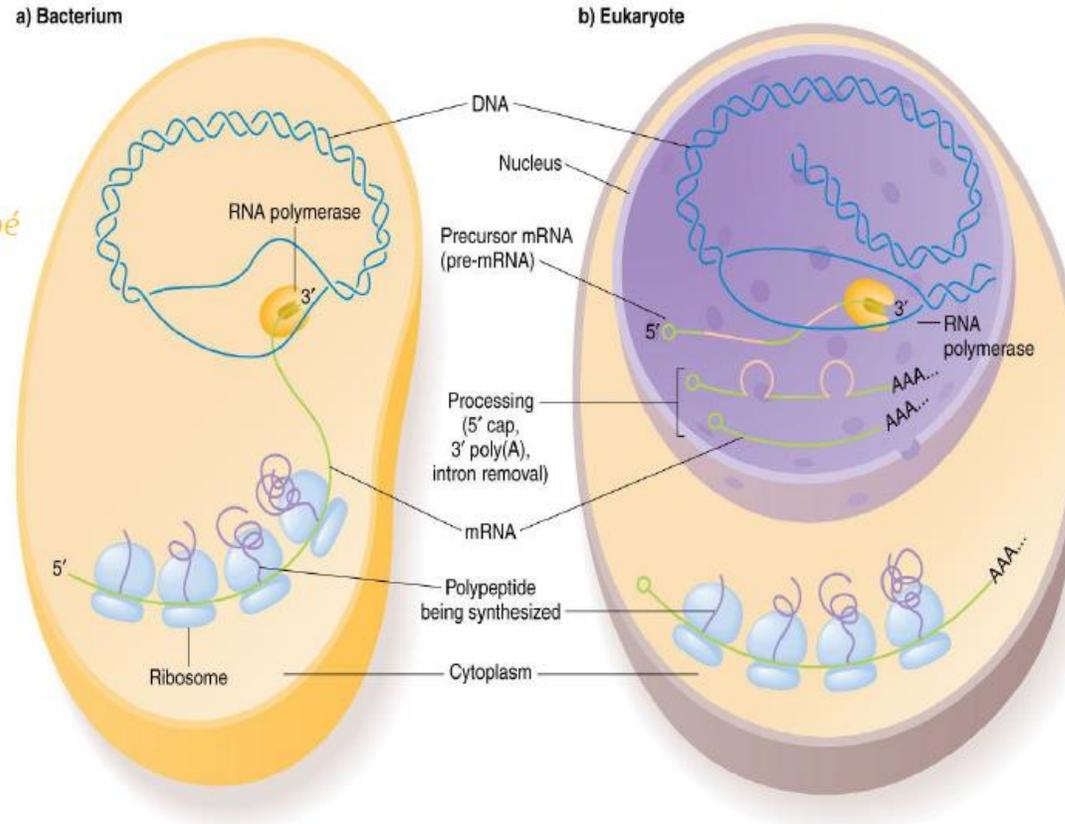
Où se trouve
l'ADN ?



Cellule bactérienne

Cellule animale

Bactéries :
Génome non enveloppé
absence de noyau,
tous les mécanismes
moléculaires se font
dans le cytoplasme



Eucaryotes :
Génome enveloppé
présence d'un noyau et
donc d'une
compartmentation des
processus
(réplication et
transcription dans le
noyau, traduction dans
le cytoplasme).

À noter qu'on a aussi
transcription et
traduction au niveau de
la mitochondrie et des
chloroplastes.

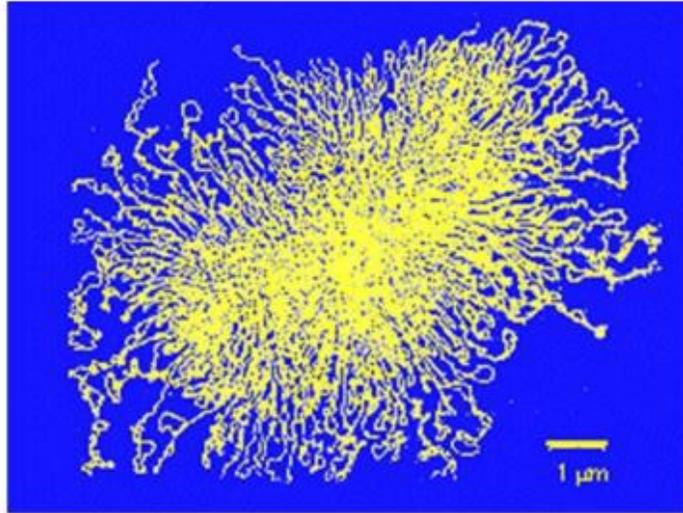
Un génome peut être stocké en
partie sous forme :

