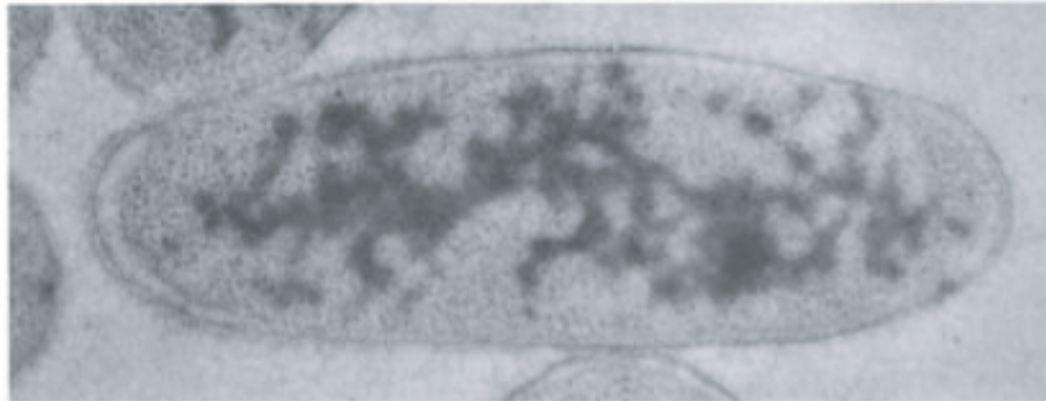


Figure 1 : le nucléoïde bactérien



E. coli marquée avec un anticorps anti ADN.

Représentation schématique du génome bactérien d'Escherichia coli (bactérie gram-)

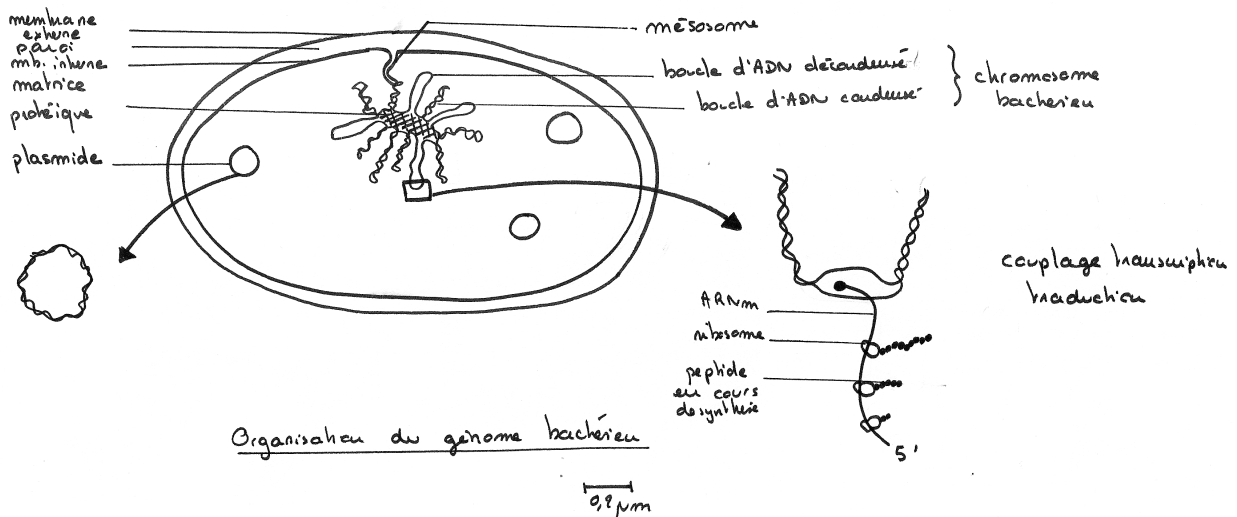
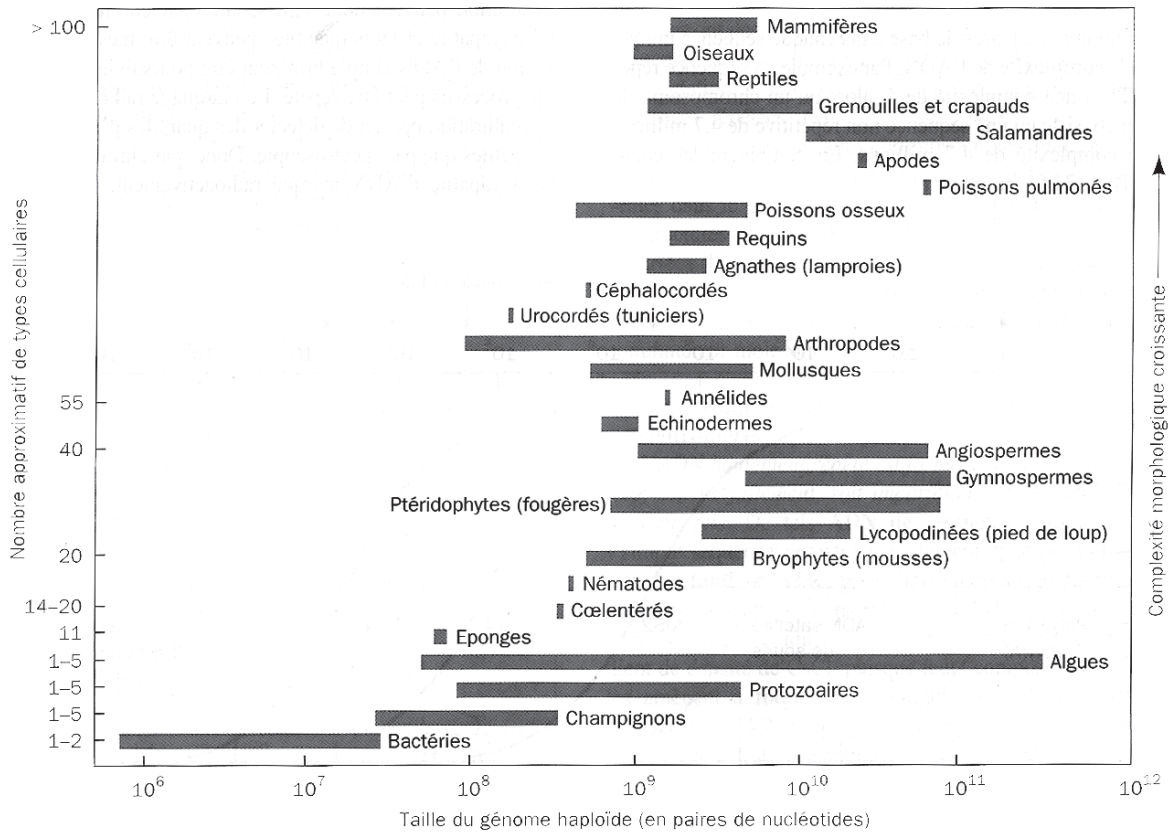


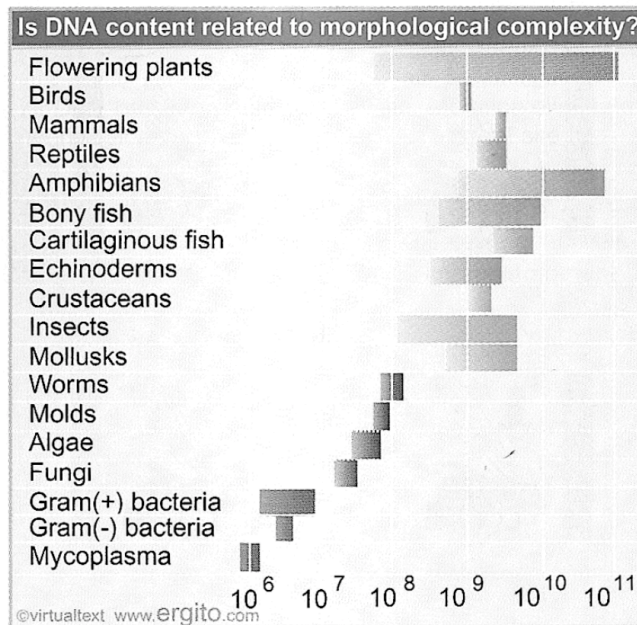
Figure 2 : Taille des génomes de quelques êtres vivants (Gene VII Lewin, Biochimie Voet)

Organismes		Nombre de séquences codantes prédites	Valeur c (Mpb)
<i>Mycoplasma</i>	procaryotes	470	0,58
<i>Rickettsia</i>		834	1,11
<i>Haemophilus</i>		1743	1,83
<i>Methanococcus</i>		1738	1,66
<i>Bacillus</i>		4100	4,2
<i>E.coli</i>		4288	4,6
<i>Saccharomyces</i> (levure)	eucaryotes	6034	13,5
<i>Drosophila</i>		12000	165
<i>Caenorhabditis</i> (nématode)		19100	97
<i>Homo sapiens</i>		20000	3200



Tailles comparées des génomes haploïdes de différentes catégories d'organismes mettant en évidence le paradoxe de la valeur C. La complexité morphologique des organismes, estimée par le

nombre de types cellulaires, est croissante du bas de la figure vers le haut. [D'après Raff, R.A. et Kaufman, T.C. *Embryos., Genes, and Evolution*, p. 314, Macmillan (1983).]



DNA content of the haploid genome is related to the morphological complexity of lower eukaryotes, but varies extensively among the higher eukaryotes. The range of DNA values within a phylum is indicated by the shaded area.

Figure 3 : techniques de séquençage (Dunod 2021)

• Méthode historique de Sanger (la seule au programme)

La méthode de **Sanger** (1977) a permis l'obtention des premières séquences. Elle est fondée sur la polymérisation d'un brin d'ADN complémentaire au brin que l'on souhaite séquencer.

Cette méthode repose sur l'utilisation de nucléotides terminateurs de chaîne : des **didésoxynucléotides** (absence de groupement OH en 3'). Ces didésoxynucléotides peuvent s'incorporer comme les désoxynucléotides, mais leur insertion provoque un arrêt immédiat de la polymérisation :

On mélange donc dans 4 tubes :

- l'ADN à séquencer
- de l'ADN polymérase
- des désoxynucléotides « normaux »
- une petite quantité de **l'un des** didésoxynucléotides, ces didésoxynucléotides sont radioactifs.

Dans chaque tube (par exemple celui qui contient ddATP), les nucléotides normaux sont en excès mais parfois, un nucléotide terminateur de chaîne vient se mettre en place et la synthèse du nouveau brin s'arrête.

En statistique, comme dans le tube on met de nombreuses copies du brin monocaténaire à séquencer, on obtient **différentes longueurs de nouveaux brins**. La plupart ont une longueur normale : ils n'ont intégré aucun nucléoside terminateur de chaîne. Mais pour ceux qui sont plus courts, on sait qu'ils se terminent tous dans ce tube par A : le nucléotide avec adénine.

Les nouveaux brins sont séparés par électrophorèse. Le contenu de chaque tube migre dans une piste différente du gel. L'emplacement des nouveaux brins est rendue visible par autoradiographie : les nucléosides terminateurs de chaînes sont radioactifs ce qui permet de visualiser les fragments.

