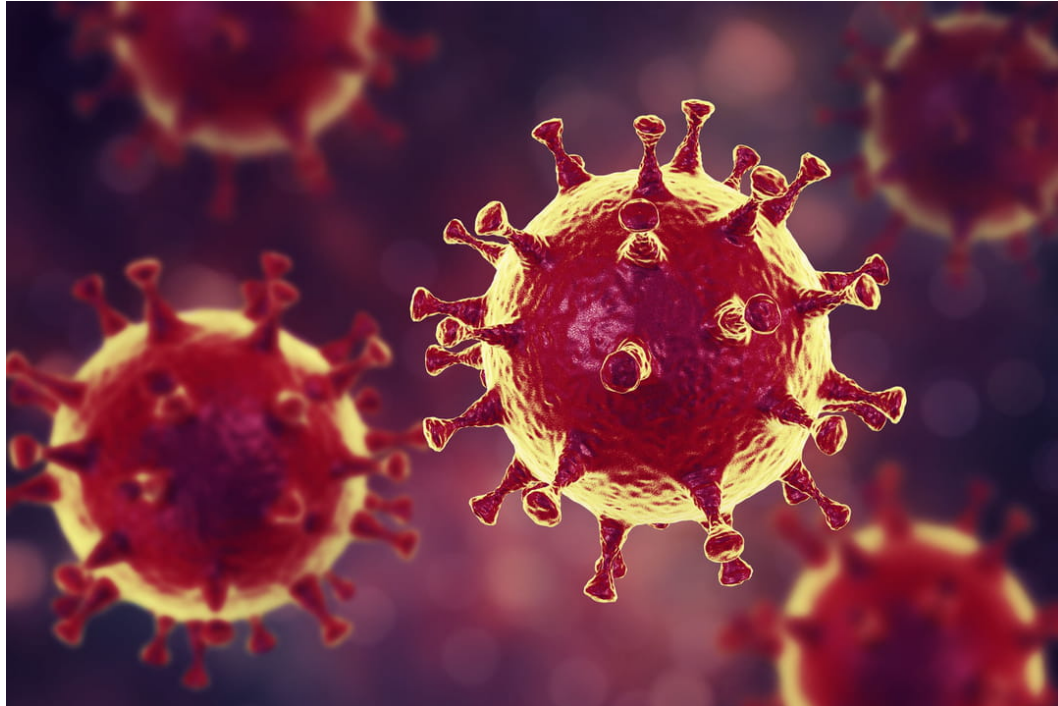


Chapitre 1 – L'organisation des génomes



Source : 123RF-Kateryna Kon

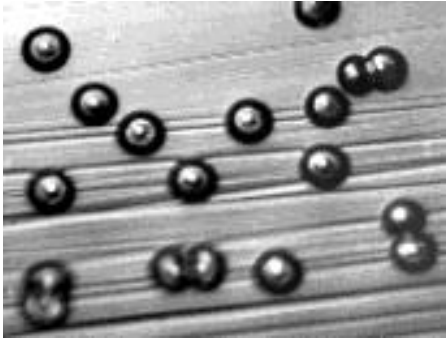
1. L'ADN, molécule universelle portant l'information génétique

... ou presque....

1.1. le support de l'information génétique

Deux souches bactériennes de Pneumocoques

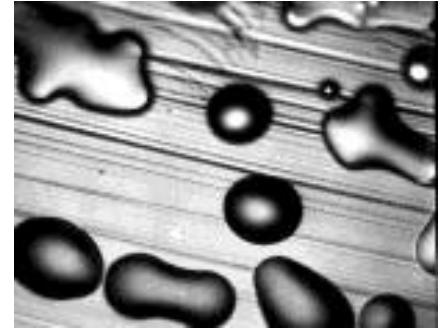
Bactéries *Streptococcus pneumoniae* (= pneumocoques)



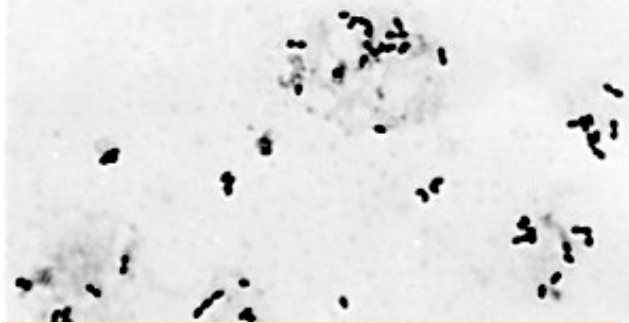
colonies rugueuses (rough R)

Aspect des colonies
sur boîte de Pétri

x 10



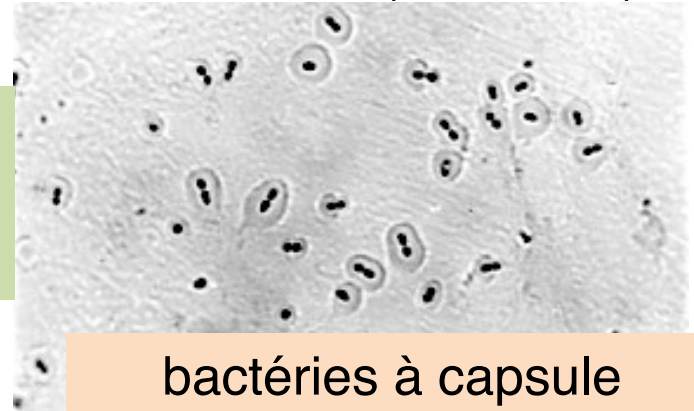
colonies lisses (smooth S)



bactéries sans capsule

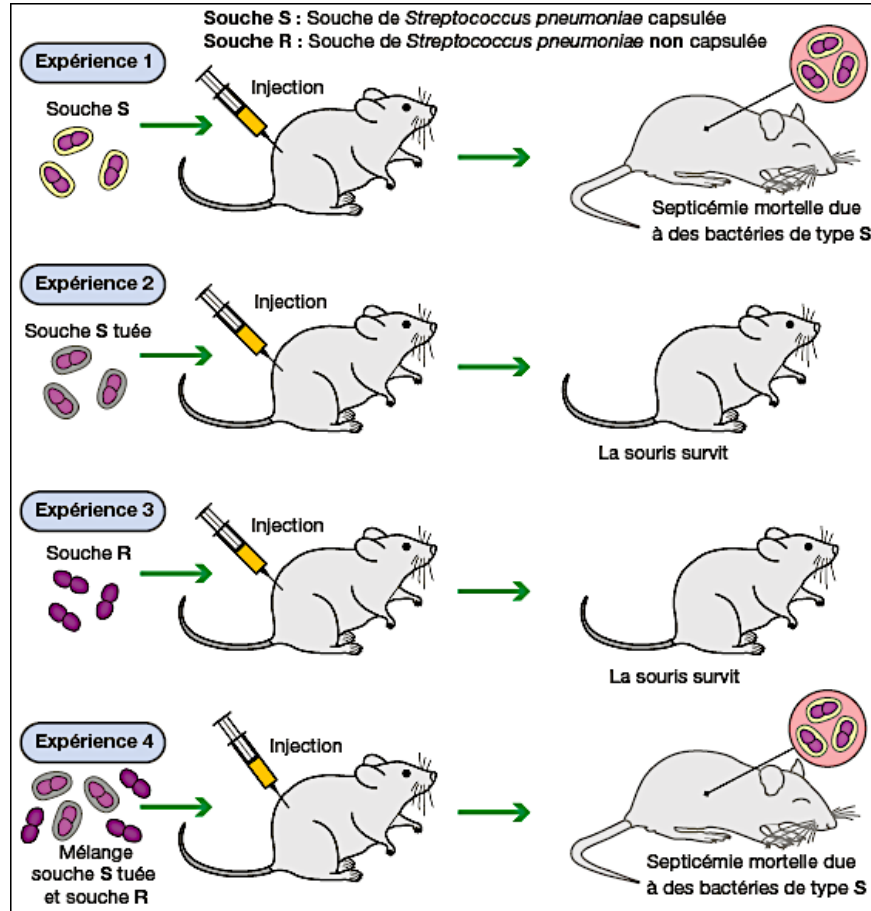
Aspect des
bactéries au
microscope

x 2 000



bactéries à capsule

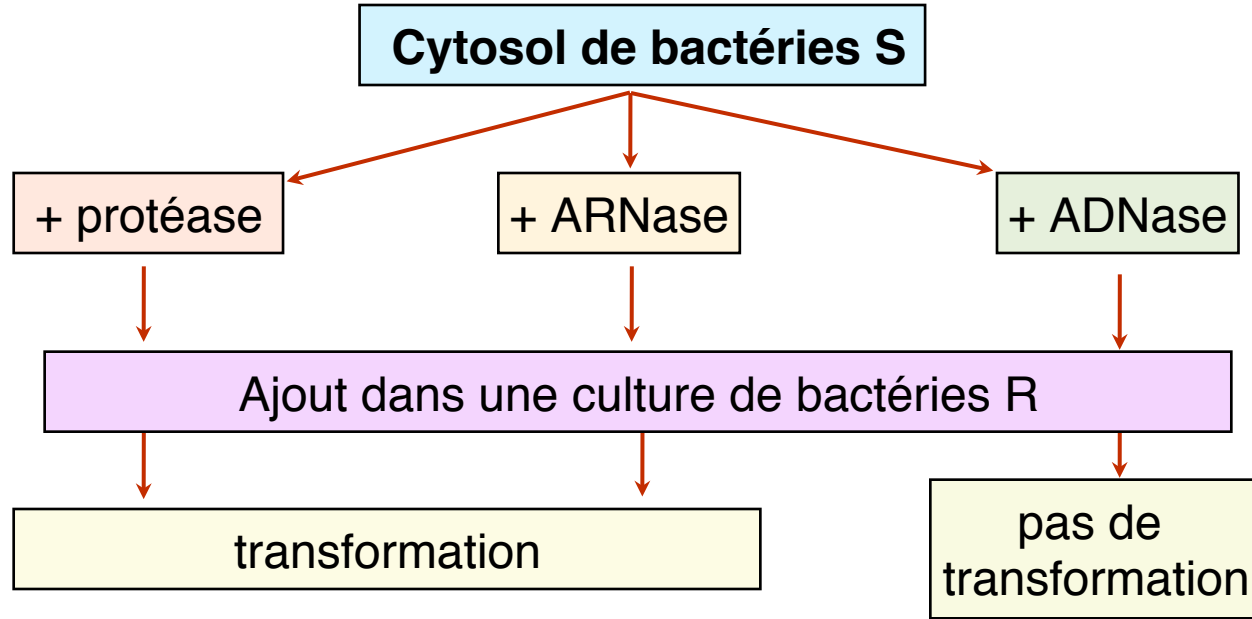
Expérience de Griffith (1928)



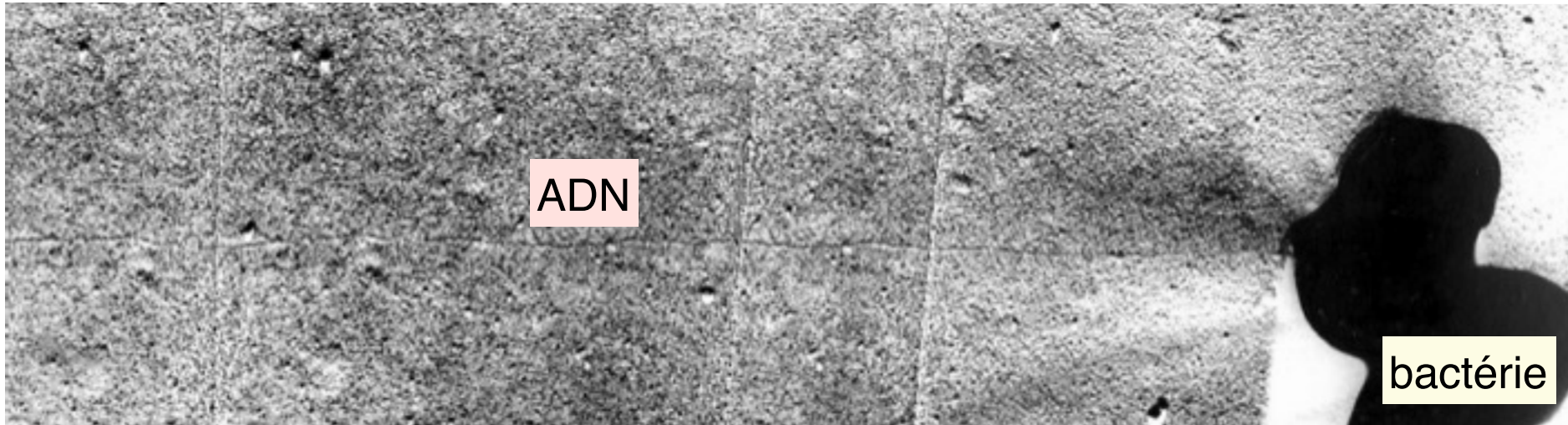
Notion de principe transformant

Expérience d'Avery, MacLeod et McCarty (1944)

5



La transformation bactérienne



This electron micrograph (courtesy of Dr. Alexander Tomasz) shows a DNA molecule entering a *Pneumococcus*. This DNA molecule — only a portion of which (scroll down) is shown here — is approximately 7 micrometers (μm) in length, long enough to include a dozen genes. The process of transformation follows the uptake of such a molecule by the bacterium.

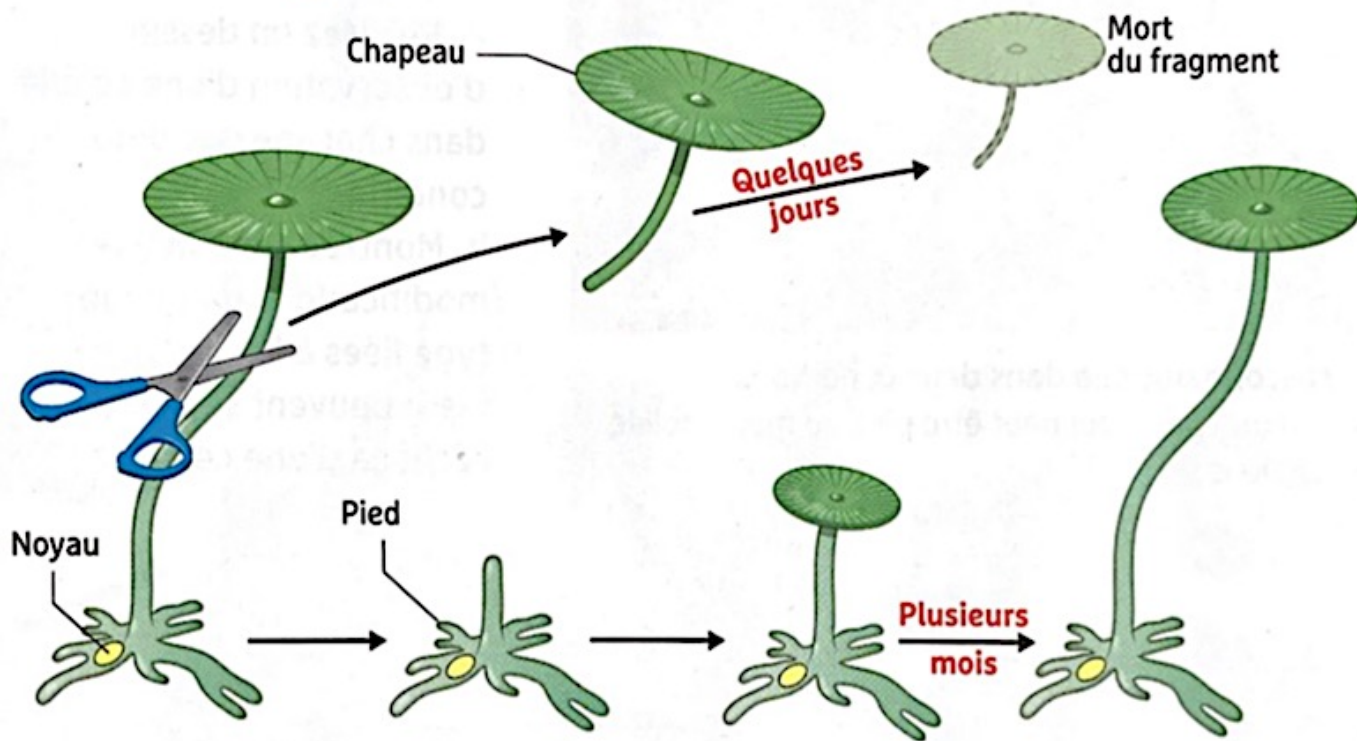
Les Acétabulaires, des cellules géantes



Polymorphisme

Le rôle du noyau

Avec ou sans chapeau...

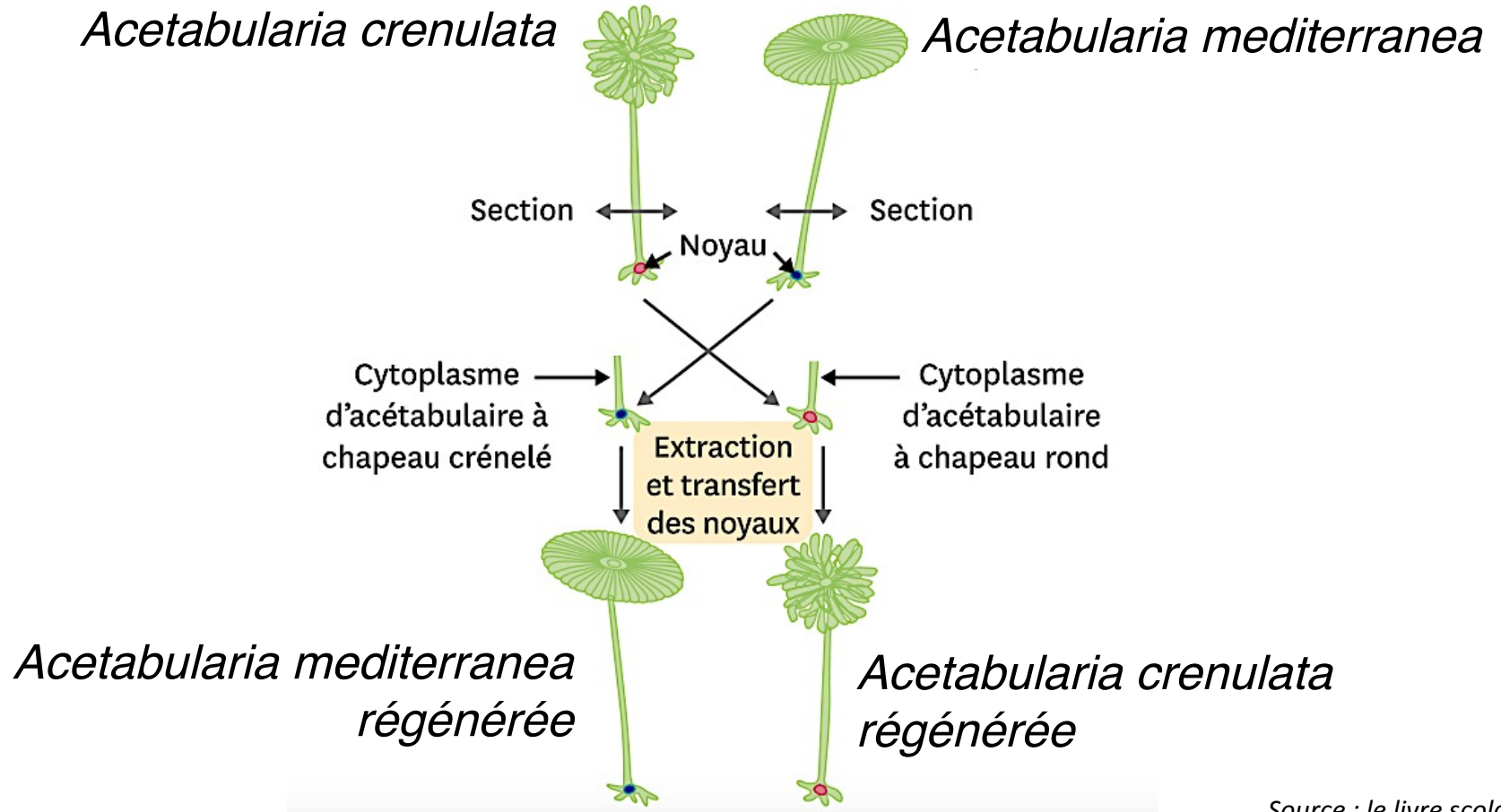


Une expérience de section réalisée sur une acétabulaire.