- 1) d'un **nucléoïde** = cas des **bactéries**
- 2) ou de **plusieurs chromosomes** : cas
des **eucaryotes**

**TRANSCRIPTION
et
TRADUCTION
COUPLEE**

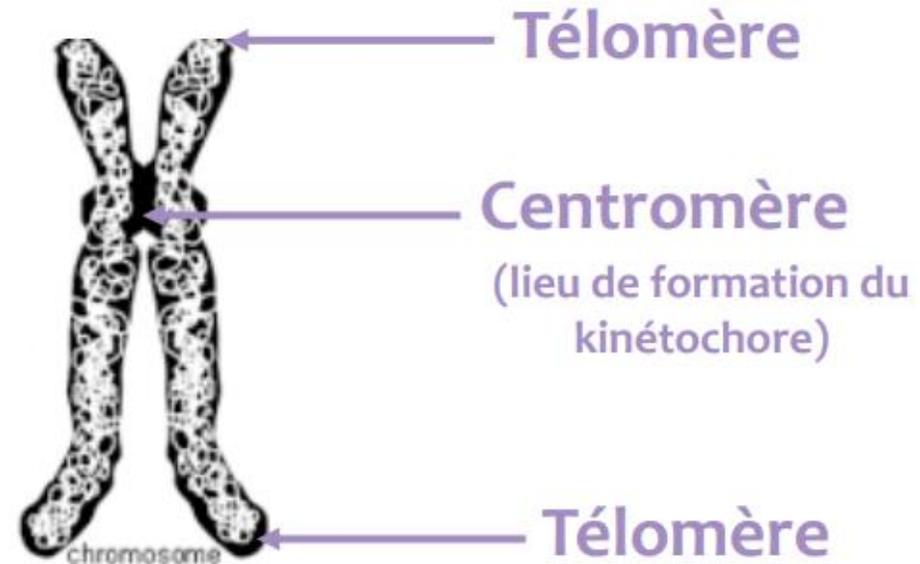
**DECOUPLAGE
TRANSCRIPTION
et TRADUCTION**

Bactéries



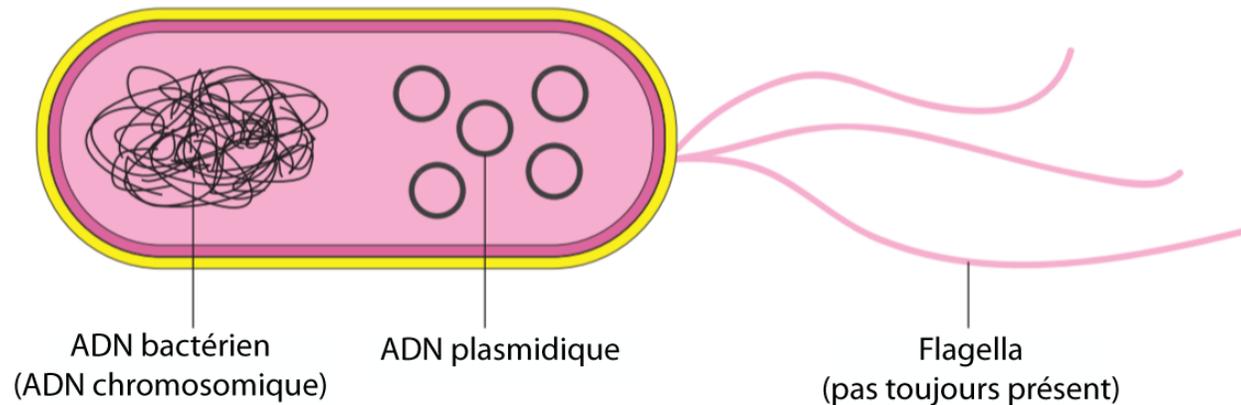
Nucléoïde

Eucaryotes



Chromosome en métaphase

Les plasmides



En plus du nucléoïde, il peut y avoir la présence de **plasmides** dans le cytoplasme.

Les plasmides sont :

- de petites molécules ADN double brins circulaires
- capables de réplication autonome non essentiels à la survie bactérienne.