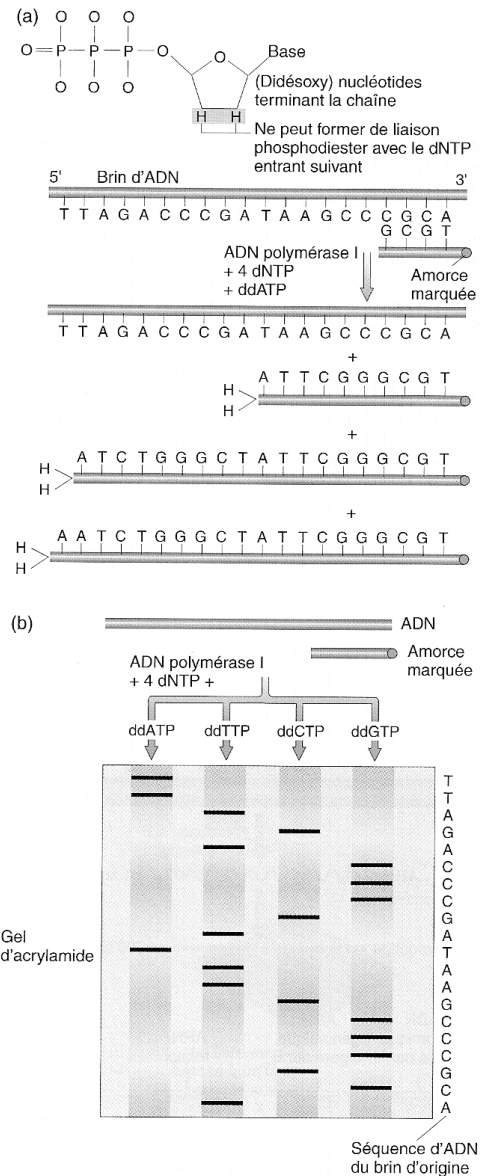
L'enchaînement des bandes permet de lire la séquence.

Pour le concours vous devez être capable de lire un gel de séquence.

• Méthodes actuelles

Méthodes basées sur la technique de Sanger

Aujourd'hui on utilise des ddNTP dont chacun est marqué par un fluorophore spécifique. La réaction de séquençage est alors conduite en un seul milieu avec le ddCTP fluorescent bleu, le ddATP fluorescent vert, le ddGTP fluorescent jaune et le ddTTP fluorescent rouge. Les fragments interrompus à un nucléotide donné porteront donc un fluorophore terminal de couleur spécifique à ce nucléotide.



La technique de séquençage utilisant des didésoxynucléotides. (a) Une amorce marquée (conçue à partir de la séquence voisine du vecteur) est utilisée pour initier la synthèse d'ADN. L'addition de quatre didésoxynucléotides différents (ici on voit du ddATP) interrompt la synthèse au hasard. (b) Les fragments résultants sont séparés par électrophorèse et soumis à une autoradiographie. (D'après J. D. Watson, M. Gilman, J. Witkowski et M. Zoller, *L'ADN recombinant*. Traduction française de la 2^e édition chez De Boeck, 1994.)

Ce sont ces fluophores qui seront détectés lors de l'électrophorèse, rendant inutile le traceur radioactif. L'électrophorèse se fait alors dans des séquenceurs capillaires : ceux-ci sont constitués de fins tubes de verre, de quelques microns de diamètres sur plusieurs dizaines de longueur (de 30 à 50 cm) et dans lesquels sont introduits l'ensemble des fragments.

La détection de la fluorescence se fait par un rayon laser qui travers le capillaire qui excite les les ddNTP fluorescents incorporées à l'ADN. Une caméra mesure l'émission de fluorescence, les différents ddNTP se distinguant par la longueur d'onde émise.

Un capillaire permet la lecture d'environ **100 kb en 1h**. Les séquenceurs utilisés actuellement comprennent de très nombreux capillaires.

Méthodes basées sur l'utilisation de nanopores

La **technologie nanopore** (Oxford Nanopore Technologies (ONT) est apparue en 2014. Elle n'utilise pas de polymérase pour la synthèse d'une copie de la molécule à séquencer mais fait passer des molécules d'ADN dans un nanopore enchâssé dans une membrane qui sépare deux compartiments contenant des solutions ioniques. Ce passage entraîne des perturbations d'un signal électrique qui est fonction du nucléotide traversant le pore. La séquence est déduite de la succession des perturbations.

Ces technologies sont dites aussi de **lectures longues** (ou *long read*), du fait de leur capacité à séquencer des molécules uniques dont la longueur peut atteindre des centaines de kb. Le facteur limitant devient la longueur des fragments d'ADN que l'on réussit préalablement à obtenir lors de l'extraction de cette molécule. Le problème devient donc de trouver des méthodes d'extraction qui laissent l'ADN le plus intact possible.

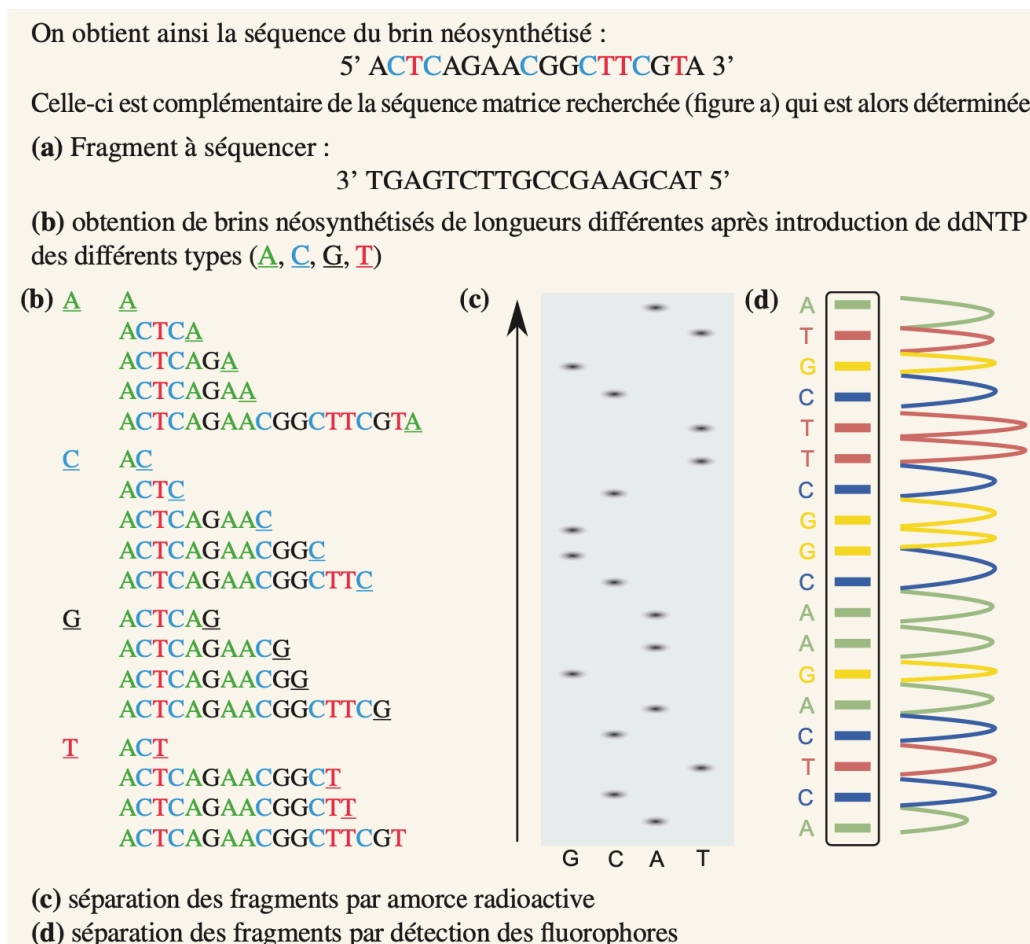


Figure 4 : annotation du génome de virus SARS
(Jean Michel et al., 2020 ; Islam et al., 2020)

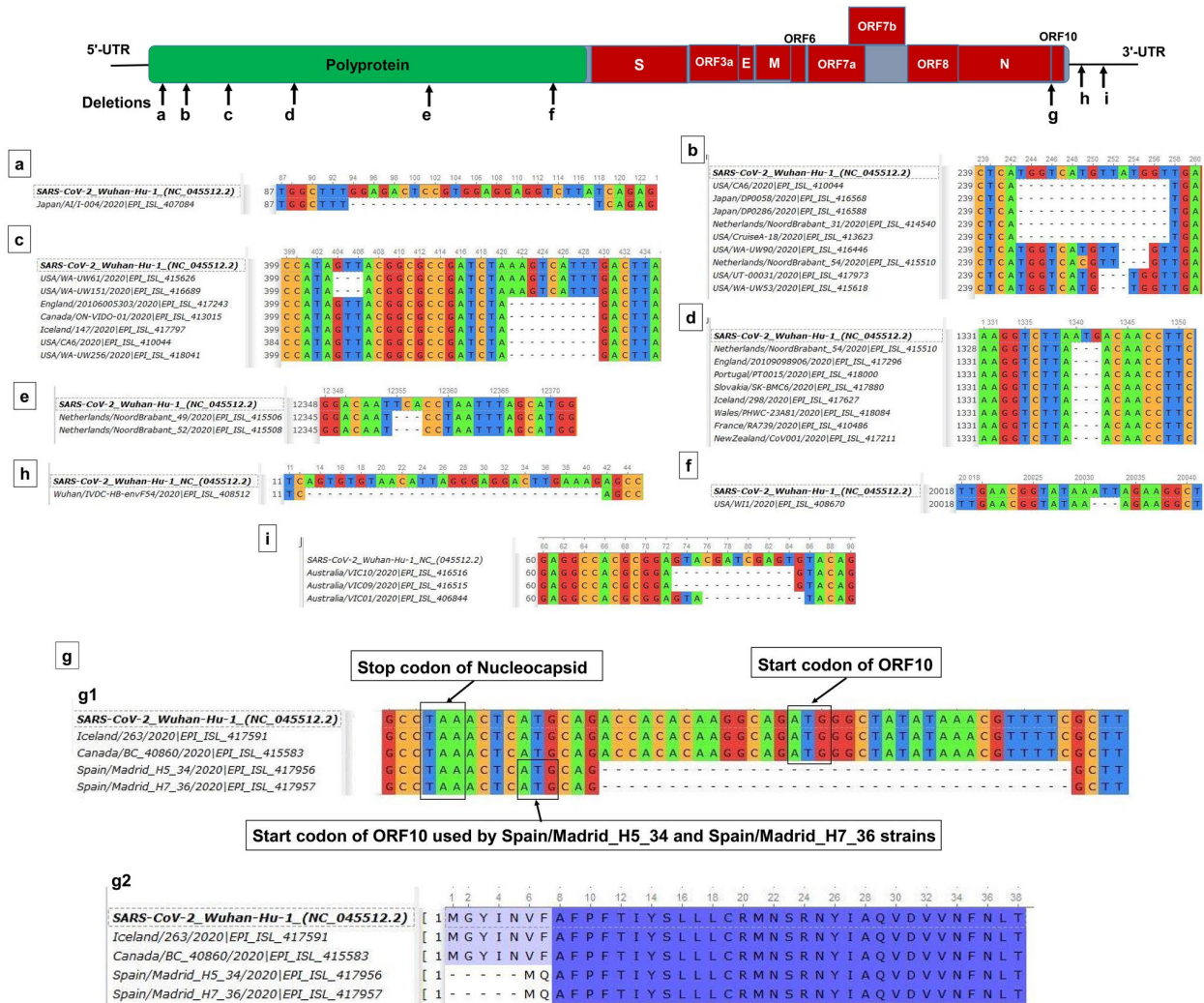


Figure 1. Genomic deletion analysis of SARS-CoV-2. Genomic deletion analysis of SARS-CoV-2 strains identified (a) 24 (nt) deletions in NSP1 in a Japanese strain; (b) 15-nt deletions in NSP1 of viral strains from USA, Japan and the Netherlands alongside three-nt deletions in USA and Netherlands; (c) three-nt deletion in NSP1 of American strains and very adjacent to that, nine-nucleotide deletion of strains from the USA, England and Canada and Iceland; (d) three-nucleotide deletions in NSP2 were observed in 99 strains from Netherlands, England, Portugal, Slovakia, Iceland, Wales, France and New Zealand (representatives from each countries were shown); (e) NSP8 undergoes three-nt deletion in Netherlandian/dutch/hollanders strains; (f) three-nucleotide deletion in NSP15 of USA strain; (g-g1) 35nt deletion, including start codon position of ORF10 of Spain strains, and the start codon in spacer position, has been used for ORF10 coding; and as a result, (g-g2) five aa residues deletion in those strains starting from position 1 to 5. Deletion of (h) 29-nt reported from Wuhan, and (i) 10-nt in 3'-UTR of strains belonged to Australia. The position of nt represents the starting position from each ORFs, for instance, position of ORF1ab was considered for the NSPs. MAFFT online tool³⁶ was used for alignment, and Unipro-UGENE⁴⁰ used for visualization.

