

ENSEIGNEMENT DE SCIENCES DE LA VIE ET DE LA TERRE (SVT)
°° SCIENCES DE LA VIE °°
>> Cours <<

Chapitre 12

Méthodes d'étude et organisation des génomes

COURS COMPLET RÉDIGÉ

Objectifs : extraits du programme

Savoirs visés	Capacités exigibles
SV-F-1-1 Organisation des génomes L'ensemble des molécules d'ADN contenues dans une cellule et l'information qu'elles portent constitue son génome. L'étude des génomes passe par une panoplie de techniques dites de biologie moléculaire. Des techniques de séquençage permettent de déterminer la séquence d'un fragment d'ADN puis de proche en proche la séquence des génomes. L'utilisation d'outils bioinformatiques permet d'identifier les différents types de séquences codantes et non codantes.	- Réaliser et analyser les résultats d'une électrophorèse d'ADN. - Interpréter l'organisation des génomes à partir des résultats de séquençage. - Exploiter les données de séquençage pour réaliser des alignements de séquences et comparer les séquences.
Précisions et limites : <i>Les principes généraux et les objectifs des différentes techniques évoquées sont à connaître mais les protocoles ne sont pas à mémoriser.</i> <i>Pour le séquençage, seul le principe de la méthode de Sanger doit être connu.</i> <i>La mise en œuvre pratique n'est exigible que pour l'électrophorèse. La maîtrise d'un logiciel d'alignement de séquences n'est pas exigible (la fiche technique du logiciel est fournie).</i>	
Chez les bactéries, le génome à localisation cytoplasmique est constitué d'un chromosome circulaire et éventuellement de plasmides. Le génome des bactéries est constitué presque exclusivement de régions codantes. Certaines sont associées à des régions régulatrices communes ce qui forme des opérons. Chez les Eucaryotes, on distingue le génome nucléaire et le génome des organites. Le génome nucléaire est constitué de chromosomes linéaires. L'ADN nucléaire des Eucaryotes est associé à des protéines dont des histones, constituant la chromatine. Il existe	- Comparer l'organisation du génome des bactéries, des Eucaryotes et des virus. - Comparer le génome cytoplasmique eucaryote et celui des bactéries. - Estimer la proportion de séquences codantes et non codantes dans les génomes des Eucaryotes, des bactéries et des virus.

différents niveaux de condensation de la chromatine. Le génome nucléaire des Eucaryotes comporte une part importante de séquences non codantes aux rôles divers. La majorité de ces séquences est répétée. Les gènes eucaryotes sont généralement morcelés. Les virus ou particules virales sont des entités nucléoprotéiques comprenant un acide nucléique (sous forme d'ADN ou d'ARN) constituant le génome viral, et des protéines. On distingue des protéines à rôle structural, formant la capsid, et parfois des protéines à rôle enzymatique. Les virus sont très divers et possèdent parfois une enveloppe lipoprotéique.	- Illustrer la diversité structurale et la diversité d'hôte des virus.
Précisions et limites : <i>On se limite à mentionner la présence de protéines structurales associées à l'ADN chez les bactéries, sans détailler l'organisation moléculaire du chromosome bactérien. On se limite à la présentation de la structure de l'opéron lactose chez E. coli.</i> <i>Aucune monographie de chaque virus n'est attendue. Il s'agit de montrer la diversité structurale (organisation structurale, taille, présence ou non d'une enveloppe, nature de l'information génétique) et la diversité d'hôte à l'aide de trois exemples, sans rentrer dans les détails des cycles de multiplication : bactériophage lambda, VMT, un coronavirus zoonotique.</i> <i>La connaissance des génomes est remobilisée dans l'étude de l'évolution.</i>	
Liens : Autres chapitres de génétique (SV-F) Compartimentation des cellules spécialisées (SV-C-2) Flux de matière et d'information dans une cellule eucaryote (SV-C-2) Structure des acides nucléiques (SV-D-2-3) Structure des protéines (SV-D-2-4) Glycosylation et phosphorylation des protéines (SV-D-2-4) Expression des gènes homéotiques au cours du développement embryonnaire (SV-H-2)	

Introduction

On appelle **information génétique** ou **patrimoine génétique** *l'ensemble des données qui assurent l'organisation, le développement et le fonctionnement d'un organisme vivant ou d'un virus, se transmettant au fil des générations d'individus.* Cette **information** est portée universellement par de l'ADN dans tous les organismes cellulaires et par de l'ADN ou de l'ARN chez les virus. **L'ensemble de l'ADN d'une cellule ou d'un organisme, ou des acides nucléiques pour un virus** constitue le **génome*** de cette entité, toutes les portions n'étant pas forcément codantes.

Le mot **génome** peut avoir plusieurs définitions selon les auteurs ou le contexte d'emploi :
1/ Il peut désigner **l'ensemble de l'information génétique d'un individu**. Quand on s'intéresse à la **variété interindividuelle**, il sera plus pertinent d'employer la notion de **génotype**.
2/ Il peut désigner **l'ensemble de l'information génétique propre à une espèce**, définition à laquelle nous faisons ici référence. Néanmoins, au sein de cette définition, certains auteurs font référence :
a) À l'ensemble des unités informatives (gènes codant les protéines) seulement.
b) À l'ensemble des unités transcrites ou transcritibles en ARN (y compris les ARN autres que messagers) (gènes au sens large).
c) À la totalité du matériel génétique, qu'il soit codant ou non. *
C'est clairement cette définition qui est aujourd'hui la plus utilisée et à laquelle fait référence le programme.

Comment s'organisent les génomes des cellules et des virus ?

Il est **absolument impératif** de revoir et de maîtriser les notions traitées dans le **chapitre 8** sur les nucléotides pour **comprendre** et **apprendre** le **présent chapitre**. J'ai remis les notions sur les acides nucléiques dans le **chapitre**.

I. La nature et l'étude des génomes

A. Les acides nucléiques, des hétéropolymères (séquencés) de nucléotides porteurs d'une information

- Les **acides nucléiques** sont de **longs polymères de nucléotides** comprenant les **ARN et l'ADN**.
- Ils sont **présents** sous forme **acide** dans les **cellules** (en réalité, c'est la **base conjuguée** au sens de BRØNSTED qui est représentée dans les **cellules**, en lien avec notamment l'**ionisation des phosphates**). Ils sont largement **hydrophiles**.
- Puisque tous les **nucléotides** ne sont **pas identiques**, les **acides nucléiques** sont des **molécules séquencées**, c'est-à-dire **caractérisées par un ordre d'enchaînement de leurs monomères** qu'on appelle ici **séquence nucléotidique**.

1. L'ADN, support universel de l'information génétique des êtres vivants (cellulaires)

- L'**ensemble des êtres vivants** (au sens des **organismes cellulaires**) a pour support de son **information génétique** l'**ADN** qui est un **acide nucléique support de l'information génétique, contenu dans le noyau des Eucaryotes ou le nucléoïde des 'procaryotes**.

a. Une molécule bicaténaire

- L'**ADN** ou **acide désoxyribonucléique** (figures 1-3) (**encadré A**) est un **acide nucléique support de l'information génétique, contenue dans le noyau des Eucaryotes ou le nucléoïde des 'procaryotes**.
- L'**ADN** est une molécule fondamentalement **bicaténaire**, c'est-à-dire qu'il est fait de **deux brins** : rigoureusement, il s'agit donc d'un **double polymère de nucléotides associés par des liaisons hydrogène**.

b. Des brins orientables : extrémités 5' et 3'

- Chaque brin peut être **orienté** : on trouve une **extrémité 5' P** (**carbone 5' associé à un groupement phosphate libre, sans nucléotide derrière**) et une **extrémité 3' OH** (**carbone 3' portant un groupement OH, non associé à un nucléotide**).

c. Structure secondaire et tertiaire : deux brins antiparallèles organisés en double hélice

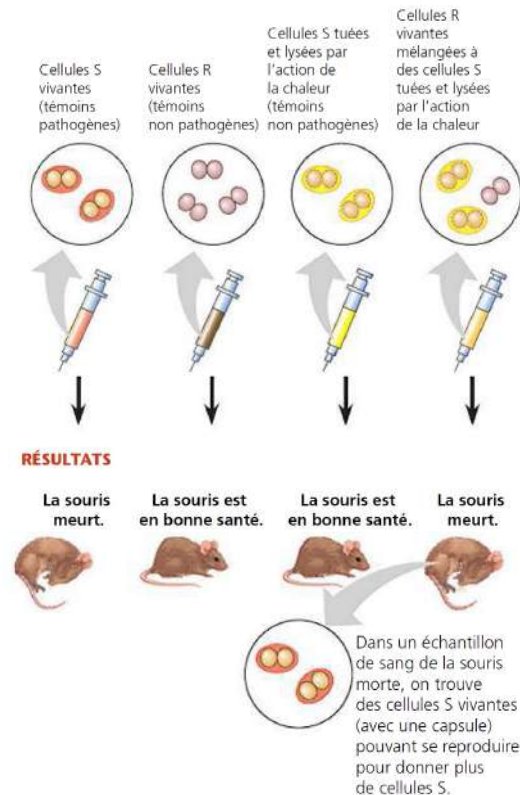
- Pour qualifier la **structure secondaire de l'ADN**, on dit qu'il présente une **structure en double hélice** car il est fait de **deux brins qui s'enroulent en hélice** ; les deux brins sont **antiparallèles**, c'est-à-dire qu'ils sont **orientés de manière inverse l'un par rapport à l'autre** (un brin 3'→5' et l'autre 5'→3'). L'hélice est, dans la forme majoritaire de l'**ADN (ADN-B)**, une hélice **droite**. Un tour de spire est réalisé toutes les **10 paires de bases** environ ; on note la présence, dans l'hélice, d'un **petit sillon** et d'un **grand sillon**.

Encadré A Quelques repères historiques dans l'identification du support de l'information génétique

Mise en évidence d'un facteur « transformant » codant l'information génétique : expériences de Frederik GRIFFITH (1879-1941) (1928)



EXPÉRIENCE Frederick Griffith a étudié deux souches de la bactérie *Streptococcus pneumoniae*. Les bactéries de la souche « S » (pour *smooth* ou lisses) causent la pneumonie chez les souris ; elles sont pathogènes parce qu'une capsule les protège contre le système immunitaire des Animaux. Les bactéries de la souche « R » (pour *rough* ou rugueuses) sont dépourvues de capsule et ne sont pas pathogènes. Afin de vérifier le pouvoir pathogène de ces deux souches, Frederick Griffith les a inoculées à des souris :



CONCLUSION Griffith en a conclu que les bactéries vivantes de la souche R ont été transformées en bactéries pathogènes de la souche S par une substance inconnue provenant des cellules mortes de la souche S. Cette substance a permis aux cellules de la souche R de fabriquer des capsules.

FIGURE a. L'expérience de GRIFFITH (1928). D'après CAMPBELL *et al.* (2012) et Wikipédia (cliché).

Identification de l'ADN comme support de l'IG : AVERY, MACLEOD & McCARTY (1944), HERSHEY & CHASE (1952)

➤ Si GRIFFITHS (1928) a pu montrer l'existence d'un **facteur porteur de l'information génétique**, un **débat** a longtemps existé pour savoir s'il s'agissait de **protéines** ou d'**acides nucléiques**. Le **travail d'Oswald AVERY, Colin M. MACLEOD & Maclyn McCARTY (1944)** (**figure b**) permet de montrer qu'il s'agit de l'ADN mais il faudra attendre l'**expérience d'Alfred HERSHEY & Martha CHASE (1952)** pour que cette réalité soit confirmée et admise par l'ensemble de la communauté scientifique (**figures c-d**).

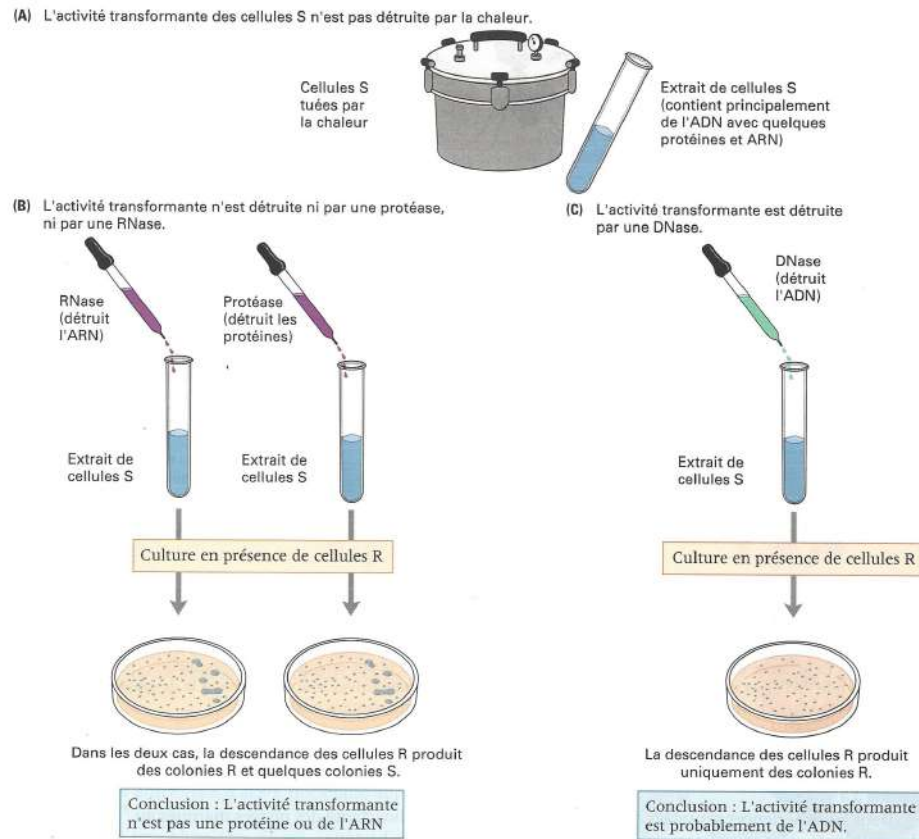
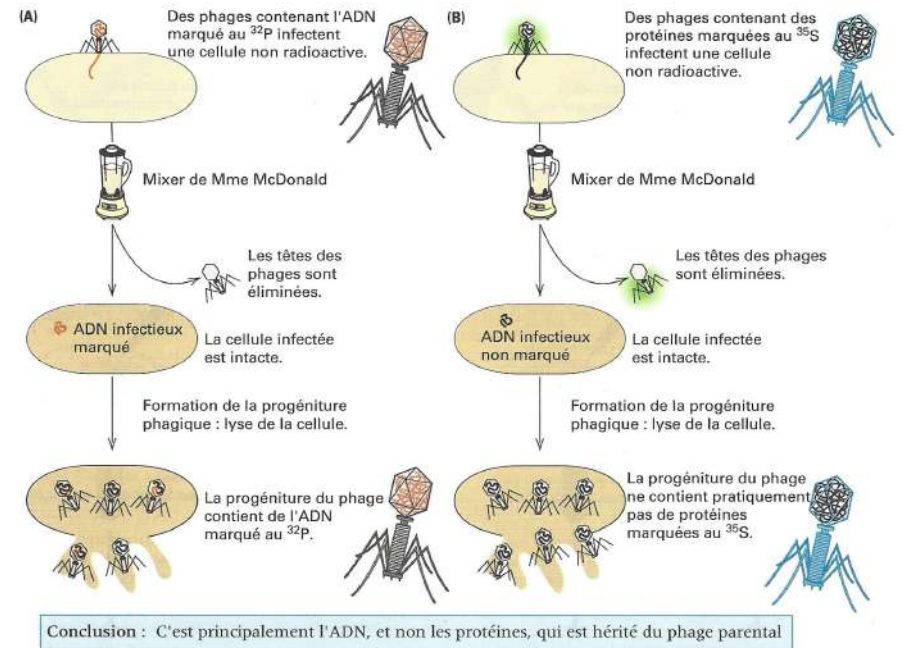


Schéma de l'expérience démontrant que l'ADN est le matériel actif dans la transformation bactérienne. (A) De l'ADN purifié extrait de cellules S tuées par la chaleur peut convertir certaines cellules vivantes R en cellules S, mais l'extrait pourrait encore contenir des traces indé-

tables de protéines et/ou d'ARN. (B) L'activité transformante n'est détruite ni par une protéase, ni par une RNase. (C) L'activité transformante est détruite par une DNase et repose donc probablement sur la présence d'ADN.

FIGURE b. Expérience d' Avery et al. (1944). D'après HARTL & JONES (2003).



Expérience (« mixer ») de Hershey-Chase, qui démontre que l'ADN, et non les protéines, est chargé de diriger la reproduction du phage T2 dans des cellules *E. coli* infectées. (A) L'ADN radioactif est transmis à la descen-

dance du phage en quantités substantielles. (B) Les protéines radioactives sont transmises à la descendance du phage en quantités négligeables.

FIGURE c. Expérience d'HERSHEY & CHASE (1952). D'après HARTL & JONES (2013).



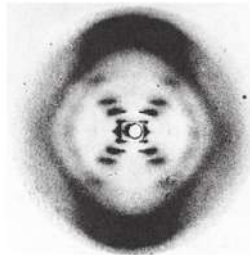
FIGURE d. Martha CHASE (1927-2003) et Alfred HERSHEY (1908-1997). <https://sites.google.com/site/dd95hersheyCHASEexperiment/> (octobre 2015)

Identification de la structure de l'ADN (F. CRICK & J. WATSON, 1953)

➤ La **structure de l'ADN** est élucidée en 1953 par **Francis CRICK (1916-2004)** & **James D. WATSON (1928-2025)** à partir d'un cliché de diffraction aux rayons X montré par **Maurice WILKINS (1916-2004)** ; ces trois personnes recevront le **prix de Nobel de physiologie-médecine en 1962**. L'histoire a longtemps **oublié Rosalind FRANKLIN (1920-1958)** qui a pourtant **produit le si précieux cliché** ayant permis cette découverte (figure e).



(a) Rosalind Franklin



(b) Radiographie de l'ADN par diffraction de rayons X produite par Rosalind Franklin

Rosalind Franklin et sa radiographie de l'ADN par diffraction de rayons X. Franklin, une spécialiste très douée en cristallographie par diffraction de rayons X, a effectué des expériences remarquables et produit la radiographie grâce à laquelle Watson et Crick ont pu découvrir la structure en double hélice de l'ADN.

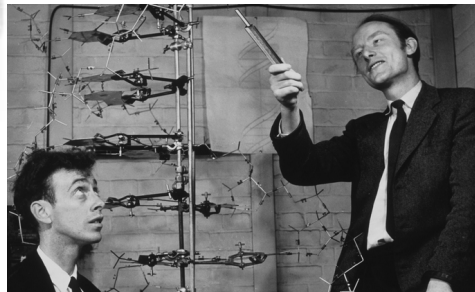


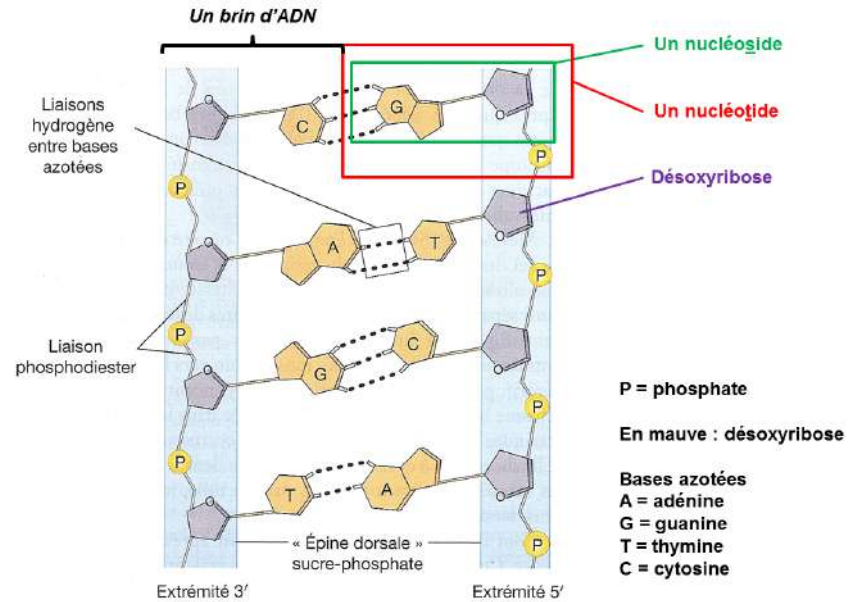
FIGURE e. Les artisans de la découverte de la structure de l'ADN : Rosalind E. FRANKLIN (1920-1958), James D. WATSON (1928) et Francis H. C. CRICK (1916-2004).

D'après CAMPBELL et al. (2012) et Wikipédia.

Importance des liaisons faibles dans la structure de l'ADN :

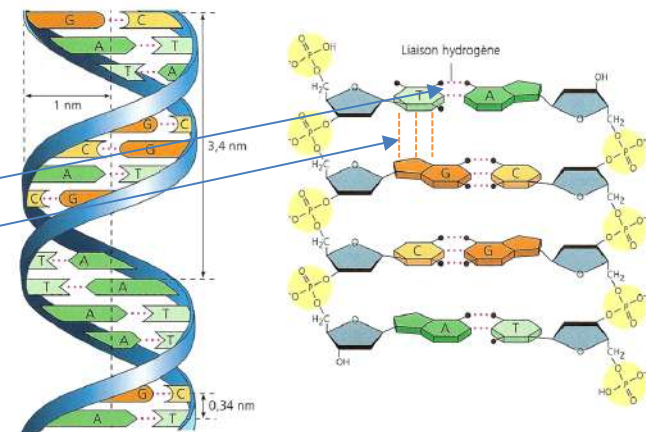
- **liaisons hydrogène** entre **bases appariées** (→ association des deux brins avec possibilité d'ouverture et de fermeture de la molécule d'ADN)
- **interactions hydrophobes** (**liaisons de VAN DER WAALS** de type **forces de LONDON**) entre les **bases empilées** verticalement (→ **cohésion** interne, autorisant toutefois une **déformabilité** et l'**hélicoïdalité**) : c'est le **base stacking** (empilement de bases).

Note : la répulsion mutuelle des squelettes ose-phosphate semble neutralisée par des cations (Mg^{2+}) et les protéines basiques (chargées positivement) s'y associant (ex. histones des Eucaryotes).



Structure et composition de l'ADN.

D'après RAVEN et al. (2007)

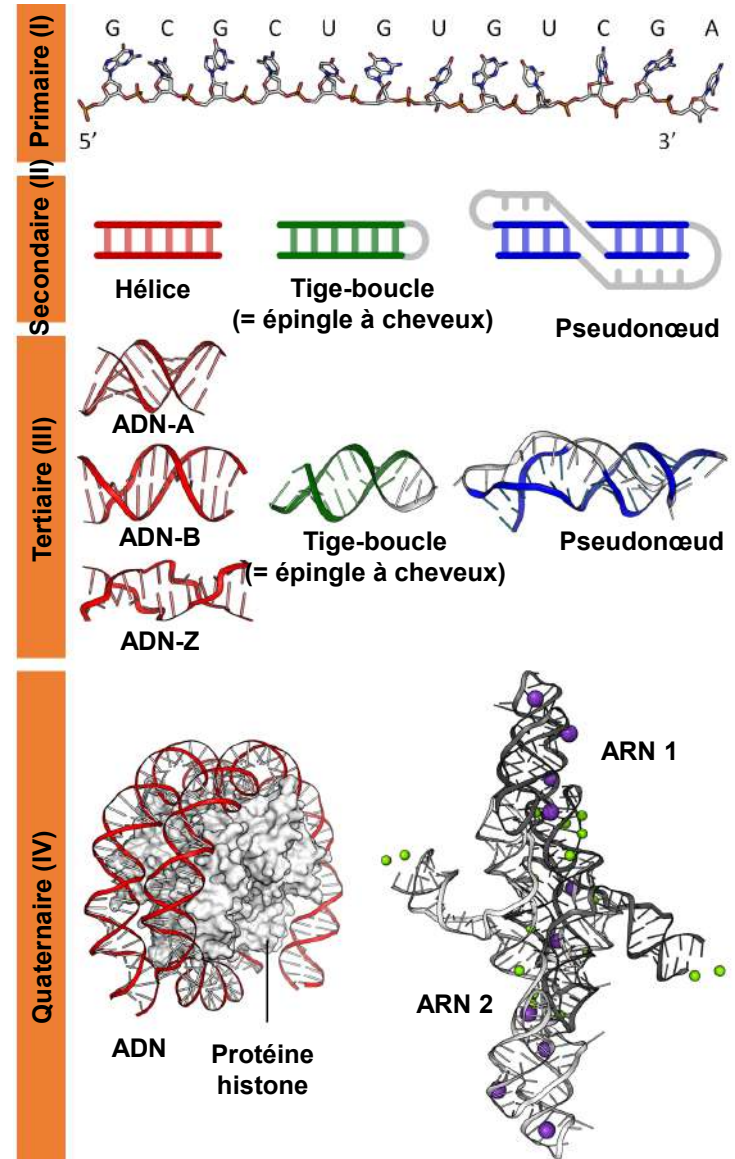


Autre vision de la constitution de l'ADN (à droite) et illustration de la structure en double hélice (à gauche).

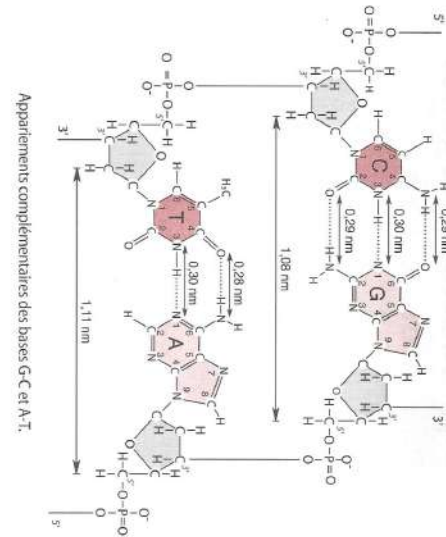
D'après CAMPBELL & REECE (2004)

▲ **FIGURE 1. Organisation de l'ADN.** Sources variées.

Encadré B Structure primaire, secondaire, tertiaire et quaternaire des acides nucléiques



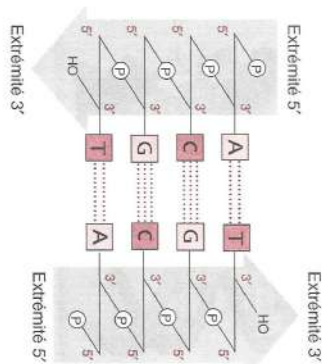
▲ FIGURE. Les niveaux de structure des acides nucléiques. Wikipédia (traduit 2021).



COMPOSITION EN BASES DE L'ADN DE DIFFÉRENTES ORIGINES.

Origine de l'ADN	Bases (%)				(A+T)/(G+C)	(A+G)/(T+C)
	A	G	C	T		
Bactériophage T7	26,0	23,8	23,6	26,6	1,11	0,99
Escherichia coli B	23,8	26,8	26,6	23,1	0,88	1,01
Neurospora	23,0	27,1	26,6	23,3	0,86	1,00
Drosophile	30,7	19,6	20,2	29,5	1,51	1,01
Saumon	28,0	22,0	20,0	27,8	1,33	1,05
Poule	28,0	22,0	21,6	28,4	1,29	1,00
Rat	28,6	21,4	21,6	28,4	1,33	1,00
Vache	27,3	22,5	22,5	27,7	1,26	0,99
Homme	29,3	20,7	20,0	30,0	1,46	1,00

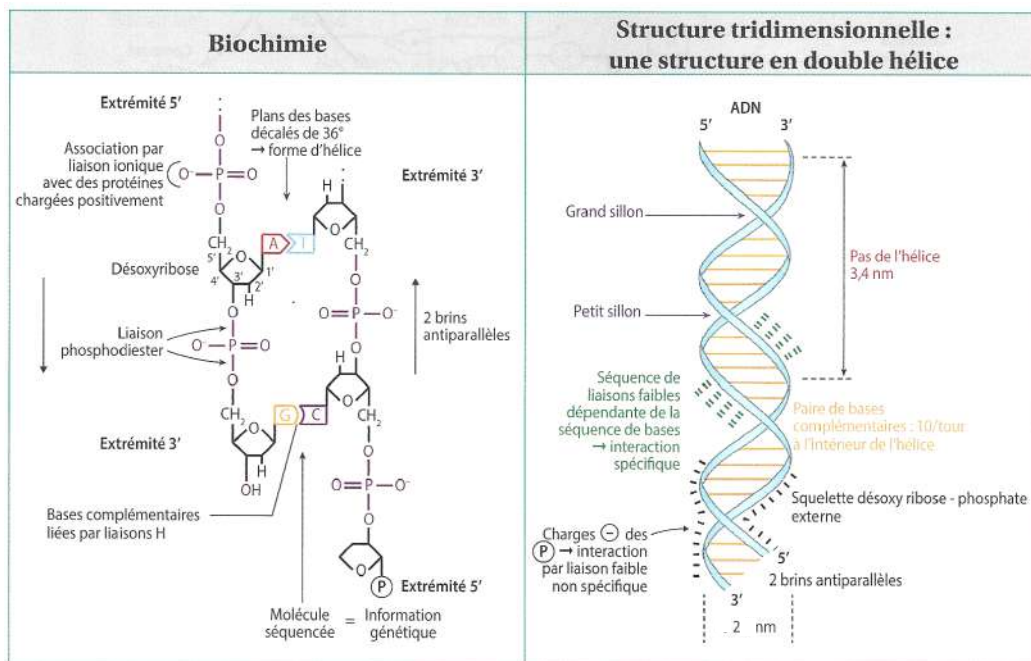
b)



Structure des appariements complémentaires de deux brins antiparallèles d'ADN. a) représentation développée; b) représentation simplifiée.

▲ FIGURE 2. Organisation de l'ADN : une autre vision. D'après PETIT & JULIEN (2007).

- Dans un **acide nucléique**, on peut définir les **niveaux de structure** suivants (figure) :
- La **structure primaire** : elle désigne le **polymère linéaire de nucléotides** et est définie par l'**ordre d'enchaînement des nucléotides au sein de l'acide nucléique (séquence nucléotidique)**. [« 1D »]
 - La **structure secondaire** : elle désigne la **bicaténarisation (→ 2 brins) locale des nucléotides sous forme d'hélice, d'épingles à cheveux ou de pseudonœud**. [« 2D »].
 - La **structure tertiaire** : elle désigne la **forme tridimensionnelle des acides nucléiques (incluant leurs structures secondaires), c'est-à-dire leur conformation réelle dans l'espace**. [3D]
- Il existe **plusieurs formes d'ADN** : l'**ADN-B**, la plus courante naturellement, et des **formes plus rares**, l'**ADN-A** et l'**ADN-Z**, rencontrées *in vitro* ou dans des **zones particulières du génome**. Elles diffèrent par la **conformation spatiale tridimensionnelle** de la **double hélice**, en lien avec **diverses caractéristiques propres des nucléotides**, l'état d'**hydratation** et l'état d'**enroulement**. Le **programme** limite l'étude de l'ADN à la **forme B**.
- La **structure quaternaire** : elle désigne **l'association de l'acide nucléique avec un autre ou d'autres acides nucléiques, ou avec une ou des protéines**.



▲ FIGURE 3. L'ADN : relations structure / propriétés chimiques / fonction(s).

D'après DAUTEL *et al.* (2021), modifié

[l'histoire des sillons n'est du reste pas très claire sur ce document]

d. Une molécule qui respecte les règles de CHARGAFF

- Les brins sont associés par des **liaisons H** entre bases azotées des nucléotides en vis-à-vis dans les brins, toujours de la même façon : **A en face de T (2 liaisons H)**, **C en face de G (3 liaisons H)**. Ce sont les **règles de complémentarité** entre nucléotides ou **règles d'appariement exclusif** des nucléotides.

Rappelons que les **liaisons hydrogène** sont des **liaisons faibles** (contrairement aux liaisons phosphodiester de nature covalente). Ces liaisons sont donc plus **labiles** et plus **faciles à rompre** que des liaisons covalentes, ce qui **permet à la molécule d'être aisément ouverte et refermée**, notamment lors de l'**expression génétique (transcription)** ou lors de la **réplication**.

- Proposées par le biologiste autrichien Erwin CHARGAFF (1905-2002), les **règles de CHARGAFF** sont les suivantes :
 - 1^{re} règle** : Une molécule d'ADN est caractérisée par une quantité équivalente de **thymine et d'adénine**, et une quantité équivalente de **guanine et de cytosine** : $A/T = C/G = 1$.
- CHARGAFF découvrait donc « en avance » (en 1950) ce qu'on baptisera les **règles de complémentarité** entre nucléotides, quoique la structure de l'ADN n'ait pas encore été élucidée et qu'il fût donc alors **impossible, à l'époque, de penser les choses en ces termes**.
- 2^e règle** : Chaque espèce est caractérisée par un rapport $(A+T)/(C+G)$ caractéristique (figure 2 : tableau).

Cela ne présage rien des séquences nucléotidiques possibles.

e. L'ADN, une molécule séquencée capable de porter une information

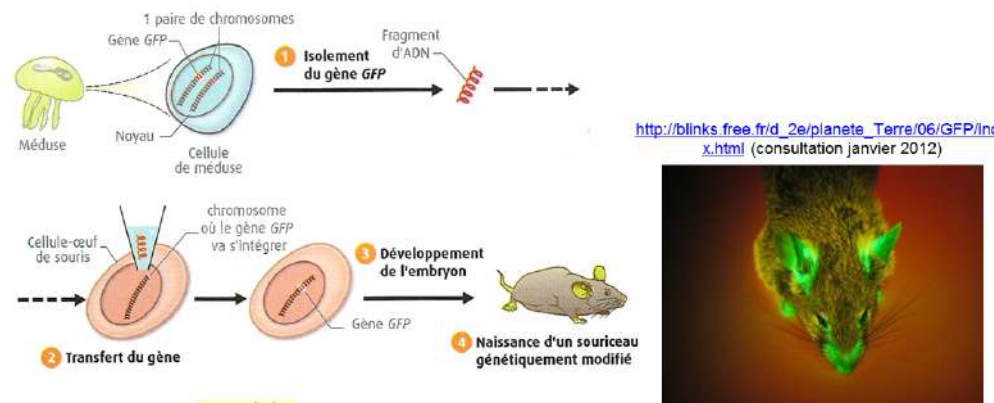
a. La capacité de porter une information : approche intuitive

- L'ADN est un **hétéropolymère** constitué de l'enchaînement de **nucléotides variables**, de **quatre types** différents. Cette molécule est donc **séquencée**, c'est-à-dire définie par la **nature et l'ordre des nucléotides qui la composent (séquence nucléotidique)**.
- Le caractère séquencé de l'ADN permet le **codage**, dans une sorte d'alphabet à **quatre lettres**, d'une information qui est l'**information génétique** de l'individu (ensemble des informations nécessaires au fonctionnement et au développement de l'individu) : **la séquence nucléotidique de l'ADN (ou de l'ARNm) code ainsi la séquence peptidique de la protéine concernée**.

Voir plus bas l'organisation des **séquences codantes** et les parties correspondantes du programme

β. Le caractère universel du codage de l'information génétique : mise en évidence par la transgénèse

- En laboratoire, il est possible d'effectuer le **transfert d'un gène (portion d'ADN) d'un organisme donneur à un organisme receveur appartenant généralement à une espèce différente** : c'est la **transgénèse** (figure 4).
Les principes de la transgénèse sont abordés dans le B.
- Sauf difficultés techniques (et il y en a !), **le caractère codé par le gène s'exprime alors dans l'organisme receveur**, même s'il n'est pas de la même espèce que l'organisme donneur : l'**information génétique est donc écrite dans un langage universel, à savoir l'ADN**.



Une expérience de transgénèse.
La transgénèse est une technique qui permet de transférer, au laboratoire, un fragment d'ADN d'un organisme à un autre. Ici, le fragment d'ADN transféré est le gène déterminant le caractère héréditaire « fluorescence de couleur verte » (gène *GFP*) chez la méduse *Aequorea victoria*. L'organisme obtenu est dit génétiquement modifié : c'est un OGM.

D'après Manuel Seconde Belin 2010.

▲ FIGURE 4. Transgénèse et universalité de la molécule d'ADN.

f. Un codage par triplets de nucléotides de l'ADN (= codons dans l'ARNm) : le code génétique

α. Un système de correspondance (quasi-) universel entre séquence nucléotidique et séquence peptidique

- Parmi les **portions codantes de l'ADN**, l'information génétique s'organise en **triplets de nucléotides (ensembles de 3 nucléotides)** qui équivalent à des **codons** sur l'ARN messager, lesquels seront lus par les **ribosomes** lors de la **traduction** et convertis en **acides aminés**. On appelle le **système de correspondance entre acides aminés et codons de l'ARN (ou triplets de l'ADN)** le **code génétique** (figure 5). Ce code est le même pour tous les êtres vivants (sauf chez quelques rares organismes qui parfois ont un ou deux codons qui diffèrent du code génétique classique) : c'est l'**universalité** du code génétique.

β. Un code ponctué : des codons initiateurs et terminateurs de la traduction

- Il existe un **codon initiateur** ou **codon start** (capable d'initier la traduction – qui code également l'AA méthionine : AUG dans l'ARNm) et des **codons stop (codons terminateurs)** (qui **arrêtent la synthèse protéique** : UAA, UAG, UGA dans l'ARNm).

γ. Un code caractérisé par sa redondance (ou dégénérescence)

- Comme il existe **64 codons possibles (4³)** mais seulement **20 acides aminés protéinogènes**, il y a des **codons qui codent le même acide aminé** : on dit que le code génétique est **redondant** ou **dégénéré**.

Cela explique l'existence de **mutations silencieuses** : **quand une mutation modifie un codon en un codon équivalent, la mutation ne modifie pas la séquence peptidique de la protéine codée**. Attention, les mutations silencieuses ne sont qu'un cas particulier de mutations neutres.

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

▲ FIGURE 5. Le code génétique (exprimé, comme presque toujours, par rapport à l'ARNm).

D'après http://www.lemm.univ-lille1.fr/biologie/biocellulaire/apprendre/chapitre9/ch9_page1.htm (octobre 2015).

g. La notion d'ORF, cadre ouvert de lecture

- La notion d'ORF est très présente dans la littérature scientifique mais pas si facile à conceptualiser parce qu'elle est rarement définie et que plusieurs définitions ont pu ou peuvent encore coexister. On peut définir :
 - Le **cadre de lecture (reading frame)** comme la **division d'une séquence nucléotidique en triplets non-recouvrants et subséquents (= qui se suivent immédiatement) de nucléotides**, base d'une possible lecture de la séquence lors de l'expression génétique.
 - Un **cadre ouvert de lecture** ou plus souvent **ORF (open reading frame)** de deux manières principales qui semblent encore coexister :
 - 1/ **Usage historique** (mais parfois encore observé, notamment en annotation de génomes) : il s'agit d'une **partie de cadre de lecture (donc d'une séquence découpée en triplets) susceptible d'être traduite en protéine**. Dans ce cas, les **introns** (s'il y en a) sont **inclus**.
 - 2/ **Usage plus récent** : il s'agit de la **partie d'un cadre de lecture effectivement codante et traduite en protéine allant du triplet start ou triplet stop (ce dernier étant généralement inclus)** ; cette dernière acception peut être pragmatiquement mise en **synonymie** du terme **cistron** ou encore **séquence codante**.

Il s'agissait là d'un **outil « pragmatique »** lors des **grands chantiers de séquençage de génomes** permettant d'**inférer la position de gènes potentiels**. ORFfinder (<https://www.ncbi.nlm.nih.gov/orffinder/>) emploie cette définition.

- Usage plus récent : il s'agit de la **partie d'un cadre de lecture effectivement codante et traduite en protéine allant du triplet start ou triplet stop (ce dernier étant généralement inclus)** ; cette dernière acception peut être pragmatiquement mise en **synonymie** du terme **cistron** ou encore **séquence codante**.

Évidemment, dans le cas de **gènes morcelés** (cas typique des **Eucaryotes**), l'**ORF peut s'étendre sur plusieurs exons**.

Dans ce second cas, **la plupart des définitions de l'ORF excluent alors clairement les introns** (et a fortiori les séquences régulatrices, les régions-exoniques non-traduites (UTR), etc.).

h. L'ADN, une molécule en interaction avec des protéines qui s'exprime, se conserve, se transmet et peut varier

- En interaction avec les **constituants alentour**, notamment des **protéines**, et selon des **modalités qui présentent des similitudes** mais **peuvent tout de même varier** entre les **Eubactéries** et les **Eucaryotes**, l'ADN est une molécule :
 - Qui **s'exprime**, notamment en **protéines**, par le biais de la **transcription** en **ARN** et la **coopération des différents types d'ARN**.
 - Dont l'**expression** est soumise à **régulation** en fonction des besoins de la cellule ou de l'organisme, en lien avec des signaux extérieurs.
 - Qui se **conserve** au travers de la **réplication** et se transmet au travers des **divisions cellulaires**.
 - Qui se peut subir des **mutations**, des **réparations post-mutationnelles** et une **variation** au fil des générations par les mécanismes de **recombinaison** (**méiose**, **parasexualité bactérienne**...).

Voir les items du programme de SVT concernés.

Parmi les **protéines associées à l'ADN**, on peut en distinguer caricaturalement deux types : les **protéines de travail** ou **protéines « fonctionnelles »** qui **assurent le dynamisme du génome** (**expression, réplication, régulation... bref, l'ensemble des « activités génétiques »**) et les **protéines de structure** qui **permettent de structurer la molécule et assurent sa condensation**.

L'interaction avec les **protéines** est :

- En **partie non spécifique** (certaines protéines, ou certaines parties de protéines données), en lien avec le **squelette ose-phosphate** qui est **régulier et riche en charges négatives** (dus aux **phosphates**), assurant la **liaison entre l'ADN et des protéines nucléaires (ou nucléoidales) chargées positivement**.
- En **partie spécifique** (certaines protéines, ou certaines parties de protéines données), en conséquence de la **séquence nucléotidique** et des **liaisons faibles** qui en découlent et qui permettent d'établir **spécifiquement une association avec certains acides aminés**.

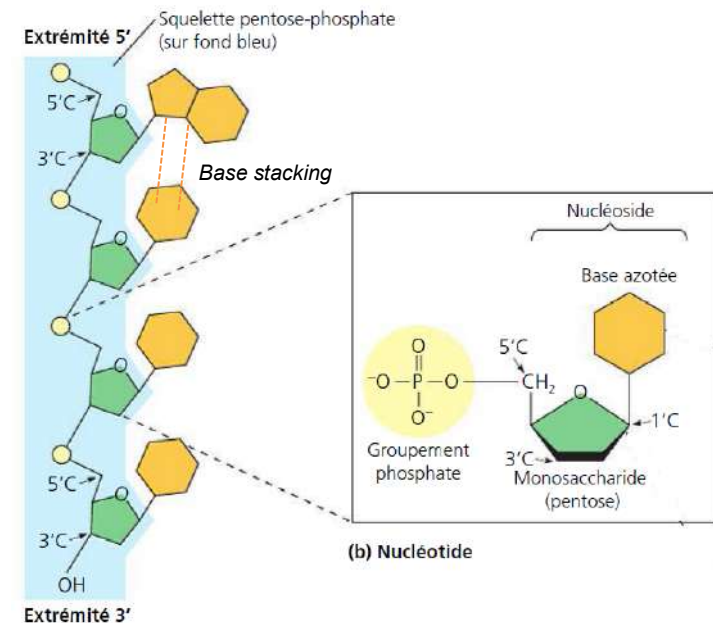
2. Les ARN, des copies plus ou moins transitoires de portions d'ADN concourant à l'expression génétique (ou supports de l'information chez certains virus)

a. Les ARN, des acides nucléiques plutôt monocaténaire (pouvant se bicaténariser) constituant des copies de petites portions d'ADN

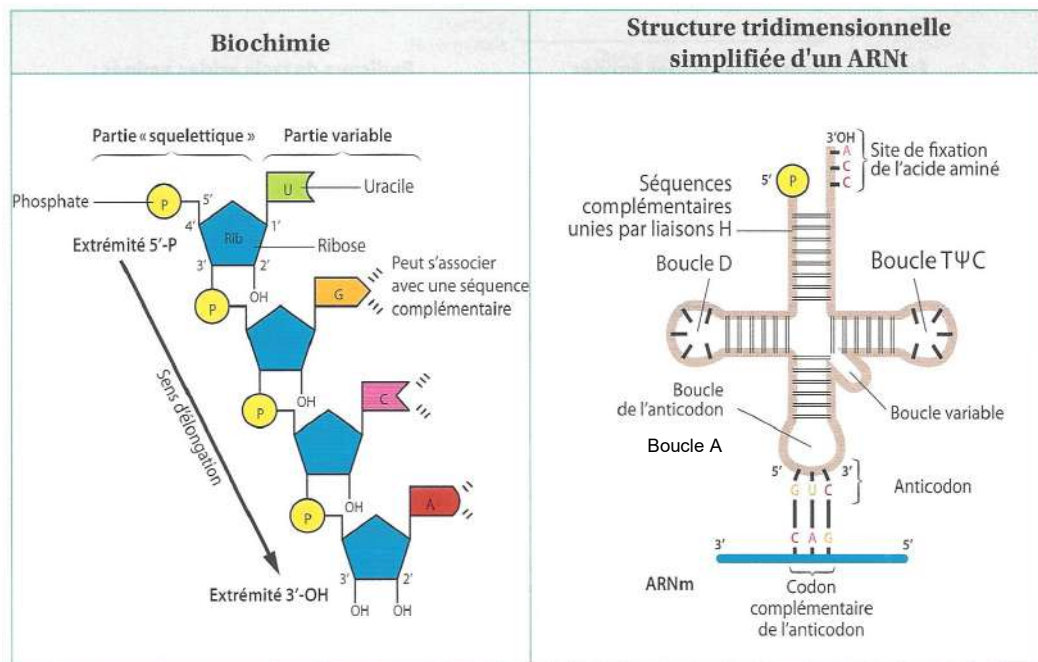
- Un **ARN** ou **acide ribonucléique** est un **polymère généralement monocaténaire de nucléotides qui participe à l'expression de l'information génétique ou son contrôle** (figures 6-7). Ce peut aussi être un **support de l'information génétique chez certains virus**.
- Les **nucléotides de l'ARN (ribonucléotides)** présentent des différences avec les **nucléotides de l'ADN (désoxyribonucléotides)** :
 - Le **pentose présent dans ces nucléotides** est le **ribose** (et non le **désoxyribose**).
 - Les **nucléotides à thymine (T) n'existent pas** ; ils sont remplacés par les **nucléotides à uracile (U)**.

- Les **règles d'appariement** entre un **ARN** et un **autre ARN**, ou bien entre un **ARN** et un **brin d'ADN** sont **les mêmes qu'entre les brins de l'ADN** : la seule différence est que **U remplace T** (on aura donc un appariement **A avec U**).
- L'ARN est fondamentalement **monocaténaire (un seul brin de nucléotides)** mais, *chez les virus*, on peut trouver des **ARN bicaténaires (comprenant deux brins)**. Du reste, **beaucoup d'ARN d'êtres vivants se replient sur eux-mêmes par endroits** et sont **« localement » bicaténaires** (exemple : **ARNt**).
- Les **ARN** permettent l'**expression de l'information génétique** chez les **êtres vivants**, mais certains **virus** ont pour **support de leur information génétique** un ou des **ARN**. *Là encore, cette situation est propre aux virus et ne se rencontre pas chez les êtres vivants.*

- Les **ARN** sont **produits à partir d'une portion d'ADN** par un processus nommé **transcription**.



▲ FIGURE 6. **Organisation de base d'un ARN.** D'après CAMPBELL et al. (2012).



▲ FIGURE 7. Les ARN avec un focus sur l'ARNt : relations structure / propriétés chimiques / fonction(s). D'après DAUTEL *et al.* (2021)

b. Des molécules variées participant à l'expression génétique

- Il existe **différents types d'ARN** que l'on peut distinguer selon différents critères. Le **tableau I** propose une introduction à cette diversité ; nous détaillerons ces aspects dans le **chapitre consacré à l'expression génétique**, fonction à laquelle coopèrent les **ARN**.

c. Étude sommaire d'un exemple : un ARN de transfert (ARNt)

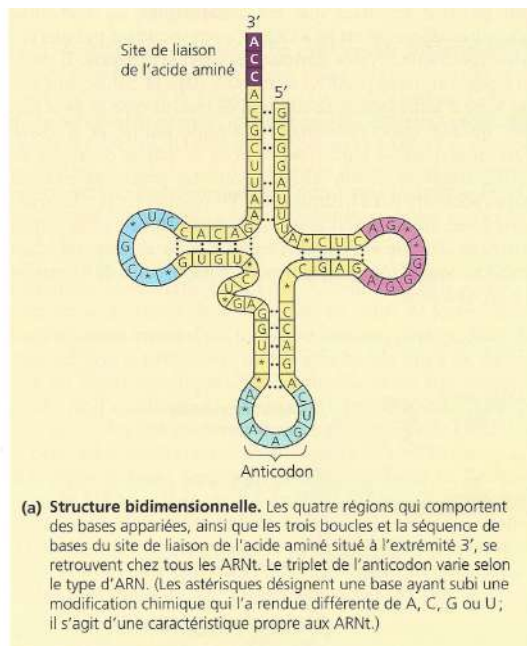
- Les **ARN de transfert (= ARNt)** (figures 8-10) sont les **ARN assurant l'acheminement des acides aminés vers le ribosome au moment de la traduction**. Ce sont des **petits ARN (75-100 nucléotides)** repliés dans l'espace et ayant une forme de **feuille de trèfle** ; ce repliement est notamment assuré par des **appariements de bases complémentaires**, ce qui rend la molécule **localement bicaténaire**, grâce à la formation de boucles en **épingles à cheveux** (tiges-boucles).

▼ TABLEAU I. La diversité des ARN. D'après SEGARRA *et al.* (2014).

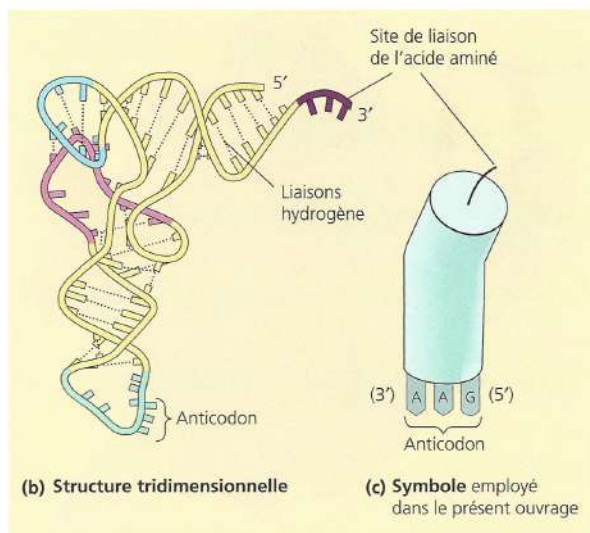
	ARNm	ARNt	ARNr	Petits ARN
Proportion	5 %	15 %	80 %	<1%
Durée de vie	Généralement courte (quelques minutes à quelques heures)	Longue	Longue	Très variable
Taille	Très variable : quelques centaines à quelques millions de nucléotides	75 à 100 nucléotides	100 à 2000 nt, taille exprimée en unité Svedberg : • eucaryotes : 28,18, 5,8 et 5S • eubactéries : 23, 16 et 5S	< 1000 nt
Structure	Linéaire, quelques structures en épingles à cheveux parfois.	Structure secondaire en feuille de trèfle, et structure tertiaire en L.	Complexe, avec association aux protéines ribosomales.	Complexe, association à de nombreuses protéines
Fonctions	« Copie de travail » de l'ADN, support manipulable par les ribosomes de l'information génétique	Adaptateur entre ARNm et acides aminés. Liaison à un acide aminé en 3'	- Structural pour la mise en forme des ribosomes. - Catalytique pour la transpeptidation (notion de ribozymes). - Positionnement du ribosome avec la séquence de Shine et Dalgarno pour eubactéries	- Catalytique (ARN de la télomérase). - Catalytique et structural pour le spliceosome. - Contrôle de l'expression de l'information génétique (ARN interférence)

- L'**extrémité 3'** (située au niveau du « pétiole ») est capable de se fixer à un **acide aminé spécifique**. Chaque **ARNt** présente, au niveau de la boucle centrale, un **anticodon** (figures 8-10) : **cet ensemble de 3 nucléotides est complémentaire d'un codon correspondant, dans le code génétique, à l'acide aminé porté par l'ARNt**. Notez qu'une **tolérance d'appariement atypique sur le troisième nucléotide est possible (effet wobble)**.
- On y trouve des **nucléosides modifiés** (exemple : **inosine**, dont la base azotée est l'**hypoxanthine**), voire des **appariements ponctuels atypiques** (dits **non canoniques**) comme **G-U** entre **anticodon** et **codon**.