Les plasmides peuvent porter des gènes de résistance à certains antibiotiques ou des gènes de sélection

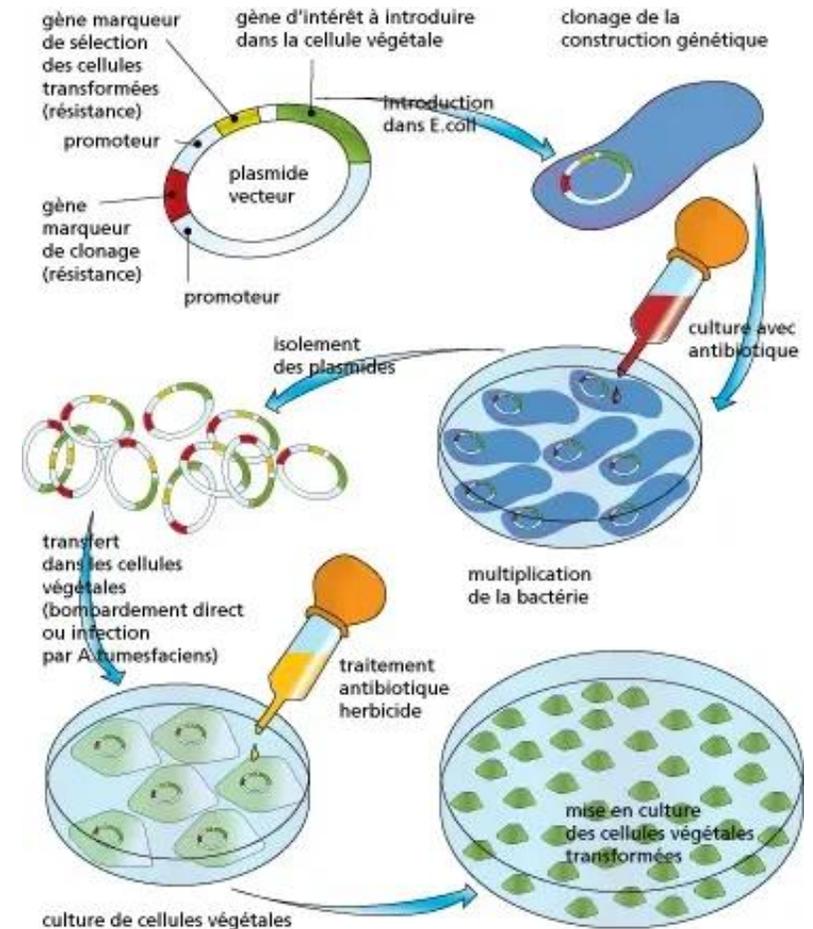
Les plasmides

Les plasmides :

- un outil important pour la manipulation d'ADN
- Possibilité de générer plusieurs centaines de copies de plasmides dans une bactérie

Il est possible d'isoler spécifiquement ces plasmides de l'ADN génomique bactérien

➔ En biotechnologies, on utilise des plasmides fortement modifiés par rapport aux plasmides « natifs »

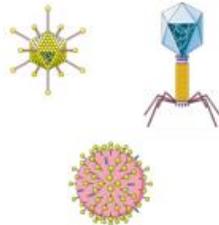


Taille des génomes

Virus

5 000 à 200 000 pb 500 000 à 2 800 000 pb

HIV 9,7 Kb
SARS-CoV2 30,1 Kb



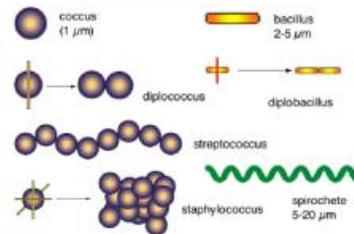
Virus géants
Mimivirus
Megavirus
Pandoravirus
Pithovirus

Procaryotes

1 chromosome
500 000 à 12 000 000 pb

+/- plasmide ~5000 pb

E. coli 4700 Kb



Eucaryotes



S. cerevisiae

14 000 000 pb



C. elegans

100 000 000 pb



D. melanogaster

170 000 000 pb



H. sapiens

3 000 000 000 pb
3 000 000 Kb
3 000 Mb



Plantes

de 50 à 300 000 Mb



Amphibiens

de 700 à 100 000 Mb



Attention aux unités !
Génome en pb – Kb – Mb ...

Taille des génomes

Comment a évolué la taille des génomes ?

- La taille d'un génome se définit comme étant égale au nombre de bases azotées contenues pour une copie d'un génome d'une espèce donnée
- La taille d'un génome chez un individu donné est identique dans tous les types cellulaires diploïdes (cellules du muscle, cellules épithéliales ...)
- L'évolution des tailles des génomes peut être due à plusieurs mécanismes (phénomène de duplication des gènes, phénomène de perte de gènes, ...)

Division du génome d'une espèce donnée en deux éléments génétiques

1) Élément génétique protéinogène =

ce sont les séquences **d'exons** qui codent des protéines.