Greffes de noyaux

Acetabularia crenulata

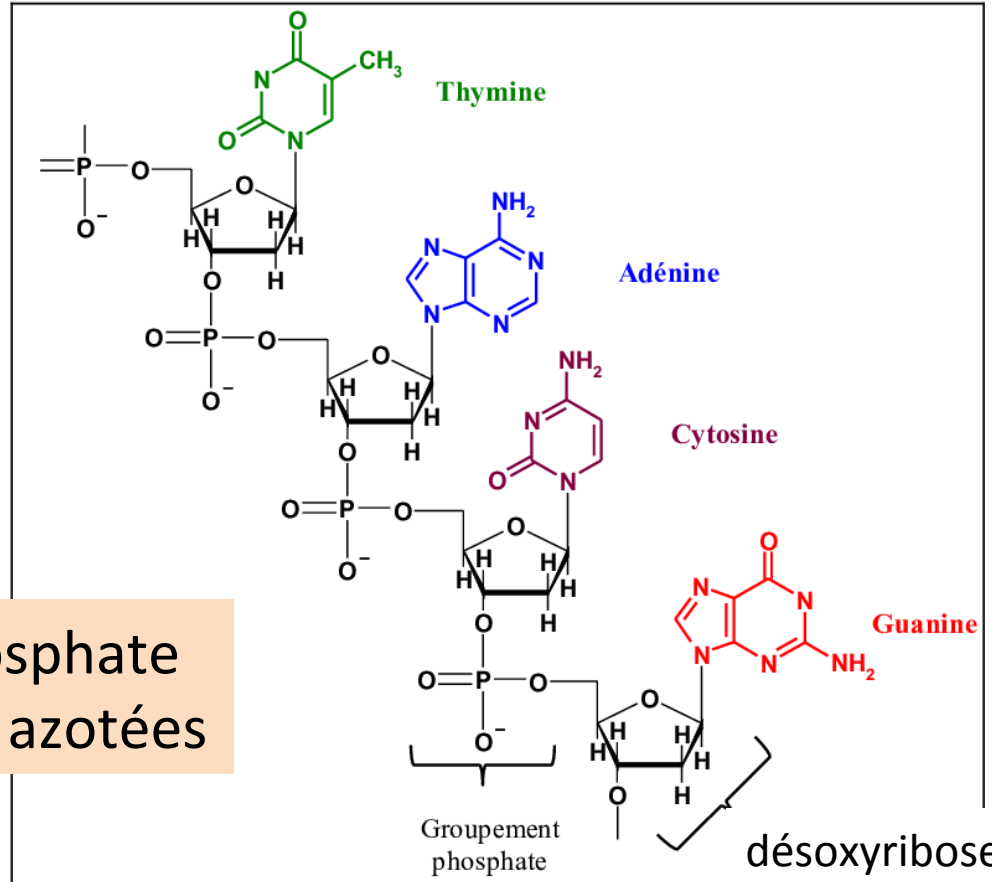
Acetabularia mediterranea



1. L'ADN, molécule universelle portant l'information génétique

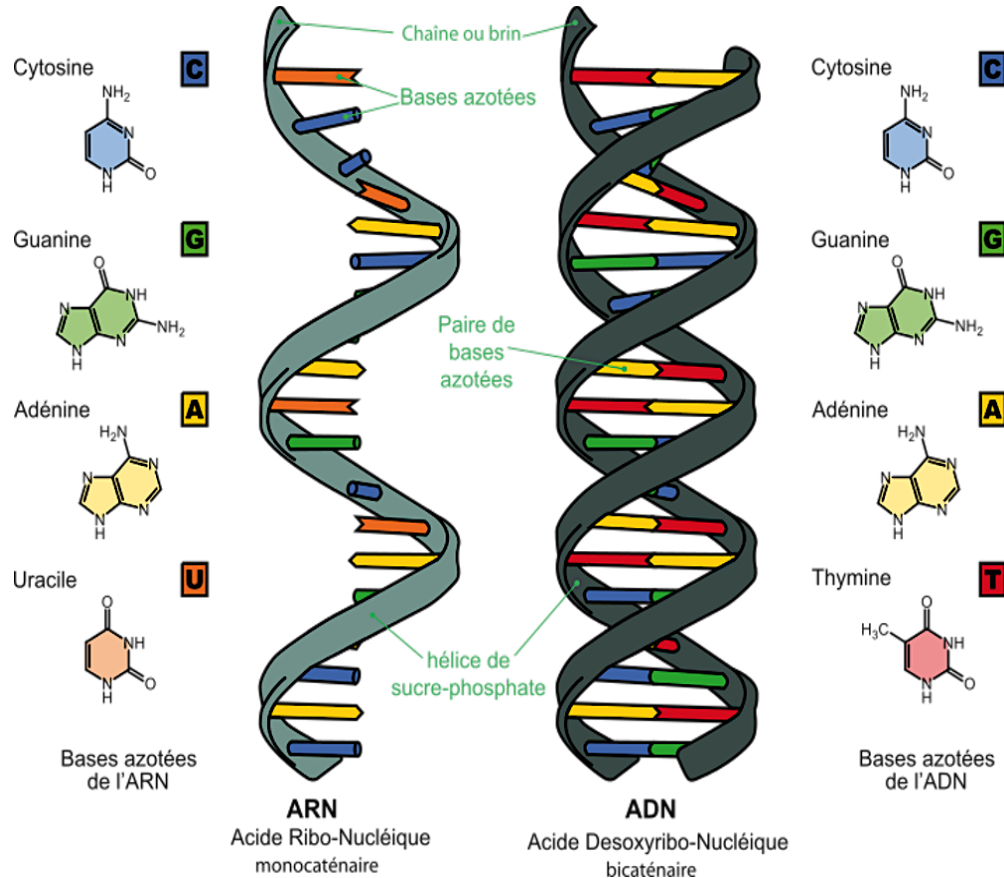
1.2. une succession ordonnée de nucléotides

Configuration d'un brin d'ADN

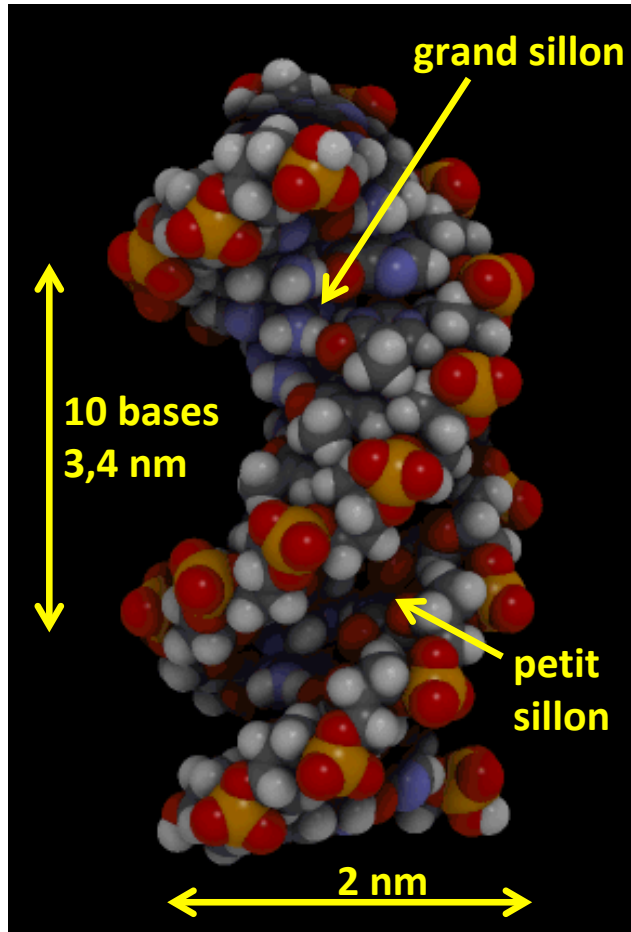


Un axe sucre-phosphate portant les bases azotées

ARN et ADN, deux hélices



L'hélice B de l'ADN



Structure condensée : grande capacité de stockage dans un faible volume

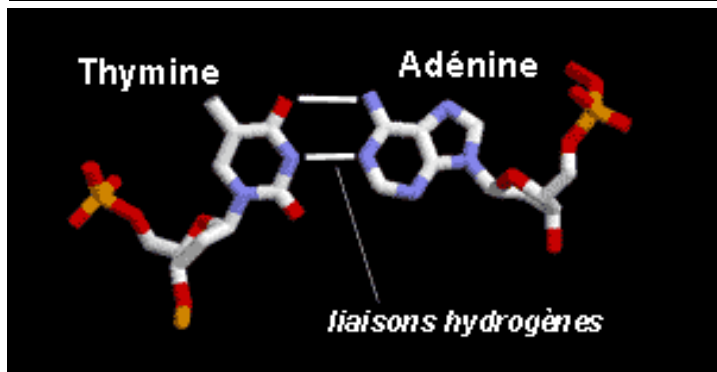
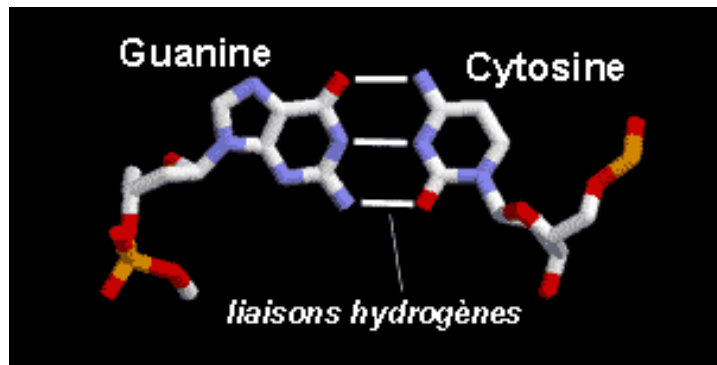
Bases protégées entre les 2 brins

Bases stabilisées par des liaisons H qui évitent la tautomérie

Structure stable grâce aux interactions (stacking, liaisons H, ioniques)

Accessibilité possible grâce aux sillons

L'association des bases azotées



$$\frac{A + G}{T + C} = 1 \quad \frac{A}{T} = \frac{G}{C} = 1$$

règle de Chargaff

U est à la place de T dans l'ARN

La séquence constitue le message

Ocytocine

Brin non transcrit : TGCTACATCCAGAACTGCCCCCTGGGC
Brin transcrit : ACGATGTAGGTCTTGACGGGGGACCCG

ADH

Brin non transcrit : TGCTACTTCCTGAACTGCCCCAAGAGGA
Brin transcrit : ACGATGAAGGACTTGACGGGTTCCTCCT

Séquence protéique correspondante

Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly

Cys-Tyr-Phe-Leu-Asn-Cys-Pro-Arg-Gly

Rôles

Ocytocine = hormone peptidique impliquée dans le contrôle de la parturition et la lactation.

ADH = hormone anti-diurétique dont la cible est le rein.

Le séquençage : méthode de Sanger

Principe : une copie de l'ADN à séquencer est réalisée en présence d'un di-désoxynucléotide phosphate qui bloque la polymérisation

ADN à séquencer (brin transcrit)



↑
synthèse du brin
d'ADN
complémentaire

dATP
 dGTP dCTP
 ddGTP (un peu) dTTP

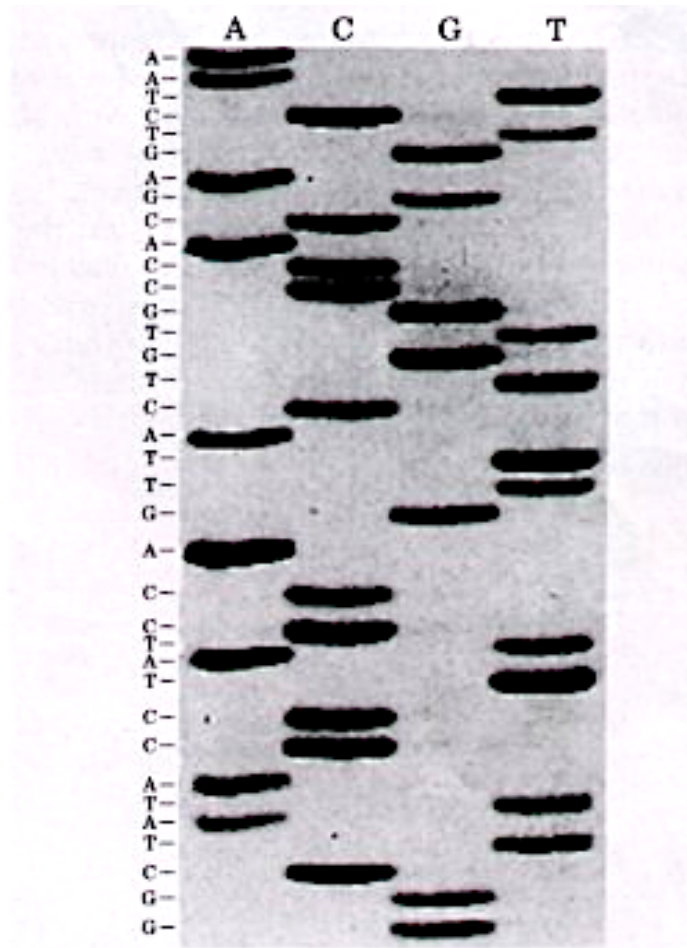
arrêt de la synthèse, par
utilisation d'un ddGTP

G utilisation d'un dGTP :
 la synthèse continue

Conclusion

de manière aléatoire, obtention d'un ensemble de fragments (de différentes tailles), arrêtés au niveau des Cytosines du brin transcrit de l'ADN (donc des Guanines du brin codant).

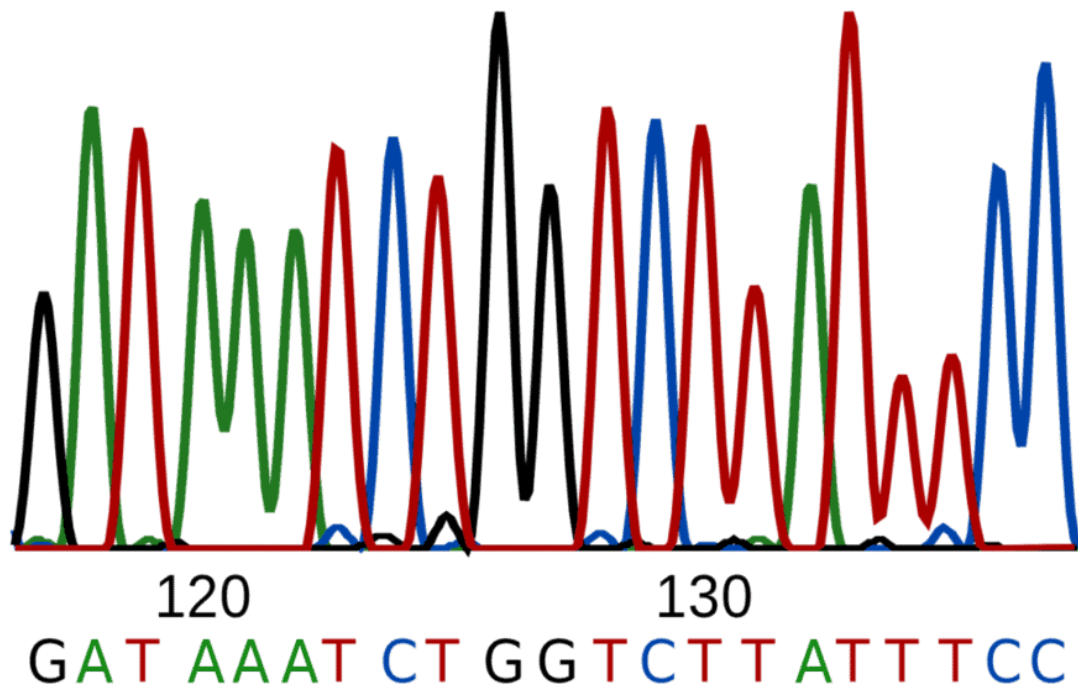
Lecture de la séquence après électrophorèse



Chaque piste contient tous les fragments qui se terminent par un A, un C, un G ou un T.

Automatisation du séquençage

1. copie de l'ADN en présence des ddNTP couplés à un fluorochrome
2. séparation des brins synthétisés par chromatographie
3. lecture de l'ADN marqué par fluorescence

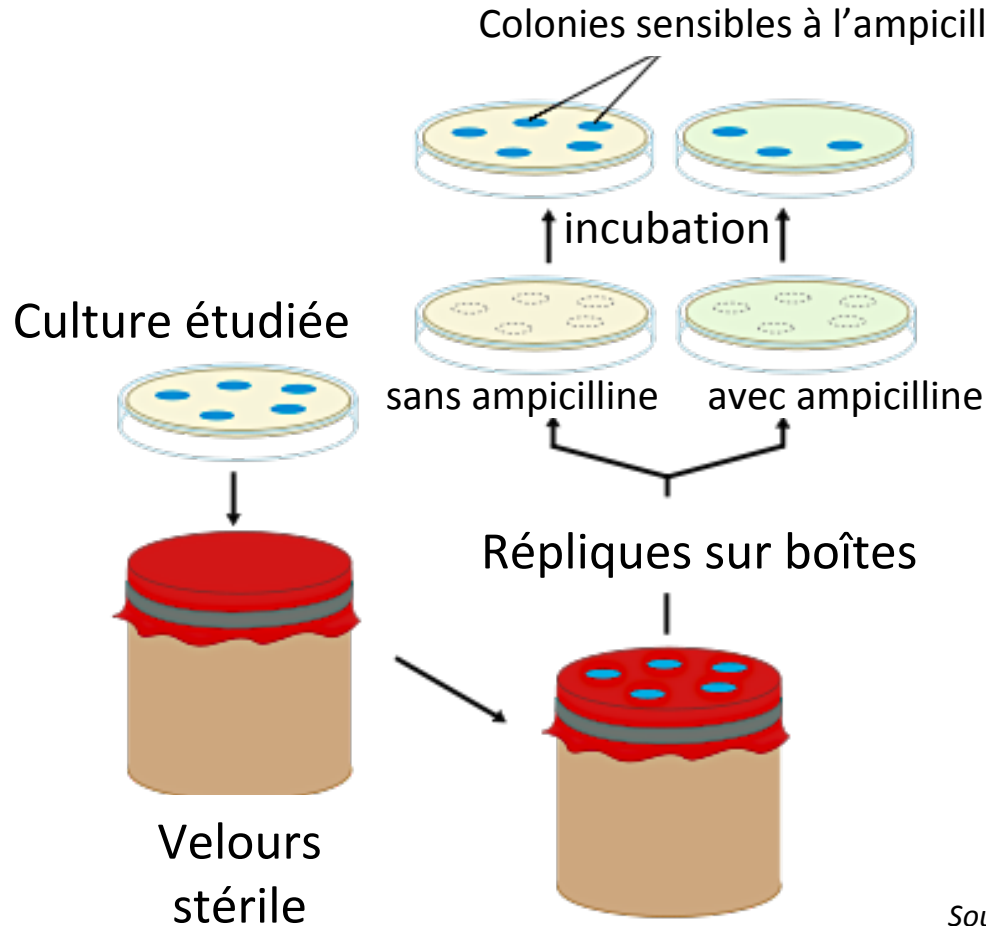


Code des fluorochromes
A T C G

1. L'ADN, molécule universelle portant l'information génétique

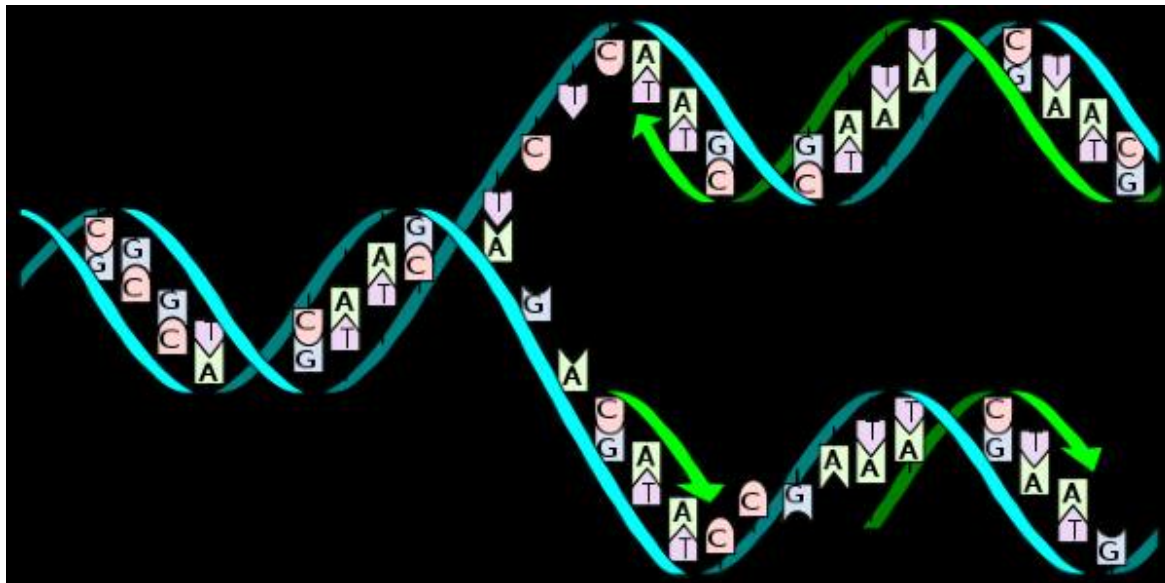
1.3. un traitement universel : réplication et transcription

La copie de l'ADN à l'identique ?



Les bactéries se divisent et leurs descendants sont identiques. Les mutations sont des erreurs de copie : le taux de mutations est très faible : fréquence de 10^{-6} à 10^{-10} .

Un modèle universel de réplication



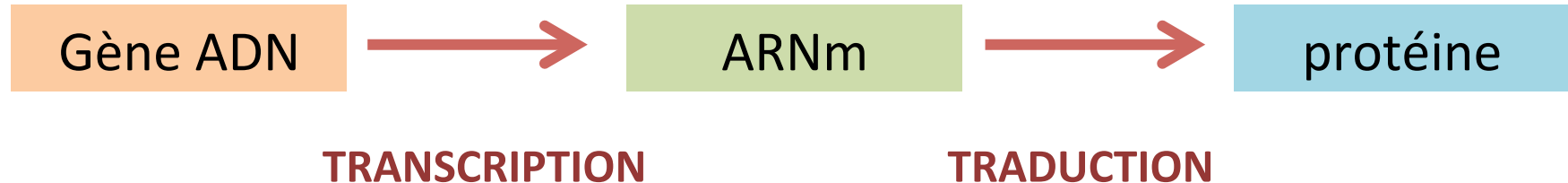
Le phénotype est lié à l'expression du génome

Attention : mécanisme très simplifié ici

La plupart des caractères est due à la présence de protéines :

- protéines d'architecture,
- enzymes permettant la réalisation d'une réaction du vivant...

Une protéine est synthétisée à partir de l'ADN.



La traduction repose sur le code génétique

		Deuxième lettre								
		U		C		A		G		
Première lettre (côté 5')	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
		UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop	A
		UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp	G
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
		CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
		AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
		AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
		AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
		GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
		GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
		GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G
		codon d'initiation				codon de terminaison				
										Troisième lettre (côté 3')

Troisième lettre (côté 3')

Notion de séquence codante

= séquence *unique* de l'ADN traduite en acides aminés



Cys – Tyr – Ile – Gln – Asn – Cys – Pro – Leu – Gly

Séquence protéique de l'ocytocine

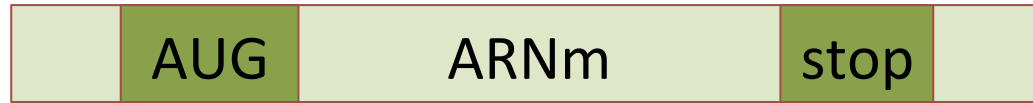
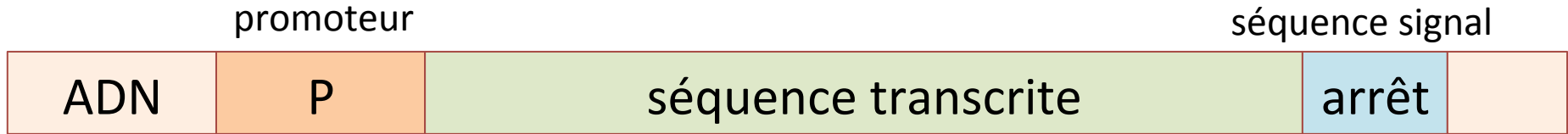
UGCUACAUCCAGAACUGCCCCCUGGGGC

Séquence d'ARNm traduite

TGCTACATCCAGAACTGCCCCCTGGGC

Séquence d'ADN codant correspondant à l'ocytocine

La structure de base d'un gène



**ATTENTION
TRÈS SIMPLIFIÉ**



Séquence codante = seulement ce qui est traduit

Quelques définitions

- **Génome** = ensemble de l'ADN d'une cellule vivante
- **Transcriptome** = ensemble des ARN contenus dans une cellule vivante
- **Protéome** = ensemble des protéines d'une cellule vivante

Les séquences codantes

Cas des humains

- D'après le **protéome**, on dénombre environ 21 000 chaînes d'acides aminés issues de 21 000 séquences d'ADN donc 21 000 séquences codantes.

Soit 1,5 % du génome !!!

- D'après le **transcriptome**, on dénombre environ entre 20 000 et 40 000 ARN différents, non traduits, correspondants à des ARN outils.