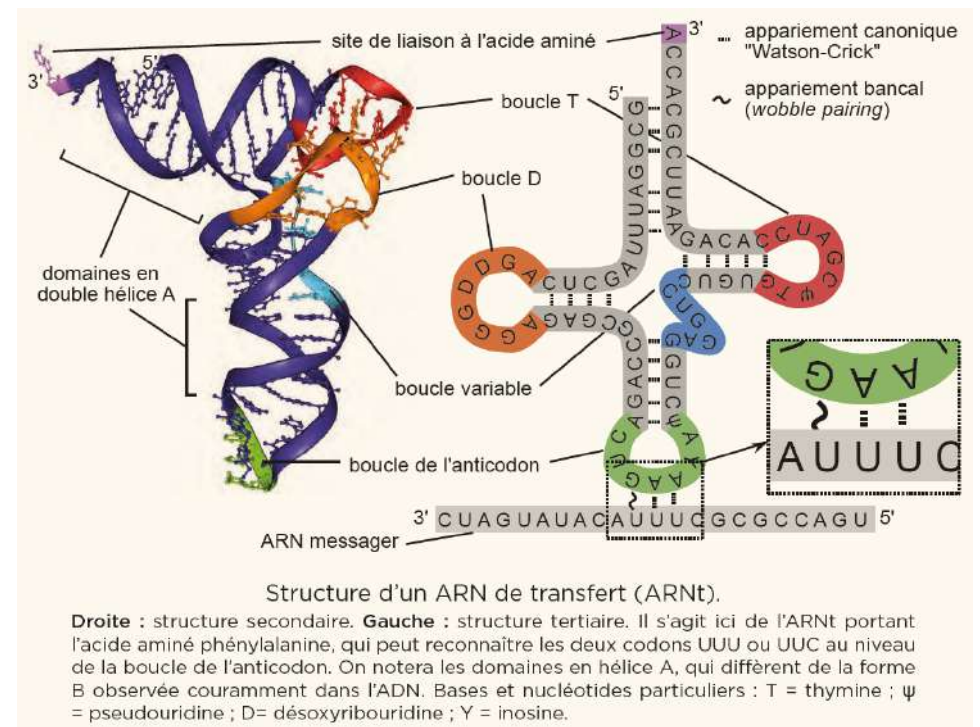
Les **ARNt** ont une **durée de vie plutôt longue** (plusieurs secondes à dizaines de secondes... c'est énorme pour un ARN !) et peuvent **servir de multiples fois**.



▲ **FIGURE 8. Structure secondaire d'un ARNt.** D'après CAMPBELL & REECE (2004).
On notera le caractère « **localement** » **bicaténaire** de la molécule.

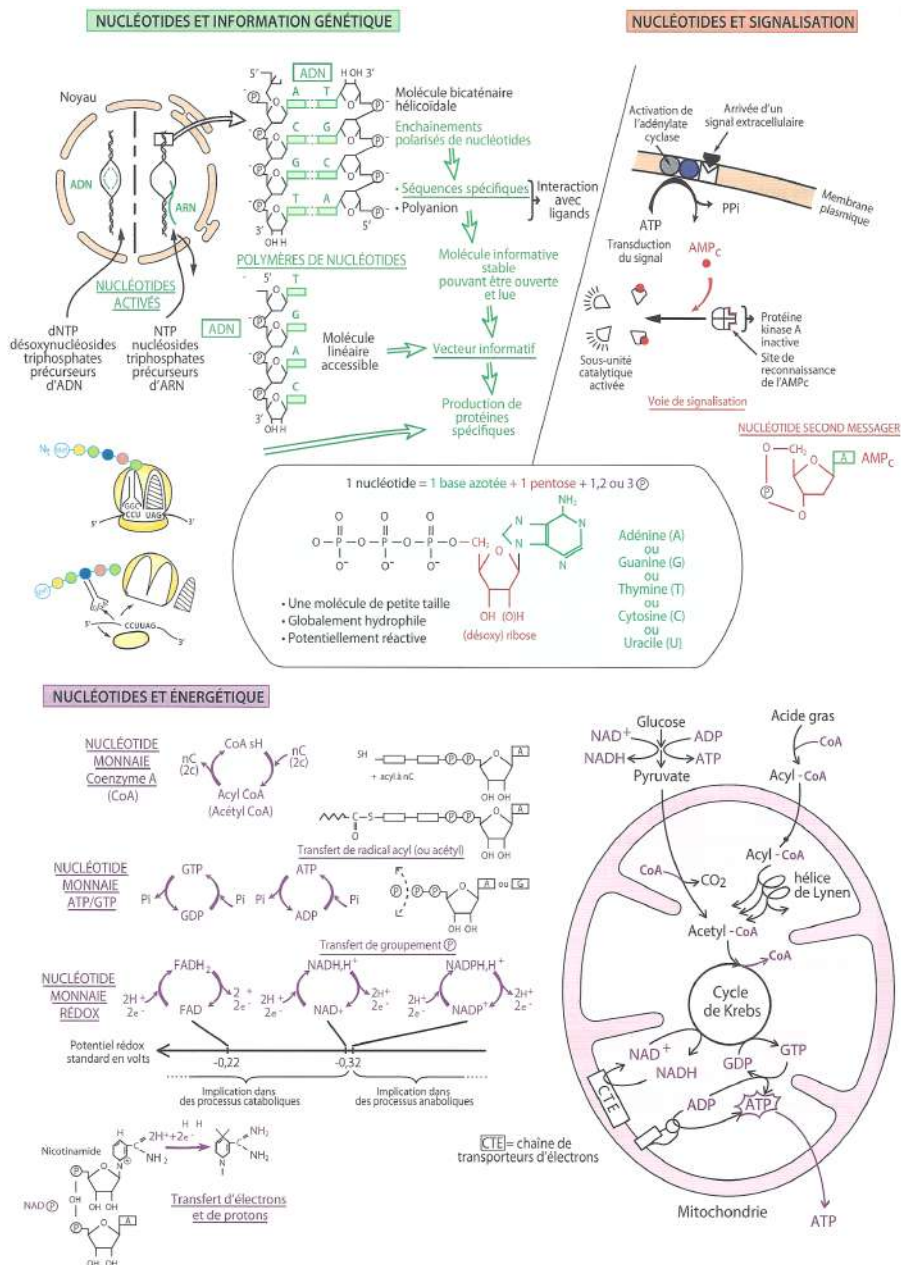


▲ **FIGURE 9. Structure tertiaire d'un ARNt.** D'après CAMPBELL & REECE (2004).
On notera le caractère « **localement** » **bicaténaire** de la molécule.



▲ **FIGURE 10. Vision d'un ARNt proposée par PERRIER, BEAUX et al. (2021).**

Dans l'ARN, on observe des appariements dit non canoniques, ou bancals (appelés en anglais *wobble pairing*) qui expliquent des repliements tridimensionnels particuliers. On notera en particulier l'appariement G-U, important pour expliquer la redondance du code génétique. La figure page précédente présente la structure d'un ARNt dont l'anticodon AAG peut reconnaître le codon UUU, par *wobble pairing*. Certains ARN possèdent des bases azotées non canoniques, ou des nucléotides modifiés (pseudouridine, hypoxanthine, thymine).



A FIGURE 11. Nucléotides (et acide nucléiques) : une vue d'ensemble.
D'après SAINTPIERRE, BORDI *et al.* (2021).

B. Des génomes étudiables par des techniques de biologie moléculaire

- L'étude des génomes passe par une panoplie de techniques dites de biologie moléculaire. Quelques-unes d'entre elles sont examinées ici. Une autre partie sera examinée dans le chapitre sur le contrôle de l'expression génétique (chapitre 15).

Capacités exigibles

- ✓ Réaliser et analyser les résultats d'une électrophorèse d'ADN.
- ✓ Interpréter l'organisation des génomes à partir des résultats de séquençage.
- ✓ Exploiter les données de séquençage pour réaliser des alignements de séquences et comparer les séquences.
- ✓ Analyser des résultats issus d'expériences de transgénèse ou de mutagenèse.]
- ✓ Analyser et interpréter des résultats expérimentaux utilisant les techniques de Southern blot, Northern blot, Western blot, hybridation in-situ ou de puce à ADN.]
- ✓ [Identifier et justifier les témoins de charge des blots.]
- ✓ [En se basant sur le fonctionnement des ADN polymérases, expliquer le principe de la PCR.]

1. Notions de techniques de biologie moléculaire et biotechnologies

- Les techniques de biologie moléculaire regroupent, au sens le plus large, l'ensemble des techniques permettant l'étude et la manipulation des molécules biologiques. Dans les faits, on entend surtout derrière ce terme les techniques modernes d'étude des macromolécules biologiques (protéines, acides nucléiques), notamment les acides nucléiques.
- Dans une approche plus appliquée aux besoins humains, on peut parler de biotechnologies pour désigner ce que l'OCDE propose de définir comme « l'application des principes scientifiques et de l'ingénierie à la transformation de matériaux par des agents biologiques pour produire des biens et services ». Les biotechnologies incluent notamment le génie génétique qui désigne l'ensemble des techniques permettant l'utilisation ou la modification du génome des organismes.

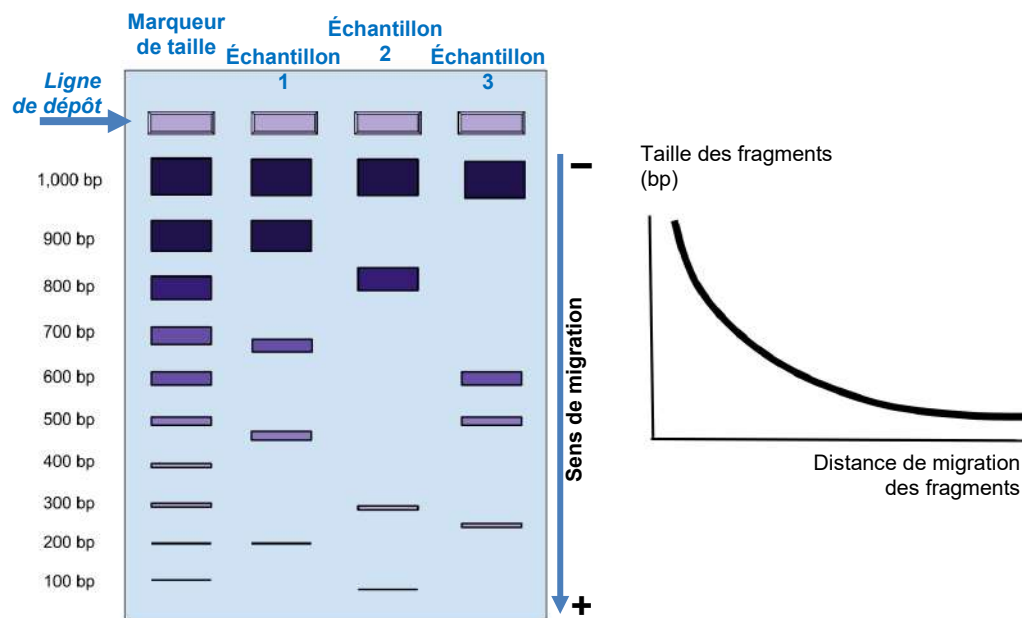
Le terme **biologie moléculaire** peut désigner plusieurs réalités selon les auteurs :

- ° Pour certains (par exemple BREUIL, 2003), il s'agit plus ou moins d'un synonyme de biochimie.
- ° Pour d'autres auteurs (par exemple BERTHET, 2006) il s'agit plus spécifiquement de l'étude des protéines et des acides nucléiques (ADN, ARN) qui sont les macromolécules du vivant.
- ° Pour la plupart des auteurs modernes et dans le monde académique, ce terme renvoie spécifiquement à l'étude des acides nucléiques et plus généralement de tous les mécanismes génétiques ayant lieu à l'échelle cellulaire et moléculaire (codage, organisme du génome, expression, régulation, conservation...); il s'agit alors d'un synonyme de **génétique moléculaire**. C'est là l'usage le plus répandu.

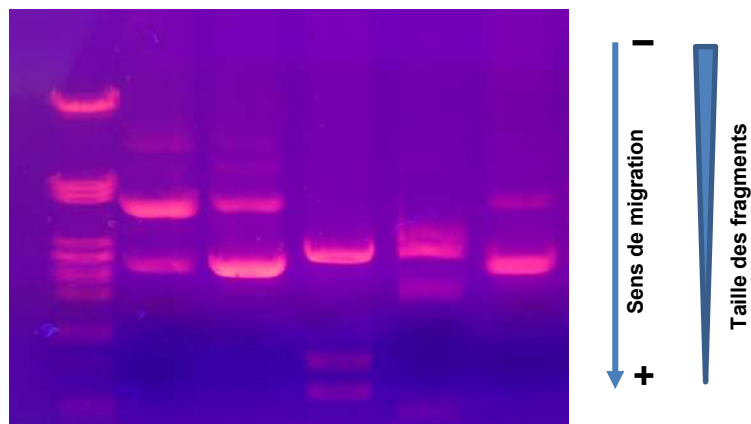
2. Les électrophorèses sur les acides nucléiques

- Une **électrophorèse** est une séparation des constituants d'un mélange selon leur capacité à migrer dans un champ électrique (figures 12-13). Cette technique est fréquemment utilisée aussi bien sur les protéines qu'avec les acides nucléiques.
- Elle se fait généralement sur gel d'agarose (acides nucléiques) ou de polyacrylamide (protéines). Des puits sont creusés dans le gel où sont déposés, par pipette à piston, les mélanges à étudier (souvent préalablement colorés ou couplés à un révélateur); un des puits contient généralement un marqueur de

taille/masse qui servira de **référence** pour évaluer la taille/masse des substances migrant sur les autres pistes.



▲ FIGURE 12. Principe d'une électrophorèse d'acides nucléiques sur gel d'agarose.
D'après SAINTPIERRE, BORDI et al. (2021).



▲ FIGURE 13. Un profil électrophorétique révélé aux UV. Wikipédia

- Dans le cas des **acides nucléiques**, le marqueur est un **marqueur de taille** : cette référence permet d'évaluer la longueur en pb des fragments d'ADN ou d'ARN présents initialement dans les puits.

- Les molécules de taille importante migrent plus lentement que celles de taille plus faible.
- Une **électrophorèse** permet l'obtention d'un **électrophorégramme** ou **profil électrophorétique** (figures 12-13).
- La **révélation** est généralement permise par l'ajout d'un **colorant d'acides nucléiques** au mélange initial ou tout autre substance que l'on peut révéler facilement (figure 13).

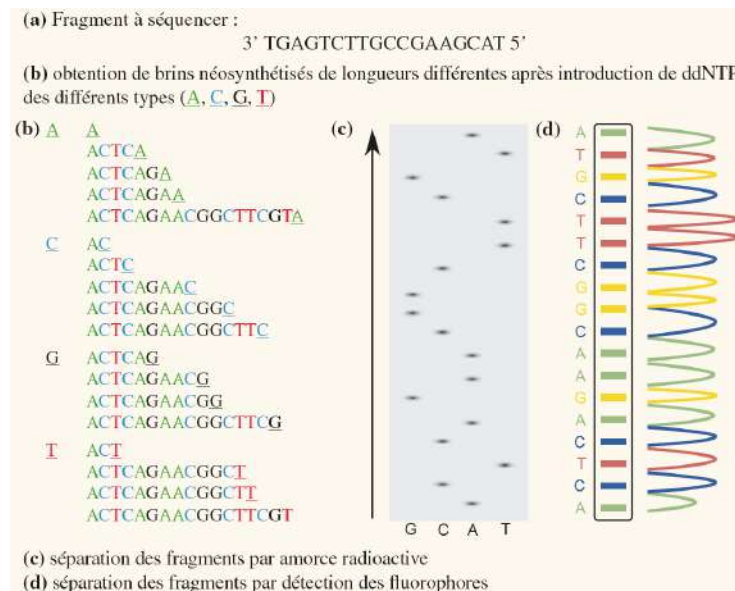
La méthode de révélation la plus utilisée pour les acides nucléiques est la révélation au **bromure d'éthidium** ou **BET**. Le **bromure d'éthidium** est un **agent d'intercalation** (= qui s'insère entre les nucléotides) couramment utilisé comme **marqueur d'acide nucléique** dans les laboratoires de biologie moléculaire. Lorsqu'il est exposé à des rayonnements ultraviolets, il devient fluorescent avec une couleur rouge-orangée, plus intense lorsqu'il est lié à un **acide nucléique**. Il est néanmoins très **dangereux** car hautement **mutagène**. Des **produits de substitution** existent.

3. Les méthodes de séquençage des génomes et leurs prolongements

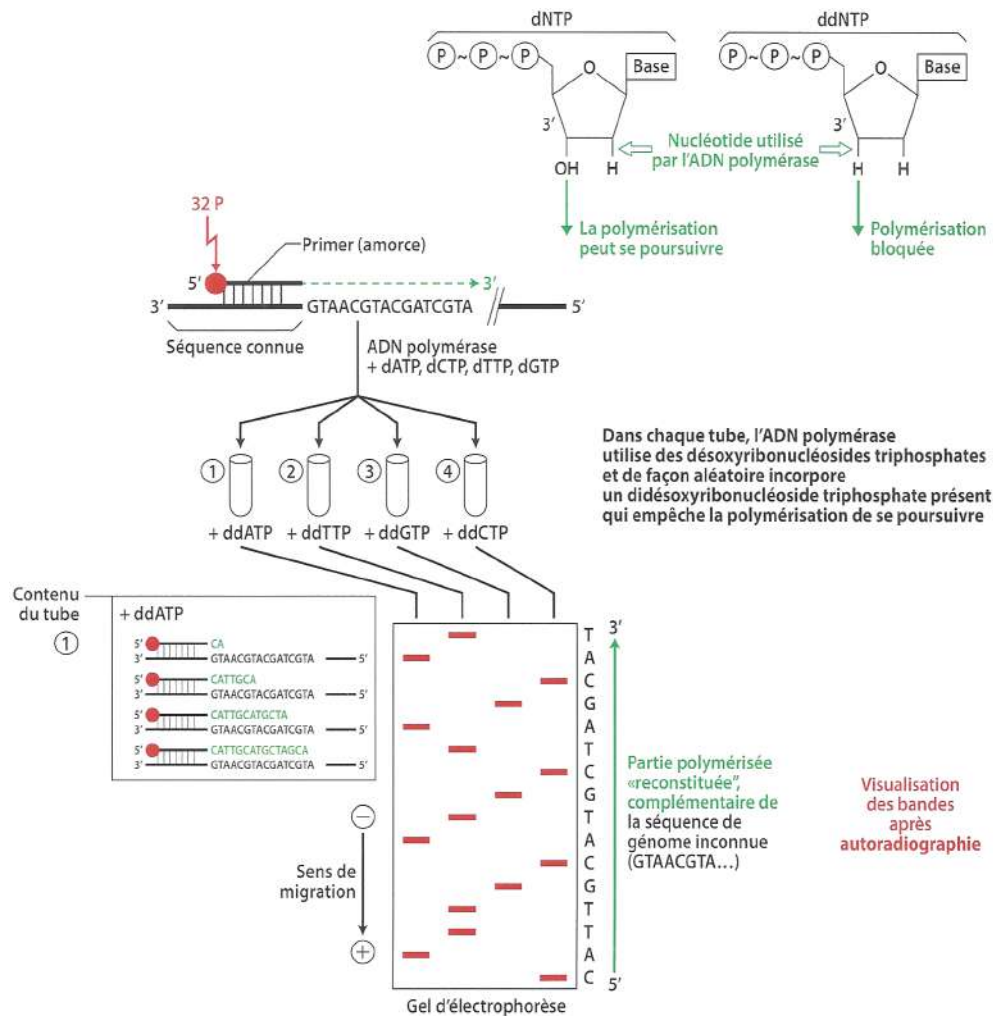
- On appelle **séquençage** la **reconstitution de l'ordre d'enchaînement des monomères d'un hétéropolymère linéaire** (= **séquence**). Dans le cas d'un **acide nucléique**, il s'agit donc de **reconstituer la séquence nucléotidique**.

Les techniques de séquençage en usage concernent des **molécules d'ADN**. Le séquençage des **ARN** suppose dans la plupart des cas leur préalable **rétrotranscription en ADN** (par une **rétrotranscriptase** ou **transcriptase inverse**).

a. Le principe d'une méthode historique de séquençage de petits fragments : la méthode enzymatique de SANGER (1977)



▲ FIGURE 14. Une illustration possible de la méthode de SANGER (révélation par autoradiographie et fluorescence). D'après DAUTEL et al. (2021).

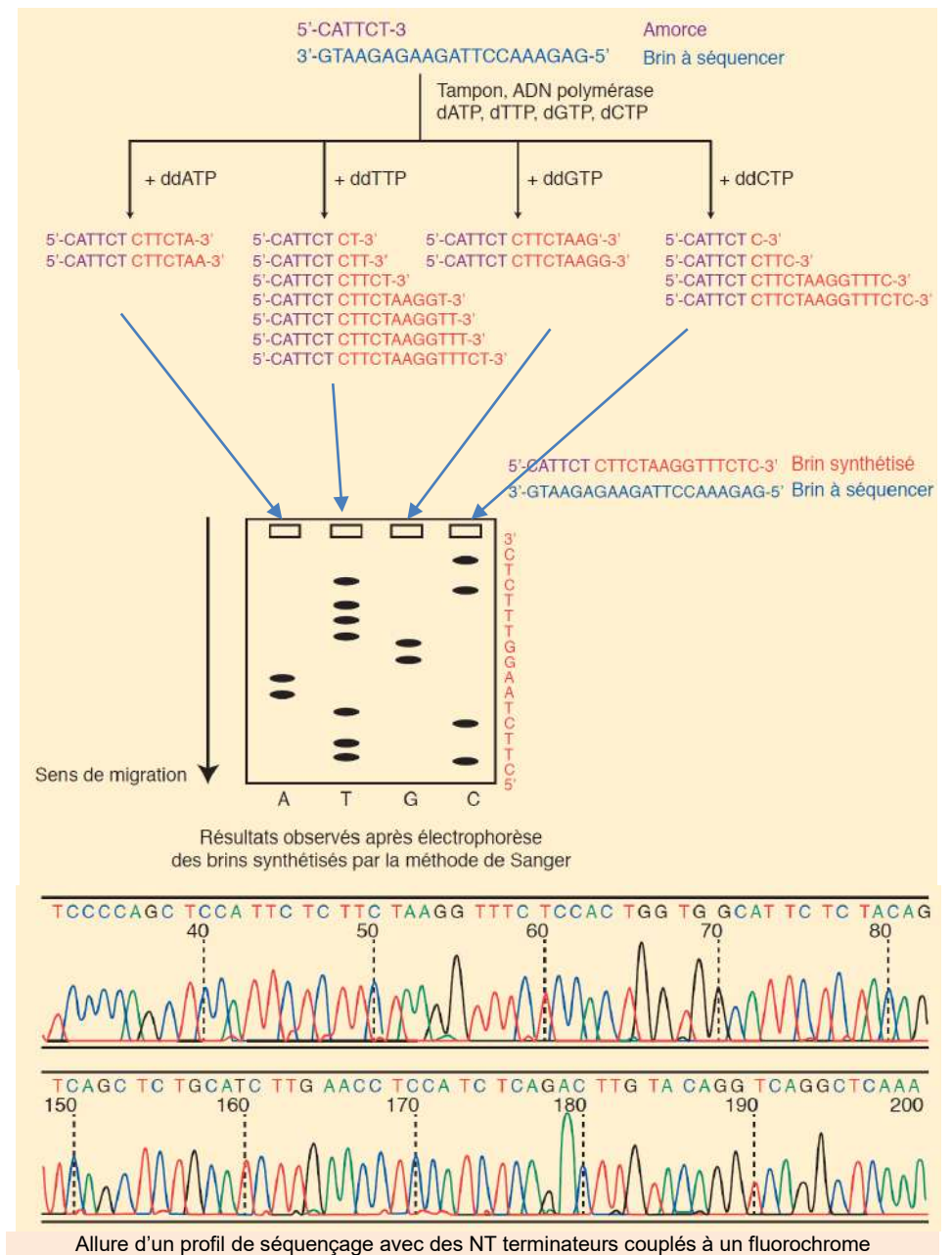


Les contenus des tubes ②, ③ et ④ contiennent des séquences de longueur variable se terminant toutes avec l'incorporation de didésoxynucleoside triphosphate, respectivement ddTTP, ddGTP et ddCTP

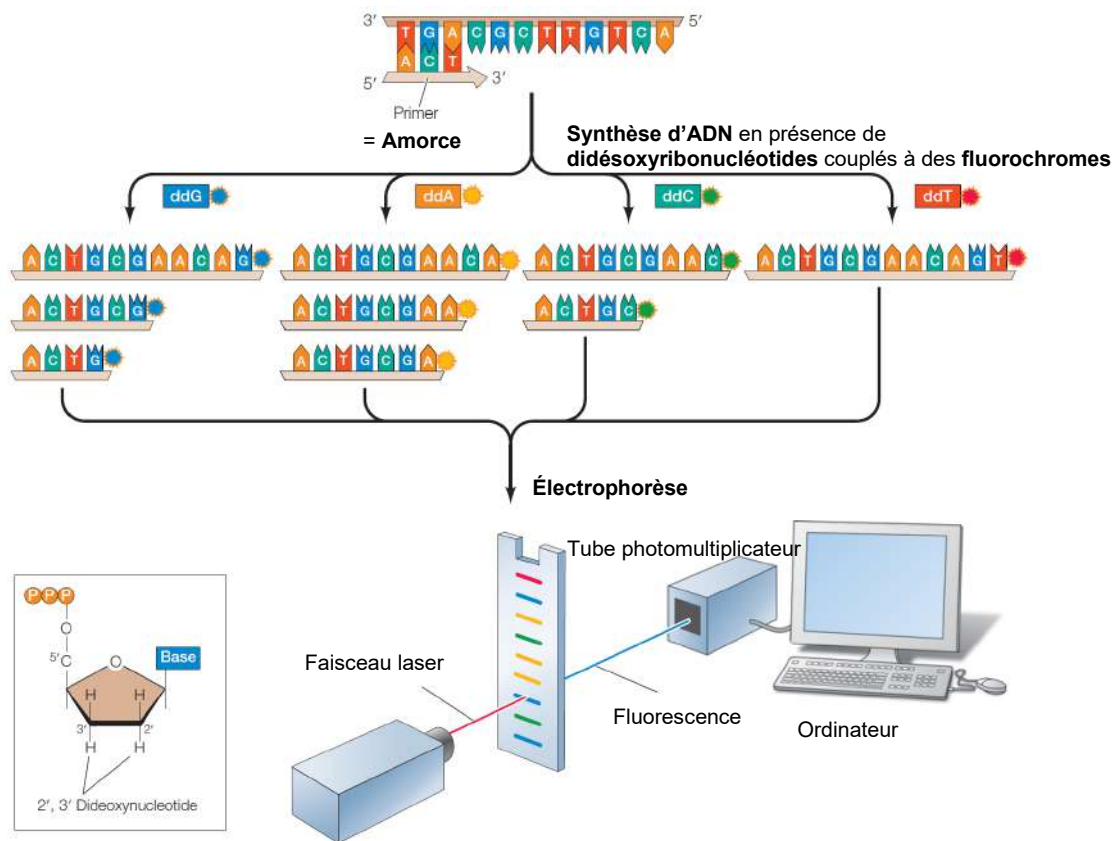
▲ **FIGURE 15. Une illustration possible de la méthode de SANGER (révélation par autoradiographie).** D'après DAUTEL *et al.* (2021).

Une animation pour comprendre le séquençage :

<https://www.dnalc.org/resources/animations/cycseq.html>



▲ **FIGURE 16. Une illustration possible de la méthode de SANGER (révélation par autoradiographie en haut, fluorescence en bas).** D'après SEGARRA *et al.* (2014).



▲ FIGURE 17. Une illustration possible de la méthode de SANGER modernisée (révélation par fluorescence). D'après COOPER (2019), traduit.

Une autre vision : <https://sciences.brussels/printemps2/pds2009/poster/biomol3.pdf> (consultation février 2022)

- Constituant encore la base de nombreux séquençages, la **méthode de séquençage de SANGER (1977)** est une **méthode de séquençage d'ADN dénaturé (monobrin)** basée sur la **synthèse du brin complémentaire** avec des **nucléotides normaux** et d'autres **bloquant la réplication (= nucléotides terminateurs)**, marqués, suivie d'une **migration électrophorétique** de fragments avec leur **brin complémentaire synthétisé**, cette migration étant d'autant plus éloignée du front de dépôt que le brin néosynthétisé est court, couplée à une **révélation par radioactivité ou fluorescence**.
- Elle s'applique sur des fragments de quelques dizaines à quelques centaines de paires de bases.
- Même si le principe de la manipulation est en réalité assez simple, la schématisation de cette technique ne parle souvent pas beaucoup aux étudiants, je vous en propose plusieurs versions (figures 14-17).

- La **méthode** peut être **résumée** comme suit :
 - L'ADN est tout d'abord **extrait, purifié et découpé** (à l'aide d'enzymes de restriction) en de multiples fragments.

L'extraction et la purification ne sont **pas vraiment au programme**, même s'il ne s'agit pas d'étapes très compliquées. La **récupération de séquences particulières** utilise des **sondes** (voir plus loin).

- [Les **séquences d'intérêt** sont aujourd'hui souvent **amplifiées par PCR** (voir plus loin)].
- L'ADN est **dénaturé** (double brin → **simple brin**).
- Les **fragments** sont placés dans **quatre tubes** (si **méthode radioactive**) ou un **seul** (si **méthode par fluorescence**) contenant un **mélange** comprenant :
 - Une **solution tampon**
 - De l'**ADN polymérase**
 - Des **amorces ADN** permettant d'initier la **réplication**, si possible correspondant à des **séquences connues** (ce qui évite de les multiplier) ;
 - Des **désoxyribonucléosides triphosphates (dNTP)** des quatre types ;
 - Dans **chaque tube** en faible quantité : un **didésoxyribonucléoside triphosphate (ddNTP)** marqué radioactivement (^{32}P) ; en cas d'usage de la méthode par fluorescence, un **seul tube** avec les **quatre ddNTP** (couplés chacun à un **fluorochrome différent**) suffit.

Les **ddNTP** ont la **capacité de stopper la polymérisation d'ADN** suite à leur incorporation dans le **brin néoformé** : ce sont des **nucléotides terminateurs**.

- On laisse ensuite le **processus de réplication** s'opérer. Les **ddNTP** vont venir se placer, au hasard, en face d'un **nucléotide** du brin matrice et **stopper le processus**, à des **endroits variables**.
- On procède à une **électrophorèse** qui permet la migration des **brins matrices appariés à leur brin complémentaire de taille variable**. Les molécules migreront d'autant plus loin que le **brin complémentaire** est court.
- Si l'on utilise la **méthode radioactive**, on réalise **quatre puits**, un pour **chaque tube** correspondant à **chaque ddNTP**.
- En cas d'usage de la **méthode par fluorescence**, un **seul puits** suffit.

- Les **profils électrophorétiques** sont ensuite soumis à **révélation**, soit par **autoradiographie**, soit par **fluorescence**, et la **lecture de la séquence** se fait du **font de migration vers le pôle de dépôt**.

Important : notez bien que le **séquençage** s'opère dans le **sens 5' → 3'** (rappelez-vous que les **NT polymérisent toujours en 3'**) et donne la séquence du **brin néoformé**. La **séquence du brin matrice** sera obtenue par application des **règles de complémentarité**.

b. L'existence de techniques plus modernes et moins coûteuses : le séquençage haut débit

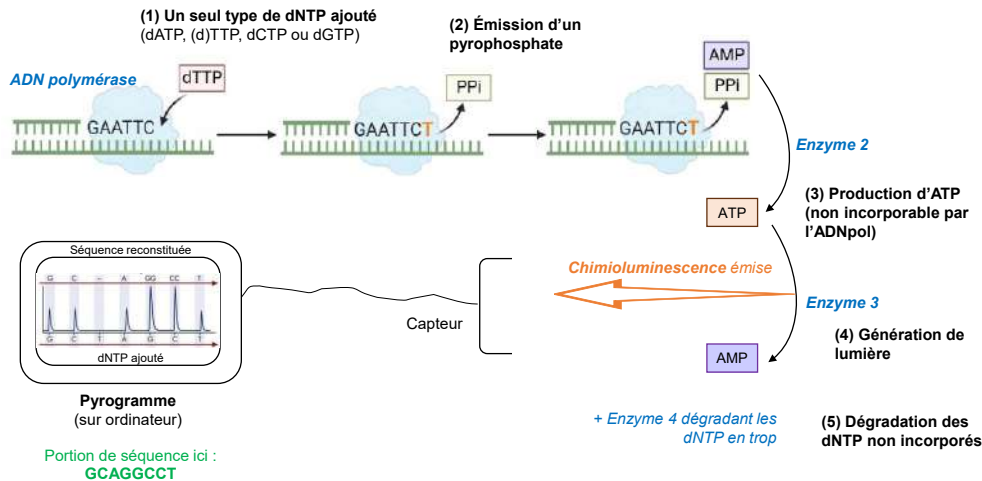
- La **méthode traditionnelle de SANGER** est **longue et coûteuse**, même si elle a permis la **plupart des séquençages** jusqu'au **début des années 2000**.
- Aujourd'hui, un **nombre considérable de techniques différentes** existe. S'inspirant plus ou moins de la méthode initiale de SANGER, elles sont :
 - De plus en plus **rapides** même si quelques **erreurs** peuvent subsister ;
 - De plus en plus **automatisées** ;
 - De plus en plus **fiables** ;
 - De moins en **moins coûteuses**.

- À côté des **méthodes de lecture courte** (*short-read sequencing methods*), dites aussi de **2^e génération**, qui requièrent toujours des **acides nucléiques plutôt courts** (plusieurs centaines de pb), des **méthodes de lecture longue** (*long-read sequencing methods*) ou **méthodes de 3^e génération** voient le jour depuis une dizaine d'années. Plus besoin de fragmenter l'ADN, le défi est alors au contraire de garder les **molécules intactes** !

Un exemple de méthode de lecture courte (2^e génération)

Le **pyroséquençage** permet, sans électrophorèse, le **séquençage en détectant le nucléotide incorporé par une ADN polymérase couplée à une enzyme chimioluminescente, au moyen de la quantification de lumière suite à la métabolisation du le pyrophosphate libéré lors de polymérisation.**

- Taille maximale des fragments : 700 pb
- Fiabilité : 99,9 %
- Coût pour 1 Gb (1 milliard de pb) : \$ 10 000
- 24 h pour lire 1 Mb (1 million de pb)



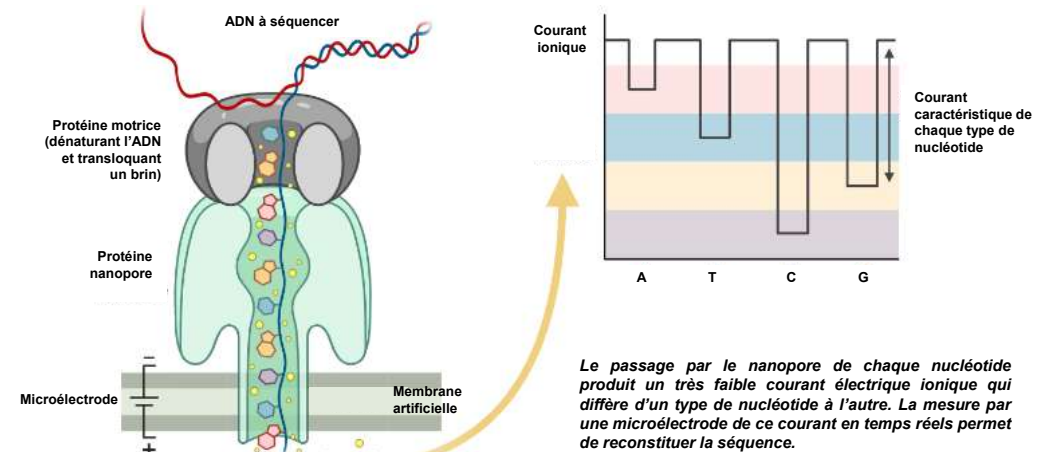
- Cette **technique** dérive de la **PCR** (qui nécessite donc un ADN dénaturé à séquençer, des amorces, des ARN polymérases... + quelques enzymes propres).
- La polymérisation se fait **nucléotide par nucléotide**.
- À **chaque étape**, on ajoute seulement un **type de désoxyribonucléoside triphosphate (dNTP)** (1).
 - ° S'il s'agit du **NTP complémentaire** qui convient pouvant être ajouté au **brin en cours d'élongation**, il est ajouté au **brin néoformé** (une ou plusieurs fois selon le nombre de bases semblables se suivant **immédiatement**) ;
 - ° S'il ne s'agit pas du **NTP complémentaire**, alors il n'y aura **pas de polymérisation**.
- En cas de polymérisation, il y a **éjection d'un PPi** (2) qui réagit avec de l'**AMP** présente dans le milieu (grâce à une enzyme 2), produisant de l'**ATP** (dans une version non incorporable à l'ADN en cours d'élongation) (3).
- Cet **ATP** est **consommé** par une **enzyme 3** (avec d'autres substrats) qui génère alors une **luminescence détectable** par une micro-caméra (4).
- (!) S'il y a **luminescence**, on peut dire que le **dNTP** a été **ajouté** : on a **élucidé une base de la séquence** (voire plusieurs si la luminescence est **élevée** et que le **dNTP** a été **incorporé plusieurs fois de suite**).
- (!) S'il n'y a **pas luminescence**, le **dNTP** n'a **pas été ajouté** et c'est un autre candidat qui convient.
- Les **dNTP surnuméraires** sont dégradés par une **enzyme 4** (5), moins rapide que la polymérase mais agissant sûrement.
- On ajoute alors un **nouveau type de dNTP**, et on **recommence le cycle**.

▲ FIGURE 17bis. **Pyroséquençage : principe simplifié [pour information]**.
Schéma compilé à partir de sources variées.

Un exemple de méthode de lecture longue (3^e génération)

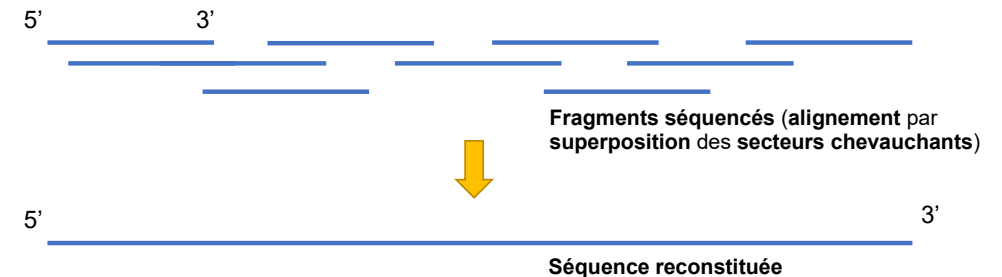
Le **séquençage nanopore** fait passer la **molécule d'ADN dans un nanopore** (ex. porine) **enchâssé dans une membrane qui sépare deux solutions ioniques** : le **passage de chaque nucléotide déclenche des perturbations ioniques qui sont fonction du type de nucléotide, la succession des perturbations étant à l'origine de la séquence reconstituée**. Cette méthode est sans polymérase ni électrophorèse.

- Taille maximale des fragments : 2,3 Mb (moy. 10-30 kb)
- Fiabilité : 92-97 % (en amélioration constante)
- Coût pour 1 Gb (1 milliard de pb) : \$ 7 (!!) à \$ 100
- 24-48 h pour lire jusqu'à 260 Gb (260 milliards de pb) !



▲ FIGURE 17ter. **Séquençage nanopore : principe simplifié [pour information]**.
Schéma BioRender, adapté.

c. La possibilité d'aligner les séquences et de rechercher les fragments chevauchants pour aboutir à une reconstitution de longues séquences



▲ FIGURE 18. **Le principe de l'alignement de séquences dans le cadre de la reconstitution d'une séquence d'ADN plus longue que les fragments séquencés (shotgun).** Original.

- On appelle **alignement des séquences** une **représentation de plusieurs séquences de macromolécules biologiques (ADN, ARN ou protéines) superposées les unes sous les autres, de manière à en faire ressortir les régions similaires ou au contraire les régions différentes**.
- Dans le cas de l'ADN, cette technique peut permettre :
 - De **reconstituer progressivement les génomes** en inférant des **séquences globales** à partir de **fragments séquencés** (figure 18) ; c'est la **méthode shotgun** (« fusil de chasse ») de reconstitution des génomes (figure 19).
 - De mettre en évidence et comparer des **zones semblables d'un génome** : **séquences répétées**, **familles multigéniques** = **gènes paralogues** (= **gènes dérivant évolutivement d'une même séquence originelle**)...
 - De **comparer des génomes** entre eux et d'y découvrir des **gènes homologues** (= **gènes partageant une même origine commune**)

Dans ce cas, il existe des **gènes orthologues** et des **gènes métallogues**... nous y reviendrons dans le chapitre sur la phylogénie.

- Notons, dans le cas du **séquençage global**, que l'existence de **séquences répétées** est un **frein à la reconstitution des génomes entiers** à partir de **petites séquences fragmentées**... Les **outils de lecture longue** en cours de perfectionnement sont donc d'un immense secours.

d. La bio-informatique au secours du séquençage et de l'annotation des génomes ; notions de génomique (structurale / fonctionnelle) et de métagenomique

- L'**outil informatique** est le seul capable de gérer les **énormes flux de données** suscités par le **séquençage à grande échelle**, qu'il s'agisse du **stockage** ou du **traitement** des séquences.
- L'**automatisation du séquençage** et l'**usage de l'informatique** ont permis la genèse d'une discipline nommée **génomique** ou **étude globale du (ou des) génome(s) dans leur organisation comme leur fonctionnement**. Celle-ci comprend **deux axes** en réalité imbriqués :
 - La **génomique structurale** qui **s'intéresse à la nature, la taille, l'organisation et la séquence des génomes**.
 - La **génomique fonctionnelle** qui **s'intéresse au fonctionnement des génomes, c'est-à-dire au rôle des séquences présentes**. Elle se base en partie sur l'**étude de l'ensemble des transcrits (= ARN) d'une cellule ou d'un organisme (transcriptome)** ou **transcriptomique**, ainsi que sur l'**étude de l'ensemble des protéines d'une cellule ou d'un organisme (protéome) ou protéomique**.

- Dans le cadre de la **génomique**, on procède à l'**annotation des génomes**, c'est-à-dire une **analyse (informatique) des génomes visant à déterminer la fonction des séquences, supposant d'identifier par exemple les séquences codantes, les familles multigéniques ou encore les séquences répétées**.
- L'**identification des séquences codantes** repose notamment sur la **recherche de promoteurs géniques**, de **codons start / stop**, de **sites d'épissage**, etc. On cherche à y identifier les **ORF potentiels** ou **séquences susceptibles d'être traduites en protéines**.

- La **recherche des fonctions géniques** peut aussi passer par l'**inactivation de gènes** ou au contraire leur **transfection dans un organisme / une cellule qui ne les exprime pas naturellement** (voir plus loin).
- La **métagenomique** consiste à **séquencer l'ensemble des génomes présents dans un habitat donné**. Cela peut aussi bien permet d'étudier le **microbiome** que

des **organismes de plus grande taille**, sans passer par les **longs délais de la systématique classique**.

e. La taille des génomes et sa non-corrélation avec la position phylogénétique des organismes : le paradoxe de la valeur C

- On peut appeler **valeur C (C-value)** la **taille d'un génome en nombre de paires de bases**.
- On appelle **paradoxe de la valeur c** le constat qu'il n'existe **pas de lien entre la position phylogénétique d'un taxon et la taille de son génome** (tableaux Ibis-II).

Certains auteurs disent plutôt qu'il n'existe **pas de lien entre la « complexité » des organismes et leur taille de génome** mais cette notion de « complexité » n'a jamais reçu de définition pertinente et unanime (encadré C).



Important : la mesure de la taille des séquences et génomes

Pour mesurer une **séquence nucléotidique d'ADN**, un **chromosome** ou un **génome**, on utilise la **paire de base** ou **pb** qui **correspond à une paire de nucléotides se faisant face au sein des deux brins d'ADN**. L'unité accepte les **préfixes multiplicateurs** : $10^3 \text{ pb} = 1 \text{ kb}$ (kilobase), $10^6 \text{ pb} = 1 \text{ Mb}$ (méga-base), $10^9 \text{ pb} = 1 \text{ Gb}$ (giga-base).

TABLEAU Ibis. Comparaison du génome de quelques organismes.

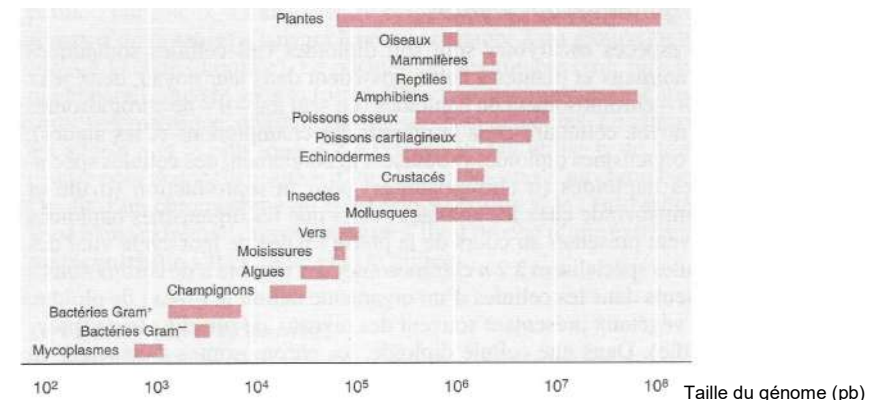
D'après SEGARRA *et al.* (2014).

Organisme	Nombre de paires de bases	Nombre de gènes estimés	Densité génique (Gènes/Mb)
<i>Mycoplasma</i>	0,6.10 ⁶	450	860
<i>E. coli</i>	4.10 ⁶	4 000	950
<i>Saccharomyces cerevisiae</i>	12.10 ⁶	6 000	480
<i>Arabidopsis thaliana</i>	142.10 ⁶	26 000	220
<i>Caenorhabditis elegans</i>	97.10 ⁶	19 000	190
<i>Drosophila melanogaster</i>	137.10 ⁶	14 000	82
<i>Homo sapiens</i>	3.10 ⁹	20 à 25 000	6.25

Taille des génomes de quelques êtres vivants procaryotes (*Mycoplasma* et *E. coli*) et eucaryotes (les autres)

TABLEAU II. Comparaison du génome de quelques organismes (en kb).

D'après PETIT & JULIEN (2007).



Encadré C « Complexité » et paradoxe de la valeur C

La « complexité » des organismes, l'impossible définition

➤ De manière surprenante, il est « évident » pour nombre de scientifiques qu'une bactérie comme *Escherichia coli* est moins complexe qu'un être humain. Pourtant, cette « évidence » repose-t-elle sur des critères objectifs ou bien sur une indéfinissable « intuition » empreinte de subjectivité et de jugement de valeur ? Sur quels critères peut-on dire qu'un organisme est plus « complexe » qu'un autre ? Le **tableau 1** montre qu'il n'existe **aucun critère vraiment opérationnel** pour définir une notion que l'on baptiserait « complexité » dans l'ensemble du vivant. Il faudrait alors **combinaison** plusieurs critères mais dans ce cas, si deux critères s'opposent, il conviendra de définir un ordre de précession des critères... qu'il faudra pouvoir justifier scientifiquement ! On voit donc que **poser la question de la « complexification » du vivant est quasi-impossible puisque la question même ce qu'est ou de ce qui fonde la « complexité » des organismes ne trouve pas de réponse satisfaisante et n'est, en tout cas, certainement pas tranchée**. La science ne peut pas évaluer quelque chose qu'on ne parvient pas à conceptualiser ni à appréhender rigoureusement. La « complexité » apparaît donc davantage comme un **jugement de valeur** que comme un concept scientifique et dès lors, il appartient à d'autres champs disciplinaires que celui de la biologie (comme la philosophie ?) de s'interroger sur ce concept.

Le paradoxe de la valeur C est-il un vrai paradoxe ?

➤ Il est dit de longue date qu'il n'existe pas de lien entre la « complexité » d'un organisme et la **taille de son génome** (nommée **valeur C** car c'est une constante pour un organisme donné), ce qu'on appelle « **paradoxe de la valeur C** ». La formulation de ce paraxode ne nous dit néanmoins pas quel autre critère est retenu pour appréhender la complexité et, comme nous avons vu qu'aucun critère d'appréciation de la complexité autre que génomique n'était vraiment objectivable (et le paradoxe ainsi formulé fait clairement référence à une complexité autre que génomique), il apparaît qu'il y aurait un « **paradoxe** » entre la **taille du génome** et quelque chose d'intuitif qu'on n'arrive pas à définir ni à appréhender rigoureusement ! D'autres auteurs, conscients du problème et de la **vacuité de la notion de complexité** des organismes, ont proposé de formuler le paradoxe différemment : **il n'existe pas de lien entre taille du génome et position phylogénétique d'un organisme**, ce qui est pour le coup **beaucoup plus rigoureux** (puisque la position phylogénétique est une réalité biologique appréhendable). En réalité, cet apparent paradoxe peut être levé en remarquant que :

- Dans certains taxons (comme les Angiospermes où la valeur C varie de 10^7 à 10^{11} pb), les phénomènes de **modification du caryotype**, notamment par **polyploïdisation**, sont **fréquents**, ce qui **augmente grandement la taille du génome**.
- Les **processus d'augmentation de la taille du génome** (duplications, transpositions, mutations chromosomiques...) sont fréquents à l'échelle de l'évolution et peuvent faire **varier** de manière importante la **taille des génomes**.
- Il existe à l'inverse des **simplifications de génomes** qui peuvent intervenir suite à la **perte d'une fonction** (cas de la **diminution du nombre de gènes chloroplastiques** dans les taxons ayant perdu les plastides ou ayant des plastides non photosynthétiques).

➤ L'évolution est un processus qui modifie les génomes en permanence, y compris entre taxons apparentés, sans obéir à une tendance régulière et uniforme. Il y a **ajustement adaptatif** de la **taille du génome** en fonction des **forces évolutives** (dont la sélection) qui s'exercent sur les organismes associés à leur milieu de vie ou aux contraintes de leur plan d'organisation.

▼ **TABLEAU 1. Les critères possibles d'évaluation de la complexité des organismes et leurs limites.**

Critère retenu	Limites du critère
Nombre de cellules	▪ Ce critère ne fonctionne pas pour comparer pour des unicellulaires entre eux (comment comparer la complexité de deux Eubactéries par exemple ?). ▪ Il est impossible de compter les cellules de la plupart des pluricellulaires : on ne les estime que de manière indirecte et spéculative par des calculs se fondant sur des extrapolations. ▪ Le nombre de cellules est souvent fonction de la taille dans beaucoup de groupes (notamment animaux), ce qui n'est pas forcément un grand indice de « complexité » : ainsi une Souris serait présentée comme moins complexe qu'un Éléphant.
Nombre de types cellulaires	▪ Des « types cellulaires » peuvent apparaître ou disparaître de manière imprévisible dans l'évolution. ▪ Sur quels critères peut-on dire de deux cellules qu'elles font partie d'un même type cellulaire ou qu'elles relèvent de deux types cellulaires différents ?
Nombre d'organes	▪ Ce critère ne permet pas de comparer des organismes qui n'ont pas déjà un plan d'organisation un temps soit peu similaire. ▪ Des organes peuvent apparaître ou disparaître assez inopinément dans l'évolution. ▪ Sur quels critères peut-on dire de deux structures qu'elles font partie d'un même organe ou qu'elles constituent deux organes différents ?
Diversité des tissus	▪ Les organismes assez proches ont globalement les mêmes types de tissus donc ne peuvent être appréhendés par ce critère. ▪ Là encore, il conviendra de définir à partir de quels critères on peut que dire deux tissus sont différents ou constituent un seul et même tissu. ▪ Ce critère est d'usage impossible chez des organismes n'ayant pas un même plan d'organisation et donc des tissus homologues.
Complexité comportementale	▪ On a parfois proposé de retenir ce critère sans qu'aucune définition plausible de la « complexité » d'un comportement n'ait jamais pu être apportée : une attention particulière a été portée sur le comportement social mais, en réalité, il s'agit plus de comparer la socialité animale à la socialité humaine prise comme référence (et considérée <i>a priori</i> comme la plus complexe) que de fournir un vrai substrat objectif. ▪ Les organismes n'ayant pas de vrai « comportement » ne peuvent être ainsi appréhendés, ce qui réduit le champ d'application de ce critère aux Métazoaires.
Complexité du système nerveux	▪ Ce critère ne fonctionne que pour les organismes eumétazoaires (qui sont pourvus d'un système nerveux). ▪ Il convient de définir ce qui fonde la complexité du système nerveux : le nombre de neurones ? la présence ou non de cellules gliales engainantes autour des axones (comme la gaine de myéline des Gnathostomes) ? La taille du cerveau (dans ce cas, le Cachalot est beaucoup plus complexe que l'Homme) ?
Nombre de gènes dans le génome	▪ Il s'agit là (enfin) d'un critère objectivement appréhendable dont l'étude est permise par les travaux de séquençage génomique dont nous disposons ▪ Les phénomènes de polyploïdisation évolutivement fréquents, particulièrement dans les génomes de groupes tels que les Angiospermes, peuvent « gonfler » le nombre de cadres de lecture détectés mais impliquent-ils vraiment une « complexification » ? ▪ Il faut être prêt à dire qu'un plant de maïs (50 000 gènes) est nettement plus complexe qu'un être humain (30 000 gènes) – ce qui n'a rien de choquant scientifiquement.
Taille du génome (valeur c)	▪ Il s'agit là encore d'un critère objectivement appréhendable. ▪ Les phénomènes de polyploïdisation peuvent là encore « gonfler » la taille des génomes mais impliquent-ils vraiment une « complexification » ? ▪ Il faut être prêt à dire qu'un unicellulaire comme l'amibe <i>Polychaos</i> (675 000 Mpb) est nettement plus complexe qu'un être humain (3 400 Mpb).

f. Le barcoding (et le metabarcoding) : des séquences d'un gène hautement conservé pour désigner les espèces

- À l'heure où la biodiversité disparaît à une vitesse fulgurante, des systématiciens ont décidé de **lancer un grand programme de séquençage visant à répertorier les espèces, connues ou inconnues**, sans passer par les **longs délais de la taxonomie classique** : le **DNA barcode (code-barres ADN)**, une **séquence ADN**

ultra conservée formant un code-barre propre à une espèce donnée malgré la possibilité d'une faible variation intraspécifique (figure 19).