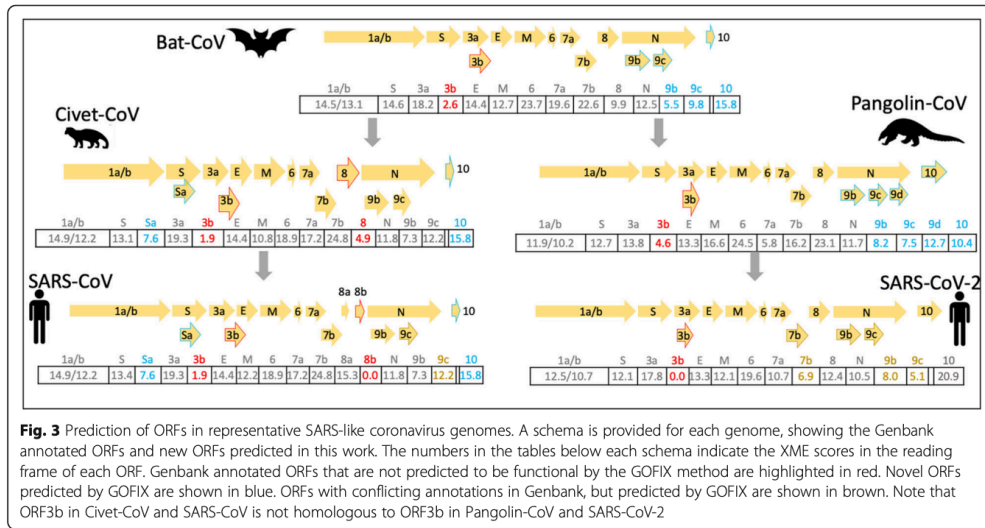


Figure 5 : les plasmides, des molécules d'ADN facultatives conférant un avantage sélectif

Type de plasmide	Propriétés
Plasmide de résistance	Résistance à un antibiotique
Plasmide de fertilité	Conjugaison et transfert d'ADN
Plasmide killer	Synthèse de toxines tuant d'autres bactéries
Plasmide métabolique	Protéines impliquées dans des phénomènes de transport de molécules
Plasmide de virulence	Pathogénicité

Figure 6 : organisation du noyau eucaryote

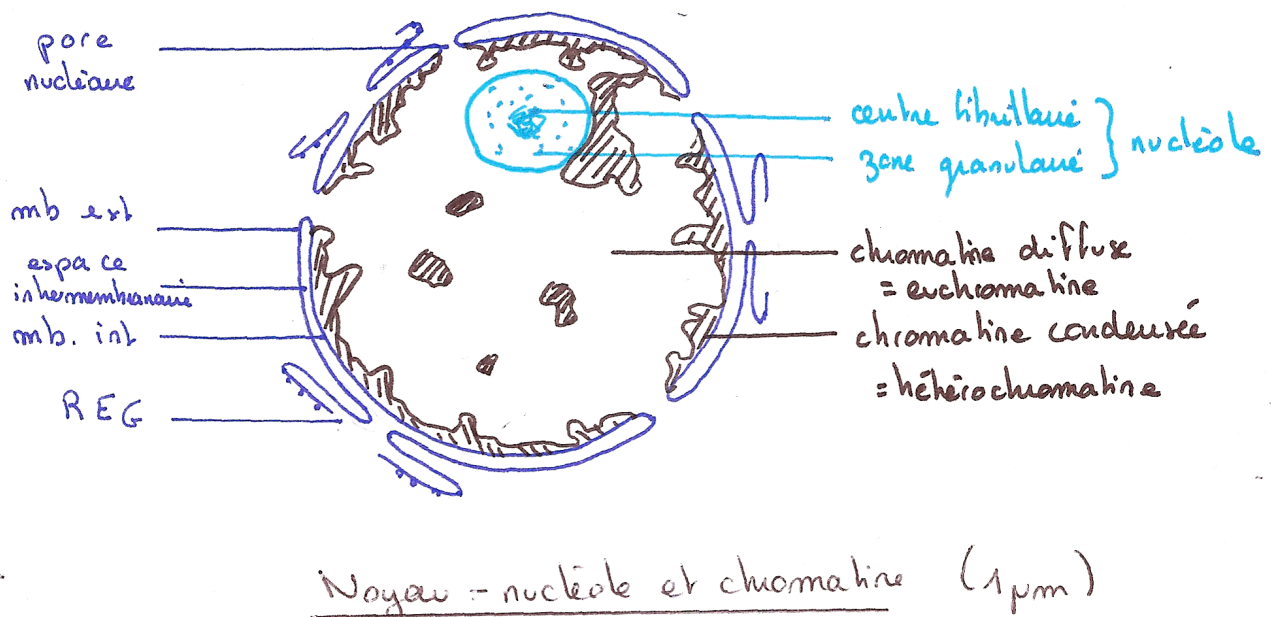


Figure 7 : coopération entre génome nucléaire et génome chloroplastique pour la synthèse de la RubisCO

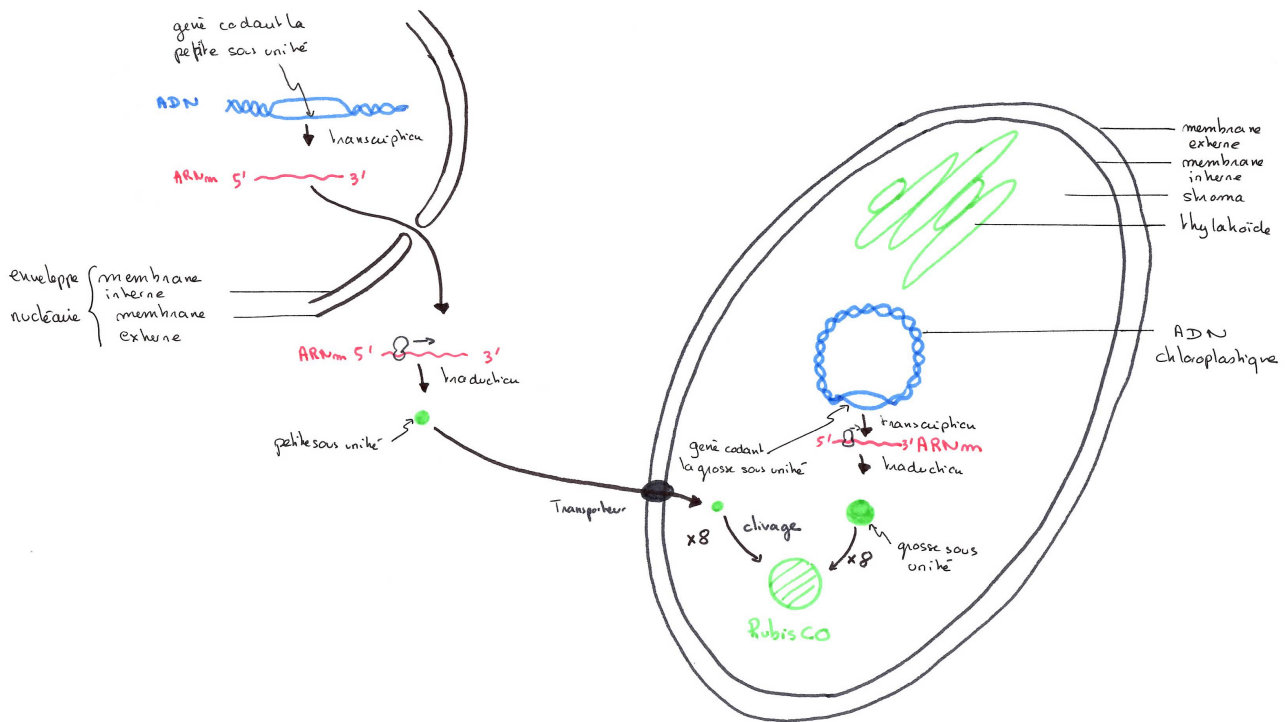
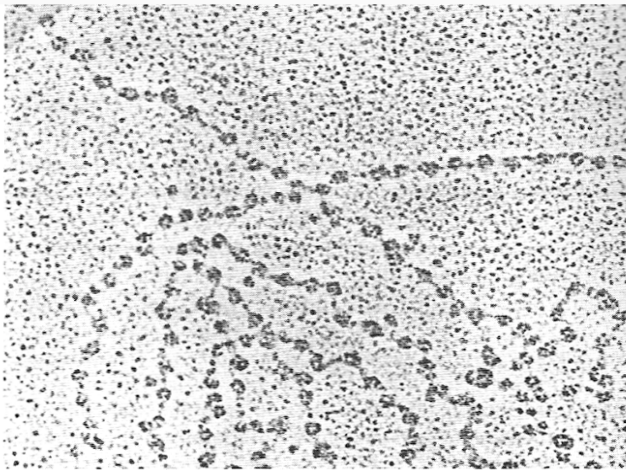
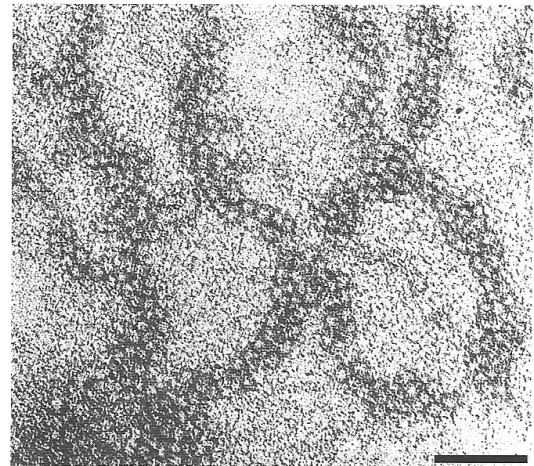


Figure 8 : Niveaux de compaction de la chromatine
(Biochimie Voet)