La part de l'ADN codant

	E.coli	Levure	Arabette	Homme
taille du génome en 10^6 pb	4,6	12,1	135	3 200
nombre de gènes <i>(avec ARN outils)</i>	4 377 <i>(4 600)</i>	5 800 <i>(6 000)</i>	27 000 <i>(36 000)</i>	21 000 <i>(60 000)</i>
% de fraction codante <i>(avec les introns)</i>	90 %	72 % <i>(75 %)</i>	29 % <i>(40 %)</i>	1,4 % <i>(25 %)</i>

Chiffres données pour un génome haploïde eucaryote

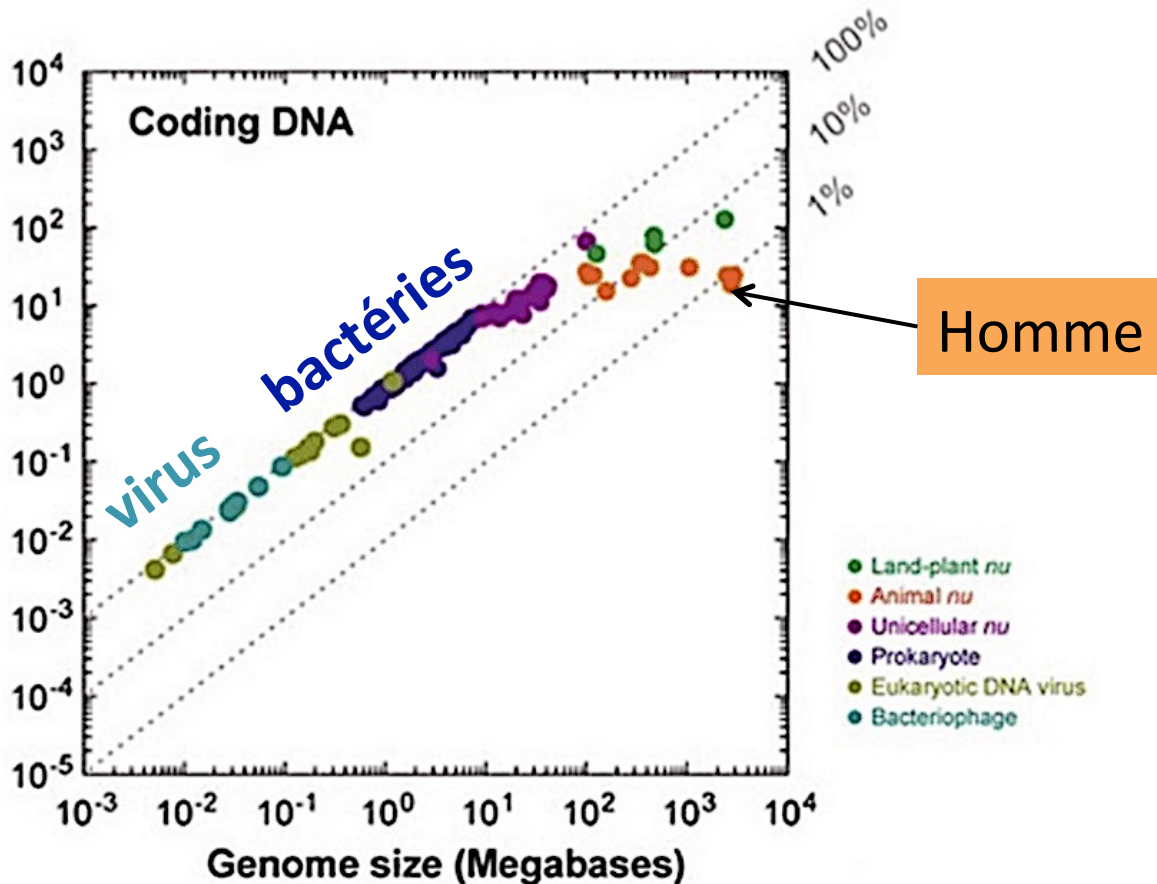
Taille du génome versus longueur d'ADN

Une cellule humaine possède un génome de $3,2 \cdot 10^9$ pb mais elle a en fait $6,4 \cdot 10^9$ pb en phase G1 puisqu'elle est diploïde.

10 pb constituent un tour d'hélice d'ADN soit 3,4 nm donc la longueur totale de l'ADN nucléaire d'une cellule humaine en phase G1 est :

$$6,4 \cdot 10^9 \times 0,34 = 2,18 \text{ m}$$

La part de l'ADN codant

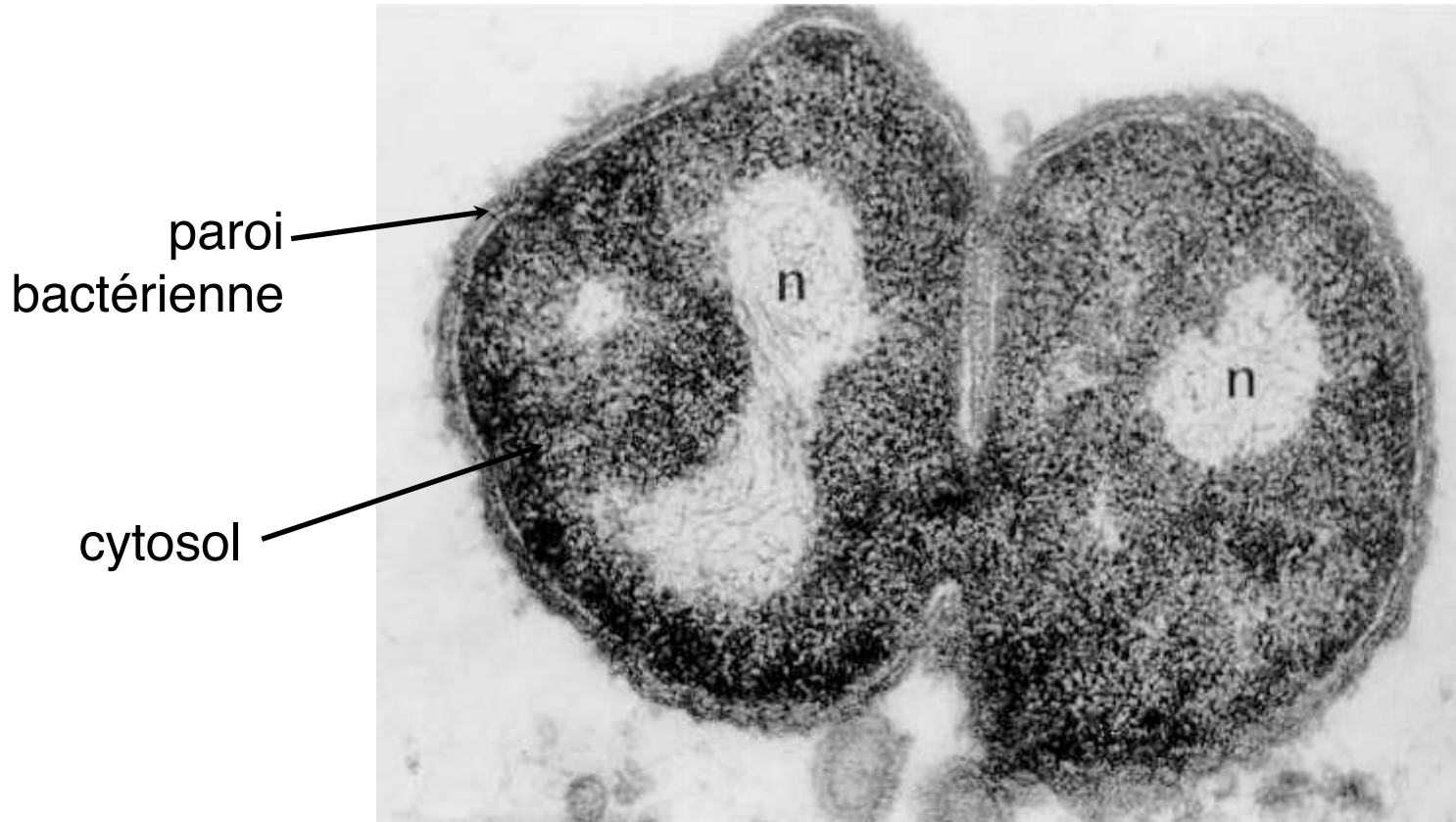


- des introns de plus en plus grands
- des séquences transcrites de plus en plus nombreuses

2. L'organisation des génomes et leur diversité

2.1. Le génome bactérien

Le nucléoïde, région contenant l'ADN



n : nucléoïde

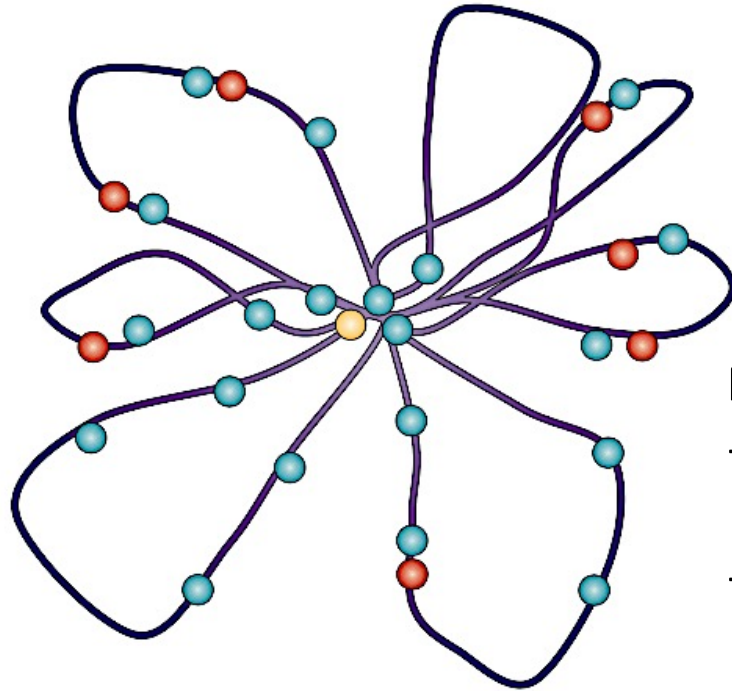
Ici 2 bactéries se séparent après division.

Le chromosome bactérien



*Observation au MET du chromosome bactérien
étalé après éclatement du colibacille*

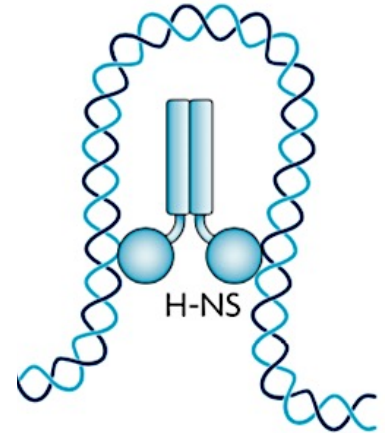
Le chromosome en boucles



Chromosome de bactérie en phase quiescente

Protéines de structure

- protéines basiques H-NS stabilisatrices
- protéines FIS induisant des supertours



● RNA polymerase
at RNA promoters

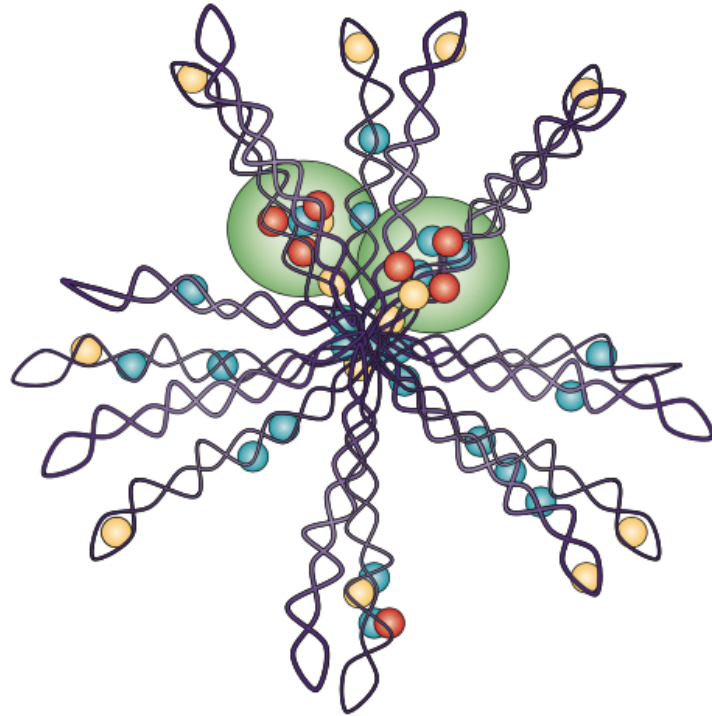
● H-NS

● Transcription
factories

● Fis

Ici : 7 boucles = 7 opérons

Le chromosome en boucles



Chromosome de bactérie en phase de croissance

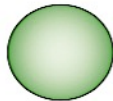
Nombreux supertours



RNA polymerase
at RNA promoters



H-NS



Transcription
factories



Fis

BILAN

Le chromosome bactérien est :

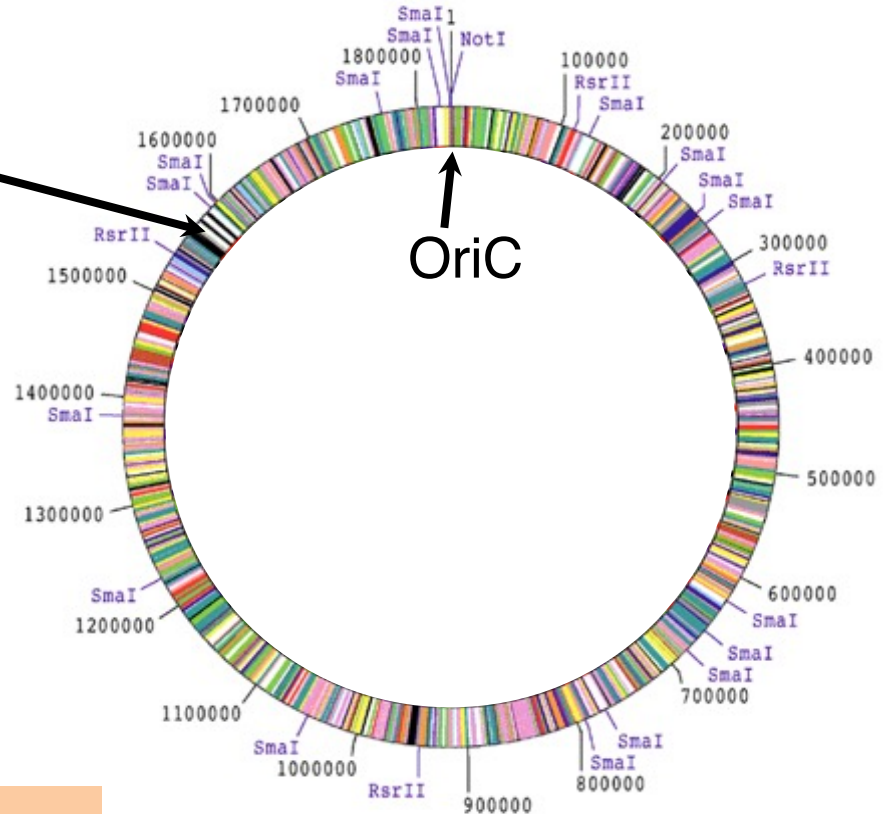
- circulaire avec de grandes boucles
- lié à des protéines de structure chargées +
- à structure dynamique
- entouré d'enzymes et de facteurs d'expression

Escherichia coli : génome de $4,6 \cdot 10^6$ pb soit 1,56 mm de long

L'ADN est majoritairement codant

séquences codantes
(couleurs selon la fonction)

Génome de
Hemophilus influenzae



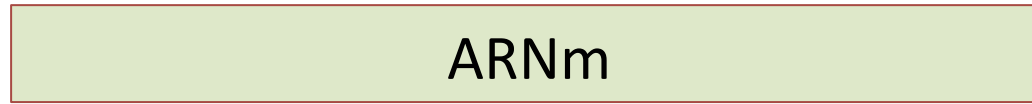
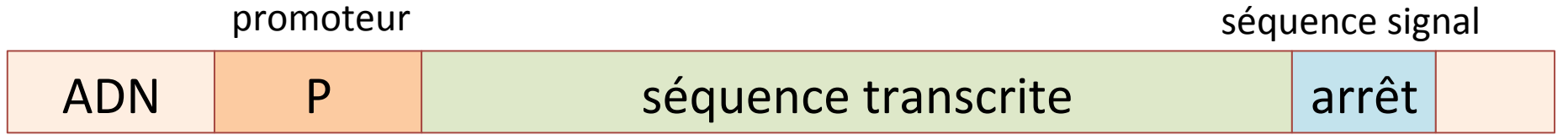
ADN codant à 90 % pour E. coli

Les séquences non codantes

- Une séquence OriC = origine de la réplication
- Séquences de contrôle de l'expression génétique
 - séquence de régulation
 - séquence promoteur
 - séquence de terminaison

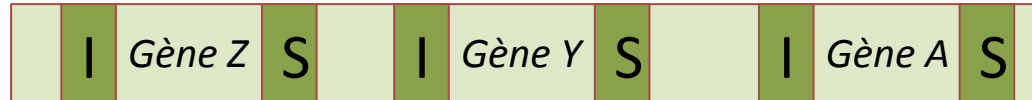
Les opérons, une organisation bactérienne

Exemple de l'opéron lactose



1 ARNm

I Codon de début de traduction



S Codon de fin de traduction

LacZ

LacY

LacA

3 protéines

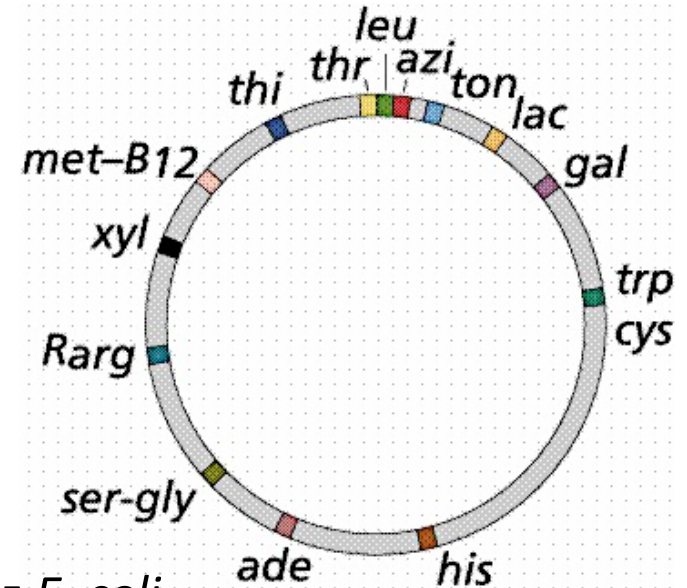
Les opérons

Opéron = groupe de gènes contrôlés par un même promoteur et transcrits en un même ARNm.

Génome d'*Escherichia coli*

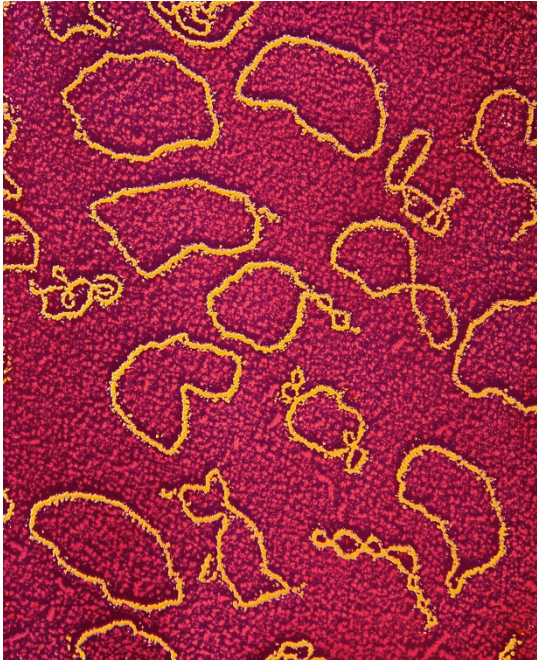
$4,6 \cdot 10^6$ pb soit 1,56 mm

4 377 gènes groupés en 40 opérons



Carte de 15 des 40 opérons décrits chez *E. coli*

Les plasmides



x 22 000

MET : plasmides isolés d'*E. coli*

Plasmide = petit ADN circulaire

10^3 à 10^5 pb

2 à 30 gènes

séquence Ori permettant sa duplication

Non indispensable à la vie bactérienne mais apporte des propriétés supplémentaires à la bactérie.