- Il s'agit de séquencer, chez un maximum de taxons, un **gène identifié** dont la séquence constitue le **code-barre d'une espèce** : chez les **Animaux**, c'est souvent le gène de la **COI, cytochrome oxydase sous-unité I, une sous-unité de protéine mitochondriale codée par environ 600 pb** ; d'autres loci sont employés (tableau III). Ces séquences peuvent en outre être utilisées en **phylogénie moléculaire** pour reconstituer des **liens de parenté**.
- Ce système de « description » ou de reconnaissance pose toutefois de **nombreux problèmes** :
 - fiabilité pour reconnaître des espèces,
 - difficulté de **mise en œuvre** sur le terrain,
 - absence totale d'information sur les taxons séquencés** s'ils n'ont pas fait l'objet d'une description classique...
 - Etc.
- Son intérêt est de permettre une **identification à grande échelle**, sans taxonomistes. Le système permet aussi de distinguer les **espèces jumelles**, **espèces qui ne peuvent pas ou très difficilement s'identifier sur des caractères structuraux**, et de mettre en évidence des **espèces cryptiques**, **espèces jumelles dont la découverte n'est permise que par le séquençage ou une autre approche moléculaire et qui n'auraient pas été découvertes sans cela**.



A FIGURE 19. Code-barres ADN de quatre Métazoaires. Wikipédia (2015).

Chacune des quatre **couleurs** représente un **type de nucléotide** (A, T, C, G).

Chez les **Métazoaires**, le **DNA barcod** est généralement la séquence de la **sous-unité 1** de la **cytochrome oxydase** (protéine mitochondriale).

▼ TABLEAU III. Séquences utilisées en barcoding [pour information].
D'après Wikipédia (2022)

Groupe	Loci
Animaux	<i>COI, Cytb, 12S, 16S</i>
Chlorobiontes	<i>matK, rbcL, psbA-trnH, ITS</i>
Bactéries	<i>COI, rpoB, 16S, cpn60, tuf, RIF, gnd</i>
Eumycètes	<i>ITS, TEF1α, RPB1, RPB2, 18S, 28S</i>
Eucaryotes unicellulaires	<i>ITS, COI, rbcL, 18S, 28S, 23S</i>

- Des **bases de données mondiales de DNA-barcods** tendant à être incrémentées très rapidement : **BOLD¹** (Barcode of Life Data System), **UNITE**, **Diat.barcode...**
- Le **métabarcoding** désigne le **barcoding en masse de l'ensemble des organismes d'un groupe vivant au sein d'un habitat naturel ou d'un échantillon**. Des campagnes de recherche de **biodiversité dans le sol**, au sein du **microbiote**, dans un **écosystème tropical**... intègrent de plus en plus cette pratique.

g. L'existence de banques de données assurant la mutualisation mondiale des portions séquencées et des protéines connues

- On assiste, depuis **plusieurs années**, à une **multiplication des bases de données moléculaires** qui assurent la **mutualisation des séquences nucléiques** ou encore des **connaissances sur les protéines** :
 - Banques de séquences** comme **GenBank²** qui recense les séquences nucléotidiques disponibles publiquement.
 - Banques de barcods** : voir **ci-dessus**.
 - Banques génomiques** où le génome entier d'une espèce est présent (exemples : *Saccharomyces Genome Database*, *FlyBase*, *The Arabidopsis Information Resource (TAIR)*, *Mouse Genome Informatics (MGI)*...). Les **données génomiques** de nombreuses espèces sont compilées par **NCBI (National Center for Biotechnology Information)** : <https://www.ncbi.nlm.nih.gov/genome/>

Une **question** traverse aujourd'hui les **biologistes moléculaires** : avec le **faible coût** et la **rapidité** actuelle du **séquençage**, **conserver de très nombreux génomes** n'est-il pas parfois **plus coûteux** que de les **re-séquencer en cas de besoin** ?

- Banques protéiques** où, au-delà des **séquences protéiques**, de très nombreuses **informations structurales** (y compris des **modèles 3D** de la **conformation native** ou des **versions mutées**) et **fonctionnelles** sont généralement présentes. Exemple : **Protein Data Bank (PDB)³**

¹ https://www.boldsystems.org/index.php/TaxBrowser_Home

² <https://www.ncbi.nlm.nih.gov/genbank/>

³ <https://www.rcsb.org/> [liens valides en février 2022]

4. L'amplification des séquences nucléotidiques

- Aujourd'hui, nul besoin de récupérer en très grande quantité de l'ADN (ou des ARN) pour procéder ensuite à son (leur) séquençage ou son (leur) étude... Des méthodes d'amplification, c'est-à-dire de polymérisation reproductrice en grande quantité d'acides nucléiques, existent. La technique PCR, avec ses (très nombreux) dérivés, constitue l'approche la plus répandue.

a. L'amplification de séquences d'ADN : la réaction de polymérisation en chaîne (PCR)

Cette technique a été publiée en 1986 par l'Américain Kary B. MULLIS (1944-2019) [Prix Nobel de Chimie 1993] et mise en œuvre avec succès en 1988 avec la Taq polymérase... Fruit de nombreuses recherches impliquant des chercheurs variés, la paternité de ce tournant majeur de la biologie moléculaire demeure largement discutée et mériterait une discussion complète.

- Produire une grande quantité de copies d'un fragment d'ADN est permis par la réaction en chaîne de la polymérase ou PCR (*Polymerase Chain Reaction*) (figures 21-22).

Cette réaction repose sur des cycles, répétés de nombreuses fois, de :

- dénaturation de l'ADN, c'est-à-dire séparation thermique des deux brins d'ADN (à 94 °C),
- fixation d'amorces (primers) [annealing] ADN sur les brins séparés (à 50-60 °C),

Le choix des amorces est crucial ; elles délimiteront la région d'ADN effectivement amplifiée (qui peut donc en pas correspondre à l'intégralité du fragment d'ADN initialement introduit).

Des PCR utilisant plusieurs amorces différentes permettent d'amplifier plusieurs séquences cibles (PCR multiplex).

- et polymérisation (élongation) de dNTP formant le brin complémentaire par une polymérase thermostable (stable à la chaleur) et agissant à chaud (72 °C), la Taq polymérase.

La Taq polymérase provient d'une Bactérie thermophile découverte dans le parc américain du Yellowstone : *Thermus aquaticus*, jadis nommée *Thermophilus aquaticus*.

- Les séquences amplifiées (= obtenues apr PCR) sont nommées des amplicons.
- Très tôt depuis les débuts de la PCR, des appareils automatisant les PCR ou thermocycleurs (figure 20) ont été mis au point. Leur coût a considérablement baissé alors que leur efficacité n'a fait qu'augmenter.



▲ FIGURE 20. Thermocycleur. Wikipédia (2022).

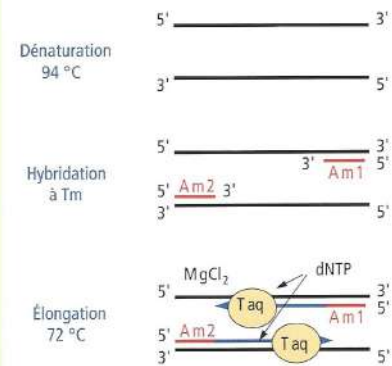
• LA RÉACTION DE POLYMERISATION EN CHAÎNE (PCR)

Le principe de la PCR (pour « Polymerase Chain Reaction » ou Réaction de Polymérisation en Chaîne a été décrit par Saiki et collaborateurs en 1985 et sa mise au point a récompensé du prix Nobel de Chimie K. Mullis en 1993. La PCR a depuis été développée pour un certain nombre d'applications.

Cette technique permet l'obtention d'un très grand nombre de copies d'un segment particulier d'ADN, nommé ADN cible. Le principe de cette technique consiste à utiliser deux amorces, courtes séquences d'ADN de 18 à 24 nucléotides, qui s'hybrident à des sites complémentaires situés en orientation inversée sur les deux brins de l'ADN et qui encadrent la région que l'on veut amplifier. L'élongation des brins complémentaires à partir des amorces est assurée par une enzyme, l'ADN polymérase. Alors que le principe de la PCR a été décrit en 1985, en 1988 la PCR devient encore plus performante avec l'emploi de la Taq polymérase dérivée d'une bactérie thermophile, *Thermus aquaticus*, isolée des sources chaudes du parc de Yellowstone et qui présente la propriété d'être stable à la chaleur.

La réaction de PCR repose sur trois étapes (figure) :

- la dénaturation de l'ADN réalisée à haute température (30 sec à 1 min à 94 °C) conduit à la séparation des brins d'ADN ;



La PCR comporte trois étapes : dénaturation, hybridation et élongation.

- l'hybridation des amorces spécifiques se réalise à une température déterminée par rapport à la température de fusion des amorces notée Tm (30 sec à 2 min à 50-60 °C) ;
- l'élongation des brins d'ADN est effectuée à 72 °C (30 sec à 1 min).

La durée de ces étapes dépend de plusieurs facteurs, comme la performance de l'appareil utilisé et la longueur des fragments à amplifier. Les produits de PCR obtenus sont désignés sous le terme d'amplicons.

Pour effectuer une PCR, on utilise un thermocycleur qui peut contenir plusieurs blocs. Chaque bloc peut contenir de 25 à 96 microtubes ou une microplaque selon les modèles.

Chaque microtube contient, outre du tampon :

- l'ADN de l'organisme à partir duquel on veut amplifier une région cible, cet ADN servant de matrice ;
- les deux amorces encadrant la région à amplifier ;
- des dNTP (dATP, dCTP, dGTP, dTTP) à partir desquels les brins complémentaires seront synthétisés ;
- du MgCl₂, cofacteur de la réaction ;
- de la Taq polymérase qui polymérise la réaction.

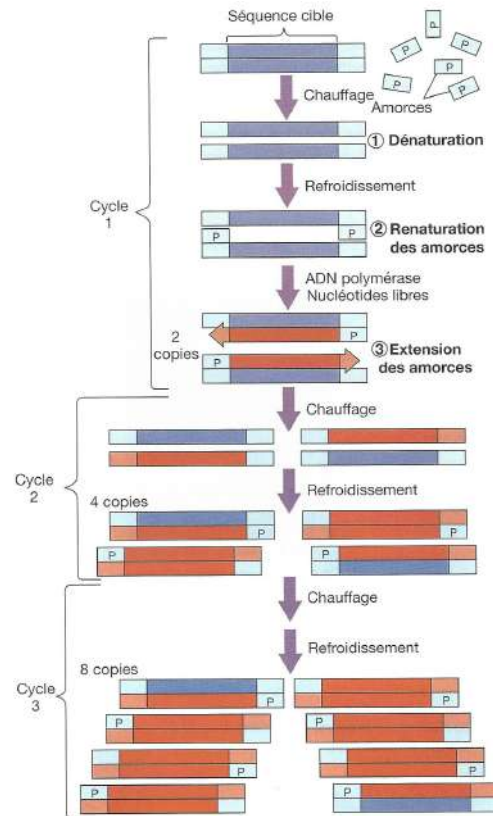
Notons que la détermination de la concentration en MgCl₂ est importante et se situe entre 1,0 et 3,0 mM. En effet, un excès de MgCl₂ provoque des hybridations non spécifiques et un manque de MgCl₂ induit une diminution de l'activité de la polymérase.

Pour chaque PCR, on établit pour un bloc un programme comprenant un cycle de trois températures (dénaturation, hybridation, élongation) répété de 20 à 40 fois. À chaque cycle, la quantité d'ADN est amplifiée par un facteur de 2, puisque chacun des deux brins d'ADN nouvellement synthétisés sert de matrice à la fabrication d'un nouveau brin complémentaire. À l'issue de la PCR, n cycles produisent en théorie 2ⁿ – (2n) amplicons.

Une séquence peut être amplifiée par un facteur de 10 millions avec une haute spécificité.

Les amplifiats obtenus après PCR sont vérifiés sur gel d'agarose. Leur taille doit correspondre à la taille attendue pour la région ciblée (figure 4.3). La RT-PCR (pour « Reverse Transcriptase PCR ») permet l'amplification d'une région donnée à partir d'ARNm. Cette réaction a comme première étape, étape de rétro-transcription (RT) la conversion des ARNm en ADN simple brin (ADNc pour ADN complémentaire) en utilisant une enzyme d'origine virale, la transcriptase inverse. Cet ADN sert ensuite de matrice pour la réaction de PCR classique.

▲ FIGURE 19. Étapes et schématisation de la PCR : une vision très simplifiée. D'après HARRY (2008).



Étape 1: dénaturation. On part d'un mélange contenant l'amorce présente en excès (généralement une séquence synthétique de 20 à 30 nucléotides) et du fragment d'ADN à amplifier. Ce mélange d'amorces et de fragments est chauffé à environ 98°C. À cette température, le fragment d'ADN bicaténaire se dissocie en brins monocatéinaires.

Étape 2: renaturation (annealing) des amorces. La solution est ensuite refroidie à 60°C environ. En refroidissant, les brins isolés d'ADN se réassocient en brins bicatéinaires. Cependant, en raison de l'important excès d'amorce, chaque brin du fragment forme des paires de bases avec une amorce complémentaire voisine de la région à amplifier, le reste du fragment restant simple.

Étape 3: extension de l'amorce. On ajoute maintenant une ADN polymérase très thermostable, la polymérase Taq (du nom de la bactérie thermophile *Thermus aquaticus*, d'où provient Taq), ainsi qu'une provision des quatre nucléotides. À partir de l'amorce, la polymérase recopie le reste du fragment comme s'il répliquait l'ADN. Quand la synthèse est terminée, l'amorce a été allongée d'une copie complémentaire de tout le fragment monocatéinaire. Les deux brins d'ADN ayant été répliqués, on a maintenant deux exemplaires du fragment d'origine.

Réaction en chaîne de la polymérase. (1) *Dénaturation*: une solution contenant des amorces et le fragment d'ADN à amplifier est chauffée pour séparer l'ADN en brins monocatéinaires. (2) *Renaturation des amorces*: la solution est refroidie et les amorces s'unissent aux séquences complémentaires de l'ADN bordant la séquence à amplifier. (3) *Extension des amorces*: l'ADN polymérase copie ensuite le reste des différents brins en commençant par l'amorce. Les étapes 1-3 sont répétées avec les brins répliqués. Ce processus se répète de nombreuses fois, doublant chaque fois le nombre de copies jusqu'à ce que le nombre de copies du fragment d'ADN soit suffisant pour l'analyse.

▲ FIGURE 20. Étapes et schématisation de la PCR : une autre vision très simplifiée.
D'après RAVEN et al. (2007).

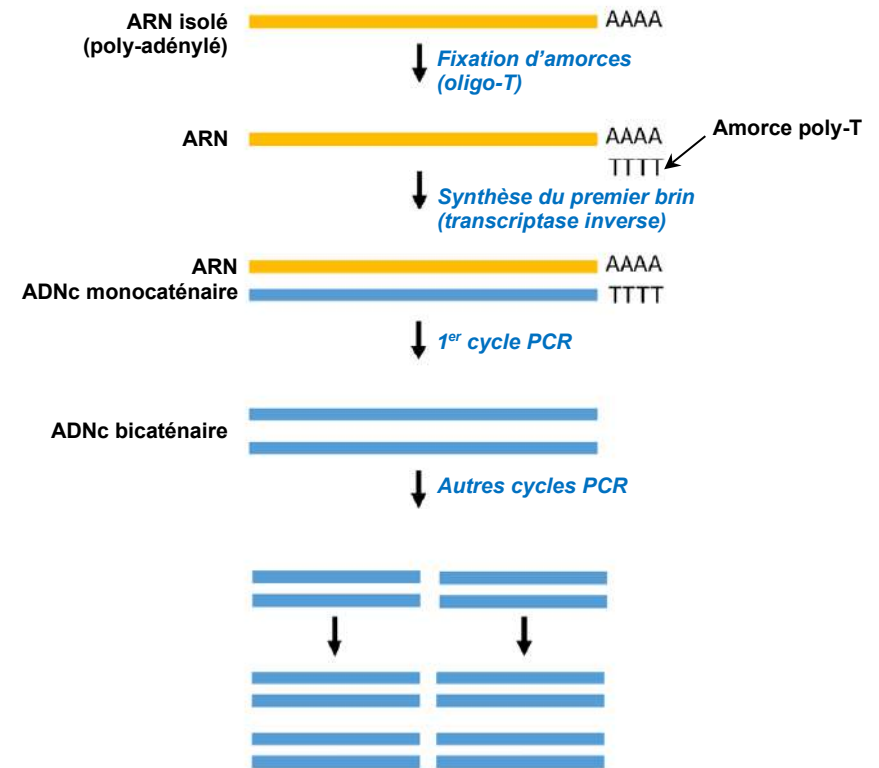
On réalise généralement entre **au maximum 20 et 40 cycles d'amplifications**, ce chiffre étant difficile à dépasser car les **réactifs (amorces, nucléotides...)** viennent tout simplement à **manquer**.

b. L'amplification de séquences d'ADN issues d'ARN : la PCR après rétrotranscription (RT-PCR)

- L'**amplification directe d'ARN** n'est pas immédiatement possible et suppose une **rétroconversion de l'ARN en ADN** ou **transcription inverse** (= **rétrotranscription**) permise par une enzyme nommée **rétrotranscriptase** ou **transcriptase inverse** (transcriptase reverse si on « anglo-saxonise »), généralement d'**origine rétrovirale**.
- Dans une **RT-PCR** (*Reverse Transcription Polymerase Chain Reaction*), **un ARN est rétrotranscrit en ADN avant que cet ADN soit à son tour soumis à une amplification par PCR** (figure 21).

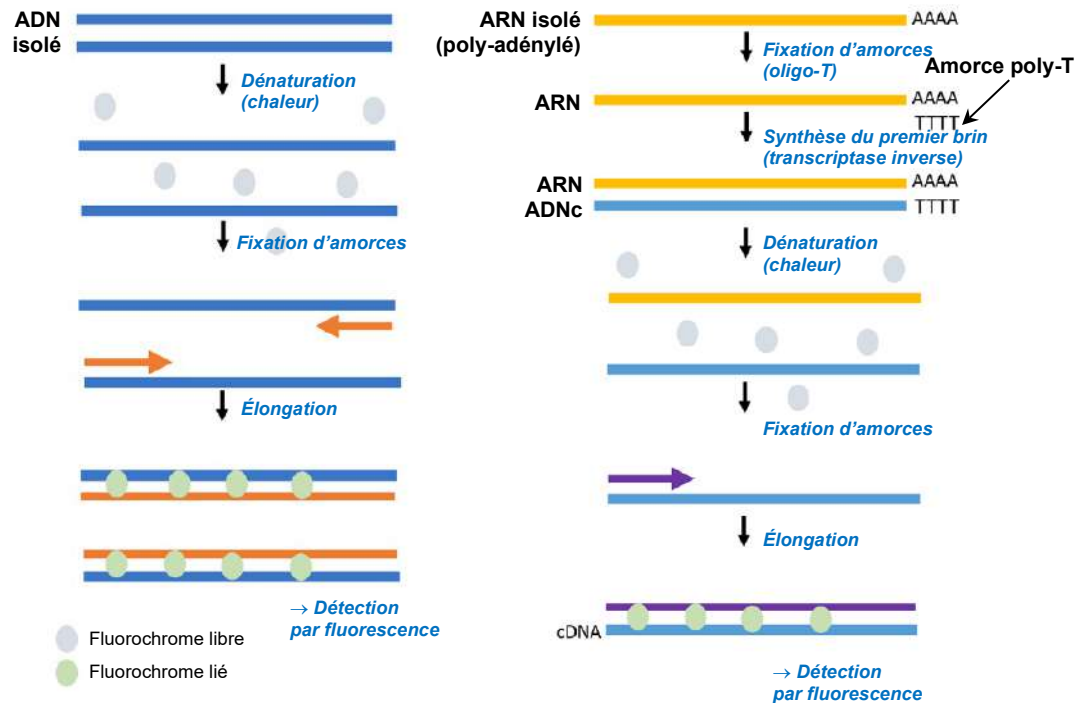
L'**ADN issu d'une rétrotranscription d'ARN** s'appelle un **ADN complémentaire (ADNc)**.

- La synthèse du **premier brin** requiert l'**ARN** (souvent préalablement **polyadénylé**) que l'on souhaite **amplifier**, des **amorces ADN**, une **transcriptase inverse**, des **dNTP**...
- D'**autres amorces (ADN)**, des **dNTP** toujours et une **ADN polymérase** permettent ensuite de produire, après **dénaturation** du **complexe ARN-ADNc**, le **second brin** et ainsi de **reconstituer un ADNc double-brin** sur lequel la **PCR classique** peut alors s'appliquer. Le **premier cycle de PCR** est souvent responsable de cette étape.



▲ FIGURE 21. Principe simplifié d'une RT-PCR. D'après ADAMS (2020), adapté / traduit.

c. Une approche quantifiée de la PCR et de la RT-PCR : la PCR et la RT-PCR quantitatives (= en temps réel) que l'on note qPCR et qRT-PCR



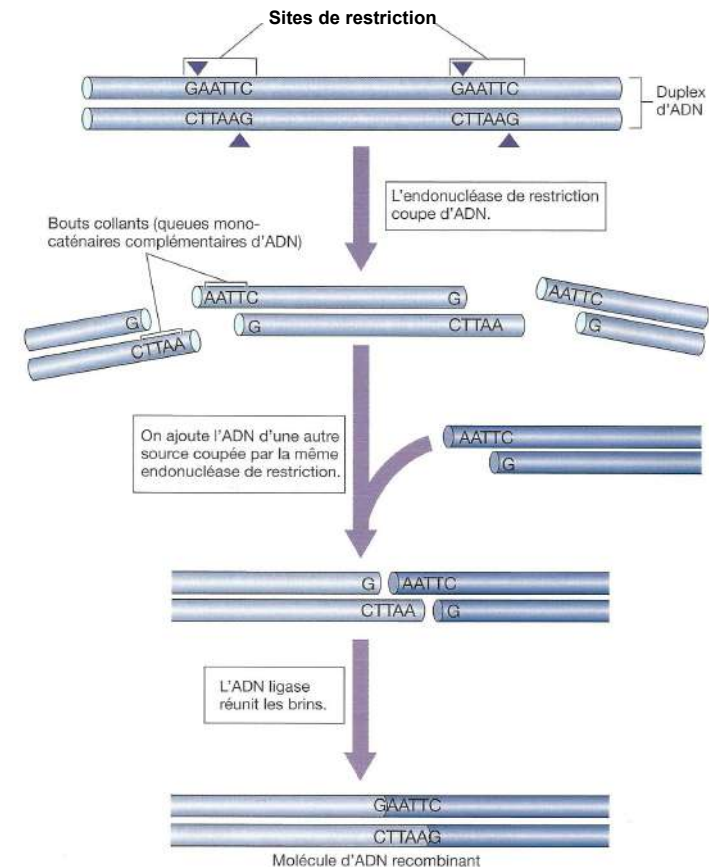
▲ FIGURE 22. Q-PCR et qRT-PCR. D'après ADAMS (2020), adapté / traduit.

- On parle d'**amplification en temps réel (real-time)** d'acides nucléiques, ou **amplification quantitative d'acides nucléiques**, notée **qPCR** pour l'ADN et **qRT-PCR** pour les ARN (le q signifiant 'quantitative' en anglais), pour désigner les **PCR ou RT-PCR permettant de mesurer, en temps réel, la quantité d'amplicons** (= nombre de copies d'ADN produits par amplification) produits grâce à un agent intercalant fluorescent (ou plus rarement du BET détecté par rayonnement UV) (figure 22).
- L'agent intercalant s'intercale à peu près régulièrement dans l'ADN double brin constitué et il suffit de mesurer la fluorescence, grâce à un capteur intégré au thermocycleur, pour en inférer la quantité d'ADN présente dans le mélange réactionnel en temps réel. La mesure intervient généralement à la fin de chaque cycle d'amplification.

Lorsque la mesure de la quantité d'amplicons est mesurée uniquement à la fin du processus, on parle de **PCR (ou RT-PCR) en point final**. La PCR (ou RT-PCR) quantitative suppose au contraire une mesure au fur et à mesure du processus, d'où la notion de « temps réel ».

5. L'utilisation d'enzymes de restriction dans le découpage de l'ADN

a. Des endonucléases bactériennes qui reconnaissent et découpent des séquences-types généralement palindromiques (sites de restriction)



Beaucoup d'endonucléases de restriction produisent des fragments d'ADN avec des « bouts collants ». L'endonucléase de restriction *EcoRI* coupe toujours la séquence GAATTC entre G et A. La même séquence se retrouvant sur les deux brins, tous deux sont coupés. Cependant, les deux séquences produites sont orientées différemment dans les deux brins. En conséquence, les queues monocaténaire sont complémentaires, elles sont « collantes ».

▲ FIGURE 24. Action des enzymes de restriction et production de bouts collants. D'après RAVEN et al. (2007).

Nucléase : enzyme capable de cliver les liaisons phosphodiester entre deux nucléotides dans un acide nucléique. Une **RNase** ou **ARNase** est une nucléase agissant sur les ARN alors qu'une **DNase** ou **ADNase** est une nucléase agissant sur l'ADN.

- Une **endonucléase** est une nucléase qui clive un acide nucléique en milieu de chaîne.
- Une **exonucléase** est une nucléase qui clive un acide nucléique par une extrémité (5' ou 3'), rarement les deux en même temps.

- Les **enzymes de restriction** sont des **protéines de type endonucléase** qui **coupent l'ADN au niveau d'une courte séquence de nucléotides caractéristique, généralement palindromique***, appelée **site de restriction** (figure 24).

* **palindromique** = qui peut se lire dans les deux sens.

Chaque **enzyme de restriction** reconnaît ainsi un **site spécifique**. Plusieurs **centaines d'enzymes** de restriction sont **actuellement connues** ; on les retrouve **naturellement** dans un **grand nombre d'espèces de Bactéries** où elles agiraient notamment comme **protection contre les bactériophages (virus bactériens)**.

Une **Bactérie** produisant une **enzyme de restriction** fabrique aussi une **deuxième enzyme** appelée **méthylase** qui est également spécifique du **site de restriction** où elle **modifie l'ADN bactérien** en lui **rajoutant un groupement méthyle** sur un ou plusieurs nucléotides. Cette **méthylation empêche la coupure du site de restriction** par l'**enzyme de restriction**. C'est la **garantie** que l'**enzyme de restriction** n'attaquera que de l'**ADN d'origine exogène** mais pas l'**ADN bactérien**.

Ce système de défense bactérienne, associant toujours les deux enzymes, s'appelle **système de restriction-modification**.

- Les **enzymes de restriction** permettent de **fragmenter l'ADN** lors de son **étude**. C'est une **étape importante** dans le **séquençage traditionnel** et de **deuxième génération**.

b. L'existence de sites de restriction en baïonnette permettant de générer des bouts collants utiles en génie génétique

▼ **TABLEAU IV. Quelques enzymes de restriction.** D'après PEYCRU *et al.* (2013).

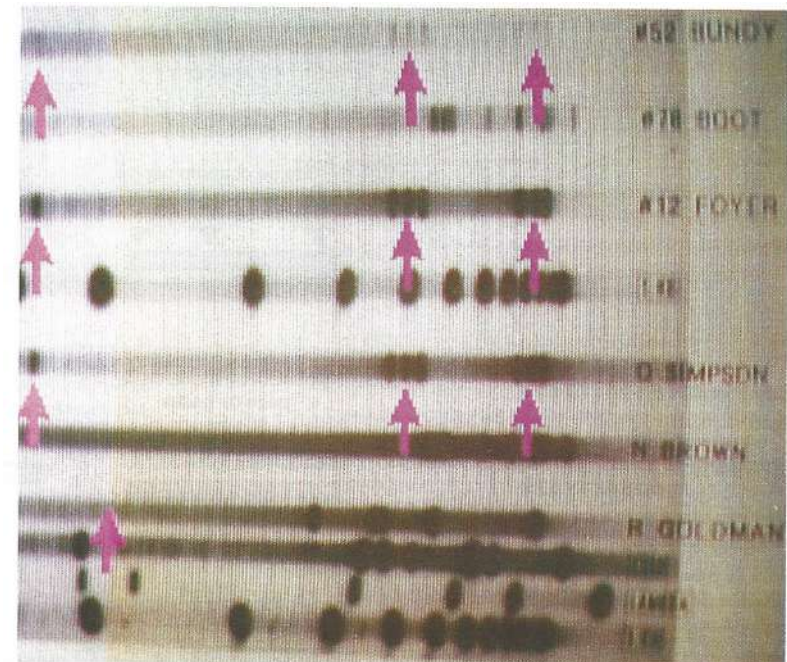
Enzyme	Micro-organisme producteur	Type de coupure	Séquence cible (généralement de type palindrome)
HpaI	<i>Haemophilus parainfluenzae</i>	Coupure franche	<pre> 5' ← G T T A A C C A A T T G → 5' </pre>
EcoRI	<i>Escherichia coli</i>		<pre> 5' ← G A A T T C C T T A A G → 5' </pre>
Hind 3	<i>Haemophilus influenzae</i>	Extrémités cohésives (coupure en baïonnette)	<pre> 5' ← A A G C T T T T C G A A → 5' </pre>
PstI	<i>Providencia stuarti</i>		<pre> 5' ← C T G C A G G A C G T C → 5' </pre>

- Les **sites de restriction** peuvent être (tableau IV) :
 - De petites **séquences coupées de manière transversalement nettes** ou **coupures franches**.
 - de petites **séquences coupées en « zigzag »**, dites **coupures en baïonnette**, ce qui produit de **courtes queues monocaténaires** pouvant **se lier entre elles** au niveau d'un **même site de restriction**.

On dit ainsi qu'une **enzyme de restriction** alors **produit des bouts collants (= extrémités cohésives)** (figure 24, tableau IV). Cette caractéristique est particulièrement **utilisée en génie génétique** pour **ajouter un gène d'intérêt** à un **ADN cible** ; on utilise généralement une **ADN ligase** pour **suturer les bouts collants** lors du processus.

c. L'utilisation d'enzymes de restriction dans la production de certaines empreintes génétiques

- On appelle **empreinte génétique** (ou **profil génétique**) un **résultat d'analyse génétique permettant l'identification d'un individu**.
- Ce type d'outil est utilisé dans les **tests de paternités**, les **enquêtes policières** (pour retrouver un **criminel** dont l'**ADN** a été récupéré sur un lieu d'exaction par exemple) mais aussi couramment dans les **études de polymorphisme** au sein des **laboratoire de recherche**.
- Des **méthodes variées de profilage génétique** existent.



Profils d'ADN de O.J. Simpson et des échantillons de sang provenant de la scène du meurtre de son ancienne épouse. On s'est servi de ces profils comme preuve lors du procès pour meurtre très médiatisé et controversé de 1995.

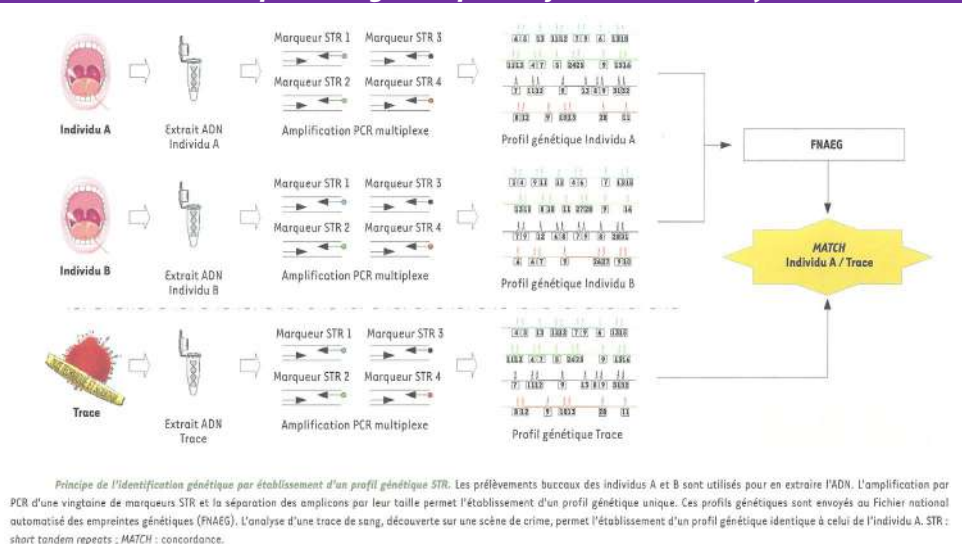
▲ **FIGURE 25. Rôle des empreintes génétiques dans les enquêtes policières.** D'après RAVEN *et al.* (2007).

- On a longtemps utilisé – et on utilise encore parfois – des **techniques faisant appel à une enzyme de restriction ou un cocktail d'enzymes de restriction découpant le génome des individus (désormais souvent après amplification PCR)**. Le génome découpé subit ensuite une **électrophorèse** et une **révélation** par des **une ou plusieurs sondes** dans le cadre d'un **Southern Blot** (voir **plus loin**). Comme les individus ont une **diversité de séquences propres**, le **nombre et la localisation de nombreux sites de restriction** est **propre à l'individu** d'où un **découpage de l'ADN** par les enzymes de restriction **propre à l'individu**, à l'origine d'un **profil génétique unique propre à chaque individu** (figure 25).

On a traditionnellement utilisé les **profils RFLP** (Restriction Fragment Length Polymorphism, polymorphisme de longueur des fragments de restriction) qui se faisaient **sans amplification** mais nécessitaient une assez **grande quantité d'ADN**. On a ensuite réalisé des **profils AFLP** (Amplified Fragment Length Polymorphism, polymorphisme de longueur des fragments amplifiés) avec une **amplification préalable de séquences**.

- Depuis **près de 30 ans**, on emploie plutôt l'**analyse STR** (short tandem repeats) (encadré D).

Encadré D Les empreintes génétiques aujourd'hui : l'analyse STR



D'après LAURENT et al. (2017)

➤ Pour réaliser des **empreintes génétiques** aujourd'hui, on utilise **l'analyse STR (short tandem repeats)** : celle-ci repose sur **l'amplification par PCR d'une vingtaine de séquences microsatellites ou STR** (séquences très répétées du génome et hautement variables longues de quelques pb) avec des amorces spécifiques couplées à un fluorochrome puis leur électrophorèse avant observation ; les pics de fluorescence sont caractéristiques des séquences présentes chez un individu. L'électrophorèse ici employée est une électrophorèse capillaire (dans un capillaire de quelques µm à dizaines de µm de diamètre), compatible avec des espèces chimiques à pouvoir de migration très rapide.

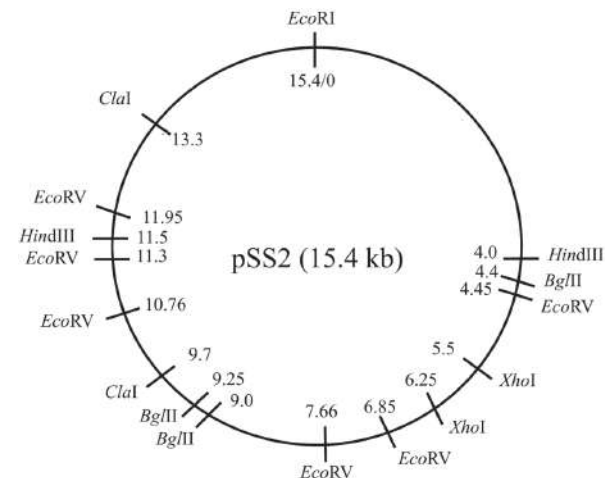
➤ Les **séquences STR utilisables** et leur possibilité de **stockage** sont codifiées par la loi.

➤ En France, un **fichier national** recense les **empreintes génétiques** ayant été mises en cause dans le cadre d'**enquêtes judiciaires** : le **FNAEG** (fichier national automatisé des empreintes génétiques).

- Le **séquençage (partiel ou total) de génome** commence à être **ponctuellement utilisé** dans certains pays.

d. L'établissement de cartes de restriction

- On appelle **carte de restriction** une **représentation schématique « cartographique » d'une unité génétique (plasmide, chromosome bactérien, génome viral, chromosome eucaryote...) sur laquelle sont reportées les sites de restriction d'une enzyme de restriction ou d'un ensemble choisi d'enzymes de restriction** (figure 26).
- Cette **carte** peut être **construite** :
 - Soit par **soumission à une enzyme**, puis **une autre, etc.**, puis **deux enzymes** ensembles, etc. de l'**unité génétique étudiée** avec, à chaque fois, **migration électrophorétique des fragments** pour connaître la **taille des fragments obtenus**. Le **croisement des données** (aujourd'hui assisté informatiquement) permet de **reconstituer progressivement la carte**.
 - Soit par **séquençage de l'unité génétique étudiée**, ce qui permet de **visualiser directement la position des sites de restriction**.



▲ FIGURE 26. **Carte de restriction du plasmide de résistance aux tétracyclines pSS2 des Bactéries Aeromonas**. D'après BALASSIANO et al. (2007).

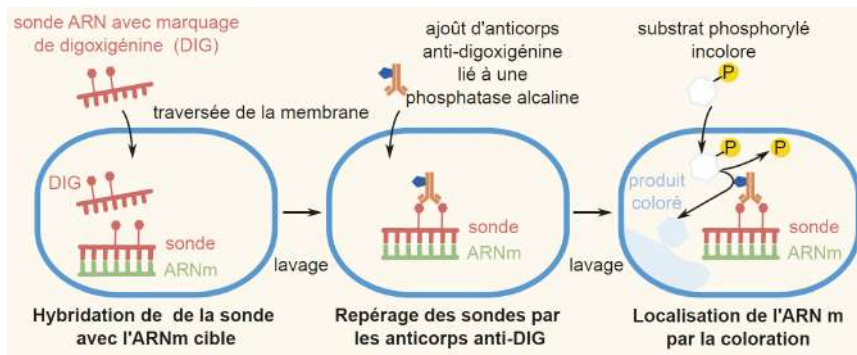
6. Des outils de recherche et de détection de séquences nucléotidiques spécifiques

- La **recherche** et la **détection de séquences** dans une **cellule** ou un **organisme** peut s'opérer selon **diverses techniques** ; nous en exposons quelques-unes ici.

a. L'hybridation *in situ* d'ARN ou d'ADN monobrin

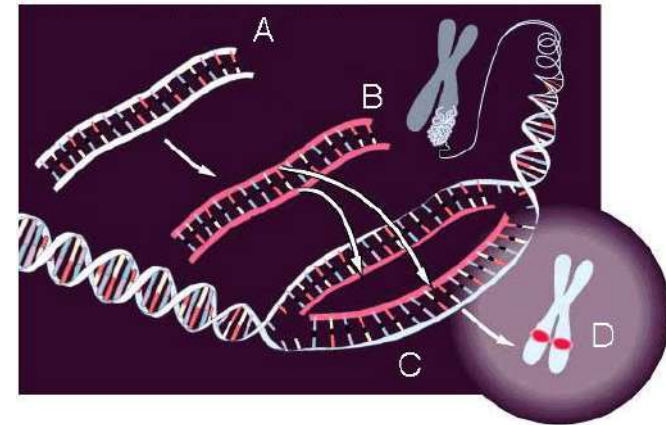
- Une **hybridation moléculaire** consiste à **associer expérimentalement deux molécules complémentaires qui s'associent ainsi spontanément**.
- Elle est dite ***in situ*** quand on la réalise **dans les structures biologiques (cellules, tissus)**.
- Une **hybridation *in situ* (HIS) d'acides nucléiques** désigne une **technique de mise en évidence, dans des tissus fixés, d'une séquence donnée d'ARN ou d'un ADN monobrin (nécessitant alors, dans ce second cas, une dénaturation préalable) par l'usage d'une sonde nucléotidique (ARN ou ADN généralement monobrin) marquée (isotopes radioactifs, produit fluorescent, enzyme, haptène...)** qui se fixe par complémentarité de base sur la séquence d'intérêt.

L'une des approches repose sur l'utilisation de digoxigénine (DIG), une molécule stéroïde fixée sur la sonde qui est un **haptène**, c'est-à-dire une **molécule susceptible de déclencher une réaction immunitaire si elle est associée à une molécule porteuse**. La DIG est soumise à des anticorps, eux-mêmes soit couplés à un fluorochrome, soit couplés à une enzyme. Dans ce second cas, il suffira alors de soumettre l'échantillon au substrat de l'enzyme et d'observer alors la formation du **produit coloré localisant la séquence d'intérêt** (figure 26).



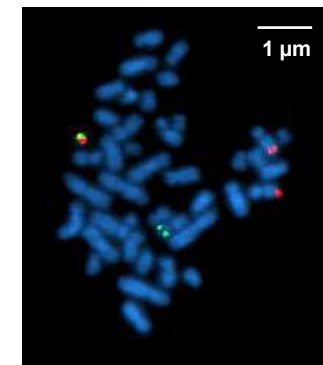
▲ FIGURE 26. **HIS utilisant la digoxigénine.**
D'après PERRIER, BEAUX *et al.* (2021).

- Cette **technique** permet de **mettre en évidence** des ARNm, des **petites séquences ADN** ou carrément des **portions chromosomiques**.
- Diverses **variantes** existent. Dans l'**hybridation *in situ* par fluorescence (FISH, Fluorescence in situ hybridization)**, de l'**ADN double brin** peut être utilisé comme sonde pour localiser des **portions chromosomiques précises** comme les **centromères**, les **télomères** ou d'autres **régions chromosomiques** (figures 27-28).



▲ FIGURE 27. **Principe de la FISH (hybridation *in situ* par fluorescence).** Wikipédia (2022)

- A**. Sonde ADN bicaténaire.
- B**. Marquage fluorescent de la sonde.
- C**. Dénaturation (par la chaleur) et hybridation de la séquence cible avec les brins de la sonde.
- D**. Possibilité d'observation : la fluorescence localise le locus de la séquence d'intérêt.



▲ FIGURE 28. **Principe de la FISH (hybridation *in situ* par fluorescence).** Wikipédia (2022)
Chromosomes métaphasiques (colorés au DAPI) montrant des sondes (fluorescence rouge et verte) fixées sur des séquences d'intérêt.

b. Des outils de recherche de séquences ADN ou d'ARN exprimés par hybridation avec des sondes : les *Blots* (buvardages avec emploi de révélateur)

a. Principe général

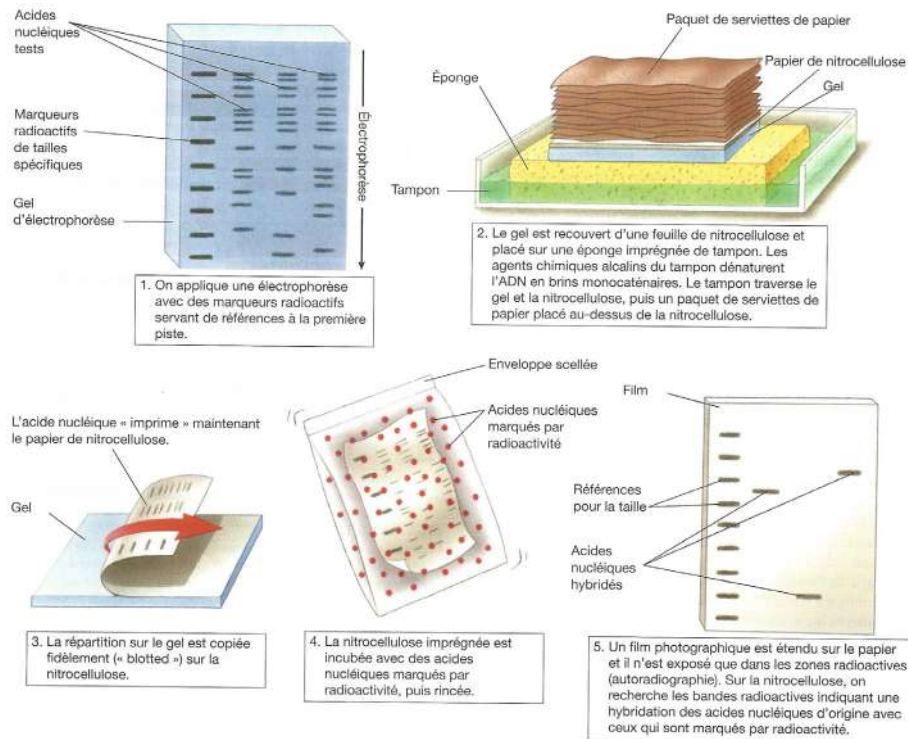
- Les ***Blots*** visent à **identifier certaines protéines ou certaines séquences particulières d'acides nucléiques après migration électrophorétique au moyen d'anticorps complémentaires de protéines (couplés à un fluorochrome, une enzyme à produit coloré...)** pour les protéines, ou de sondes ADN ou ARN (radiomarquées) pour les acides nucléiques.

- La technique est parfois qualifiée de « **buvar dage** » en français (traduction discutable de **blot** qui signifie « marque, tache ») puisqu'elle suppose la **réalisation d'une réplique du profil électrophorétique** sur une **membrane de nitrocellulose** : celle-ci est produite en faisant **migrer par capillarité** sur la **membrane** les molécules ayant **migré dans le gel d'électrophorèse** posé en dessous et saturé en tampon (figures 29-30).
- Les **taches obtenues** sont d'autant plus **intenses** que la **molécule révélée** par la sonde est en **quantité importante** dans l'échantillon d'origine.

L'importance du témoin de charge

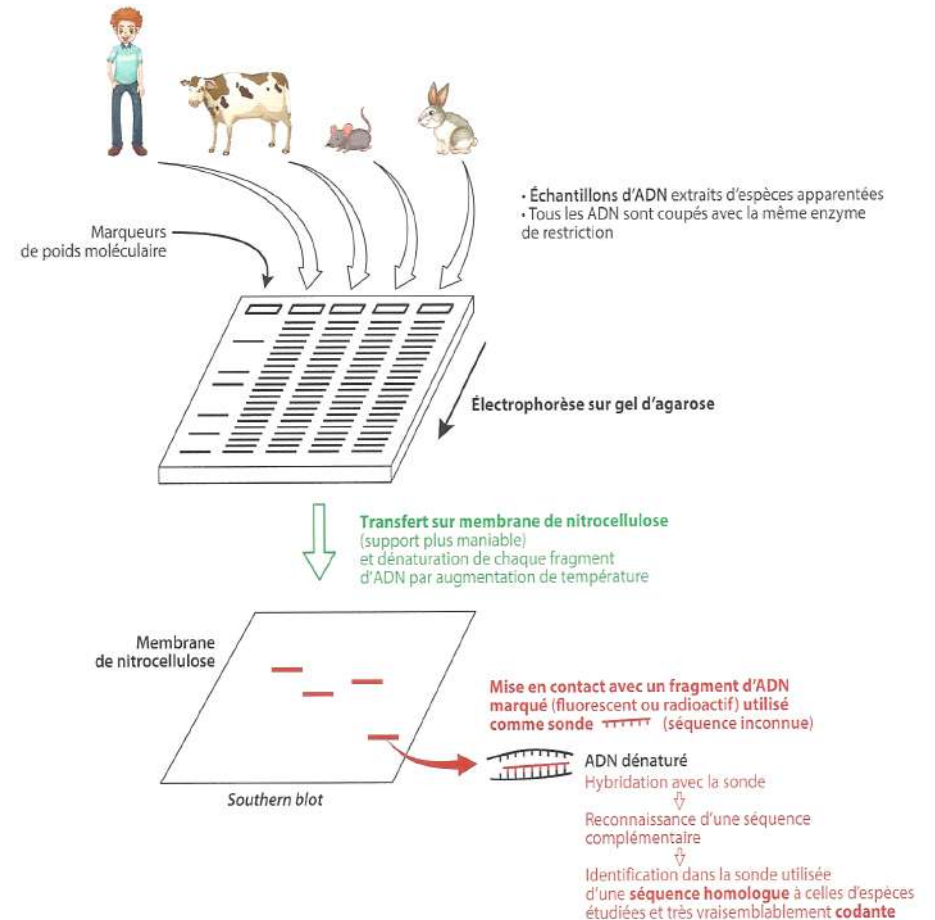
Un **témoin de charge** qui est **une molécule (ARN ou protéine en résultant) dont l'expression est constante dans la situation d'étude**. Il a deux intérêts :

- Il constitue un **contrôle de purification (témoin d'extraction)** garantissant le bon déroulement de l'extraction des ARN ou protéines, et donc la validité du Blot.
- Il peut servir dans les tentatives de quantification d'expression. En effet, si cette molécule est régulièrement et constitutivement exprimée indépendamment de la situation expérimentale, on pourra alors normaliser l'expression génétique constatée des **molécules d'intérêt** sur celle du **témoin de charge**.



La méthode « Southern blot ». E.M. Southern a mis au point cette méthode en 1975 pour mettre en évidence des fragments d'ADN d'intérêt dans un mélange complexe contenant de nombreux autres fragments de même taille. L'ADN est séparé sur un gel, puis transféré (« blotté ») sur un support solide comme du papier de nitrocellulose ou une membrane de nylon. Ce support est ensuite incubé avec une copie monocaténaire du gène d'intérêt qui s'hybride à la tache où se trouve un fragment possédant une séquence complémentaire. Les fragments d'intérêt sont identifiés sur le support par la position des bandes radioactives.

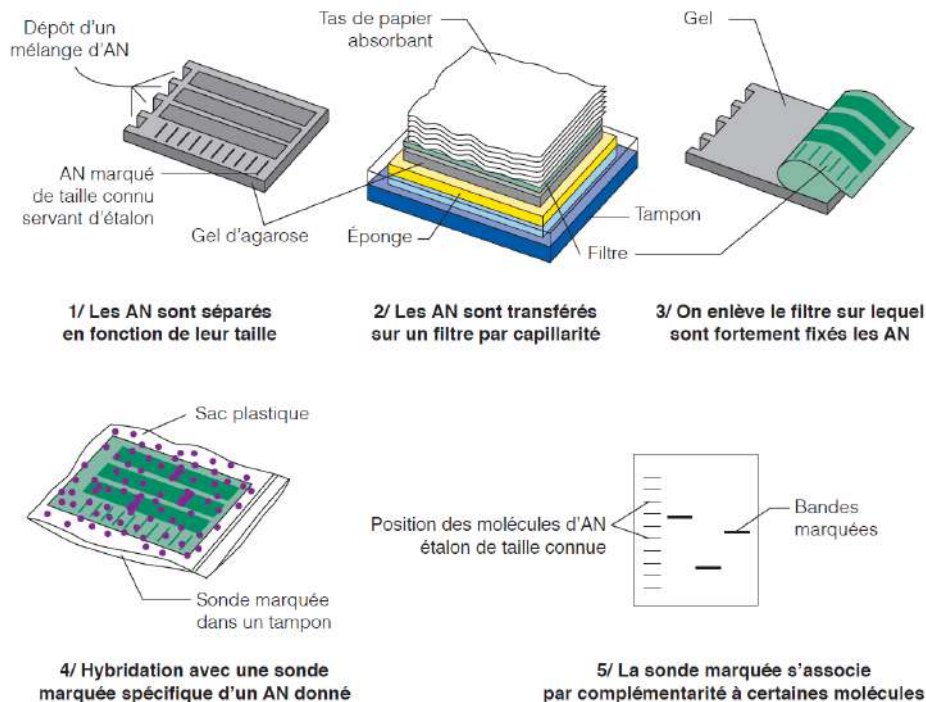
▲ FIGURE 29. **Le Southern Blot (étude des ADN)**. D'après RAVEN *et al.* (2007).



▲ FIGURE 30. **Le Southern Blot (étude des ADN) : une autre vision**. D'après DAUTEL *et al.* (2021).
(!) Les marqueurs de tailles sont généralement radioactifs, ce qui manque sur cette figure...

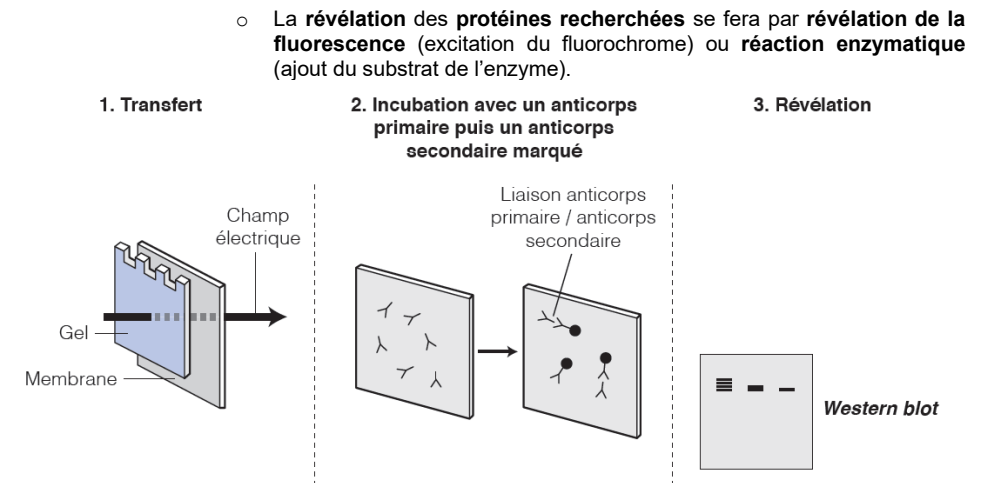
β. Southern Blot (ADN), Northern Blot (ARN) et Western Blot (protéines) : diversité des Blots

- On distingue **trois types de « Blots »** selon la nature chimique de la molécule étudiée :
 - Southern Blot** (figures 29-31) : il permet l'étude des ADN. Il comprend les étapes suivantes :
 - Électrophorèse des fragments d'ADN à étudier (+ un marqueur de taille).
 - Buvar dage du profil électrophorétique : une réplique du profil est réalisée sur une **feuille de nitrocellulose** par migration d'une solution tampon.



▲ FIGURE 31. Le Southern Blot (étude des ADN) et le Northern Blot (étude des ARN) : une vision simplifiée. D'après SEGARRA *et al.* (2014).
AN = acide nucléique.