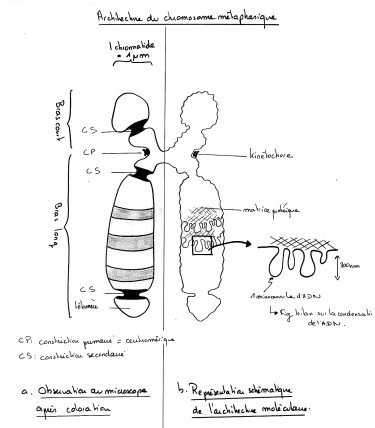
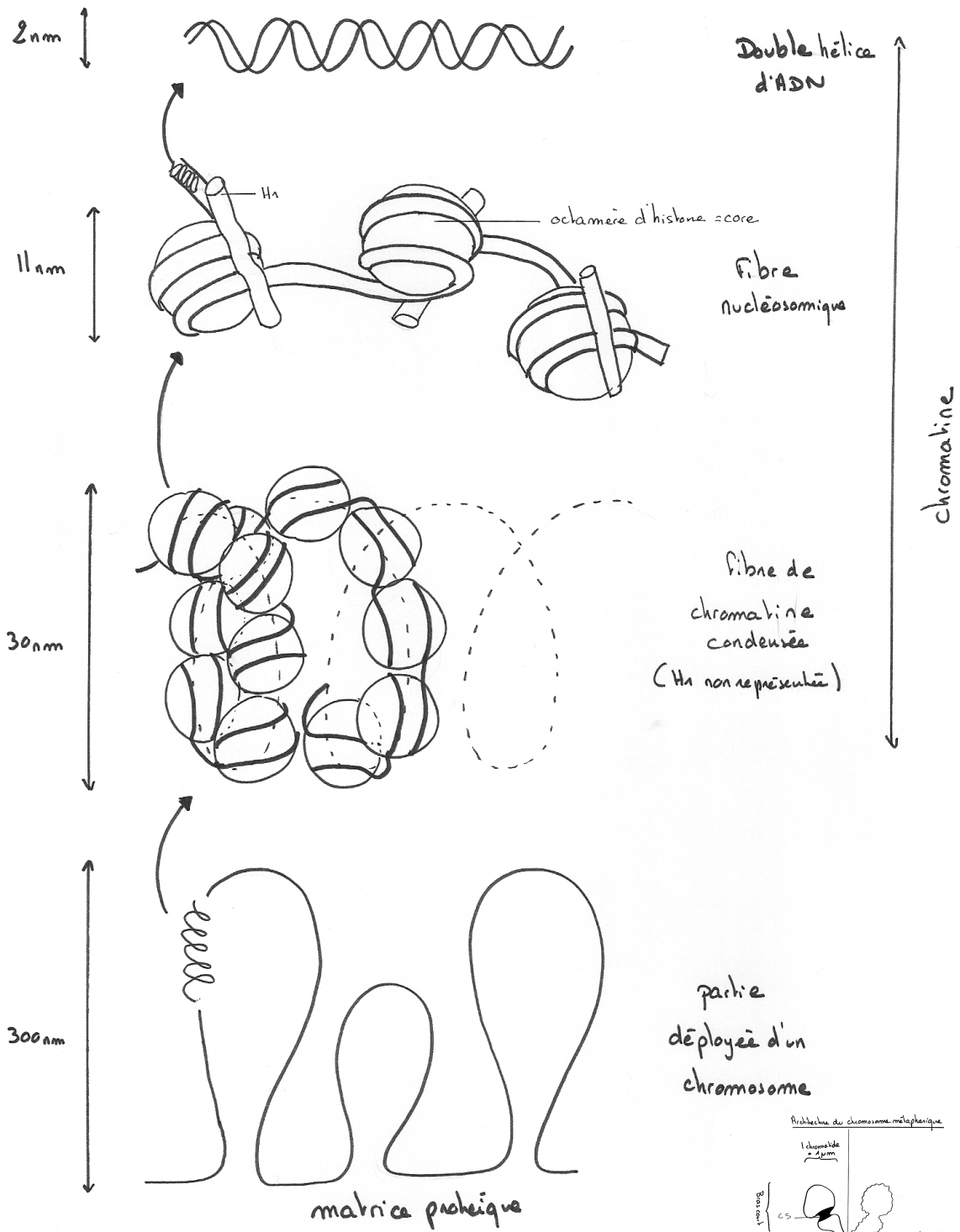


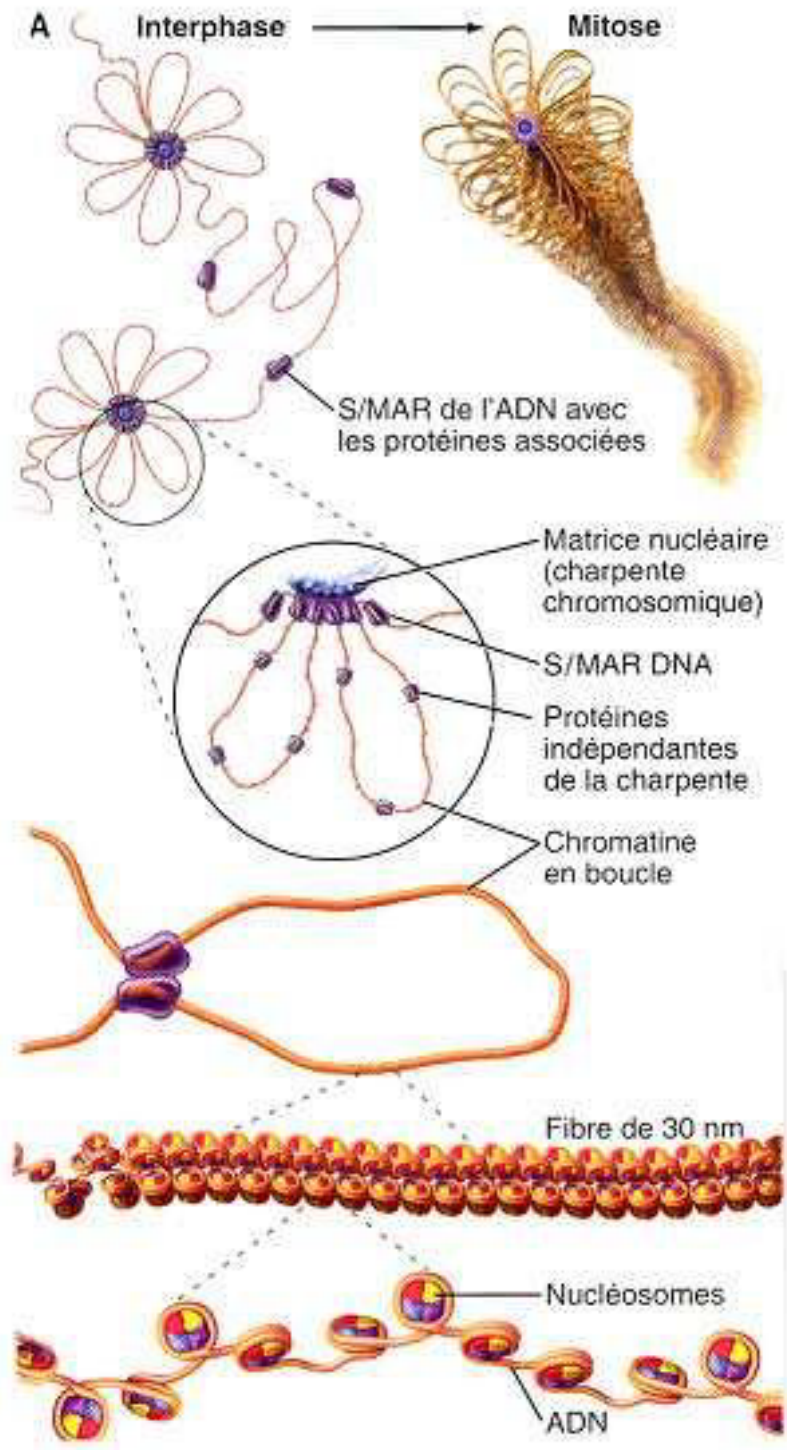
Micrographie électronique de chromatine de *D. melanogaster* montrant que ses fibres de 10 nm sont des chapelets de nucléosomes. [Avec la permission d'Oscar L. Miller, Jr., Université de Virginie.]



Micrographie électronique des filaments de chromatine de 30 nm. Notez que les filaments ont un diamètre correspondant à deux ou trois nucléosomes. La barre représente 1000 Å. [Avec la permission de Jérôme B. Rattner, Université de Calgary, Canada.]

Figure 9 : Bilan sur la compaction de la chromatine





Données récentes sur compaction de la chromatine

(<https://www.umontpellier.fr/articles/de-nouvelles-structures-identifiees-au-sein-de-nos-chromosomes>)

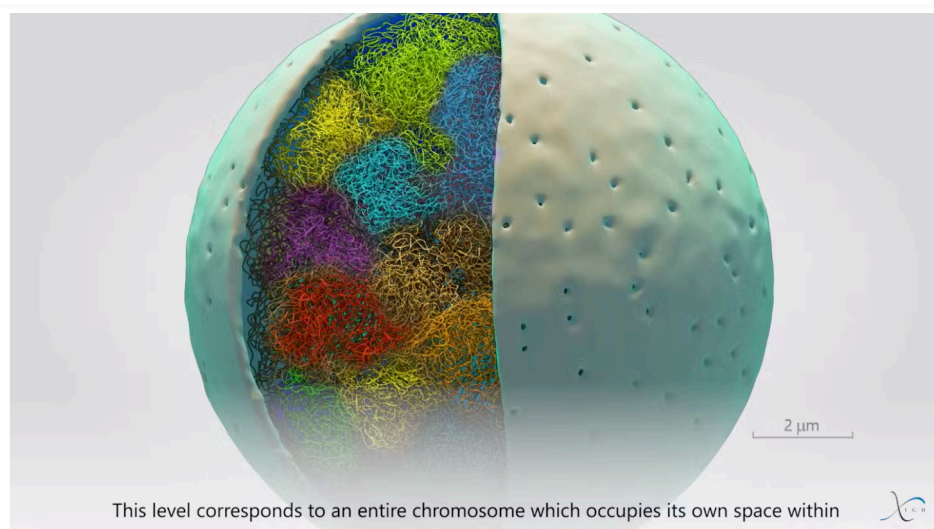
- Le premier niveau de compaction est le **nucléosome**.
- Le deuxième niveau d'organisation correspond à des groupements de quelques nucléosomes, comme des petits amas que l'on appelle en anglais « **nucleosome clutches** » (ils ont été ainsi nommés par les auteurs de leur découverte en analogie aux œufs trouvés dans les nids couvés). Les nucléosomes ne sont donc pas groupés de manière régulière comme on le pensait, mais plutôt en paquets irréguliers.

Ces « nucleosome clutches » se regroupent ensuite pour former une structure appelée « **Chromatin Nanodomain** », ou CND, qui inclue approximativement 100 000 à 200 000 paires de base d'ADN, formant des gros amas irréguliers de nucléosomes de 150 à 300 nm de large. Ces deux niveaux ont été découverts récemment (respectivement en 2015 et 2020), grâce à la microscopie en super-résolution, capable de résoudre des structures de 20 à 100 nm.

- Le niveau suivant de cette organisation est appelé TAD, en anglais « **Topologically Associating Domain** » pour « domaine d'association topologique », identifiés en 2012 avec la méthode moléculaire du Hi-C. Les TADs sont composés de plusieurs CNDs, formant des super amas de nucléosomes d'environ 500 nm de large. Ils comprennent ainsi des tailles variables d'ADN, avec une moyenne d'environ 1 mégabase (1 million de paires de bases).

Les **TADs vont préserver les gènes et leurs régions régulatrices** dans un même environnement moléculaire, ce qui peut être propice à leur expression (c'est-à-dire leur lecture pour conduire à la production d'une protéine) dans un type cellulaire donné où leur activité est nécessaire. Grâce à leurs frontières, ils vont également permettre de bien séparer les gènes entre eux, pour éviter que des gènes actifs influencent d'autres gènes inactifs dans un type cellulaire donné.