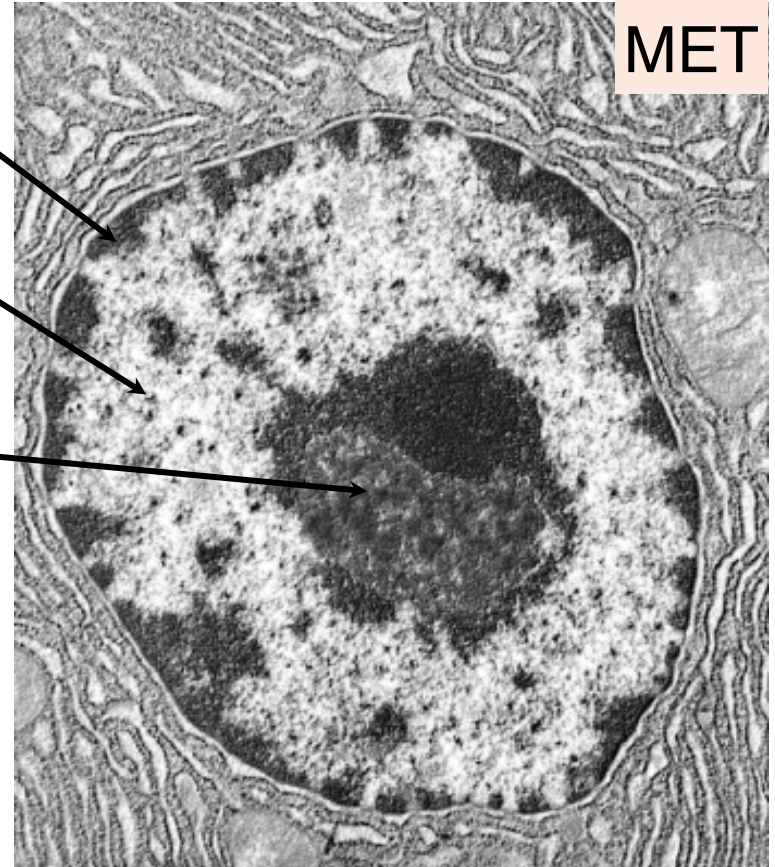
2. L'organisation des génomes et leur diversité

2.2. Le génome eucaryote

Le génome nucléaire



acinus pancréatique



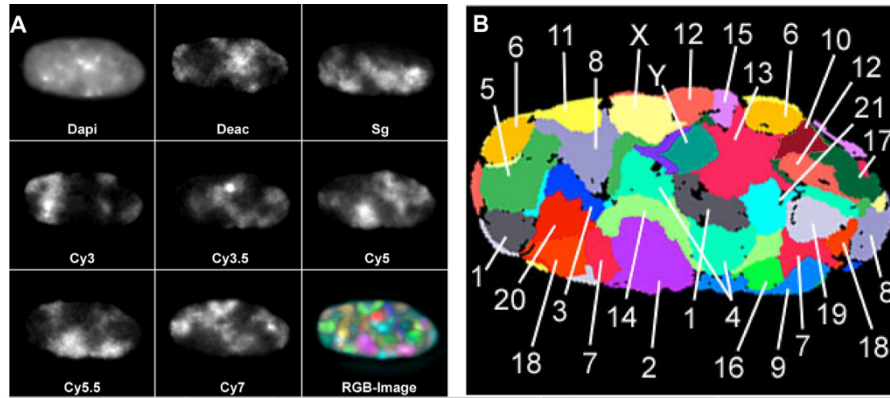
MET

Composition de la chromatine

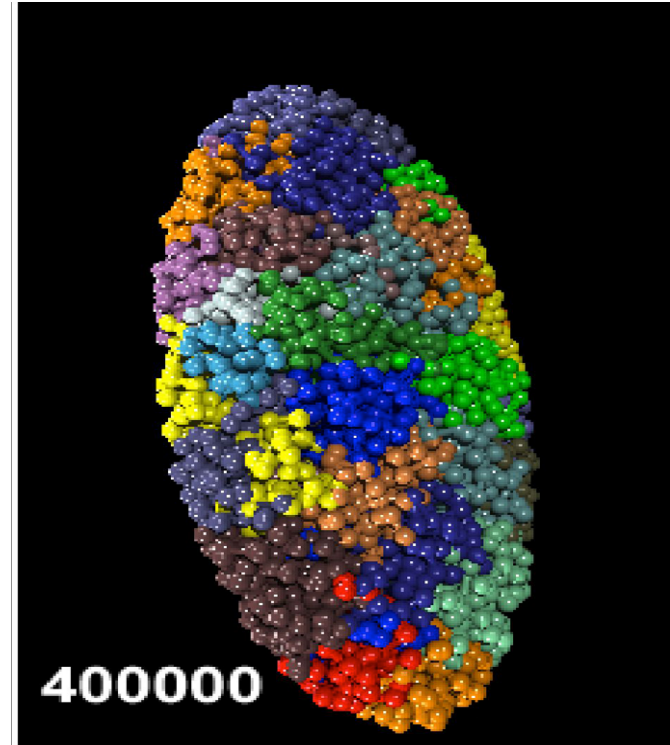
30% ADN 5% ARN
30% protéines histones
25% protéines autres
10 % autres molécules (nucléotides libres, phosphates...)

Les territoires chromosomiques

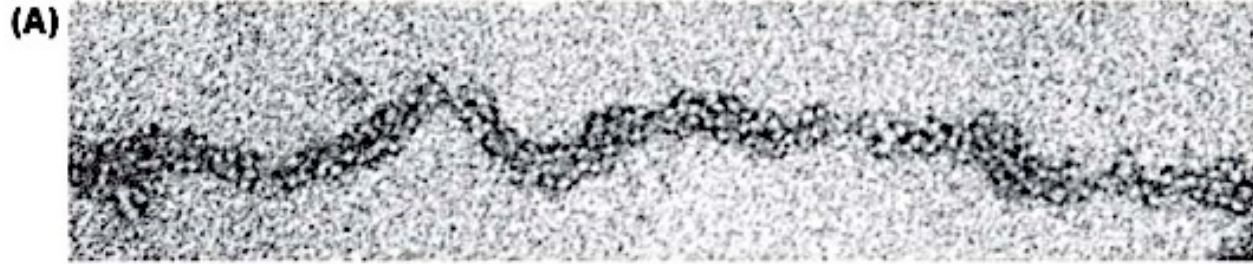
Les chromosomes linéaires forment des régions nucléaires propres



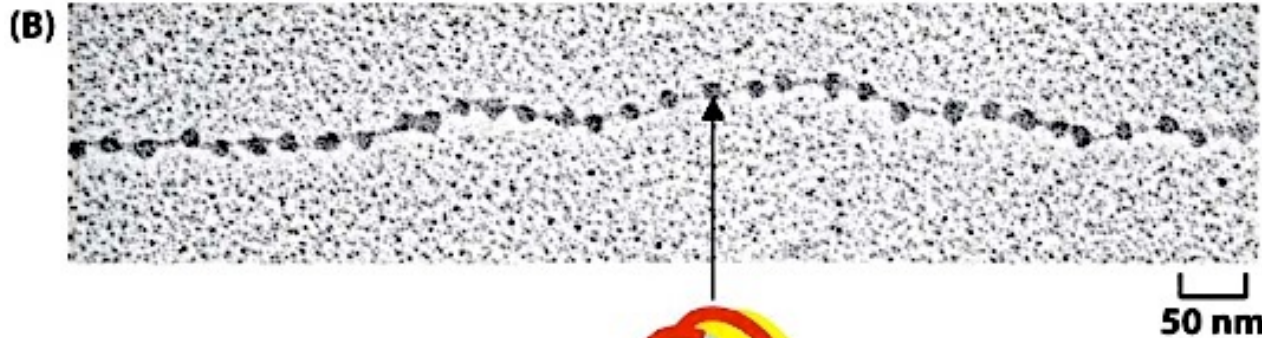
Technique de fluorescence permettant de visualiser les zones correspondant à chaque chromosome.



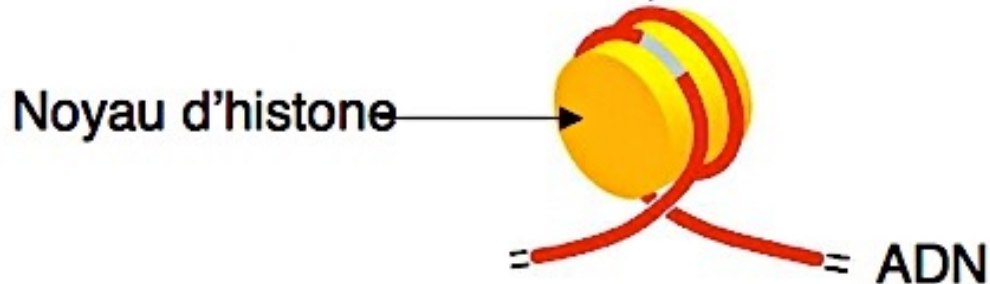
Observation de l'euchromatine



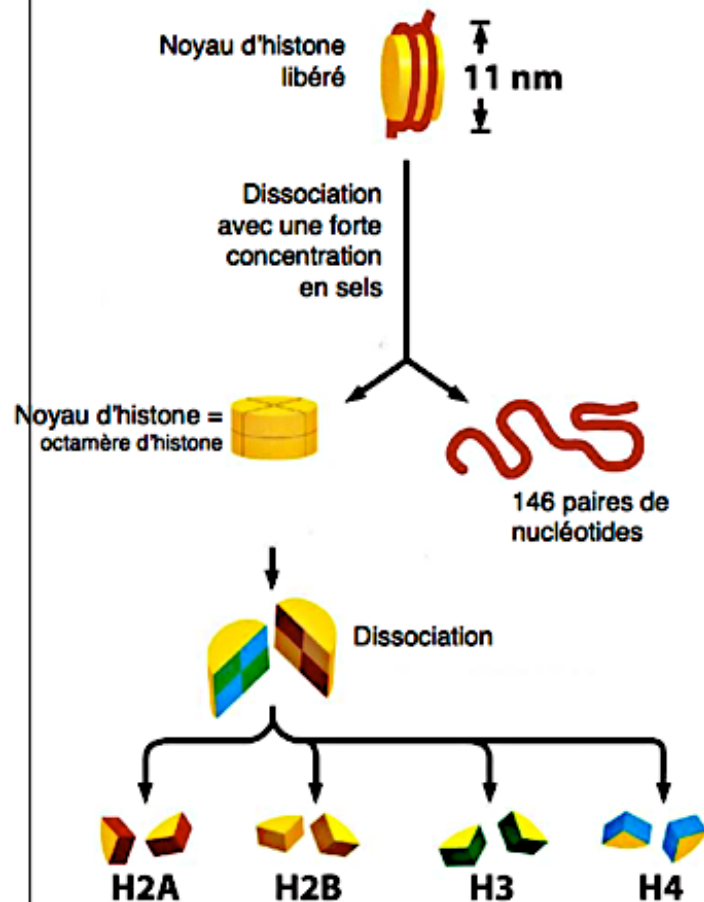
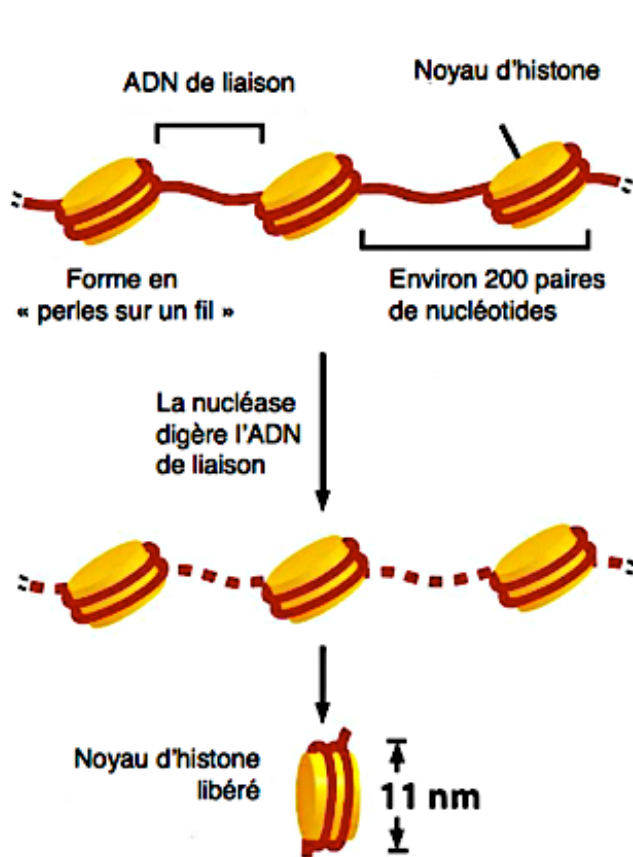
fibre nucléosomique de
30 nm de diamètre



nucéofilament de
11 nm de diamètre

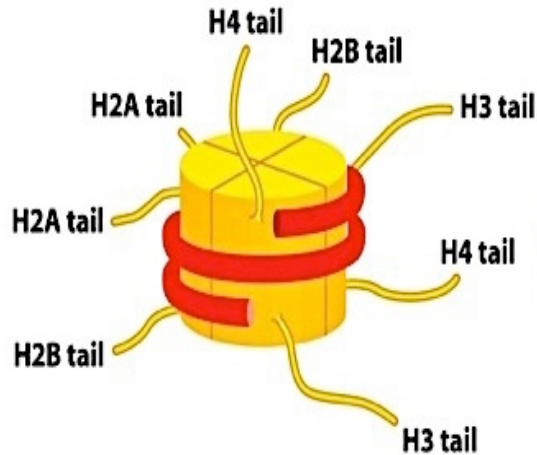


Le nucléosome, unité de structure



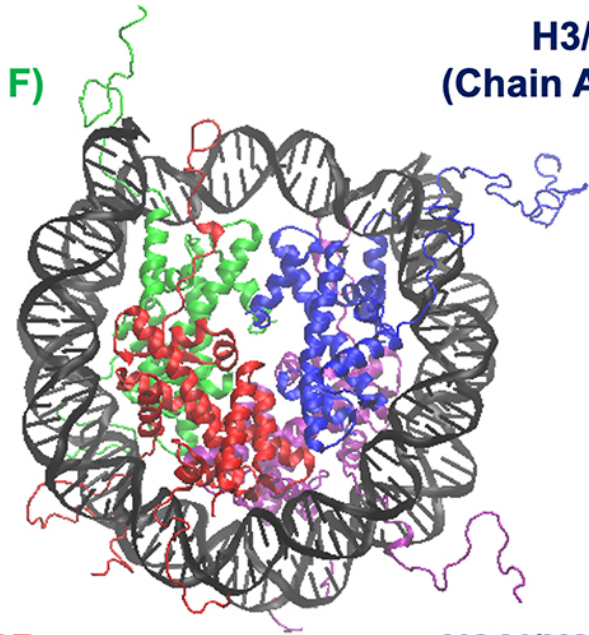
Les histones, protéines basiques

Noyau du nucléosome = 8 histones



H3'/H4'
(Chain E and F)

H3/H4
(Chain A and B)

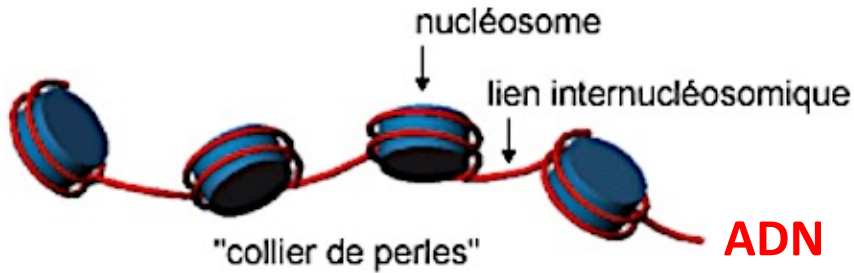


H2A/H2B
(Chain C and D)

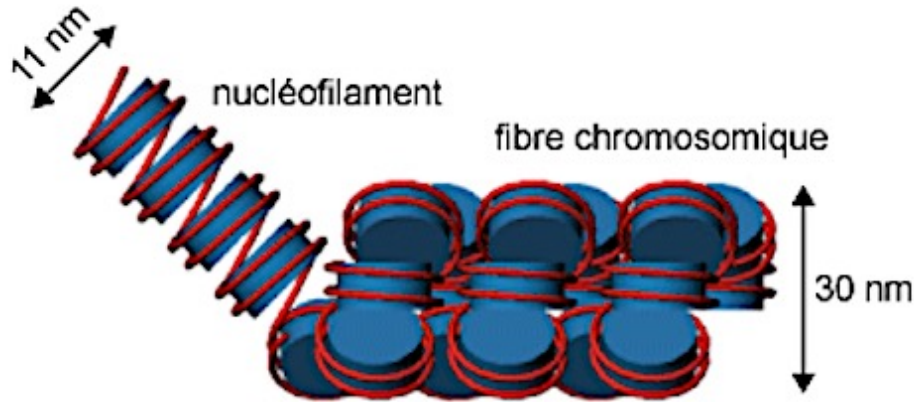
H2A'/H2B'
(Chain G and H)

ADN
150 pb enroulées
+ 50 pb linker

La compaction de l'ADN



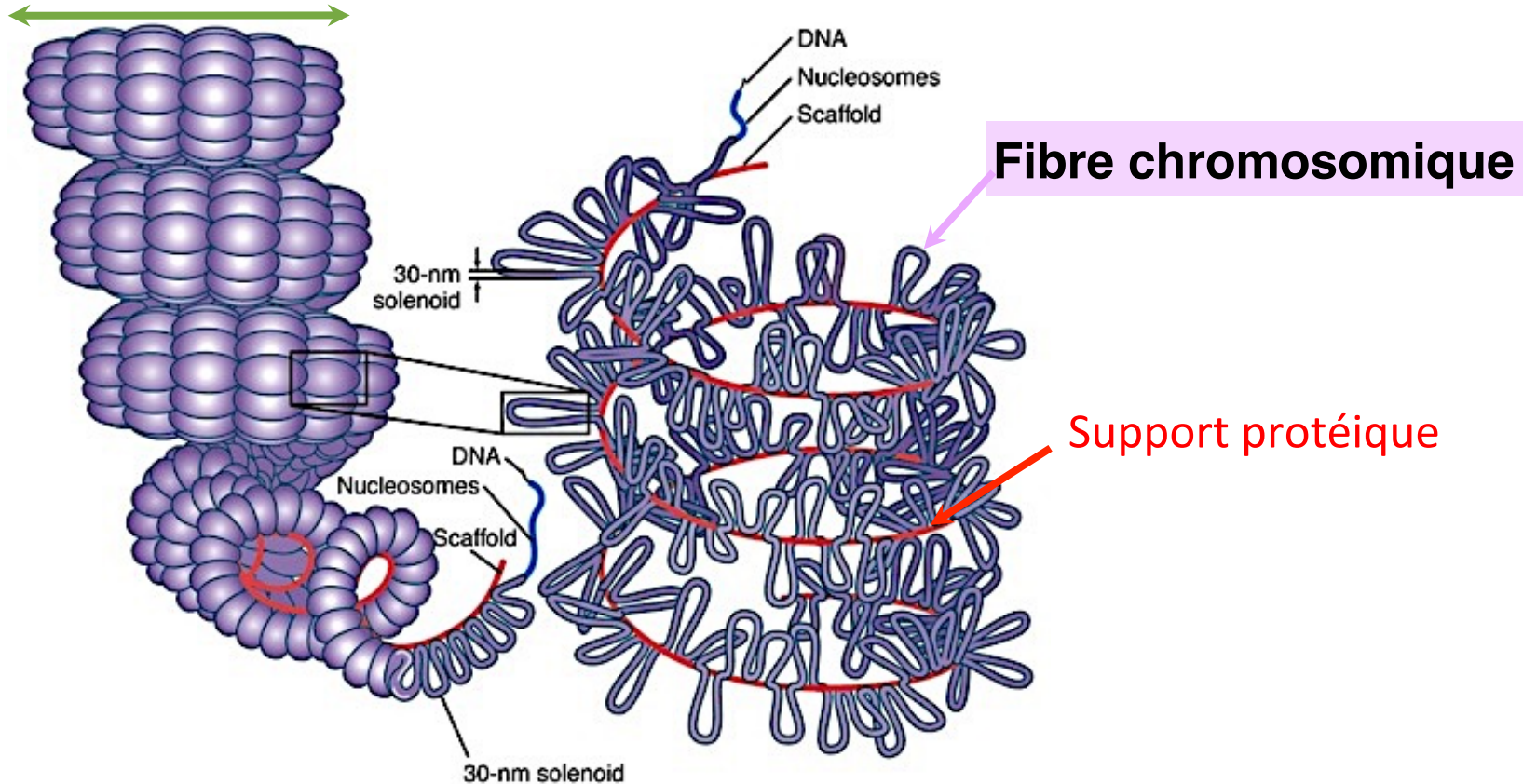
Nucléofilament
raccourcissement d'un facteur 7



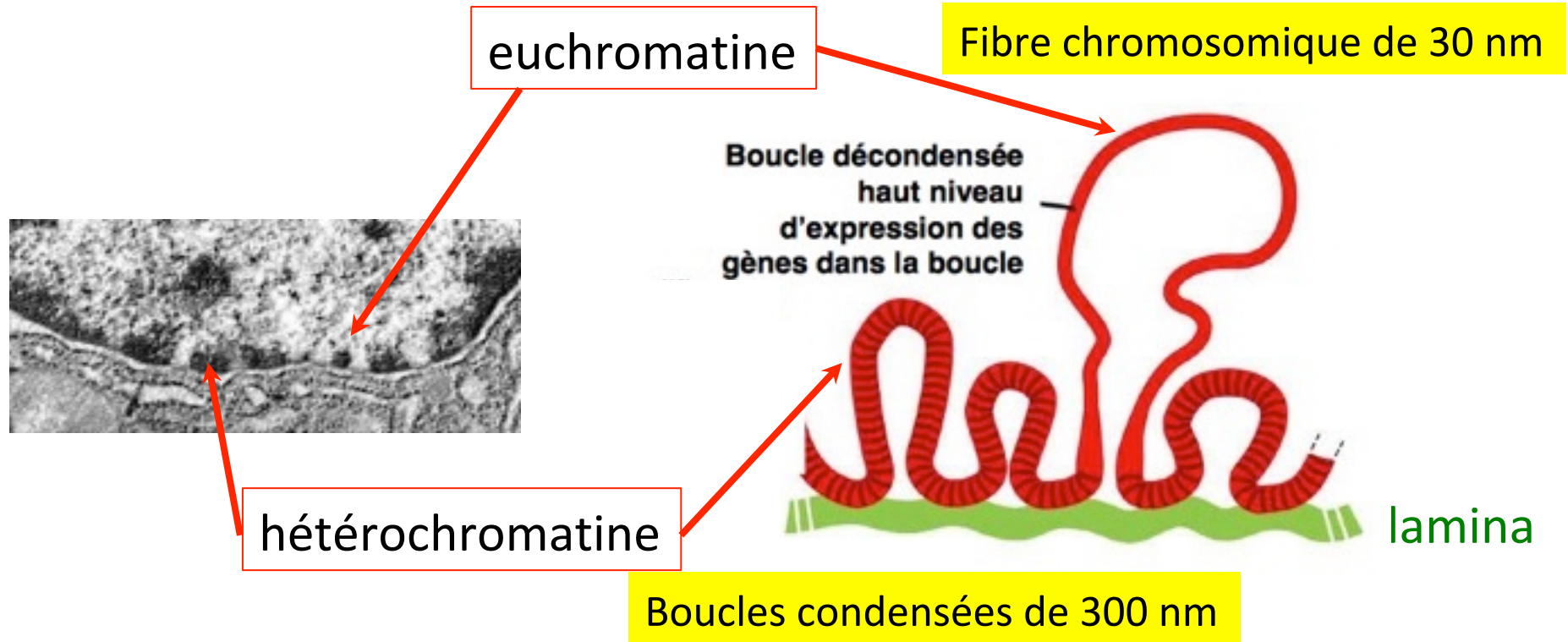
Fibre chromosomique
raccourcissement d'un facteur 6

L'hétérochromatine

boucles condensées (300 nm)