- Incubation de la nitrocellulose avec des sondes ADN monobrin (ou d'ARN) marquées (souvent radioactivement) complémentaires des séquences ADN recherchées qui s'hybrident avec ces séquences cibles.
- Révélation (recherche de la radioactivité) sur la nitrocellulose pour localiser les **séquences cibles**.
- **Northern Blot** (figure 31) : il permet l'étude des ARN. Il présente le même principe que le Southern Blot sauf que les séquences cibles sont des ARN ; les sondes peuvent être des ARN ou des ADN monobrans.
- **Western Blot** (figure 32) : il permet l'étude des protéines qui sont révélées par immunodétection après buvardage. On désigne parfois cette technique par le terme « **immunobuvardage** ».
 - Après migration électrophorétique, un buvardage du profil obtenu (Blot) est réalisé soit par un champ électrique, soit au moyen d'un solvant tampon comme pour les autres Blots.
 - Cette réplique est alors soumise à un anticorps anti-protéine cible (anticorps primaire) ; on utilise généralement ensuite un anticorps secondaire (anti-Ac primaire) le plus souvent couplé à un fluorochrome, une enzyme... Il arrive toutefois qu'il n'y ait qu'un seul anticorps qui permette directement la révélation.



▲ FIGURE 32. Principe du Western Blot (étude des protéines). D'après SEGARRA *et al.* (2014).

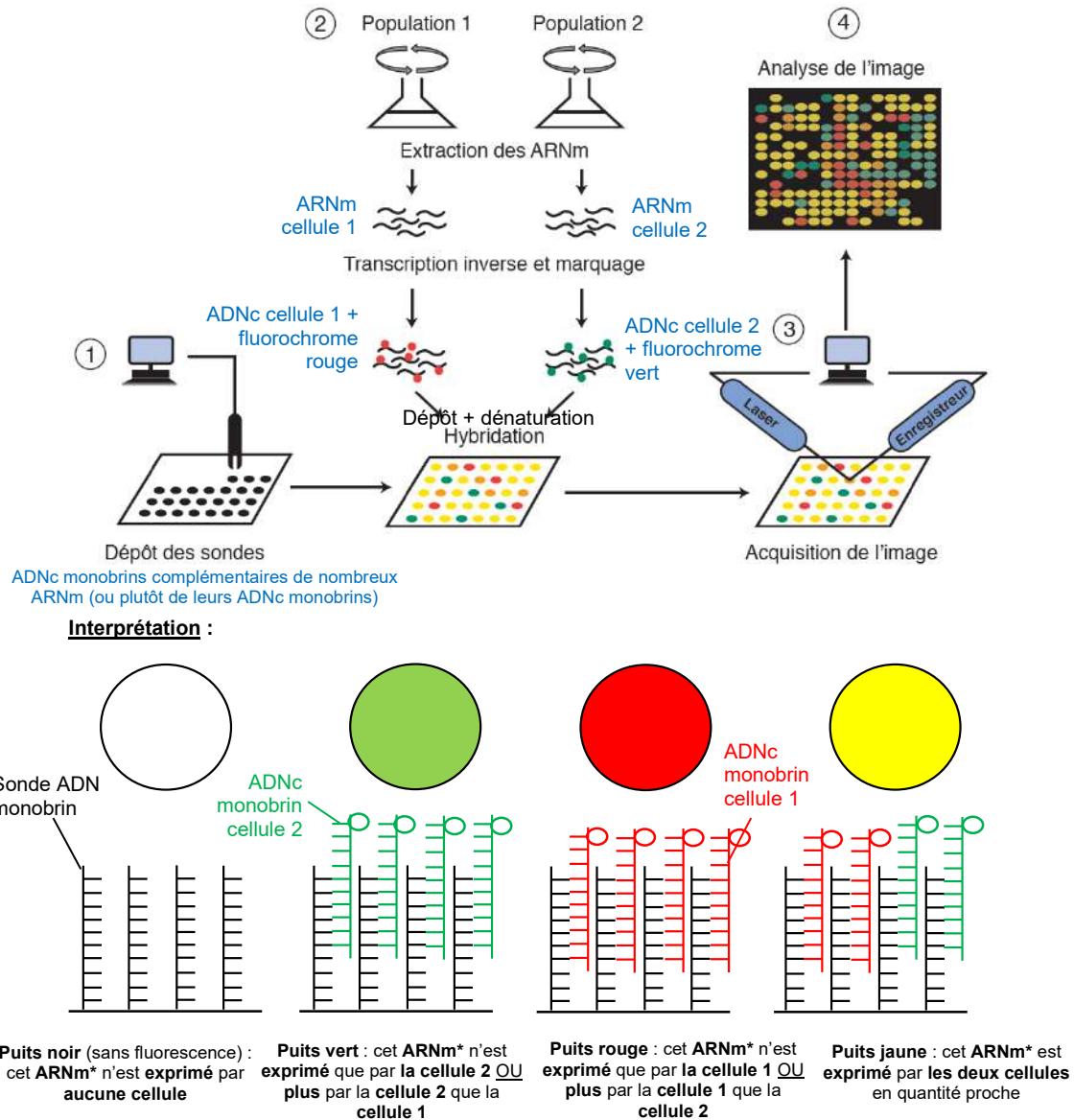
Le **Southern Blot** est le premier test inventé en 1975 par le Britannique **Edwin M. SOUTHERN** (né en 1938). Par la suite, les autres tests ont été **humoristiquement** baptisés avec d'autres **points cardinaux** en hommage à ce scientifique. **L'Eastern Blot n'existe pas !**

c. Vers une automatisation de la recherche des ARN exprimés : les puces à ADN (biopuces), outil majeur en transcriptomique

- Le **transcriptome** est l'ensemble des ARN présents dans une cellule (ou un ensemble de cellules) à un moment donné. Son étude, ou **transcriptomique**, est possible par un **séquençage intensif des ARN** ou encore des **puces à ADN**.
- Les **puces à ADN** ou **biopuces** (figure 33) sont des **plaques rigides (en verre ou plastique)** dont la surface est subdivisée en **quelques milliers à quelques dizaines de milliers** de zones de dépôt d'acides nucléiques (**spots**) répartis sur quelques cm².
- **Chaque** zone de dépôt comprend une sonde à ADN monobrin (ADN dénaturé), dont une **séquence précise**. Celle-ci est déposée, par un procédé assisté par ordinateur, des **ADN monobrans généralement construits à partir d'ARN rétrotranscrits ou de manière purement synthétique**.
- Les ARNm extraits du lysat de cellules à étudier sont eux aussi soumis à une **rétrotranscription en ADNc** (et souvent **amplifiés** par **PCR**). Ces ADNc sont couplés à des **fluorochromes** de type **cyanines** (ou jadis parfois **radioactifs**) ; ils sont ensuite ajoutés sur la **plaque** et **dénaturés**. La **révélation par fluorescence** permet de savoir si les **fragments d'ADNc déposés** se sont ou non **hybridés avec les sondes de la puce**, et donc si **telle ou telle séquence est exprimée** par une **population de cellules**. La lecture et le traitement des résultats sont informatisés.

Des **biopuces modernes** permettent de travailler avec des **ADN double brin**.

Notons qu'on peut étudier, en jouant sur **diverses couleurs**, **plusieurs populations de cellules conjointement** (souvent deux).



▲ FIGURE 33. **Principe des puces à ADN.** D'après SEGARRA *et al.* (2014), complété.
 Sous-entendu : * l'ARNm correspondant à l'ADNc obtenu par **rétrotranscription** et fixé dans le puits.

7. La production d'organismes génétiquement modifiés (OGM) : la transgénèse

a. Principes généraux et étapes de la transgénèse

α. Notion de transgénèse, gène d'intérêt et avantage sélectif

- La **transgénèse** est le **transfert, opéré en laboratoire, d'un gène d'intérêt (= transgène = gène hétérologue) depuis un organisme donneur vers un organisme receveur appartenant généralement à une espèce différente.**

Vous verrez parfois le terme de « **transfection** », de périmètre un peu variable en fonction des auteurs, pour désigner le **fait d'ajouter une portion exogène d'ADN (voire d'ARN, pour certains auteurs) dans des cellules eucaryotes** ; cette définition est parfois – mais pas toujours ! – complétée par la mention « **sans utiliser de vecteur viral** ».

- Le **gène d'intérêt (= transgène)** exprimé **permet alors à l'organisme receveur de bénéficier d'une nouvelle caractéristique génétique sur laquelle il est sélectionné** et qui **présente généralement un intérêt pour l'être humain** : on dit que la **transgénèse** confère un **avantage sélectif** au profit de l'espèce humaine (à ne pas confondre avec un **avantage sélectif au sens évolutif**).

β. Principales étapes du processus

- La **transgénèse** au sens le plus large comprend les étapes suivantes :
 - Isolément du gène d'intérêt à l'aide d'une **enzyme de restriction** et son identification grâce à une **sonde moléculaire**.
 - Amplification par **PCR** de ce gène.
 - [Construction de l'**ADN recombinant (ADNr)** (= **ADN artificiel obtenu en laboratoire par réunion de matériel génétique issu d'origines différentes**) en insérant le gène d'intérêt dans un **vecteur** comme un **plasmide** // parfois intégration directe dans le génome cible]
 - Insertion du gène d'intérêt dans le **génome** de l'organisme cible
 - Criblage** : **sélection des cellules ou organismes exprimant l'ADN recombinant (souches transformées)**.

γ. Intérêts et risques de la transgénèse

- La **transgénèse** est un **outil important** en **recherche scientifique** où elle a permis d'immenses **progrès**. C'est aussi un outil utilisé par l'**industrie pharmaceutique** (production par des **bactéries** d'**hormones**, de **vitamines**...), le **monde agricole et agro-alimentaire** (production de **variétés végétales résistantes** aux ravageurs ou aux herbicides, production de **variétés à croissance plus rapide** et/ou **plus importante**...).
- Toutefois, des **risques** (**environnementaux**, **sanitaires** voire **économiques** à cause de certains monopoles notamment dans les semences) existent et leur **évaluation** n'est pas forcément toujours réalisée sérieusement (**encadré E : exemple des plantes transgéniques**).

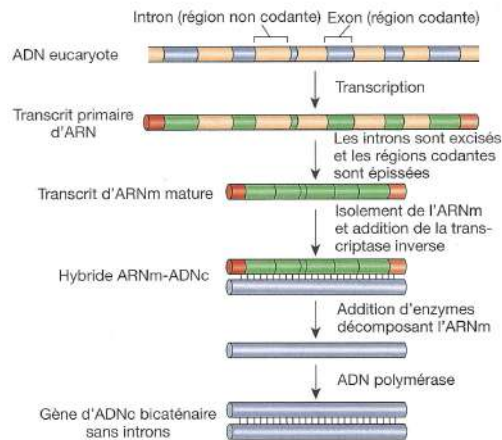
b. Isolement du gène d'intérêt

- Le **gène d'intérêt** est généralement obtenu par **fragmentation du génome** par des **enzymes de restriction** et **Southern Blot** sur les fragments obtenus.
- Notons que, si l'on souhaite exprimer un **gène morcelé eucaryote** chez des **Eubactéries**, on produit un **gène artificiel sans introns** généralement au moyen de l'**ARNm mature** du gène **réroconverti** (par la **transcriptase inverse**) en **ADNc monocaténaire** (**ADN complémentaire monobrin immédiatement obtenu**) puis (après action d'une **polymérase**) en **ADNc double brin** (figure 34).

Encadré E Les OGM en agriculture : entre bénéfices et risques

	Bénéfices	Risques
Agriculture	- Augmentation du rendement - Diminution de la charge de travail : moins de traitements chimiques et mécaniques - Plus de sécurité : moins d'utilisation de pesticides toxiques - Diversification des débouchés (ex : pharmacie)	- Dépendance de l'agriculteur à des firmes agro-chimiques - Contraintes pour les agriculteurs pour la coexistence de filières OGM et non OGM - Contamination de champs non transgéniques
Société	- Produits avec de nouvelles qualités organoleptiques et nutritionnelles - Production de nombreuses molécules (plante-usine)	- Problèmes éthiques : brevetabilité du vivant - Risques sanitaires encore peu cernés
Environnement	- Diminution des intrants (pesticides) : baisse de la pollution agricole des sols et des nappes phréatiques - Meilleur respect de la faune - Phytoremédiation	- Flux des transgènes hors culture - Risques de perturbation des équilibres biologiques

D'après MEYER *et al.* (2008)



Production d'ADNc. On isole un transcrit mature d'ARN du cytoplasme d'une cellule. On utilise une enzyme, la transcriptase inverse, pour fabriquer un brin d'ADN complémentaire de l'ARNm mature. Cet ADN sert de modèle pour l'ADN polymérase, enzyme qui assemble en face de lui un brin complémentaire d'ADN et donne l'ADNc, ADN bicaténaire représentant l'ARNm sans introns.

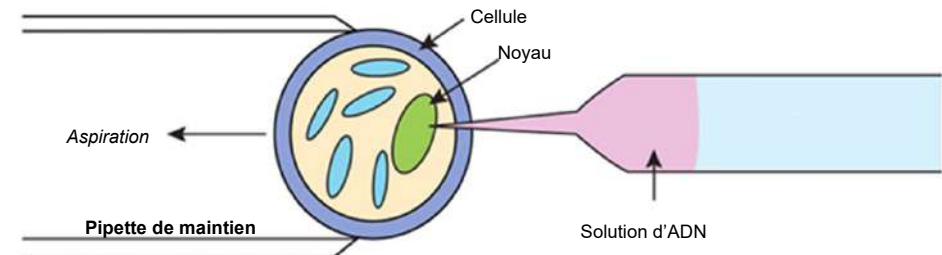
▲ FIGURE 34. **Technique de l'ADNc (ADN complémentaire).** D'après RAVEN *et al.* (2007).

c. Insertion directe ou indirecte du gène d'intérêt (transgène) dans le génome cible

- Deux solutions se présentent, avec souvent l'inconvénient que l'on maîtrise mal à quel endroit le gène d'intérêt (= **transgène**) va s'insérer :
 - Soit le gène est inséré directement dans le génome cible.
 - Soit le gène est inséré dans un **vecteur** qui permettra l'insertion du gène d'intérêt dans la cellule cible et son expression.

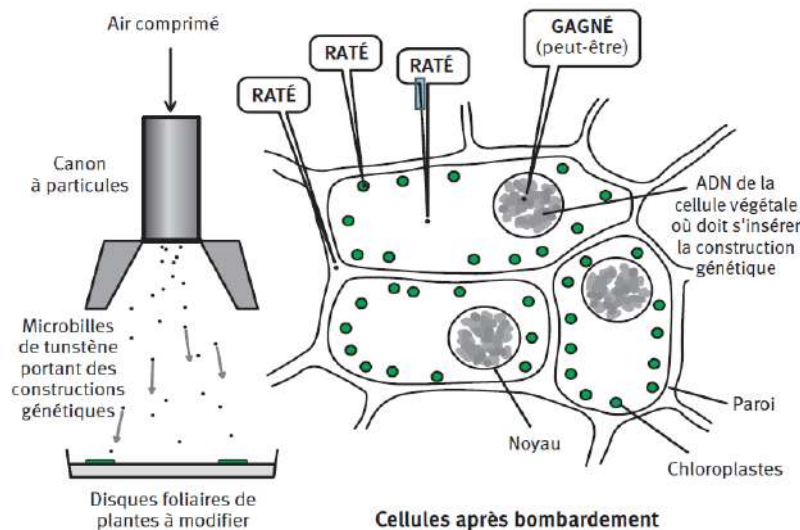
a. Quelques modalités classiques d'insertion directe : électroporation, micro-injection et biolistique

- On peut citer :
 - L'**électroporation** : sur des **cellules bactériennes**, des **cellules végétales sans paroi (protoplastes)** ou des **cellules embryonnaires (ou zygotes) d'Animaux**, on applique un **champ électrique qui déstabilise les membranes**, ce qui permet l'entrée d'ADN extracellulaire dans la cellule puis le noyau.
 - La **micro-injection** (figure 35) : utilisée chez les Animaux mais aussi sur des **protoplastes végétaux**, il s'agit d'**injecter une solution contenant l'ADN d'intérêt directement dans le noyau des cellules à modifier**.
 - La **biolistique** (contraction de « bio- » et « balistique ») (figure 36) : utilisée sur les **organes végétaux** (feuilles notamment, ou bien **ensembles de cellules méristématiques obtenus artificiellement** qu'on nomme **cals**), il s'agit de **faire pénétrer les ARN d'intérêts par le biais de billes d'or ou de tungstène propulsés sous pression au moyen d'un canon à particules**. On régénère ensuite une **plante entière** par **multiplication végétative artificielle** des cellules exprimant le gène d'intérêt et ne semblant **pas déstabilisées** (car hélas le gène peut s'insérer n'importe où, y compris au milieu de **séquences codantes** qui sont alors **endommagées**).



▲ FIGURE 35. **Micro-injection d'ADN.**

<http://nptel.ac.in/courses/102103016/module3/lec24/3.html> (consultation décembre 2015).



▲ FIGURE 36. **Biolistique**. © Académie en ligne (consultation décembre 2015).

β. Transformation de la cellule cible via un vecteur plasmidique ou viral

- On peut également **transformer génétiquement** des cellules en produisant un **vecteur (plasmide recombiné, génome viral recombiné)** (figure 37).

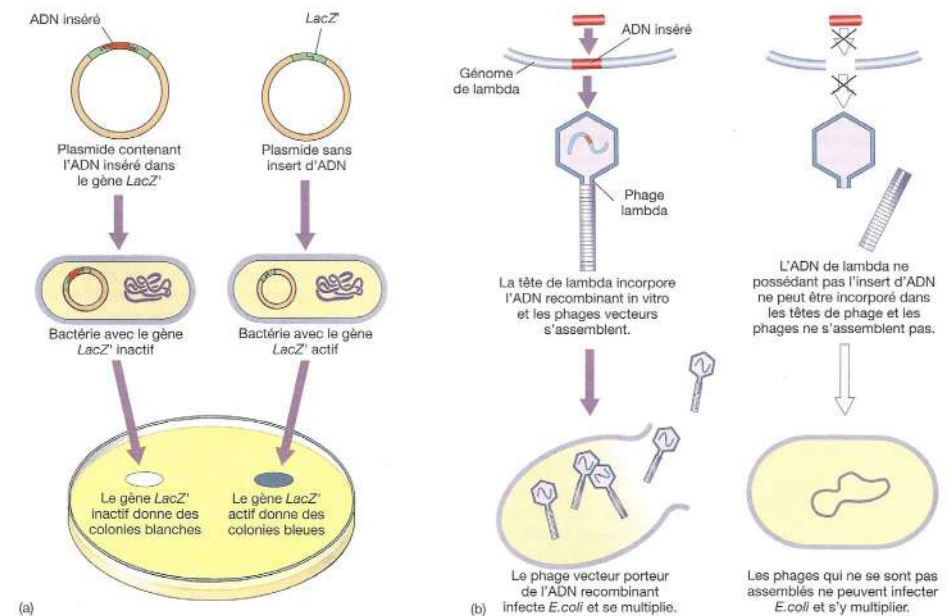
i. Utilisation de plasmides

- Les **plasmides recombinants** sont des **plasmides intégrant un gène d'intérêt et produits au moyen d'enzymes de restriction en ajoutant le gène d'intérêt et également, très souvent, un gène rapporteur**. Ce dernier est un **gène dont l'activité est facile à observer ou tester (gène d'une enzyme comme la luciférase, gène GFP, gène de résistance aux antibiotiques...)** qui permettra de **savoir si la transformation a réussi ou échoué**.
- Cette technique est **très utilisée chez les Bactéries**.
- Cette technique peut aussi être utilisée sur **certaines plantes sensibles à la Bactérie *Agrobacterium tumefaciens* (= *Rhizobium radiobacter*)** responsable de tumeurs. Cette Bactérie possède un **plasmide *Ti* (Tumor inducing Plasmid)** qui est **capable de s'insérer dans la 'plante' et de se transmettre ensuite aux cellules végétales produites par divisions cellulaires tout en provoquant des tumeurs végétales**. On utilise alors des *Agrobacterium* dont le **plasmide a été modifié** : suppression ou **inactivation des gènes de virulence**, ajout du **gène d'intérêt** et d'un **gène rapporteur...** (figure 38).

Je continue d'utiliser le genre *Agrobacterium* comme cela correspond à mon vieux schéma mais le nom actuel est bien *Rhizobium radiobacter*.

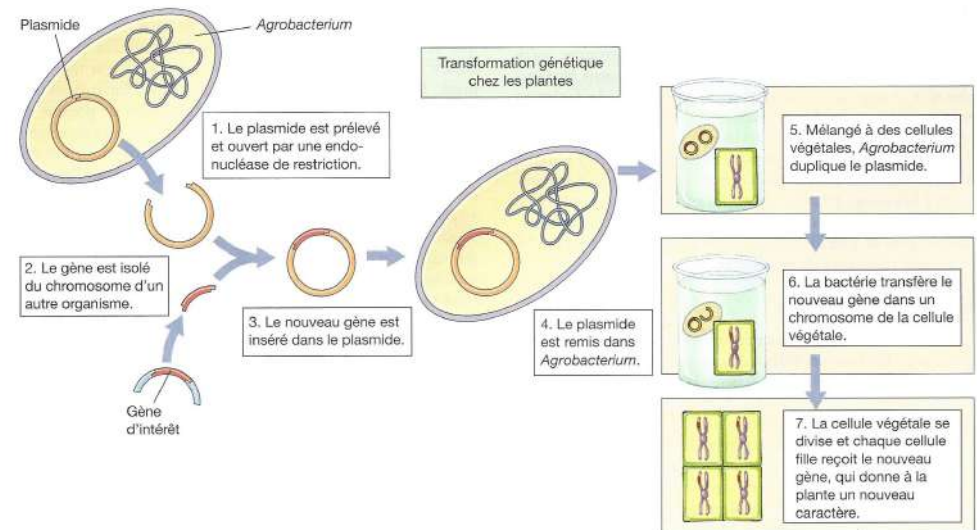
ii. Utilisation de virus

- Chez les **Bactéries**, on utilise un **Bactériophage** (figure 37) dont le **génom est modifié** : insertion du **gène d'intérêt** et du **gène rapporteur**, **favorisation du cycle lysogénique...** Il y a ainsi **modification durable du génome cible**. On peut aussi utiliser le **cycle lytique** (figure 37) simplement pour **multiplier les vecteurs**.



Utilisation des plasmides et des phages comme vecteurs. (a) On peut cribler les bactéries hôtes pour la présence des plasmides recombinants en observant la présence d'un phénotype chez l'hôte, comme l'absence de coloration bleue si la bactérie est étalée sur un milieu contenant X-gal. (b) On peut sélectionner les phages vecteurs possédant l'ADN recombinant parce qu'ils sont capables de s'assembler in vitro, d'infecter un hôte et de s'y multiplier.

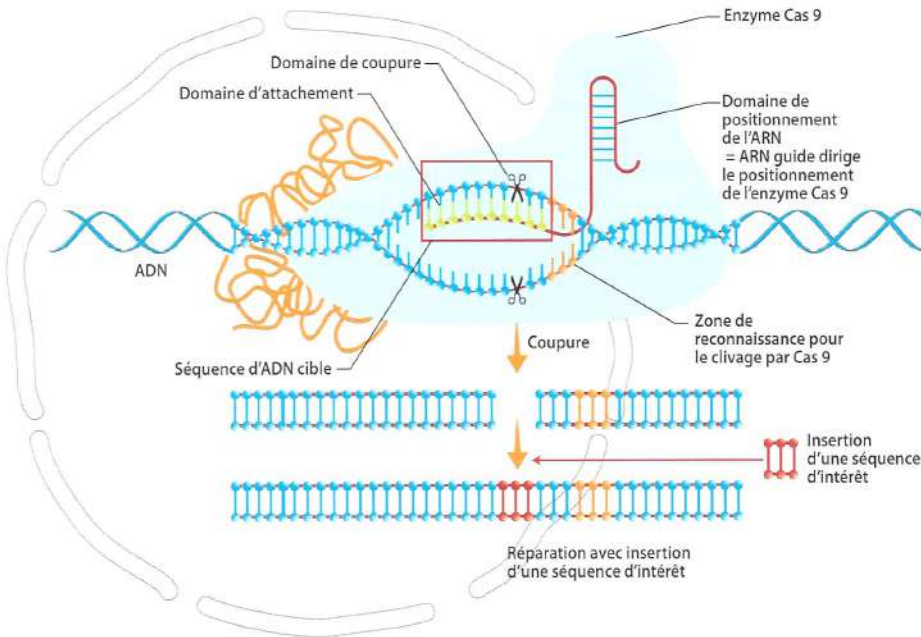
▲ FIGURE 37. **Plasmides et phages, des vecteurs**. D'après RAVEN et al. (2007). Notons que la production d'un OGM stable par bactériophages suppose une induction du cycle lysogénique.



▲ FIGURE 38. **Transformation génétique végétale par plasmide *Ti* modifié d'*Agrobacterium* (*Rhizobium radiobacter*)**. D'après RAVEN et al. (2007).

- Chez les **Eucaryotes**, on peut utiliser un **rétrovirus modifié** : insertion du **gène d'intérêt** et du **gène rapporteur**, **suppression** ou **inhibition** des **gènes de virulence**... Il y a là encore **modification durable** du génome cible. *Notons qu'il s'agit là d'une technique également employée en thérapie génique.*
- Inconvénient** de ces techniques : l'insertion du **provirus** est **aléatoire** et fait courir le **risque** d'une **altération de séquence codante**.

y. Une technique en plein essor d'édition du génome : les ciseaux génétiques (cas de la protéine CRISPR Cas9)



▲ FIGURE 39. Génie technique utilisant Crispr Cas9. D'après DAUTEL et al. (2021).

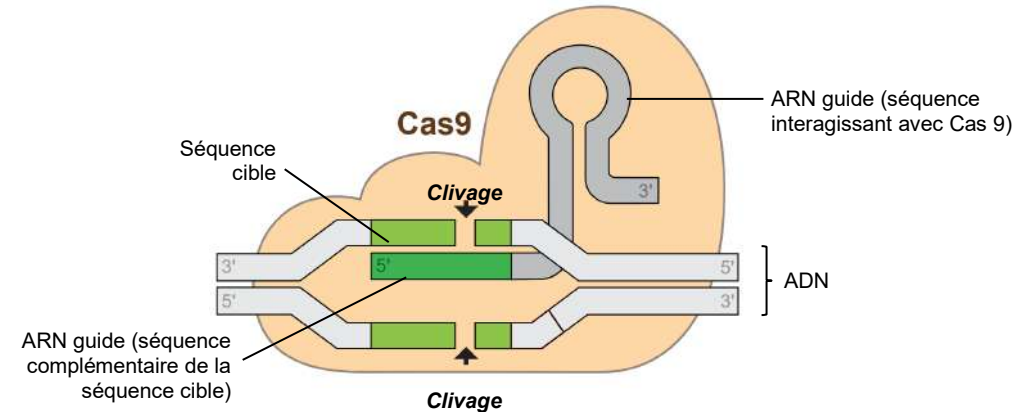
- La **protéine Cas9** (CRISPR* associated protein 9) ou **CRISPR Cas9** est une **protéine à activité endonucléasique d'origine bactérienne** – **initialement à fonction antivirale** – capable de **couper l'ADN en presque n'importe quel point précis guidée par un ARN guide modifiable par l'expérimentateur**.

* **CRISPR** : Clustered Regularly Interspaced Short Palindromic Repeats, **Courtes répétitions palindromiques groupées et régulièrement espacées** = **familles de séquences répétées dans l'ADN où des répétitions directes courtes (de 21 à 37 paires de bases) sont régulièrement espacées par des séquences appelées spacer, généralement uniques, de 20 à 40 paires de bases**. En réalité, **CRISPR Cas 9** peut se fixer ailleurs que sur des **séquences CRISPR** dès lors qu'on modifie son **ARN guide**.

- Le terme **ciseaux génétiques** parfois employé fait référence à l'idée que **cet outil génétique permet de couper l'ADN un endroit précis et choisi, comme les ciseaux permettent de couper volontairement un objet de manière choisie**.

Cette technique relève des **méthodes d'édition du génome, ensemble de techniques visant à modifier le génome en un point précis**, là où la **transgénèse classique** présente un caractère d'insertion aléatoire et imprévisible. D'autres **techniques** existent mais l'usage de **Cas9** est la plus efficace et prometteuse parmi les **approches actuellement disponibles**.

- En **génie génétique**, la **protéine Cas9** peut être artificiellement chargée avec une **séquence d'ARN guide complémentaire d'une séquence** que le biologiste cherche à **éditer (modifier)**. Le **complexe enzymatique** ouvre la **molécule d'ADN** et **coupe** alors la **séquence cible** suite à sa **reconnaissance** par l'**ARN guide** (figures 39-40).



▲ FIGURE 39. Une autre représentation de Cas9. Wikipédia (2022), adapté.

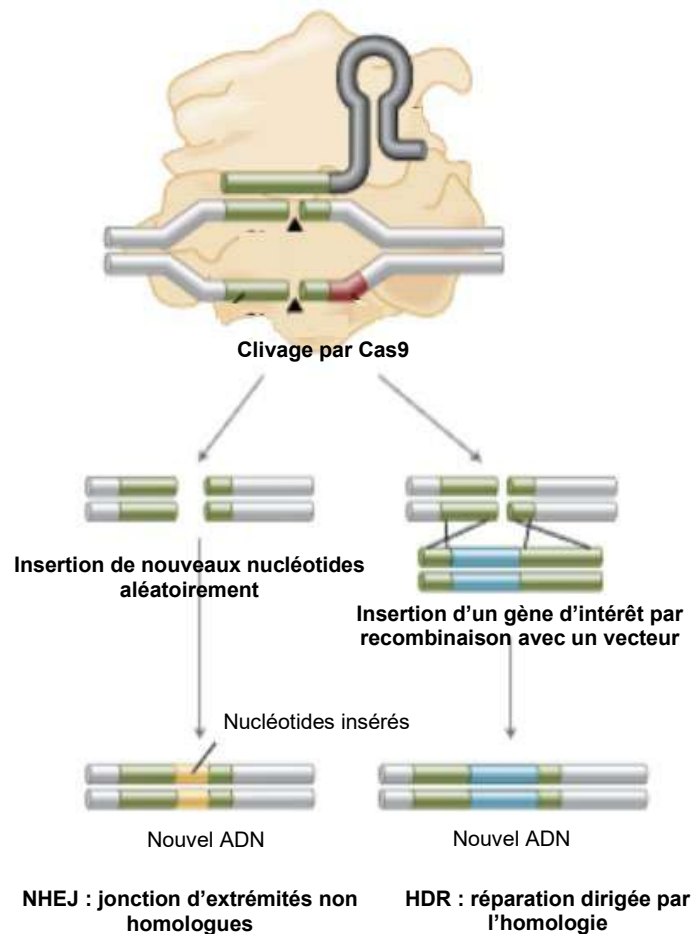
- Il est ensuite possible d'insérer, par **d'autres systèmes enzymatiques** reconnaissant la **coupure**, une **séquence d'intérêt** portée par exemple par un **plasmide**. La technique permet donc d'**introduire un nouveau gène** mais aussi, plus souvent, de **réparer** ou **modifier** un **gène** (voire de l'**inactiver**)... L'utilisation de cet outil en **transgénèse** devient **courante** et son emploi en **thérapie génique** est **prometteur**.
- Dans le **détail**, l'insertion de **nucléotides** ou de **gènes** après **coupure** se fait par deux **mécanismes principaux et naturels** que le **biotechnologiste** peut **détourner** à son avantage (figure 40) :

- La **jonction d'extrémités non homologues** ou **NHEJ** (*Non-Homologous End-Joining*) : **mécanisme hautement conservé dans le monde vivant, des systèmes enzymatiques complètent aléatoirement avec quelques paires de bases la zone cassée d'ADN**. La séquence initiale n'est pas restaurée et des **nouveautés génétiques** peuvent **apparaître** ou bien, si la **nouvelle séquence** est **peu efficace** ou **non codante**, rien de notable ne se passe.

En **génie génétique**, cette méthode permet de faire, par exemple, de la **mutagenèse dirigée**.

- La **réparation dirigée par l'homologie** ou **HDR** (*Homology directed repair*) : **mécanisme non encore totalement élucidé où un ADN présent dans la cellule (dans un vecteur, en cas de génie génétique, ou naturellement dans une autre portion du génome) est inséré au niveau de la coupure dans le seul cas où le nouvel ADN est encadré par des séquences homologues des séquences encadrant la coupure de l'ADN initial, ces séquences homologues étant échangées entre les deux entités génétiques**.

Cette méthode permet d'**insérer**, par exemple, un **transgène** en recourant à un **vecteur plasmidique**.



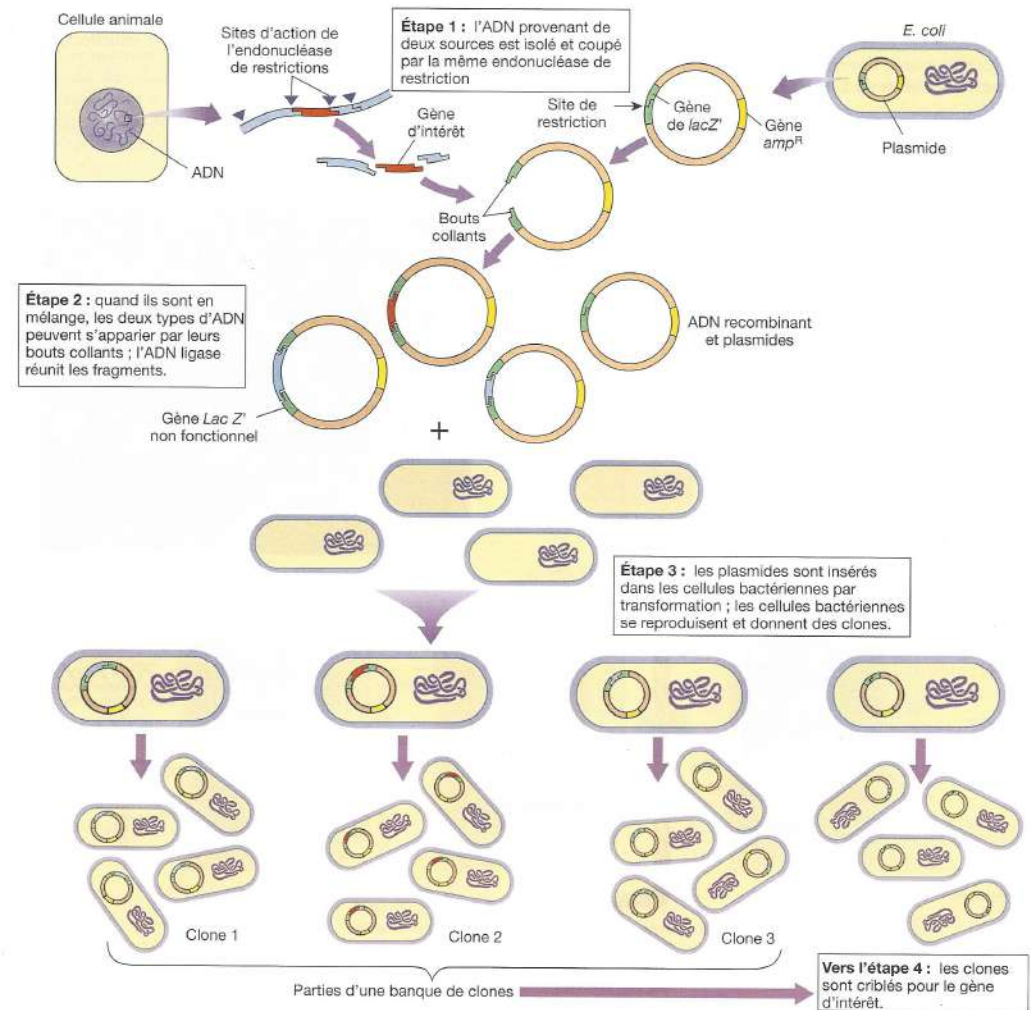
▲ FIGURE 40. **Édition et modification du génome par Cas9.** <https://www.neb.com/tools-and-resources/feature-articles/crispr-cas9-and-targeted-genome-editing-a-new-era-in-molecular-biology> (consultation février 2022), modifié et adapté.

d. Clonage des cellules après insertion du transgène

a. Principe général

- Très souvent, les **cellules** ou **organismes** ayant subi la transgénèse sont **clonés**, c'est-à-dire qu'on **produit des cellules ou organismes génétiquement identiques**, ce qui permet au passage de multiplier les séquences d'intérêt.

Un **clone** est, au sens premier, un **ensemble de cellules génétiquement identiques**. Chez les **organismes pluricellulaires**, on peut parler de **clones** pour désigner des **individus porteurs du même patrimoine génétique**.



Étapes d'une expérience de génie génétique. À l'étape 1, l'ADN contenant le gène d'intérêt (dans ce cas, provenant d'une cellule animale) et l'ADN d'un plasmide sont coupés par la même endonuclease de restriction. Les gènes *amp^R* et *lacZ'* se trouvent dans le plasmide et servent à cribler un clone (étape 4). À l'étape 2, les deux types d'ADN sont mélangés et s'apparient par leurs bouts collants. À l'étape 3, l'ADN recombinant est introduit dans une cellule bactérienne qui se reproduit et donne des clones. À l'étape 4, les clones bactériens seront criblés en fonction du gène d'intérêt.

▲ FIGURE 41. **Étapes d'une expérience de génie génétique incluant l'étape de clonage.**
D'après RAVEN *et al.* (2007).

- Dans tous les cas, les **cellules** ou **organismes** qu'on a essayé de **modifier génétiquement** seront (le plus souvent **après clonage**) **testés de manière à vérifier que la transgénèse a réussi** : c'est le **criblage** (voir E).

β. Cas des Bactéries

- Dans le cas des **Bactéries**, on laisse simplement les cellules **se reproduire naturellement** (figure 41).

γ. Cas des 'plantes'

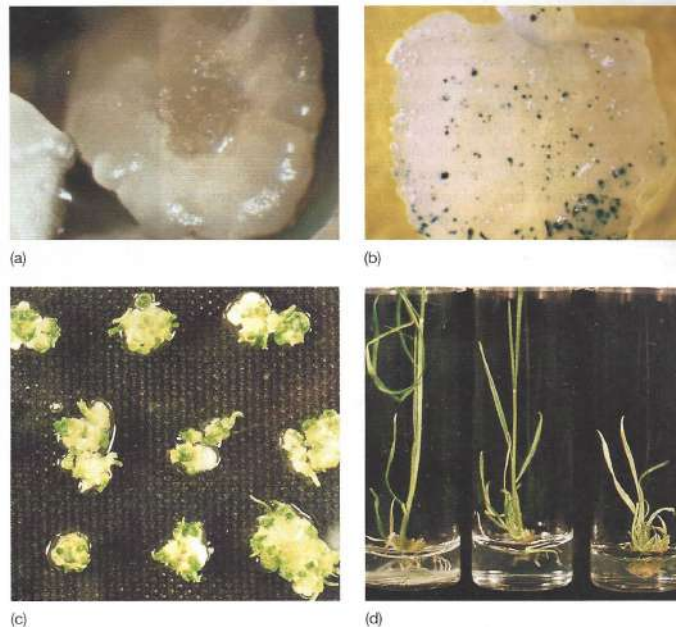
- Dans le cas des '**plantes**', on **cultive artificiellement les cellules** qu'on suppose **transformées** sur un **gel** en jouant sur **différentes hormones végétales**. Les cellules sont d'abord poussées à se **dédifférencier** en **cellules méristématiques** (il arrive que ce soit à ce stade qu'on pratique la transgénèse) puis à se **multiplier** ; enfin, on sépare les **ensembles artificiels de cellules méristématiques** et l'on pousse artificiellement (pas des hormones végétales) ces ensembles de cellules à se **différencier en plantes entières** (figure 42).

δ. Cas des Animaux

- Chez les **Animaux**, la **transgénèse** a souvent lieu sur des **zygotes** ou de **très jeunes cellules embryonnaires**. On laisse ensuite généralement **se développer** des **embryons transformés** mais sur **seulement quelques divisions** : en effet, alors que les **cellules** sont encore **totipotentes**, on les sépare artificiellement et **chaque cellule** donnera individuellement un **nouvel embryon**.

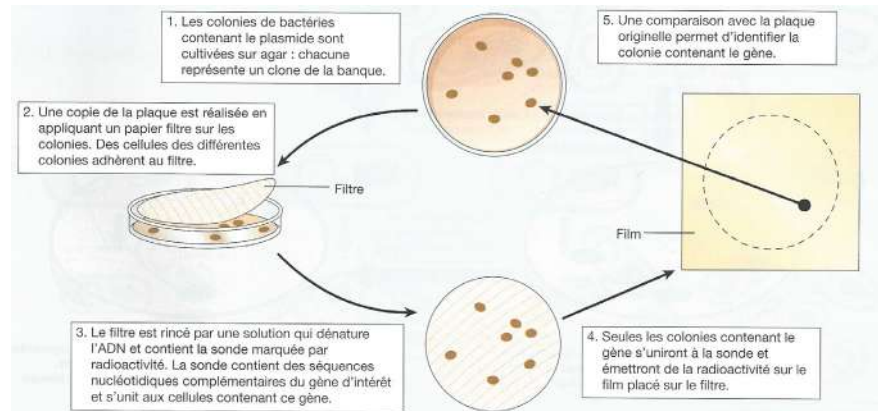
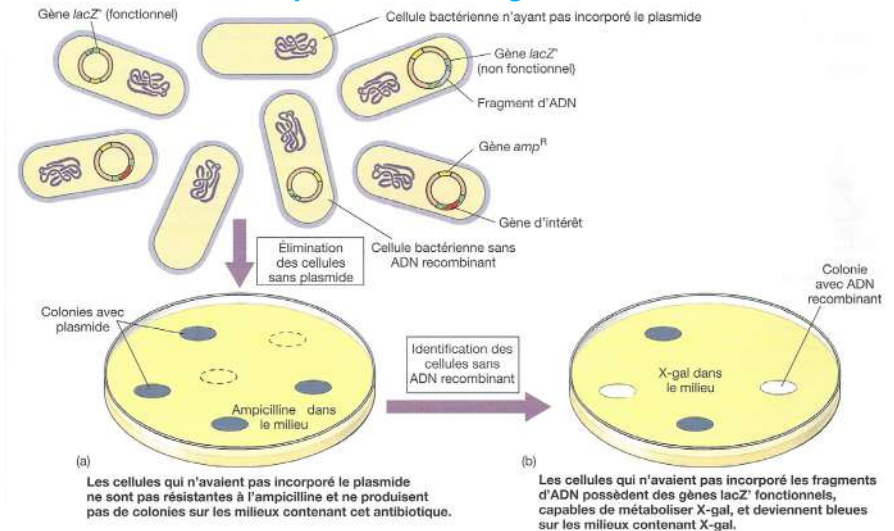
Après **criblage**, lorsque des **individus transgéniques et viables** ont été obtenus, les **nouveaux animaux transgéniques** sont produits en **croisant les animaux transgéniques** existants entre eux par **reproduction sexuée classique**.

Développement d'une plante transgénique. L'ADN contenant un gène de résistance à l'herbicide a été introduit chez le blé (*Triticum aestivum*). L'ADN contient aussi le gène *GUS*, qui sert de marqueur. Le gène *GUS* produit une enzyme qui catalyse la transformation d'une solution limpide en un colorant bleu. (a) Tissu embryonnaire immédiatement avant l'insertion de l'ADN étranger. (b) après le transfert, les cellules du cal contenant l'ADN étranger se reconnaissent à la couleur produite par le gène *GUS* (taches bleues). (c) Formation de tiges feuillées sur des cals transgéniques cultivés sur un milieu sélectif. Ici, le gène de résistance à l'herbicide permet le développement des cellules résistantes sur le milieu sélectif contenant l'herbicide. (d) Croissance comparée, sur le milieu sélectif, de plantes transgéniques possédant le gène de résistance (à gauche) et d'une plante non transformée (à droite).



▲ **FIGURE 42. Implication de la multiplication végétative artificielle dans la mise au point d'une plante transgénique (+ criblage en d).** D'après RAVEN et al. (2007).

e. Criblage génétique : sélection des souches transformées par vérification de l'insertion et de l'expression des transgènes



▲ **FIGURE 43. Criblage génétique : exemple chez des Bactéries transgéniques.** D'après RAVEN et al. (2007).

- Le **criblage** consiste à **tester les cellules ou individus obtenus après transgénèse afin de vérifier le succès (ou l'échec) du processus**. Il s'agit **soit de vérifier directement si les souches d'organismes modifiés expriment bien le gène d'intérêt, soit de vérifier si elles expriment le gène rapporteur dont on déduit le succès de la transgénèse** (figures 42-43).

8. L'élucidation de la fonction des gènes

- L'élucidation de la **fonction d'un gène** passe par diverses **techniques de génie génétique** : quelques-unes sont ici présentées. Une partie d'entre elles consiste à **invalider le gène et rechercher la fonction / la protéine manquantes ou diminuées** en conséquence.

a. La possibilité de muter le gène étudié ici en vue de son invalidation : la mutagenèse aléatoire ou dirigée

- On appelle **mutagenèse** le **processus d'apparition d'une mutation (généralement ponctuelle) dans l'information génétique**.
- Le terme peut aussi faire référence aux **techniques de génie génétique visant à provoquer artificiellement et volontairement la mutation d'une séquence**.

α. La mutagenèse aléatoire : le déclenchement de mutations non maîtrisées par des agents mutagènes (en des lieux variables) ou la PCR mutagène (sur une séquence précise)

i. Mutagenèse aléatoire par des agents mutagènes : des mutations à localisation et en nombre imprévisibles

- On appelle **agents mutagènes** des **agents capables d'augmenter le taux de mutation au sein de l'information génétique**. Il peut s'agir d'**agents physiques** (rayonnements UV ou ionisants) ou d'**agents chimiques** (analogues de bases, agents alkylants...).
- La **soumission** d'une cellule ou d'un ensemble de cellules à des **agents mutagènes** provoque **des mutations en nombre variable et non prévisible** qui, si elles ne sont **pas corrigées**, peuvent **persister** dans les **lignées cellulaires**.

Un excès de mutations et/ou une seule mutation mal placée peut conduire à des organismes non viables.

- L'**identification ultérieure des mutations** se fait par **identification de phénotypes mutants** même si ce travail est **fastidieux** et parfois **difficile** si l'expression des **modifications phénotypiques** est **discrète**.

ii. L'utilisation de la PCR mutagène : des mutations favorisées lors l'amplification d'une séquence

- L'**amplification par PCR** est **fiable** mais possède un **taux d'erreur minime**. Ce taux d'erreur peut être **artificiellement augmenté** par le manipulateur en modifiant la **solution tampon**, en **déséquilibrant** les quantités de **NTP**... Des **amplicons mutés** apparaissent alors et peuvent ensuite être **insérés** dans des **organismes**, *via* des vecteurs ou non, par **transgénèse**.
- Si l'on **sait** ici **quelle séquence** est **touchée** (celle qu'on a amplifiée), on ne peut **pas prédire** la **localisation** ni le **nombre** des **mutations** engendrées qui, de plus, peuvent **différer** entre les **différents amplicons**.

β. La mutagenèse dirigée : le déclenchement de mutations ciblées par PCR mutagène

- La **mutagenèse dirigée (par PCR)** est une **technique visant à modifier une séquence nucléotidique amplifiée par PCR en utilisant des amorces complémentaires de l'ADN cible à l'exception du (ou des) nucléotide(s) que l'on souhaite muter, ce qui engendre une amplification de séquences ayant la mutation spécifique souhaitée à l'endroit souhaité**.
- Des **variantes** jouant sur la **longueur des amorces** (on peut en utiliser plusieurs ou au contraire employer des **méga-amorces** de très grande taille) ou les **systèmes enzymatiques** éventuellement ajoutés au mélange existent.
- Les **amplicons** peuvent ensuite être **insérés** dans des **organismes** par **transgénèse**.

γ. L'édition génomique par le système CRISPR Cas9

- Comme nous l'avons vu **plus haut**, l'**édition du génome** par **Cas9** permet de produire des **allèles nouveaux / mutés**, notamment par **NEHJ**. Cette technique peut même être **utilisée in vivo** et constitue la voie de « **mutagenèse dirigée** » (même si le terme est normalement plutôt à réserver au **cas exposé en β**) qui sera sans doute privilégiée dans les **années à venir**.

b. La possibilité d'invalider un gène par l'introduction d'ARN interférents

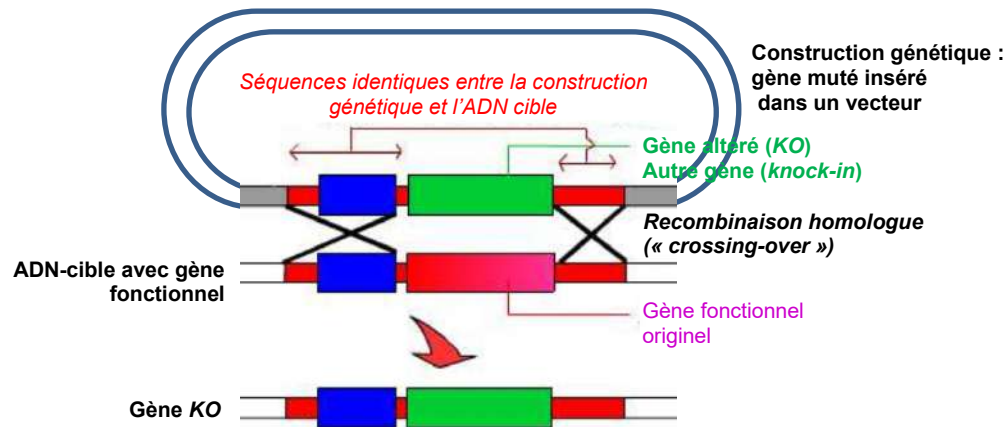
- L'**injection d'ARN** (ou d'**ADN monobrans**) **antisens** (= avec une **séquence complémentaire de l'ARNm visé**) permet, par **interférence**, de **déclencher la lyse** de l'**ARNm d'intérêt** et donc **d'empêcher son expression** sous forme de **protéine**.

c. La possibilité d'éteindre un gène par recombinaison homologue avec un allèle muté non fonctionnel (**knock-out KO**) ou de le remplacer par un autre (**knock-in**)

- Une **recombinaison homologue** est un processus où **deux portions d'ADN sont échangées entre deux molécules d'ADN grâce à un enjambement (crossing-over) au niveau de séquences semblables (homologues) encadrant les séquences subissant l'échange** (figure 44).
- On détourne généralement l'**appareil enzymatique** impliquée dans la **réparation dirigée par l'homologie (HDR)** **vue page 30** en utilisant un **plasmide** (présentant le **gène à insérer** encadré par des **séquences homologues des séquences encadrant le gène à remplacer**) introduit dans la cellule.
- Un **knock-out de gène** est une **extinction de gène typiquement par recombinaison homologue avec remplacement de l'allèle fonctionnel par un allèle muté invalidé**. Ici, il s'agit de **remplacer un ADN génique fonctionnel par un ADN génique muté ou incomplet** conduisant à **invalider l'expression du gène** dont on cherche à étudier le rôle.
- Un **knock-in de gène** est un **remplacement de gène typiquement par recombinaison homologue par un autre gène différent (par exemple un gène rapporteur), ce qui invalide le gène ôté mais assure l'expression à sa place du gène inséré souvent placé sous les mêmes séquences régulatrices**.

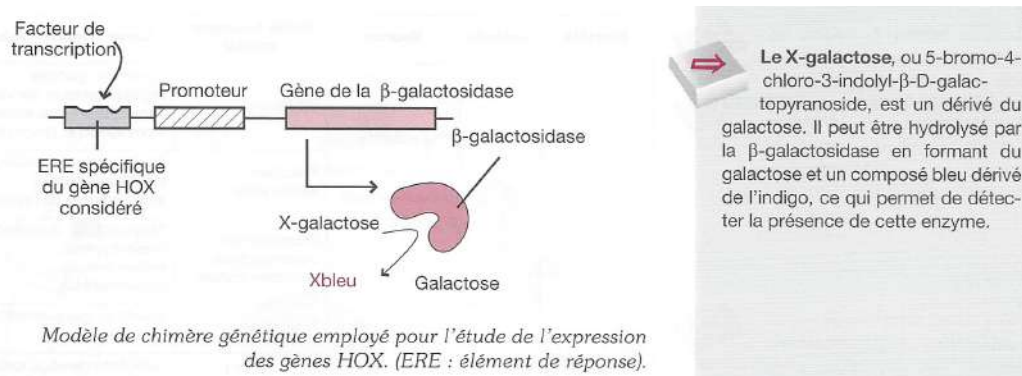
Dans ce cas, si le **gène de substitution** est un **gène exprimant une fluorescence**, un **composé coloré** ou encore une **activité enzymatique aisément détectable** (par soumission au substrat), alors on peut **localiser les lieux d'expression du gène originel**, voire parfois la **quantifier** (voir **d**).

- On utilise des **cellules-souches embryonnaires** dans lesquelles le **gène muté** est inséré au moyen d'un **vecteur** (entré dans le noyau par électroporation) qui réalise un **crossing-over** avec l'ADN-cible qu'il remplace.