- Plusieurs TADs vont ensuite se regrouper en **deux compartiments distincts**, le compartiment « A » qui contient principalement des gènes actifs et le compartiment « B », qui inclut principalement des gènes inactifs. Ce **regroupement en compartiments** permet de renforcer les fonctions du génome.
- Enfin, à l'extrémité de cette échelle d'organisation, on retrouve les territoires chromosomiques, qui permettent aux chromosomes d'être individualisés les uns par rapport aux autres au sein des noyaux.



Chaque chromosome (1 chromosome = 1 couleur) occupe son propre territoire dans le noyau des cellules.

Frédéric Bantignies, Giacomo Cavalli, Fourni par l'auteur

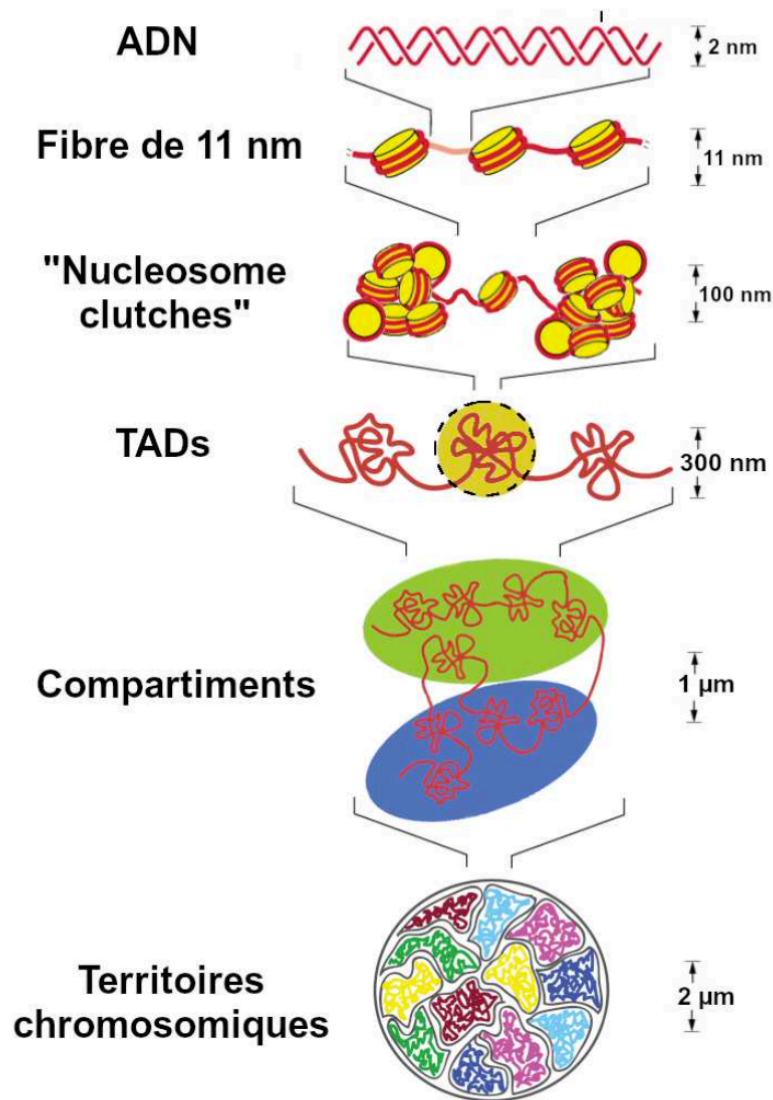


Figure 1 : Représentation schématique des différents niveaux d'organisation de la chromatine dans une cellule en interphase. L'acide désoxyribonucléique ou ADN (en rouge) s'associe aux protéines histones (en jaune) pour former des nucléosomes en fibre de 11 nanomètres. Cette fibre se replie par la suite en groupes de taille et densité en nucléosomes hétérogènes qui ont été définis comme des « nucleosome clutches ». Cette organisation se replie encore pour former les domaines topologiques ou 'Topologically Associating Domains' (TADs, rond de couleur jaune) se repliant sur eux-mêmes en compartiments (ellipses bleues et vertes), et finalement formant les territoires chromosomiques. Schéma modifié d'après (Schilit, 2011) et (Matharu and Ahituv, 2015).

Figure 10 : le chromosome métaphasique des Eucaryotes

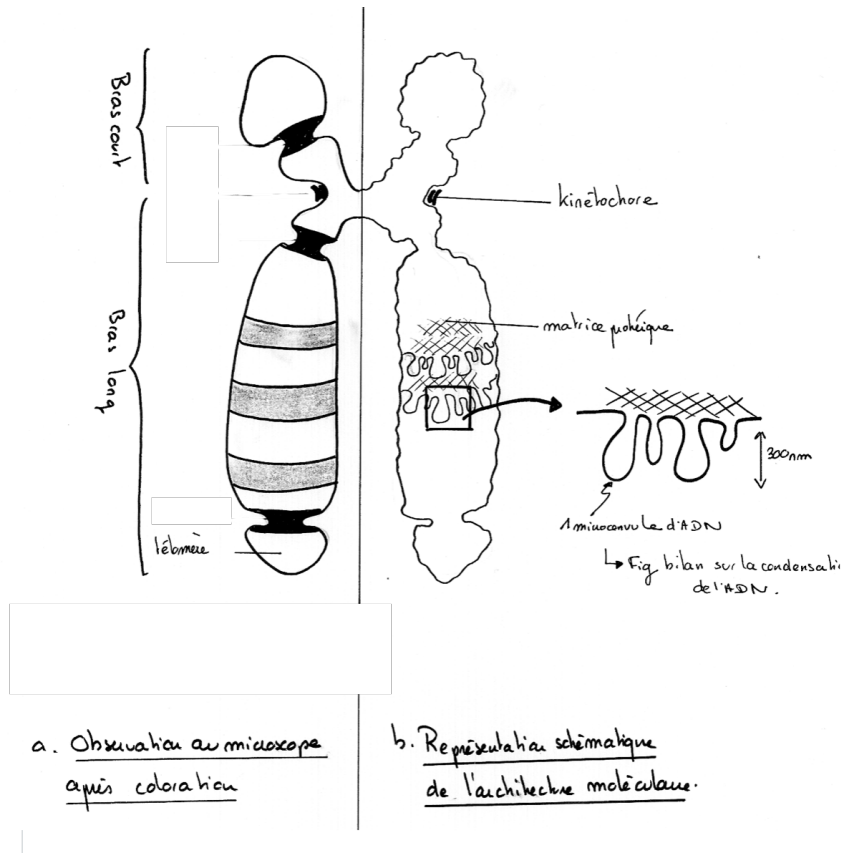


Figure 11 : structure de l'opéron lactose

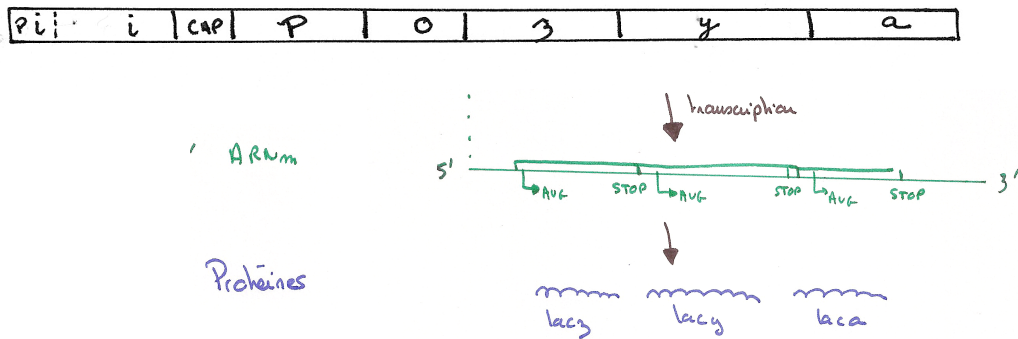


Figure 12 : étude de la croissance d'une population bactérienne sur différents substrats

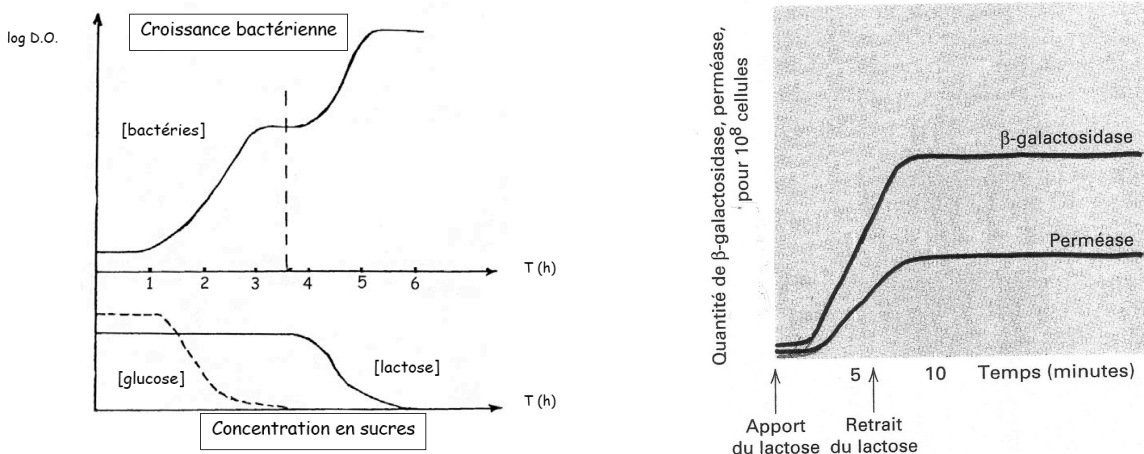


Figure 13 : étude de la complexité des génomes par des expériences de réassociation
(www.edu.upmc)

Méthode : l'ADN est découpé en fragments de 300 à 10 000 pb puis dénaturé. On le laisse ensuite se réhybrider.

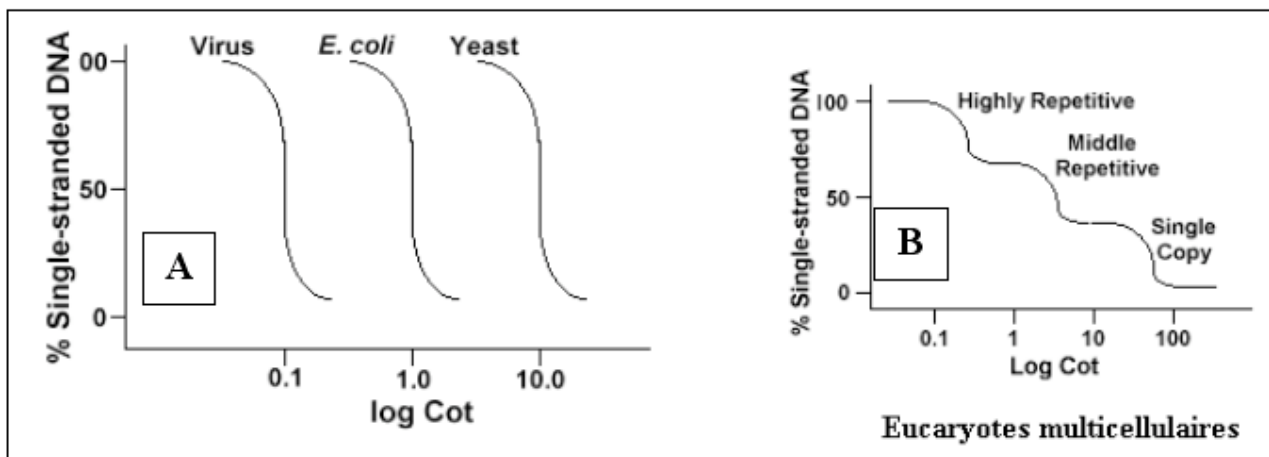
L'étape déterminant la vitesse de renaturation (=hybridation) est la rencontre entre séquences complémentaires : cette étape est directement dépendante du nombre de séquences identiques dans le génome, c'est-à-dire le nombre de séquences répétées. Une fois que les séquences complémentaires se sont rencontrées par diffusion au hasard, elles s'hybrident rapidement. et forment des molécules double brin.