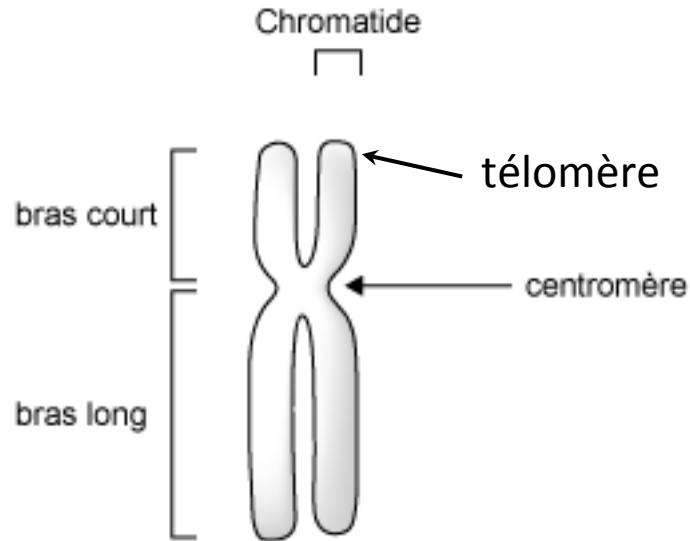


Eu- et hétérochromatine



➤ Hétérochromatine constitutive et facultative

La prophase : une compaction supplémentaire

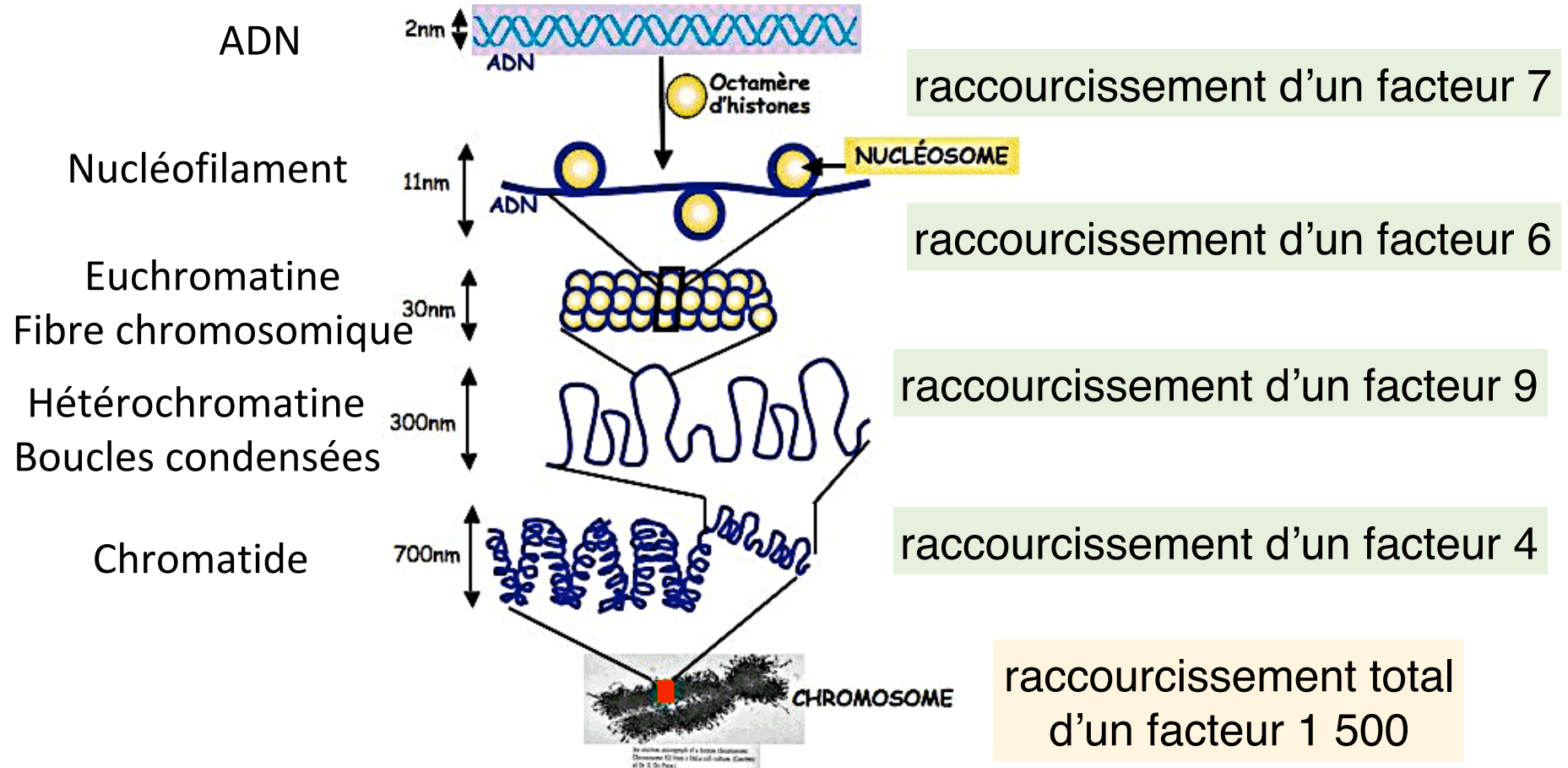


Chromosome métaphasique

2 chromatides issues de la réplication

Compaction sur un squelette protéique

BILAN



BILAN

État compacté : au minimum fibre chromosomique

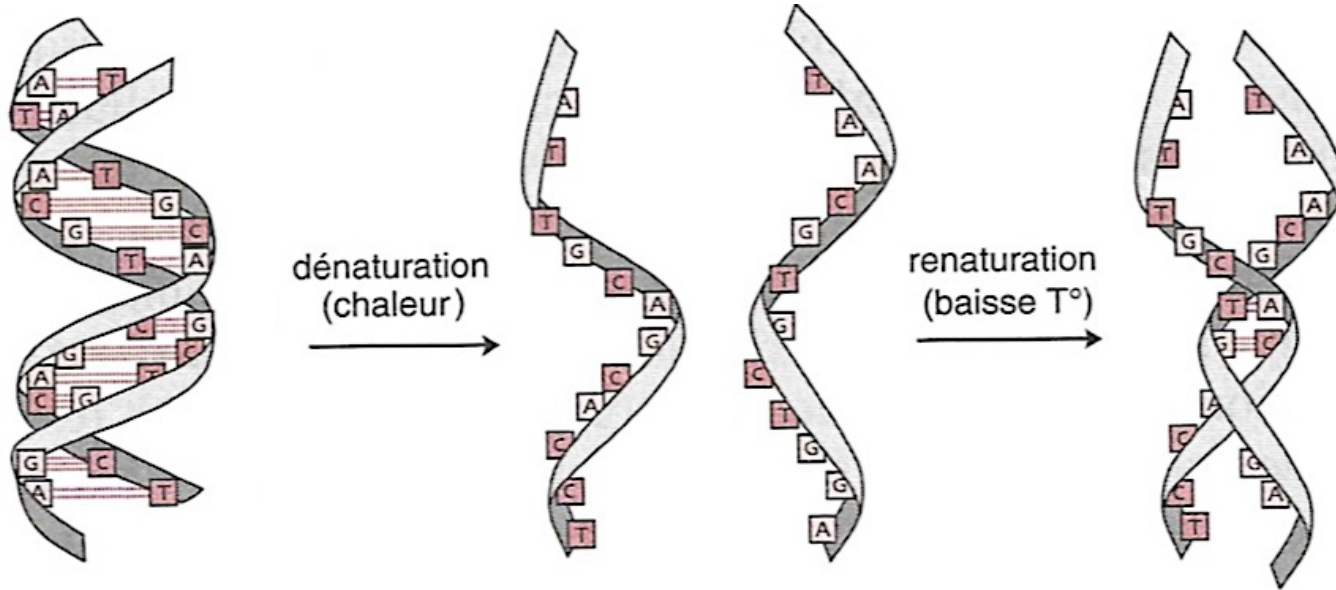
Lien avec l'expression génétique (accessibilité de l'ADN)

État dynamique de la compaction :

- hétérochromatine constitutive
- hétérochromatine facultative
- des facteurs de régulation de l'état de compaction = épigénétique

Structure la plus condensée = chromosome métaphasique

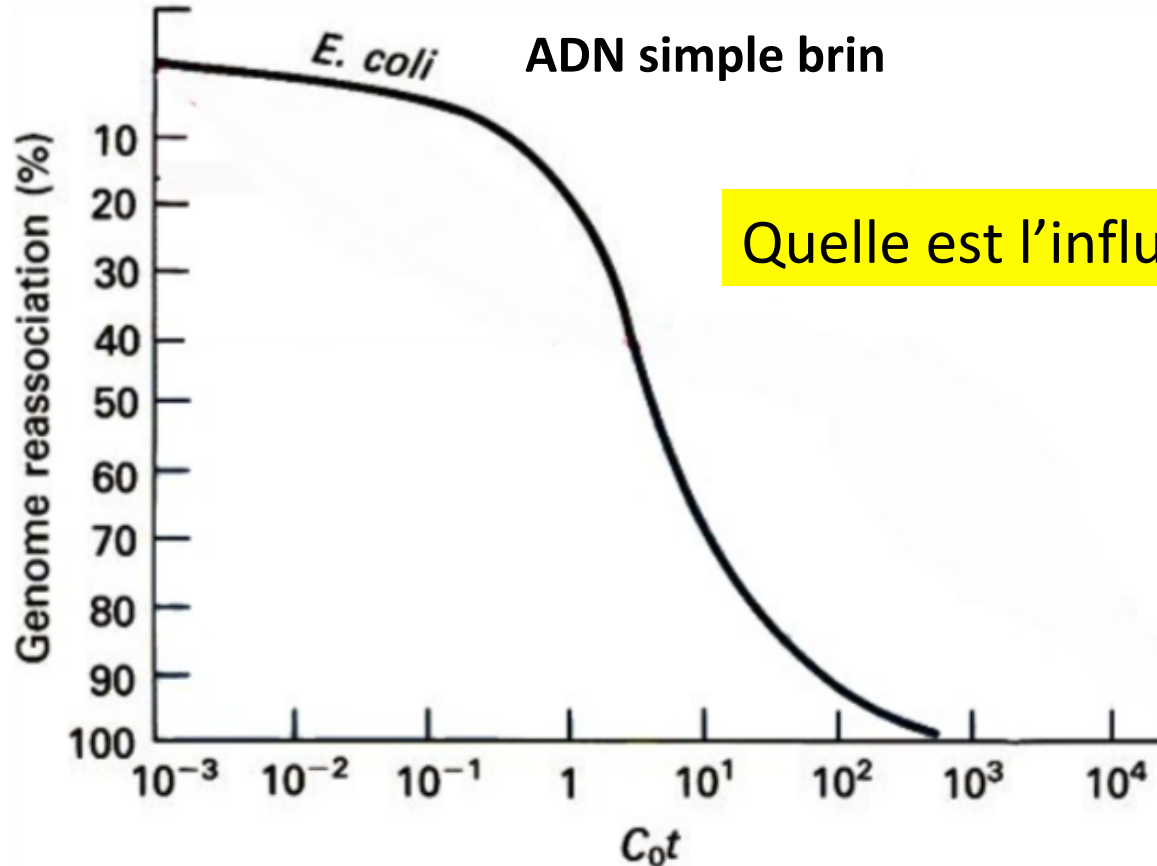
Analyser les séquences d'ADN



Dénaturer l'ADN = séparer les 2 brins

L'ADN absorbe à 260 nm : absorption différente entre ADN double brin et ADN simple brin.

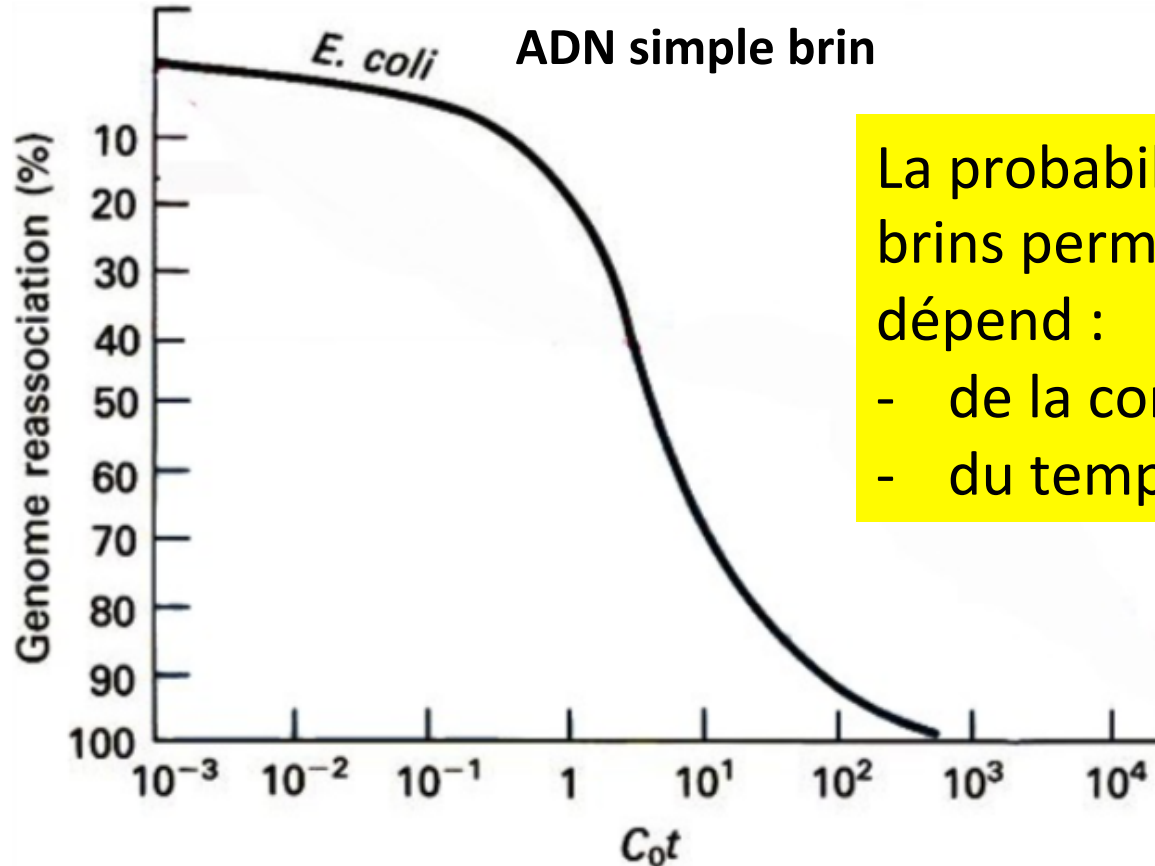
Expérience de dénaturation et renaturation



Quelle est l'influence de C_0 et de t ?

$C_0 = 10^{-3} \text{ mol.L}^{-1}$
 t en secondes

Expérience de dénaturation et renaturation



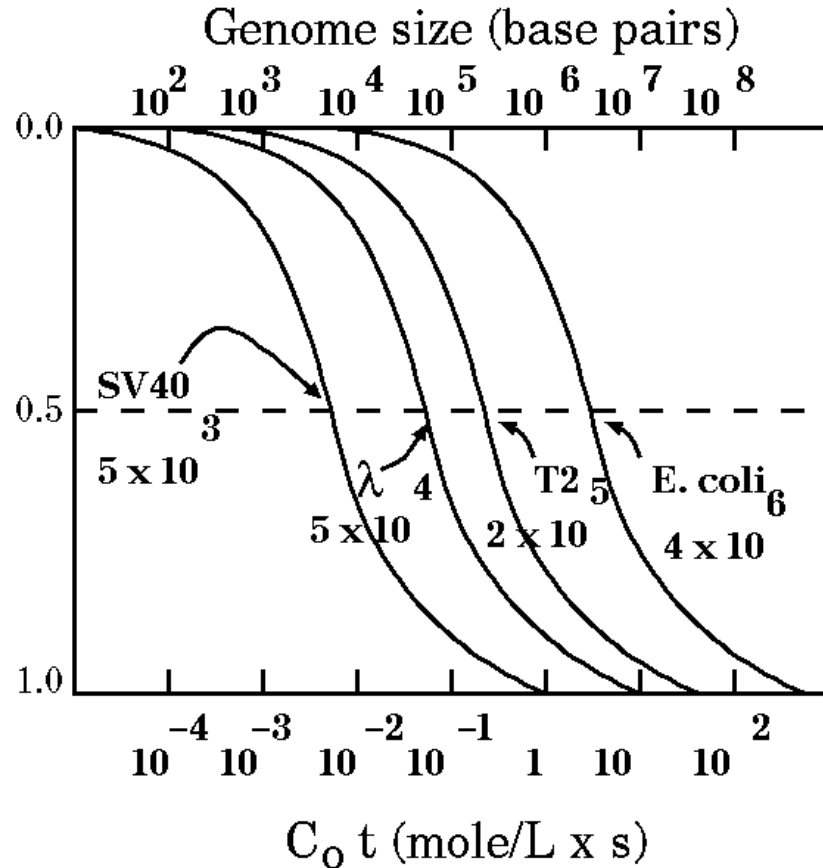
La probabilité de rencontre des 2 brins permettant la renaturation dépend :

- de la concentration en ADN C_0
- du temps de contact.

$$C_0 = 10^{-3} \text{ mol.L}^{-1}$$

t en secondes

Influence de la taille du génome



Virus testés

SV40 : génome de $5 \cdot 10^3$ pb

Bactériophage λ : génome de $5 \cdot 10^4$ pb

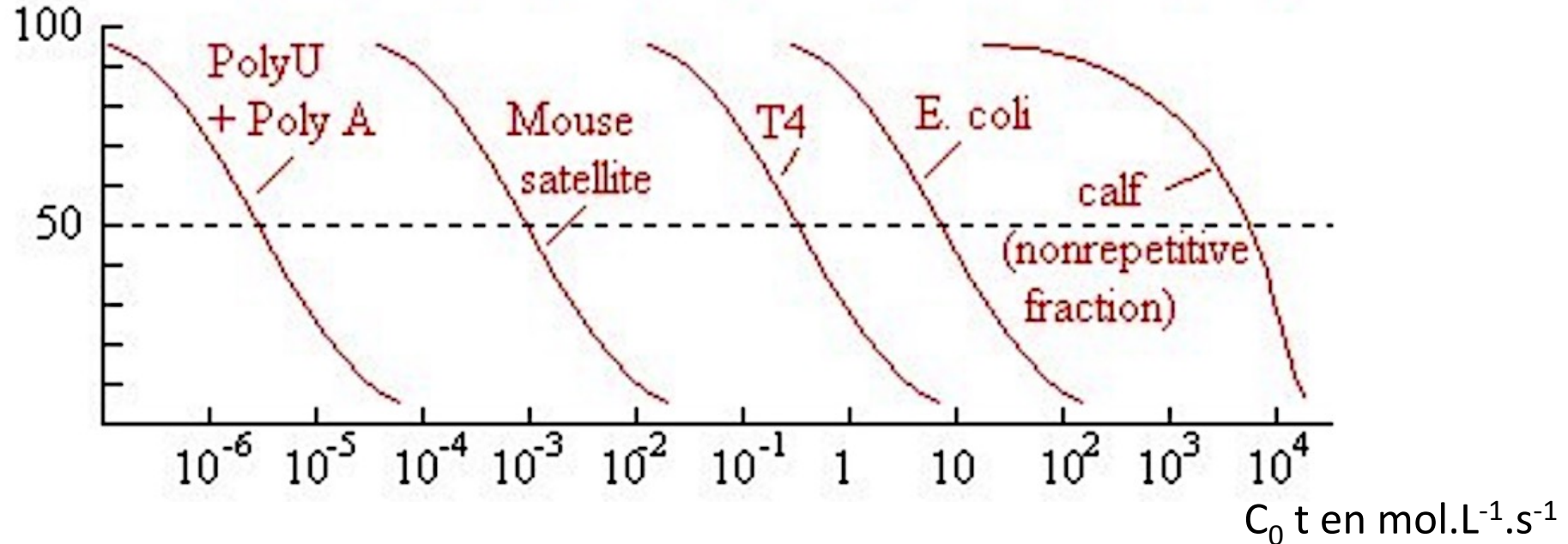
Bactériophage T2 : génome de $2 \cdot 10^5$ pb