▲ FIGURE 44. Principe du **knock-out** et du **knock-in**.
Document G. FURELAUD (*Planet-vie.ens.fr*), modifié.

d. La possibilité d'exprimer un (trans)gène rapporteur ayant même séquence régulatrice qu'un gène dont on cherche à localiser et dater l'expression



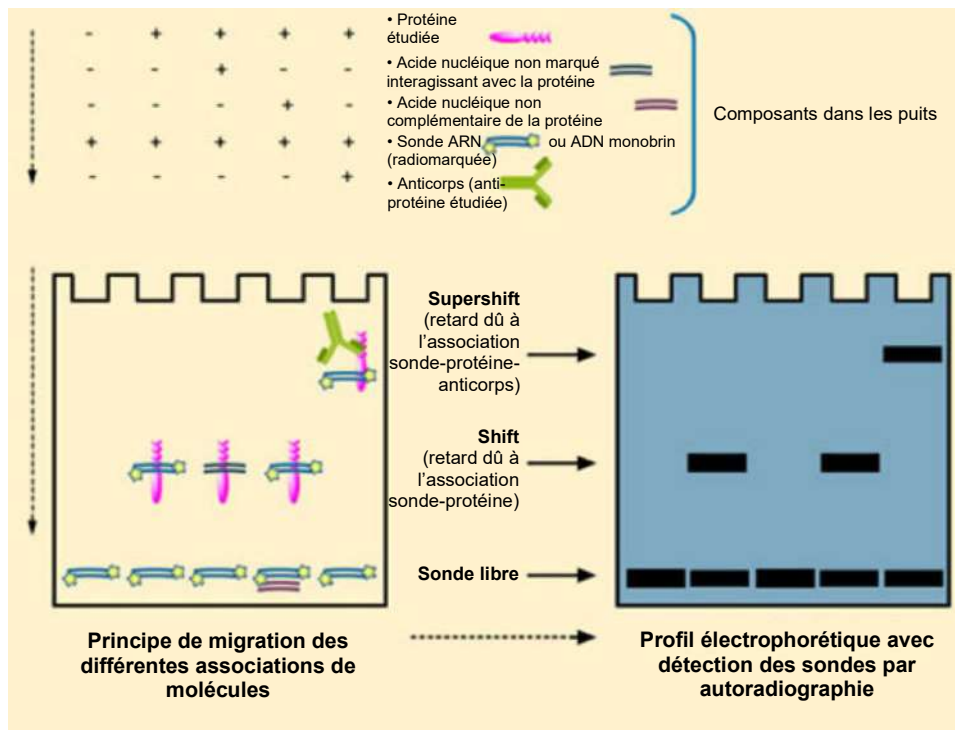
▲ FIGURE 45. **Fonctionnement d'un gène rapporteur : le gène de la bêta-gal.**
D'après GODINOT *et al.* (2010).

- Souvent produits par **knock-in**, des **organismes génétiquement modifiés** capables d'exprimer un **gène** que l'on place sous la **dépendance des mêmes séquences régulatrices (éléments de réponse)** – souvent des **séquences enhancer** – que le **gène** dont on veut **localiser l'expression**.

- Ainsi, la **séquence enhancer** sera activée pour les **deux gènes en même temps** :
 - le **gène rapporteur**
 - le **gène d'étude** dont on veut déterminer **quand et où il s'exprime**
- Le **gène rapporteur** est souvent un **gène** produisant un **composé coloré** (par fluorescence, activité enzymatique...).
Exemple : le **gène Lac Z** de la **bêta-galactosidase** permet, en présence de **X-galactose**, la formation d'un **composé bleu** (figure 45).

e. La mise en évidence d'une interaction entre une protéine et un acide nucléique par la technique du retard sur gel (EMSA)

- La technique du **retard sur gel** ou **EMSA** (*electrophoretic mobility shift assay*, **test de décalage / retard de mobilité électrophorétique**) (figure 46) est un **Blot conjoint de protéines et acides nucléiques** visant à **mettre en évidence l'interaction entre une protéine (ou un ensemble de protéines) et un acide nucléique (ADN ou ARN)**, révélé par sonde radioactive, au moyen de la mise en évidence d'un **retard de migration (shift)** en cas de formation du complexe protéine-acide nucléique.
- Des anticorps** contre la (ou les) **protéines étudiées** peuvent être ajoutés, induisant alors la formation d'un **supercomplexe anticorps-protéine-acide nucléique** dont le **retard de migration (supershift, « super-retard »)** est encore plus important que le **shift** du complexe protéine-acide nucléique.
- L'absence de retard sur gel indique que la **protéine et l'acide nucléique n'interagissent pas** alors que la **présence d'un retard de migration** permet de conclure à l'existence de l'interaction.



▲ FIGURE 46. Le principe de l'EMSA ou retard sur gel.

Fond de schéma : <http://www.cloud-clone.com/incrementServices/IS043.html> (consultation mars 2022).

II. Le génome des Bactéries

- On s'intéresse à caractériser ici le **génome des Bactéries** – au sens des **Eubactéries**, donc **excluant les 'archées'**.

Capacités exigibles

- ✓ **Comparer** l'organisation du génome des bactéries, des Eucaryotes et des virus.
- ✓ **Comparer** le génome cytoplasmique eucaryote et celui des bactéries.
- ✓ **Estimer** la proportion de séquences codantes et non codantes dans les génomes des Eucaryotes, des bactéries et des virus.

A. Un génome généralement circulaire et localisé dans le cytoplasme

- L'ensemble du **génome bactérien** a une **localisation cytoplasmique** et peut se présenter sous **deux formes** (figure 47). Le génome n'est donc **pas confiné** / **protégé dans un compartiment particulier**.



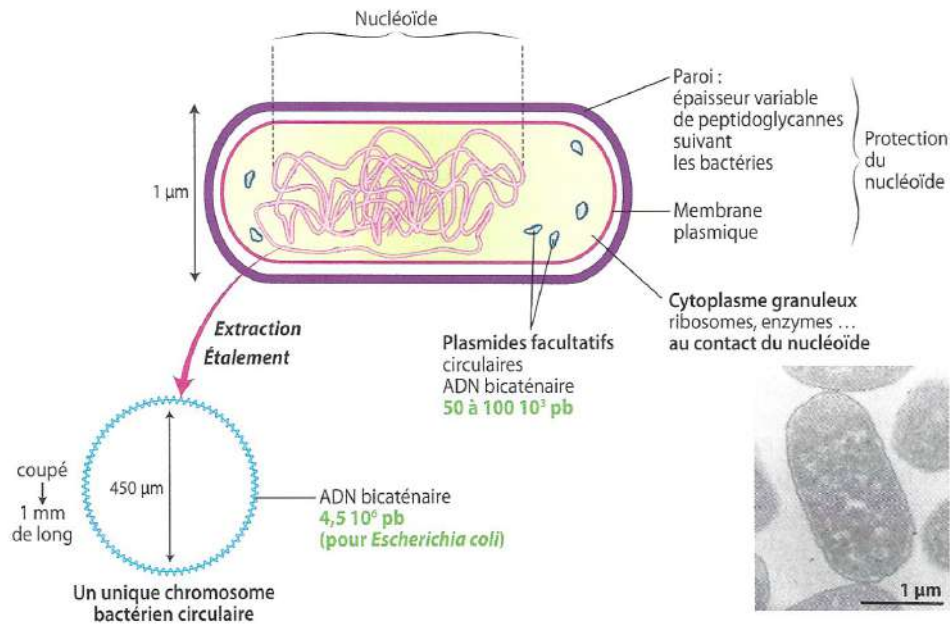
▼ TABLEAU V. Comparaison des génomes eucaryotes et eubactériens.

Inspiré de PEYCRU *et al.* (2013).

	Localisation	Organisation de l'ADN	Nombre de molécules d'ADN	Taille du génome	Séquences répétées non codantes	Structure des gènes
Eucaryotes	Noyau (= génome nucléaire)	Linéaire	Plusieurs différentes (état souvent dominant : diploïdie)	Élevée : généralement de 10^8 à 10^{11} pb Homme : $3,2 \times 10^9$ pb (cas de certaines 'amibes' : 10^{11} à 10^{12} !!)	Proportion souvent élevée Homme 50 % du génome (mais peut tomber à quelques % chez certaines espèces)	Introns + exons (gènes morcelés) Quelques espèces avec très peu d'introns (ex. levure <i>Saccharomyces cerevisiae</i>)
	Organites semi-autonomes (= génome extra-nucléaire)	Circulaire	1 en plusieurs copies Mitochondrie humaine : 1 à 15 copies, souvent 5	Petite : de 10^4 à 10^6 ADNmt humain : $16,6 \times 10^3$ pb	Proportion nulle ou faible	Pas d'introns Quelques gènes avec introns auto-épissables chez certaines espèces
Bactéries	Cytoplasme : chromosome bactérien (dans nucléoïde)	Souvent circulaire (linéaire chez quelques espèces)	1 (très rarement plusieurs)	Moyenne : de 10^6 à 10^7 pb <i>E. coli</i> : $4,64 \times 10^6$ pb (Mycoplasmes : 10^5 pb)	Proportion nulle	Pas d'introns
	Cytoplasme : Plasmides (génome extra-chromosomique)	Circulaire	Nombre variable	Très petite : 10^3 à 10^5 pb (en moyenne quelques milliers de pb... mais exceptions notoires !)	Proportion nulle	Pas d'introns

Bilan (adapté du programme)

- ✓ L'ensemble des **molécules d'ADN** contenues dans une **cellule** et l'**information** qu'elles portent constitue son **génome**.
- ✓ L'**étude des génomes** passe par une panoplie de **techniques** dites de **biologie moléculaire**.
- ✓ Des **techniques de séquençage** permettent de déterminer la **séquence** d'un **fragment d'ADN** puis de proche en proche la **séquence des génomes**.
- ✓ L'utilisation d'**outils bioinformatiques** permet d'identifier les **différents types de séquences codantes et non codantes**.



Un cliché de microscopie électronique d'*Escherichia coli* et le schéma d'interprétation d'une bactérie.

▲ FIGURE 47. **Le génome bactérien : vue d'ensemble.** D'après DAUTEL *et al.* (2021).

1. Un « chromosome bactérien » le plus souvent circulaire et unique situé dans une zone nommée nucléotide : le génome principal

a. Nature et taille du chromosome bactérien

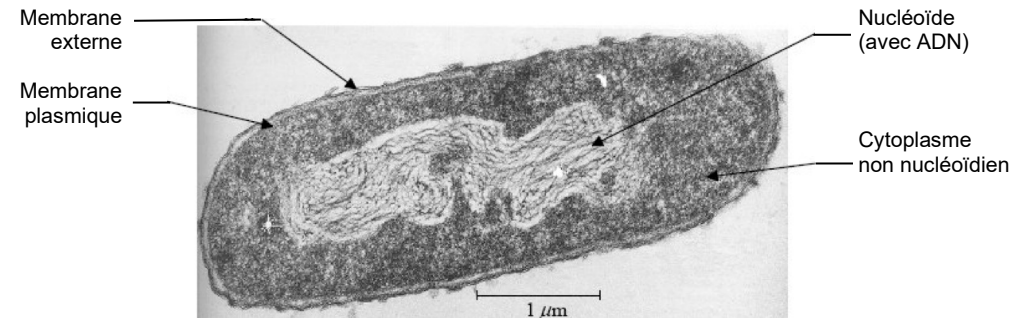
- Le **chromosome bactérien** (parfois appelé **génomophore**) est une **molécule d'ADN présente en un seul exemplaire (exceptionnellement plusieurs), sauf en cas de réplication où il y a évidemment doublement de l'ADN, de taille souvent inférieure à l'ADN eucaryote et le plus souvent circulaire** (mais on trouve des chromosomes bactériens linéaires chez quelques espèces bactériennes).
⇒ Le **génome** est donc globalement **non fragmenté**.
- Les **gènes portés par le chromosome bactérien** sont **indispensables au fonctionnement cellulaire**. On dit que le **chromosome bactérien** correspond au **génome principal** de la bactérie, en ce sens qu'**il comprend à la fois l'essentiel de l'information génétique et la portion indispensable au fonctionnement cellulaire de cette information génétique**.
- Le **chromosome bactérien** présente généralement :
 - Une **taille** comprise entre **1000 kb = 1 Mb à 5000 kb = 5 Mb**.
 - Un **nombre de gènes** autour de **1000** (parfois plusieurs centaines) à **6000 gènes**.

Escherichia coli :

- 4,64 Mb = 4 640 kb
- 4300 gènes.

b. Localisation fine (nucléotide) et organisation du chromosome bactérien (condensation modérée et protéines de structure)

- Ce **chromosome bactérien** est **situé dans une zone du cytosol** appelée **nucléotide** qui correspond, pour certains auteurs, au chromosome bactérien lui-même [c'est un peu pénible, d'ailleurs...]. Cette zone est **parfois visible** dans les clichés de MET et se reconnaît alors par son **aspect filamenteux** (lié à la **présence d'ADN**) et **relativement clair** comparativement au **reste du cytoplasme** (figure 48).



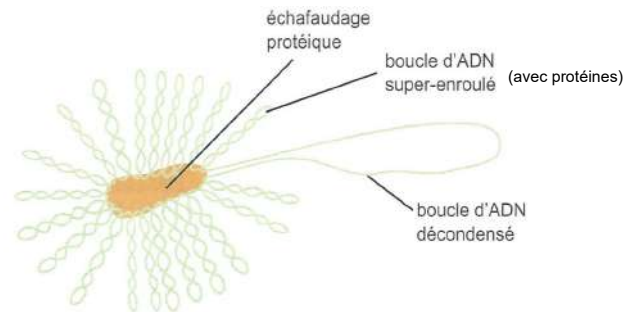
▲ FIGURE 48. **Escherichia coli (MET).**

<http://tblsbio.blogspot.fr/2009/09/prokaryotes-are-everywhere.html> (septembre 2015)

- Le **chromosome bactérien** est constitué d'**ADN « nu »** (absence de protéines de type histone*) **circulaire (rarement linéaire) et replié en une cinquantaine de boucles associées à un noyau protéique** (figures 49-50).

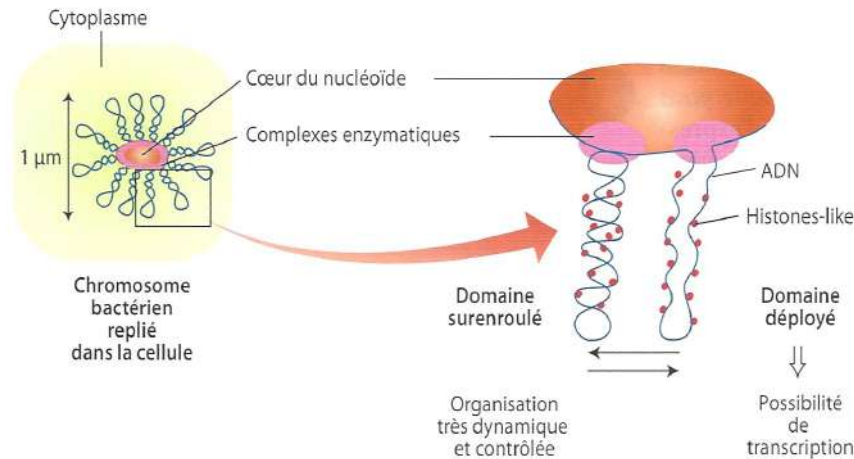
* Malgré l'absence d'histones, mais il existe des **protéines associées** au chromosome bactérien tout de même, notamment **celles qui interviennent dans le fonctionnement génomique** et des **protéines de structure** dont celles qu'on appelle **NAP (« histone-like »)**, Nucleoid Associated Proteins, agissant sur l'état de condensation du chromosome.

- La **condensation** est permise par un **surenroulement** de l'ADN permis par les **NAP** ; des **topoisomérases** gèrent également le **degré d'enroulement** et donc de **compaction** de l'ADN (figures 49-50) qui **conditionnerait**, comme chez les **Eucaryotes**, l'**accessibilité des gènes** à la transcription. La **condensation** reste néanmoins très **modérée** par rapport à celle de l'ADN eucaryote.
- Notons que les **protéines du nucléotide** présentent **beaucoup d'AA basiques** (ionisés positivement) qui font aisément des **liaisons avec les charges négatives** de l'ADN.



Organisation de l'ADN dans le chromosome bactérien
(D'après T. Brown : Genomes - second edition, Wiley-Liss, 2003).

▲ FIGURE 49. Organisation du chromosome bactérien. D'après PEYCRU *et al.* (2013).



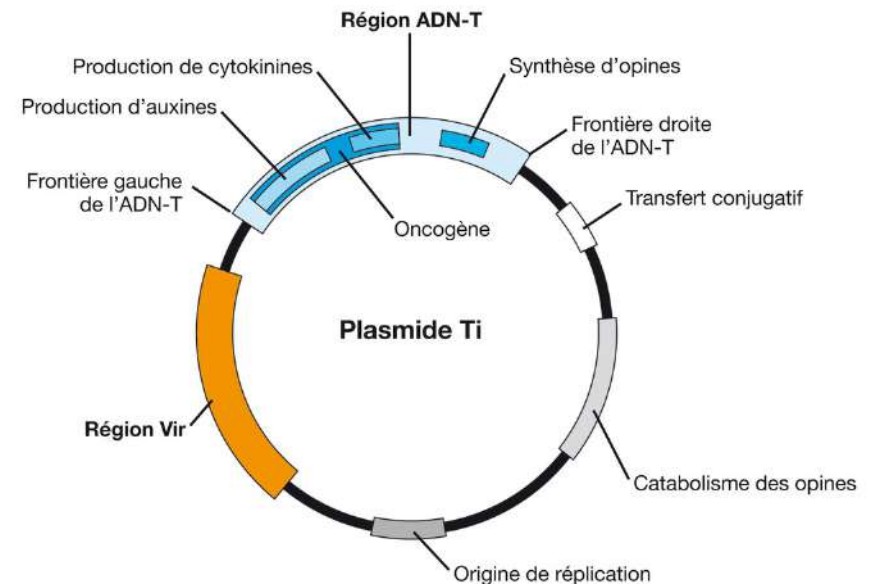
▲ FIGURE 50. Organisation du chromosome bactérien. D'après DAUTEL *et al.* (2021).

2. Des plasmides, petites unités génétiques fréquentes (mais non systématiques), circulaires, autonomes et transmissibles : le génome facultatif ou accessoire

- Les **plasmides** sont des **petits ADN bactériens circulaires, surnuméraires et capables de s'autorépliquer de façon autonome** (figure 51 + revoir figure 26 page 23).

Des plasmides peuvent s'intégrer au génome bactérien : ce sont les **épisomes**.

- Non indispensables, inexistantes chez certaines espèces, ils ne sont pas toujours présents – mais peuvent être présents en de nombreux exemplaires lorsqu'il y en a – et portent généralement des gènes apportant un avantage sélectif aux Bactéries détentrices dans des conditions données (exemple : résistance à un antibiotique, virulence...) :** c'est le **génome accessoire** ou **facultatif** de la Bactérie.
- Ils **s'échangent aisément entre Bactéries via des pili sexuels** par un mécanisme nommé **conjugaison**.
- Leur taille est en général **de quelques centaines à quelques dizaines de milliers de pb**, mais des **mégaplasmides** peuvent être bien **plus grands**, comme le **plasmide Ti des Agrobacterium** (100 kb à.... 2000 kb !).



▲ FIGURE 51. Le plasmide Ti d'une Agrobactérie (un gros plasmide !) et quelques gènes classiquement portés. D'après A. CAMPEAU-PÉLOQUIN, S. ROY & G. CHABOT (<https://monde.ccdmd.gc.ca/ressource/?id=121442&demande=desc>) (consultation mars 2022)

B. Un génome largement codant et des gènes souvent groupés en opérons polycistroniques

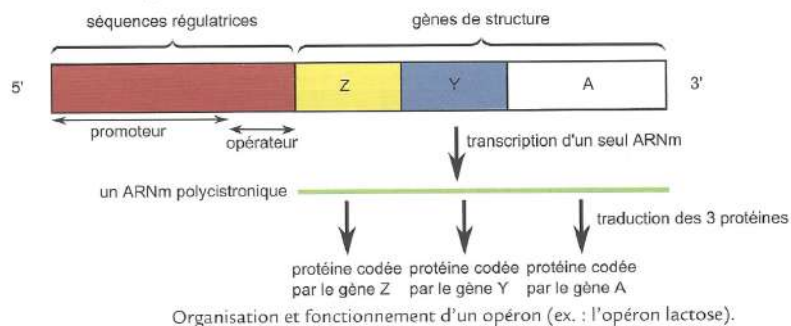
1. Peu de séquences répétées et non codantes : un génome largement codant et plutôt compact

- Contrairement aux **Eucaryotes** chez lesquels on trouve une **grande quantité d'ADN non codant** et beaucoup de **séquences répétitives**, chez les **Bactéries**, l'ADN est **majoritairement codant (près de 90 %)** et les **rares répétitions** sont surtout dues à des **séquences conservées** comme les **promoteurs de gènes**, les **gènes codant des ARNt...**
- Les **gènes** sont **séparés** par des quantités d'ADN espaceur très faibles.
- Le **génome bactérien** est donc assez **compact**.

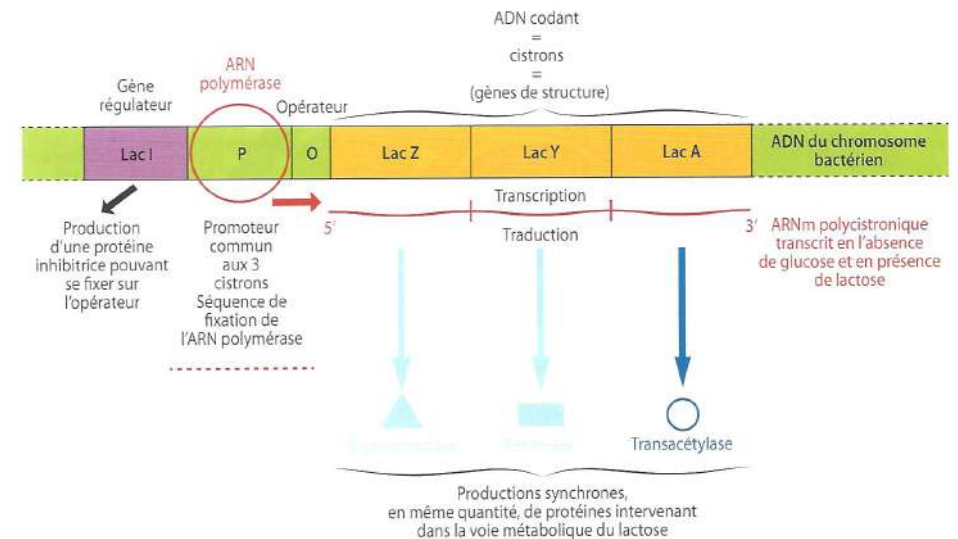
2. Des gènes non morcelés (= sans introns) et souvent organisés en opérons, unités de transcription polycistroniques dépendant des mêmes unités régulatrices

- Les **séquences codantes** sont souvent organisées en **opérons** (cas de près de **60 % des gènes chez E. coli**) : un **opéron** est un **ensemble de gènes bactériens disposés immédiatement à la suite les uns des autres et qui sont gouvernés par une même séquence régulatrice** (figures 52-54). Ces opérons sont donc des **ensembles polycistroniques**, c'est-à-dire **comportant plusieurs séquences codant des protéines (cistrons)**.

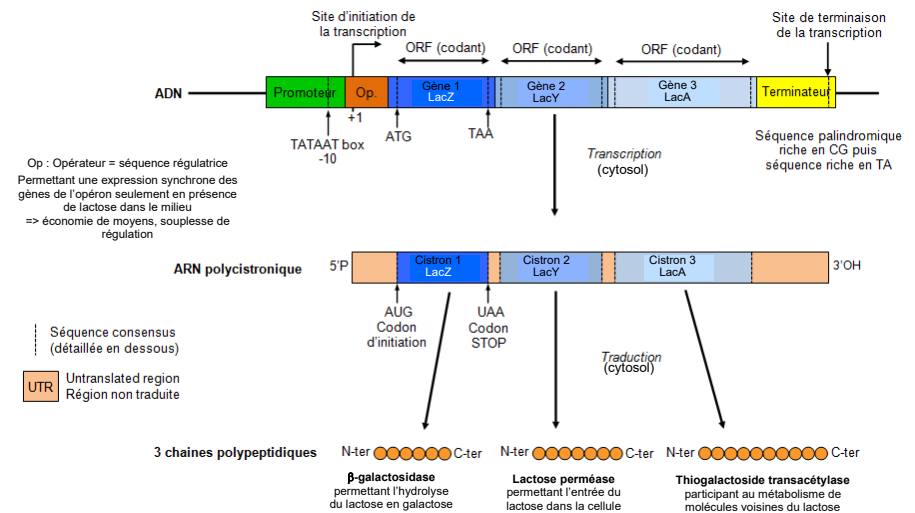
Chez *Escherichia coli*, 59 % des gènes sont regroupés au sein d'opérons - ensemble de gènes de structure mis bout à bout (figure 4) et placés sous le contrôle des mêmes séquences régulatrices (ex : promoteur, opérateur) ainsi que ces séquences - tandis que les 41 % restants sont dispersés. Ces gènes codent chacun une protéine qui, ensemble, assurent le plus souvent une même fonction (ex : 3 protéines participant au métabolisme du lactose pour l'opéron lactose). Lorsqu'il y a activation de l'expression de ces gènes de structure, un seul ARNm est produit mais il permet la synthèse des différentes protéines. C'est donc un **ARNm polycistronique** (poly = plusieurs, cistron = unité codant une chaîne polypeptidique). Les protéines sont toutes synthétisées en même temps ce qui garantit une **plasticité phénotypique rapide** à un changement de l'environnement pour l'eubactérie (ex : absence de glucose mais présence de lactose dans le milieu). Inversement, lorsqu'il y a inhibition de l'expression, c'est l'ensemble des protéines qui cessera ou presque d'être exprimé. Un opéron fonctionne donc sur un mode « on-off » qui permet une économie d'énergie en ne synthétisant les protéines que quand elle en a besoin. Par contre, cela ne permet pas une modulation indépendante de l'expression de chaque gène d'un opéron.



▲ FIGURE 52. **Les opérons, base du génome bactérien : l'exemple de l'opéron lactose.**
D'après PEYCRU *et al.* (2013).



▲ FIGURE 53. **Les opérons, base du génome bactérien : l'exemple de l'opéron lactose.**
D'après PEYCRU *et al.* (2013).



▲ FIGURE 54. **L'opéron lactose : des gènes aux protéines.**
Document S. FABRE (Lycée Chateaubriand, Rennes)
Voir le **chapitre sur l'expression génétique**.

III. Le génome des Eucaryotes

- On s'intéresse désormais au **génome des Eucaryotes**, qui se divise entre un **génome principal** située dans le **noyau** et un **génome extranucléaire** situé dans les **organites semi-autonomes**.

Capacités exigibles

- ✓ **Comparer** l'organisation du génome des bactéries, des Eucaryotes et des virus.
- ✓ **Comparer** le génome cytoplasmique eucaryote et celui des bactéries.
- ✓ **Estimer** la proportion de séquences codantes et non codantes dans les génomes des Eucaryotes, des bactéries et des virus.

A. Localisation, organisation et structuration du génome eucaryote

1. Vue d'ensemble : un génome protégé comprenant une composante principale nucléaire et une composante mineure cytoplasmique

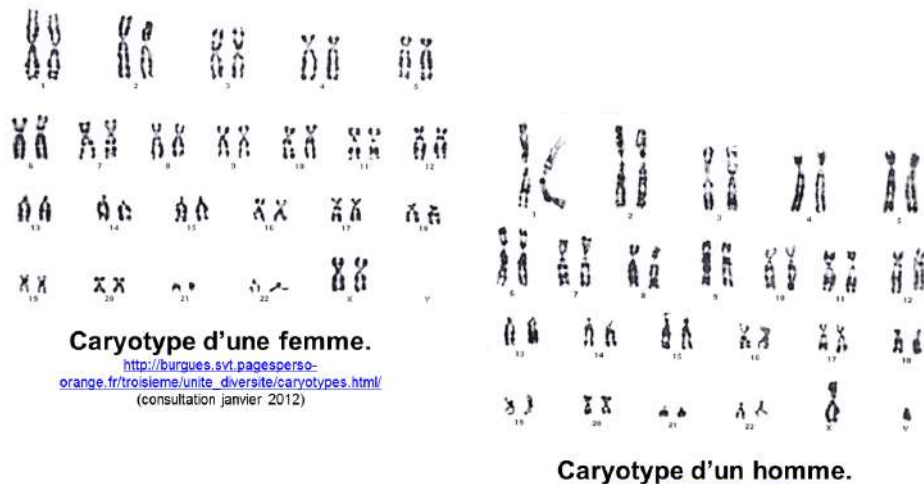
- Le **génome des Eucaryotes** est **protégé** dans des **organites** et comprend (revoir le **tableau V page 35**) :
 - Le **génome nucléaire** : des **chromosomes** au sens strict, **longs filaments d'ADN associés à des protéines présentant des niveaux variables de condensation**. La **taille totale** est **très variable** mais généralement nettement **plus importante** que celle du **génome bactérien** (un peu plus de **3 Gb** chez l'homme).

Chez les Eucaryotes, il existe généralement une **phase haploïde** du cycle de vie où **les chromosomes sont en un seul exemplaire** (ex. gamètes chez les Mammifères) et une **phase diploïde** où **les chromosomes sont en deux exemplaires** (ex. toutes les autres cellules chez les Mammifères). Les chromosomes sont alors représentés en paires de **chromosomes homologues** ou **autosomes** (**chromosomes de même taille, portant les mêmes gènes**) auxquels s'ajoute, chez de nombreux Animaux, une paire de **chromosomes sexuels** ou **gonosomes**. On appelle **caryotype** soit le **cliché des chromosomes d'une cellule ou d'un individu classés par paires** (**figure 55**), soit le **nombre de chromosomes d'une espèce, d'un individu ou d'une cellule** (= « **garniture chromosomique** », terme désuet à éviter).

- Le **génome extranucléaire (= cytoplasmique)** : il s'agit du **génome des organites semi-autonomes (plastides, mitochondries)** situé dans la **matrice** ou le **stroma**. Il s'agit de **petits chromosomes circulaires de type bactérien**, dans leur structure comme leur fonctionnement (ce qui confirme la **théorie de l'endosymbiose**), qui codent **une partie des polypeptides de l'organite**. Notons que **la majorité des polypeptides mitochondriaux ou plastidiaux sont codés par le génome nucléaire** (cas de **70 %** environ des **polypeptides mitochondriaux** chez l'Homme) : **de nombreux gènes ont été transférés de l'organite vers le noyau au cours de l'évolution**. Ces génomes sont d'ailleurs **très petits** (ex. **génome mitochondrial humain env. 16 kb**). On trouve souvent ce **génome** en de **multiples exemplaires**.

Bilan (adapté du programme)

- ✓ Chez les **bactéries**, le **génome à localisation cytoplasmique** est constitué d'un **chromosome circulaire** et éventuellement de **plasmides**.
- ✓ Le **génome des bactéries** est constitué **presque exclusivement** de **régions codantes**.
- ✓ Certaines sont **associées** à des **régions régulatrices communes**, ce qui forme des **opérons**.



Caryotype d'une femme.

http://burgues.svt.pagesperso-orange.fr/troisieme/unite_diversite/caryotypes.html/
(consultation janvier 2012)

Caryotype d'un homme.

http://burgues.svt.pagesperso-orange.fr/troisieme/unite_diversite/caryotypes.html/
(consultation janvier 2012)

A FIGURE 55. Caryotype humain.

[Caryotype humain : 23 paires de chromosomes : 22 paires d'autosomes, 1 paire de gonosomes]

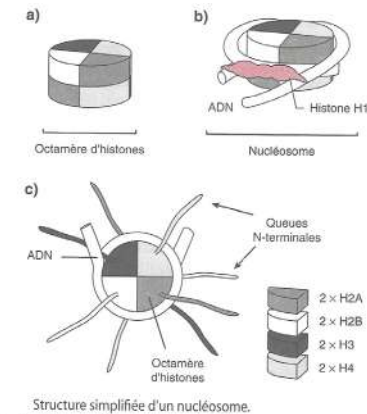
Notation du caryotype humain diploïde : $2n = 46$ [si c'était un gamète haploïde : $n = 23$]

2. Le génome nucléaire, une entité composée de chromosomes linéaires pouvant présenter des niveaux variables de condensation

a. Organisation fondamentale : des chromosomes comportant une ou deux chromatide(s) impliquant des protéines associées à l'ADN (dont les histones)

- Le **génome nucléaire eucaryote** est fondamentalement **fragmenté en de multiples chromosomes linéaires** (ex. 8 chez la Drosophile, 46 chez l'Homme).
- On appelle **chromatide** une **molécule d'ADN eucaryote associée à des protéines**.
- Un **chromosome** est constitué d'une **chromatide (chromosome simple** – en phase G1 du cycle cellulaire) ou bien de **deux chromatides (chromosome double = dupliqué** – en phase G2 du cycle cellulaire). Dans ce second cas, il s'agit de deux **chromatides identiques (chromatides sœurs)** (aux erreurs de réplication près).
- On y trouve **deux types de protéines associées** :
 - les **protéines de travail**, permettant le **fonctionnement et la dynamique du génome : la transcription, la réplication...**
 - les **protéines de structure**, assurant l'**organisation du génome**, dont les **histones (figure 56)**, protéines formant des octamères nommés **nucléosomes autour desquels s'enroule l'ADN eucaryote (ou archéen)**.

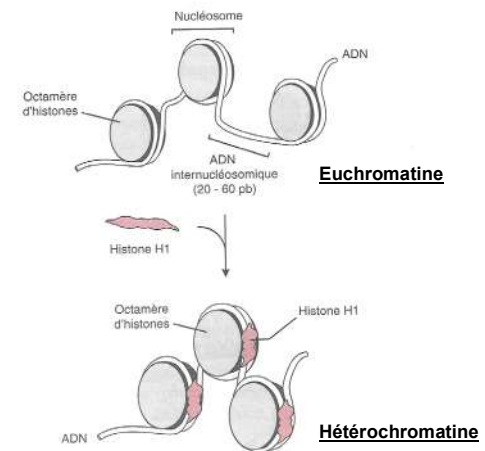
Ce sont toujours des **protéines fortement basiques (chargées positivement)**, ce qui facilite les liaisons à l'**ADN acide / chargé négativement**.



Structure simplifiée d'un nucléosome.
a) l'octamère d'histones ; b) l'enroulement de l'ADN autour de l'octamère et sa solidification par l'histone H1 ; c) protubérance vers l'extérieur des queues N-terminales d'histones de l'octamère.

Chacune des classes d'histones est constituée de petites protéines riches en amino-acides chargés positivement (lysine et arginine). Leurs séquences en amino-acides sont restées remarquablement identiques au cours de l'évolution. Les **histones H2A, H2B, H3 et H4** (11 à 15 kd) sont en quantité égale et chacune en double exemplaire, elles forment un **octamère** d'histones (Fig.) constituant un « cœur » protéique

discoïde de 8 protéines autour duquel l'**ADN nucléosomal** d'une taille de 147 paires de bases est enroulé (Fig.). L'histone **H1** (20 kd) ne fait pas partie du « cœur ». Elle est attachée d'une part, à l'ADN de liaison situé entre les nucléosomes et, d'autre part, à l'ADN nucléosomal ce qui a pour effet de solidifier l'ensemble (Fig.).

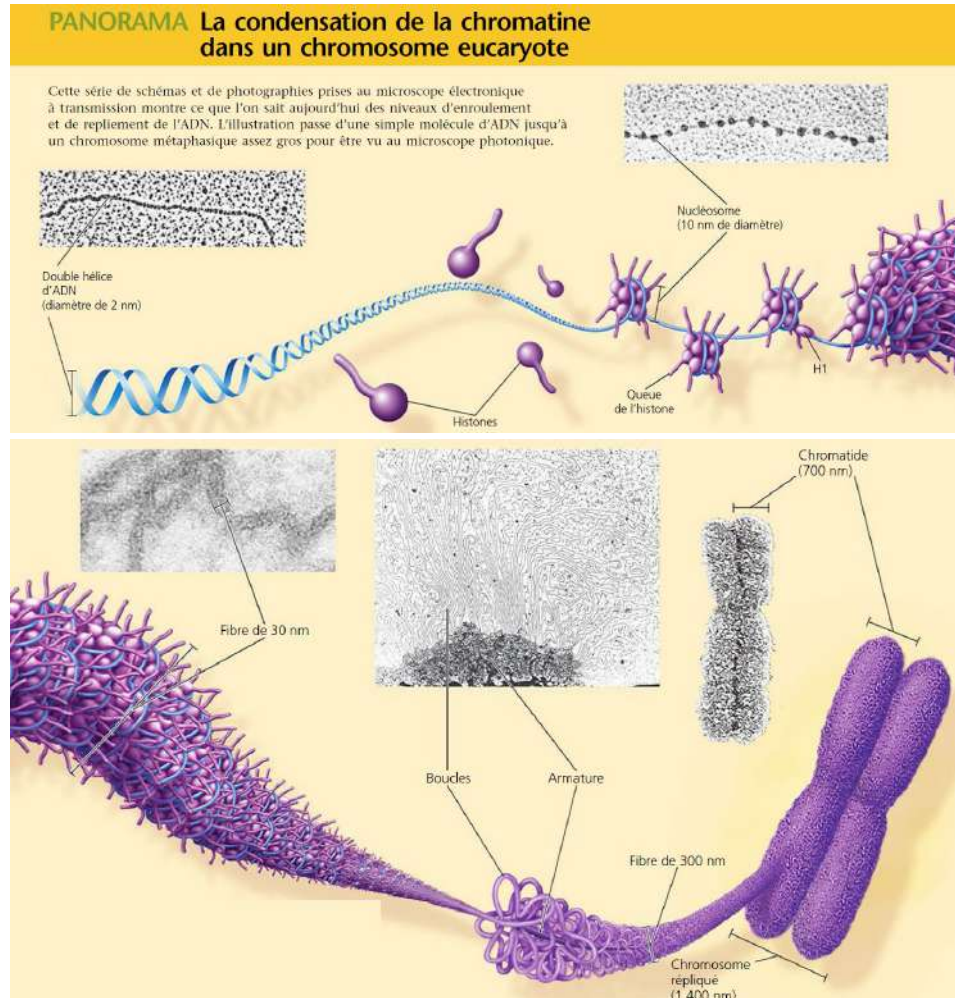


Représentation simplifiée d'un enchaînement de trois nucléosomes montrant le rôle de l'histone H1 dans le resserrement de l'ADN autour de l'octamère d'histones.

A FIGURE 56. Histones et nucléosomes. D'après PETIT & JULIEN (2007).

- On appelle **nucléosome** l'assemblage régulier d'histones et d'ADN (figure 56). Le **premier niveau de condensation** ainsi formé s'appelle **fibre nucléosomique** (ou **nucléofilament**, ou encore « collier de perles ») (figure 56).
Vous devez savoir **représenter simplement un nucléosome** et **expliquer sa constitution**.

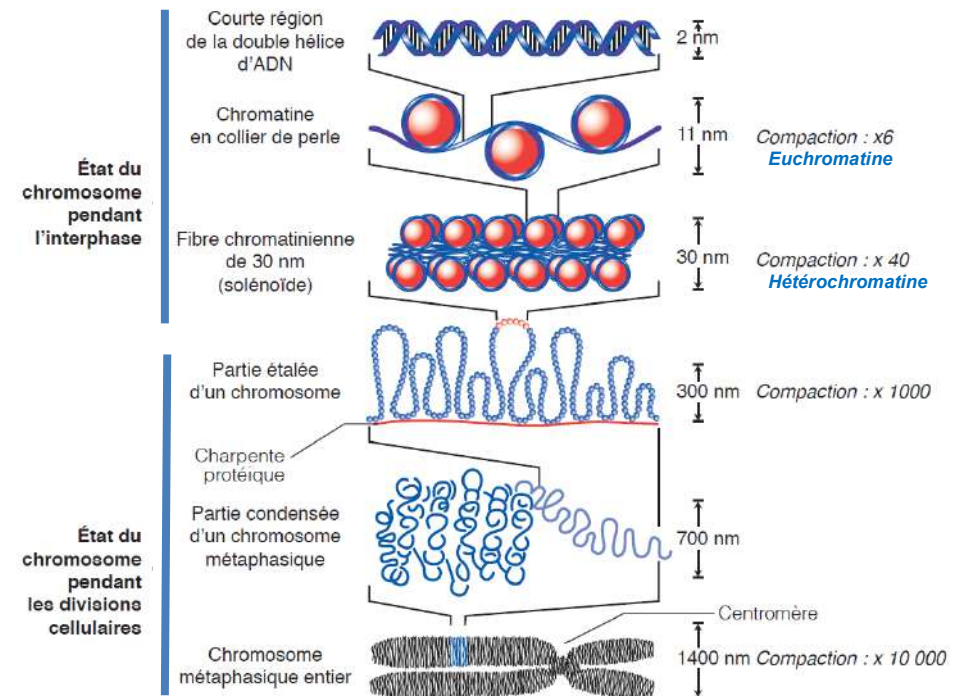
b. Degrés de condensation de l'ADN



▲ FIGURE 57. **Niveaux de condensation de l'ADN chez les Eucaryotes.**

D'après CAMPBELL et al. (2012).

[Attention à savoir reconnaître ces niveaux de compaction sur les **électronographies** + maîtriser les **échelles** !]



▲ FIGURE 58. **Niveaux de condensation de l'ADN eucaryote (vision simple).**

D'après SEGARRA et al. (2014), adapté.

- L'ADN eucaryote peut se compacter à des niveaux divers imbriqués selon le moment du cycle cellulaire.
- On peut citer les niveaux de compaction suivants de l'ADN (figures 57-58) :

Des discussions existent sur l'organisation fine de ces structures.

- Fibre nucléosomique** (déjà abordée ci-dessus) (facteur de compaction : 6) [correspond à l'euchromatine].
- Fibre épaisse = chromatinienne**, dite en solénoïde, ou **fibre de 30 nm constituée de l'enroulement d'environ 6 nucléosomes en hélice** ; cette structure est stabilisée par les histones H1 (facteur de compaction : 40) [correspond à l'hétérochromatine].

Au-delà, les niveaux de compaction suivants ne sont atteints que lors de la condensation de l'ADN au moment des divisions cellulaires. C'est d'ailleurs seulement lors des divisions que les chromosomes sont visibles et individualisés.

- Fibre compacte de 300 nm** formée par la réalisation de boucles qui s'associent sur un cœur protéique non composé d'histones.
- Cette fibre s'enroule ensuite en domaines d'ordre de grandeur 700 nm que l'on retrouve au sein des chromosomes condensés (largeur environ 1400 nm), le chromosome métaphasique étant le point culminant de la condensation.

c. L'ADN en interphase

α. Notions de chromatine (euchromatine vs. hétérochromatine) et de nucléole(s)

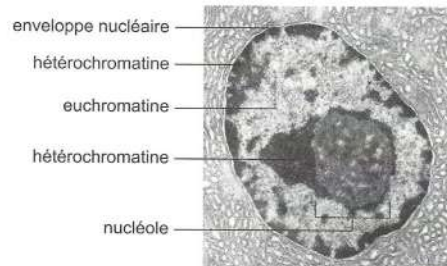
- Dans un **noyau en interphase** (figure 59), on trouve de la **chromatine** (essentiel de l'ADN et protéines associées) et un ou plusieurs **nucléoles** (où sont synthétisés les ARNr et où sont assemblées les sous-unités ribosomiques).
Attention : ne pas confondre chromatine et chromatide !
- On peut ensuite distinguer au sein de la chromatine, en fonction du **niveau de densité aux électrons au MET**, l'**euchromatine** et l'**hétérochromatine** ; **seule la première est accessible aux ARN polymérases qui réalisent la transcription**.

La chromatine baigne dans le nucléoplasme.

La microscopie électronique à transmission permet de reconnaître deux catégories bien différentes :

- l'**hétérochromatine** (ou chromatine dense, très opaque aux électrons) localisée essentiellement en périphérie contre la face interne de l'enveloppe nucléaire et autour du ou des nucléoles.
- l'**euchromatine** dispersée dans le nucléoplasme.

Sur le plan fonctionnel, l'hétérochromatine est inaccessible aux ARN pol, à la différence de l'euchromatine ; c'est pourquoi l'hétérochromatine est qualifiée d'inactive contrairement à l'euchromatine.

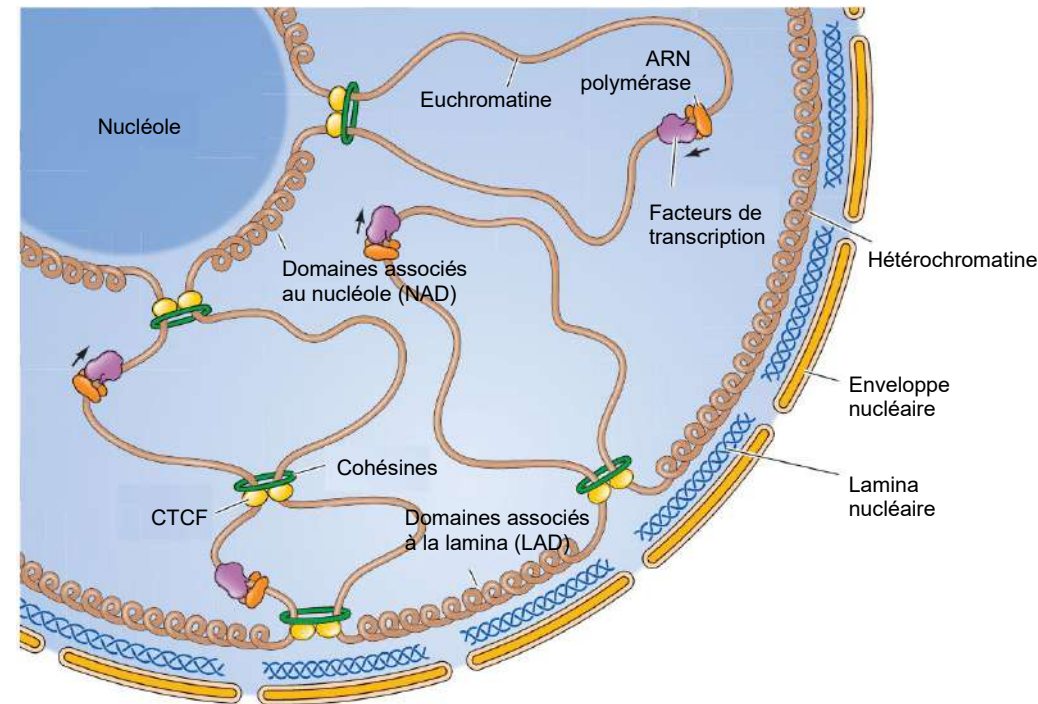


Noyau cellulaire et chromatine observés au M.E.T.
(Cliché J. André labo, BC4, Orsay, « Atlas de Biologie cellulaire », J.-C. Callen, J.-C. Rolland, A. et D. Szöllösi, 5e éd. Dunod, 2001).»

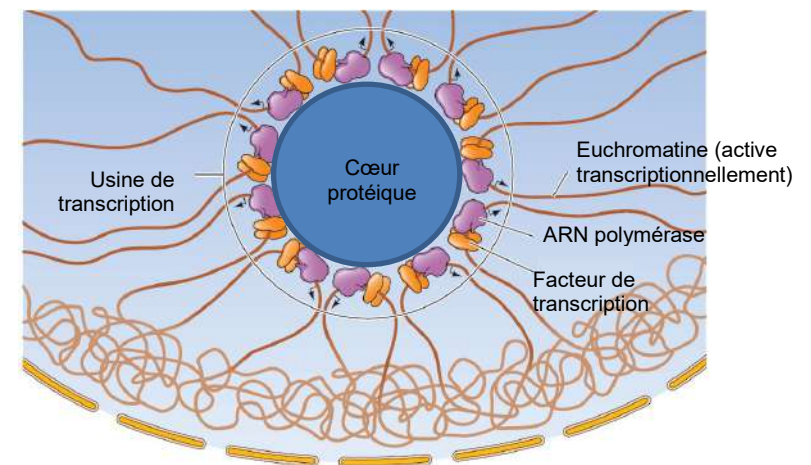
▲ FIGURE 59. **Allure du noyau au MET (taille env. 4 μm).** D'après PEYCRU et al. (2013).

β. Une association de l'hétérochromatine avec la lamina nucléaire (LAD) ou le nucléole (NAD), et un regroupement fonctionnel des zones d'expression dans l'euchromatine (usines à transcription)

- Les **boucles de l'hétérochromatine** sont **transcriptionnellement inactives** et **associées de manière non covalente, soit à la lamina nucléaire (LAD : lamina-associated domains), soit au nucléole (NAD : nucleolus-associated domains)** (figure 60).
- Les **domaines euchromatiniens** sont **transcriptionnellement actifs** et sont **organisés en unités décondensées** où les **gènes transcrits** sont **regroupés** dans des **zones riches en enzymes, protéines et autres molécules** (ex. NTP) nécessaires à la transcription (figure 61). On appelle **usines de transcription (transcription factories)** ces **regroupements d'ARN polymérases et de facteurs de transcription** dans lesquels **plusieurs domaines de chromatine active sont transcrits simultanément et qui seraient fixés sur un cœur protéique**.
- Les **limites de domaines**, c'est-à-dire les **zones de transition** entre les **domaines euchromatiniens** et **hétérochromatiniens**, sont structurées par le **CTCF** (contraction de **CCCTCT-binding Factor**), un **facteur spécifique de transcription qui intervient aussi en l'occurrence dans la limitation des domaines chromatinien**, et des **protéines de condensation particulières** nommées **cohésines** (figure 60).

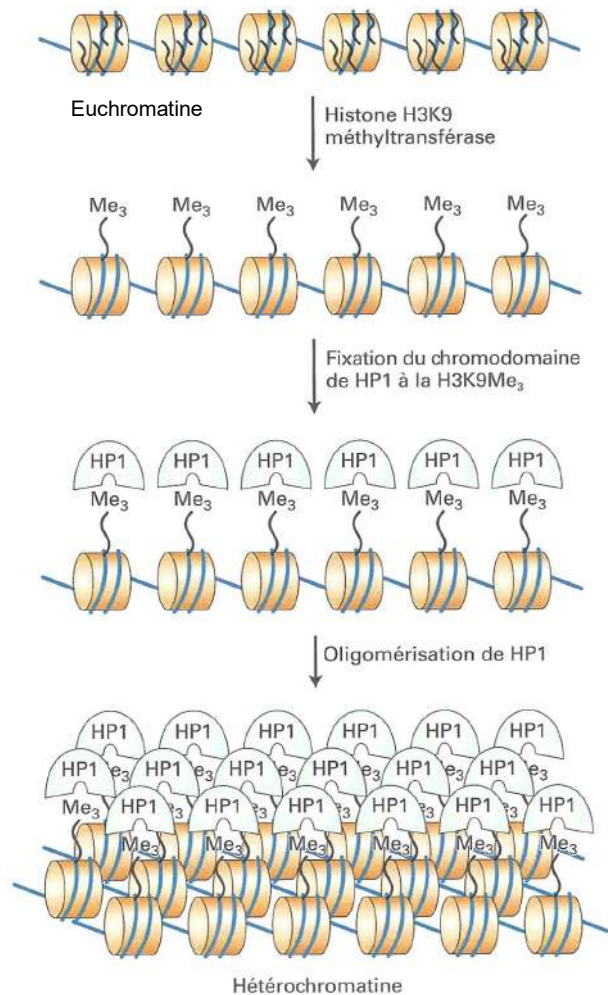


▲ FIGURE 60. **Distribution de la chromatine active et inactive [pour information ?].**
D'après COOPER (2019), adapté / traduit.



▲ FIGURE 61. **Les usines de transcription (transcription factories) [pour information ?].**
D'après COOPER (2019), adapté / traduit, modifié.

γ. L'importance de l'histone H1 et du complexe de remodelage de la chromatine dans la transition euchromatine / hétérochromatine



▲ FIGURE 62. Un modèle de transition eu-/hétérochromatine médié par la méthylation des histones H3 et le recrutement des histones H1 [pour information ?].

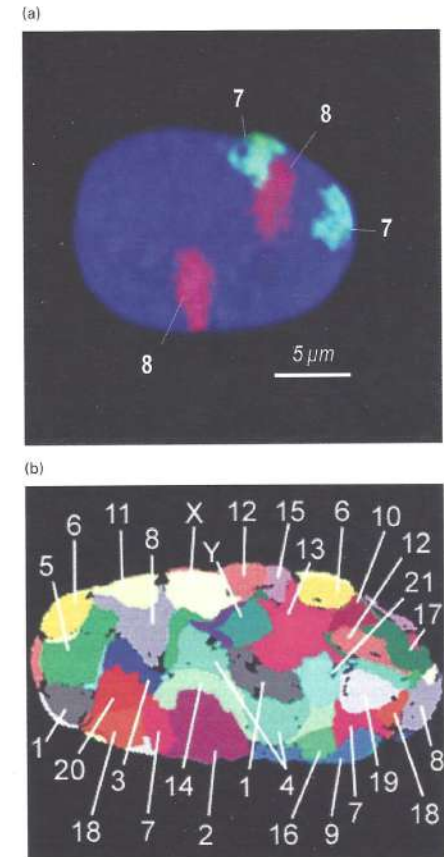
D'après LODISH *et al.* (2014).

- Des **complexes enzymatiques** permettent de **modifier les histones** et ces **modifications** (par exemple : **méthylation de l'extrémité N-terminale** de l'histone H3, **figure 62**) peuvent entraîner le **recrutement d'histones H1** à l'origine de la condensation de l'euchromatine en hétérochromatine.

On précisera ces aspects dans le chapitre sur le contrôle de l'expression génétique.

δ. Une répartition des chromosomes décondensés non aléatoire et localisée : notion de territoire chromosomique

- Un **territoire chromosomique** est un **volume nucléaire occupé préférentiellement par un chromosome décondensé en interphase** (**figure 63**). La **taille de ces territoires** et leur **localisation** peuvent **légèrement varier** en fonction de l'état de la chromatine et de l'activité génétique. Ces territoires ont été mis en évidence par **hybridation *in situ* de type FISH**.