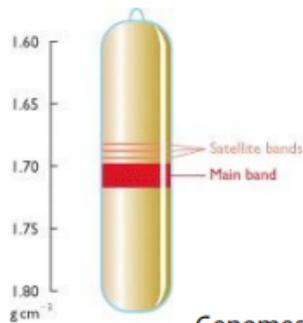
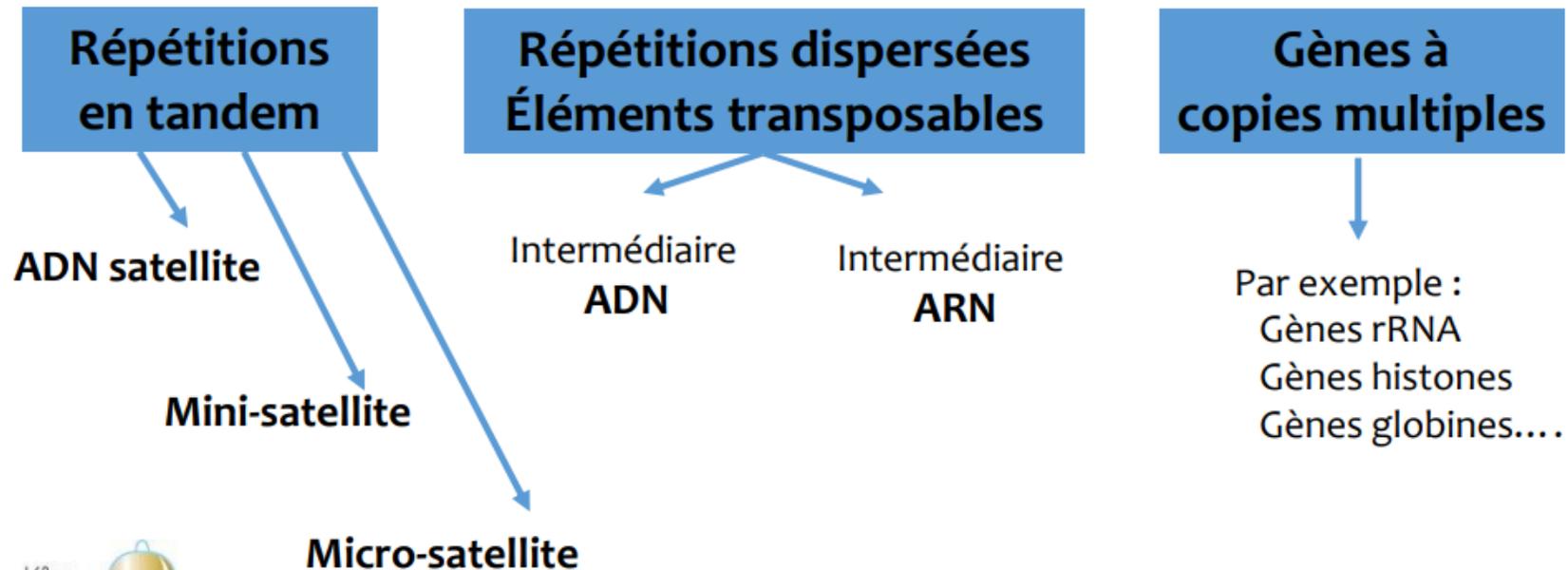
2% de séquences codantes dans le génome humain

2) Élément génétique non protéinogène = regroupe plusieurs types de séquences comme :

- Les séquences **d'introns** (transcrites mais non traduites en protéines)
- Les séquences codant des **ARN** de transfert (ARNt), ARN ribosomiques (ARNr), et d'autres petits ARN
- Les **séquences répétées** dont les éléments **transposables**
- Les **pseudo-gènes....**