Bactérie

E. coli : génome de $4 \cdot 10^6$ pb

Influence de la présence d'ADN répété

% ADN simple brin

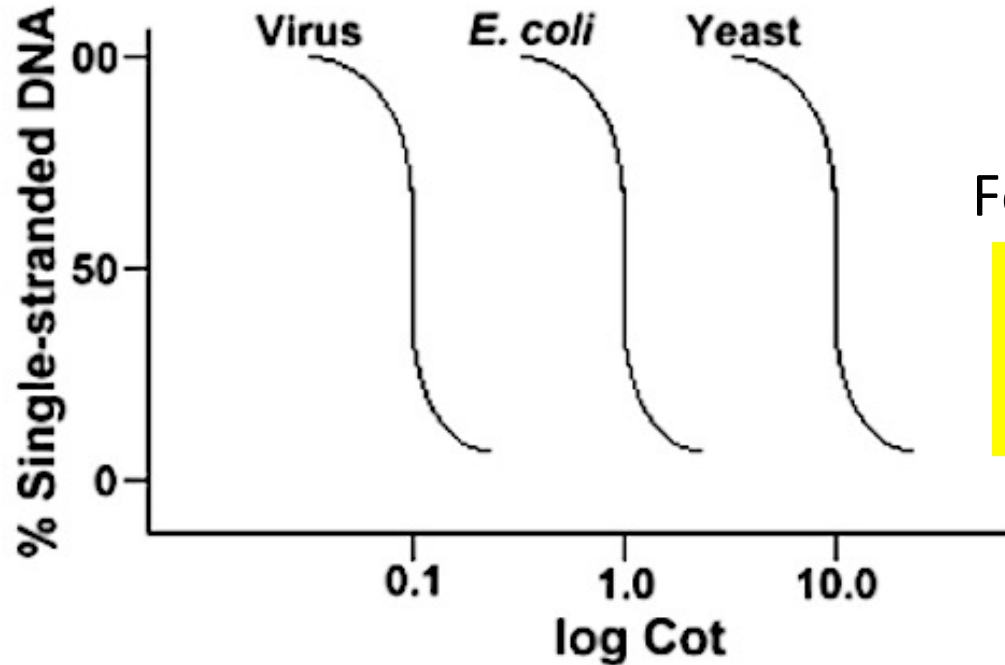


PolyU + PolyA = ADN synthétique artificiel totalement répété en A ou U

Mouse satellite = séquence retrouvée en grand nombre dans le génome de souris

Calf = séquence unique de génome de Veau

Application aux formes de vie



Formes de vie unicellulaires

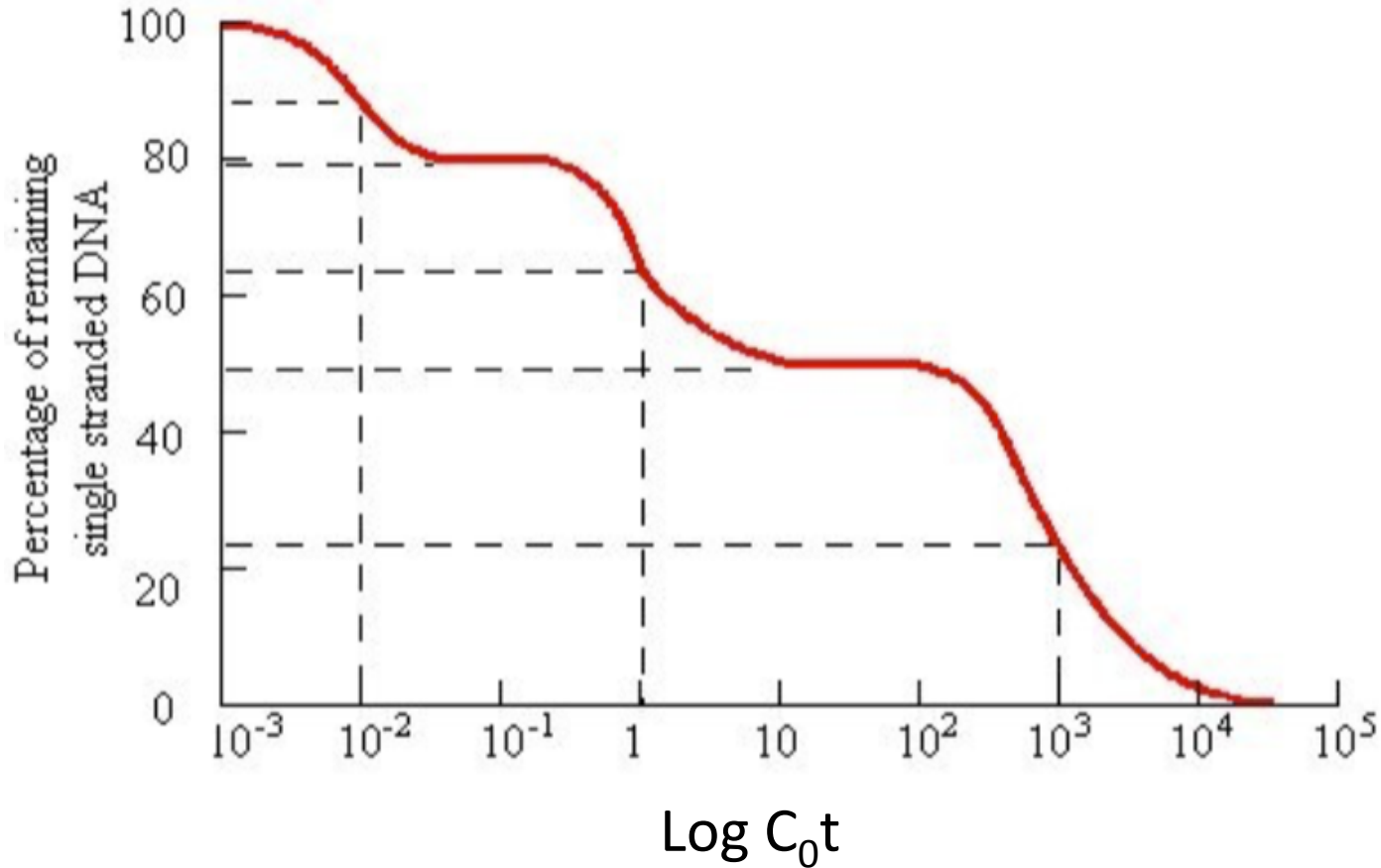
Génome de virus : $5 \cdot 10^4$ pb

Génome *E. coli* : $4,6 \cdot 10^6$ pb

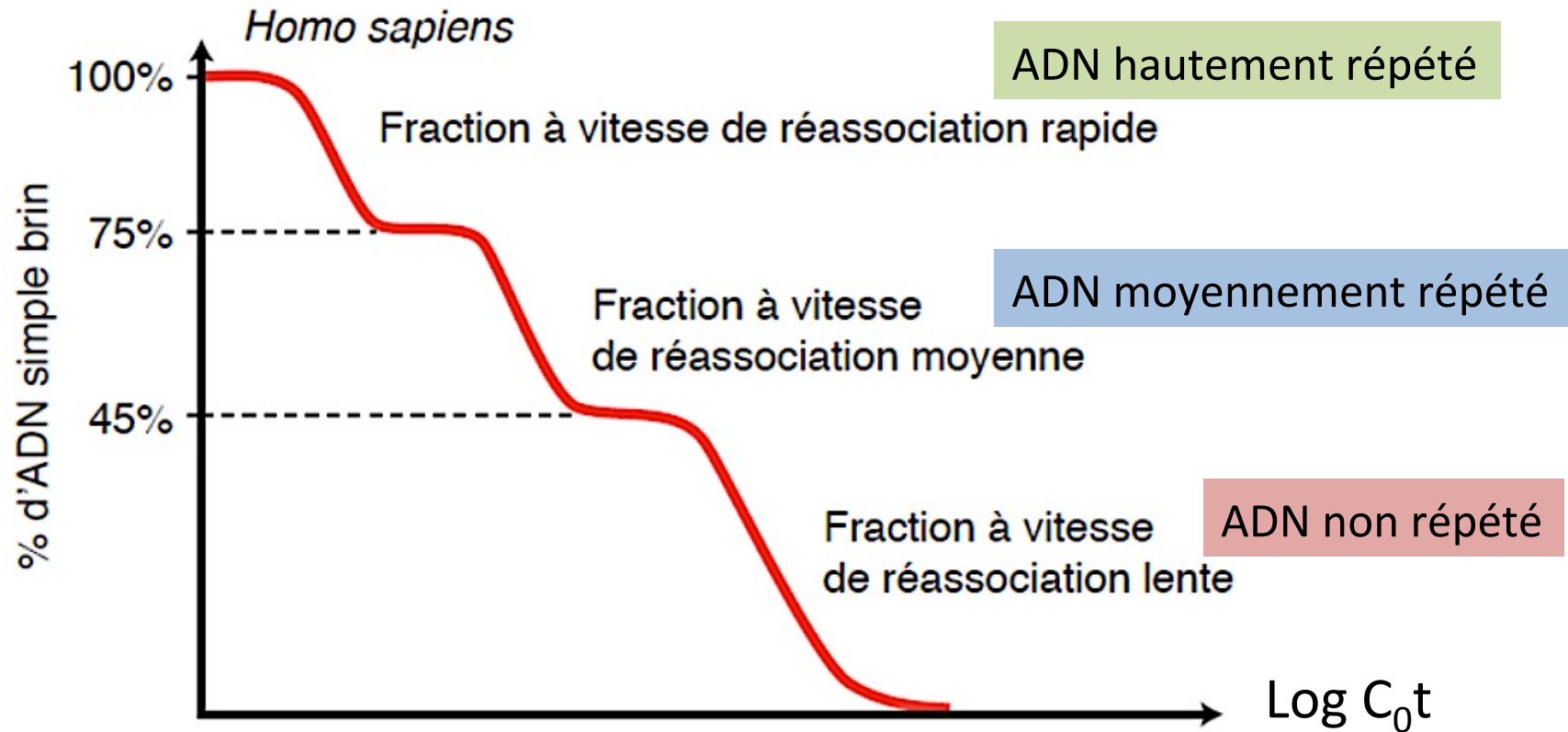
Génome de Levure : $12 \cdot 10^6$ pb

Levure = Eucaryote unicellulaire : peu de séquences répétées
ADN codant = 68 % du génome

Application aux Mammifères



Application aux Eucaryotes : *Homo sapiens*



ADN répété et non répété

La proportion d'ADN répété dépend beaucoup d'un Eucaryote à un autre

10 à 25 % d'ADN hautement répété

25 à 30 % d'ADN moyennement répété

45 % d'ADN non répété

L'ADN hautement répété

- **ADN satellite**

blocs d'ADN répétés en tandem constituant les séquences :

- des télomères : séquences riches en C et G renforçant la stabilité des extrémités des chromosomes ;
- des centromères : séquences répétées de 171 pb permettant la liaison des chromosomes au fuseau mitotique

- **ADN mini et microsatellite**

courtes séquences répétées de motifs polymorphiques (utilisés dans la reconnaissance des individus et espèces).

Centromères : séquences CEN



répétition de ce motif sur 10^6 pb

Télomères : séquences TEL

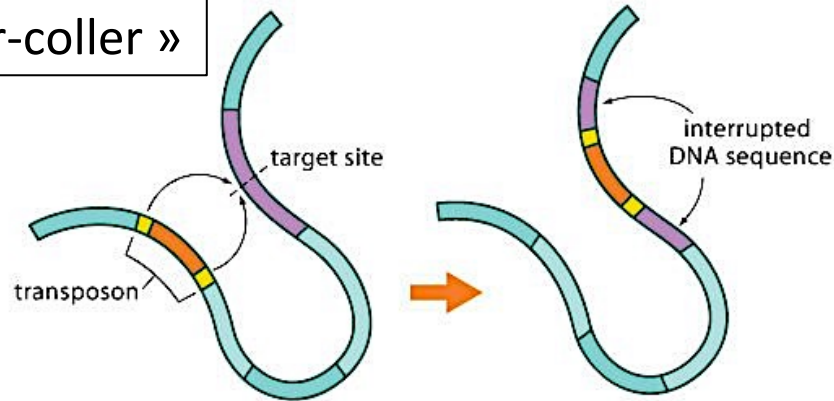
Organisme	Nature des répétitions	Longueur du télomère
Levure	TGGG	400 pb
souris <i>Mus musculus</i>	TTAGGG	150 000 pb
<i>Arabidopsis thaliana</i>	TTTAGGG	9 000 pb
<i>Homo sapiens</i>	TTAGGG	10 000 pb

L'ADN moyennement répété

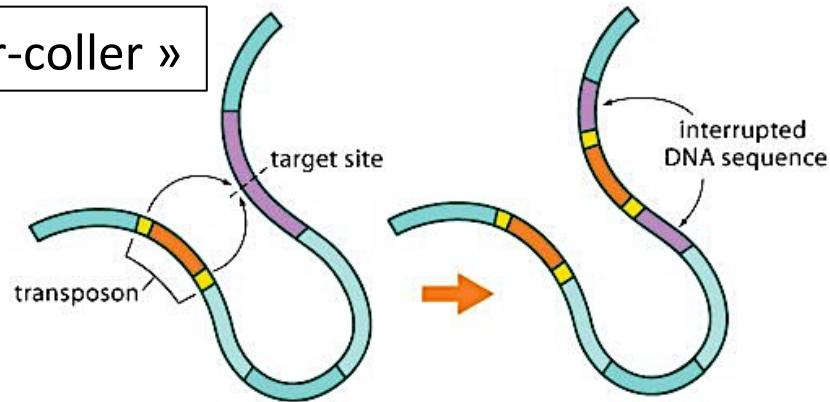
- **Éléments transposables : transposons et rétrotransposons**
séquences mobiles dans le génome, héritées d'ADN viral
séquences courtes (SINE) ou longues (LINE) copiées des milliers de fois.
- **Séquences codant pour les ARN outils**
ARNr en tandem, ARNt (jusqu'à 20 000 copies), ARNsn...
- **Séquences codant pour les protéines histones**
20 copies de chaque gène → **ADN codant**

La transposition

Type « couper-coller »

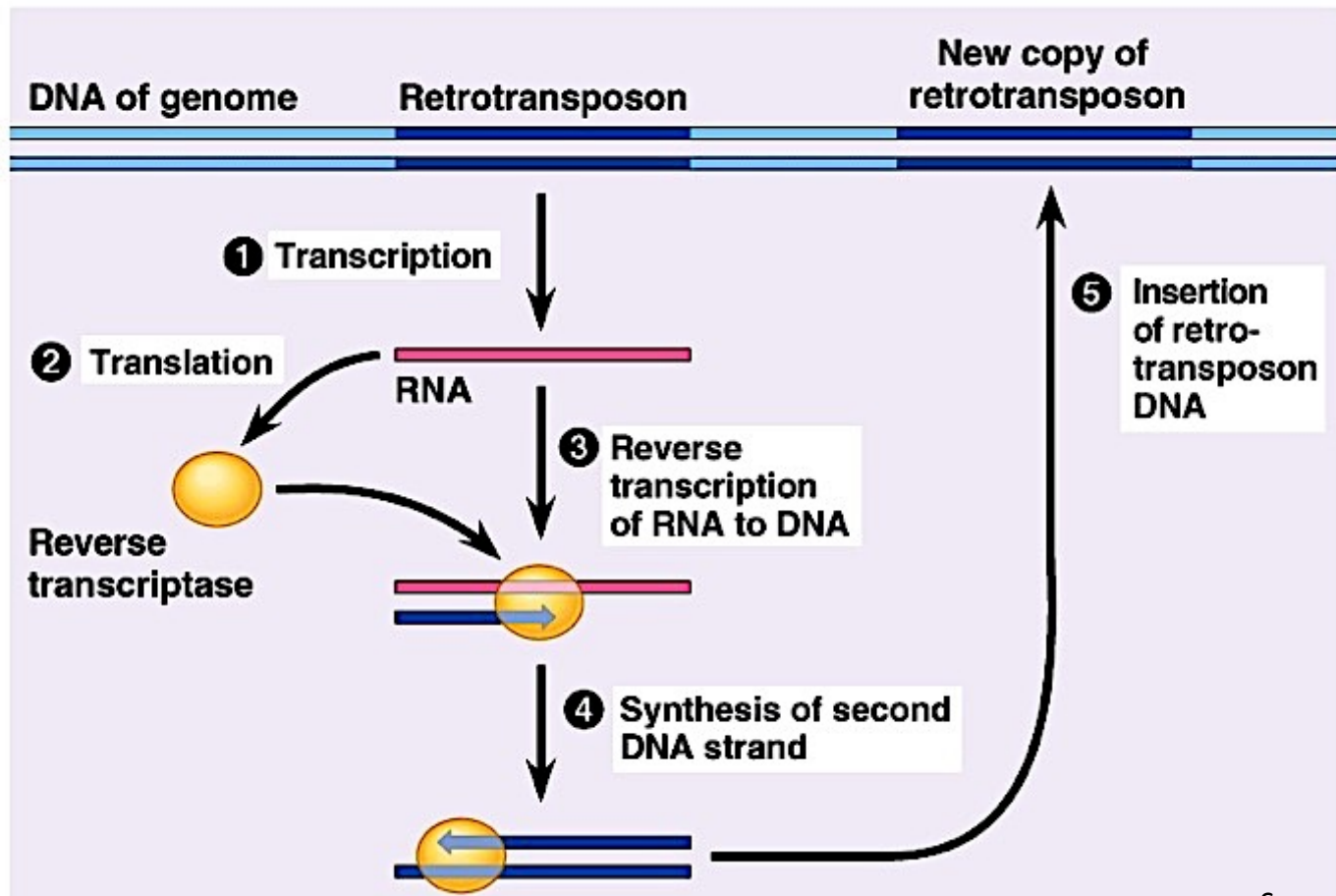


Type « copier-coller »

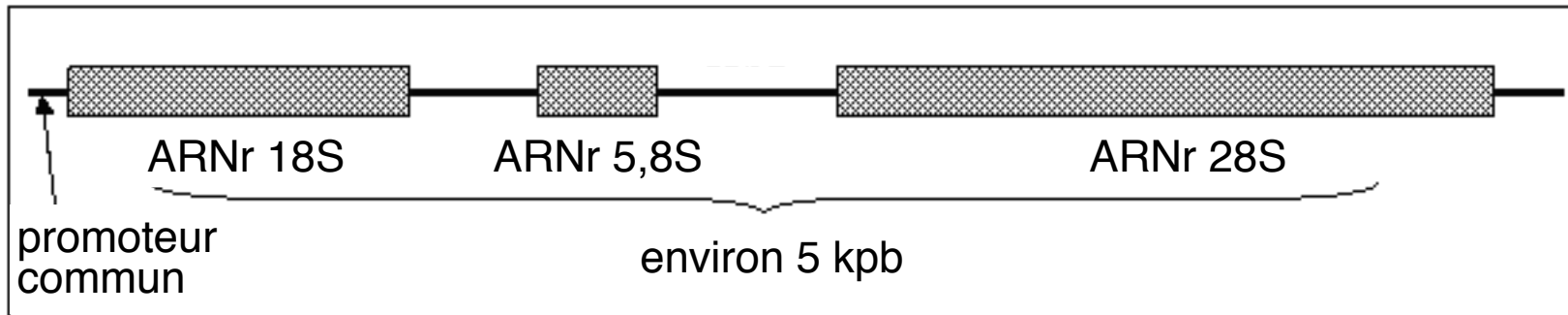


Conduit à une
multiplication des
exemplaires

La rétrotransposition

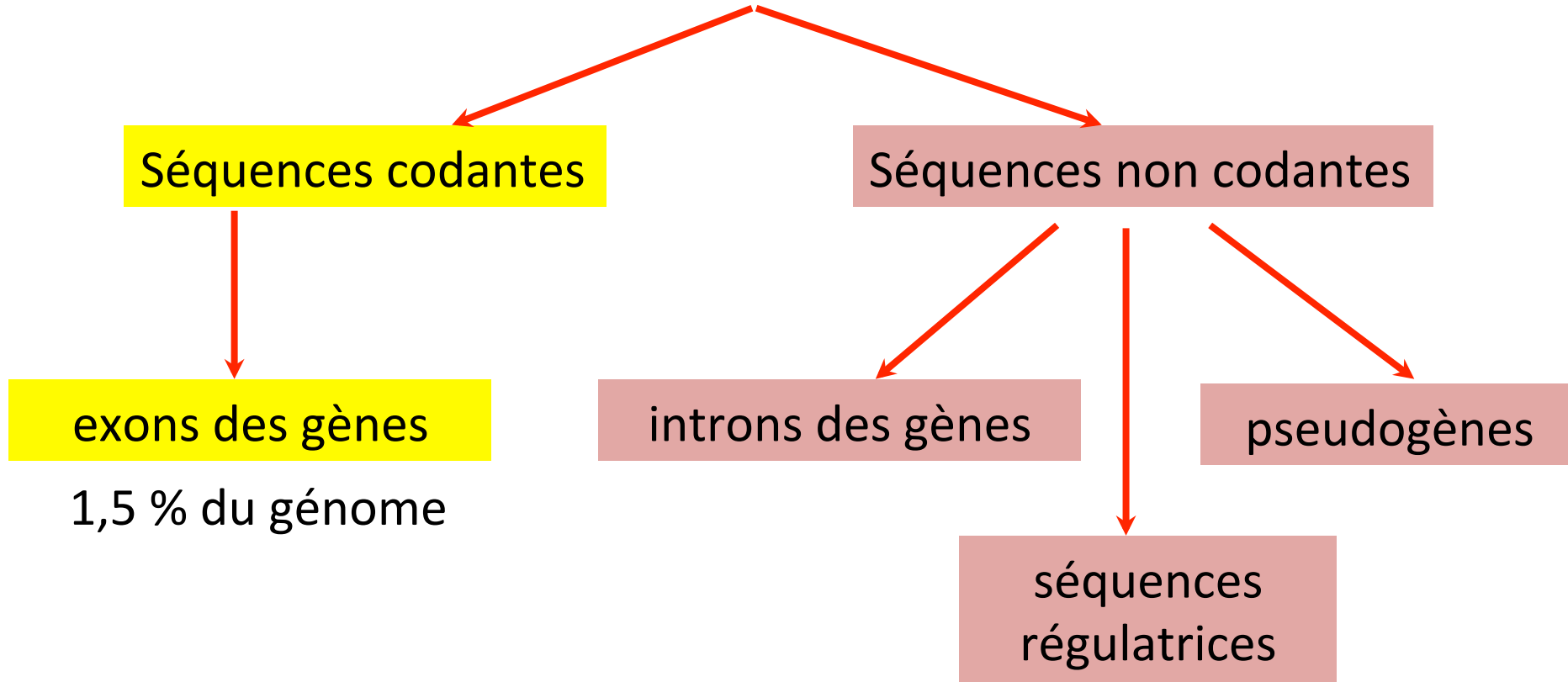


Les séquences des ARNr

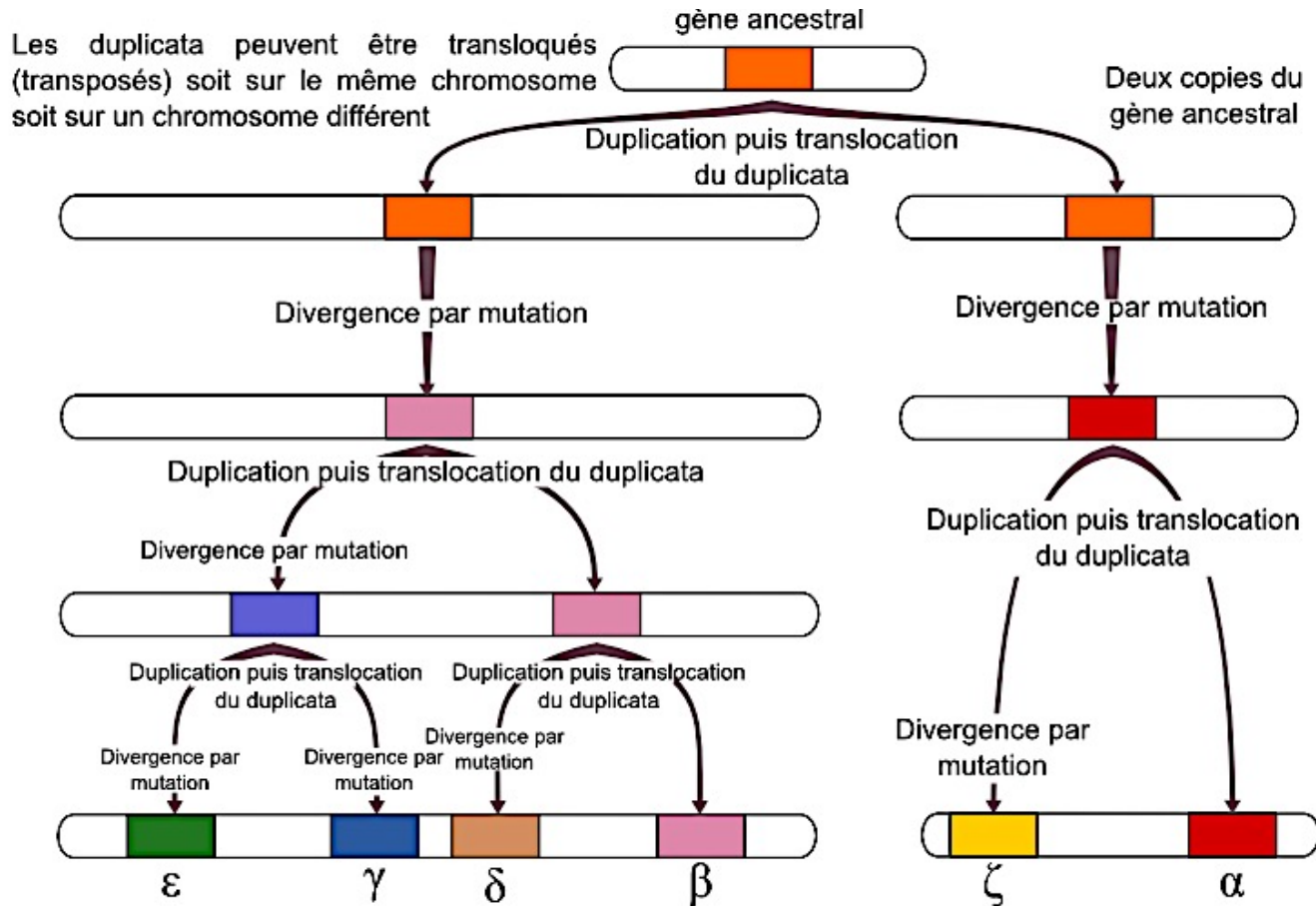


	élément du ribosome	ARNr	protéines	Ribosome
Bactéries	petite sous-unité 30S	16 S	21 sortes	70 S
	grande sous-unité 50S	23 S + 5 S	31	
Eucaryotes	petite sous-unité 40S	18 S	33	80 S
	grande sous-unité 60S	28 S + 5,8 S	50	