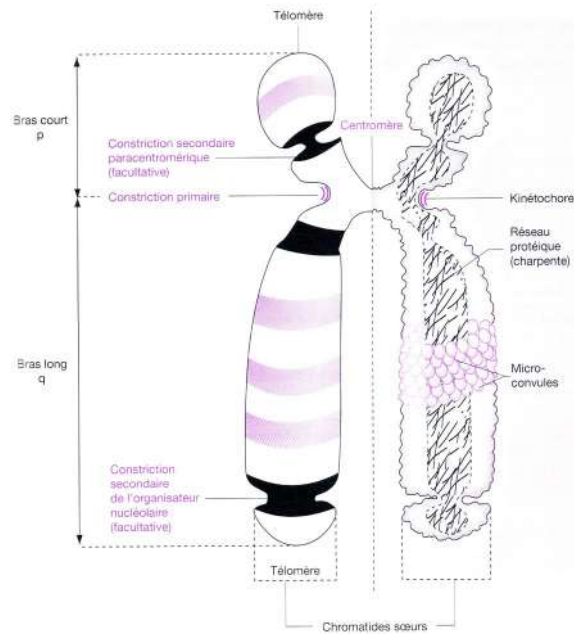
EXPÉRIENCE Pendant l'interphase, les chromosomes humains restent dans des territoires non chevauchants du noyau. Des fibroblastes humains fixés en interphase ont subi une hybridation *in situ* avec des sondes marquées par fluorescence, spécifiques de séquences situées sur toute la longueur des chromosomes humains 7 (cyan) et 8 (rose). L'ADN est marqué en bleu par du DAPI. Dans la cellule diploïde, chacun des deux chromosomes 7 et des deux chromosomes 8 est limité à un territoire ou domaine dans le noyau, au lieu de s'étendre dans la totalité du noyau. (b) Même chose qu'en (a) à l'exception du fait que les sondes de coloration chromosomique spécifiques de chaque chromosome ont été hybridées pour révéler la position de presque tous les chromosomes dans un fibroblaste d'homme. Certains de ces chromosomes sont invisibles dans cette coupe optique qui passe à travers le noyau. [Partie (a) aimablement communiquée par Drs. I. Solovoi & T. Cremer. Partie (b) d'après A. Bolzer *et al.*, 2005, *PLOS Biol* 3:826.]

▲ FIGURE 63. Les territoires chromosomiques. D'après LODISH *et al.* (2014).

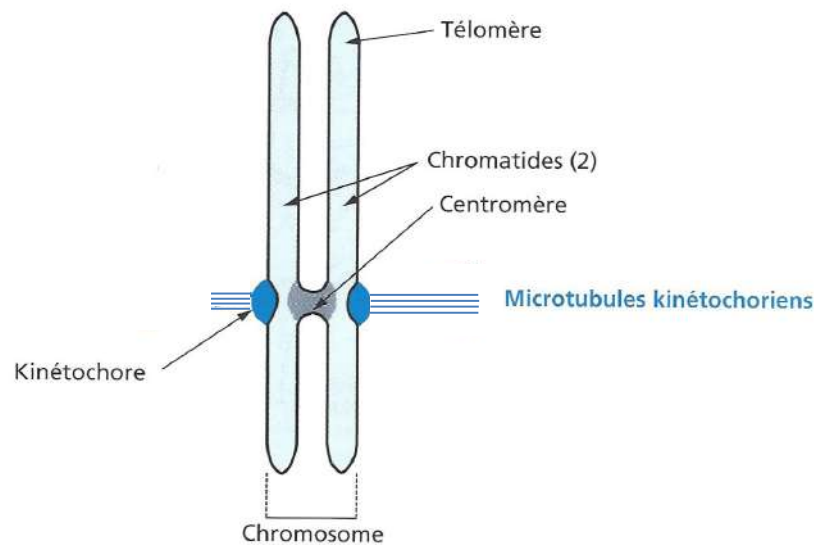
d. Le chromosome mitotique (cas du chromosome métaphasique)

- Dans un **chromosome condensé** (**figures 64-65**), nécessairement **double** avant l'anaphase, on constate que **les deux chromatides sont associées par un centromère**. On appelle **télomères** les **extrémités chromosomiques**. Le **kinétochore** est un **complexe protéique où viennent s'attacher les fuseaux de division** en prophase.

Notez que télomères et centromères existent sur les chromosomes décondensés mais ne se distinguent alors pas en microscopie (sauf usage de FISH).



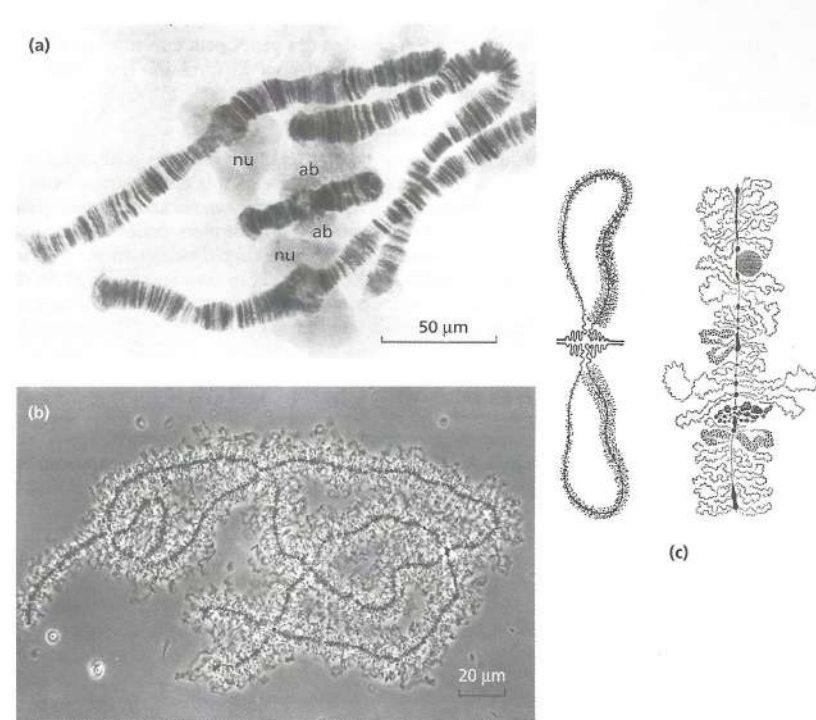
▲ FIGURE 64. Le chromosome métaphasique. D'après BREUIL (2007).



▲ FIGURE 65. Organisation d'un chromosome double condensé. D'après PEYCRU *et al.* (2010a), légèrement modifié.

Encadré F Des chromosomes particuliers

(Au-delà du programme : pour information)



Des chromosomes particuliers.

(a) **Chromosomes polyténiques**, glandes salivaires d'une larve d'insecte (Chironome) x 900, (Cliché W. Beermann). Dans les noyaux des cellules des glandes salivaires, plusieurs cycles de réplication de l'ADN se succèdent sans séparation des chromosomes fils. Les chromosomes polyténiques ou chromosomes géants qui en résultent rassemblent des faisceaux de chromosomes interphasiques. Les bandes sombres sont des régions riches en ADN condensé et les inter-bandes claires ne contiennent que 15 % de l'ADN du chromosome. Les puffs ou anneaux de Balbiani (ab) sont des zones où l'ADN est décondensé ; à ce niveau, l'incorporation d'uracile tritié ou d'uridine tritiée démontre l'existence d'une transcription intense. Les bandes sombres sont donc des zones où sont localisés des gènes ; les puffs sont des zones où les gènes sont fonctionnels. Au niveau de certains puffs sont synthétisés les ARN ribosomiques ; ces puffs correspondent donc aux nucléoles (nu).

(b) **Chromosomes en écouvillon**, ovocyte de triton, contraste de phase x 900, (Cliché J. Gall). Une paire de chromosomes homologues a été isolée d'un noyau d'ovocyte en prophase de méiose. A ce stade les chromosomes présentent des expansions latérales en forme de boucles ; à ce niveau, l'incorporation d'uracile tritié ou d'uridine tritiée démontre l'existence d'une transcription intense.

(c) Schémas d'interprétation (chromosomes en écouvillon).

FIGURE a. Chromosomes polytènes (= polyténiques = géants) et chromosomes en écouvillon. D'après PEYCRU *et al.* (2010a).

Les chromosomes géants (chromosomes polyténiques)

➤ Ce sont des chromosomes **observables** dans les **cellules larvaires de glandes salivaires d'Insectes Diptères** (figure a.a). Il s'agit de **chromosomes constitués du regroupement de très nombreux chromosomes interphasiques ayant subi de multiples réplifications en restant solidaires** (un chromosome polyténique peut compter jusqu'à 512, voire 1024 copies !). On y distingue, en microscopie électronique, des **bandes sombres** hautement condensées et des **interbandes** moins condensées. Les **puffs** ou **anneaux de BALBIANI** (ab sur la figure a) sont des **zones décondensées d'intense transcription** qui ont permis de **démontrer que l'état décondensé était indispensable à la transcription des gènes** ; certains puffs constituent le **nucléole**. L'état polyténique permet une **synthèse protéique très intense**.

Les chromosomes en écouvillon

➤ Ce sont des chromosomes **observables** dans les **ovocytes d'Amphibiens en prophase I de méiose (stade diplotène)** montrant également le **lien entre décondensation et transcription** (figure a.b-c). Ces **ovocytes** accumulent une grande quantité de **réserves** utiles à la **gamétogenèse** et au **développement embryonnaire**. Les chromosomes en écouvillon **présentent des boucles disposées en vis-à-vis où l'ADN est décondensé** ; cette organisation lui donne un **aspect plumeux**, d'où l'expression écouvillon.

- La **morphologie** des chromosomes peut varier : beaucoup d'auteurs présentent les **chromosomes géants** ou les **chromosomes en écouvillon**, souvent abordés en classe préparatoires. Notez qu'ils ne sont pas explicitement au programme : les données de l'**encadré F** sont donc proposées seulement à titre d'information... mais on ne sait jamais avec ce concours ! ☺

e. Une taille variable du génome nucléaire (en moyenne grossièrement de l'ordre du Gb)

- Le **génome nucléaire** peut présenter une **taille variable** (tableau VI), ce qui vaut aussi pour le **nombre de gènes**.

Saccharomyces cerevisiae (levure Ascomycète) :

- 13 Mb = 13 000 kb
- 6200 gènes.

Homo sapiens (Homme) :

- 3,2 Gb = 3 200 Mb = 3 200 000 kb
- 20 000 / 25 000 gènes.

Polychaos dubium (amibe) :

- 670 Gb = 670 000 Mb = 670 000 000 kb
- ?? gènes.

▼ TABLEAU VI. **Taille de quelques génomes eucaryotes.** D'après PERRIER, BEAUX et al. (2021).

Les données, surtout relatives au nombre de gènes, peuvent varier notablement d'un auteur à l'autre.

Organisme	Type	Nombre de chromosomes par jeu haploïde et nombre de lots haploïdes	Taille du génome (Mb)	Nombre de gènes
<i>Saccharomyces cerevisiae</i>	Ascomycète	16	13	6 200
<i>Arabidopsis thaliana</i>	Angiosperme (Brassicacées)	5 × 2	120	25 500
Riz	Angiosperme (Poacée)	12 × 2	430	Environ 40 000
Blé	Angiosperme (Poacée)	7 × 6	17 000	
<i>Drosophila</i>	Insecte	4 × 2	180	13 400
<i>Mus musculus</i>	Mammifère	20 × 2	3 000	Environ 20 000
<i>Homo sapiens</i>	Mammifère	23 × 2	3 400	21 000

3. Le génome extranucléaire, un ensemble de chromosomes de type bactérien codant très partiellement les polypeptides des organites semi-autonomes

- Qu'il s'agisse des **plastides** ou des **mitochondries**, le **génome des organites semi-autonomes** (tableau VII) présente les caractéristiques suivantes :
 - Il est **protégé** dans les **organites semi-autonomes** (dans le **stroma plastidial** ou la **matrice mitochondriale**).
- Contrairement à celui des **Bactéries**, le **génome eucaryote** – même **extranucléaire** – n'est donc **jamais cytosolique**.
- Il est de **taille plutôt faible**, de l'ordre de 10^4 à 10^6 pb (ex. ADN mitochondrial humain : 16,6 kb / 37 gènes dont 13 polypeptidiques ; ADN plastidial *A. thaliana* : 155 kb / 87 gènes protéiques).
- Il est de **type procaryote** (ADN circulaire, gènes sans introns, parfois en **opérons**, traduction **cotranscriptionnelle**...), même si certains peuvent présenter, notamment dans les plastides, des **introns auto-épissables**.
- Il est **exprimé** dans l'**organite semi-autonome** lui-même.
- Il code au maximum **30 % des polypeptides locaux** (ADNmt humain : 5-10 %).
- Les autres **polypeptides** sont codés par le **génome nucléaire** et **importés** dans l'**organite** après **synthèse cytosolique** ; les **gènes nucléaires** en question sont **issus de transferts horizontaux** de gènes depuis le **génome de l'organite** vers le **génome nucléaire** au cours de l'évolution.
- Il n'y a **pas de séquences répétées**.
- Il est représenté par de **multiples copies** (du même ADN).

▼ TABLEAU VII. **Caractéristiques des génomes mitochondrial et chloroplastique.**

D'après PERRIER, BEAUX *et al.* (2021).

Dans ce domaine aussi, les **données**, surtout relatives au **nombre de gènes**, peuvent **varier** notablement d'un auteur à l'autre.

Caractères du génome mitochondrial.

Taille	- <i>Saccharomyces cerevisiae</i> : 75 kb (35 gènes). - <i>Arabidopsis thaliana</i> : 367 kb (284 gènes). - Homme : 16,569 kb (37 gènes dont 13 codant des protéines de la phosphorylation oxydative ; 22 codant des ARNt et 2 des ARNr mitochondriaux).
Caractère du génome et information	- Le génome possède toute l'information nécessaire à la traduction des protéines (ARNm, ARNt, ARNr). - Le nombre d'ARNt (20) requis pour la synthèse de protéines mitochondriales est inférieur au nombre d'ARNt (32) requis pour traduire les ARN produits dans le noyau. Cette économie est permise par un flottement plus net de l'appariement des anticodons des ARNt, expliquant les différences avec le code génétique nucléaire. - Chez l'homme, tout l'ADN mitochondrial est codant mais chez certains végétaux et mycètes, il possède des séquences non codantes (introns).
Relation avec le génome nucléaire	- Le génome de petite taille ne permet le codage que de 5 à 10 % des protéines de la mitochondrie (protéines de la chaîne respiratoire). - Les autres protéines du système de phosphorylation oxydative sont codées par des gènes nucléaires. L'ARNm est traduit hors de la mitochondrie par les ribosomes cytosoliques. - Les protéines synthétisées sont transportées dans la mitochondrie et le système est assemblé dans la membrane mitochondriale interne.

Caractères du génome chloroplastique.

Taille	- De 15 à 500 kb (155 kb pour <i>Arabidopsis thaliana</i>), avec de nombreuses copies. - 20 à 35 copies par chloroplaste pour <i>A. thaliana</i>, soit pour l'ensemble d'une cellule de 1 000 à 1700 copies
Caractère du génome et information	- Chez <i>Arabidopsis</i>, le nombre de protéines codées est de 87, dont la moitié est impliquée dans les processus photosynthétiques. - Les gènes portés codent pour des protéines et des ARN (ARNr et ARNt). Une partie des gènes chloroplastiques comporte des introns, produisant alors des ARNm épissés avant traduction.
Relation avec le génome nucléaire	- Une grande partie des protéines est codée par le génome nucléaire et est importée depuis le cytosol. - La Rubisco est constituée de sous-unités codées les unes par le génome nucléaire, les autres par le génome chloroplastique.

B. Des gènes nucléaires morcelés et monocistroniques

1. Notions de gène, cistron, locus, allèle, hétérozygotie, homozygotie, dominance et récessivité

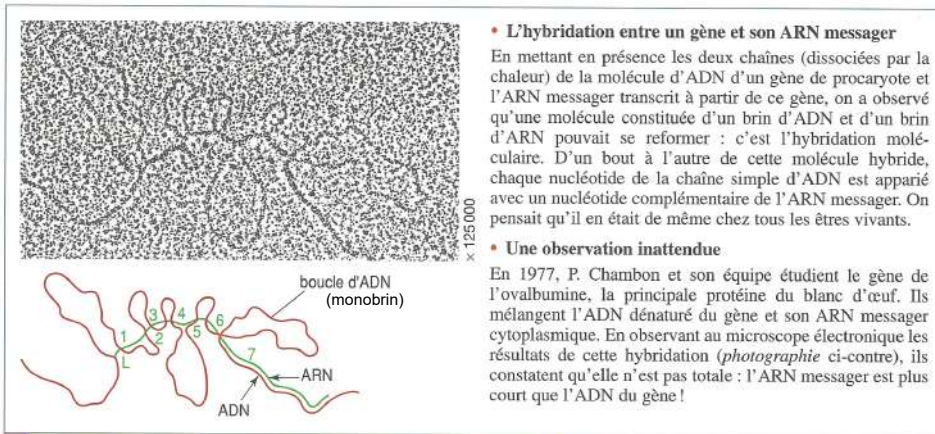
- Un **gène** est fondamentalement un **fragment d'ADN qui code un ARN (c'est donc une unité de transcription), souvent celui d'une protéine** (c'est donc une sorte de « plan de montage » de protéine). La **partie effectivement codante d'un gène** (la séquence **codant rigoureusement la protéine**) peut être appelée un **cistron** ou **séquence codante** (*coding sequence* CDS).
- Un gène occupe une **position précise sur son chromosome** qu'on peut nommer **locus**.
- Les **différentes versions d'un gène** (c'est-à-dire les **séquences nucléotidiques possibles de ce gène**) constituent des **allèles**.

Chez une espèce **diploïde** qui possède **deux fois le même gène** (un sur chaque chromosome homologue), on dit que l'individu est **homozygote s'il possède deux fois le même allèle** et **hétérozygote s'il possède deux allèles différents**. Dans le cas d'un individu hétérozygote pour un gène donné, on appelle **allèle dominant celui qui s'exprime** et **allèle récessif celui qui ne s'exprime pas**. Il arrive que les deux allèles soient **coexprimés**, on les dira alors **codominants**.

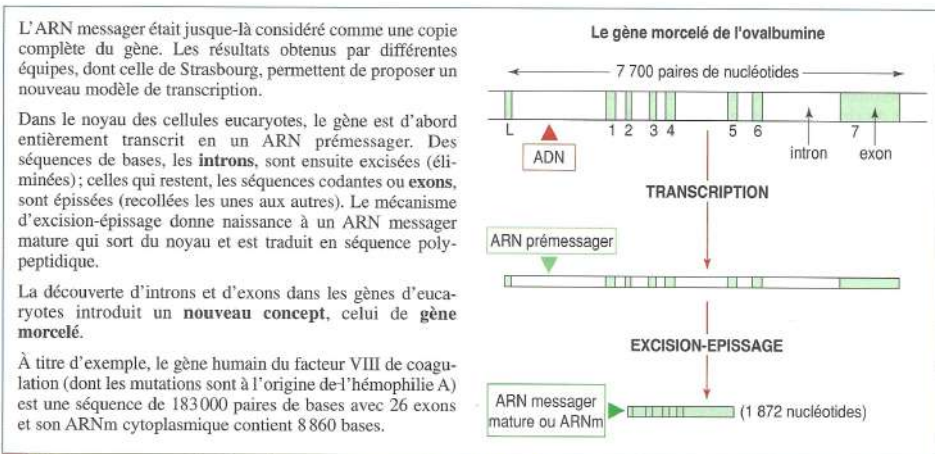
2. Des gènes morcelés (= gènes mosaïques) chez les Eucaryotes avec des régions non codantes (introns) séparant les portions codantes ou non traduites (exons)

a. Mise en évidence des introns par hybridation ARN-ADN monobrin (1977)

- En mettant en relation un fragment d'**un brin d'ADN dénaturé par la chaleur** (= **dont les deux brins ont été artificiellement séparés thermiquement**) et l'**ARN messager correspondant à son expression (technique d'hybridation, basée sur la complémentarité entre nucléotides)**, l'équipe du médecin strasbourgeois Philippe CHAMBON (né en 1931) met en évidence l'existence de **séquences dans l'ADN qui ne se retrouvent pas dans l'ARNm (figure 66)**. Il s'agit des **introns, séquences non codantes qui subissent une excision lors de la maturation de l'ARNm**.
- Hélas, cette équipe n'aura pas le prix Nobel car les introns avaient été découverts un an plus tôt sur un **ADN de virus** par des chercheurs américains (Philipp A. SHARP [1944] & Richard J. ROBERTS [1943] – Nobel 1993).



Un résultat surprenant obtenu par des chercheurs de Strasbourg en 1977.



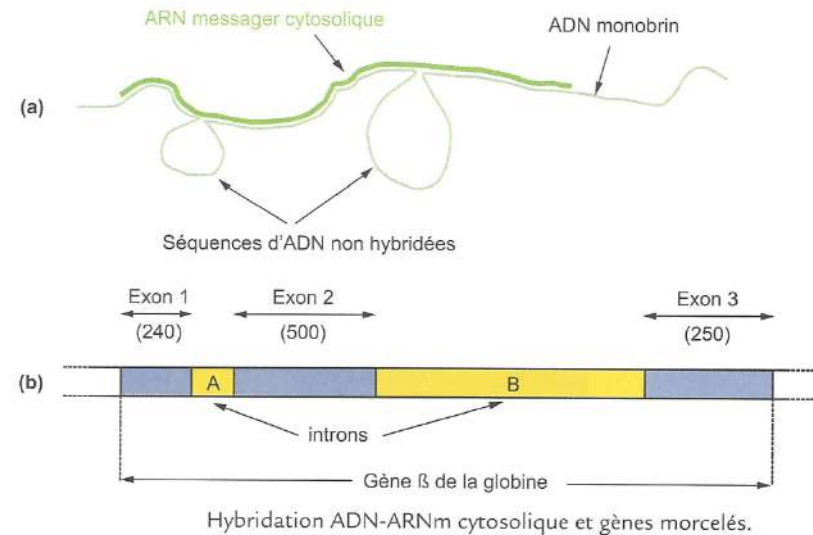
L'ARN messager cytoplasmique n'est pas une copie complète du gène.

▲ **FIGURE 66. La découverte des introns chez les Eucaryotes et l'idée d'excision-épissage des ARN pré-messagers.** Manuel de Terminale S, spécialité SVT (dir. R. TAVERNIER & C. LIZEAUX, Bordas, Paris, 2002).

b. Organisation des gènes eucaryotes

- Les **gènes eucaryotes** sont fondamentalement **monocistroniques** : ils sont **composés d'une seule séquence codante de protéine**. En revanche, cette séquence est fragmentée en **exons** (**portions présentes dans l'ARNm final**) interrompues de **portions non codantes excisées** nommées **introns** (figures 67-69). La **transcription** d'un ADN produit ainsi un **ARN pré-messager (ARNpm)** qui subit une maturation caractérisée par l'**élimination des introns (excision)** et le **raboutage des exons (épissage)** : il y a alors obtention d'un **ARN messager (ARNm)**.

Les données ont été acquises grâce à la technique d'hybridation moléculaire entre un segment d'ADN génomique et l'ARNm cytosolique codé par ce même segment d'ADN (figure 17.5). Cette technique a été appliquée aux gènes de l'ovalbumine et de la globine β . L'ADN génomique est rendu monocaténaire par dénaturation thermique puis il est mis en présence de son ARNm cytosolique. La microscopie électronique à transmission permet de visualiser l'appariement éventuel de ces deux molécules a priori complémentaires.



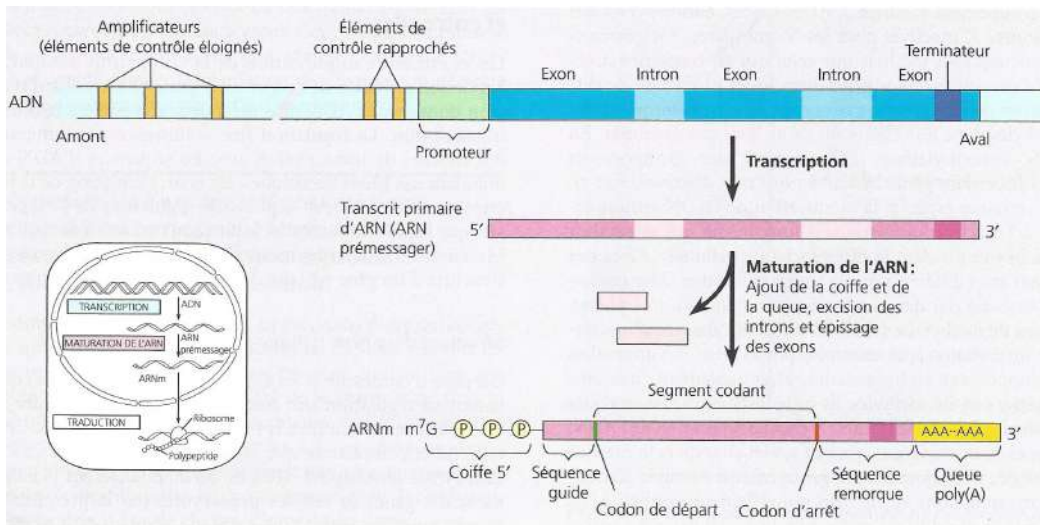
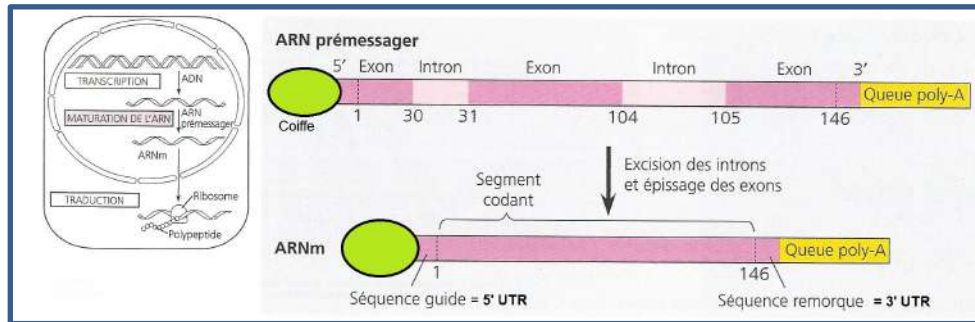
Cas du gène β de la globine des mammifères. Il comprend 3 exons et 2 introns ; l'ARNm final correspond aux seuls exons. Les chiffres indiquent les nombres de paires de bases, respectivement 120 et 550 pour les introns A et B. (a) Figures d'hybridation (d'après observation au MET), (b) Gène β de la globine.

Plusieurs constats s'imposent :

- l'ADN du gène est beaucoup plus long que l'ARNm ;
- les deux extrémités de l'ADN n'ont pas d'équivalent chez l'ARNm ;
- l'ADN montre des secteurs appariés à l'ARNm séparés par des boucles d'ADN dépourvues de segment complémentaire chez l'ARNm.

▲ **FIGURE 67. Une autre représentation des expériences d'hybridation ARN-ADN et du morcellement des gènes. Les portions régulatrices ne sont pas représentées.** D'après PEYCRU *et al.* (2013).

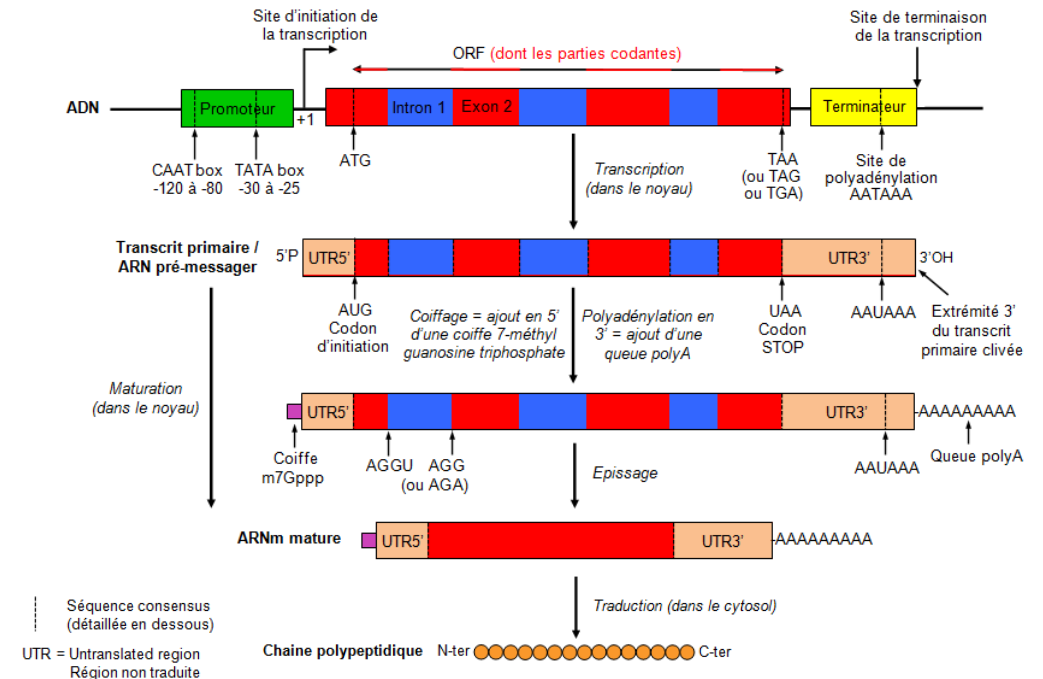
- **IMPORTANT.** Les gènes eucaryotes présentent également des **séquences régulatrices** en amont voire en aval des **portions codantes** (figure 68). *L'organisation des gènes eucaryotes sera précisée plus tard, notamment dans le chapitre sur l'expression génétique.*
- Le cas des **transcrits issus de gènes codant des protéines** chez les eucaryotes est particulièrement intéressant. Les modifications sont **nucléaires** et, là encore, **majoritairement co-transcriptionnelles**. Le **transcrit primaire** est appelé **ARN pré-messager (ARNpm)** et la **structure finale après maturation** constitue l'**ARN messager (ARNm)** à proprement parler (figure 68-69).



▲ FIGURE 68. **Le gène eucaryote : une vision précisée.** D'après CAMPBELL & REECE (2004).

On notera qu'en amont du premier nucléotide codant et en aval du dernier nucléotide codant, on trouve des **séquences non traduites** ou **UTR** (UnTranslated Regions). On les appelle respectivement **séquence 5' UTR** (ou **séquence guide**) et **séquence 3' UTR** (ou **séquence remorque**). La première, notamment, permet le **positionnement du ribosome** et peut participer au **contrôle de l'expression génétique**.

Soyons précis ! >> Si l'on veut être rigoureux, un **exon** n'est **pas exactement une portion nécessairement codante** mais plutôt **une portion d'ADN transcrite en ARNpm et qui demeure dans l'ARNm final après excision-épissage**. Au sein de ces exons, on trouve ainsi en amont du codon start et en aval du codon stop des **portions qui ne seront pas traduites en acides aminés**, que l'on peut donc considérer comme **non codantes (séquences UTR)**.

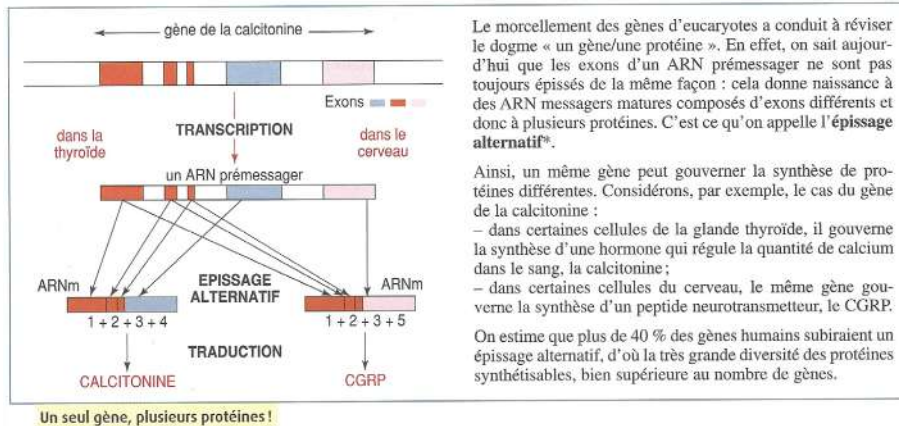


▲ FIGURE 69. **Le gène eucaryote.** Document S. FABRE (Lycée Chateaubriand, Rennes)
L'ORF est ici comprise dans sa première définition et inclut, pour ma collègue, les introns.

- Remarque : **certaines gènes eucaryotes peuvent coder plusieurs protéines** par le biais de l'**épissage alternatif** (encadré G).

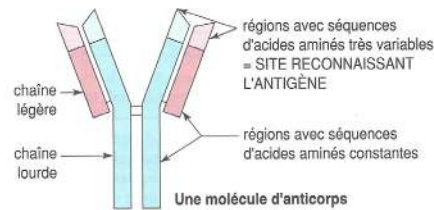
Encadré G Un gène, plusieurs protéines ?

➤ La **figure** ci-dessous présente la notion d'**épissage alternatif** et initie à la **génétique** très complexe **des anticorps**.



Vous avez vu ou vous verrez en immunologie que l'organisme est susceptible de fabriquer des milliards d'anticorps différents capables chacun de reconnaître un antigène et un seul. Les anticorps, synthétisés par les lymphocytes B, sont formés par 4 chaînes polypeptidiques et c'est une région déterminée des chaînes, dont la séquence d'acides aminés est différente d'un anticorps à un autre, qui permet cette reconnaissance spécifique.

La variabilité des anticorps est codée génétiquement. Or, chez l'homme, il n'y a pas des milliards de gènes mais sans doute seulement entre 20 000 et 30 000 gènes. Lors de la formation d'un lymphocyte B à partir d'une cellule souche, différents segments de gènes répartis en différents ensembles sur les chromosomes sont assemblés au hasard. Chaque lymphocyte B né d'une cellule souche possède donc un réarrangement unique de segments de gènes ; cet assem-



blage constitue son gène fonctionnel qui gouverne la synthèse d'un anticorps. Deux lymphocytes B issus de deux cellules souches différentes produisent donc des anticorps différents.

Un exemple plus complexe : une combinatoire entre segments d'ADN.

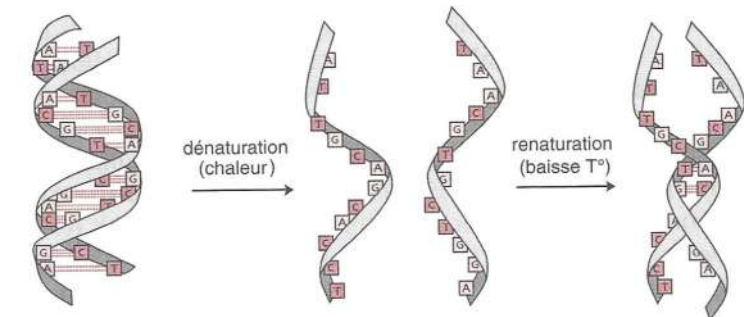
FIGURE a. Épissage alternatif (en haut). Génétique des anticorps (en bas).

Manuel de Terminale S, spécialité SVT (dir. R. TAVERNIER & C. LIZEAUX, Bordas, Paris, 2002).

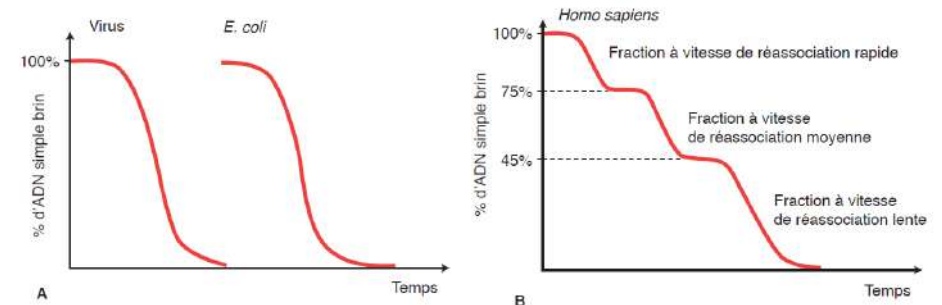
C. La présence importante de séquences répétées et d'ADN non codant chez les Eucaryotes

1. Mise en évidence des séquences répétitives par les expériences de dénaturation-renaturation de l'ADN

- Par la **chaleur** (et éventuellement de l'urée), on peut **dénaturer** l'ADN, c'est-à-dire **en séparer les deux brins** (suppression des liaisons H). La **baisse de température** aboutit à la **reformation de la structure bicaténaire** (**renaturation**).
- On procède à cette **expérience de dénaturation-renaturation** sur des **génomes** préalablement **digérés** par des **enzymes de restriction** en suivant la **cinétique de renaturation** (figures 70-71). La **renaturation de l'ADN** est mesurée par l'**absorption des UV**, plus forte lorsque les **bases azotées** ne sont **pas appariées**.



▲ FIGURE 70. Principe de la dénaturation-renaturation de l'ADN. D'après PETIT & JULIEN (2007).



Les cinétiques monophasiques de réassociation de l'ADN dénaturé d'organismes sont caractéristiques de génome ne présentant pas de séquences répétées (A). Chez l'Homme (B), apparaissent trois paliers correspondant aux trois fractions retrouvées : plus le réappariement est rapide, plus les séquences se réhybrident sont présentes en grand nombre.

▲ FIGURE 71. Résultats des expériences de dénaturation-renaturation de l'ADN d'un virus, d'*E. coli* et de l'Homme. D'après SEGARRA *et al.* (2014).

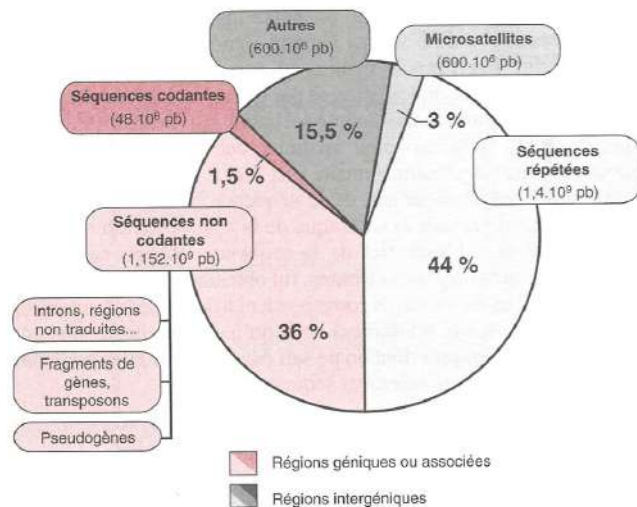
- Deux situations (figure 70) sont alors possibles :
 - Si les deux brins se renaturent à une vitesse qui ne subit pas de changements brusques, on en déduit que chaque brin ne peut s'apparier qu'avec son brin d'origine porteur de la séquence complémentaire. C'est le cas des **Bactéries** (mais aussi de la plupart des **virus**) chez lesquelles on admet qu'il y a **très peu de séquences répétées**.

Notons que, **chez les Eubactéries, on considère que la majorité de l'ADN est codant**. Chez *E. coli* par exemple, **90 % de l'ADN** serait transcrit en ARNm. Les **10 % restants** correspondent en majorité à des **portions régulatrices** (+ un peu d'ADN espaceur et des **séquences transposables**) ainsi que les séquences codant les **ARNr, ARNt...**

- Si les deux brins se renaturent à une vitesse qui présente des paliers, on en déduit que **les portions qui se renaturent en premier sont hautement répétées** (elles trouvent facilement une séquence complémentaire) alors que **les séquences qui se renaturent en dernier ne sont présentes qu'en un seul exemplaire** (elles ne peuvent s'associer qu'avec leur séquence complémentaire d'origine). C'est le cas chez beaucoup d'**Eucaryotes**.

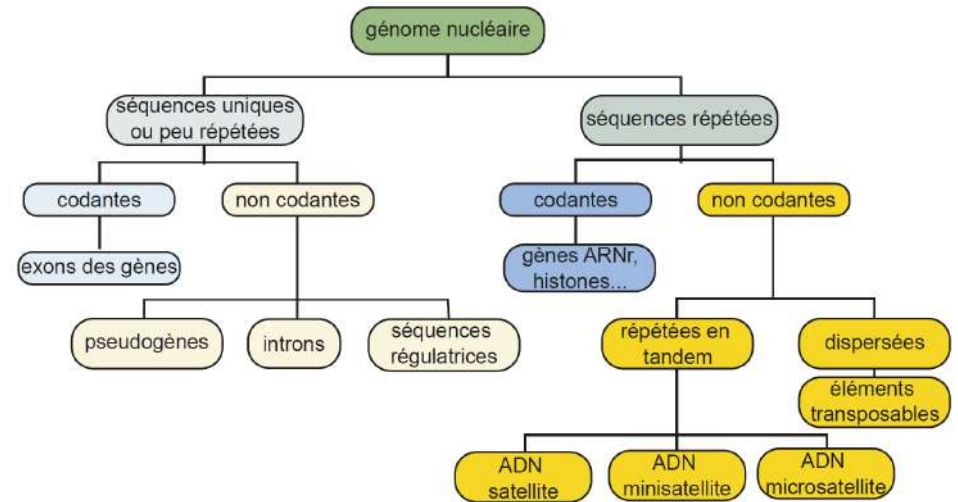
2. Nature des séquences répétitives des Eucaryotes

- La **proportion de séquences répétitives** varie grandement d'un Eucaryote à un autre (figures 72-73 + tableau VIII). Si le **génome humain** en comprend **55 %**, celui de la **Levure du boulanger *Saccharomyces cerevisiae*** n'en comprend **presque pas**. Les **pourcentages évalués** – de même que la typologie – **diffèrent** assez notablement d'un auteur à l'autre.



Différents types de séquences au sein du génome humain.
La majorité des séquences ne codent pas des protéines. Seules 1,5 % des séquences soit 48 mégabases sur 3 200 pour le génome complet, correspondent à des exons au sein de gènes codants.

▲ FIGURE 72. **Composition du génome nucléaire humain : une vision possible.**
D'après PETIT & JULIEN (2007).



▲ FIGURE 73. **Composition d'un génome nucléaire eucaryote : une vision possible.**
D'après PERRIER, BEAUX *et al.* (2021).

- Quelques éléments à retenir au-delà du **tableau** :
 - Le **génome correspondant à des séquences codantes (CDS) de protéines**, donc des **exons hors UTR**, serait de l'ordre d'**1,5 % du génome humain**.
 - Les **gènes**, notamment **protéiques**, peuvent être **extrêmement longs** : un **gène protéique humaine** mesure environ **27 000 pb en moyenne** mais seuls **1300 pb** sont nécessaires à la **synthèse d'une protéine moyenne de 430 acides aminés**. **Beaucoup d'ADN génique** correspond donc à des **introns** qui seront **excisés**. En moyenne, un **gène humain** comprend **8,8 exons** et **7,8 introns** avec des **introns** souvent nettement **plus longs** que les **exons**.
 - Notons que les **séquences codant des ARN non messagers** sont considérées par certains auteurs comme « non codantes », restreignant la notion d'**ADN codant (ADNc)** aux **exons protéiques**.
 - Pour désigner l'**ADN non codant à fonction inconnue**, certains auteurs parlent d'**ADN « poubelle » (junk DNA)**, notamment celui sans fonction connue, mais des **fonctions** tendent à être élucidées à mesure que la recherche avance.

Bilan (adapté du programme)

- ✓ Chez les **Eucaryotes**, on distingue le **génome nucléaire** et le **génome des organites**. Le **génome nucléaire** est constitué de **chromosomes linéaires**.
- ✓ L'**ADN nucléaire des Eucaryotes** est associé à des **protéines** dont des **histones**, constituant la **chromatine**. Il existe différents **niveaux de condensation** de la **chromatine**.
- ✓ Le **génome nucléaire des Eucaryotes** comporte une part importante de **séquences non codantes** aux **rôles divers**. La majorité de ces **séquences** est **répétée**.
- ✓ Les **gènes eucaryotes** sont généralement **morcelés**.

▼ **TABLEAU VIII. Typologie de l'ADN nucléaire (centrée sur le génome humain).**

Librement mais fortement inspiré d'un document de S. FABRE (Lycée Chateaubriand, Rennes), reprenant en partie des données de PERRIER, BEAUX et al. (2021) ou de Wikipédia

Des **séquences en tandem** sont des **motifs nucléotidiques répétitifs disposés les uns à la suite des autres de manière immédiatement adjacente**.

Les proportions estimées varient beaucoup selon les auteurs et sont plus qu'indicatives...

Niveau de répétition	Expression	Nature et caractéristiques	
Séquences hautement répétées (ADN satellite sens large) Homme : ~ 10 %	Non transcrites	➤ ADN satellite sens strict • Séquences CEN (centromériques) : séquences de 171 pb, s'associant au complexe kinétochorion. • Séquences TEL (télomériques) : séquences de pb (TTAGG), intervenant dans la protection des extrémités chromosomiques et l'association à la lamina ou au nucléole.	
		➤ ADN minisatellite (VTNR, variable number of tandem repeat) : fragments de 1 à 5 kb formés de motifs répétés de 15 à 100 pb, dispersés dans le génome, à fonction peu évidente (régulation de l'expression ?) et à fort polymorphisme. ➤ ADN microsatellite (STR, short tandem repetition) : fragments inférieurs à 150 pb formés de motifs répétés de quelques pb (1-4), dispersés dans le génome, à fonction inconnue et à fort polymorphisme. Les empreintes génétiques sont aujourd'hui basées sur ces séquences.	
Séquences moyennement répétées Homme : ~ 45 %	Non transcrites (quand un ARN est formé, c'est une copie ARN)	➤ Éléments génétiques mobiles (transposons s. l. = éléments transposables) : éléments génétiques capables de se déplacer au sein du génome de manière autonome par un phénomène nommé transposition (encadré H). 46 % du génome humain* • Rétrotransposons [classe I] : séquences d'ADN copiées en ARN (copie du rétrotransposon), rétro-transcrit ensuite par une transcriptase inverse en ADN pouvant s'insérer dans l'ADN. - On y trouve : - les LINE (Long Interspersed Nuclear Elements, longs éléments nucléaires intercalés) : 5000-7000 pb, quelques milliers de copies complètes, quelques dizaines de milliers de copies tronquées - les SINE (Short Interspersed Nuclear Elements, courts éléments nucléaires intercalés) : 130-700 pb, près d'un million de copies - le rétrotransposon L1 qui occupe à lui seul 17 % du génome (500 000 copies) • Transposons au sens strict [classe II] : séquences d'ADN excisées avant d'être insérées dans l'ADN cible ou bien séquences d'ADN répliquées directement en ADN avant d'être insérées. On y trouve : - les MITE (Miniatures Inverted Repeats Transposable Elements, Éléments Transposables à Répétitions Inversées Miniatures) : environ 400 pb bordées par des séquences inversées répétées	
	Transcrites	➤ Gènes répétés en tandem : • gènes répétés codant des protéines : histones (200 copies), actine... • gènes codant des ARN non messagers : ARNt de transfert (20 000 copies), ARN ribosomiques, ARN interférents...	
Séquences uniques (très peu ou pas répétées) Homme : ~ 45 %	Transcrites	➤ Gènes codant des protéines non (ou peu) répétés, incluant les introns [20 000 – 25 000 chez l'Homme] • gènes de ménage : gènes protéiques exprimés par tous les types cellulaires (ex. codant les enzymes de la glycolyse). • gènes tissu-spécifiques : gènes protéiques exprimés dans certains types cellulaires (ex. codant l'insuline) (!) La plupart sont en un seul exemplaire, mais il existe des gènes apparentés dérivant d'une séquence initiale et perdurant en divers exemplaires dans le génome : ce sont des gènes paralogues qui forment une famille multigénique. Chaque gène peut alors acquérir une fonction propre.	
	Non transcrites	➤ Séquences régulatrices spécifiques ➤ Pseudogènes : anciens gènes fonctionnels devenus inactifs au sein d'un génome, du fait d'altérations génétiques comme des mutations les rendant incapables de conduire à la production d'une protéine. ➤ Éléments transposables peu / pas répétés (*à inclure dans les 46 % cités) ➤ Autres séquences d'ADN espaceur non répété...	

Encadré H Les éléments transposables

➤ Les **éléments transposables** sont de **courtes séquences d'ADN de 5000 à 10000 pb capables de se propager de manière autonome dans le génome grâce à des enzymes de transport (transposases)**.

➤ On distingue :

a) Les **rétrotransposons (classe I)** qui sont **copiés en ARN (copie du rétrotransposon), rétro-transcrit ensuite par une transcriptase inverse en ADN pouvant s'insérer dans l'ADN**.

NB La **transposition** est alors **réplicative** et suppose en outre une étape de **transcription** puis **rétrotranscription**.

b) Les **transposons sensu stricto (classe II)** qui sont des **séquences d'ADN excisées avant d'être insérées dans l'ADN cible ou bien séquences d'ADN répliquées directement en ADN avant d'être insérées**.

NB La **transposition** peut être **réplicative** (« copier coller ») ou **non-réplicative** (= par excision-insertion) (« couper coller »).

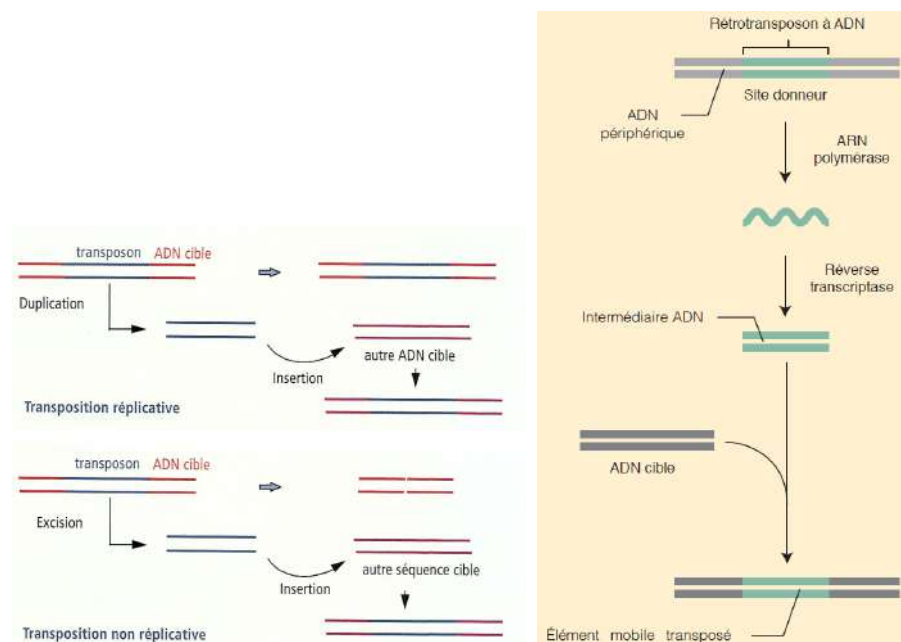


FIGURE a. **La transposition sensu stricto (à gauche) et la rétrotransposition (à droite).**

D'après HARRY (2008) et SEGARRA et al. (2014)

➤ La **transposition** peut entraîner des **réarrangements chromosomiques** et certains auteurs pensent qu'ils interviennent **plus souvent** que les **dysfonctionnements de méiose**. Des gènes peuvent aussi être **coupés** ou **modifiés** suite au **phénomène (figure b)**.

IV. Les virus et leur génome

- Les **virus** (du lat. *virus*, poison) sont des **entités sub-biologiques comprenant un acide nucléique et des protéines, ne réalisant pas de métabolisme autonome et devant parasiter le système d'expression des cellules dans le cadre de leur reproduction et leur survie.**

Les virus sont-ils des êtres vivants ?

Il n'existe **pas forcément de réponse formelle** à cette question qui mériterait un long développement.

- Comme les **êtres vivants**, les **virus** sont **composés** de matière organique : **protéines, acides nucléiques, glucides**, parfois **lipides**...
- Comme les **êtres vivants**, les **virus** possèdent un **génome** et **se reproduisent**.
- Contrairement aux êtres vivants, les **virus** sont **dépourvus de cellule, organites ou métabolisme**, ce qu'on peut nommer un **état acellulaire** ou **acaryote** (voir plus loin).
- Contrairement aux êtres vivants, ils sont **incapables** de réaliser une **activité métabolique autonome**, d'exprimer de l'information **génétique** ou de **se reproduire** sans utiliser et détourner le **matériel enzymatique** des **cellules** qu'ils infectent. Ce sont donc des **parasites obligatoires de cellules** dont ils **détournent la machinerie cellulaire**, notamment celle qui permet l'**expression génétique**.

Capacités exigibles

- ✓ **Comparer** l'organisation du génome des bactéries, des Eucaryotes et des virus.
- ✓ **Estimer** la proportion de séquences codantes et non codantes dans les génomes des Eucaryotes, des bactéries et des virus.
- ✓ **Illustrer** la diversité structurale et la diversité d'hôte des virus

A. Les virions (= particules virales), état libre des virus

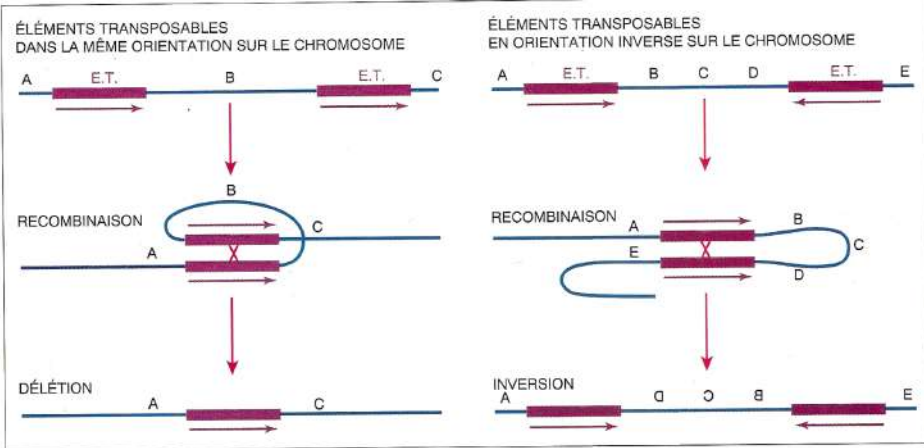
- On appelle **virion** ou **particule virale** une **structure biologique acellulaire de petite taille (généralement comprise entre 50 et 250 nm) comprenant une information génétique enfermée dans une structure macromoléculaire protectrice et capable d'infecter une cellule**. Il s'agit donc de l'état libre du virus.