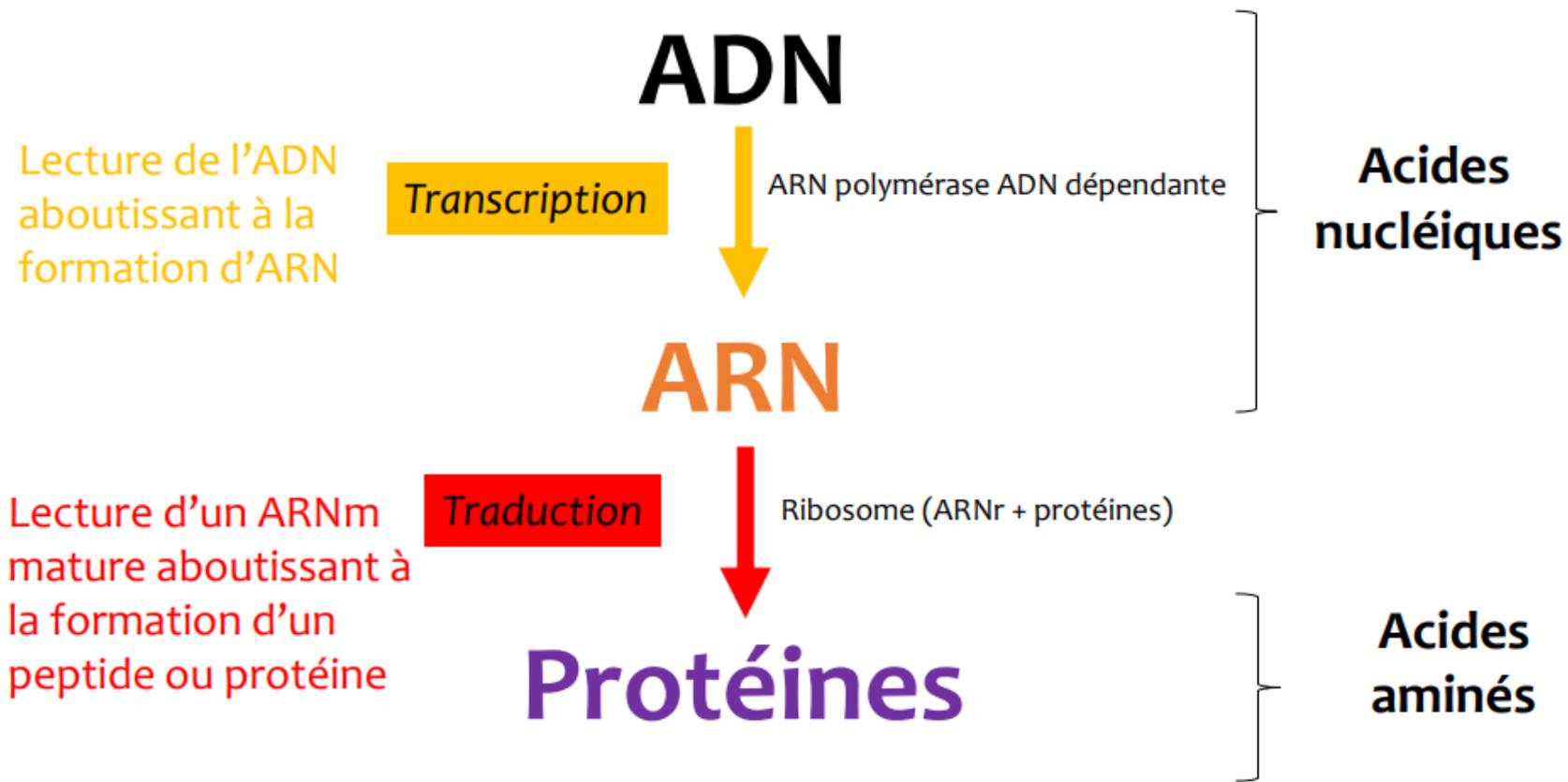
98% de séquences non codantes dans le génome humain