L'ADN non répété

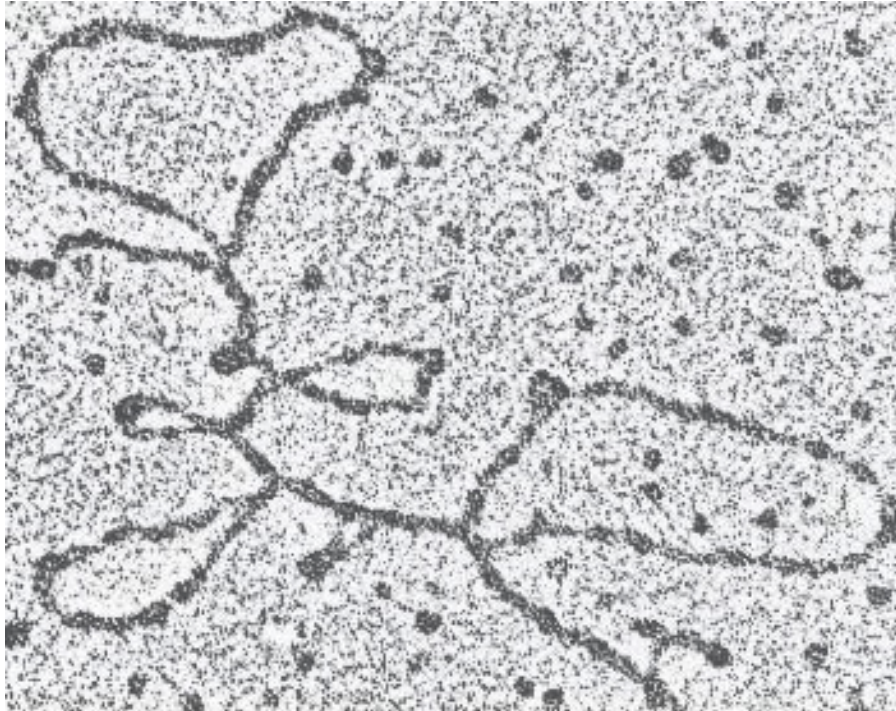


Les familles multigéniques : les globines

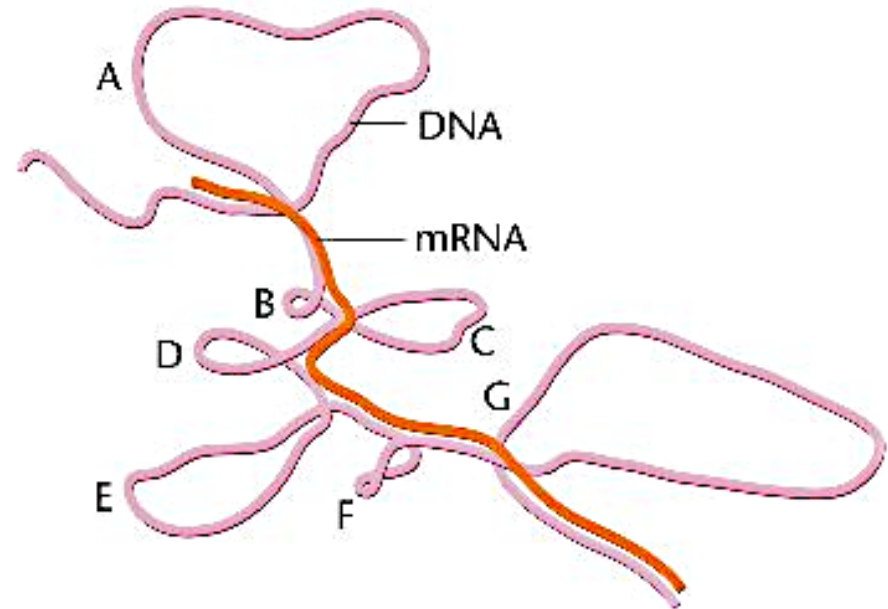


Les gènes morcelés des Eucaryotes

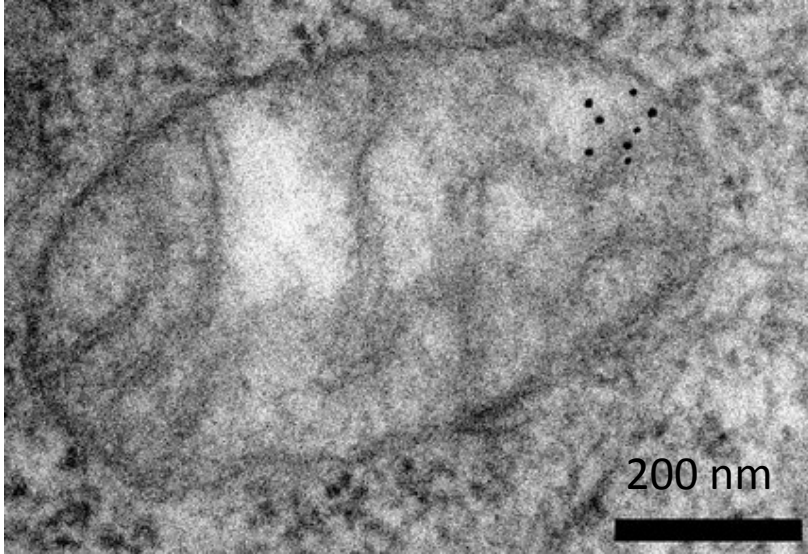
ADN du gène de l'ovalbumine de Poule mis au contact de l'ARNm



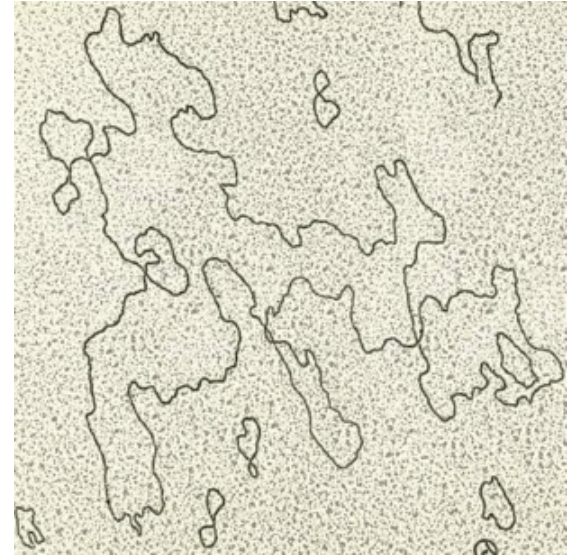
Interprétation
hybridation ADN-ARNm



Le génome cytoplasmique



Mitochondrie avec marquage de l'ADN
jusqu'à 10 copies du chromosome

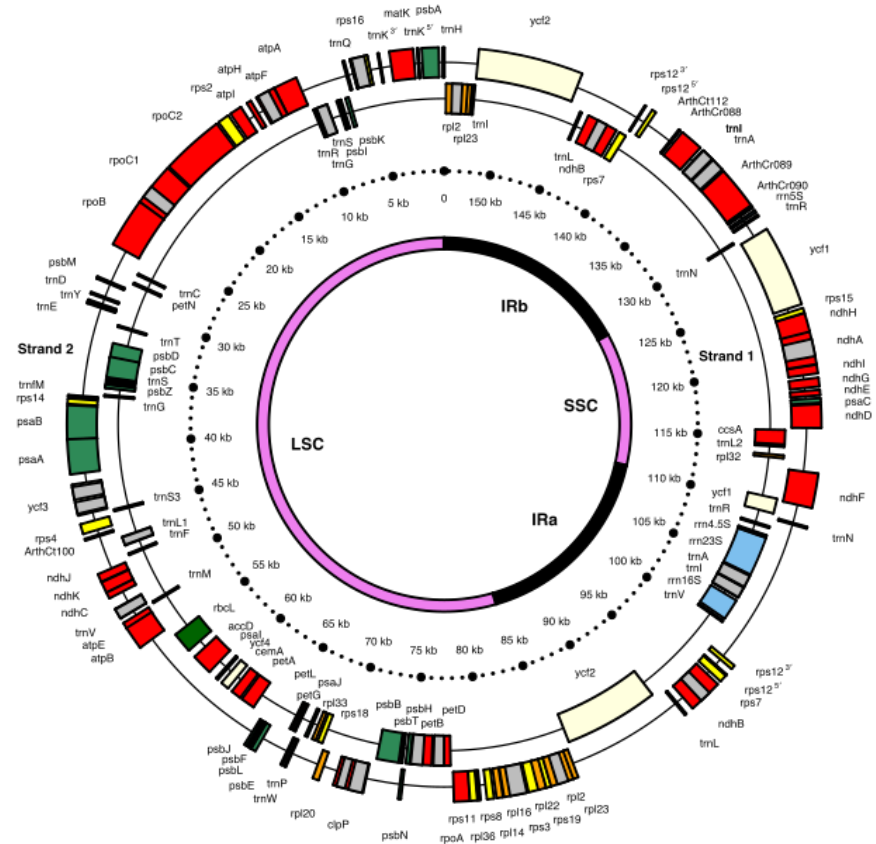


chromosome de chloroplaste
isolé et étalé observé au MET

Le chromosome chloroplastique

ADN de 70 à 200 kb
quelques séquences répétées

- 50 gènes de protéines
- 4 gènes des ARNr
- 30 gènes d'ARNt

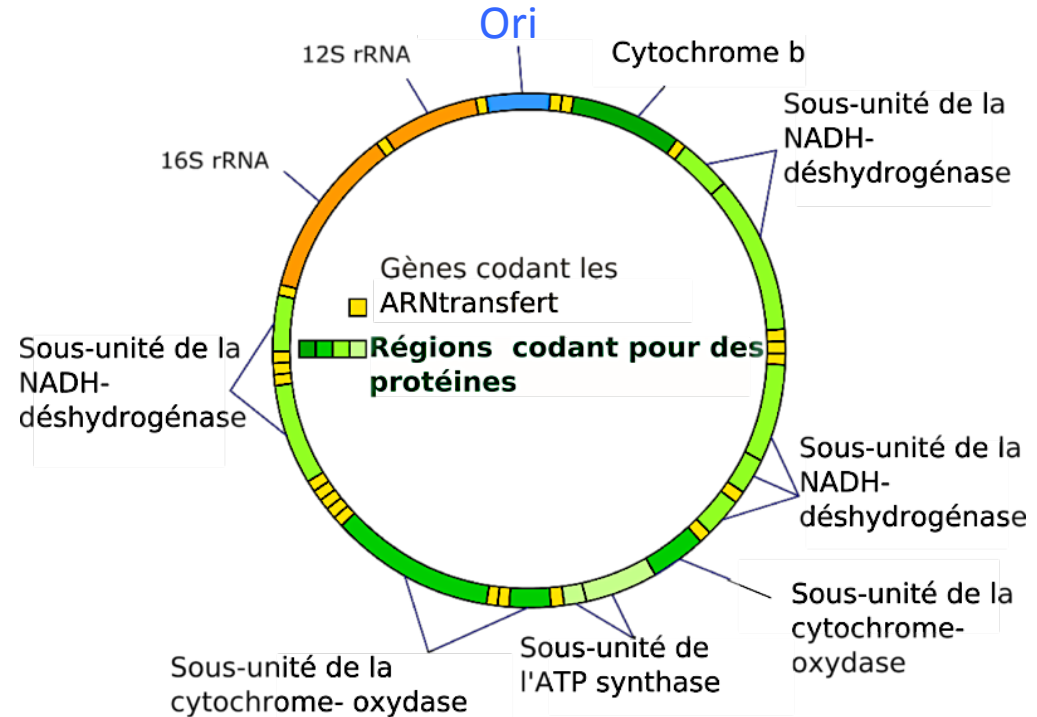


Le chromosome mitochondrial

Mitochondrie humaine

16 kb (= 16 000pb)

- 13 polypeptides impliqués dans la chaîne respiratoire
- 22 ARNt
- 2 ARNr



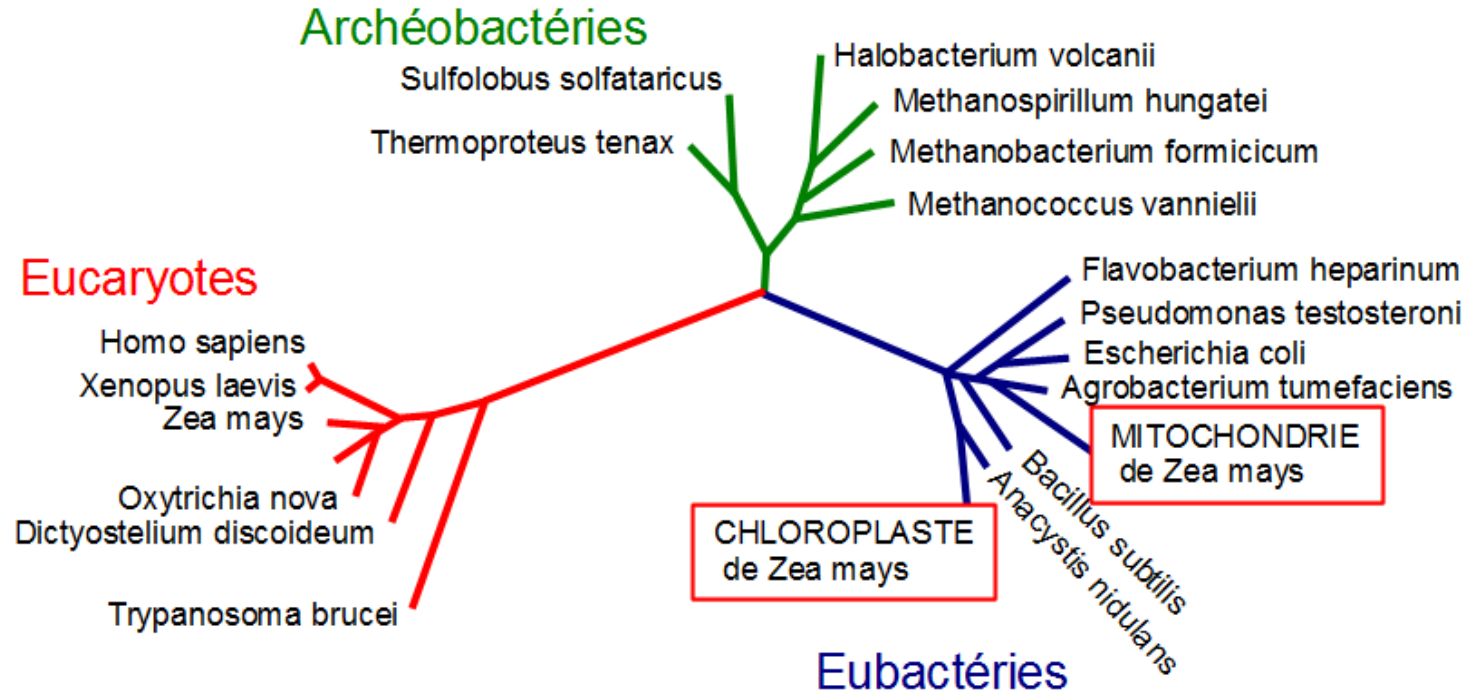
1 500 séquences nucléaires codent les protéines mitochondriales

Un chromosome aux allures bactériennes

- Chromosome circulaire dans un nucléoïde comportant des protéines basiques
- ADN codant
- Pas d'exons
- Séquences des ARNr homologues à celles des bactéries

Notion d'organites semi-autonomes

Une origine bactérienne

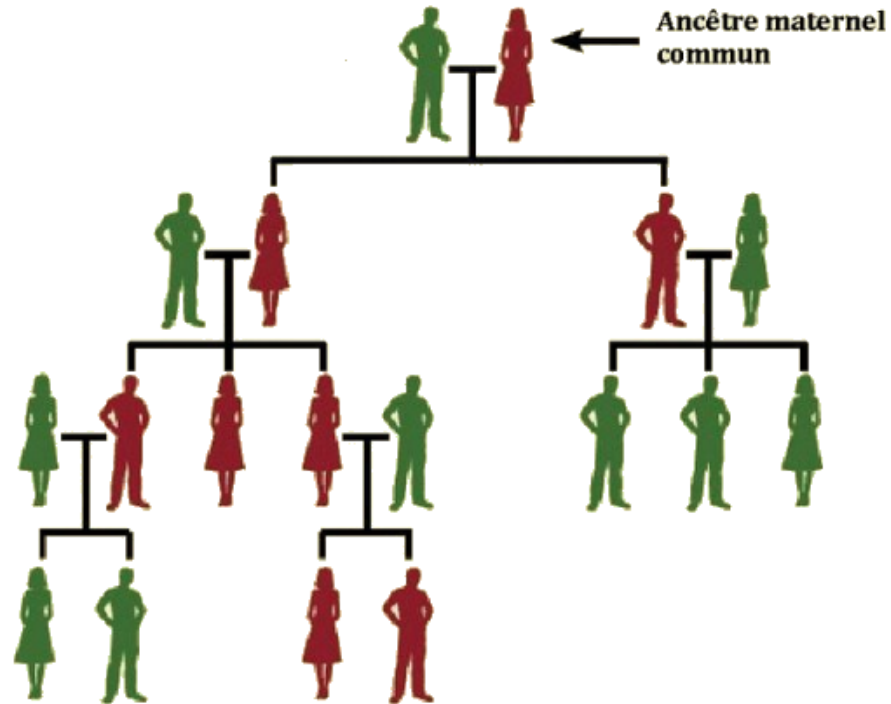


Arbre obtenu à partir de 21 séquences complètes d'ARNr 16S et 18S de Procaryotes, d'Eucaryotes, de mitochondries et de chloroplastes.

Hérédité cytoplasmique

Rouge = atteint

Vert = non atteint



Neuropathie optique de Leber

Suivre la transmission génétique

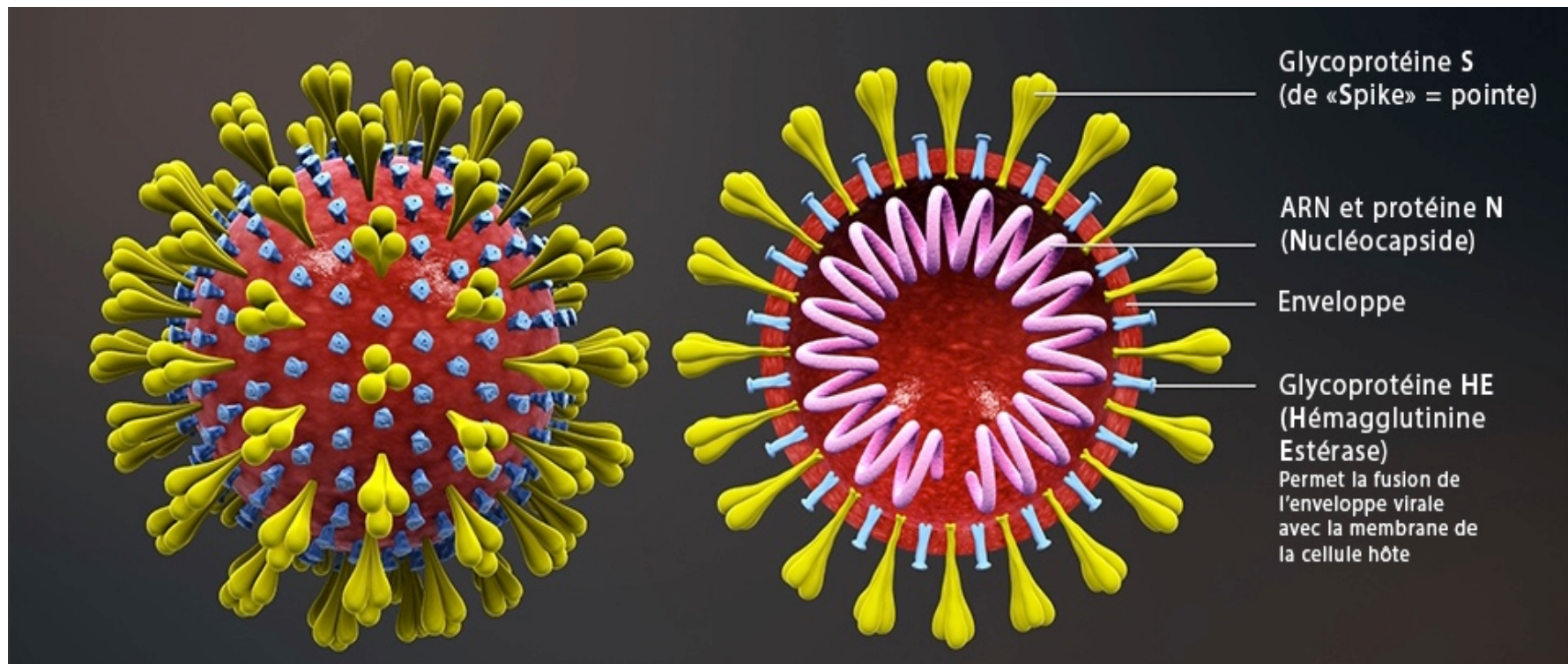
Hérédité maternelle : suivi de l'ADN mitochondrial

Hérédité paternelle : suivi du chromosome Y

2. L'organisation des génomes et leur diversité

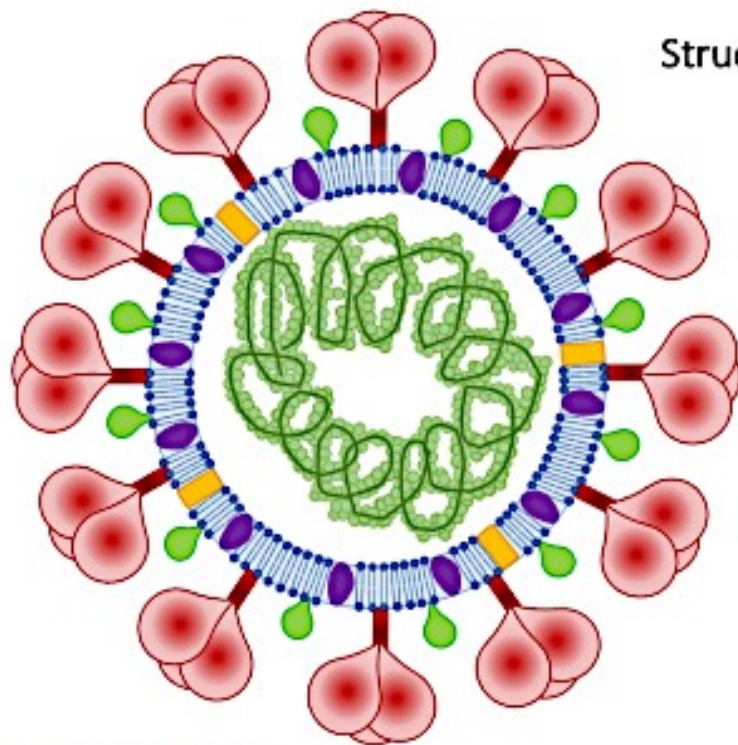
2.3. Les génomes viraux

Le Sars-Cov2



De l'ARN dans une enveloppe

Structure schématisée du SARS-CoV-2



Glycoprotéine de surface
(S, spike)