1. Organisation : une information génétique enfermée dans une capsid et éventuellement une enveloppe

a. Une information génétique portée par un acide nucléique

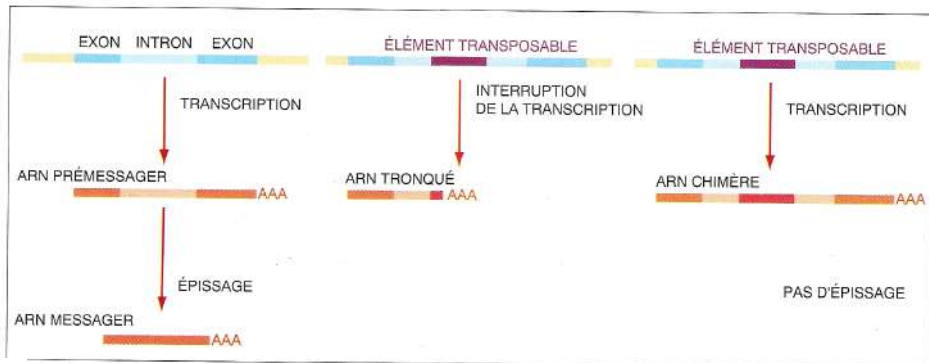
a. Un acide nucléique de nature variable qui sert de base à la classification

- Il existe une **grande diversité de molécules porteuses du génome viral (tableau IX)**, généralement **linéaire** :
 - ADNbc (bicaténaire) ((cas des **bactériophages**))
 - ADNmc (monocaténaire) [**plutôt rare**]
 - ARNbc (bicaténaire) [**plutôt rare**]
 - ARNmc (monocaténaire). Dans ce dernier cas, trois situations peuvent se présenter :
 - L'**ARN viral** est **immédiatement traductible**, comme un **ARNm**, en **protéines** : c'est un **ARN à polarité positive** ou **ARN +** ((cas du **VMT** ou des **Coronavirus**))
 - L'**ARN viral** est **copié en ARN+** qui est alors traduit : l'ARN viral initial est donc un **ARN à polarité négative** ou **ARN -** ((cas du **Virus de la grippe-influenza**))



REMANIEMENTS CHROMOSOMIQUES causés par la présence d'éléments transposables dans le génome. Un événement de recombinaison entre les séquences homologues de deux éléments transposables entraîne,

selon leur orientation respective, la **délétion** ou l'**inversion** des séquences qu'ils encadrent. En modifiant l'environnement des gènes, une inversion peut altérer leur expression.



EFFETS DE L'INSERTION D'UN ÉLÉMENT transposable sur l'expression d'un gène. Le gène normal est d'abord transcrit en une molécule d'ARN pré-messager, qui subit ensuite l'épissage de la séquence intronique.

Inséré dans l'intron, l'élément transposable peut interrompre prématurément la transcription (production d'un ARN tronqué) ou empêcher l'épissage de la séquence ARN intronique (production d'un ARN chimère).

FIGURE b. Quelques conséquences possible de l'insertion d'éléments transposables dans le génome. D'après BAZIN *et al.* in LE GUYADER (1998)

▼ **TABLEAU IX. Classification des virus (avec des exemples de virus animaux, avec signalement des virus au programme).** D'après CAMPBELL & REECE (2004).

Classification, selon le type d'acide nucléique, des Virus parasites d'Animaux	
CLASSE*	
Famille (genre)	Exemples, maladies
I. ADN bicaténaire	
<i>Papovaviridae</i> (Papovavirus)	Papillomes (chez l'Humain, verrues, cancer du col utérin); polyomes (tumeurs chez certaines espèces animales)
<i>Adenoviridae</i> (Mastadenovirus, Aviadenovirus)	Maladies respiratoires; certains virus provoquent des tumeurs chez les Mammifères (Mastadenovirus) ou les Oiseaux (Aviadenovirus).
<i>Herpesviridae</i> (Simplexvirus, Varicellovirus)	Herpès labial (Herpès simplex I); herpès génital (Herpès simplex II); varicelle, zona (Herpèsvirus III humain); mononucléose, lymphome de Burkitt (Virus d'Epstein-Barr)
<i>Poxviridae</i> (Orthopoxvirus)	Variole; vaccine
II. ADN monocaténaire	
<i>Parvoviridae</i> (Parvovirus)	Roséole (à noter que la croissance des Parvovirus dépend généralement d'une infection simultanée par des Adénovirus)
III. ARN bicaténaire	
<i>Reoviridae</i> (Orthoreovirus)	Diarrhée; maladies respiratoires bénignes
IV. ARN monocaténaire pouvant jouer le rôle d'ARNm ARN +	
<i>Picornaviridae</i> (Enterovirus, Rhinovirus)	Poliomyélite et maladies intestinales (Enterovirus); rhume (Rhinovirus)
<i>Togaviridae</i> (Rubivirus, Alphavirus)	Rubéole (Rubivirus); fièvre jaune; encéphalite (Alphavirus)
V. ARN monocaténaire servant de matrice pour l'ARNm ARN -	
<i>Rhabdoviridae</i> (Lyssavirus)	Rage
<i>Paramyxoviridae</i> (Morbillivirus, Rubulavirus)	Rougeole (Morbillivirus); oreillons (Rubulavirus)
<i>Orthomyxoviridae</i> (Virus de l'influenza)	Grippe (Virus de l'influenza de type A, B ou C)
VI. ARN monocaténaire servant de matrice pour la synthèse de l'ADN ARN + rétrotranscrit	
<i>Retroviridae</i> (Lentivirus)	Leucémie, lymphome (Virus oncogènes à ARN); sida (VIH, Lentivirus)
* Les familles à l'intérieur de chaque classe (ou groupe, selon certaines taxinomies) diffèrent surtout par la structure de la capside et par la présence ou l'absence d'une enveloppe membraneuse. Nous avons identifié un ou deux genres seulement par famille.	

- Chez les **rétrovirus**, l'ARN (bien que de sens +), n'est pas traduit mais **rétrotranscrit en ADN double brin** (par des **transcriptases inverses** apportées par le virion) ((cas du VIH)).

- L'information génétique est stabilisée par des **protéines basiques** associées. Il peut y avoir **une ou plusieurs molécules** d'acides nucléiques.

Signalons que ce n'est pas l'état bicaténaire ou monocaténaire qui définit si un acide nucléique est un ADN ou un ARN mais bien le **type de pentose** (désoxyribose ou ribose). Il se trouve que, chez les **êtres vivants cellulaires**, l'ADN est **bicaténaire** et l'ARN est (basiquement) **monocaténaire**... mais ce n'est pas forcément le cas chez les virus.

Pour information

- Les **virus** sont regroupés artificiellement en « **groupes** » (Groupe I, Groupe II... jusqu'à groupe VI) qui se basent sur le **type d'information génétique** (tableau IX). La classification comprend ensuite **divers rangs taxonomiques** (classes, ordres, famille, sous-famille, genre, espèce) basées sur les **caractéristiques structurales des virions** et le **fonctionnement viral**.
- Des tentatives de **classification phylogénétique** existent mais se heurtent à la **complexe évolution virale** (voir plus loin).

β. Une information réduite compactée et composée de quelques gènes à quelques dizaines, avec une possibilité de chevauchement des cadres de lecture

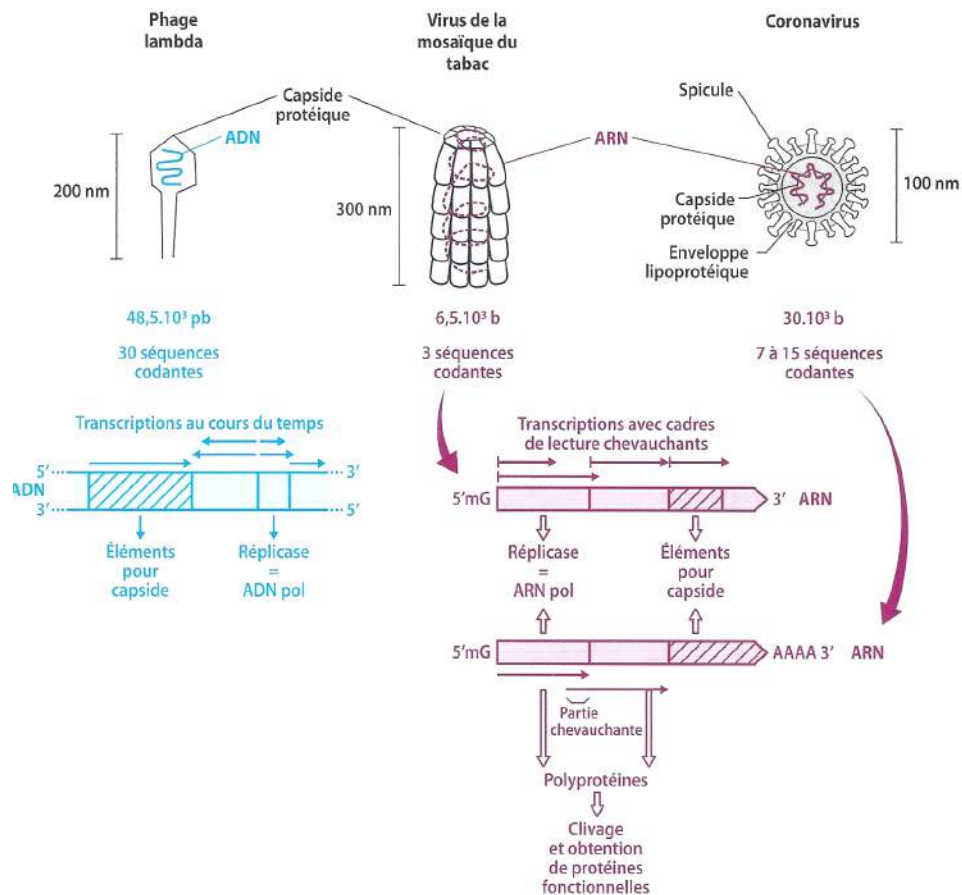
- La **taille du génome**
 - diffère considérablement au sein des **virus à ADN** (3 à 300 kb),
Record : **virus géants d'amibes avec un génome de l'ordre de la Mb** (encadré I)
 - alors qu'elle est comprise **entre 10 et 20 kb** pour la plupart des **virus à ARN**

Ces valeurs sont à **comparer** aux tailles des **génomés vivants** (*E. coli* 4,34 Mb, *H. sapiens* 3,4 Gb). Les **virus** ont donc, **globalement**, des **génomés** notoirement **plus petits**.

- Voilà **trois exemples** que vous pouvez retenir (tableau X + figure 74).

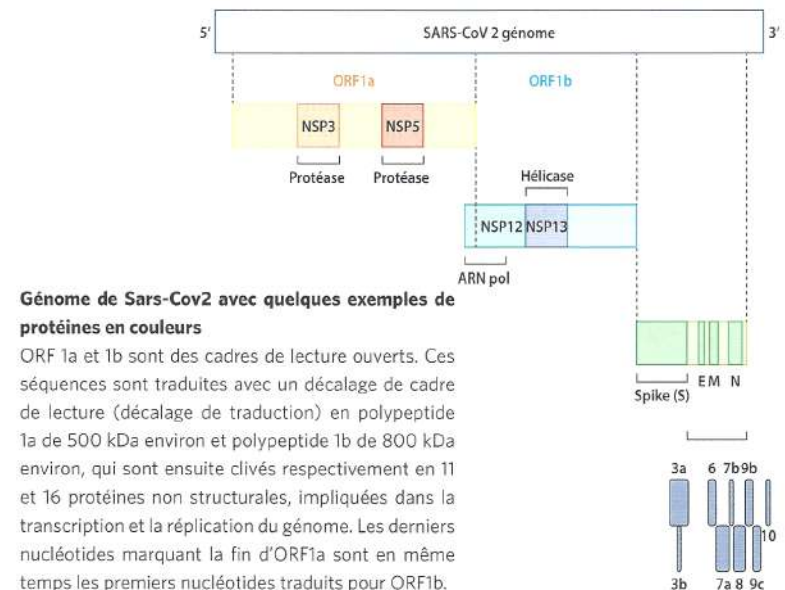
▼ **TABLEAU X. Le génome des trois virus au programme.**

Virus	Bactériophage (exemple : phage λ)	VMT (Virus de la Mosaïque du Tabac)	Coronavirus (Sars Cov-2)
Type de génome	ADN double brin (linéaire mais à bouts collants, autorisant la cyclisation)	ARN (simple brin) +	ARN (simple brin) +
Taille	48,5 kb	6,4 kb	27-32 kb
Nombre de gènes	50 à 70 gènes protéiques (selon les auteurs)	3 gènes (une réplicase, la protéine de la capside et une protéine de mouvement)	13 gènes protéiques (protéine Spike, protéine de nucléocapside...) + 2 ORF chevauchants codant 2 longs polypeptides produisant chacun 16 protéines fonctionnelles après clivage (figures 75-76)



▲ FIGURE 74. Génomes viraux. D'après DAUTEL *et al.* (2021).

- La capacité plutôt réduite de **codage des génomes viraux** est souvent compensée par un **chevauchement des cadres ouverts de lecture** (figures 75-76) voire par le **phénomène d'épissage alternatif des ARN messagers** (découvert chez les adénovirus). Les **introns** peuvent en effet exister chez les **virus d'Eucaryotes**.
- Le **génome viral**, **codant à plus de 90 %**, comprend souvent de **quelques gènes à quelques dizaines**. Les **protéines codées** peuvent être des **protéines de la capsid** ou de **l'enveloppe**, des **enzymes** synthétisant **d'autres constituants**, des **enzymes** impliquées dans la **réplication** et le **cycle viral**...

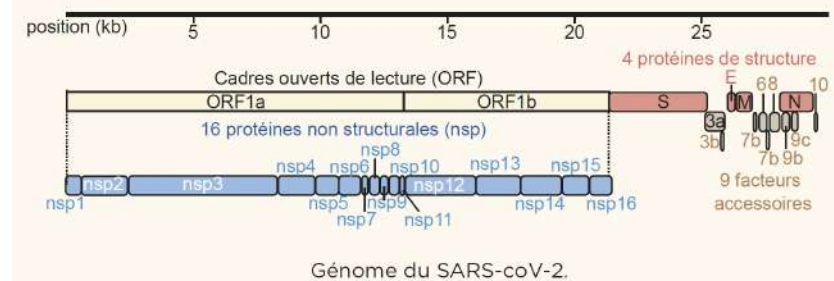


Génome de Sars-Cov2 avec quelques exemples de protéines en couleurs

ORF 1a et 1b sont des cadres de lecture ouverts. Ces séquences sont traduites avec un décalage de cadre de lecture (décalage de traduction) en polypeptide 1a de 500 kDa environ et polypeptide 1b de 800 kDa environ, qui sont ensuite clivés respectivement en 11 et 16 protéines non structurales, impliquées dans la transcription et la réplication du génome. Les derniers nucléotides marquant la fin d'ORF1a sont en même temps les premiers nucléotides traduits pour ORF1b.

▲ FIGURE 75. Génome du Sars-Cov-2. D'après DAUTEL *et al.* (2021).

Le génome du SARS-CoV-2 est un ARN de 30 kb qui code notamment pour 4 protéines de structure (S, E, M, N) dont la protéine S (pour spike) qui permet l'entrée du virus dans les cellules épithéliales en se fixant sur une molécule de celles-ci et la protéine N (nucléoprotéine) : elle se lie à l'ARN et forme autour de lui la nucléocapside.



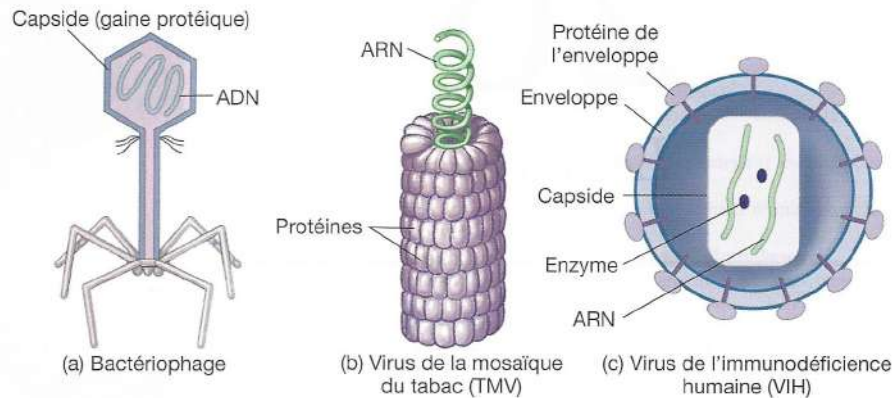
Génome du SARS-coV-2.

Le génome comprend aussi deux ORF, cadres ouverts de lecture (*open reading frame*) : Chacun correspond à une séquence codante non interrompue par un codon stop. La traduction produit deux longs polypeptides ensuite scindés en protéines non structurales (nsp) impliquées dans la réplication du virus.

▲ FIGURE 76. Génome du Sars-Cov-2. D'après PERRIER, BEAUX *et al.* (2021).

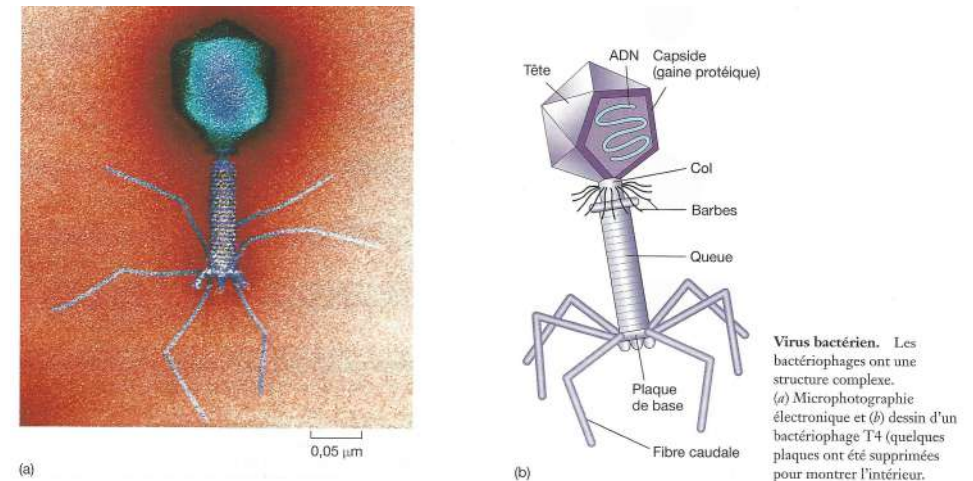
b. La présence d'une capside protéique protectrice

- La **capside** (figures 77-80) est une **structure protéique dans laquelle est enfermé le matériel génétique des virions**. On peut y trouver des **enzymes** (ex. dans le VIH : transcriptases inverses permettant la **rétrotranscription** de l'ARN viral en ADN).
- La **capside** a souvent une **organisation géométrique** : ex. **icosaédrique** pour le phage, **hélicoïdale** pour le VMT. Chez les **Coronavirus**, on trouve une **nucléocapside**, **structure composée d'ARN et de protéines de capsid** **immédiatement associées à l'acide nucléique**.
- La capsid peut contenir des **protéines** ou **assemblages protéiques** utiles dans l'**infection** (exemple : **fibres caudales des certains Bactériophage** permettant d'agripper la Bactérie).

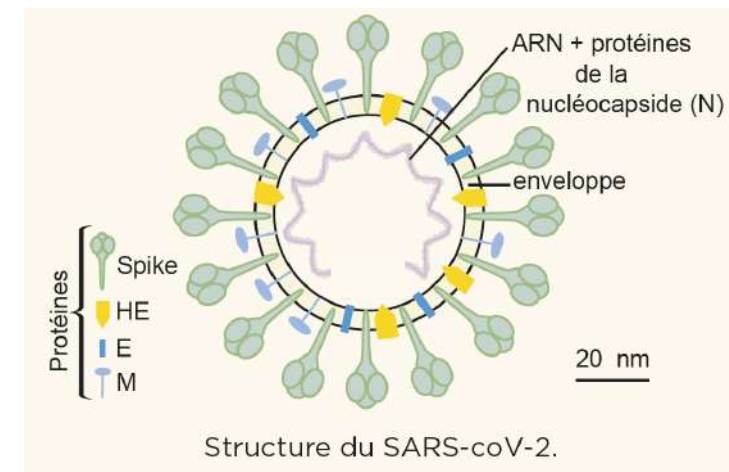


Structure de virus bactérien, végétal et animal. (a) Les virus bactériens, ou bactériophages, ont souvent une structure complexe. (b) Le TMV infecte les plantes et consiste en 2130 molécules protéiques identiques (*en pourpre*) formant une enveloppe cylindrique autour de l'unique brin d'ARN (*en vert*). La colonne vertébrale d'ARN détermine la forme du virus et est protégée par des molécules identiques de protéine qui l'entourent étroitement. (c) Chez les virus de l'immunodéficience humaine (VIH), l'ARN central est maintenu à l'intérieur d'une capsid entourée d'une enveloppe protéique.

▲ **FIGURE 77. Trois virions de trois virus différents.** Notez que certains auteurs représentent deux capsides pour le VIH (une **capside interne** et une **capside externe** nommée **matrice** sous la membrane). D'après RAVEN *et al.* (2007).

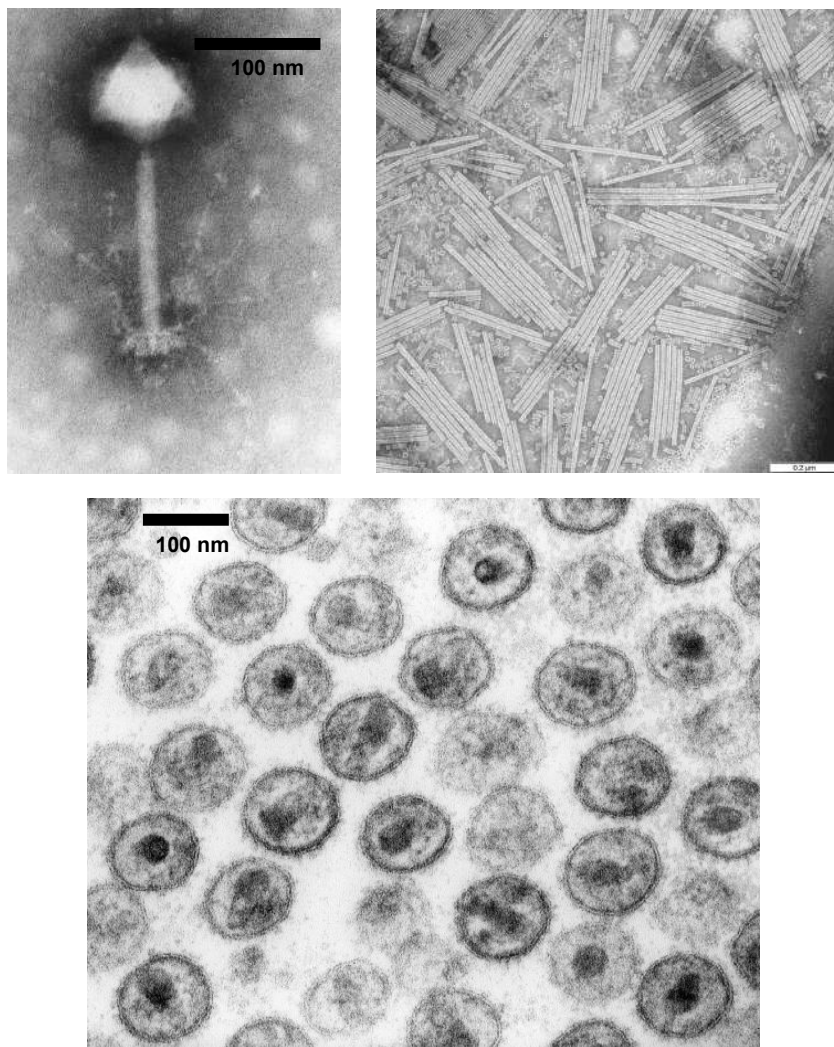


▲ **FIGURE 78. Gros plan sur un Bactériophage T4.** D'après RAVEN *et al.* (2007).



▲ **FIGURE 79. Structure du SARS-Cov2 (env. 100 nm), adapté.** D'après PERRIER, BEAUX *et al.* (2021).

Protéine *spike* = péplomère = spicule = protéine de pointe
HE = hémagglutinine estérase | E = Protéine d'enveloppe | M = Protéine de membrane



▲ FIGURE 80. Trois types de virions au microscope électronique : bactériophage S-PM2 (env. 200 nm), VMT (largeur env. 18 nm) et VIH (taille env. 100 nm).
D'après Wikipédia (consultation novembre 2015).

c. La présence fréquente d'une enveloppe (= péplos) de nature membranaire

- On appelle **enveloppe** ou **péplos** (figure 79) un **reste de membrane plasmique de cellule infectée enfermant la capside de certains virions**. Cette structure peut contenir des **protéines** utilisées dans l'infection des cellules (exemples : **spike** des **Coronavirus**, gp120 du VIH).
- Les **virus sans enveloppe** (ex. phage, VMT...) sont appelés **virus nus**.

2. Les virions, entités composées de matière organique mais acellulaires et « acaryotes »

- Les **virions** sont composés d'eau et de **matière organique** : **acides nucléiques**, **protéines** + **constituants membranaires** en cas d'enveloppe (protéines, lipides, glucides).
- Les **virions** n'ayant **pas d'activité propre**, les virus sont **dépendants des cellules** sur le plan **énergétique** (utilisation de l'énergie des cellules infectées) et **métabolique** (utilisation des **enzymes** – dont les enzymes de l'expression génétique – et **métabolites** de la cellule).
- Dépourvus de cellules**, on les dit **acellulaires**.
- Dépourvus de noyau et plus généralement de tout compartiment cellulaire**, on peut les dire **acaryotes**.

3. Les virions, parasites pathogènes de cellules de tous groupes

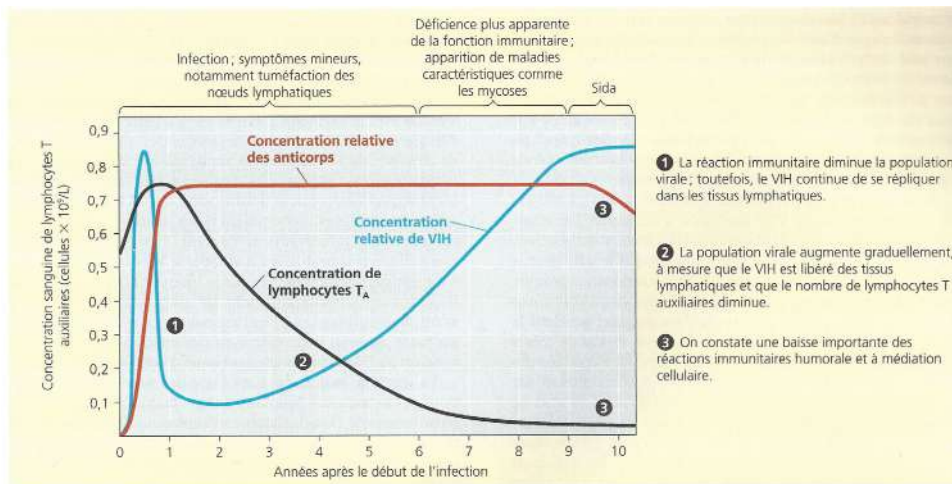
- Les **virus** parasitent **tous les types d'êtres vivants** (Archées, Eubactéries, 'plantes', Animaux...).

On a même découvert en 2008 des « **virus de virus** » ou **virophages**.

- Ils déclenchent généralement des **symptômes pathogènes** dans l'**organisme infecté**, pouvant finalement parfois conduire à la mort de l'individu :
 - Exemple du VMT : **déverdissement des cellules atteintes**, **modification de la structure des cellules** (donnant un **aspect gaufré** aux feuilles qui sont par ailleurs maculées de **taches blanchâtres**) (figure 81). *Notons que le virus se déplace par les sèves mais aussi entre cellules proches grâce aux plasmodesmes, ce qui lui permet d'infecter toute la plante.*
 - Exemple du VIH (figure 82) : **diminution progressive du nombre de lymphocytes T4** aboutissant finalement à un état de **système immunitaire déficient** qu'on nomme **SIDA** (*syndrome de l'immunodéficience acquise*) où se développent des **maladies opportunistes** (*pathologies dues à des agents infectieux qui prolifèrent face à des défenses immunitaires faibles*) qui finissent généralement par emporter le patient.
 - Exemple du SARS-Cov2 : ce **virus** a déclenché une **pandémie humaine mondiale** apparue fin 2019 et encore présente à ce jour. Le virus touche divers **types cellulaires** des **voies respiratoires** mais aussi parfois des **cellules digestives, nerveuses, immunitaires**...
 - Exemple du phage lambda : en cas de **cycle lysogénique** (voir plus loin), il y a **lyse** de la **Bactérie hôte** par **éclatement** lors de la **sortie du virus**.



▲ FIGURE 81. Tabac atteint par la mosaïque.
D'après Wikipédia (consultation novembre 2015).



▲ FIGURE 82. Effet du VIH sur le fonctionnement immunitaire humain.
D'après CAMPBELL & REECE (2004).

4. Les virus, entités évolutives à l'origine discutée [pour information]

a. Des entités génétiques qui subissent des mutations

- Comme les êtres vivants, les **virus évoluent** (les gènes des virus accumulent des **mutations**) et il est donc possible de **réaliser des phylogénies moléculaires** pour envisager leur **apparentement**.

b. Petit détour sur les origines du vivant (organismes cellulaires)

- De **nombreuses hypothèses** existent sur l'**origine du vivant** et il est souvent admis que le vivant proviendrait d'une **évolution chimique prébiotique (précédant l'état cellulaire)** ayant donné les **molécules organiques** qui se seraient peu à peu organisées.
- L'apparition de l'**information génétique** (vraisemblablement **sous forme ARN** mais il n'y a pas unanimité) est, pour la plupart des auteurs, clairement **antérieure** à l'apparition des **cellules**.

c. L'origine des virus : diversité des hypothèses possibles

- Un certain nombre d'**hypothèses** quant à la place des **virus** dans cette évolution est avancé. On peut en citer trois principales :
 - Les virus seraient contemporains des premières cellules et auraient émergé une seule fois** à l'époque où devaient foisonner de **nombreuses formes de vie « précellulaires »** dont la plupart n'ont pas persisté et ont aujourd'hui disparu : **virus et organismes actuels (ces derniers dérivant de LUCA) seraient ainsi les deux formes d'organisation biologiques ayant persisté et évolué jusqu'à nos jours.**
 - Les virus seraient préexistants au vivant.** De nombreuses **protéines** présentes à la fois chez les **organismes** et chez les **virus** auraient été transmises au vivant par **transferts horizontaux**, comme les protéines impliquées dans la **réplication de l'ADN** par exemple. Cette hypothèse est **difficile à défendre** pour nombre d'auteurs qui soulignent que **les virus ne sont pas des entités autonomes** et qu'il paraît difficile d'envisager des virus sans que les hôtes n'aient préalablement existé. *Il se pourrait toutefois que les premiers hôtes soient des formes de vie inconnues n'ayant pas laissé de descendance jusqu'à nos jours.*
 - Les virus seraient apparus de multiples fois (et continueraient d'apparaître ?), en même temps ou après les premières cellules, à partir de souches bactériennes ou eucaryotes (voire d'Archées)** en acquérant une **relative autonomie** à partir des génomes d'êtres vivants dont ils sont issus qui auraient connu une **réduction drastique de leur taille**. Certains auteurs assimilent les **gènes viraux** à des **séquences mobiles** dans les génomes d'organismes vivants (un peu à l'image des transposons) se propageant par **infection de cellules**, les **virions** en étant les **vecteurs**.

d. La difficulté de brancher les virus sur l'arbre du vivant

- S'il paraît **difficile de trancher entre ces trois positionnements** (qui demeurent hypothétiques et largement spéculatifs), il est toutefois bien établi aujourd'hui que de **nombreux gènes** sont **communs** entre **organismes vivants** et **entités virales**, ce qui démontre que de **nombreux transferts horizontaux** sont intervenus et interviennent au cours de l'évolution **entre les virus et les cellules qu'ils parasitent**. L'étude de ces **gènes viraux** et de leurs **homologues** chez les êtres vivants permet d'inclure les **virus** dans des **phylogénies** avec des **organismes vivants**. On obtient ainsi les **liens de parenté entre gènes** qui permettent de

positionner des **hypothèses de transferts horizontaux** sans toutefois que ces liens ne représentent nécessairement les liens des organismes et virus concernés.

- L'existence de **virus géants** (**encadré I**) semblant jeter un **pont entre monde viral et monde cellulaire**.

Certaines phylogénies tendent à montrer l'unité du monde viral et à le présenter comme une branche phylétique singulière alors que d'autres phylogénies suggéreraient que les virus sont un ensemble polyphylétique de lignées se branchant en de multiples endroits sur l'arbre du vivant (il y aurait alors eu convergence de l'acquisition de l'état viral qui peut alors être vu comme un état de vie régressé).

Encadré I Les virus géants ou girus (giant virus)

(Au-delà du programme : pour information)

➤ On connaît depuis quelques années des **virus géants** (parfois appelés **girus**, contraction de l'anglais *giant virus*) qui présentent un **nombre de gènes très important**, largement supérieur à ce qu'on trouve dans la plupart des autres virus : un virus géant est caractérisé par une **taille supérieure à 0,2 µm** et un génome formé de **plus de 300 000 pb** (là où les virus « traditionnels » ont un génome de taille souvent située autour de quelques milliers voire dizaines de milliers de pb maximum).

➤ Les plus fréquents sont les **grands virus nucléocytoplasmiques** (NCLDV pour *nucleocytoplasmic large DNA viruses*), en particulier ceux **associés aux 'amibes'** comme les *Mimivirus* ou encore les *Pandoravirus* récemment découverts en 2013 (il y aurait jusqu'à 2500 gènes au sein de ces derniers).

➤ C'est sur les girus qu'ont été basées la plupart des phylogénies essayant de positionner le monde viral par rapport au monde vivant.

➤ Il est intéressant de noter que si les **virus amibiens** sont ceux avec les **plus gros génomes viraux connus**, les **'amibes'** sont également de leur côté les **organismes vivants chez lesquels on a trouvé les génomes de plus grande taille**.

B. Infection, parasitisme et reproduction des virus

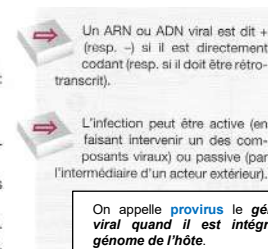
1. Fonctionnement des virus : une vue d'ensemble

- Voir **tableau XI**.

✓ **TABEAU XI. Notion de cycle viral et comparaison de celui du VMT, du VIH et d'un phage.**
D'après CAMPBELL & REECE (2004).

Quelles sont les principales étapes du cycle viral ?

- Le cycle viral peut être décomposé en quatre à cinq étapes :
 - l'infection ;
 - l'intégration éventuelle dans le génome de la cellule hôte ;
 - l'expression de l'information génétique virale grâce à la machinerie cellulaire de la cellule hôte ;
 - la reconstitution de virions (matériel génétique viral, capsid, et protéines éventuelles) ;
 - la propagation du virus (par sortie de la cellule) et l'infection d'autres cellules.
- Chaque virus présente des variantes de tout ou partie de ces étapes (Tableau 5).



	VMT	VIH	Phage	SARS-Cov-2
Infection	Passive par des insectes piqueurs.	Active par reconnaissance de protéines membranaires (CD4 des lymphocytes) et fusion des deux membranes.	Active par destruction de la paroi (lysozyme) et injection de l'ADN viral sous forme linéaire.	Reconnaissance des récepteurs ACE2 et endocytose des virions
Intégration dans le génome	Non.	Rétrotranscription de l'ARN viral en ADN _{cc} puis intégration irréversible à l'aide de l'intégrase virale = provirus.	Intégration réversible, uniquement dans la voie lysogénique = prophage.	Non.
Expression	Traduction précoce (répliques brin + et brin -). Réplication de l'ARN viral. Traduction tardive (capsomères, translocase).	Transcription et traduction du provirus.	Transcription précoce (méthylases, ADNases, régulation des gènes tardifs). Transcription tardive (capsid et ADN polymérase virale).	Traduction précoce (répliques brin + et brin -). Réplication de l'ARN viral. Traduction parallèle (passage par le REG et le GOLGI)
Reconstitution des virions	Autoassemblage. Jusqu'à 60 millions de virus par cellule.	Association des protéines et de l'ARN viraux. Insertion de protéines virales dans la membrane plasmique.	Intervention de protéines d'assemblage.	Assemblage au niveau du REG et du GOLGI → vésicules d'exocytose.
Propagation	De proche en proche par les plasmodesmes.	Bourgeonnement des virus à partir de la membrane plasmique. Éclatement de la cellule.	Lyse de la bactérie par choc osmotique (lysozyme).	Exocytose dans les voies respiratoires et transmission aéroportée.

Cycle viral.

2. Modalités de pénétration des virus dans les cellules hôtes (infection)

a. Première possibilité : pénétration passive (cas du VMT)

- Le **VMT** ne possède **pas d'enzymes capables de dégrader la paroi cellulaire** : il se propage par des **blessures**.
- Les **phytophages** (comme les Pucerons) peuvent aussi servir de **vecteurs**.

b. Deuxième possibilité : fixation sur la cellule hôte (cas du SARS-Cov2 et des phages)

- Cas du phage : **fixation** sur une protéine spécifique de l'hôte, déclenchant l'**injection de l'ADN viral** dans la cellule.
- Cas du VIH : présence de la **protéine gp120** de l'enveloppe qui se fixe au **marqueur CD4** des lymphocytes T *helper* (auxiliaires) provoquant la **fusion** des

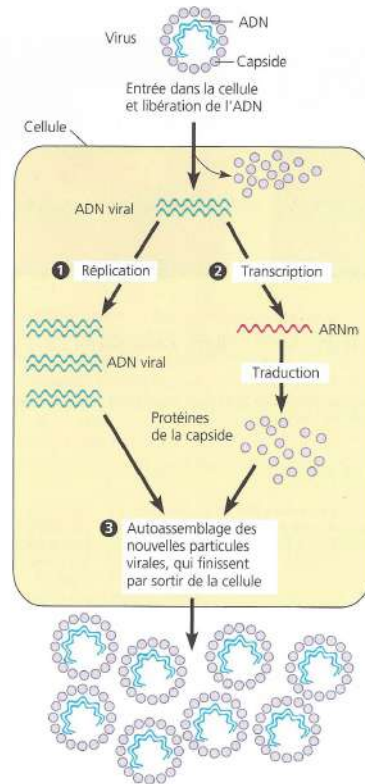
membranes du virus et de la cellule. La capside pénètre dans la cellule et s'y désagrège.

- Cas du SARS-Cov2 : fixation de la protéine spike sur une protéine spécifique de l'hôte (ex. ACE2, *angiotensin-converting enzyme 2*, enzyme de conversion de l'angiotensine 2), déclenchant l'endocytose des virions.

3. Les cycles viraux des exemples étudiés en détail

- On peut appeler **cycle viral** l'ensemble des étapes allant de l'infection d'une cellule à la production de nouveaux virions (figure 83).

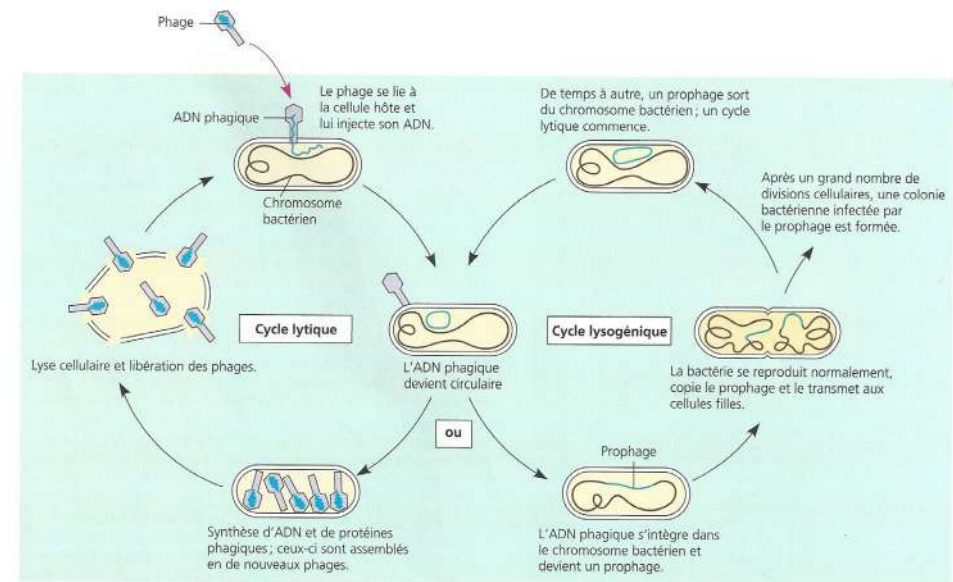
a. Vue d'ensemble



Représentation simplifiée du cycle de répllication d'un virus. Un virus est un parasite intracellulaire obligatoire qui se multiplie grâce aux structures et aux fonctions métaboliques de la cellule hôte. Dans cet exemple de cycle de répllication viral, le plus simple de tous, le parasite est un virus à ADN dont la capsid ne comporte qu'une seule sorte de protéine. 1. Après avoir pénétré dans la cellule, l'ADN viral utilise des nucléotides et des enzymes de l'hôte pour se répliquer. 2. Il produit les protéines de sa propre capsid (suivant la transcription et la traduction) à l'aide d'autres matériaux et d'autres structures de l'hôte. 3. L'ADN viral nouvellement formé et les protéines de la capsid s'assemblent ensuite pour former de nouvelles particules virales, qui quittent la cellule.

▲ FIGURE 83. **Cycle viral.** D'après CAMPBELL & REECE (2004).

b. Les deux cycles des Bactériophages (cycle lytique, cycle lysogénique)



Le cycle lytique et le cycle lysogénique chez un phage tempéré, le Phage λ . Après avoir pénétré dans la cellule bactérienne, l'ADN d'un phage λ peut soit s'intégrer au chromosome bactérien (cycle lysogénique), soit

commander immédiatement la production d'un grand nombre de phages λ (cycle lytique). Dans la plupart des cas, il suit le cycle lytique. Cependant, une fois le cycle lysogénique amorcé, le prophage peut demeurer dans le chromosome de la cellule

hôte pendant de nombreuses générations. Le Phage λ n'a qu'une seule fibre caudale, qui est courte et qui n'apparaît pas sur ce schéma.

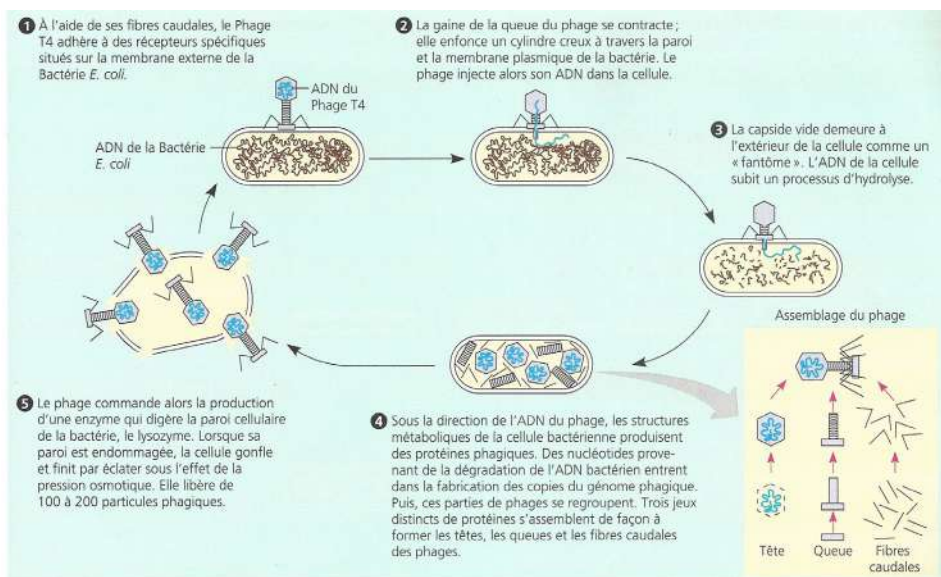
▲ FIGURE 84. **Les deux cycles des Bactériophages : cas du phage λ .**

D'après CAMPBELL & REECE (2004).

a. Le cycle lytique, un cycle sans provirus qui aboutit à la libération de nouveaux phages par lyse bactérienne

- Le **cycle lytique** (figures 84-85) est le **cycle virulent** du phage (= **cycle conduisant à la multiplication du virus et aboutissant à la lyse de la cellule**) :
 - **Infection** : elle est permise par :
 - l'**accrochage du phage sur la bactérie** au moyen de ses **fibres caudales**... mais certains bactériophages n'en ont pas, voire n'ont même pas de queue !
 - et l'**injection du génome dans la Bactérie** : la **gaine caudale** serait **contractile** chez nombre de Bactériophages (cas envisagé figure 9, mais cela n'explique pas ce qui se passe chez les phages sans queue) ; l'intervention d'**enzymes digérant la paroi bactérien** – les **lysozymes** – est en outre citée par de nombreux auteurs.
 - **Circularisation du génome** grâce à une **ligase** (enzyme souvent apportée par le phage). Le **génome viral cyclisé** est alors traité comme un **plasmide** par l'hôte.
 - **Intégration du génome viral dans le génome de l'hôte** : pas dans ce cycle.

- **Transcription précoce** : l'ADN phagique est transcrit par les ARN polymérases de l'hôte, ce qui permet la production de protéines (enzymes notamment) indispensables à la **réplication** de l'ADN phagique.
- **Transcription tardive** : les multiples copies de l'ADN phagique sont ensuite à nouveau transcrites, ce qui permet notamment la production de protéines structurales du virus.
- **Reconstitution des virions** : il y a assemblage des particules virales avec encapsidation de l'ADN phagique.
- **Lyse cellulaire et libération des virions** : la perturbation de l'activité génétique de l'hôte par la présence du virus entraînerait un moindre renouvellement des constituants moléculaires de la cellule (ce qu'on appelle le **turn-over moléculaire**), y compris au niveau de la membrane et de la paroi, ce qui finirait par provoquer la lyse cellulaire, libérant ainsi les virions. Notons que certains auteurs avancent là encore l'intervention du **lysozyme** qui digérerait cette fois de l'intérieur les limites bactériennes (cas de la figure 9).



Cycle lytique du Phage T4.
Le Phage T4 possède environ 100 gènes, qui sont transcrits et traduits par les structures d'une cellule hôte. Une fois que celle-ci est infectée, l'un des premiers gènes du phage à être traduit code pour

une enzyme qui découpe l'ADN de la cellule hôte (étape 3). L'ADN du phage n'est pas découpé, parce qu'il contient une forme modifiée de cytosine que l'enzyme ne reconnaît pas. L'ensemble du cycle lytique – à partir du contact entre le phage et la

surface de la bactérie jusqu'à la lyse de la cellule – ne dure que de 20 à 30 minutes à 37 °C.

▲ FIGURE 85. **Focus sur le cycle lytique : cas du phage T4.** D'après CAMPBELL & REECE (2004).

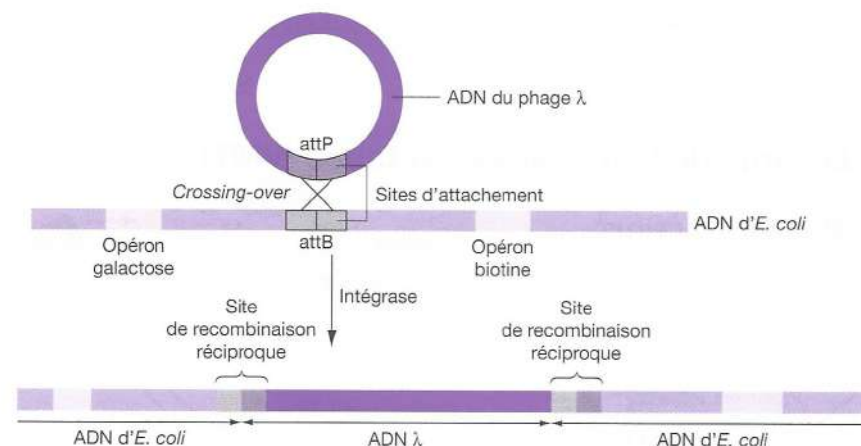
β. Le cycle lysogénique, un cycle latent avec provirus (prophage)

α. Un cycle qui coexiste avec le cycle lytique

- Le **cycle lysogénique** (figures 84) est un **cycle latent du virus qui se met en place chez certains phages sous des conditions encore mal élucidées**. Néanmoins, on a mis en évidence l'existence de gènes qui activent ou inhibent l'un ou l'autre des cycles.
- Les **phages qui présentent ce cycle** peuvent alors être appelés **phages tempérés**.
- Il arrive que certains phages n'entrent plus jamais en cycle lytique, devenant ainsi avirulents.
- Notons que les phages sont d'importants vecteurs de transferts géniques entre bactéries (on parle de **transduction génétique**) et cette propriété peut être utilisée en génie génétique.
- Le **passage d'un cycle à l'autre** s'appelle **commutation**.

β. Modalités du cycle lysogénique

- Le **cycle lysogénique** (figure 84) comprend :
 - **Infection** : comme dans le cycle précédent.
 - **Circularisation du génome** comme dans le cycle précédent.
 - **Transcription précoce** : l'ADN phagique est transcrit par les ARN polymérases de l'hôte, ce qui permet la production d'un **répresseur** du cycle lytique et d'une **intégrase**.
 - **Intégration du génome viral dans le génome de l'hôte** : l'intégrase permet l'intégration du **génome viral** dans le **génome de l'hôte**. Le génome viral est donc un **provirus** (qu'on peut ici appeler **prophage**). L'intégration se fait au moyen de **sites d'attachement** présents dans le génome viral et le génome bactérien (figure 86).



Intégration du phage λ dans le chromosome d'E. coli

Le mécanisme d'intégration est une recombinaison réciproque (crossing-over) qui s'effectue entre un site d'attachement spécifique présent sur l'ADN circulaire de λ et une séquence correspondante située entre les opérons galactose et biotine du chromosome bactérien. Par le mécanisme inverse, le prophage λ quitte le chromosome et se réplique, le cycle devient lytique.

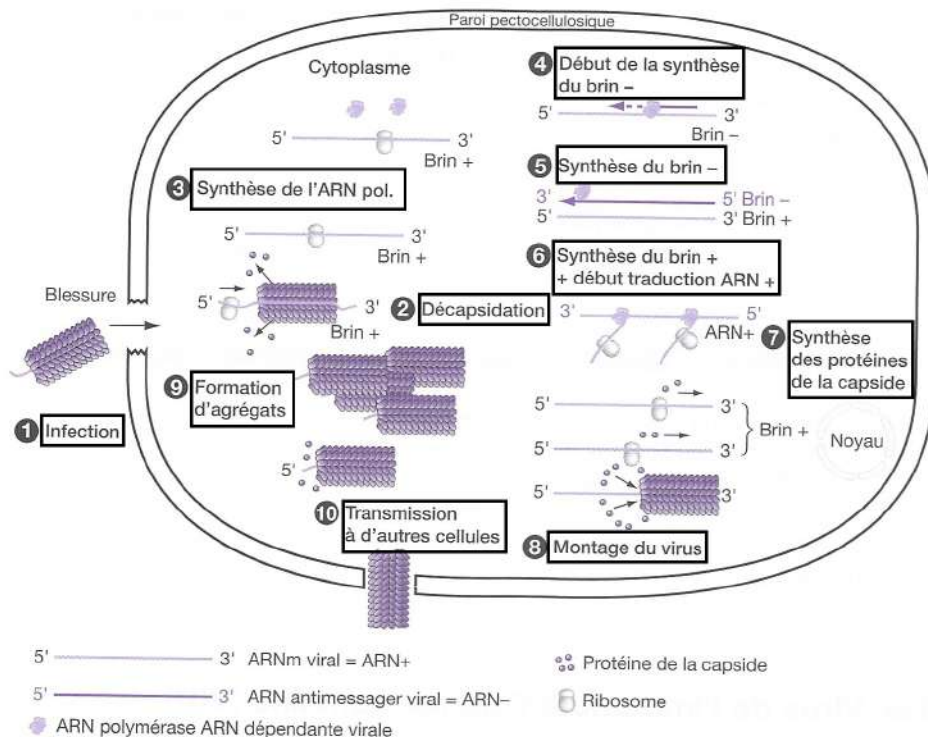
▲ FIGURE 86. **Intégration du génome phagique dans la Bactérie hôte.** D'après BREUIL (2007).

- **Inactivité du phage et multiplication bactérienne** : pendant de nombreux cycles, la **bactérie** se multiplie par **divisions cellulaires** sans que le phage n'entre en activité.
- **Sortie du cycle lysogénique et entrée en phase lytique (commutation)** : il arrive qu'une bactérie sorte du **cycle lysogénique** et entre en **cycle lytique**.