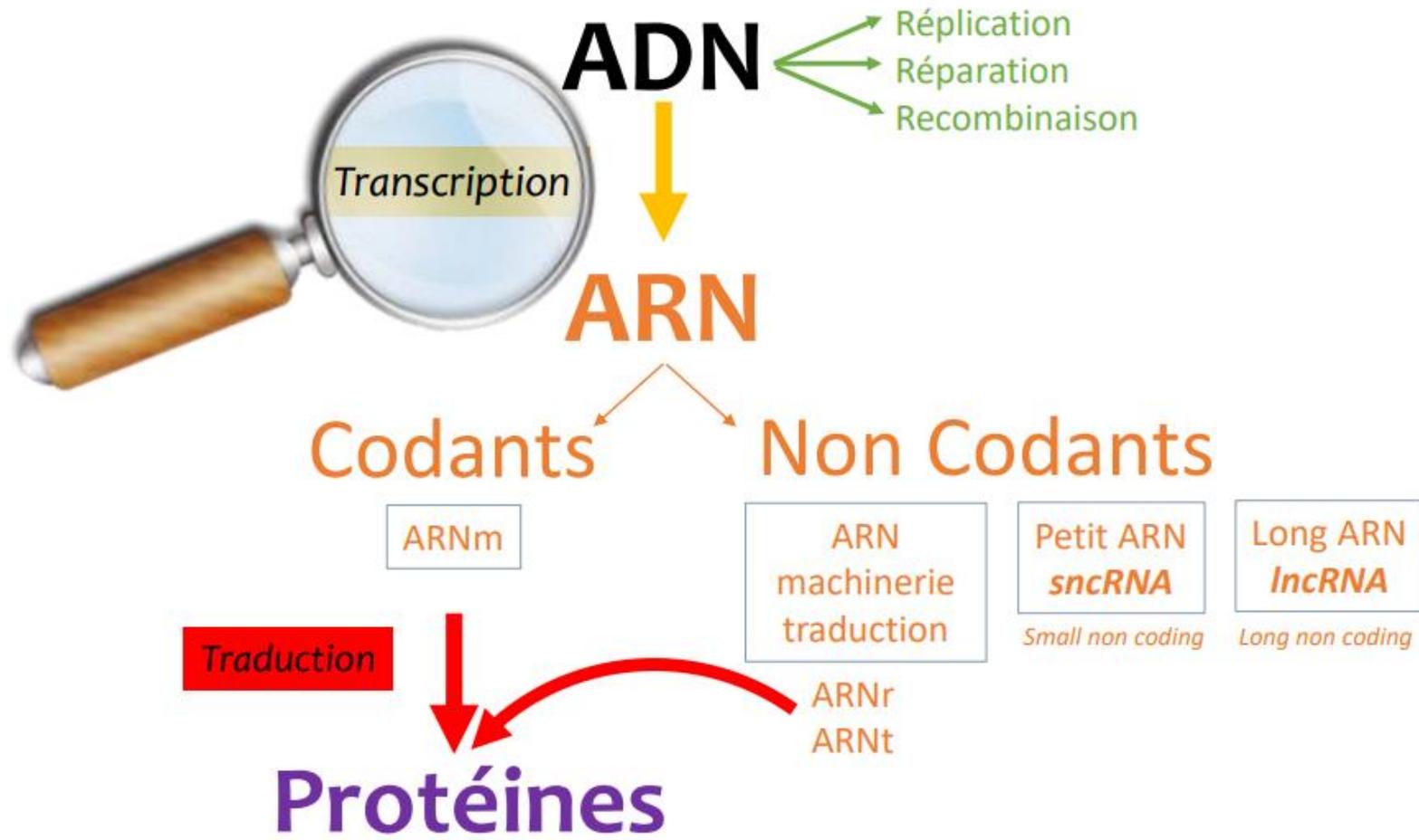
Séquences répétées



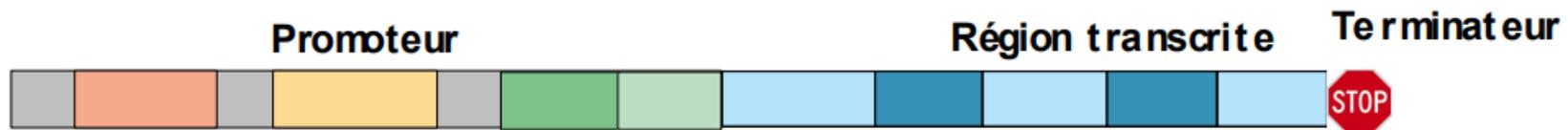
Genomes. 2nd edition. Brown TA

Maryline Moulin 2023





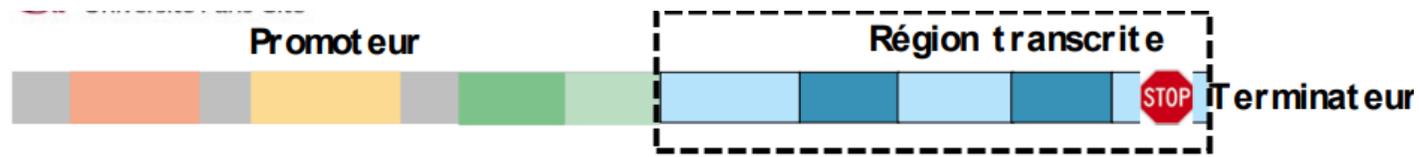
Rappel de la structure du gène



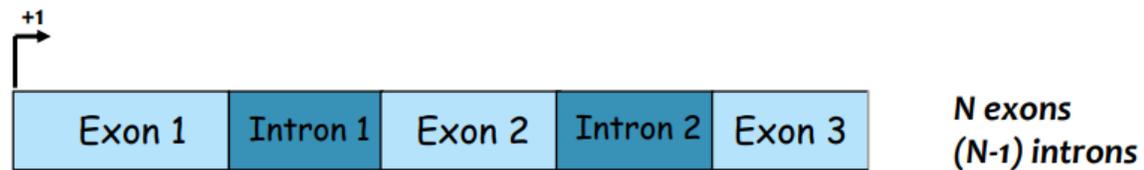
Un gène comporte trois « éléments » essentiels :

- 1- Le **promoteur**, le point de rencontre de tous les facteurs protéiques pour démarrer la transcription et bien positionner l'ARN polymérase
- 2- La **région transcrite**, qui sera la région de lecture pour aboutir à la formation d'ARN
- 3- Le **terminateur** où la transcription s'arrête et l'ARN est libéré

Rappel de la structure du gène



Chez les eucaryotes :



La région transcrite chez les eucaryotes est formée d'une longue séquence de nucléotides qui est divisée en introns et exons.