Hémagglutinine estérase (HE)



Protéine de membrane (M)



Protéine d'enveloppe (E)



Phospholipides



Protéine de nucléocapside (N)



ARN génomique (monocaténaire
de polarité positive)

(A. Gautheret-Dejean 2020)

Deux enveloppes protectrices

CAPSIDE PROTÉIQUE

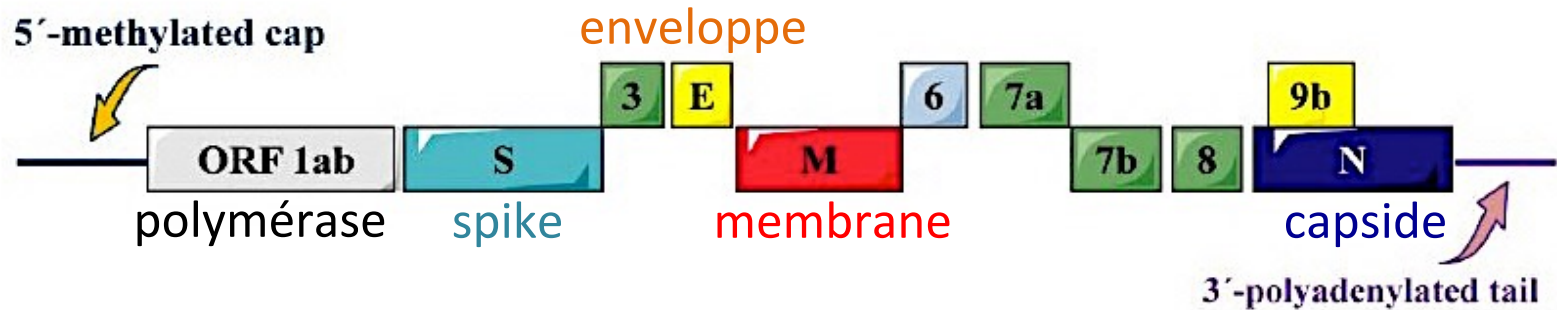
ARN associé à des protéines N : enveloppe protéique = **capside**

ENVELOPPE

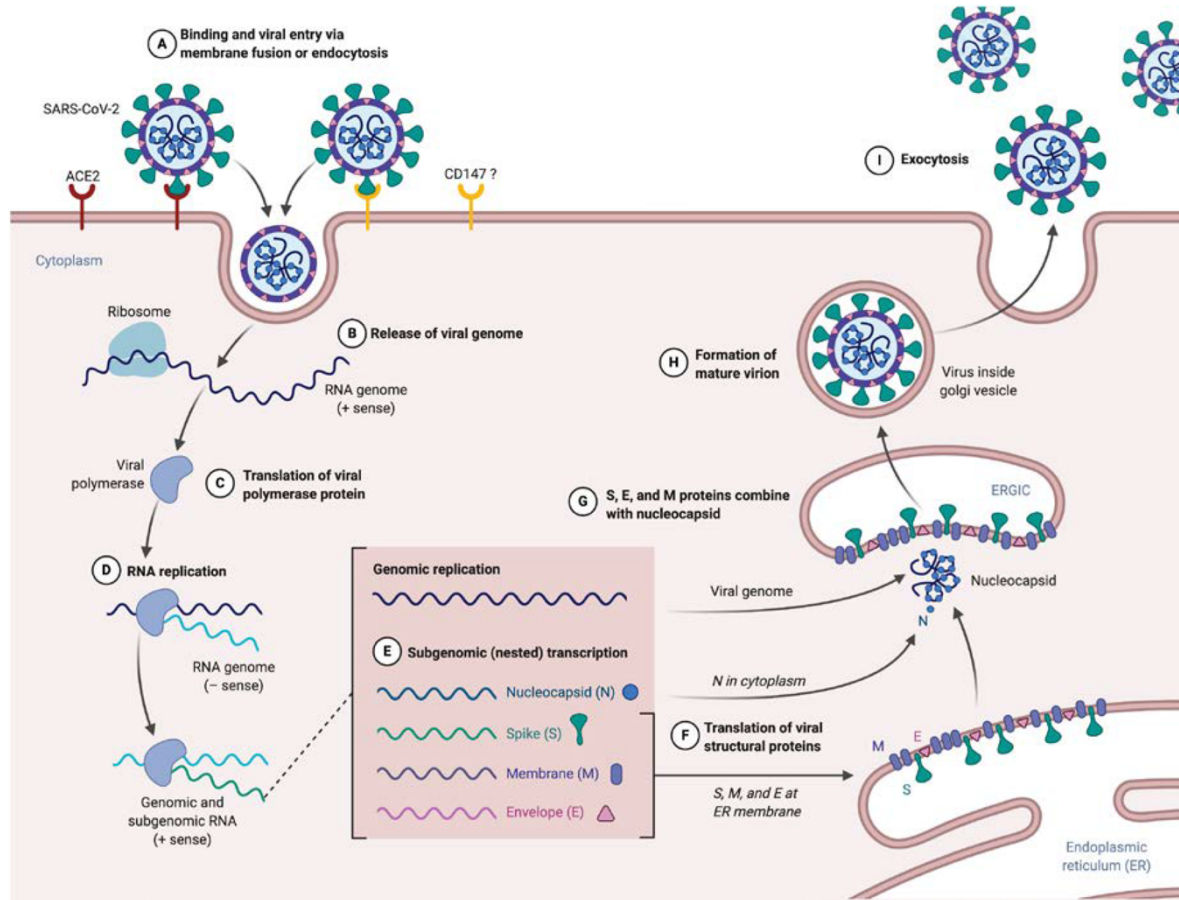
bicouche lipidique renfermant des protéines de 4 types intervenant dans le cycle viral : HE, S, M et E.

Un ARN portant l'information génétique minimale

Grand ARN (29,8 kb) en un brin pouvant être directement traduit.
ARN possédant des repères de l'expression eucaryote.



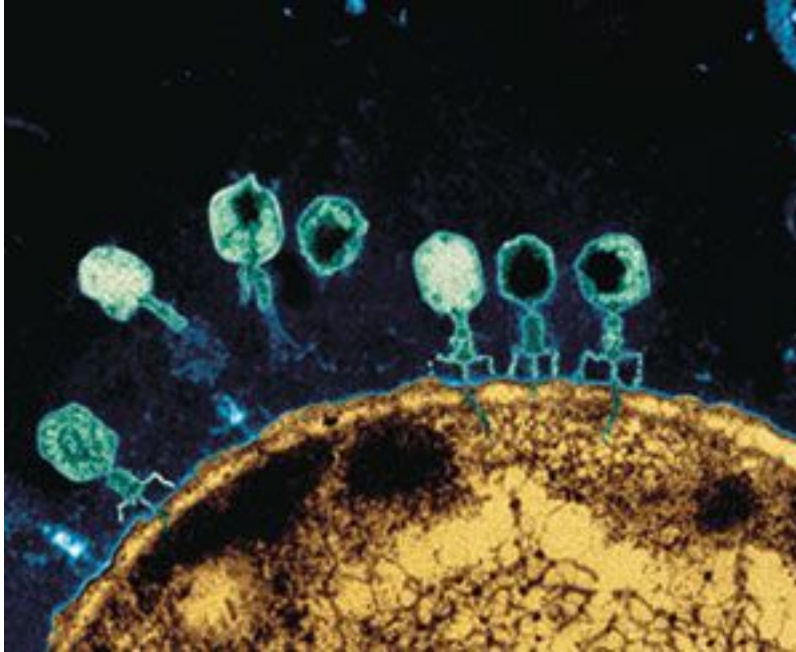
Un cycle viral lytique



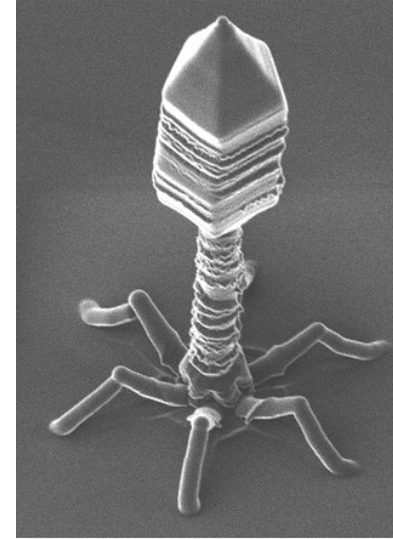
- entrée du virus
- traduction de l'ARN => protéines virales + polymérase
- la polymérase copie l'ARN en ARN complémentaire puis l'ARN complémentaire en nouvel ARN viral
- association des protéines virales et de l'ARN viral au sein du RE
- exocytose de nouveaux virions.

Une diversité virale selon l'acide nucléique

Virus à ADN : exemple des bactériophages

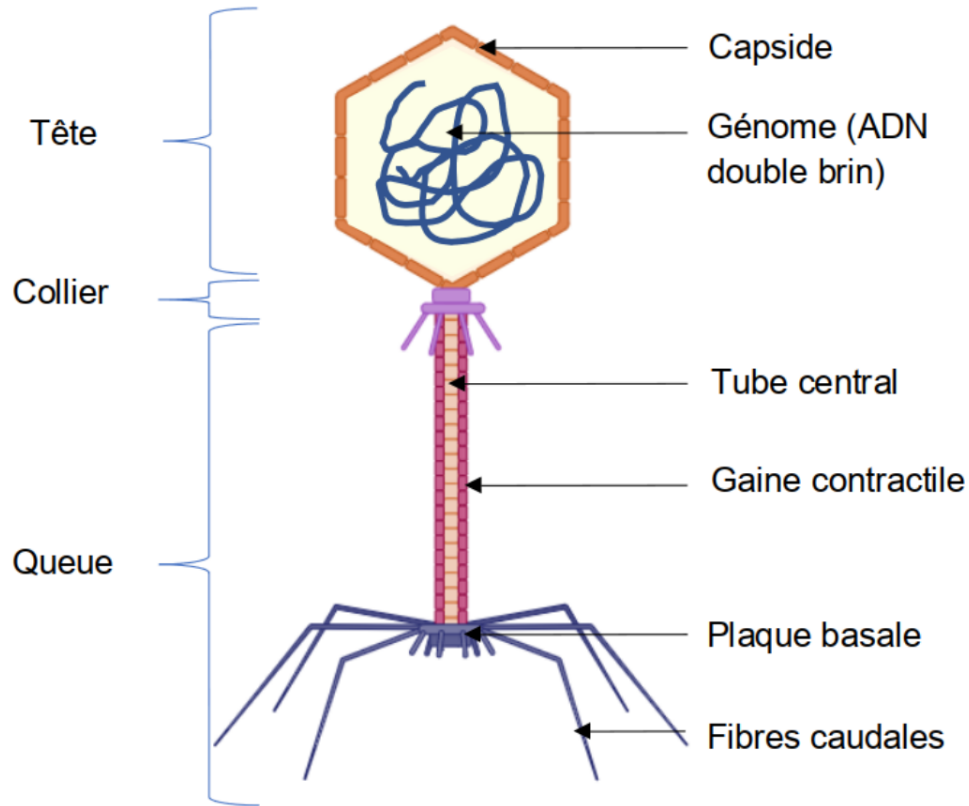


Bactériophages liés à une bactérie
(MET, fausses couleurs)



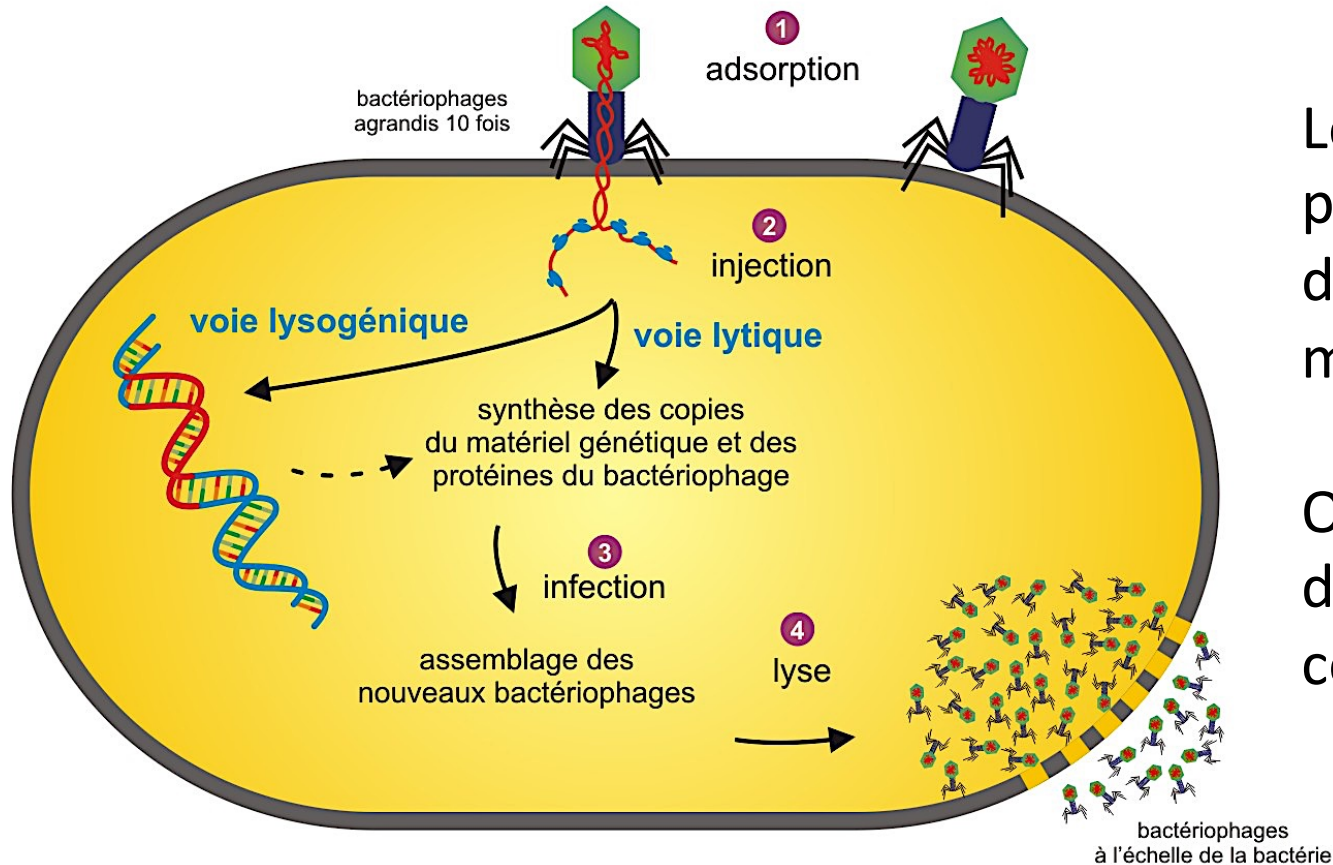
Bactériophage
(MEB)
hauteur 200 nm

Le bactériophage



ADN linéaire double brin
dans une capside protéique

Les cycles viraux

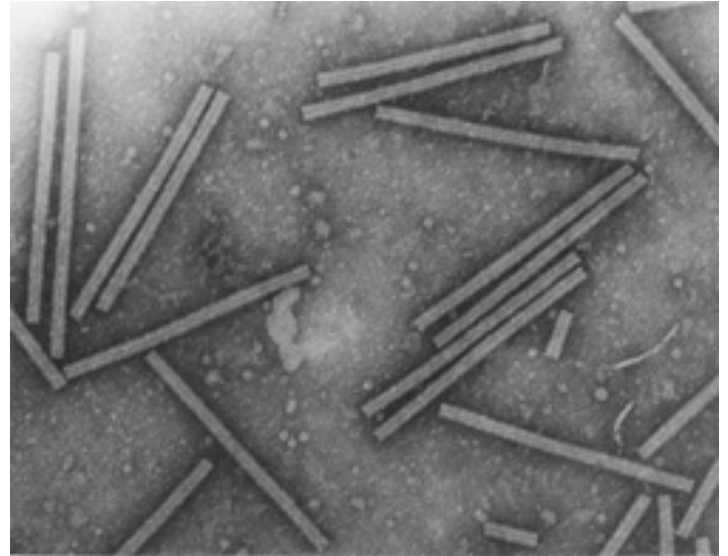
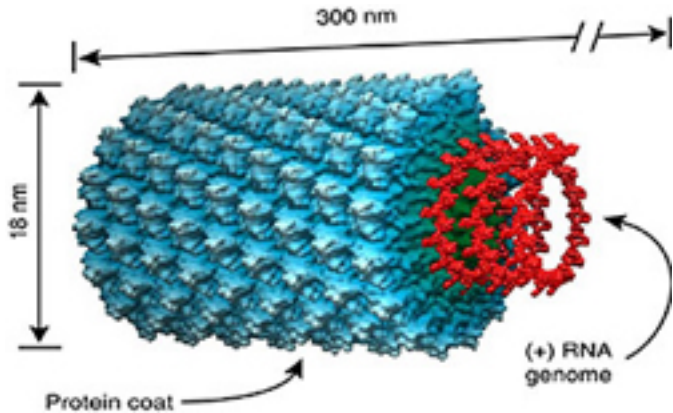


Le virus est multiplié par la cellule hôte dont il a détourné la machinerie cellulaire.

Ou bien il s'intègre dans le génome de la cellule hôte.

Une diversité virale selon l'acide nucléique

Virus à ARN : exemple des virus de mosaïque

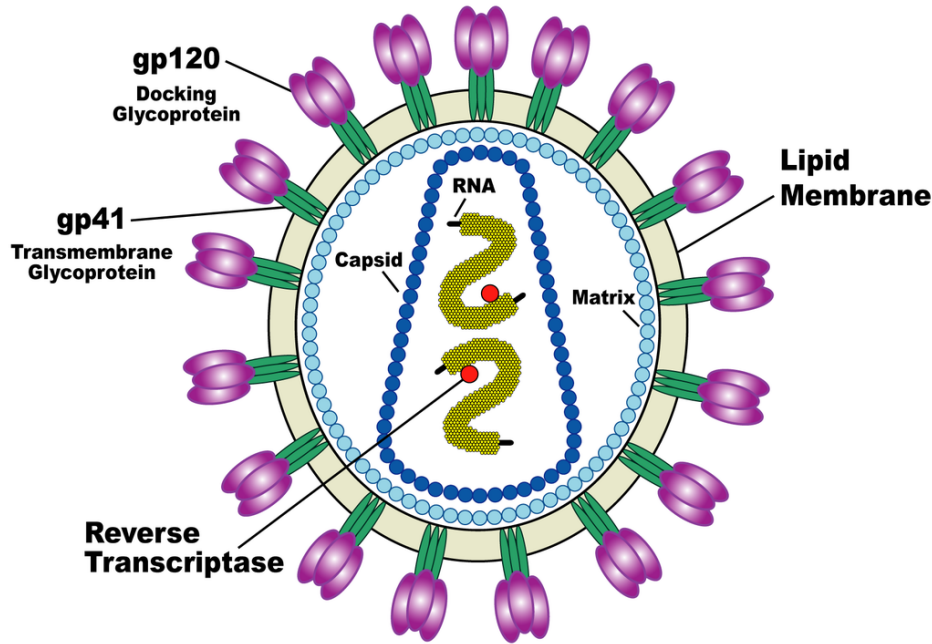


ARN simple brin de 6 400 pb

VMT en long bâtonnet de 300 nm de long

Une diversité virale selon l'acide nucléique

Virus à ARN : exemple des rétrovirus



VIH

- Entrée du virus
- ARN copié en ADN par une reverse-transcriptase
- ADN intégré dans le génome de la cellule hôte
- Synthèse de protéines virales
- Transcription de l'ADN en ARN viral
- Assemblage de nouveaux virions

BILAN

- Virus = **virion** ou **prophage**
- Virion de très petite taille ($< 1 \mu\text{m}$)
- Forme de vie réduite au minimum :
 - un acide nucléique (ADN ou ARN) codant à plus de 90 %
 - une enveloppe (**capside** protéique +/- membrane)
 - parfois des enzymes associées (rétrovirus)
- Pas de métabolisme propre ni de gène d'enzyme métabolique
- Parasite obligatoire

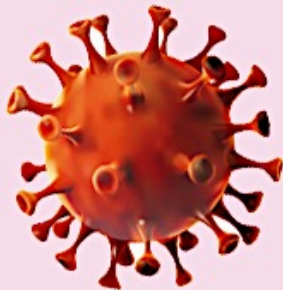
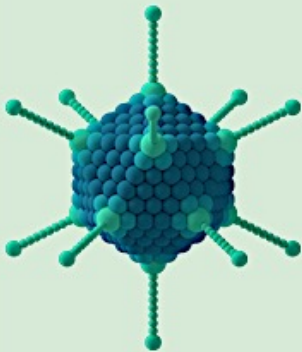
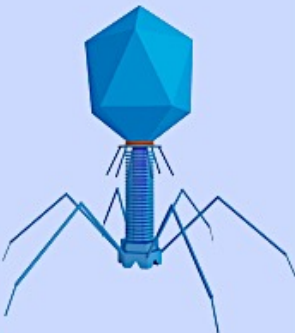
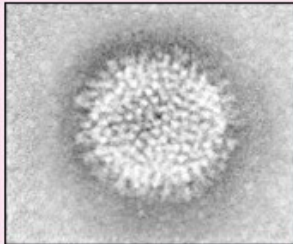
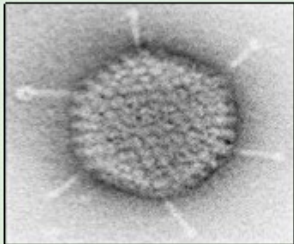
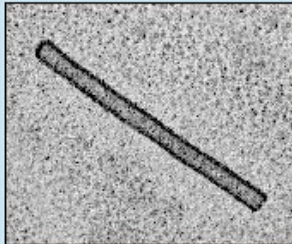
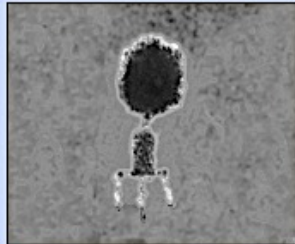
Diversité des virus et de leurs hôtes

Bactéries parasitées par des virus à ADN tels les bactériophages, de formes variées.

Cellules eucaryotes animales parasitées autant par des virus à ADN (papillomavirus) que des virus à ARN (coronavirus, VIH).

Cellules eucaryotes végétales majoritairement parasitées par des virus à ARN : virus des mosaïques.

Diversité des formes de virus

Sphérique	Polyhédrique	Hélice	Complexe
			
			
Coronavirus	Adénovirus	Mosaïque du Tabac	Bactériophage

Adaptation: T.lombry

Source : T. Lombry

CONCLUSION

Acides nucléiques = base du vivant

Notion de macromolécule séquencée portant l'information
Stabilité et compacité de la molécule.

Transmission ?

Expression ?