La **lysogénie** est le **fait, pour une bactérie, de porter dans son génome le génome d'un Bactériophage** (la bactérie est qualifiée de « **lysogène** »).

c. Cycle du Virus de la mosaïque du tabac (VMT)

- Le **cycle du Virus de la mosaïque du tabac (VMT)** se déroule sur 1 à 3 jours et comprend les étapes suivantes (figure 87-88) :

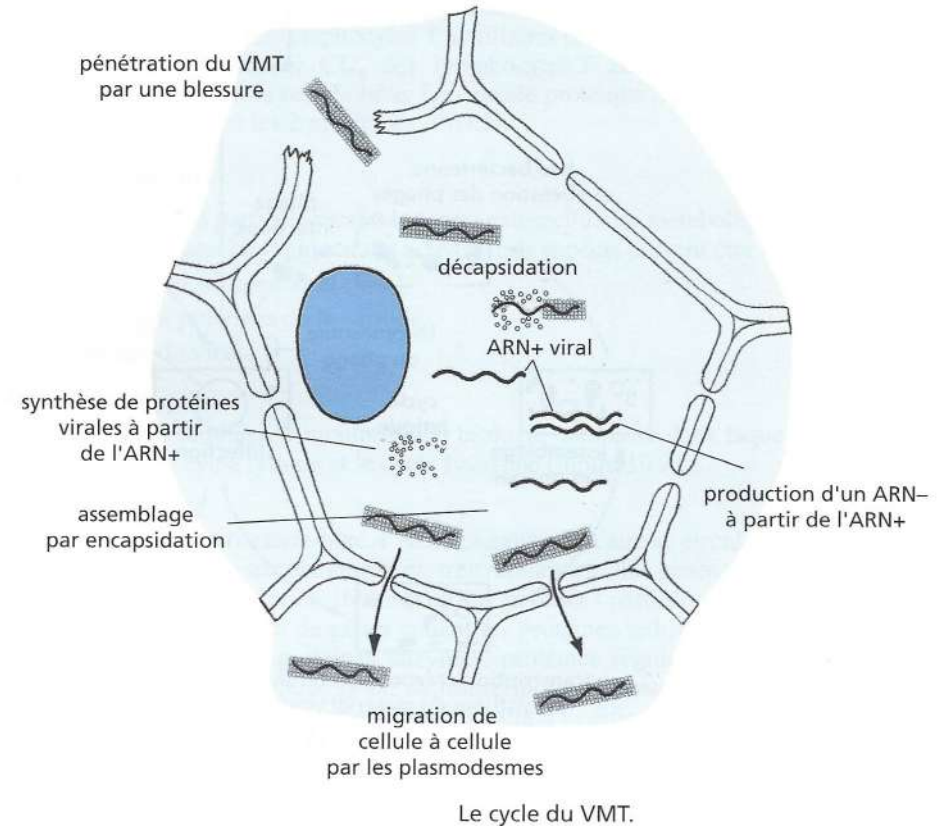


Cycle du VMT

▲ FIGURE 87. **Cycle du VMT : version longue.** D'après BREUIL (2007), corrigé.

- **Infection** : elle est permise par une **blessure** et/ou un **vecteur phytophage**.
- **Intégration du génome viral dans le génome de l'hôte** : *inexistante*.
- **Décapsidation et traduction** : l'**ARN viral** (noté +) est **libéré de la capsid** et **traduit par les ribosomes** de l'hôte, ce qui permet la production d'une **ARN polymérase virale**.

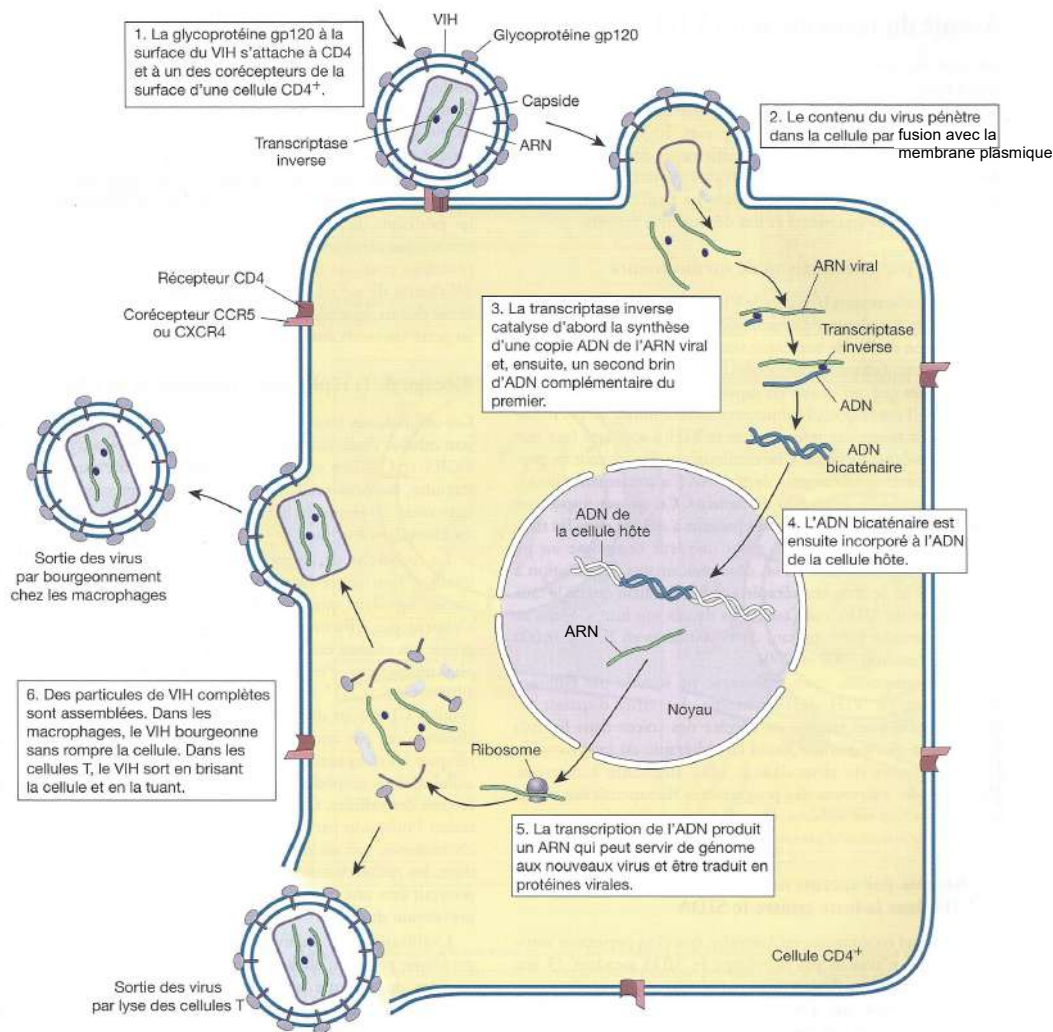
- **Réplication de l'ARN viral** : l'**ARN +** est alors répliqué par l'**ARN polymérase** en une multitude d'**ARN complémentaires** (notés **ARN -**). Ces **ARN -** servent ensuite de **matrice** à la synthèse d'**ARN +** par ces mêmes **ARN polymérases**.
- **Synthèse des protéines phagiques** : les **ARN +** sont **traduits** par les **ribosomes** de l'hôte en **protéines virales**, notamment celles de la capsid.
- **Reconstitution des virions** : il y a **assemblage des particules virales** avec **encapsidation de l'ADN viral**.
- **Libération des virions** : les **virus** se propagent par les **sèves** ou par les **plasmodesmes** à d'autres cellules.



▲ FIGURE 88. **Cycle du VMT : version courte.** D'après PEYCRU *et al.* (2010a).

d. Cycle du virus de l'immunodéficience humaine (VIH) [pour information]

- Le **VIH** est un **rétrovirus**, c'est-à-dire un **virus à ARN** présentant une **transcriptase inverse** (ou **rétrotranscriptase**) permettant la **rétrotranscription de l'ARN viral en ADN** pouvant **s'intégrer au génome de l'hôte**. Son cycle est résumé très simplement par la **figure 89**.
- Notons que ces virus pourraient avoir un **rôle évolutif, ajoutant des gènes aux génomes**.
- Cette propriété est utilisée en **génie génétique**.

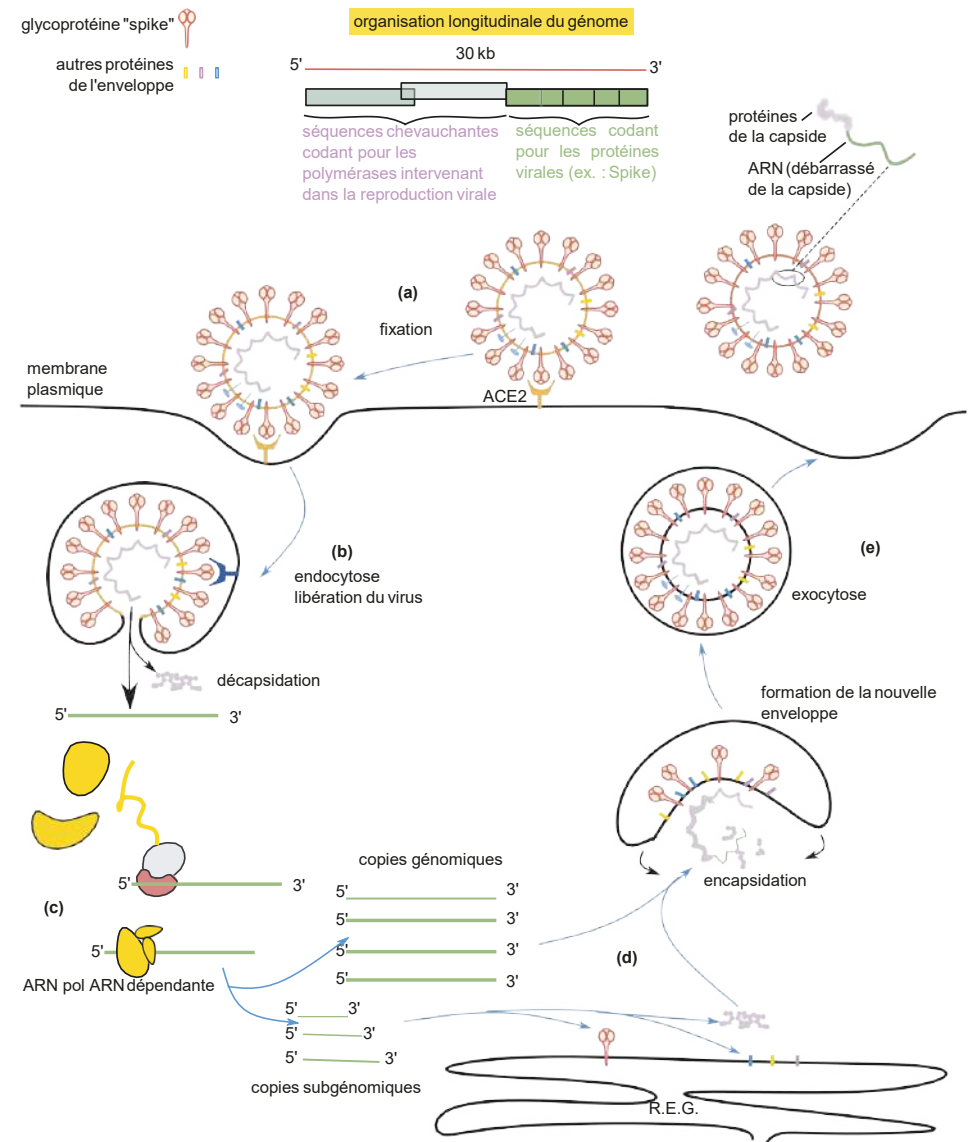


Cycle d'infection par le VIH. Le cycle débute et se clôture avec des particules de VIH libres dans le sang humain. Ces virus libres infectent les lymphocytes possédant des récepteurs CD4, les cellules appelées CD4⁺.

▲ **FIGURE 89. Cycle du VIH.** D'après RAVEN *et al.* (2007), corrigé.

*Ce schéma est très indicatif et assez vieux... Il semblerait que la **décapsidation** n'intervienne que tardivement et que la **rétrotranscription** se fasse dans la capside.*

e. Cycle du SARS-Cov-2



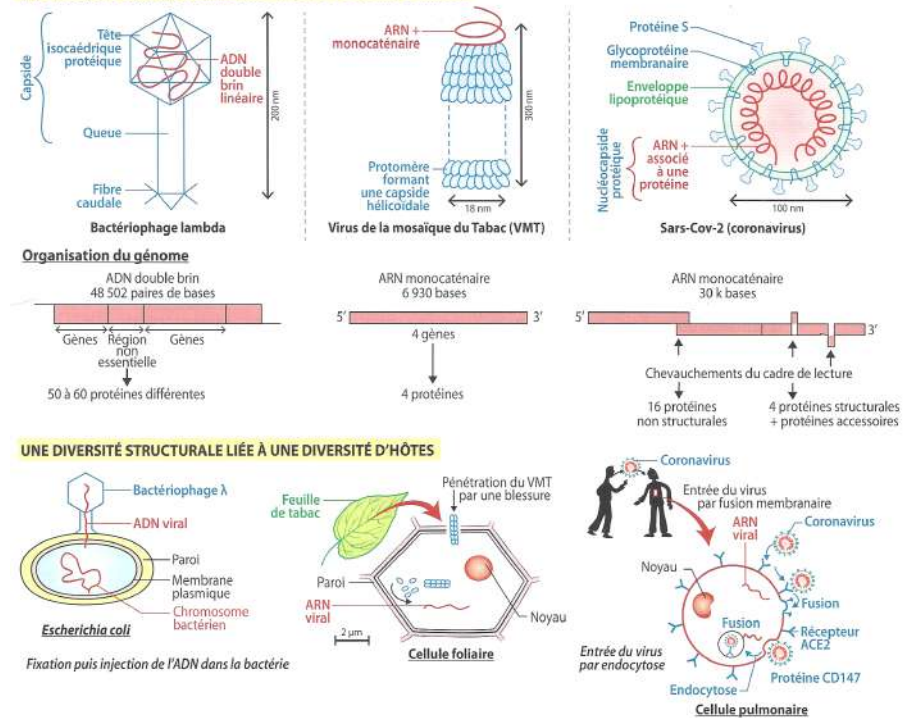
▲ **FIGURE 90. Cycle du SARS-Cov-2.** D'après PERRIER, BEAUX *et al.* (2021).

- Le **cycle viral** du **SARS-Cov2** (figure 90) peut être résumé comme suit :
 - **Infection** : elle est permise par la **fixation de la protéine S (spike)** sur des **récepteurs** de la future **cellule hôte**, ce qui déclenche l'**endocytose des virions**.
 - **Intégration du génome viral dans le génome de l'hôte** : *inexistante*.
 - **Décapsidation et traduction** : l'**ARN viral (noté +)** est **libéré de la capsid** et **traduit par les ribosomes** de l'hôte, ce qui permet la production d'une **ARN polymérase virale**.
 - **Réplication de l'ARN viral** : l'**ARN +** est alors répliqué par l'**ARN polymérase** en une multitude d'**ARN complémentaires (notés ARN -)**. Ces ARN - servent ensuite de **matrice** à la synthèse d'**ARN +** par ces mêmes **ARN polymérases**.

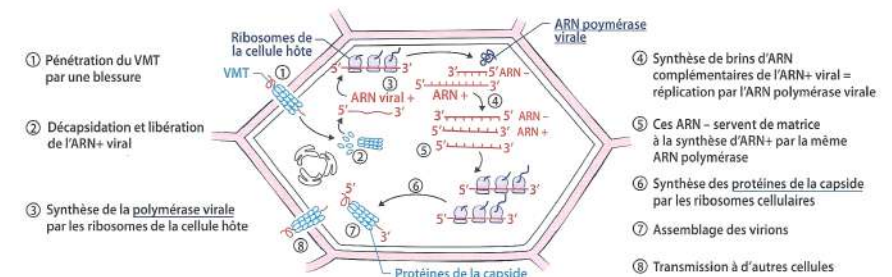
Deux types d'ARN+ sont produits : des **ARN+ complets** (dits **génomiques**) prêts à être encapsidés, et des **ARN+ tronqués** (dits **subgénomiques**) qui sont traduits en protéines virales par les **ribosomes** de la **cellule hôte** avec adressage dans le REG.

- **Synthèse des protéines phagiques** : les **ARN +** sont traduits par les **ribosomes** de l'hôte en **protéines virales**, notamment celles de la **capside**, de l'**enveloppe**... adressées de manière **cotranscriptionnelle** dans le REG où elles sont intégrées à la **membrane**, sauf les **protéines de nucléocapside** qui restent **cytosoliques**.
- **Reconstitution des virions** : il y a **assemblage des particules virales** dans l'appareil de GOLGI avec **encapsidation de l'ADN viral**.
- **Libération des virions** : les **virions**, présents dans des **vésicules d'exocytose**, sont libérés par **exocytose**.

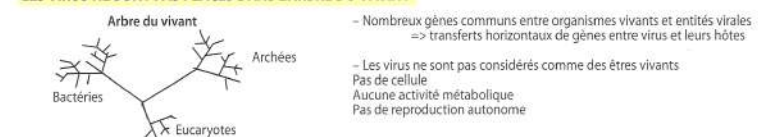
LES VIRUS, DES ENTITÉS NUCLÉOPROTÉIQUES À GÉNOME COMPACT



LES VIRUS, PARASITES INTRACELLULAIRES OBLIGATOIRES, DÉTOURNENT LA MACHINERIE D'EXPRESSION DE L'HÔTE



LES VIRUS NE SONT PAS PLACÉS DANS L'ARBRE DU VIVANT

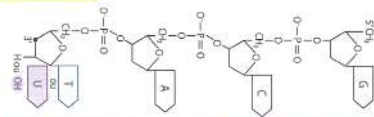


▲ FIGURE 91. **Bilan sur les virus.** D'après SAINTPIERRE, BORDI *et al.* (2021).

Bilan (adapté du programme)

- ✓ Les **virus** ou **particules virales** sont des **entités nucléoprotéiques** comprenant un **acide nucléique** (sous forme d'**ADN** ou d'**ARN**) constituant le **génome viral**, et des **protéines**.
- ✓ On distingue des **protéines à rôle structural**, formant la **capside**, et parfois des **protéines à rôle enzymatique**.
- ✓ Les **virus** sont **très divers** et possèdent parfois une **enveloppe lipoprotéique**.

UN GÉNOME D'ACIDE NUCLÉIQUE

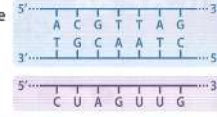


LOCALISATION

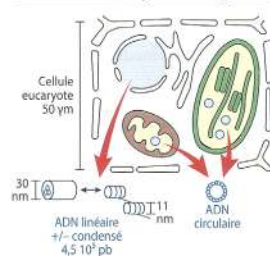
CONTENU INFORMATIF

RÉPLICATION ET TRANSMISSION

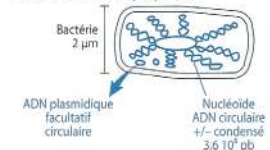
Un génome d'ADN ou d'ARN



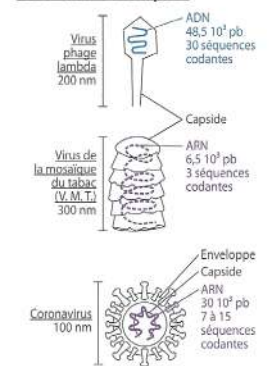
Génome dans le noyau et les organites



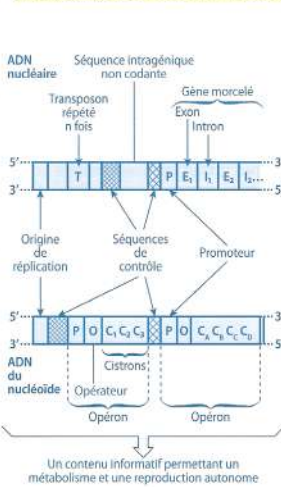
Génome dans le cytoplasme



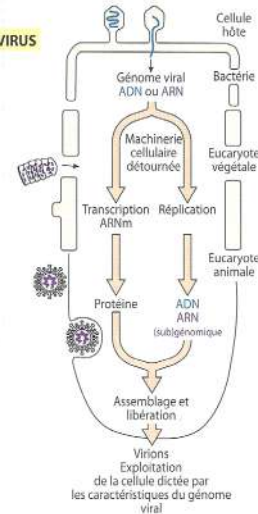
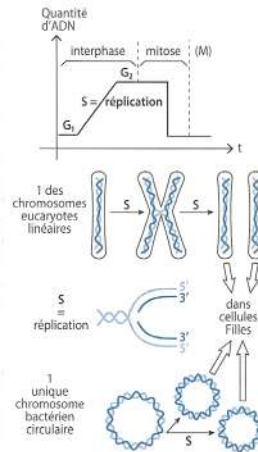
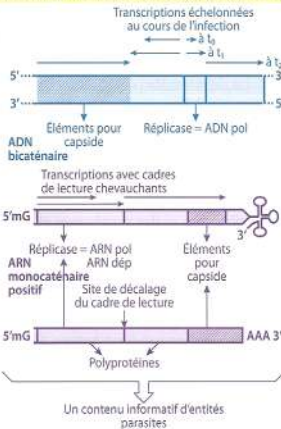
Génome dans une capsid



UN GÉNOME D'ADN POUR LES CELLULES



UN GÉNOME DE DIFFÉRENTES NATURES DANS LES VIRUS



Il est conseillé de maîtriser les **grandes lignes du plan**

Le **plan** ne doit pas être perçu comme un carcan figé, ou comme un modèle de plan de dissertation à ré-utiliser en devoir, mais bien comme un **outil d'apprentissage et de structuration des concepts importants**. Vous pouvez en recopier les grandes lignes ou annexer le plan du polycopié directement.

Il est conseillé de réaliser un **lexique des principales définitions**.

Il est conseillé de reproduire les **schémas (et tableaux) majeurs** :

Liste indicative.

- Support de l'IG

[° **Expériences** de GRIFFITH (1928), AVERY MACLEOD & MCCARTY (1944), HERSHEY & CHASE (1952) ?]

° Nucléotide

° ADN

° ARN (notamment ARNt)

- Techniques

° **Électrophorèse**

° **Séquençage** (méthode de **SANGER** : autoradiographie / fluorescence)

° **Alignement** de séquences

° **PCR** et ses variantes (qPCR, RT-PCR, qRT-PCR)

° **Enzymes de restriction** (coupures en baïonnette et bouts collants)

° **Hybridation in situ**, **FISH**

° **Southern Blot**, **Northern Blot**, **Western Blot**

° **Puces à ADN**

° **Avantages / inconvénients** de la **transgénèse**

° Technique de l'**ADNc**

[° **Techniques variées** de **transgénèse** à avoir compris]

° **CRISPR Cas9**

° **NEHJ**, **HDR**

[° Principe du **criblage**]

[° Techniques de **mutagenèse** à avoir compris]

° **Knock-out**, **knock-in**

° **Gène rapporteur**

° **Retard sur gel** (EMSA)

- **Génome bactérien**

° **Tableau comparatif** Bactéries / Eucaryotes

° **Génome bactérien**, **chromosome bactérien**

° **Opéron** (exemple opéron lactose)

- **Génome eucaryote**

° **Histones** et **fibre nucléosomique**

° Rôle de **H1**

° **Niveaux de condensation** de l'ADN

[° Transition **eu-/hétérochromatine**]

° **Chromosome double condensé**

° **Hybridation gène / ARNm**

° **Gène eucaryote**

° Courbes de **renaturation** de l'ADN

° **Tableau de la diversité des séquences**

▲ FIGURE 92. **Bilan sur les génomes**. D'après SAINTPIERRE, BORDI *et al.* (2021).

- Virus

- ° Tableau du **génome des trois virus** au programme
- ° **Virions** au programme
- ° **Cycles des virus** au programme

Vous devez en outre **savoir / pouvoir** exploiter des **données expérimentales** faisant appel aux **techniques au programme**.
(!) N'oubliez pas d'apprendre quelques **données chiffrées** !

Références

- ADAMS, G. (2020). A beginner's guide to RT-PCR, qPCR and RT-qPCR. *The Biochemist*, **42** (3) : 1-6. <http://doi.org/10.1042/BIO20200034>
- ALBERTS, B., A. JOHNSON, J. LEWIS, M. RAFF, K. ROBERTS & P. WALTER (2004). *Biologie moléculaire de la cellule. Quatrième édition*. Flammarion, Paris. Traduction de la quatrième édition américaine (2002) par F. LE SUEUR-ALMOSNI. Première édition américaine 1983 (1986 1^{re} édition française).
- ALBERTS, B., A. JOHNSON, J. LEWIS, M. RAFF, K. ROBERTS & P. WALTER (2008). *Molecular Biology of the Cell. Fifth Edition*. 1st edition 1983. Garland Science, New York (NY), USA & Abingdon, UK.
- ALBERTS, B., A. JOHNSON, J. LEWIS, D. MORGAN, M. RAFF, K. ROBERTS & P. WALTER (2015). *Molecular Biology of the Cell. Sixth Edition*. 1st edition 1983. Garland Science, New York (NY), USA & Abingdon, UK.
- ALBERTS, B., D. BRAY, K. HOPKIN, A. JOHNSON, J. LEWIS, M. RAFF, K. ROBERTS & P. WALTER (2015). *Essential Cell Biology. Fourth Edition*. 1st edition 1998. Garland Science, New York (NY), USA & Abingdon, UK.
- BALASSIANO, I. T. M. DO CARMO DE FREIRE BASTOS, D. JANNUZZI MADUREIRA, I. GRIPP DA SILVA, Â. CORRÊA DE FREITAS-ALMEIDA & S. SOARES DE OLIVEIRA (2007). The involvement of *tetA* and *tetE* tetracycline resistance genes in plasmid. *Memorias do Instituto Oswaldo Cruz*, **102** (7): 861-866. <https://doi.org/10.1590/S0074-02762007005000121>
- and chromosomal resistance of Aeromonas in Brazilian strains
- BERTHET, J. (2006). *Dictionnaire de Biologie*. De Boeck Université, Bruxelles (Belgique).
- BOUJARD, D. (dir.). B. ANSELME, C. CULLIN & CÉLINE RAGUÉNÉS-NICOL (2015). *Biologie cellulaire et moléculaire. Tout le cours en fiches. Licence. PACES. CAPES. 2^e édition* (1^{re} édition 2012), Dunod, Paris.
- BREUIL, M. (2007). *Biologie 1^{re} année BCPST-véto*. Tec & Doc, Lavoisier, Paris.
- BREUIL, M. (2009). *Biologie 2^e année BCPST-véto*. Tec & Doc, Lavoisier, Paris.
- CALLEN, J.-C. (2005). *Biologie cellulaire. Des molécules aux organismes*. Dunod, Paris, 2^e édition (1^{re} édition 1999).
- CAMPBELL, N. A. & J. B. REECE (2004). *Biologie*. De Boeck Université, Bruxelles, 2^e édition (1^{re} édition 1995).
- [CAMPBELL, N. A.], J. B. REECE, L. A. URY, M. L. CAIN, S. A. WASSERMAN, P. V. MINORSKY, R. B. JACKSON (2012). *Campbell Biologie*. Adaptation française J. FAUCHER & R. LACHAÎNE. Pearson, Paris (4e édition).
- CLARK, D. P., N. J. PAZDERNIK & M. R. MCGEEHEE (2019). *Molecular Biology*. Third edition. Academic Press, Cambridge (MA), USA, Elsevier, Amsterdam.
- COOPER, G. M. (2019). *Cell. A Molecular Approach*. 8th edition, Sinauer / Oxford University Press, Oxford (GB).
- DAUTEL, O. (dir.), A. PROUST, M. ALGRAIN, C. BORDI, A. HELME-GUIZON, F. SAINTPIERRE, M. VABRE & C. BOGGIO (2017). *Biologie Géologie BCPST 1^{re} année*. Vuibert, Paris.
- DAUTEL, O. (dir.), C. BORDI, F. SAINTPIERRE, M. ALGRAIN-PITAVY, M. QUERTINIEZ, A. PROUST, M. VABRE A. HELME-GUIZON & B. MOLLIER (2019). *Biologie Géologie BCPST 2^e année*. Vuibert, Paris.
- DAUTEL, O. (dir.), M. ALGRAIN-PITAVY, C. BORDI, A. HELME-GUIZON, B. MOLLIER, A. PROUST, M. QUERTINIEZ, F. SAINTPIERRE & M. VABRE (2021). *Prépas scientifiques BCPST 1^{re} année. Biologie Géologie. Tout-en-un*. Vuibert, Paris.
- DENÈUD, J., T. FERROIR, O. GUIPPONI, H. MOREAU, M. PAULHIAC-PISON, M.-L. PONS & F. TEJEDOR (2011). *Biologie-Géologie BCPST-véto 2^e année*. Tec & Doc, Lavoisier, Paris.
- DENÈUD, J., C. GODINOT, O. GUIPPONI, H. MOREAU, M. PAULHIAC-PISON & F. TEJEDOR (2013). *Biologie-Géologie BCPST-véto 1^{re} année*. Tec & Doc, Lavoisier, Paris.
- DENÈUD, J., C. GODINOT, O. GUIPPONI, H. MOREAU, M. PAULHIAC-PISON, M.-L. PONS & F. TEJEDOR (2014). *Biologie-Géologie BCPST-véto 2^e année*. Tec & Doc, Lavoisier, Paris.
- GODINOT, C., H. MOREAU, M. PAULHIAC-PISON & F. TEJEDOR (2010). *Biologie-Géologie 1^{re} année BCPST-véto*. Tec & Doc, Lavoisier, Paris.
- HARRY, M. (2008). *Génétique moléculaire et évolutive*. Maloine, Paris, 2^e édition (1^{re} édition 2001).
- HARTL, D. L. & E. W. JONES (2003). *Génétique. Les grands principes*. Traduction E. DEQUIER, S. DUHARCOURT, D. JUTIER, A. LE ROUZIC, G. PAHLAVAN & N. SERRANO. Dunod, Paris, 3^e édition.
- LAFON, C. (2003). *La biologie autrement. 100 questions de synthèse*. Ellipses, Paris.

- LATRUFFE, N. (dir.), F. BLEICHER-BARDETTI, B. DUCLOS & J. VAMECQ (2014). *Biochimie. Tout le cours en fiches. Licence. PACES-UE1. CAPES*. Dunod, Paris.
- LAURENT, F.-X., G. VIBRAC, A. RUBIO, M.-T. THÉVENOT & L. PÈNE (2017). Les nouvelles technologies d'analyses ADN au service des enquêtes judiciaires. *Médecine/Sciences*, **33** (11) : 971-978. <http://doi.org/10.1051/medsci/20173311014>
- LE GUYADER, H. (dir.) (1998). *L'Évolution*. Belin-Pour la Science, Paris.
- LELIÈVRE, É., J. DENÈUD, J. ROQUES, É. HAMARD-PÉRON & M. AIRAUD (2018). *Biologie*. Dunod, Paris.
- LODISH, H., A. BERK, C. A. KAISER, M. KRIEGER, A. BRETSCHER, H. PLOEGH & A. AMON (2014). *Biologie moléculaire de la cellule*. Traduction P. L. MASSON & C. SANLAVILLE, 4^e édition, De Boeck, Louvain-la-Neuve (B).
- MEYER, S., C. REEB & R. BOSDEVEIX (2008). *Botanique. Biologie et physiologie végétales*. Maloine, Paris, 2^e édition (1^{re} édition 2004).
- MORÈRE, J.-L., R. PUJOL (coord.), J.-C. CALLEN, L. CHESNOY, J.-P. DUPONT, A.-M. GIBERT-TANGAPREGASSOM, G. RICOU, N. TOUZET (dir.) et collaborateurs (2003). *Dictionnaire raisonné de Biologie*. Frison-Roche, Paris.
- PERRIER, C. & J.-F. BEAUX (dir.), A. BOUFFIER, L. BOUGEOIS, P. CARRÈRE, T. DARRIBÈRE, J. DÉMARET-NICOLAS, A. EMOND, S. MAURY, O. MONNIER, T. SOUBAYA, A. VERGNAUD & A. WOEHRLÉ (2021). *Biologie-Géologie BCPST 1. Tout-en-un*. Dunod, Malakoff (F).
- PETIT, J.-M. & R. JULIEN (2007). *Mini-manuel de Génétique*. Dunod, Paris.
- PETIT, J.-M., S. ARICO & R. JULIEN, 2013. *Mini manuel de Biologie cellulaire*. Dunod, Paris, 3^e édition (1^{re} édition 2007).
- PEYCRU, P. (dir.), J.-F. FOGELGESANG, D. GRANDPERRIN, B. AUGÈRE, J.-C. BAEHR, C. PERRIER, J.-M. DUPIN & C. VAN DER REST (2010a). *Biologie tout-en-un BCPST 1^{re} année*. Dunod, Paris, 2^e édition (2009), réimpression corrigée (2010) (1^{re} édition 2006).
- PEYCRU, P. (dir.), J.-C. BAEHR, F. CARIOU, D. GRANDPERRIN, C. PERRIER, J.-F. FOGELGESANG & J.-M. DUPIN (2010b). *Biologie tout-en-un BCPST 2^e année*. Dunod, Paris, 2^e édition (1^{re} édition 2007).
- PEYCRU, P., D. GRANDPERRIN, C. PERRIER (dir.), B. AUGÈRE, T. DARRIBÈRE, J.-M. DUPIN, C. ESCUYER J.-F. FOGELGESANG, & C. VAN DER REST (2013). *Biologie tout-en-un BCPST 1^{re} année*. Dunod, Paris, 3^e édition (1^{re} édition 2006).
- PEYCRU, P., D. GRANDPERRIN, C. PERRIER (dir.), B. AUGÈRE, J.-F. BEAUX, F. CARIOU, P. CARRÈRE, T. DARRIBÈRE, J.-M. DUPIN, C. ESCUYER, J.-F. FOGELGESANG, S. MAURY, É. QUÉINNEC, E. SALGUEIRO & C. VAN DER REST (2014). *Biologie tout-en-un BCPST 2^e année*. Dunod, Paris, 3^e édition (1^{re} édition 2007).
- PEYCRU, P., D. GRANDPERRIN, C. PERRIER (dir.), B. AUGÈRE, T. DARRIBÈRE, J.-M. DUPIN, C. ESCUYER, J.-F. FOGELGESANG, & C. VAN DER REST (2017). *Biologie tout-en-un BCPST 1^{re} année*. Dunod, Paris, 4^e édition (1^{re} édition 2006).
- PEYCRU, P., D. GRANDPERRIN, C. PERRIER (dir.), B. AUGÈRE, J.-F. BEAUX, F. CARIOU, P. CARRÈRE, T. DARRIBÈRE, J.-M. DUPIN, C. ESCUYER, J.-F. FOGELGESANG, S. MAURY, É. QUÉINNEC, E. SALGUEIRO & C. VAN DER REST (2018). *Biologie tout-en-un BCPST 2^e année*. Dunod, Paris, 3^e édition (1^{re} édition 2007).
- POULIZAC, J.-A. (1999). *La variabilité génétique*. Ellipses, Paris.
- RAVEN, P. H., G. B. JOHNSON, J. B. LOSOS, S. S. SINGER (2007). *Biologie*. De Boeck, Bruxelles.
- RICHARD, D. (dir.), P. CHEVALET, S. FOURNEL, N. GIRAUD, F. GROS, P. LAURENTI, F. PRADÈRE & T. SOUBAYA (2012). *Biologie. Tout le cours en fiches. Licence. CAPES. Prépas*. Dunod, Paris, 2^e édition (1^{re} édition 2010).
- SAINTPIERRE, F., C. BORDI (dir.), M. ALGRAIN, Y. KRAUSS, I. MOLLIÈRE & H. CLAUCÉ (2017). *Mémento Biologie BCPST 1^{re} et 2^e années*. Vuibert, Paris.
- SAINTPIERRE, F., C. BORDI (dir.), M. ALGRAIN-PITAVY, A. DENIS, L. GERAY & I. MOLLIÈRE (2021). *Mémento Biologie BCPST 1^{re} et 2^e années*. Vuibert, Paris, 2^e édition (1^{re} édition 2017).
- SEGARRA, J. (dir.), É. CHAUVET, C. COLSON-PROCH, M. HUILLE, M. LABROUSSE, F. LOUET, F. METZ & E. PIÈTRE (2014). *Biologie BCPST 1^{re} année*. Ellipses, Paris.
- SEGARRA, J., E. PIÈTRE (dir.), G. BAILLY, O. CHASSAING, D. FAVRE, T. JEAN, F. METZ & C. MEUNIER (2015). *Biologie BCPST 2^e année*. Ellipses, Paris.
- TAGU, D., S. JAUBERT-POSSAMAI & A. MÈREAU (dir.) (2018). *Principes des techniques de biologie moléculaire et génomique*. 3^e édition revue et augmentée, Quae, Versailles (F).
- VIGNAIS, P. (2001). *La Biologie des origines à nos jours. Une Histoire des idées et des hommes*. « Grenoble Sciences », EDP Sciences, Les Ulis.
- VIGNAIS, P. (2006). *Science expérimentale et connaissance du Vivant. La Méthode et les concepts*. « Grenoble Sciences », EDP Sciences, Les Ulis.

Plan du chapitre

Objectifs : extraits du programme	1
Introduction	1
I. La nature et l'étude des génomes	2
A. Les acides nucléiques, des hétéropolymères (séquencés) de nucléotides porteurs d'une information	2
1. L'ADN, support universel de l'information génétique des êtres vivants (cellulaires)	2
a. Une molécule bicaténaire	2
b. Des brins orientables : extrémités 5' et 3'	2
c. Structure secondaire et tertiaire : deux brins antiparallèles organisés en double hélice	2
d. Une molécule qui respecte les règles de CHARGAFF	6
e. L'ADN, une molécule séquencée capable de porter une information	6
α. La capacité de porter une information : approche intuitive	6
β. Le caractère universel du codage de l'information génétique : mise en évidence par la transgénèse	6
f. Un codage par triplets de nucléotides de l'ADN (= codons dans l'ARNm) : le code génétique	7
α. Un système de correspondance (quasi-) universel entre séquence nucléotidique et séquence peptidique	7
β. Un code ponctué : des codons initiateurs et terminateurs de la traduction	7
γ. Un code caractérisé par sa redondance (ou dégénérescence)	7
g. La notion d'ORF, cadre ouvert de lecture	7
h. L'ADN, une molécule en interaction avec des protéines qui s'exprime, se conserve, se transmet et peut varier	8
2. Les ARN, des copies plus ou moins transitoires de portions d'ADN concourant à l'expression génétique (ou supports de l'information chez certains virus)	8
a. Les ARN, des acides nucléiques plutôt monocaténaires (pouvant se bicaténeriser) constituant des copies de petites portions d'ADN	8
b. Des molécules variées participant à l'expression génétique	9
c. Étude sommaire d'un exemple : un ARN de transfert (ARNt)	9
B. Des génomes étudiables par des techniques de biologie moléculaire	11
1. Notions de techniques de biologie moléculaire et biotechnologies	11
2. Les électrophorèses sur les acides nucléiques	11
3. Les méthodes de séquençage des génomes et leurs prolongements	12
a. Le principe d'une méthode historique de séquençage de petits fragments : la méthode enzymatique de SANGER (1977)	12
b. L'existence de techniques plus modernes et moins coûteuses : le séquençage haut débit	14
c. La possibilité d'aligner les séquences et de rechercher les fragments chevauchants pour aboutir à une reconstitution de longues séquences	15
d. La bio-informatique au secours du séquençage et de l'annotation des génomes ; notions de génomique (structurale / fonctionnelle) et de métagénomique	16
e. La taille des génomes et sa non-corrélation avec la position phylogénétique des organismes : le paradoxe de la valeur C	16
f. Le <i>barcoding</i> (et le <i>metabarcoding</i>) : des séquences d'un gène hautement conservé pour désigner les espèces	17
g. L'existence de banques de données assurant la mutualisation mondiale des portions séquencées et des protéines connues	18
4. L'amplification des séquences nucléotidiques	19
a. L'amplification de séquences d'ADN : la réaction de polymérisation en chaîne (PCR)	19
b. L'amplification de séquences d'ADN issues d'ARN : la PCR après rétrotranscription (RT-PCR)	20
c. Une approche quantifiée de la PCR et de la RT-PCR : la PCR et la RT-PCR quantitatives (= en temps réel) que l'on note qPCR et qRT-PCR	21
5. L'utilisation d'enzymes de restriction dans le découpage de l'ADN	21
a. Des endonucléases bactériennes qui reconnaissent et découpent des séquences-types généralement palindromiques (sites de restriction)	21
b. L'existence de sites de restriction en baïonnette permettant de générer des bouts collants utiles en génie génétique	22
c. L'utilisation d'enzymes de restriction dans la production de certaines empreintes génétiques	22
d. L'établissement de cartes de restriction	23
6. Des outils de recherche et de détection de séquences nucléotidiques spécifiques	24
a. L'hybridation <i>in situ</i> d'ARN ou d'ADN monobrin	24
b. Des outils de recherche de séquences ADN ou d'ARN exprimés par hybridation avec des sondes : les <i>Blots</i> (buvardages avec emploi de révélateur)	24
α. Principe général	24
β. <i>Southern Blot</i> (ADN), <i>Northern Blot</i> (ARN) et <i>Western Blot</i> (protéines) : diversité des <i>Blots</i>	25
c. Vers une automatisation de la recherche des ARN exprimés : les puces à ADN (biopuces), outil majeur en transcriptomique	26
7. La production d'organismes génétiquement modifiés (OGM) : la transgénèse	27
a. Principes généraux et étapes de la transgénèse	27
α. Notion de transgénèse, gène d'intérêt et avantage sélectif	27
β. Principales étapes du processus	27
γ. Intérêts et risques de la transgénèse	27
b. Isolement du gène d'intérêt	27
c. Insertion directe ou indirecte du gène d'intérêt (transgène) dans le génome cible	28
α. Quelques modalités classiques d'insertion directe : électroporation, micro-injection et biolistique	28
β. Transformation de la cellule cible via un vecteur plasmidique ou viral	29
i. Utilisation de plasmides	29
ii. Utilisation de virus	29
γ. Une technique en plein essor d'édition du génome : les ciseaux génétiques (cas de la protéine CRISPR Cas9)	30
d. Clonage des cellules après insertion du transgène	31
α. Principe général	31
β. Cas des Bactéries	32
γ. Cas des 'plantes'	32
δ. Cas des Animaux	32
e. Criblage génétique : sélection des souches transformées par vérification de l'insertion et de l'expression des transgènes	32
8. L'élucidation de la fonction des gènes	33
a. La possibilité de muter le gène étudié ici en vue de son invalidation : la mutagenèse aléatoire ou dirigée	33
α. La mutagenèse aléatoire : le déclenchement de mutations non maîtrisées par des agents mutagènes (en des lieux variables) ou la PCR mutagène (sur une séquence précise)	33
i. Mutagenèse aléatoire par des agents mutagènes : des mutations à localisation et en nombre imprévisibles	33
ii. L'utilisation de la PCR mutagène : des mutations favorisées lors l'amplification d'une séquence	33
β. La mutagenèse dirigée : le déclenchement de mutations ciblées par PCR mutagène	33
γ. L'édition génomique par le système CRISPR Cas9	33
b. La possibilité d'invalider un gène par l'introduction d'ARN interférents	33
c. La possibilité d'éteindre un gène par recombinaison homologue avec un allèle muté non fonctionnel (<i>knock-out</i> KO) ou de le remplacer par un autre (<i>knock-in</i>)	33

d. La possibilité d'exprimer un (trans)gène rapporteur ayant même séquence régulatrice qu'un gène dont on cherche à localiser et dater l'expression	34
e. La mise en évidence d'une interaction entre une protéine et un acide nucléique par la technique du retard sur gel (EMSA)	34
II. Le génome des Bactéries	35
A. Un génome généralement circulaire et localisé dans le cytoplasme	35
1. Un « chromosome bactérien » le plus souvent circulaire et unique situé dans une zone nommée nucléoïde : le génome principal	36
a. Nature et taille du chromosome bactérien	36
b. Localisation fine (nucléoïde) et organisation du chromosome bactérien (condensation modérée et protéines de structure)	36
2. Des plasmides, petites unités génétiques fréquentes (mais non systématiques), circulaires, autonomes et transmissibles : le génome facultatif ou accessoire	37
B. Un génome largement codant et des gènes souvent groupés en opérons polycistroniques	38
1. Peu de séquences répétées et non codantes : un génome largement codant et plutôt compact	38
2. Des gènes non morcelés (= sans introns) et souvent organisés en opérons, unités de transcription polycistroniques dépendant des mêmes unités régulatrices	38
III. Le génome des Eucaryotes	39
A. Localisation, organisation et structuration du génome eucaryote	39
1. Vue d'ensemble : un génome protégé comprenant une composante principale nucléaire et une composante mineure cytoplasmique	39
2. Le génome nucléaire, une entité composée de chromosomes linéaires pouvant présenter des niveaux variables de condensation	40
a. Organisation fondamentale : des chromosomes comportant une ou deux chromatide(s) impliquant des protéines associées à l'ADN (dont les histones)	40
b. Degrés de condensation de l'ADN	41
c. L'ADN en interphase	42
α. Notions de chromatine (euchromatine vs. hétérochromatine) et de nucléole(s)	42
β. Une association de l'hétérochromatine avec la lamina nucléaire (LAD) ou le nucléole (NAD), et un regroupement fonctionnel des zones d'expression dans l'euchromatine (usines à transcription)	42
γ. L'importance de l'histone H1 et du complexe de remodelage de la chromatine dans la transition euchromatine / hétérochromatine	43
δ. Une répartition des chromosomes décondensés non aléatoire et localisée : notion de territoire chromosomique	43
d. Le chromosome mitotique (cas du chromosome métaphasique)	43
e. Une taille variable du génome nucléaire (en moyenne grossièrement de l'ordre du Gb)	45
3. Le génome extranucléaire, un ensemble de chromosomes de type bactérien codant très partiellement les polypeptides des organites semi-autonomes	45
B. Des gènes nucléaires morcelés et monocistroniques	46
1. Notions de gène, cistron, locus, allèle, hétérozygotie, homozygotie, dominance et récessivité	46
2. Des gènes morcelés (= gènes mosaïques) chez les Eucaryotes avec des régions non codantes (introns) séparant les portions codantes ou non traduites (exons)	46
a. Mise en évidence des introns par hybridation ARN-ADN monobrin (1977)	46
b. Organisation des gènes eucaryotes	47
C. La présence importante de séquences répétées et d'ADN non codant chez les Eucaryotes	49

1. Mise en évidence des séquences répétitives par les expériences de dénaturation-renaturation de l'ADN	49
2. Nature des séquences répétitives des Eucaryotes	50
IV. Les virus et leur génome	52
A. Les virions (= particules virales), état libre des virus	52
1. Organisation : une information génétique enfermée dans une capsid et éventuellement une enveloppe	52
a. Une information génétique portée par un acide nucléique	52
α. Un acide nucléique de nature variable qui sert de base à la classification	52
β. Une information réduite compactée et composée de quelques gènes à quelques dizaines, avec une possibilité de chevauchement des cadres de lecture	53
b. La présence d'une capsid protéique protectrice	55
c. La présence fréquente d'une enveloppe (= péplos) de nature membranaire	56
2. Les virions, entités composées de matière organique mais acellulaires et « acaryotes »	56
3. Les virions, parasites pathogènes de cellules de tous groupes	56
4. Les virus, entités évolutives à l'origine discutée [pour information]	57
a. Des entités génétiques qui subissent des mutations	57
b. Petit détour sur les origines du vivant (organismes cellulaires)	57
c. L'origine des virus : diversité des hypothèses possibles	57
d. La difficulté de brancher les virus sur l'arbre du vivant	57
B. Infection, parasitisme et reproduction des virus	58
1. Fonctionnement des virus : une vue d'ensemble	58
2. Modalités de pénétration des virus dans les cellules hôtes (infection)	58
a. Première possibilité : pénétration passive (cas du VMT)	58
b. Deuxième possibilité : fixation sur la cellule hôte (cas du SARS-Cov2 et des phages)	58
3. Les cycles viraux des exemples étudiés en détail	59
a. Vue d'ensemble	59
b. Les deux cycles des Bactériophages (cycle lytique, cycle lysogénique)	59
α. Le cycle lytique, un cycle sans provirus qui aboutit à la libération de nouveaux phages par lyse bactérienne	59
β. Le cycle lysogénique, un cycle latent avec provirus (prophage)	60
α. Un cycle qui coexiste avec le cycle lytique	60
β. Modalités du cycle lysogénique	60
c. Cycle du Virus de la mosaïque du tabac (VMT)	61
d. Cycle du virus de l'immunodéficience humaine (VIH) [pour information]	61
e. Cycle du SARS-Cov-2	62
Bilan	64
Pour faire une fiche de révision : quelques pistes	64
Références	65
Plan du chapitre	66
Plan simplifié du chapitre (3 niveaux de plan)	68
Plan très simplifié du chapitre (2 niveaux de plan)	69

Plan simplifié du chapitre (3 niveaux de plan)

Objectifs : extraits du programme	1
Introduction	1
I. La nature et l'étude des génomes	2
A. Les acides nucléiques, des hétéropolymères (séquencés) de nucléotides porteurs d'une information	2
1. L'ADN, support universel de l'information génétique des êtres vivants (cellulaires)	2
2. Les ARN, des copies plus ou moins transitoires de portions d'ADN concourant à l'expression génétique (ou supports de l'information chez certains virus)	8
B. Des génomes étudiables par des techniques de biologie moléculaire	11
1. Notions de techniques de biologie moléculaire et biotechnologies	11
2. Les électrophorèses sur les acides nucléiques	11
3. Les méthodes de séquençage des génomes et leurs prolongements	12
4. L'amplification des séquences nucléotidiques	19
5. L'utilisation d'enzymes de restriction dans le découpage de l'ADN	21
6. Des outils de recherche et de détection de séquences nucléotidiques spécifiques	24
7. La production d'organismes génétiquement modifiés (OGM) : la transgénèse	27
8. L'élucidation de la fonction des gènes	33
II. Le génome des Bactéries	35
A. Un génome généralement circulaire et localisé dans le cytoplasme	35
1. Un « chromosome bactérien » le plus souvent circulaire et unique situé dans une zone nommée nucléoïde : le génome principal	36
2. Des plasmides, petites unités génétiques fréquentes (mais non systématiques), circulaires, autonomes et transmissibles : le génome facultatif ou accessoire	37
B. Un génome largement codant et des gènes souvent groupés en opérons polycistroniques	38
1. Peu de séquences répétées et non codantes : un génome largement codant et plutôt compact	38
2. Des gènes non morcelés (= sans introns) et souvent organisés en opérons, unités de transcription polycistroniques dépendant des mêmes unités régulatrices	38
III. Le génome des Eucaryotes	39
A. Localisation, organisation et structuration du génome eucaryote	39
1. Vue d'ensemble : un génome protégé comprenant une composante principale nucléaire et une composante mineure cytoplasmique	39
2. Le génome nucléaire, une entité composée de chromosomes linéaires pouvant présenter des niveaux variables de condensation	40
3. Le génome extranucléaire, un ensemble de chromosomes de type bactérien codant très partiellement les polypeptides des organites semi-autonomes	45
B. Des gènes nucléaires morcelés et monocistroniques	46
1. Notions de gène, cistron, locus, allèle, hétérozygotie, homozygotie, dominance et récessivité	46
2. Des gènes morcelés (= gènes mosaïques) chez les Eucaryotes avec des régions non codantes (introns) séparant les portions codantes ou non traduites (exons)	46
C. La présence importante de séquences répétées et d'ADN non codant chez les Eucaryotes	49
1. Mise en évidence des séquences répétitives par les expériences de dénaturation-renaturation de l'ADN	49
2. Nature des séquences répétitives des Eucaryotes	50

IV. Les virus et leur génome	52
A. Les virions (= particules virales), état libre des virus	52
1. Organisation : une information génétique enfermée dans une capsid et éventuellement une enveloppe	52
2. Les virions, entités composées de matière organique mais acellulaires et « acaryotes »	56
3. Les virions, parasites pathogènes de cellules de tous groupes	56
4. Les virus, entités évolutives à l'origine discutée [pour information]	57
B. Infection, parasitisme et reproduction des virus	58
1. Fonctionnement des virus : une vue d'ensemble	58
2. Modalités de pénétration des virus dans les cellules hôtes (infection)	58
3. Les cycles viraux des exemples étudiés en détail	59
Bilan	64
Pour faire une fiche de révision : quelques pistes	64
Références	65
Plan du chapitre	66
Plan simplifié du chapitre (3 niveaux de plan)	68
Plan très simplifié du chapitre (2 niveaux de plan)	69

Plan très simplifié du chapitre (2 niveaux de plan)

Objectifs : extraits du programme	1
Introduction	1
I. La nature et l'étude des génomes	2
A. Les acides nucléiques, des hétéropolymères (séquencés) de nucléotides porteurs d'une information	2
B. Des génomes étudiables par des techniques de biologie moléculaire	11
II. Le génome des Bactéries	35
A. Un génome généralement circulaire et localisé dans le cytoplasme	35
B. Un génome largement codant et des gènes souvent groupés en opérons polycistroniques	38
III. Le génome des Eucaryotes	39
A. Localisation, organisation et structuration du génome eucaryote	39
B. Des gènes nucléaires morcelés et monocistroniques	46
C. La présence importante de séquences répétées et d'ADN non codant chez les Eucaryotes	49
IV. Les virus et leur génome	52
A. Les virions (= particules virales), état libre des virus	52
B. Infection, parasitisme et reproduction des virus	58
Bilan	64
Pour faire une fiche de révision : quelques pistes	64
Références	65
Plan du chapitre	66
Plan simplifié du chapitre (3 niveaux de plan)	68
Plan très simplifié du chapitre (2 niveaux de plan)	69

© Tanguy JEAN. Les textes et les figures originales sont la propriété de l'auteur. Les figures extraites d'autres sources restent évidemment la propriété des auteurs ou éditeurs originaux.
Document produit en février 2022 • Dernière actualisation : janvier 2026.
Contact : Tanguy.Jean4@gmail.com
Adresse de téléchargement : <https://www.svt-tanguy-jean.com/>



Ces données sont placées sous licence *Creative Commons Attribution – Pas d'Utilisation commerciale 4.0 CC BY NC* qui autorise la reproduction et la diffusion du document, à condition d'en citer explicitement la source et de ne pas en faire d'utilisation commerciale.