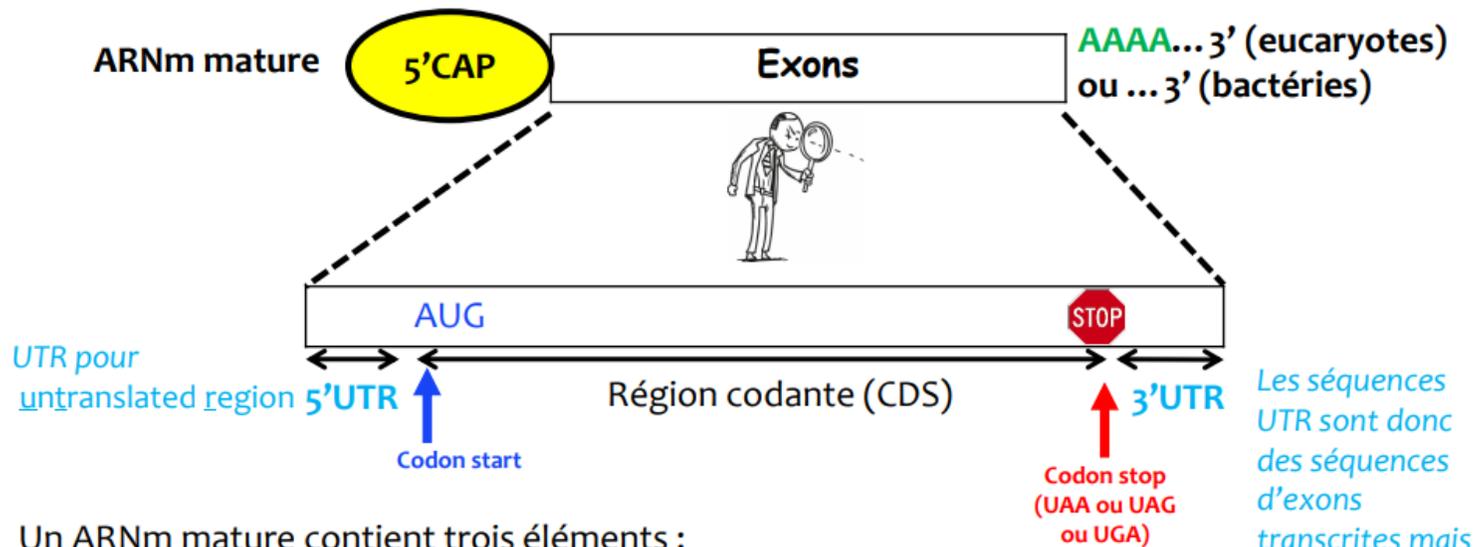
Les gènes sont dits « morcelés » : formé de régions codantes pour des protéines (exons) et de régions non-codantes pour des protéines (introns).

Chez les bactéries :

Dans la **très grande majorité** des cas, la région transcrite chez les bactéries est formée d'une longue séquence de nucléotides qui ne subit pas de maturation.

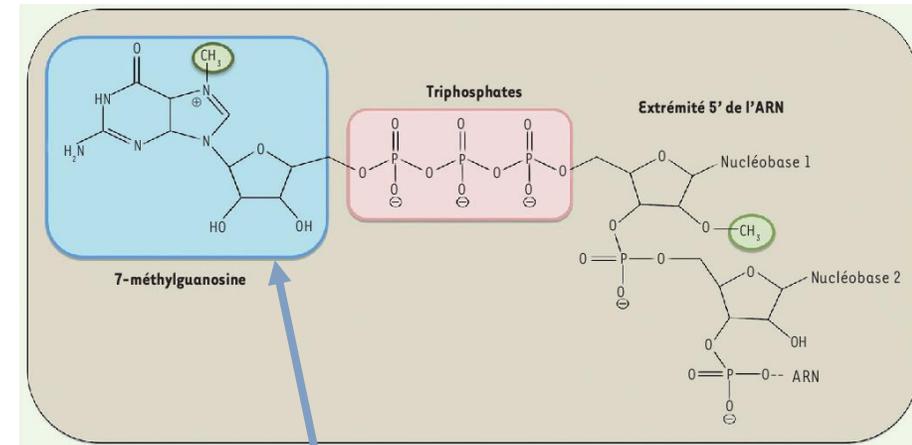
Les gènes bactériens ne sont donc pas « morcelés » !!