A retenir : **plus une séquence est répétée, plus la réhybridation est rapide.**

Présentation des résultats : courbe Cot

Sur l'axe des Y on porte le pourcentage d'ADN qui reste simple brin par rapport à la concentration totale d'ADN. L'axe des X porte une échelle logarithmique du produit de la concentration initiale de l'ADN (en moles/litre) par le temps écoulé (en secondes). Ainsi une solution avec une forte concentration en ADN incubée pdt un temps court aura le même Cot qu'une solution ayant une faible concentration en ADN incubée pendant un temps long.

Ces courbes sont appelées courbes de Cot. Une courbe lisse est l'indice d'une renaturation progressive et régulière de l'ADN. C'est ce qui est observé pour des virus, des bactéries et pour des cellules de levure. Par contre lorsque l'on prend de l'ADN d'organismes multicellulaires (*Homo sapiens*), l'aspect est totalement différent. La courbe observée semble le résultat de la mise à bout de trois courbes. Ce qui indique que dans le premier cas il y avait un seul type d'ADN alors que dans le second trois types d'ADN différents doivent coexister : le premier type se renature très rapidement, le second se renature nettement plus lentement, quant au troisième il ne se renature que très lentement.



ADN hautement répété : ADN non codant des centromères et télomères

ADN moyennement répété : rétrotransposons, gènes codant pour les ARNr et ARNt

ADN non répété : essentiellement ADN codant

Bilan : les différences de taille entre génomes ne reflètent pas le contenu en gènes : elles sont essentiellement dues à la présence de séquences répétées.

Pour l'Homme :

- **50% du génome se réhybride très lentement : cela correspond à du génome non ou très peu répété**
 - **ADN fonctionnel** : gènes protéiques, parfois au sein de familles multigéniques (ex gènes de globines ou d'opsines, gènes homéotiques...)
 - **ADN sans fonction connue** (non informatif ?) : Introns ou ADN espaceur
- **45% du génome se réhybride plus rapidement : séquences moyennement répétées**

Cet ADN correspond à deux types de séquences : codantes ou non codantes

- **ADN codant** : gènes très répétés dont la cellule a besoin en grande quantité comme les histones (gènes disposés en quintet et répétés en tandem 10 ou 20 fois), gènes codant pour les ARNt, gènes ARNr (250gènes).
- **ADN non codant et sans fonction connue** : la plupart sont des éléments **génétiques mobiles** qui se déplacent par copier/coller ou couper/coller et s'insèrent à n'importe quel endroit du génome en emportant parfois avec eux une séquence voisine (un gène entier, une séquence régulatrice ou simplement un exon, qui va alors, être répété avec l'élément génétique mobile).

Ces **éléments génétiques mobiles** représentent près de la **moitié du génome humain**. Il contiennent toujours :

- **des séquences répétées** sur leurs bordures ce qui permet aux algorithmes d'annotations du génome de les rechercher (ex les rétrovirus sont bordés par des séquences LTR : Long Terminal Repeat)
- des séquences nécessaires à leur transposition (séquence codant pour une **transposase**).

Il en existe différents types :

- **les transposons** à ADN : séquences d'ADN mobiles qui se déplacent dans le génome par couper coller (excision intégration) ou copier coller (copie puis intégration).
- **les rétrotransposons** : transcrits en ARN puis **rétro-transcrits** grâce à une **transcriptase inverse** en ADN mono puis double brin et se réinsérant n'importe où dans le génome grâce à une intégrase ce qui peut être à l'origine de cancers ou de maladies génétiques.
- **les rétrovirus endogènes** : anciens rétrovirus ayant infecté nos génomes et dont nous conservons la trace.

- **5% du génome se réhybride très rapidement : séquences très répétées souvent en tandem**

Cet ADN correspond à deux types de séquences : ADN fonctionnel ou non fonctionnel.

ADN fonctionnel :

- **télomères** : chez l'homme ce sont des répétition TTAGGG sur une longueur de 3 à 20kb (protège le chromosome de la digestion par des exonucléases, permet l'attachement à la membrane nucléaire...)
- **centromère** (chez l'homme ADN α -satellite = 171 pb répété, chez kangourou satellite = 50% du génome)
- îlots CpG de contrôle de la compaction de l'ADN par méthylation (épigénétique)

ADN sans fonction connue : parfois à l'origine de certaines maladies

Ces petites séquences d'ADN répété (**microsatellites** ou **minisatellites**) sont dues à des rippages de l'ADN pol qui bégaie lorsqu'elle doit polymériser une séquence répétée courte (et qui rajoute ou enlève encore certaines répétitions). Le nombre de répétitions est donc très variable : 2 individus n'en ont jamais exactement le même nombre, c'est pourquoi elles sont utilisées pour les empreintes génétiques.

Figure 14 : les gènes eucaryotes sont morcelés

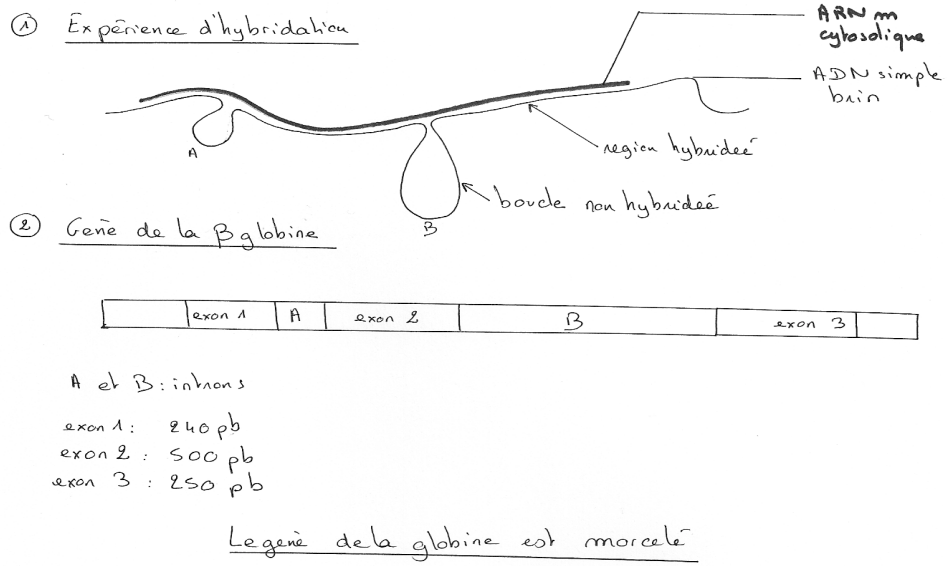


Figure 15 : 3 virus au programme
(Dunod 2021)

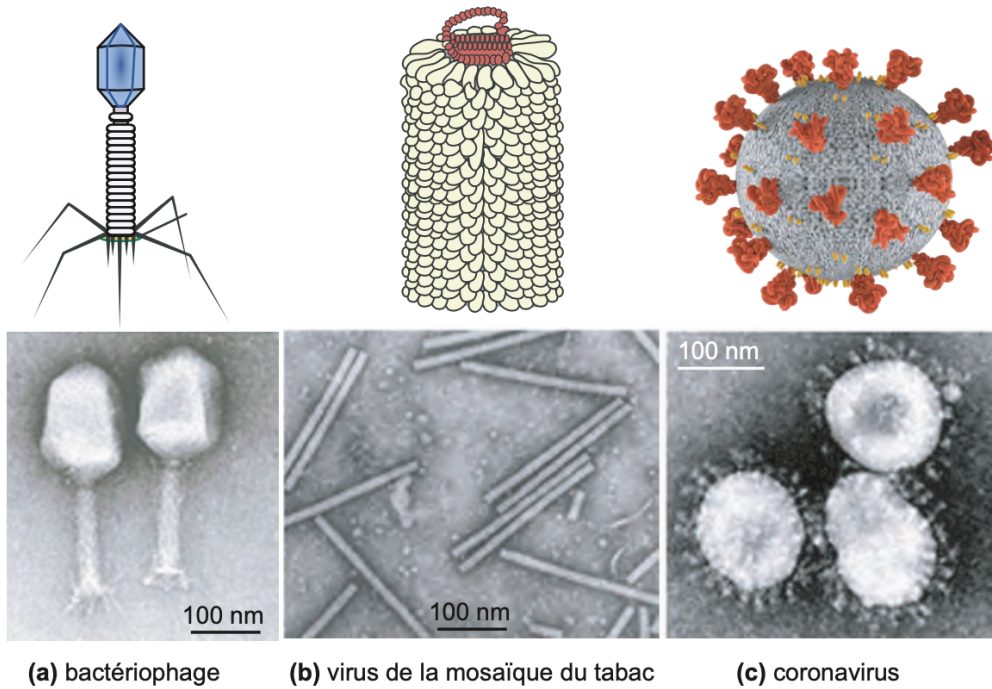
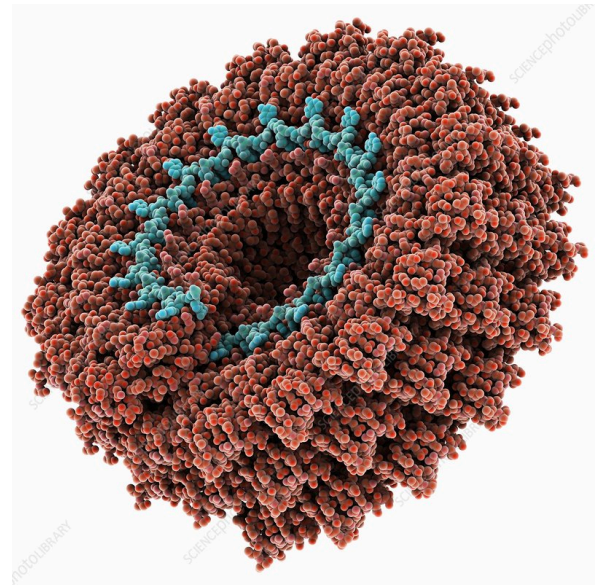


Figure 13.7 Les différents types de virus.

- (a) Phage HY01 (MET). Source : Lee H, Ku H-J, Lee D-H, Kim Y-T, Shin H, Ryu S, et al. (2016) « Characterization and Genomic Study of the Novel Bacteriophage HY01 Infecting Both Escherichia coli O157:H7 and Shigella Flexneri ».
- (b) Source :1994 Rothamsted Experimental Station <https://www.rothamsted.ac.uk/our-science>
- (c) Source : CDC/Dr. Fred Murphy Centers for Disease Control and Prevention’s Public Health Image Library, #4814.