Structure globale des ARNm matures



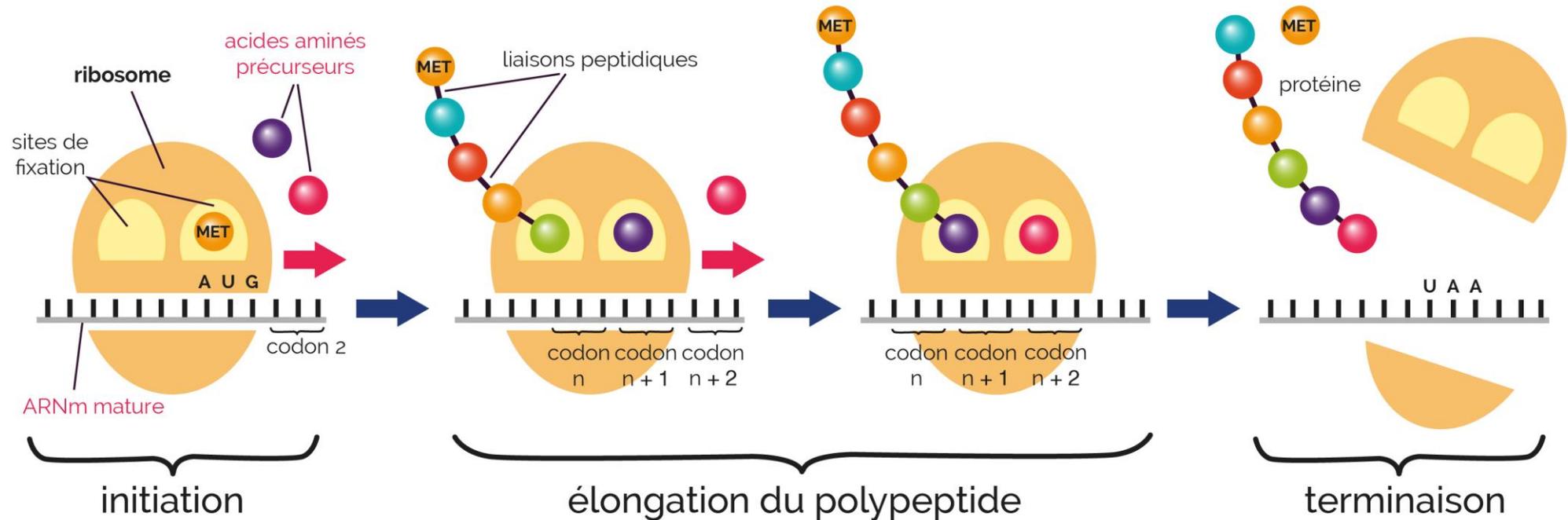
Un ARNm mature contient trois éléments :

- Une **coiffe (ou CAP)** au côté 5'
- Une **queue polyA** au côté 3' (dans le cas eucaryote)
- Une séquence nucléotidique pouvant être divisée en trois parties :
 - 1) Le **5' UTR**, la séquence de nucléotides qui vient juste avant (en 5') le **codon start AUG** de traduction
 - 2) La **région codante (CDS)** qui commence par le **codon start AUG** et finit par le **codon stop de traduction (UAA, UAG ou UGA)**
 - 3) Le **3'UTR**, la séquence de nucléotides qui vient juste après (en 3') le **codon stop**



Coiffe (CAP)

Traduction



Code génétique

		Deuxième nucléotide								
		U		C		A		G		
Premier nucléotide	U	UUU	phényl-alanine	UCU	sérine	UAU	tyrosine	UGU	cystéine	U C A G
		UUC		UCC		UAC		UGC		
		UUA	leucine	UCA		UAA	STOP	UGA	STOP	
		UUG		UCG		UAG		UGG		
	C	CUU	leucine	CCU	proline	CAU	histidine	CGU	arginine	U C A G
		CUC		CCC		CAC		CGC		
		CUA		CCA		CAA	glutamine	CGA		
		CUG		CCG		CAG		CGG		
	A	AUU	isoleucine	ACU	thréonine	AAU	asparagine	AGU	sérine	U C A G
		AUC		ACC		AAC		AGC		
		AUA		ACA		AAA	lysine	AGA	arginine	
		AUG	méthionine	ACG		AAG		AGG		
	G	GUU	valine	GCU	alanine	GAU	acide aspartique	GGU	glycine	U C A G
		GUC		GCC		GAC		GGC		
		GUA		GCA		GAA	acide glutamique	GGA		
		GUG		GCG		GAG		GGG		

Troisième nucléotide