Figure 16 : cycle du VMT (hors programme)
 (/www.sciencephoto.com, <http://ephytia.inra.fr/>)



Cycle du VMT dans une cellule
du mésophylle

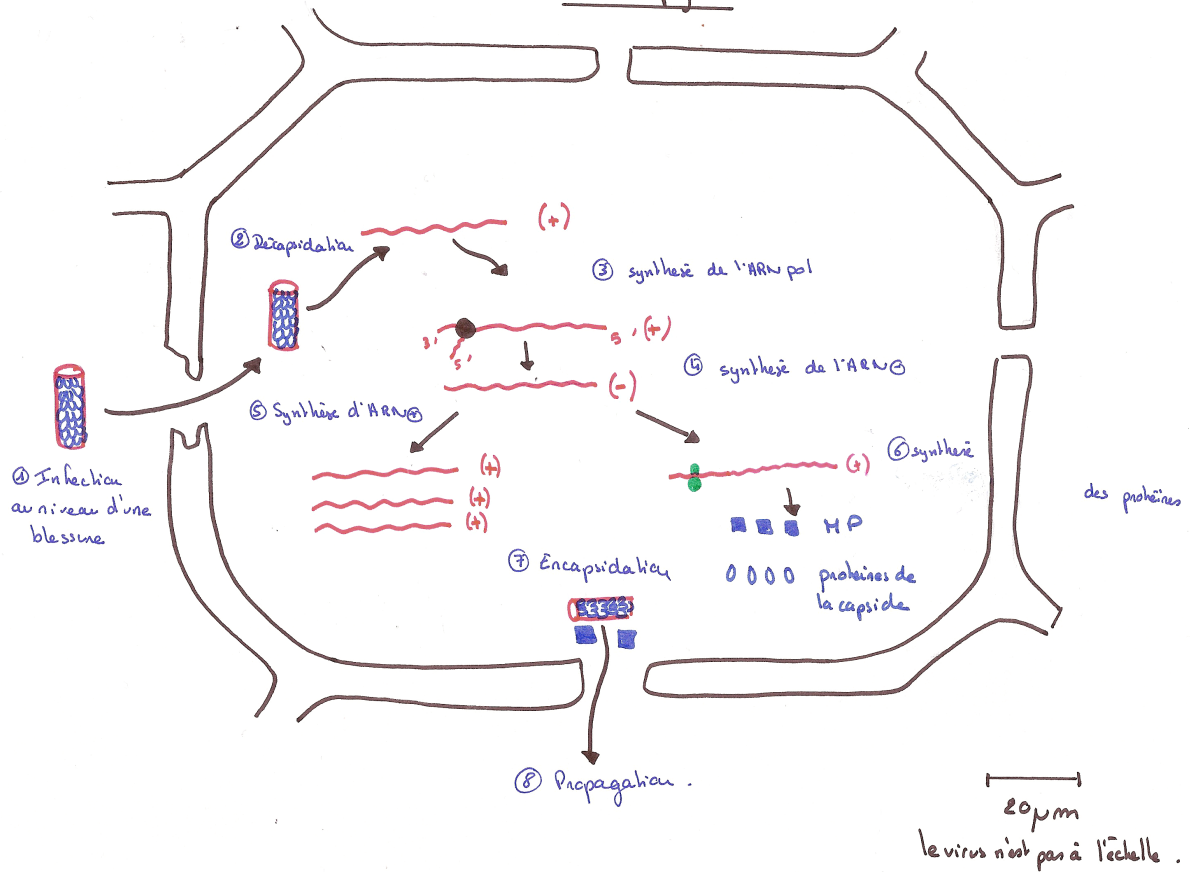
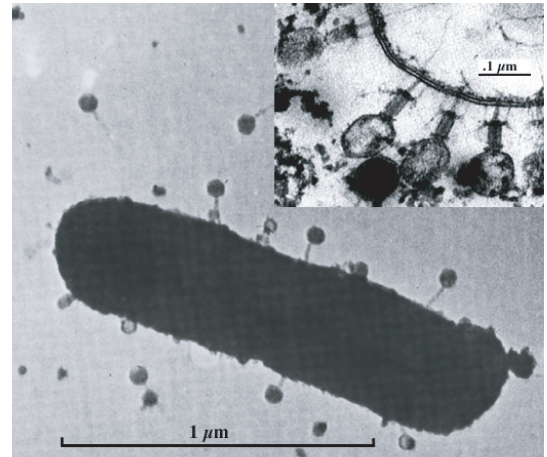
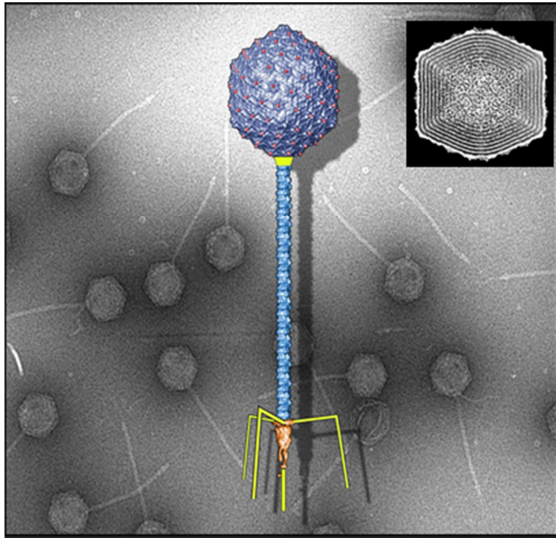


Figure 17 : cycle du bactériophage lambda
(Preux, 2013)



Fixation du phage lambda sur E. coli

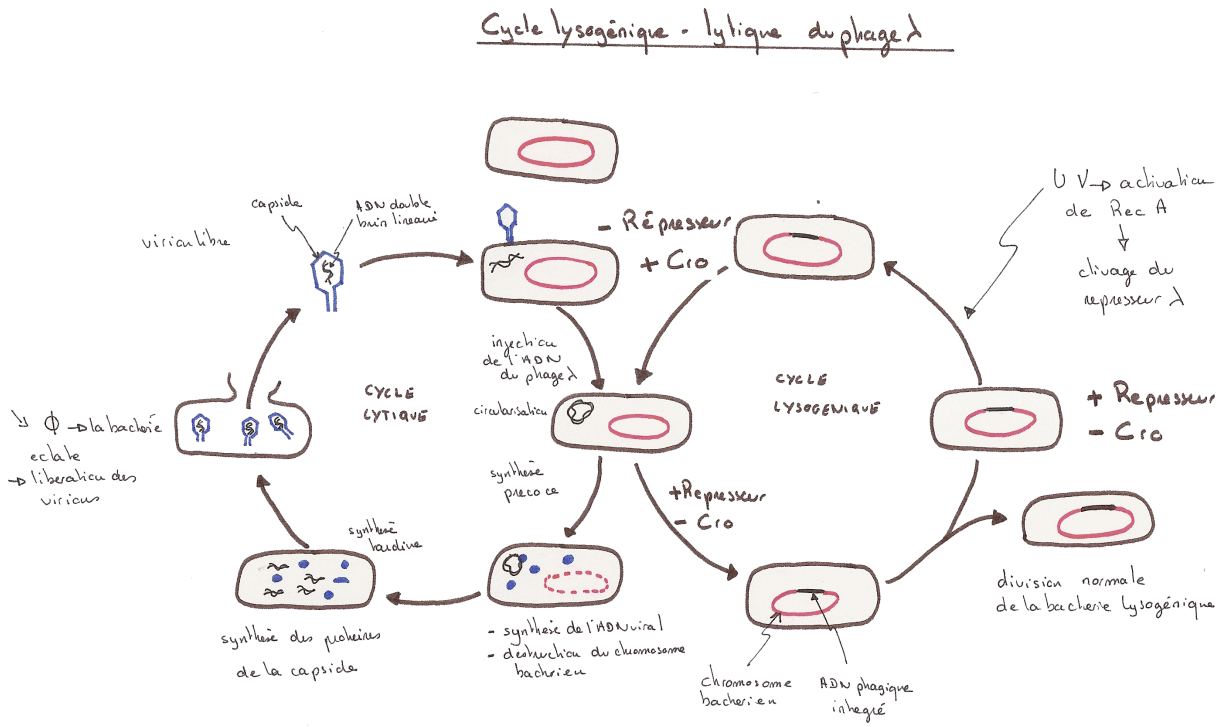
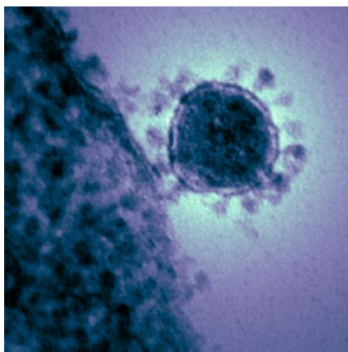
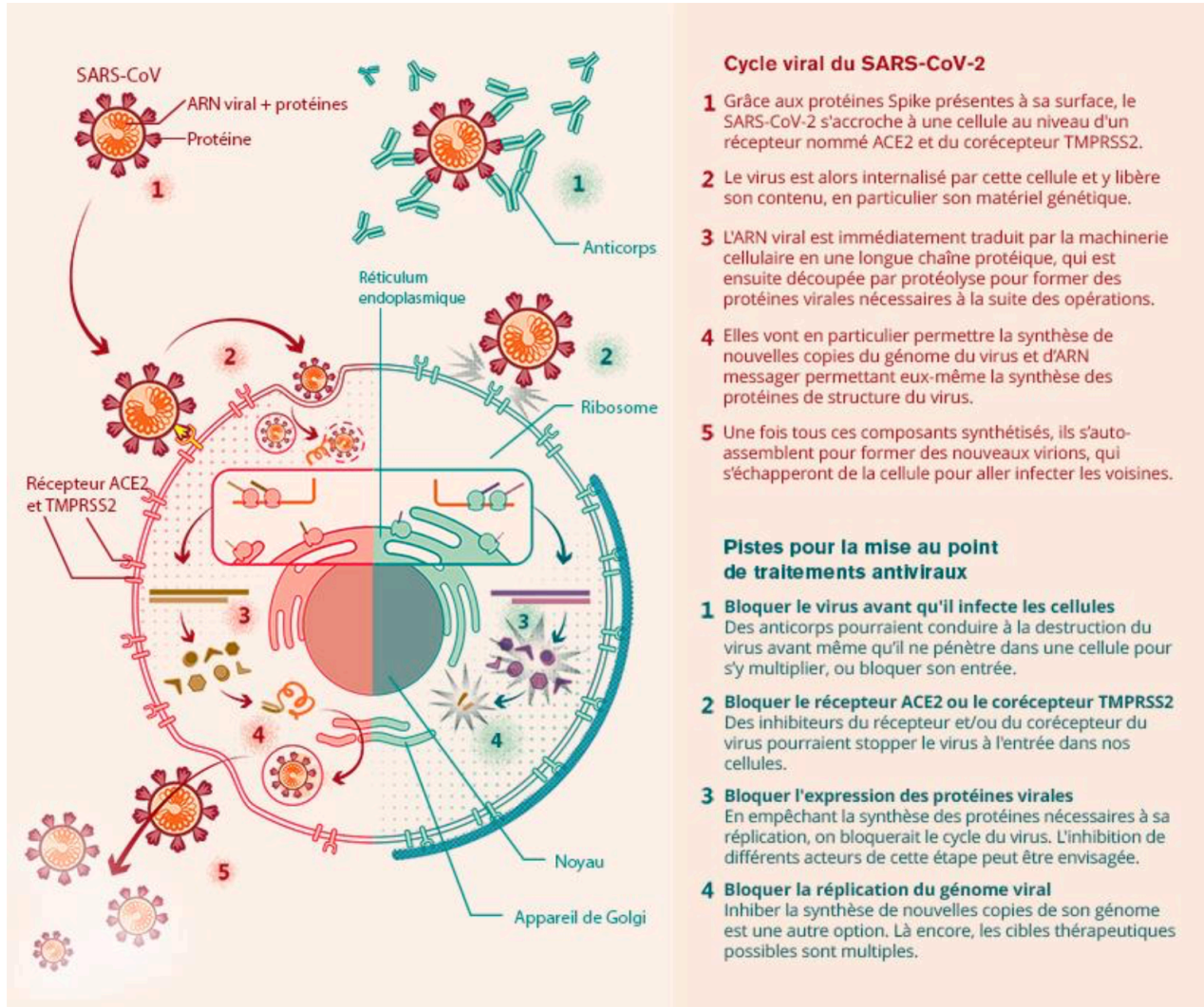


Figure 18 : cycle du SARS-CoV-2
(Heliot, 2021)



microscopie électronique à transmission colorée du MERS-CoV

A

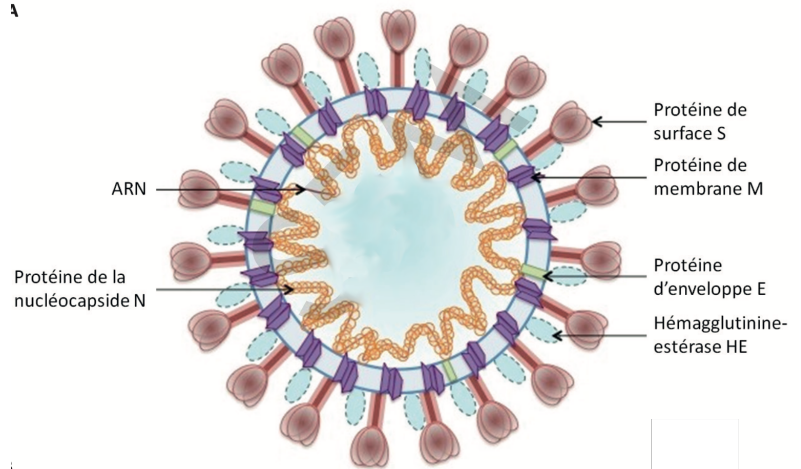
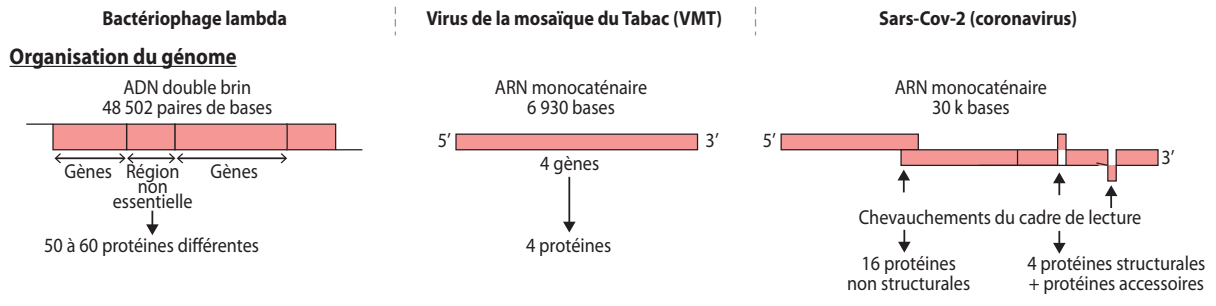


Figure 19 : organisation des génomes des 3 virus
(Memento, Dunod, 2021)



ORF1a et 1b chevauchantes dans le génome du SARS-CoV-2 (Vabret et Gouilh)

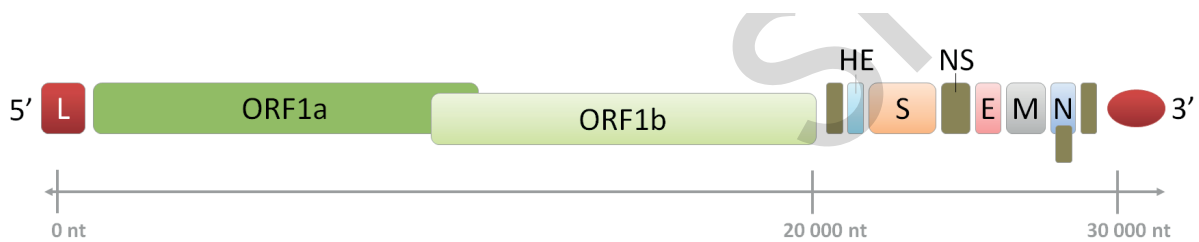


Figure 4 : Représentation schématique d'un génome de *Betacoronavirus* de clade A (HCoV-OC43)
Le génome du HCoV-OC43 comporte 31 728 nucléotides (nt). Les extrémités 5' (L = séquence *leader*, rectangle rouge) et 3' (queue polyA, cercle rouge) sont non codantes. Les 2 premiers tiers du génome sont constitués de 2 ORF chevauchantes, ORF1a et ORF1b, codant le complexe de réplication / transcription. Les gènes codant les protéines de structure sont toujours dans le même ordre : HE/S/E/M/N. Les ORF codant les protéines non structurales (en marron clair sur le schéma) sont en nombre et en position variables selon les espèces de coronavirus.

Figure 20 : une diversité de structure adaptée aux modalités d'infection des hôtes
(Memento, Dunod, 2021)

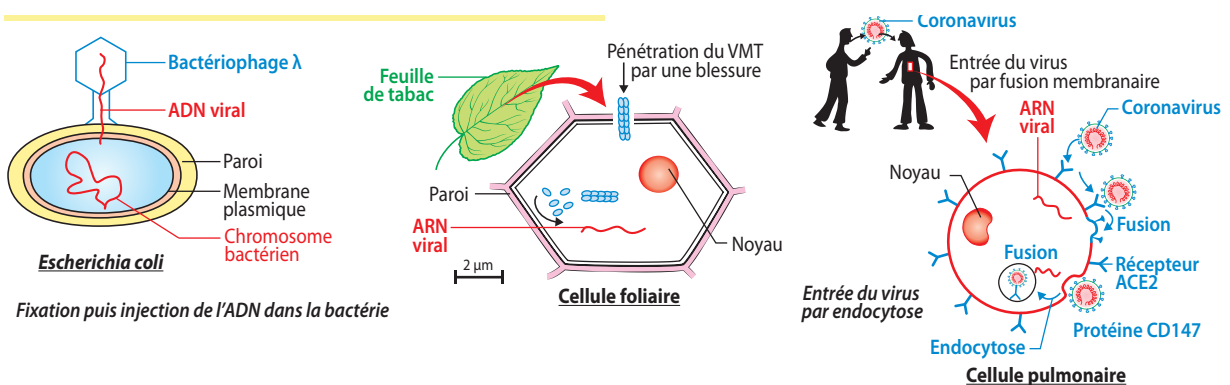
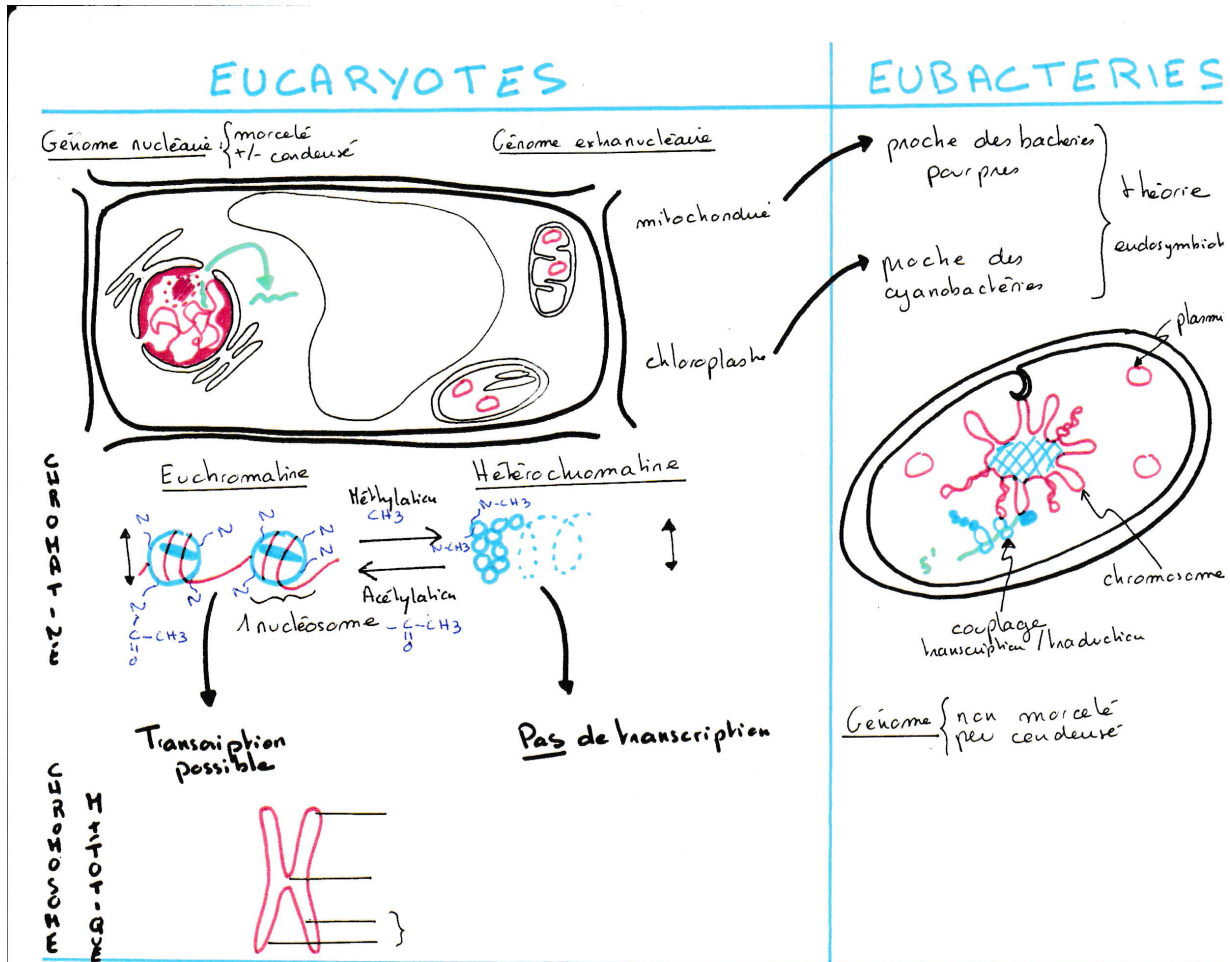


Schéma Bilan sur les géomes eucaryotes et procaryotes



Homme: - ph ↔ m / 4
 - géome
 - géome

E. coli: - densité de gènes
 - gènes non

ORGANISATION DES SEQUENCES CODANTES

- séquences uniques

- gènes en tandem → → →

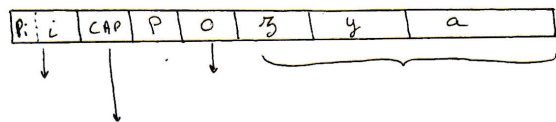
- familles multigéniques

↓

↓

↓

- opérons: exemple



ORGANISATION DES SEQUENCES NON CODANTES

- transposons

- séquences satellites (centromères/télocènes)

Très rare (promoteurs / opérateurs...)

Schéma Bilan sur les 3 virus du programme

